January 2012

1-deoxy-D-xylulose-5-phosphate Synthase (DXS) Mechanistic Study and its Implication in the Development of Novel Antibiotics and Antimalarials

Sumit Handa
University of South Florida, shanda@mail.usf.edu

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the American Studies Commons, and the Biochemistry Commons

Scholar Commons Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
1-deoxy-D-xylulose-5-phosphate Synthase (DXS) Mechanistic Study and its Implication in the Development of Novel Antibiotics and Antimalarials

By

Sumit Handa

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Department of Chemistry
College of Arts and Sciences
University of South Florida

Major Professor: David J. Merkler, Ph.D.
Wayne C. Guida, Ph.D.
John H. Adams, Ph.D.
Randy W. Larsen, Ph.D.
Jon C. Antilla, Ph.D.

Date of Approval:
April 04, 2012

Keywords: non mevalonate pathway, D. radiodurans, P. vivax, 1-deoxy-D-xylulose-5-phosphate Synthase.

Copyright © 2012, Sumit Handa
Dedication

I want to dedicate this thesis to my parents Mr. Vinod Kumar Handa and Mrs. Renu Handa. I would also like to thank my sister Mrs. Bhawna Suri and my fiancée Ms. Nupur Agrawal for the incredible support over the years.
Acknowledgment

I would like to acknowledge Dr. David J. Merkler for his continuous support and trust in me. It’s his guidance and encouragement that helped me during my graduate work. I would also like to thank my committee member Dr. John H. Adams, Dr. Wayne C. Guida, Dr. Randy W. Larsen and Dr. Jon C. Antilla for their support and supervision over the years.

During my stay here as graduate student I met so many fine scientists and I am heartily thankful to them for all the suggestions and the driving force in field of science and life in general. I would also like to once again thank Dr. John Adams for letting me work in his lab and his group members, Dr. Naresh Singh, Mr. Christopher Campbell and especially Dr. Francis Ntumngia for introducing me into the field of molecular biology. I would also like to acknowledge the support provided by Mrs. Divya Ramamoorthy from Dr. Wayne Guida’s lab.

I want to thank the entire previous and current lab members from Dr. Merkler’s lab: Dr. Emma Farrell, Dr. Will Lowe, Jacob Shafer, Milena Ivkovic, Shikha Mahajan, Kristen Jeffries, Matthew Battistini, Perry Mitchell. I want to especially acknowledge Daniel Dempsey for all the support and helpful discussion in science. I also want to thank wonderful undergraduate students Tyler Spradling and Nancy Cook. I also extend my acknowledgment to the Florida Center of Excellence for Biomolecular Identification and
Targeted therapeutics (FCoE-BITT) for the Graduate Multidisciplinary Scholar (GMS) award.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>i</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>Abstract</td>
<td>x</td>
</tr>
<tr>
<td>Chapter One: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Isoprenoids biological relevance and its characterization</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Mevalonate pathway for isoprenoid synthesis</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Non-mevalonate pathway for isoprenoid synthesis</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Non-mevalonate pathway as potential drug target</td>
<td>18</td>
</tr>
<tr>
<td>1.5 References</td>
<td>25</td>
</tr>
<tr>
<td>Chapter Two: 1-deoxy-D-xylulose-5-phosphate synthase (DXS) Reaction Mechanism</td>
<td>30</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>30</td>
</tr>
<tr>
<td>2.1.1 Isoprenoids biosynthesis through MVA and NMVA</td>
<td>30</td>
</tr>
<tr>
<td>2.1.2 Targeting NMVA pathway</td>
<td>32</td>
</tr>
<tr>
<td>2.1.3 Biochemistry of NMVA pathway</td>
<td>33</td>
</tr>
<tr>
<td>2.1.4 DXS a rate limiting enzyme in NMVA pathway</td>
<td>34</td>
</tr>
<tr>
<td>2.1.5 DXS reaction mechanism</td>
<td>37</td>
</tr>
<tr>
<td>2.2 Materials and Methods</td>
<td>39</td>
</tr>
<tr>
<td>2.2.1 Materials</td>
<td>39</td>
</tr>
<tr>
<td>2.2.2 Cloning of recombinant D. rad DXS</td>
<td>39</td>
</tr>
<tr>
<td>2.2.3 Production of DXS mutants using site-directed mutagenesis</td>
<td>40</td>
</tr>
<tr>
<td>2.2.4 Production of DXS mutants using overlap extension method</td>
<td>41</td>
</tr>
<tr>
<td>2.2.5 Overexpression and purification of DXS and the DXS mutants</td>
<td>42</td>
</tr>
<tr>
<td>2.2.6 Overexpression and purification of DXR</td>
<td>44</td>
</tr>
<tr>
<td>2.2.7 Determination of D-GLP and DXP concentration</td>
<td>44</td>
</tr>
<tr>
<td>2.2.8 DXS-DXR coupled assay</td>
<td>45</td>
</tr>
<tr>
<td>2.2.9 Product inhibition of DXS by DXP</td>
<td>46</td>
</tr>
<tr>
<td>2.2.10 Inhibition of DXS by Fluoropyruvate (F-Pyr)</td>
<td>47</td>
</tr>
<tr>
<td>2.2.11 Analyzing the initial velocity kinetic data</td>
<td>48</td>
</tr>
<tr>
<td>2.3 Results and Discussion</td>
<td>49</td>
</tr>
<tr>
<td>2.3.1 Overexpression, purification of DXS and coupled assay</td>
<td>49</td>
</tr>
<tr>
<td>2.3.2 Determination of the kinetic parameters for DXS-WT and the mutant DXS proteins</td>
<td>50</td>
</tr>
</tbody>
</table>
2.3.3 Pyruvate and D-GLP binding study
2.4 Summary
2.5 References

Chapter Three: Overexpression of \textit{P. vivax} DXS in \textit{E. coli}

3.1 Introduction
3.1.1 Malaria life cycle
3.1.2 Compartmentalization of metabolic pathways in \textit{Plasmodium} species
3.1.3 Targeting non-mevalonate pathway in \textit{Plasmodium} species

3.2 Materials and Methods
3.2.1 Materials
3.2.2 Cloning of recombinant \textit{P. vivax} DXS from \textit{P. vivax} genomic DNA
3.2.3 Overexpression and purification of \textit{P. vivax} DXS
3.2.4 Cloning of recombinant \textit{P. vivax} DXS and constructs without signal and transit peptide from codon optimized \textit{P. vivax} dxs gene
3.2.5 Overexpression of full length and truncated \textit{P. vivax} DXS
3.2.6 Ni-NTA Affinity chromatography purification
3.2.7 Gel filtration chromatography purification of SH-2 construct
3.2.8 Determination of D-GLP concentration
3.2.9 Preparation of Apo-DXS SH-2
3.2.10 DXS-DXR coupled assay
3.2.11 DXS-HPLC assay
3.2.12 Equation for analyzing the kinetic data
3.2.13 Inductively coupled mass spectrometry (ICP-MS) for \textit{Mg}^{2+} estimation

3.3 Results and Discussion
3.3.1 Cloning and expression of full length \textit{P. vivax} DXS using genomic DNA
3.3.2 Cloning and expression of full length \textit{P. vivax} DXS using codon optimized DNA
3.3.3 Cloning and expression of truncated \textit{P. vivax} DXS
3.3.4 Purification of \textit{P. vivax} DXS SH-2
3.3.5 Characterization of \textit{P. vivax} DXS SH-2

3.4 Summary
3.5 References

Chapter Four: Conclusion and Future Directions
4.1 Conclusion
4.2 Future directions
4.2.1 Lead compound for \textit{P. vivax} DXS inhibition
4.2.2 Crystal structure of \textit{P. vivax} DXS

4.3 References
Appendices
Appendix 1: Abbreviations and Figures 129
Appendix 2: Production of the Catalytic Core of Human Peptidylglycine α-Hydroxylating Monooxygenase (hPHMcc) in *E. coli* 144
List of Tables

Table 1.1 Examples of some biologically important isoprenoids ........................................... 3

Table 1.2 Distribution of metabolic enzymes involved in the biosynthesis of isoprenoids ....... 21

Table 2.1 PCR primers used to create the DXS mutants ...................................................... 42

Table 2.2 Steady-state Kinetic Parameters for DXS WT and DXS Mutants for D-GLP ...... 57

Table 2.3 Steady-state Kinetic Parameters for DXS WT and DXS Mutants for pyruvate .... 60

Table 2.4 Comparison of the Activity of the Mutant DXS Protein relative to the DXS WT .... 61

Table 2.5 Observed inhibition pattern and inhibition constant for DXS with DXP and fluoropyruvate ................................................................. 67

Table 3.1 Example of apicoplast targeting proteins ................................................................. 82

Table 3.2 Primer used for the DXS constructs ........................................................................ 89

Table 3.3 Predicted cleavage site between signal and transit peptide .................................. 101

Table 3.4 N-terminal amino acid truncated for SH-2 to SH-9 ............................................. 104
Table 3.5 Comparison of steady state kinetic constants for pyruvate 114

Table 3.6 Comparison of steady state kinetic constants for D-GLP 115
List of Figures

Figure 1.1 The Isoprenoids Precursors: Isopentyl Pyrophosphate (A), Dimethylallyl Pyrophosphate (B) 1

Figure 1.2 The Mevalonate-dependent production of IPP and DMAPP 7

Figure 1.3 Labeling pattern of D-GLP, Pyruvate and Acetyl-CoenzymeA 9

Figure 1.4 Reaction catalyzed by DXS 12

Figure 1.5 Reaction catalyzed by DXR 13

Figure 1.6 Reaction catalyzed by IspD 14

Figure 1.7 Reaction catalyzed by IspE 14

Figure 1.8 Reaction catalyzed by IspF 15

Figure 1.9 Reaction catalyzed by IspG 16

Figure 1.10 Reaction catalyzed by IspH 17

Figure 2.1 The Non-Mevalonate Dependent Biosynthesis of IPP and DMAPP 34

Figure 2.2 Crystal structure of DXS from *D. radiodurans* 36

Figure 2.3 Proposed mechanisms for the reaction catalyzed by DXS 38

Figure 2.4 SDS-PAGE, Lane-1 Marker, Lane-2 *D. rad* DXS, and Lane-3 *E. coli* DXR 50

Figure 2.5 Schematic drawing of the *D. rad* DXS depicting bound TPP with domain I (yellow), II (red), and III (blue) 51

Figure 2.6 Sequence comparison of DXS from *H. influenza, V. colerae,*
$P. \text{vivax}, P. \text{falciparum}, D. \text{radiodurans}, B. \text{anthracis}$ and $M. \text{tuberculosis}$ using Clustal W2

Figure 2.7 The Proposed Reaction Mechanism of DXS

Figure 2.8 Stereo view of the DXS active site depicting active C-2 of the thiazolium ring

Figure 2.9 Double reciprocal analysis of initial velocities of pyruvate at different fixed concentration of D-GLP

Figure 2.10 Double reciprocal analysis of initial velocities of D-GLP at different fixed concentration of pyruvate

Figure 2.11 Double reciprocal analysis of initial velocities of DXS Inhibition by DXP at different initial concentration of D-GLP

Figure 2.12 Double reciprocal analysis of initial velocities of DXS- Inhibition by DXP at different fixed concentration of pyruvate

Figure 2.13 Double reciprocal analysis of initial velocities of DXS Inhibition by F-Pyr at different fixed concentration of pyruvate

Figure 2.14 Double reciprocal analysis of initial velocities of DXS Inhibition by F-Pyr at different fixed concentration of D-GLP

Figure 2.15 Random sequential pathway proposed for the reaction Catalyzed by DXS

Figure 3.1 Life cycle of Malaria parasite in Anopheles mosquito and humans

Figure 3.2 Schematic diagram of apicoplast targeting protein in plasmodium species
Figure 3.3  (A) *P. vivax* dxs gene encoded on chromosome-12
(highlighted in yellow color), (B) DNA gel showing-
Marker (lane 1) and *P. vivax* dxs gene isolated
using genomic DNA as template (lane 2)

Figure 3.4  SDS-PAGE analysis of *P. vivax* DXS in *E. coli*. MW
standard (lane 1), cell lysate (lane 2), Flow through
from Ni-NTA resin (lane 3), binding buffer wash (lane 4),
wash buffer wash (lane 5), and purified *P. vivax* DXS (lane 6)

Figure 3.5  pUC 57 vector with codon optimized *P. vivax* dxs gene

Figure 3.6  SDS-PAGE analysis of *P. vivax* DXS SH-1 in *E. coli*. MW
standard (lane 1), cell lysate (lane 2), Flow through from
Ni-NTA resin (lane 3), binding buffer wash (lane 4), wash
buffer wash (lane 5), and Elution *P. vivax* DXS (lane 6-9)

Figure 3.7  N-terminal Sequence comparison of DXS from *P. berghei*,
*P. yoelii*, *P. falciparum*, *P. vivax*, and *P. knowlesi*

Figure 3.8  Clustal W2 alignment of *E. coli*, *D. rad.* and *P. vivax* DXS

Figure 3.9  DNA gel showing cloned vector digested with *NdeI* and
*XhoI* restriction enzymes

Figure 3.10 SDS-PAGE analysis of *P. vivax* DXS SH-2 affinity purification

Figure 3.11 Gel filtration chromatography of *P. vivax* DXS SH-2 on
Sephadex G-200

Figure 3.12 Determination of the molecular mass of *P. vivax* DXS SH-2
by gel filtration chromatography
Figure 3.13 (A) SDS-PAGE analysis of *P. vivax* DXS SH-2: MW standard (lane 1), 10 µg of affinity purified enzyme (lane 2), and 10 µg of gel filtration purified enzyme. (B) Western blot analysis of *P. vivax* DXS SH-2: 1 µg of gel filtration purified enzyme (lane 1), 1 µg of affinity purified enzyme (lane 2), and MW standard (lane 3) 111

Figure 3.14 pH optimum determination of *P. vivax* DXS SH-2 using HPLC assay 112

Figure 3.15 Steady state kinetics of *P. vivax* DXS SH-2 upon reconstitution with TPP 113
Abstract

Isoprenoids are the largest family of biologically active compounds, synthesized by five carbon subunits namely isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). For long time the mevalonate-dependent (MVA) pathway has been considered as the sole source of IPP and DMAPP, until recently a new non-mevalonate dependent (NMVA) pathway was discovered. This new pathway utilizes entirely different set of enzymes for isoprenoids synthesis and don’t have any homologues in humans. NMVA pathway is the only source of isoprenoids for certain eubacteria, parasite and plants. Absence of the NMVA pathway in higher organisms has opened a new platform for the development of novel antibiotics and antimalarials.

1-deoxy-D-xylulose-5-phosphate synthase (DXS), the first enzyme in NMVA pathway has been reported as the rate limiting enzyme in the synthesis of IPP and DMAPP and has been the center of interest for inhibitor development. Reaction mechanism of thiamine pyrophosphate (TPP) and Mg\(^{2+}\) dependent DXS enzyme has been studied in this report. Using steady state kinetics analysis, product inhibition and dead end inhibitor, the mechanism of substrate (pyruvate and D-glyceraldehyde-3-phosphate) addition was studied. Due to different domain organization in DXS as compared to the
other TPP dependent enzyme, the mechanism of addition was found to be random sequential rather than ping-pong mechanism.

Based on bioinformatics tool and in vitro studies it has been established that NMVA exists in all the plasmodium species, thus making the enzymes involved in NMVA as an alluring target for new antimalarial drugs. All the plasmodium species and other member of the phylum apicomplexa harbor apicoplast an organelle which is homologous to the chloroplast of plants and algae. All the enzymes from NMVA pathway translocate to apicoplast from nucleus through a secretory pathway using signaling and transit peptide. In this study DXS from \textit{P. vivax} has been cloned and expressed in \textit{E. coli} using genomic DNA and codon optimized synthetic DNA as a source. Expression of full length DXS with signal and transit peptide as well as mature protein without these peptide using serial deletion has been studied. Kinetic parameters of \textit{P.vivax} DXS have been calculated and found to be comparable to the DXS from other species.
Chapter 1: Introduction

1. Introduction

1.1 Isoprenoids biological relevance and its characterization

Isoprenoids (or terpenoids) belong to the largest family of small and complex biologically active compounds, widespread throughout the living system. Isoprene a C$_5$ carbon skeleton (methylbuta-1, 3-diene) that is the precursor of all the isoprenoids through cyclization, rearrangement, and oxidation. They have diverse biological roles, functioning as pigments, plant hormones, vitamins (A, D, E and K), sterols, carotenoids, chlorophyll, components of the cell membrane and intermediate in the biosynthesis of glycoproteins and glycolipids (Sacchettini and Poulter 1997). Isoprene itself is not used for the production of isoprenoids in vivo; instead, the biological forms of isoprene are two branched phosphorylated five carbon precursors namely isopentyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Figure 1.1).

Figure 1.1: The Isoprenoids Precursors: Isopentyl Pyrophosphate (A), Dimethylallyl Pyrophosphate (B).
The library of known isoprenoids seems to increase daily as we expand our knowledge of natural products. List of some of the important isopenoids and their role is listed below:

Table 1.1: Examples of some biologically important isoprenoids.

<table>
<thead>
<tr>
<th>Isoprenoids</th>
<th>Importance or Role</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Geranyl pyrophosphate" /></td>
<td>Precursor of geraniol required for terpenes biosynthesis.</td>
</tr>
<tr>
<td><img src="image2" alt="Farnesyl pyrophosphate" /></td>
<td>Intermediate in the HMG-CoA reductase pathway used for the biosynthesis of terpenes, sterols.</td>
</tr>
<tr>
<td><img src="image3" alt="Cholesterol" /></td>
<td>Required for cell membrane fluidity and precursor of bile acids, steroid hormones and vitamin D.</td>
</tr>
<tr>
<td><img src="image4" alt="Vitamin E" /></td>
<td>Antioxidant to reduce the production of reaction oxygen species formation during lipid oxidation.</td>
</tr>
<tr>
<td>Chemical Structure</td>
<td>Description</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><img src="image" alt="Ubiquinone" /></td>
<td>Part of the electron transport chain for cellular respiration.</td>
</tr>
<tr>
<td><img src="image" alt="Abscisic acid" /></td>
<td>Plant hormone required for development processes.</td>
</tr>
<tr>
<td><img src="image" alt="Retinol" /></td>
<td>Form of vitamin A, required for vision.</td>
</tr>
<tr>
<td><img src="image" alt="Limonene" /></td>
<td>Cyclic terpene used in cosmetic industry for citrus flavor.</td>
</tr>
</tbody>
</table>

Isoprenoids has been classified under six major categories:

1. **Monoterpenoids**: They contain two isoprene units linked in a head to tail fashion. These are the major component of the fragrant oil from leaves, flowers, and fruits (e.g. limonene and nerol)

2. **Sesquiterpenoids**: These terpenoids consists of three isoprene units to form 15 carbon cyclic and acyclic compounds (e.g. Farnesol, Abscisic acid)
3. Diterpenoids: These compounds comprised of 20 carbon atom derived from geranyl geranoil pyrophosphate (a 10 carbon unit). Some of the important and well characterized diterpenes include retinol (vitamin A), the phytohormone gibberllin, tetrahydrocannabinol.

4. Sesterterpenoids: These are derived from 25 carbon framework geranyl-farnesol pyrophosphate (a 10 and 15 carbon unit respectively). An example ophiobolin A, a fungal metabolite.

5. Triterpenoids: These are the largest group of natural isoprenoid consisting of six isoprenoids units and include sterols such as cholesterol and lanosterol. Squalene being the precursor of all triterpenoids.

6. Carotenoids: These consist of eight isoprenoid units to make \( \text{C}_{40} \) chains with conjugated double bonds. These function in photosynthesis because of their efficient light absorption properties and also prevent photo-oxidative damage in the cell. Examples include lycopene and \( \alpha-, \beta-, \text{and } \gamma- \text{-carotene.}
1.2 Mevalonate Pathway for isoprenoid synthesis

The mevalonate-dependent (MVA) pathway of isoprenoids biosynthesis has dominated research in this area for more than 5 decades as this pathway was considered as the only route to IPP and DMAPP in living systems. Interest in the MVA pathway has always been high due to its connection to cholesterol biosynthesis. Three Nobel Prizes have been awarded for understanding the structure, biosynthesis, and homeostasis of cholesterol: (A. Windaus in 1928, K. Bloch and F. Lynen in 1964, M. Brown and J. Goldstein in 1985). Pioneering work by K. Bloch and D. Rittenburg (1942) showed that cholesterol ring and side chains were labeled with deuterium when cells were feed with deuterium labeled acetate (Bloch and Rittenberg 1942), and in 1956, Folker and group showed that it’s the IPP unit which is the direct precursor of the cholesterol and not the acetate (Tavormina, Gibbs and Huff 1956). In late 1980’ M. Brown and J. Goldstein provided an elegant explanation of feedback regulation of cholesterol biosynthesis and its removal from plasma based in low density lipoprotein (LDL) receptor (Brown and Goldstein 1986).

Mevalonate is the most important intermediate in the bioconversion of acetyl-CoA to IPP and DMAPP (Figure 1.2). IPP and DMAPP, to a lesser extent, are then used for the biosynthesis of cholesterol and other isoprenoids. The first step of the MVA pathway involves the condensation of acetyl-CoA and acetoacetyl-CoA to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), a reaction catalyzed by HMG-CoA synthase (HMGS). The next step involves HMG-CoA reductase (HMGR) catalyzed irreversible
reduction of HMG-CoA to mevalonate, the first committed step in isoprenoid biosynthesis. Mevalonate is phosphorylated in two steps to yield mevalonate-5-diphosphate which is subsequently decarboxylated in an ATP-dependant step to yield IPP, phosphate, and CO₂. IPP is the exclusion product from mevalonate via the MVA pathway and is enzymatically converted to DMAPP by isopentenyl diphosphate isomerase.
Figure 1.2. The Mevalonate-Dependant Production of IPP and DMAPP (A) and the Subsequent Conversion of IPP and DMAPP to Cholesterol (B). The condensation of acetoacetyl-CoA and acetyl-CoA catalyzed by HMG-CoA synthase (HMGS) (enzyme 1). The NADPH-dependent reduction of HMG-CoA to mevalonate by HMGR (enzyme 2). Phosphorylation mevalonate at the 5-hydroxyl as catalyzed by mevalonate kinase (MK) (enzyme 3). Formation of the 5-pyrophosphate catalyzed by phosphomevalonate kinase (PMP) (enzyme 4). The reaction of pyrophosphomevalonate decarboxylase (DMPD) to form IPP (enzyme 5), and the isomerization of IPP to DMAPP catalyzed by isopentenyl pyrophosphate isomerase (IDI) (enzyme 6). For cholesterol synthesis (B): condensation of IPP and DMAPP yield geranyl pyrophosphate, which undergoes addition of an additional IPP to form the 15 carbon intermediate, farnesyl pyrophosphate. Reductive addition of two molecule of farnesyl pyrophosphate molecules leads squalene, which undergoes NADPH mediated reduction to form lanosterol, followed by a multistep transformation to cholesterol.

Squalene, a C30 intermediate in the cholesterol biosynthetic pathway, is ultimately synthesized from six molecules of IPP; included as an example of the further metabolic elaboration of IPP and DMAPP (Figure 1.2) (Karl 2004).

High serum cholesterol correlates to high blood pressure and a greater risk for cardiovascular disease and stroke (Grundy et al. 1998). One significant outcome from our understanding of the MVA pathway leading to cholesterol is the development of safe and effective drugs to lower serum cholesterol level by inhibiting the synthesis of MVA pathway (Endo 1992). This is accomplished by the therapeutical use of HMGR inhibitors, a class of drugs known as the statins. Early work from the Endo group led to the isolation of mevastatin from *Penicillium citrinum* as potent inhibitors of HMGR (Endo, Kuroda and Tanzawa 1976). Later, a group in Merck isolated lovastatin from *Aspergillus terreus*, structurally similar to mevastatin, but more a potent HMGR inhibitor (K_i = 1.0 nM) (Alberts et al. 1980). Currently there are several statins available in the market for
the treatment of high cholesterol and generate highest revenue as pharmaceutical drug
((atorvastatin (Lipitor), fluvastatin (Lescol), lovastatin (Mevacor), and pravastatin
(Pravachol)).

1.3 Non-Mevalonate Pathway for isoprenoid synthesis

Pioneering work by Rohmer and his coworker led to the discovery of the non-
mevalonate (NMVA) pathway for the biosynthesis of IPP and DMAPP derived
isoprenoids (Rohmer et al. 1993). They studied the incorporation of universally $^{13}$C
labeled glucose [U-$^{13}$C$_6$] in hapnoid of *Rhodopseudomonas palustris* and
*Rhodopseudomonas acidophila*, an equivalent of cholesterol for bacteria to maintains the
cell fluidity (Flesch and Rohmer 1988). They found that the labeling was not consistent
with MVA pathway, and it consists of three carbon atoms from [U-$^{13}$C$_6$] glucose rather
than two carbon atoms from acetate, a precursor of mevalonate (Figure 1.3).

Definitive proof of the involvement of pyruvate and glyceraldehyde-3-phosphate
(GLP) was done using $^{13}$C labeled pyruvate and $^{13}$C labeled glycerol (Rohmer et al.
1996), mutant of *E. coli* defective in triose phosphate metabolism were chosen to avoid
the inter-conversion of pyruvate and glycerol. When *E. coli* was incubated with $^{13}$C
labeled pyruvate and unlabeled glycerol, all the mutants were found to have labeled C$_2$
subunit of isoprenoid, while none of the $^{13}$C glycerol was incorporated in C$_2$ subunit of
isoprenoid skeleton when *E. coli* was culture in $^{13}$C labeled glycerol and unlabeled pyruvate. While, C$_3$ subunit of isoprenoid was only incorporated form $^{13}$C pyruvate with *E. coli* defective in glycerol kinase, glycerol phosphate dehydrogenase, triose phosphate isomerase mutants and from $^{13}$C glycerol with enolase, phosphoglycerate kinase and glyceraldehyde phosphate dehydrogenase defective mutants.

Figure 1.3: (A) Labeling pattern of Glyceraldehyde-3-phosphate (GLP), Pyruvate and Acetyl-CoenzymeA starting from labeled glucose during glycolysis and pyruvate dehydrogenase reaction, (B) labeling pattern in IPP synthesis by mevalonate pathway using acetyl-coA, (C) labeling pattern in IPP synthesis by non-mevalonate pathway using GLP and pyruvate.
In another attempt to prove that mevalonate is not the precursor for the isoprenoids in certain eubacteria and plants, *E. coli* was incubated with $[2^{-13}C]$, $[2,3^{-13}C_2]$ and $[2,4^{-13}C_2]$-d-deoxyxylulose and its integration into the ubiquinone suggested that an alternate pathway for exists for IPP synthesis (Putra et al. 1998). These experiments pointed GLP as the only precursor of the C3 subunit in isoprenoid skeleton.

There are more than 20,000 plants terpenes have been reported and the origin has always been associated with MVA pathway. While, study done by Arigoni and Schwarz in *Ginkgo biloba* (maidenhair tree) showed that higher plants have both MVA and NMVA pathways and separated by compartmentalization (Eisenreich et al. 1998). Higher plants have an unusual balance for the terpenoids synthesis through both MVA and NMVA pathways. It has been shown that sterols and ubiquinone is biosynthesized by MVA pathway, while variety of monoterpenes and diterpenes such as menthone, thymol, geraniol, THC, β-carotene, lutein, etc. are synthesized by NMVA pathway.

Assignment of MVA and NMVA pathways to bacteria, archaea, plants, fungi and animal was been done using genome mining. All the genes of MVA pathway were known before the whole genome sequence were published, and were later used as a template to assign the pathway utilized for isoprenoid synthesis. In case of NMVA pathway, the first two enzymes (DXS and DXR) sequence were known by the end of the last century and later used as a template to find the prospective orthologs in the genome database (Gräwert et al. 2011). Since 1990, NMVA pathway has been explored in to a
great extent and genome mining through search of open reading frame has leads to the elucidation of all the enzymes involved in this pathway. Plethora of information from this new mechanism of isoprenoid synthesis can be used to develop new family of antibiotics, herbicides, etc. Another great opportunity for the better life of mankind can be achieved by targeting *Plasmodium falciparum* parasite, responsible for Malaria. This pathway presents an analogy to the MVA pathway, where targeting the HMGR has leads to the development of cholesterol lowering drugs. All the seven enzymes involved in this pathway will be discussed in detail in here:

1. **1-deoxy-D-xylulose-5-phosphate synthase (DXS):** This enzyme catalyzes the thiamine and Mg$^{2+}$ dependent condensation of pyruvate and GLP to from 1-deoxy-D-xylulose-5-phosphate (Figure 1.4) (Lange et al. 1998, Eisenreich et al. 1998). DXS (EC 2.2.1.7) belongs to the family of enzymes which uses thiamine pyrophosphate (TPP) as a co-factor and shares a high homology with pyruvate dehydrogenase (EC 1.2.4.1), pyruvate decarboxylase (EC 4.1.1.1), and transketolase (EC 2.2.1.1). To understand the role of DXS as a rate limiting enzyme a transgenic *Arabidopsis thaliana* was constructed to over and under express this enzyme (Estévez et al. 2001). The effect of expression on chlorophyll, α-tocopherol, carotenoids, abscisic acid was studied. This study clearly showed that the transgenic *A. thaliana* expressing higher amount of DXS showed elevated level of isoprenoids under observation and reduced level for lower expressing DXS as compared to wild type, which suggest DXS as the committed and rate limiting enzyme in NMVA pathway.
2. **1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR):** This enzyme (EC 1.1.1.267) catalyzes the NADPH and Mg$^{2+}$/Mn$^{2+}$ dependent conversion of 1-deoxy-D-xylulose-5-phosphate (DXP) to 2C-Methyl-D-erythritol-4-phosphate (MEP) (Figure 1.5). This step involves the rearrangement and two electron reduction of linear carbon skeleton to branched chain isoprenoid precursor, shown by the heavy isotope labeling study (Reuter et al. 2002). DXR has been cloned, over expressed and crystallized from many organisms including *E. coli* (Steinbacher et al. 2003), *M. tuberculosis* (Henriksson et al. 2007), and *Z. mobilis* (Ricagno et al. 2004). Crystal structure of DXR shows that each subunit of the homodimer consists of N-terminal NADPH binding domain connected with active domain pocket. Great detail of mechanistic study has been performed for *E. coli* DXR (Fox and Poulter 2005, Koppisch et al. 2001), using fosmidomycin and dihydro-NADPH as a dead end inhibitor to show that DXR undergoes an ordered sequential mechanism with NADPH binds before DXP.
3. **4-Diphosphocytidyl-2C-methyl-D-erythritol 4-phosphate synthase (IspD):** IspD (EC 2.7.7.60) is the third enzyme in the NMVA pathway catalyzing the formation of 4-diphosphocytidyl-2-C-methylerythritol (CDP-ME) from MEP and cytidyl triphosphate (CTP) (Figure 1.6). Discovery of IspD represents an example of advancement in biochemical and comparative genomic analysis. In order to search for the subsequent enzyme that utilizes MEP as substrate, *E. coli* crude cell lysate was incubated with MEP led to the conversion to CDP-ME by reaction with CTP, a preferred nucleotide triphosphate over ATP, UTP and GTP (Richard et al. 2001).

The crystal structure of IspD from *E. coli* with substrate complex shows the coordination of α, β and γ phosphate of CTP with Mg$^{2+}$. The possible mechanism includes the formation of reactive cytidyl monophosphate (CMP) upon loss of pyrophosphate (ppi) and subsequent reaction with 4-phosphate group of MEP to form CDP-ME. A recent pulse chase experiment (Richard et al. 2004), using [2-14C] MEP and [2-14C] CTP, it was shown that CTP binding precedes to that of MEP, which points towards an ordered sequential mechanism.
Figure 1.6: Reaction catalyzed by IspD

4. **4-Diphosphocytidyl-2C-methyl-D-erythritol Kinase (IspE):** This enzyme (EC 2.7.1.148) catalyzes Mg$^{2+}$ dependent phosphorylation at 2-hydroxy position of CDP-ME using ATP as second substrate (Figure 1.7). IspE shares high homology with homoserine kinase, mevalonate kinase and phospho-mevalonate kinase (collectively called GHMP) (Miallau et al. 2003). Crystal structure of *E. coli* IspE with CDP-ME and non-hydrolysable from of ATP displayed a α/β fold characteristic of the GHMP kinase superfamily.

Figure 1.7: Reaction catalyzed by IspE

5. **2C-Methyl-D-erythritol 2,4 diphosphate synthase (IspF):** This enzyme (EC 4.6.1.12) catalyzes the conversion of 4-diphosphocytidyl-2C-methyl-D-erythritol 2-
phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate (cMEDP) (Figure 1.8) and release of cytidyl monophosphate (CMP) (Kemp, Bond and Hunter 2002). IspF from *Plasmodium falciparum* has been also shown to catalyze the conversion of CDP-ME or 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2-phospho-2C-methyl-D-erythritol 3,4-cyclic monophosphate (Figure 1.8) at lower rate (~10%) (Rohdich et al. 2001), and it has been associated to the structural perturbation of the catalytic domain. In another report *Mesorhizobium loti* has an open reading frame of IspD and IspF as a

![Figure 1.8: Reactions catalyzed by IspF](image)

Figure 1.8: Reactions catalyzed by IspF
fusion protein and IspE on other part of the chromosome (Testa et al. 2006). Incubation of the bi-functional IspDF with MEP, CTP and ATP only yield CDP-ME and not cMEDP. When the same constituent were incubated with IspE then there was a formation of cMEDP.

6. 2C-Methyl-D-erythritol 2, 4-cyclodiphosphate reductase (IspG): This enzyme (EC 1.17.7.1) catalyzes the conversion of cMEDP to 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (Figure 1.9). Discovery of this enzyme was confirmed by recombinant hyper-expression of DXS, DXR, IspD, IspE and IspF in E. coli (Hecht et al. 2001). E. coli strain grown in media with $^{13}$C labeled 1-deoxy-d-xylulose (DX) led to the accumulation of cyclic pyrophosphate precursor. Addition of IspG gene in this system leads the conversion of DX to 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate as characterized by NMR (Zepeck et al. 2005).

Figure 1.9: Reaction catalyzed by IspG
7. **1-Hydroxy-2-methyl-butenyl 4-diphosphate reductase (IspH):** This enzyme (EC 1.17.1.2) catalyzes the conversion of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate to IPP and DMAPP (Figure 1.10). IspH was discovered using the same strategy as that for IspG. *E. coli* strain bearing all the enzyme of the NMVA pathway converted the $^{13}$C labeled DX to IPP and DMAPP. IspH possesses a redox-active [4Fe-4S]$^{2+}$ cluster for radical mediated cleavage of C-OH bond (Rohdich, Bacher and Eisenreich 2004). Due to the sensitivity towards the oxygen (in solution and atmospheric) all the characterization step for IspH were performed under anaerobic conditions.

![Figure 1.10: Reaction catalyzed by IspH](image)

Figure 1.10: Reaction catalyzed by IspH
1.4 Non-Mevalonate Pathway as a potential drug target

Sequencing of human genome represents a milestone achievement by the scientific community. Since then the process of whole genome sequencing of various organisms has gone up by an astonishingly high rate. The technology in terms of genomics and software have advanced so much that now a day the whole genome from an individual takes only 1-2 weeks. There are commercial ventures (GenomeQuest, Complete genomics, etc.) (Roach et al. 2010), which supports the whole genome sequence with accuracy of 99.99% and much lower price ($1,800) as compared to the human genome project which costs around $4 billion. There are many aspects and benefits associated with genome sequencing and have been used for the better life of mankind. Early detection of genetic discrepancies, rational drug design, diagnosis of diseases, and gene therapy collectively termed as molecular medicine has revolutionized the field of biomedical research. Understanding anthropology and evolution through comparative genomics has led to the association of diseases and traits.

Molecular medicines have helped the human race to overcome various diseases caused by bacteria, parasites and yeast infection. But, lately we have been facing an emerging trend in resistance against antibiotics and anti-parasitic drugs. Several of the last resort antibiotics such as vancomycin have been shown to non-effective against Staphylococcus aureus, bacteria responsible for wound and surgical infection in hospitals (Leeb 2004).
All the major pharmaceutical companies have currently stopped working on the development of new antibiotics because of the lower percentage of profit generated to the investment. These companies have shifted to the development of drugs for high cholesterol, rheumatoid arthritis, depression, erectile dysfunction, etc. and all the new antibiotics in the development in just a mere modification of old antibiotics. Tropical diseases such as malaria, tuberculosis, leishmaniasis, chagas, etc. are the most neglected human health problem. Malaria being the most profound human problems with highest annual morbidity and several hundred million reported cases worldwide.

Malaria primarily affects people living below the poverty line and couldn’t afford medication. In addition to the high annual mortality, malaria causes dramatic economic loss, hardship due to the inability of acutely infected individuals to work and in children it can lead to life-long learning impairment (Breman, Egan and Keusch 2001). The widespread nature of *P. vivax* compared to *P. falciparum* and evidence of mounting resistance to current drugs in the plasmodium species makes it vital to discover and develop new drugs for these species. Out of the four species of the malaria parasite, the majority of morbidity is caused by *Plasmodium falciparum* and *Plasmodium vivax*. *P. falciparum* is the most studied species because it has the highest rate of mortality and its widespread resistance to anti-malarial drugs. Drugs like Chloroquine and Fansidar now suffer from worldwide resistance problem; this growing concern in the treatment of malaria is because of poor implementation of therapies, toxicity of existing therapeutics, and the small number of available therapeutics.
Molecular basis of antimalarial resistance has been cause primarily because of the mutation in drug target which reduces the sensitivity. Some drugs such as chloroquine and mefloquine don’t have any protein targets which can mutate, these classes of drugs become resistant through the mutation in transporters (Woodrow and Krishna 2006). Elevated level of transporters such as \textit{pfmdr1} (\textit{Plasmodium falciparum} multi drug resistant gene-1) effectively reduces the drug concentration at target site (Ruetz et al. 1996). Arylaminoalcohol resistance is one example in which there is an elevated gene copy number of \textit{pfmdr 1}. Increasing drug resistance for chloroquine, mefloquine, arylaminoalcohol and artemisinins in sub-saharan Africa, South Asia and America presents a perplexing problem which can only be solved by studying the new drug target in \textit{plasmodium} species.

Antibiotics and antimalarial drug resistance has been under study for more than two decades because of the high mortality associated with it. Non-mevalonate pathway as discussed earlier is one of the last evolutionary pathway which can be considered as an alternative drug target because of the non-existence in higher species. Enzymes involved in the synthesis of isoprenoids using NMVA pathway don’t have any orthologous in mammalian system. \textit{Arabdopsis thaliana} and \textit{Listeria monozytogenes} are some of the exceptions which use both the pathways for the isoprenoids synthesis. Complete genome sequence analysis of microbial species has found the consistence presence of NMVA pathway with few anomalies (Table 1.2), but the overall consensus represents this as a promising target.
Table 1.2: Distribution of metabolic enzymes involved in the biosynthesis of isoprenoids

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Non-Mevalonate Pathway</th>
<th>Mevalonate Pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DXS</td>
<td>DXR</td>
</tr>
<tr>
<td><em>Deinococcus radiodurans</em></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Mycoplasma genitalium</em></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Listeria monocytogenes</em></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Plasmodium falciparum</em></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*(+ = Present, - = Absent)

**Table adapted from Eisenreich et al (Eisenreich et al. 2004)

All the current antibiotics work predominantly by targeting the biosynthesis of macromolecular component (DNA, RNA, cell wall) of bacteria cell, while few of them actually target the metabolic enzymes (folic acid biosynthesis). In a bioinformatics analysis by Arigoni et al. it was found that approximately 30 *E. coli* genes with orthologous in other bacterial species are essential for the survival (Arigoni et al. 1998). NMVA pathway metabolic enzymes are among these set of genes, which are found in
pathogenic bacteria including *Mycobacterium tuberculosis* and *Helicobacter pylori*. *E. coli* carrying the deletion of NMVA pathway genes can to be only rescued by the supply of exogenous mevalonate (Altincicek et al. 2001), indicating the critical requirement of NMVA pathway for survival.

Along with bacteria some protozoal genomes e.g. *Trypanosoma cruzi*, *Leishmania major*, *Toxoplasma gondii*, *Cryptosporidium parvum*, *Plasmodium falciparum* and *Plasmodium vivax* have to been shown to have orthologous of the NMVA pathway enzymes. A long N-terminal leader/signaling sequence have been reported to present in the genes from *plasmodium*, which is believed to target the enzymes to the apicoplast. Apicoplast are supposed to be derived from secondary endosymbiosis of a plastid bearing red algae (Yeh and DeRisi 2011). Apicoplast in *plasmodium* has lost its photosynthetic function and most of its genome is transferred to the nucleus over the course of evolution, and it’s function is necessary for the survival of *plasmodium* in intraerythrocytic and intrahepatic stages. It has also been shown that fosmidomycin inhibited NMVA pathway can be rescued by the exogenous supply of IPP and DMAPP (Yeh and DeRisi 2011). In summary, NMVA pathway represents a highly potential target for the development of novel antibiotics and antimalarial.
1.5 References:


determinants of resistance and their clinical significance. *Cellular and Molecular
Life Sciences*, 63, 1586-1596.

Yeh, E. & J. L. DeRisi (2011) Chemical Rescue of Malaria Parasites Lacking an
Apicoplast Defines Organelle Function in Blood-Stage *Plasmodium falciparum*.  

Zepeck, F., T. Gräwert, J. Kaiser, N. Schramek, W. Eisenreich, A. Bacher & F. Rohdich
(2005) Biosynthesis of Isoprenoids. Purification and Properties of IspG Protein
Chapter 2: 1-deoxy-D-xylulose-5-phosphate synthase (DXS) Reaction Mechanism

2.1 Introduction

2.1.1 Isoprenoids biosynthesis through MVA and NMVA pathways

Isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are the precursor for isoprenoids, which is the largest family of biologically active compounds. Example of important isoprenoids include the ubiquiniones, sterols, dichols, triterpenes, and chlorophyll (Eisenreich et al. 1998). The isoprenoids family of molecule represents the most diversified set of chemical entities known to man and is of immense importance. Cholesterol one of the most intensly studied molecule derived from the IPP and DMAPP, serves as a precursors to steroids such as, the glucocorticoids, the androgens, the mineralocorticoids, the gestagens and estrogen (Eisenreich et al. 2004). The natural products Taxol, a relative recent addition to the isoprenoid family is clinical useful drug for treatment of mammary and ovarian tumors (Holmes et al. 1991).

The biomedical importance of the isoprenoids has always intrigued the scientific community with question “How these molecules are synthesized in the cell?” This intriguing question was answered for the Cholesterol in research that spanned the 1940’s and the 1950’s. The conversion of acetyl-CoA to cholesterol is outlined in every standard
undergraduate textbook and takes place in four stages. Stage 1 is the conversion of 3-acetyl-CoA thioesters to CO$_2$ and the 5-carbon acid, mevalonate (HOCH$_2$-CH$_2$-COH(CH$_3$)-CH$_2$-COO$^-$). Stage 2 is conversion of mevalonate to the activated isoprenes, IPP and DMAPP. Stage 3 is the conversion of IPP and DMAPP to the linear 30-carbon intermediate, squalene. The conversion of IPP and DMAPP to squalene involves the condensation of IPP and DMAPP to produce the 10-carbon geranyl pyrophosphate and pyrophosphate. A reaction between IPP and geranyl pyrophosphate leads to farnesyl pyrophosphate, a 15-carbon intermediate. Two farnesyl pyrophosphates react to generate the 30-carbon squalene. Stage 4 is cyclization of squalene to a fused ring 30-carbon sterol which is subsequently "trimmed down" to the 27-carbon cholesterol (Figure 1.2) (Vance and Van den Bosch 2000).

For many years, it was thought that IPP, DMAPP, and all the biologically-occurring isoprenoid were derived solely from this mevalonate-dependent (MVA) pathway. Recently, an alternative pathway to IPP, DMAPP, and the isoprenoids has been uncovered that exists in many eubacteria, green algae, protozoa, and plants. In this alternative pathway, IPP is derived not from mevalonate, but from pyruvate and glyceraldehydes 3-phosphate. A 5-carbon sugar phosphate, xylulose-5-phosphate is a key intermediate in the mevalonate-independent pathway of isoprenoid biosynthesis (Figure 2.1) (Rohmer et al. 1996). The enzymes catalyzing the reactions of the mevalonate-independent pathway or non-mevalonate pathway (NMVA) are completely different than the enzymes of MVA pathway of isoprenoid biosynthesis.
2.1.2 Targeting non-mevalonate pathway

Many of the human pathogens produce their isoprenoids exclusively via the NMVA pathway or a combination of the MVA and NMVA pathways. Of particular relevance to my research, the *Plasmodium* species that cause malaria are solely dependent upon the NMVA pathway for the production of IPP, DMAPP, and their isoprenoids (Cassera et al. 2004). Thus, inhibitors to the enzymes in the *Plasmodium* NMVA pathway are likely to be quite useful to treat malaria.

New treatments for malaria are essential as there has been an alarming increase in the rate of resistance to standard drug therapies in the last 10 years. The problem of drug resistance in the malaria parasite needs to be addressed before it become unsolvable. Most of the antimalarial and antibiotics in the market were developed more than 20 year ago and have merely been modified to develop more recent antimalarial drugs (Leeb 2004, Rohdich, Bacher and Eisenreich 2004). The antimalariais predominately work by targeting the biosynthesis of critical macromolecules (such as DNA, RNA, and the cell wall) (Sahu, Sahu and Kohli 2008). More promising drugs can be designed by targeting the enzymes involved in the key steps of physiological processes unique to *Plasmodium* parasite. The enzymes of the NMVA pathway are excellent targets for the development of new antimalarial drug because there are no homologues to these enzymes in man and other mammals. Because the NMVA pathway is solely responsible for IPP, DMAPP, and the isoprenoids in a number of human and animal pathogens and in problematic
weeds (Table 1.2), the enzymes of the NMVA pathway are intriguing targets for the development of novel drugs and herbicides.

2.1.3 Biochemistry of non-mevalonate pathway

The first committed step in the NMVA pathway (Figure 2.1) involves the condensation of pyruvate (Pyr) and D-glyceraldehydes-3-phosphate (D-GLP) to 1-deoxy-D-xylulose-5-phosphate (DXP) and CO₂ catalyzed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS). Next step involves the NADPH reductive rearrangement of DXP to 2-C-methyl-D-erythritol 4-phosphate (MEP) catalyzed by D-xylulose-5-phosphate reductoisomerase (DXR). In the sequential steps MEP is conjugated with cytidine monophosphate (CMP), followed by phosphorylation of the C2 hydroxyl group, elimination of CMP and cyclization to form 2C-methyl-D-erythritol 2, 4-cyclodiphosphate. The enzymes catalyzing these reactions are IspD, IspE, IspE and IspF respectively. The last steps involve the formation of 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate (HMBPP) and then to IPP and DMAPP by IspG and IspH, respectively (Figure 2.1). The IspH protein catalyzes the reduction of HMBPP to IPP and DMAPP at ratio of ~ 6 IPP to 1 DMAPP (significantly different from the thermodynamic equilibrium of ~1 IPP/3 DMAPP) (Gräwert et al. 2010). The interconversion of IPP and DMAPP is catalyzed by isopentenyl diphosphate isomerase (the Idi protein), which exists two species dependent forms, IDI-I and IDI-II (Kaneda et al. 2001, Barkley, Desai and Poulter 2004). Idi may be expressed in Plasmodium and (Mohanty and Srinivasan 2009),
if so, is likely present in the cytoplasm as IPP alone is able to rescue apicoplast-minus parasites (Yeh and DeRisi 2011).

![Diagram of Non-Mevalonate Dependent Biosynthesis of IPP and DMAPP]

Figure 2.1: The Non-Mevalonate Dependent Biosynthesis of IPP and DMAPP

2.1.4 DXS a rate limiting enzyme in NMVA pathway

DXS catalyzes a thiamine pyrophosphate (TPP) and divalent metal ion (Mg$^{2+}$ or Mn$^{2+}$) dependent transketolase-type condensation reaction and that is rate-limiting step in NMVA pathway. DXS from *Rhodobacter capsulatus*, *Mentha x piperita*, *Mycobacterium*
tuberculosis, Arabdopsis thaliana, and Escherichia coli have been studied and characterized (Estévez et al. 2000, Lange et al. 1998, Hahn et al. 2001, Mao et al. 2008, Lois et al. 1998). DXP, the product of the DXS reaction, is not only used for the production of IPP, DMAPP, and the isoprenoids, but is also the biosynthetic precursor to thiamine (vitamin B₁) and pyridoxal (vitamin B₆) in Plasmodium and other species (Müller, Hyde and Wrenger 2010). Both the MVA and NMVA pathways exist in higher plants, with majority of the IPP supply in cytoplasm being produced by the MVA pathway, while most of the the IPP supply in the plastid coming from the NMVA pathway. Up- and down-regulation of DXS expression in A. thaliana (Estévez et al. 2000, Estévez et al. 2001) results in the commensurate changes in the level of the isoprenoids: up-regulation increases the cellular levels of the isoprenoids while down-regulation suppresses the isoprenoid levels. Alteration in the expression and regulation of DXS in A.thaliana demonstrates that that DXS is essential for normal growth and provides additional evidence that DXS catalyzes the rate-limiting step in the NMVA pathway.

Recombinant DXS has been overexpressed from many species and its crystal structure from Escherichia coli and Deinococcus radiodurans has been reported (Xiang et al. 2007). Relative to the other dimeric TPP dependent enzyme, DXS is unusual in its domain organization and active site location within the monomer. Monomer unit of TPP dependent enzymes typically consists of three domains with active site located at the interface between the individual monomers and domain I of one monomer is located above domain II and III of the other monomer. However DXS exhibits a novel domain
organization with active site located within a monomer and domain I of one monomer is located right above domain II and III of the same monomer (Figure 2.2).

Figure 2.2: Crystal structure of DXS from *D. radiodurans*. This structure illustrates the dimer with one active site per monomer. Each active site contains a bound TPP (shown as red and orange colored balls).
2.1.5 DXS reaction mechanism

DXS belongs to TPP dependent transketolase, pyruvate decarboxylase and carboligase family which uses ping-pong mechanism for catalysis, where release of the first product is preceded by the binding of second substrate (Figure 2.3A) (Liu et al. 2001). A ordered sequential kinetic mechanism has been reported for *Rhodobacter capsulatus* DXS, with pyruvate binding first. D-GLP then bind to form the active DXS-Pyr-GLP complex which is then followed by the release of CO₂ and DXP in that order (Figure 2.3B) (Eubanks and Poulter 2003). In another report using single-molecule force spectroscopy (SMFS) nano sensor technique it was found that D-GLP binding was enhanced by 1.7-folds in the presence of pyruvate using *E. coli* DXS, also suggesting the reaction to proceed though ordered sequential mechanism (Sisquella et al. 2010). A recent conflicting report by Meyers and group have proposed that the reaction catalyzed by *E. coli* DXS proceeds through random sequential mechanism using tryptophan fluorescence quenching study (Figure 2.3C) (Brammer et al. 2011). In this study it was found that binding of the first substrate is independent of the binding of the second substrate and binding constants for pyruvate and D-GLP (individually) were calculated in absence of the second substrate.
The focus of the present study is to comprehend the active site pocket of *D. rad* DXS using DXS mutants and elucidate the mechanism of substrate binding using steady state kinetic analysis, product inhibition with DXP, and a dead end inhibitor (fluoropyruvate, F-Pyr). This work will lead to a better understanding of the DXS-catalyzed reaction to aid in the rational design of DXS inhibitors for the treatment of malaria and other human diseases.
2.2 Materials and Methods:

2.2.1 Materials:

Thiamine pyrophosphate, pyruvate, glyceraldehyde-3-phosphate, 1-deoxy-D-xylulose-5-phosphate sodium salt, bovine serum albumin and LB-broth were purchased from Sigma Aldrich. NADPH was from Alexis Biochemical’s, Ni-NTA resin was from Invitrogen, and β-mercaptoethanol (β-Me) was from Fisher. *E. coli* XL-10 cells, deoxynucleotide mix PCR grade, *pfu*Ultra hotstart DNA polymerase, QuikChange II site directed mutagenesis kit and acetonitrile (HPLC grade) were purchased from Agilent. The DNA vectors pET28a(+) and pET15b(+) and *E. coli* BL-21 B(DE3) cells were purchased from EMD Biosciences. DNA sequencing and primers were purchased from MWG Operon. All the other reagents were of the highest quality commercially available.

2.2.2 Cloning of recombinant *D. radiodurans* 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and *E. coli* 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR):

A codon optimized synthetic *D.radiodurnas* dxs gene with 5’-*NdeI* and 3’-*XhoI* restriction sites in pMK vector was purchased from Geneart, Germany. The dxs gene was used to clone cloned in pET28a(+) vector (*kanamycin* resistance) with N-terminal His-tag to yield pET28a(+)DXS plasmid. Briefly, pET 28a(+) vector and pMK vector were
digested with *NdeI* and *XhoI* restriction enzyme for 1 hour each at 37°C, followed by
digestion of pET28a(+) with Antarctic phosphatase for 30 mins at 37°C. The digested
samples were then run on an agarose gel (0.7%) and the appropriate bands were extracted
for the overnight ligation at 16°C. An aliquot of the ligation reaction (4 µL) was used to
transform *E. coli* XL-10 cells and plated on LB plates with kanamycin. Only colonies
with vector pET28a(+) - DXS grew on the plate and were cultured. Plasmid was extracted
from the cells cultured overnight and used for primary screening to confirm the insertion
of dxs gene in pET28a(+) vector. Cloning was further confirmed by DNA sequencing at
MWG Operon. Codon optimized *E. coli* dxr gene with 5’-*NdeI* and 3’-*BamHI* restriction
sites was used to clone in pET15b(+) with N-terminal His tag using the same protocol
described above.

2.2.3 Production of DXS mutant by site-directed mutagenesis

Site directed mutagenesis was carried out using the QuikChange II site directed
mutagenesis kit. Briefly, the mutagenesis mixture consists of 50-100 ng plasmid
pET28a(+) - DXS as a template, 1X PCR reaction buffer, 0.4 mM each of the forward and
reverse primer (Table 2.1), 0.25 mM dNTP mixture, 5 µL Quik solution, and 2.5 units of
*pfu*Ultra hotstart polymerase in a 50 µL reaction. The PCR thermo cycler settings used
were: 95°C for 4 min., 18 cycles of denaturation at 95°C for 30 sec., annealing at gradient
temperature between 58-61°C for 30 sec., elongation at 72°C for 16 min. and one final
elongation step at 72°C for 10 min. Samples were kept at 4°C, before digestion with
DpnI at 37° C for 2 hrs. These digested reactions were stored at -80° C before transformation of XL-10 cells. The sequence of the mutant plasmids DNA were confirmed by DNA sequencing.

### 2.2.4 Production of DXS mutant using the overlap extension method

The overlap extension method was used to produce the DXS mutant that were difficult to create via site directed mutagenesis. In this method, two PCR reactions were performed using set of primer in which one of set of primers possessed a point mutation and has an overlap (>15 base pairs) with the primer of the second PCR reaction (Appendix figure S-28). PCR products were extracted and used as a template for the next round of PCR in order to create the gene of interest with a point mutation. This product was ligated into pET28a(+) vector using the NdeI and XhoI restriction sites as discussed in section 2.2.2 and confirmed by DNA sequencing.

Table 2.1: PCR primers used to create the DXS mutants

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>E373A</td>
<td>Sense</td>
<td>5’ GAT GTG GCC ATT GCG GCG GAA GTG GCG GTT ACC A 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ TGG TAA CGC CCA CTT CGG CCG CAA TGC CCA CAT C 3’</td>
</tr>
<tr>
<td>H82A</td>
<td>Sense</td>
<td>5’ TTT GAT GTG GCC GCT CAG GCG TAT GC 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GCA TAC GCC TGA GCG CCC ACA TCA AA 3’</td>
</tr>
<tr>
<td>H304A</td>
<td>Sense</td>
<td>5’ GAT CCG ATT TAT TGG GCT GGT GCG GCG AAA TTT 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ AAA TTT CGC CGG ACC AGC CCA ATA AAT CGG ATC 3’</td>
</tr>
<tr>
<td>Y395F</td>
<td>Sense</td>
<td>5’ CGT CCG GTG GTG GCG ATT TTT AGC ACC TTT C 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GAA AGG TGC TAA AAA TCG CCA CCA CCG GAC G 3’</td>
</tr>
</tbody>
</table>
2.2.5 Overexpression and purification of DXS and the DXS mutants

Both wild type and mutant DXS plasmid were transformed in *E. coli* BL-21 B(DE3) cells and used for protein expression. An overnight culture of *E. coli* in LB broth
containing 50 µg/ml kanamycin was diluted 100-folds, cultured at 37 °C until the OD$_{600}$ reached ~ 0.6, and then cooled down to 20 °C. Expression was induced by the addition of 0.5mM isopropyl β-D-1-thiogalactopyranoside (IPTG), after shaking for 6 hrs cells at 20 °C, the cells were harvested by centrifugation at 6,000g for 10 min. at 4 °C and finally stored at -80 °C before purification. Cells were thawed and all the purification steps were performed at 4 °C. Cells were resuspended in binding buffer (20 mM Tris, 500 mM NaCl, 5 mM imidazole, 10 mM β-Me, pH = 7.5) supplemented with 1 mM phenylmethanesulfonylfluoride (PMSF), 4 µg/mL leupeptin and 2 µg/mL pepstatin, sonicated using a Heat systems W-380 ultrasonic processor, and centrifuged at 16,000g for 20 min. to remove cell debris. The supernatant from the cell lysate was applied to Ni-NTA resin (1.5 cm x 5 cm) which had been equilibrated with binding buffer at rate of 1 mL/min and non-bound proteins eluted from the column by first washing with 5 column volume of binding buffer followed by 20 column volume of wash buffer (20 mM Tris, 500 mM NaCl, 60 mM imidazole, 10 mM β-Me, pH = 7.5). DXS (wildtype and the desired mutants) was eluted using elution buffer (20 mM Tris, 500 mM NaCl, 250 mM imidazole, 10 mM β-Me, pH = 7.5). Fractions containing protein were dialyzed at 4 °C overnight against 2L of 20 mM Tris, 100 mM NaCl, 10 mM β-Me, pH = 7.5 and concentrated using an Amicon ultra centrifugation filter. Final yield from the expression was 7-8 mg/L of culture. Enzyme was flash frozen in liquid nitrogen, stored at -80 °C and the purity of the DXS (wildtype or the mutant proteins) were evaluated by SDS-PAGE.
2.2.6 Overexpression and purification of DXR

DXR was expressed in *E. coli* BL-21 B(DE3) cells. An overnight culture of *E. coli* in LB broth containing 100 µg/ml *ampicillin* was diluted 100-fold and cultured at 37 °C with continuous shaking. The temperature of the culture was lowered to 30 °C when the OD$_{600}$ reached ~ 0.6 and DXR expression was induced by the addition of 1 mM IPTG. Cells were cultured for another 5 hours at 30°C, harvested by centrifugation at 6,000g, and stored at -80 °C. The protocol mentioned above for DXS purification was used to purify recombinant DXR. The final yield from the expression was 9-10 mg/L of culture.

2.2.7 Determination of D-GLP and 1-Deoxy-D-xylulose-5-phosphate (DXP) concentration

D,L-GLP was obtained from Sigma Aldrich as a suspension in water. Concentration of D-GLP was measured spectrophotometrically using glyceraldehyde-3-phosphate dehydrogenase (GAPDH). A reaction mixture (0.5 mL) containing 30 mM sodium pyrophosphate buffer (pH = 8.6), 5 mM L-cysteine, 0.6 mM NAD$^+$, 0.6 mM sodium arsenate, and D,L-GLP at concentration less than 0.2 mM were incubated at 30 °C for 10 min, the reaction initiated with 2 µg of GAPDH (Novus Biotechnology), and the progress of the reaction monitored spectrophotometrically at 340 nm. The final
concentration of D-GLP was calculated based on the production of NADPH ($\varepsilon_{340} = 6,220 \text{M}^{-1} \text{cm}^{-1}$).

DXP was purchased from Sigma Aldrich as the sodium salt. The salt was resuspended in 10 mM HEPES buffer (pH = 6.0), and the concentration was measured spectrophotometrically using DXR. A reaction mixture (0.5 mL) containing of 100 mM HEPES (pH = 8.0), 1 mg/mL BSA, 1.5 mM MnCl$_2$, 0.3 mM NADPH, and DXP at concentration less than 0.1 mM were incubated at 37°C for 5 min, the reaction initiated with 1.25 µg of DXR, and the progress of the reaction monitored spectrophotometrically at 340 nm. Final concentration of DXP was calculated based on the consumption of NADPH ($\varepsilon_{340} = 6,220 \text{M}^{-1} \text{cm}^{-1}$).

2.2.8 DXS-DXR coupled assay

The DXS-DXR coupled assay solution consists of 100 mM HEPES buffer (pH = 8.0), 100 mM NaCl, 1 mg/ml BSA, 1 mM thiamine pyrophosphate (TPP), 1.5 mM MnCl$_2$, 2 mM β-Me, 0.15 mM NADPH, 0.2 mg/ml DXR, and varying concentrations of pyruvate or D-GLP. Steady-state kinetics studies were performed by varying pyruvate or D-GLP at fixed saturating concentration of the co-substrate. A DXS-DXR reaction solution was incubated at 37 °C for 5 min, the reaction was initiated by addition of 358 nM DXS, and the progress of the reaction monitored spectrophotometrically at 340 nm.
for the consumption of NADPH. NADPH consumption directly corresponds to the production of DXP. Assays were performed in triplicate and the data were fitted using equation 1 as discussed in section 2.11 to calculate the steady-state kinetic parameters.

2.2.9 Product inhibition of DXS by DXP

Product inhibition is studied using a modification of the HPLC assay (Brammer and Meyers 2009). This assay involves the modification of the unreacted substrate and product with 2,4-dinitrophenylhydrazine (2,4-DNP) followed by separation and quantification of the resulting 2,4-DNP adducts. A 0.25 mL reaction mixture consisting of 100 mM HEPES buffer (pH = 8.0), 80 mM NaCl, 1 mg/ml BSA, 5 mM MgCl₂, 1 mM thiamine pyrophosphate, 2 mM β-Me, and variable concentrations of of pyruvate, D-GLP, and DXP was incubated at 37 °C for 5 min. The reaction was initiated by addition of 573 nM DXS and, at the desired time, an aliquot (50 µL) was removed and quenched into equal volume of cold methanol, and incubated at 4°C for 20 min. Samples were centrifuged and a 50 µL aliquot of the supernatant was incubated with equal volume of 2,4-DNP reagent (100 mM 2,4-DNP in 2N H₂SO₄). After a 5min. incubation, the pH was adjusted to 5-7 by addition of 36 µL 4M Tris buffer (pH = 10.0), and then centrifuged to remove excess 2,4-DNP. Samples were stored at -20°C before HPLC analysis. The DNP derivatives of pyruvate, D-GLP, and DXP were separated using a 5 µm Discovery® C₁₈ column (25 cm × 4.6 mm) developed using the following conditions: flow rate = 1.5 mL/min; solvent A, 100 mM ammonium acetate, 0.05% TFA (pH = 4.6); solvent B,
acetonitrile, 0.05% TFA; 20% - 35% solvent B over 12 min and then 35 – 60% solvent B over 2 min. When varying [pyruvate] at one fixed [D-GLP], the concentrations used for the initial velocity kinetic experiments were as follows: D-GLP, 1.0 mM; pyruvate, 0.26, 0.50, 0.75, or 4.0 mM; and DXP, 0, 0.2, 0.3, or 0.4 mM. The DXP hydrazone peak area was measured and the DXP concentration was determined against a standard curve generated using DXP hydrazone. When varying [D-GLP] at one fixed [pyruvate], the concentrations used for the initial velocity kinetic experiments were as follows: pyruvate, 0.3 mM; D-GLP, 0.05, 0.1, 0.25, and 0.50 mM; and DXP, 0, 0.2, 0.3, or 0.4 mM. The D-GLP hydrazone peak area was measured and the D-GLP concentration was determined compared against a standard curve generated using D-GLP hydrazone.

Assays were performed in duplicate and the data was fitted and analyzed using equations discussed in section 2.11.

2.2.10 Inhibition of DXS by Fluoropyruvate (F-Pyr)

Initial reaction velocities were determined using the DXS-DXR coupled assay as detailed in Section 2.8. When varying [pyruvate] at one fixed [D-GLP], the concentrations used for the initial velocity kinetic experiments were as follows: D-GLP, 0.5 mM; pyruvate, 0.1, 0.18, 0.5, or 2 mM; and F-Pyr, 0, 25, 50, 75, or 100 µM. When varying [D-GLP] at one fixed [pyruvate], the concentrations used for the initial velocity kinetic experiments were as follows: pyruvate, 2 mM; D-GLP, 0.025, 0.05, 0.125, or 0.25 mM; and F-Pyr, 0, 25, 50, 75, or 100 µM. Assays were performed using 115nM of *D. radiodurans* DXS in
triplicate and the data were fitted using nonlinear least-squares regression analysis using
equation discussed in section 2.11.

2.2.11 Analyzing the initial velocity kinetic data

The steady state initial velocity for DXS measured at various concentrations of
pyruvate and D-GLP were fit to eq 2 using nonlinear regression analysis in Sigma-Plot

Equation 1 \( v = \frac{V_{max} [S]}{K_m + [S]} \)

Equation 2 \( v = \frac{V_{max} [Pyr][GLP]}{K_{iPyr}K_{mGLP} + K_{mGLP}[Pyr]+K_{mPyr}[GLP]+[Pyr][GLP]} \)

Equation 3 \( v = \frac{V_{max} [Pyr][GLP]}{K_{iPyr}K_{mGLP} + K_{mGLP}[Pyr]+[Pyr][GLP]} \)

Equation 4 \( v = \frac{V_{max} [Pyr]}{K_{mPyr} (1+[\text{Ki}])+[Pyr]} \)

Equation 5 \( v = \frac{V_{max} [Pyr][GLP]}{K_{mGLP} (1+[\text{Ki}])+[GLP] (1+[\text{Ki}])} \)

Equation 6 \( \text{Slope} = (K_{mGLP}/V_{max}) (1 + K_{iPyr}/[Pyr]) \)
Data for inhibition of fluoropyruvate against pyruvate and D-GLP were fit to equation 4 and 5 respectively at fixed concentration of the second substrate, where [I] is the concentration of fluoropyruvate and Ki is the inhibition constant with respect to fluoropyruvate. Data for product inhibition (DXP) was fit to equation 4 and 5 for pyruvate and D-GLP respectively, and in this case [I] represents the concentration of DXP and Ki is the inhibition constant associated with it.

2.3 Results and Discussion

2.3.1 Overexpression, purification of DXS and coupled assay

The wild type *D. radiodurans* dxs gene (DXS-WT) with a N-terminal His6 tag was codon optimized (codon adaptive index value = 0.98), synthesized by GeneArt and cloned in pET28a(+) vector using *NdeI* and *XhoI* restriction sites. A series of DXS mutants were designed near the active site pocket (Xiang et al. 2007) using DXS-WT as template by the Quikchange site-directed mutagenesis method or the overlap extension method (Ho et al. 1989). DXS-WT and the mutant enzymes (all possessing N-terminal His6 tag) were affinity purified using the Ni-NTA resin and the expression yield varied from 6-8 mg per liter of culture, with >95% purity (Figure 2.4 and supplementary figure S3).
There is no convenient, HTS-compatible assay to directly measure DXS activity. Therefore, *E. coli* DXR was cloned, over-expressed, and purified to establish a straightforward coupled assay to measure DXS activity. The DXS reaction is coupled to the DXR catalyzed NADPH-dependent reduction of DXP to MEP, which is monitored by the decrease in $A_{340}$.

2.3.2 Determination of the kinetic parameters for DXS-WT and the mutant DXS proteins

DXS is a TPP and Mg$^{2+}$/Mn$^{2+}$-dependent enzyme which belongs to the family of transketolase (TK) family of enzyme existing as dimer (Hawkins, Borges and Perham 1989). The folding pattern and dimer organization in the active site pocket for TPP binding in crystal structure of *D. radiodurans* and *E. coli* DXS has been reported to be different than that observed for transketolase and pyruvate dehydrogenase (other members of the transketolase family) (Xiang et al. 2007). *D. radiodurans* DXS has three domains: domain I, residue 1-319; domain II, residue 320-495, and domain III, residue 496-629. The TPP molecule amino-pyrimidine ring interacts with domain II and the pyrophosphate moiety interacts with domain I of the same monomer (Figure 2.5). While in case of TK and PDH, which also have three domains; the active site is located between
domain I of one monomer and domain II of the other monomer. This difference in domain organization has been associated to the relatively short linker between domain I and II (20 residues) in DXS, as compared to that of TK and PDH (95 residues) and could be the reason for different kinetic mechanism behavior as compared to the traditional TPP dependent enzymes.

Figure 2.5: Schematic drawing of the *D. rad* DXS depicting bound TPP with domain I (yellow), II (red), and III (blue) of the same monomer.
Figure 2.6: Sequence comparison of DXS from *H. influenza*, *V. colerae*, *P. vivax* (after gap removal), *P. falciparum* (after gap removal), *D. radiodurans*, *B. anthracis* and *M. tuberculosis* using Clustal W2. (gap = sequence of residues in *plasmodium*, which shares no homology with DXS from other species).
Amino acids in DXS involved in the binding of the cofactor TPP and the substrate D-GLP are highly conserved in all the species which uses NMVA pathway for isoprenoid biosynthesis. A sequence comparison of *Haemophilus influenza*, *Vibrio colerae*, *Plasmodium vivax*, *Plasmodium falciparum*, *Deinococcus radiodurnas*, *Bacillus anthracis* and *Mycobacterium tuberculosis* shows the conserved region for TPP and D-GLP binding (Figure 2.6). In DXS, C-2 atom of thiazolium ring is exposed to the solvent while rest of the molecule is buried inside tightly bound within domain I and II. The amino-pyrimidine ring of TPP interacts through carboxylate of Glu 373, π-stacking interaction with Phe 398, van der Waals interaction with Ile 371 and Ile 187 (all part of domain II).

Major role played by domain I is to support the pyrophosphate moiety of TPP through polar interaction with amino acid. The highly conserved His 52 (or His 49 in *E. coli*) in TPP utilizing enzyme has been proposed to play role in distributed proton transfer during catalysis. *E. coli* DXS mutant H49Q has been shown to be deficient in catalyzing the conversion of pyruvate and D-GLP to DXP (Querol et al. 2001). In contrast domain II plays major role in activating the thiazolium ring in TPP. Glu 373 of this domain is involved in the hydrogen bonding with N-1 atom of amino-pyrimidine ring and may help in the activation of the C-2 of the thiazolium ring though formation of catalytically potent ylide system. Replacement of Glu 373 with alanine leads to the completely inactive enzyme (Xiang et al. 2007), this residue is equivalent to Glu 56 oxalyl-CoA decarboxylase, another TPP dependent enzyme and mutation of this residue to alanine
also led to complete inactive enzyme (Berthold et al. 2007). Carboxylate group of Glu 373 is ~ 1.8 Å apart from the hydrogen of N1’ of TPP,
abstraction of this hydrogen by Glu 373 leads to the formation of the 1’, 4’-iminopyrimidine tautomer of TPP to facilitate the deprotonation of C-2 of thiazolium ring. After generation of this ylide, the C-2 carbon attacks the carboxylic group of pyruvate through nucleophilic addition reaction, which later undergoes decarboxylation to generate an activated C-2 pyruvate. This activated C-2 of pyruvate further attacks the aldehyde group of D-GLP through nucleophilic addition and later undergoes rearrangement to yield DXP and activated ylide (Figure 2.7).

Replacement of Asn 183 by alanine leads to an inactive enzyme, because of the role played by Asn 183 in the coordination of Mg$^{2+}$/Mn$^{2+}$. Mutation of Asn 181 to Ala is an active mutant with the same $K_m$ as wildtype, but with a 50% decrease in $k_{cat}/K_M$ (Table 2.2 and 2.3). This data suggest that Asn 181 has role in binding TPP in an optimum configuration for catalysis. On the other hand D-GLP binding was studied by mutating the residue which are involved in direct interaction with the substrate: His 304, Tyr 395, Arg 423, and Asp 430 (Figure 2.8). In previous report the conversion of Arg 401 and Arg 480 to alanine each lead to completely inactive R401A and R480A DXS mutants, suggesting that these two residues are important in catalysis, most likely by binding to the phosphate moiety of D-GLP (Xiang et al. 2007).

For Tyr 395 the role of hydroxyl group as well as the importance of aromatic ring has been studied here. Hydroxyl group interacts directly with D-GLP, as the mutant Y395A leads to 10-fold increase in the value of Km and 25-fold decrease in the value of
This mutant was rescued to a certain extent by Y395F mutant which showed 15-fold lower \( k_{\text{cat}}/K_m \) value (Table 2.2), indicating the importance of both hydroxyl group and aromatic system.

Table 2.2: Steady-state Kinetic Parameters for DXS WT and DXS Mutants for D-GLP

<table>
<thead>
<tr>
<th></th>
<th>Steady-state Kinetics Parameters for D-GLP*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( K_m ) (mM)</td>
</tr>
<tr>
<td>DXS WT</td>
<td>0.05 ± 0.01</td>
</tr>
<tr>
<td>H82A</td>
<td>0.03 ± 0.01</td>
</tr>
<tr>
<td>Y395F</td>
<td>0.59 ± 0.04</td>
</tr>
<tr>
<td>Y395A</td>
<td>0.67 ± 0.05</td>
</tr>
<tr>
<td>D430A</td>
<td>0.12 ± 0.01</td>
</tr>
<tr>
<td>H434A</td>
<td>0.23 ± 0.01</td>
</tr>
<tr>
<td>R423A</td>
<td>12.3 ± 3.2</td>
</tr>
<tr>
<td>R423K</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td>H304A</td>
<td>0.08 ± 0.02</td>
</tr>
<tr>
<td>N181A</td>
<td>0.03 ± 0.01</td>
</tr>
<tr>
<td>N183A</td>
<td>0</td>
</tr>
</tbody>
</table>

(*Apparent kinetic constants were measured by varying the initial concentration of DGLP at 2.0 mM pyruvate).
The other amino acids studied here were Asp 430, H434 and Arg 423. The mutants D430A and H434A have exhibit relatively small changes in the kinetic parameters, while R423A mutant showed a significant increase in $K_m$ value (250-folds) from 0.05 mM to 12.3 mM and decrease in the value of $k_{cat}/K_m$ value (1000-folds) relative to the DXS WT, suggesting that Arg 423 recognizes the phosphate moiety of D-GLP. Note that the $K_m$ value for DGLP for the R423A mutant is approximately equivalent to the $K_m$ for glyceraldehyde in DXS WT (14mM) (Hahn et al. 2001), providing additional evidence that the guanidino moiety of R423 interacts with the phosphate of D-GLP. D-GLP is a substrate for the R423K mutant with a 10-fold increases in the value of $K_m$ and 13-fold decrease in the value of $k_{cat}/K_m$. The $k_{cat}/K_m$ value for the R423K mutant is approximately 100-fold higher than the $k_{cat}/K_m$ value for the R423A mutant, again consistent with the importance of positive charge of the guanidine group in stabilizing the negative charge of D-GLP phosphate.

The previous section detailed how the DXS mutants affected the kinetic parameters measured when D-GLP was the variable substrate. The focus of this section is the effect of DXS mutation on the kinetic parameters measured when pyruvate is the variable substrate. The H434A mutants showed a significant increase in the value of $K_m$ (6-fold) and no significant changes in the $k_{cat}/K_m$ values for observed (Table 2.3, 2.4). H82A mutant showed only 5% turnover as compared to DXS WT and no change in the value of $K_m$ for pyruvate and DGLP, indicating the importance of this histidine as proton acceptor/donor. While H304A mutant showed change in $K_m$ (6-fold higher), $k_{cat}/K_m$ (45-fold lower), and only 11% turnover indicating the possibility of histidine as proton
acceptor/donor and binding of pyruvate through ionic interactions. Mutant DXS Y395A, Y395F, R423K, and D430A showed no significant difference in the $k_{cat}/K_m$ value for pyruvate, because of their role in supporting D-GLP binding as discussed previously. Steady-state parameters for R423A were not determined because the saturating concentration of D-GLP (~10 $K_m$) required was very high ($K_m = 12.3$ mM).

Figure 2.8: Stereo view of the DXS active site depicting active C-2 of the thiazolium ring, residues from domain I (H82, D154, N181, N183, and H304) and domain II (E373, Y395, R401, R423, D430, H434, and R480).
Table 2.3: Steady-state Kinetic Parameters for DXS WT and DXS Mutants for pyruvate

<table>
<thead>
<tr>
<th>Steady-state Kinetics Parameters for Pyruvate**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Km (mM)</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>DXS WT</td>
</tr>
<tr>
<td>H82A</td>
</tr>
<tr>
<td>Y395A</td>
</tr>
<tr>
<td>Y395F</td>
</tr>
<tr>
<td>D430A</td>
</tr>
<tr>
<td>H434A</td>
</tr>
<tr>
<td>R423A</td>
</tr>
<tr>
<td>R423K</td>
</tr>
<tr>
<td>H304A</td>
</tr>
<tr>
<td>N181A</td>
</tr>
<tr>
<td>N183A</td>
</tr>
</tbody>
</table>

(**Apparent kinetic constants were measured by varying the initial concentration of pyruvate at 0.5 mM D-GLP, aNot determined).
Table 2.4: Comparison of the Activity of the Mutant DXS Protein relative to the DXS WT.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Relative $V_{\text{max}}$ value (%) with D-GLP as varying substrate*</th>
<th>Relative $V_{\text{max}}$ value (%) with Pyr as varying substrate**</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS WT</td>
<td>100 ± 4</td>
<td>100 ± 4</td>
</tr>
<tr>
<td>E373A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H82A</td>
<td>4.8 ± 0.2</td>
<td>5.1 ± 0.2</td>
</tr>
<tr>
<td>Y395A</td>
<td>50.7 ± 0.9</td>
<td>39.8 ± 0.3</td>
</tr>
<tr>
<td>Y395F</td>
<td>77.6 ± 2.3</td>
<td>48.3 ± 0.7</td>
</tr>
<tr>
<td>D430A</td>
<td>99.8 ± 2.4</td>
<td>97.5 ± 1.9</td>
</tr>
<tr>
<td>H434A</td>
<td>125.2 ± 3.7</td>
<td>133 ± 3</td>
</tr>
<tr>
<td>R423A</td>
<td>22.3 ± 3.9</td>
<td>----&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>R423K</td>
<td>80.4 ± 2.4</td>
<td>53.6 ± 1.6</td>
</tr>
<tr>
<td>H304A</td>
<td>11.2 ± 0.7</td>
<td>13.1 ± 1.7</td>
</tr>
<tr>
<td>N181A</td>
<td>42.7 ± 0.9</td>
<td>44.8 ± 1.6</td>
</tr>
<tr>
<td>N183A</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(*Apparent kinetic constants were measured by varying the initial concentration of DGLP at 4.0 mM pyruvate, **Apparent kinetic constants were measured by varying the initial concentration of pyruvate at 0.5 mM D-GLP, <sup>a</sup>Not determined).
2.3.3 Pyruvate and D-GLP binding study

Family of enzymes which uses the TPP cofactor has been well studied. Steady-state kinetics, dead-end inhibition, product inhibition and crystallization studies have been used to elucidate the mechanism of substrate binding and reaction catalysis (Frank, Leeper and Luisi 2007). All TPP-dependent enzymes catalyzes two successive half reactions, the first step involves the attack of activated TPP (ylide system at C-2) to first substrate. The next step could occur in three possible ways, the most common being the classical ping-pong mechanism where the first product is released generating a metastable enzyme-intermediate complex. Product formation follows after onto the second substrate. Another possible second half reaction proceeds through an ordered mechanism, in which the first product can only be released if the second substrate is bound to the active site. For an ordered mechanism like this, the first substrate binds tightly and irreversibly, while the second substrate could only bind in the presence of first substrate. The last possible second half reaction could proceeds through random sequential mechanism, where binding of both the substrates are reversible and independent of each other (Figure 2.3)

Most of the known TPP-dependent enzymes like phosphoketolase-2 and pyruvate dehydrogenase (Yevenes and Frey 2008, Liu et al. 2001) follow a ping-pong mechanism. For phosphoketolase-2, first step involves the binding of either fructose-6-phosphate or xylulose-5-phosphate followed by release of either erythrose-4-phosphate or D-GLP, respectively, while proton acts as an acceptor in the second step resulting in the release of
acetyl phosphate. The kinetic mechanism of the substrate binding has been studied previously for DXS from *Escherichia coli*, *Rhodobacter encapsulates* and *Haemophilus influenzae* (Eubanks and Poulter 2003, Brammer et al. 2011, Matsue et al. 2010). Work from the Poulter lab demonstrates that the *R. capsulatus* DXS has an ordered sequential mechanism (Eubanks and Poulter 2003), using dead-end inhibition by fluoropyruvate and $^{14}$CO$_2$ trapping experiments. These researchers found that the conversion of pyruvate to $^{14}$CO$_2$ is inefficient in absence of D-GLP, eliminating the ping-pong mechanism a classical pathway used by many TPP-dependent enzymes. Studies on *E. coli* using methyl-acetylphosphonate (MAP) as a dead-end inhibitor and tryptophan fluorescence quenching experiment demonstrate the kinetic mechanism for this enzyme is random sequential (Brammer et al. 2011). The binding of pyruvate and D-GLP to *E. coli* DXS was found to be reversible and completely independent of each other. This is in contrast to the ordered sequential mechanism of *R. capsulatus* DXS, which requires that D-GLP can only bind to the enzyme in presence of pyruvate.

In this study of *D. radiodurans* DXS, a double reciprocal analysis of the substrates were used. Along with these studies, we also carried out product inhibition using DXP and dead-end inhibition using fluoropyruvate (F-pyr) to aid in defining the kinetic mechanism. Double reciprocal analysis at lower concentrations of D-GLP (0.01-0.03mM) and pyruvate (0.02mM-0.06mM) at different, but fixed concentrations of the other substrate revealed a non-competitive pattern with change in both slope and intercept (Figure 2.9, 2.10), strongly suggesting a random sequential mechanism. These results demonstrate that both substrates bind to DXS independent of each other. For an ordered
mechanism, we would have observed a competitive pattern for D-GLP and an uncompetitive pattern for pyruvate in case of ping-pong mechanism.

Figure 2.9: Double reciprocal analysis of initial velocities of pyruvate at different fixed concentration of D-GLP. Kinetic analysis was performed at sub-saturating concentration of D-GLP: 0.01 mM (●), 0.014 mM (○), 0.02 mM (▼), and 0.03 mM (△).
Figure 2.10: Double reciprocal analysis of initial velocities of D-GLP at different fixed concentration of pyruvate. Kinetic analysis was performed at sub-saturating concentration of pyruvate: 0.02 mM (●), 0.03 mM (○), 0.04 mM (▼), and 0.06 mM (△).

Product inhibition studies were next used to determine the mechanism of substrate binding and product release in the *D. radiodurans* DXS catalyzed reaction. DXP was used as product inhibitor against pyruvate and D-GLP. We found that DXP is a weak inhibitor with ~60% inhibition at concentration of 0.4 mM, yielding a Ki in low millimolar range (Table 2.5). Product inhibition by DXP, as we report here, provides a possible explanation of change in double reciprocal plots at higher concentration of pyruvate and D-GLP for *R. capsulatus* and *E. coli* DXS as compared to the plot at lower concentration (Brammer et al. 2011, Matsue et al. 2010). DXP showed a competitive pattern vs. pyruvate indicating that DXP competes for the same enzyme form as pyruvate (Figure 2.11), despite the chemical resemblance between DXP and D-GLP. When D-GLP was the varied substrate at fixed sub-saturating concentration of pyruvate DXP yielded a
non-competitive inhibition pattern (Figure 2.12). This shows that DXP and D-GLP do not bind to the same enzyme form. At a saturating concentration of pyruvate, no inhibition was detected by DXP when D-GLP was varied substrate. On the other hand, DXP inhibition was still observed at saturating concentration of D-GLP when pyruvate was the variable substrate.

Figure 2.11: Double reciprocal analysis of initial velocities of DXS inhibition by DXP at different initial concentration of D-GLP. Kinetic analysis were performed at following concentration of DXP: 0 mM (●), 0.2 mM (○), 0.3 mM (▼), and 0.4 mM (▼), and the initial concentration of D-GLP was fixed at 1 mM.
Figure 2.12: Double reciprocal analysis of initial velocities of DXS-inhibition by DXP at different fixed concentration of pyruvate. Kinetic analysis were performed at following concentration of DXP: 0 mM (●), 0.2 mM (○), 0.3 mM (▼), and 0.4 mM (▽), and the initial concentration of D-GLP was fixed at 0.3 mM.

Table 2.5: Observed inhibition pattern and inhibition constant for DXS with DXP and fluoropyruvate.

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>Substrate</th>
<th>Pattern</th>
<th>Ki (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXP</td>
<td>Pyr</td>
<td>Competitive</td>
<td>128.7 ± 32.1</td>
</tr>
<tr>
<td>DXP</td>
<td>D-GLP</td>
<td>Non-competitive</td>
<td>828.3 ± 155.5</td>
</tr>
<tr>
<td>F-Pyr</td>
<td>Pyr</td>
<td>Competitive</td>
<td>3.3 ± 0.4</td>
</tr>
<tr>
<td>F-Pyr</td>
<td>D-GLP</td>
<td>Non-competitive</td>
<td>57.1 ± 3.2</td>
</tr>
</tbody>
</table>
F-Pyr was also used as a dead-end inhibitor to study the mechanism of binding in *D. radiodurans* DXS. F-Pyr is competitive vs. pyruvate and non-competitive vs. D-GLP (Figure 2.13 and 2.14) and exhibits a Ki = 3.3 ± 0.4 M when pyruvate was the variable substrate. For an ordered or random sequential mechanism, F-pyr vs. D-GLP would yield a non-competitive pattern, while ping-pong mechanism would yield an

![Figure 2.13: Double reciprocal analysis of initial velocities of DXS inhibition by F-Pyr at different fixed concentration of pyruvate. Kinetic analysis were performed at following concentration of F-Pyr: 0 µM (●), 25 µM (o), 50 µM (▼), 75 µM (∇), and 100 µM (■), and the initial concentration of D-GLP was fixed at 0.5 mM.](image-url)
uncompetitive pattern. A result from inhibition study by F-Pyr with respect to D-GLP was difficult to distinguish between non-competitive and uncompetitive. However, the inhibition pattern by DXP and the double reciprocal studies of varying both substrates, the best fit to data is a random sequential mechanism of *D. radiodurans* DXS. This random sequential mechanism can degenerate to ordered sequential mechanism when the rate of reaction is more through path A than path B (Figure 2.15).
2.4 Summary

Non-mevalonate pathway represents a unique opportunity for the development of novel inhibitors to treat drug resistant pathogens, both bacteria and parasites. The enzymes involved in NMVA pathway have no resemblance or homology homology to the enzymes of the mevalonate-dependent pathway found in man and other mammals. The reaction catalyzed by DXS is the rate determining step (Estévez et al. 2001) of isoprenoid synthesis in A. thaliana as well as the source of DXP for the synthesis of vitamins B1 and B6 in Plasmodium sp. (Müller et al. 2010). In order to develop a novel inhibitor/lead against DXS, the reaction mechanism of DXS was studied using site-directed mutagenesis, double reciprocal studies, product inhibition and dead-end inhibitor.
2.5 References:


Querol, J., M. Rodríguez-Concepción, A. Boronat & S. Imperial (2001) Essential Role of Residue H49 for Activity of Escherichia coli 1-Deoxy--xylulose 5-Phosphate Synthase, the Enzyme Catalyzing the First Step of the 2-C-Methyl--erythritol 4-Phosphate Pathway for Isoprenoid Synthesis. *Biochemical and Biophysical Research Communications*, 289, 155-160.


3.1 Introduction

3.1.1 Malaria life cycle

Recent report by World Health Organization (WHO) have pointed out the increasing number of reported cases of various infectious diseases (WHO 2010), specially the most dreadful diseases known to human kind such as malaria (~81 millions) and tuberculosis (~5.8 millions) globally. Malaria is one of the most appalling forms of disease with shocking increase in number of incident in twenty first century (Petersen, Eastman and Lanzer 2011a, Carlton, Sina and Adams 2011). These higher number of cases of malaria has been associated with increasing drug resistance in plasmodium species in African, South Asian and South American continent (Ruetz et al. 1996).

There has been 156 named species of plasmodium genus which infects vertebrates, with *P. falciparum, P. vivax, P. ovale, P. knowlesi* and *P. malar*iae being the five major species responsible for malaria in humans (http://www.cdc.gov/malaria/about/disease.html ). The malaria parasite lives a complex life cycle in two hosts (*Anopheles* mosquito and vertebrates). In first stage, a parasite infected female *Anopheles* mosquito transmit the sporozites (infectious form of
plasmodium) while feeding on human host (Figure 3.1). These sporozoites infect the liver cells and asexually reproduces to multinuclear schizonts and later rupture to releases merozoites. These merozoites infect the red blood cells (RBC), where they again reproduce asexually into multinuclear schizonts by feeding on the cytoplasm and hemoglobin. Upon maturation of the schizonts, the RBC ruptures with the release of merozoits which then invades more RBC. A small percentage of the merozoites that

Figure 3.1: Life cycle of Malaria parasite in *Anopheles* mosquito and humans (Adopted from Center for Disease Control and Prevention).
invades the erythrocytes do not develop into schizonts, but instead differentiate into micro and macrogametocytes (male and female cells, respectively). These gametocytes have no activity within the human host and must be ingested into a female *Anopheles* mosquito for sexual reproduction. Once in the stomach of the mosquito, the micro- and macrogametocytes fuse to form an oocyst, which fills with the infectious sporozoites.

These sporozoites migrate to the salivary gland in the female *Anopheles* after oocyst rupture, ready to infect the next victim bitten by the mosquito. The lifecycle of *Plasmodium*, ultimately involving both mosquito and mammalian hosts, coupled to the complex interplay of the hepatic and erythrocyte stages of the protozoan infection present quite a challenge to the development of new antimalarial drugs. In case of *P. vivax* and *P. ovale* a dormant stage of hypnozites can form, which can replaces by invading the bloodstream in weeks or even year later.

### 3.1.2 Compartmentalization of metabolic pathways in plasmodium species

The whole genome sequence of *P. falciparum* and *P. vivax* has been published recently and gave vital information about the metabolic pathway plasmodium species used for survival and resistance (Carlton et al. 2008, Gardner et al. 2002). Infection by *P. falciparum* has been mostly dominant in African continent, while *P. vivax* is widespread
in Asian and American continent because *P. vivax* cannot infect duffy blood group negative reticulocytes (Mendis et al. 2001), a trait found in people from Africa. Both the plasmodium species share some common genomic features such as the total number of genes (~ 5400), while differ in other features such as GC % content (19.4 % in *falciparum* and 42.3 % in *vivax*). Using bioinformatics tools and sequencing results, some of the important biological pathways such as complete type II fatty acid synthesis, isopentenyl pyrophosphate (IPP) synthesis, iron sulfur cluster assembly, and a fragmented haem synthesis have been identified and localized in this parasite.

All the plasmodium species and other member of the phylum apicomplexa harbor apicoplast, an organelle which is homologous to the chloroplast of plants and algae (Köhler et al. 1997, McFadden et al. 1996). This organelle consists of 35-kb genome which encodes 30 proteins mainly consisting of complete set of tRNAs, ribosomal proteins and several unidentified open reading frame (Sato 2011, McFadden 2011). Homologs to plant plastid most of the apicoplast targeted proteins are encoded by nuclear gene and transported to the organelle through a two steps process. It has been estimated that approximately 800 nuclear encoded protein may be targeted to apicoplast (Waller et al. 2000). Protein targeted to apicoplast consists of bipartite N-terminal pre-sequence consisting of two functional domains: (a) signal peptide and (b) transit peptide (Figure 3.2). Using acyl carrier protein (ACP) as an example, the role of signal and transit peptide has been studied previously (Waller et al. 2000, Zuegge et al. 2001). When transit peptide from ACP was fused to N-terminal of green fluorescence protein (GFP), it leads to the accumulation of GFP in cytoplasm indicating the importance of signal peptide. On
contrary, when just the signal peptide was fused to the N-terminal of GFP all the GFP protein was accumulated within the parasitophorous vacuole, not in apicoplast.

![Diagram of apicoplast targeting protein](image)

**Figure 3.2**: Schematic diagram of apicoplast targeting protein in plasmodium species

This ordered two steps process shows that the signal peptide is important for protein to enter within the secretory pathway, then transit peptide play role in directing into the apicoplast, and finally transit peptide cleaves off to leave mature protein in apicoplast. Over the past years various software’s have been developed to differentiate between protein targeted to apicoplast, mitochondria and cytoplasm using the analogy to the plastid targeting protein used by plants. Software’s such as SignalP, PSORT, PATS and PlasmoAP has been developed to identify the signal peptide (Petersen et al. 2011b, Nakai and Kanehisa 1992, Zuegge et al. 2001). But, so far there is no clear consensus regarding the size and cleavage of transit peptide from mature protein. There are two ways to characterize the transit peptide presence: (a) net charge at the N-terminal (basic versus acidic), (b) presence of enriched aspargine and lysine residue (Foth et al. 2003, Ponpuak et al. 2007). Based on the mature apicoplast protein analyzed from *P.*
*falciparum* it has been shown that transit peptide can varies anywhere between 50 to over few hundreds amino acid (Table 3.1).

Table 3.1: Example of apicoplast targeting proteins

<table>
<thead>
<tr>
<th>Apicoplast targeting sequence in <em>P. falciparum</em></th>
<th>Size of the transit peptide</th>
<th>Comment/ References</th>
</tr>
</thead>
<tbody>
<tr>
<td>l-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)</td>
<td>49</td>
<td>(Jomaa et al. 1999)</td>
</tr>
<tr>
<td>rRNA methylase</td>
<td>54</td>
<td>Putative</td>
</tr>
<tr>
<td>queuine tRNA-ribosyltransferase</td>
<td>302</td>
<td>Putative</td>
</tr>
<tr>
<td>Alanine-tRNA ligase</td>
<td>367</td>
<td>Putative</td>
</tr>
<tr>
<td>Acyl carrier protein</td>
<td>78</td>
<td>(Waller et al. 2000)</td>
</tr>
</tbody>
</table>

Apicoplast consists of four outer membranes and the mechanism of translocation across these membranes is poorly understood. Cleavage of signal peptide and transit peptide while passing through these membrane is not understood either (Tonkin, Kalanon and McFadden 2008). In plants, stromal processing peptidase (SPP) has been characterized for the cleavage of transit peptide, and recently a putative non-plant stromal processing peptidase from *P. falciparum* has been proposed to do the similar chemistry as plant SPP (Richter and Lamppa 1999, van Dooren et al. 2002). The cleavage site for transit peptide can be found out by serial deletion of gene from N-terminal of the mature protein based on Clustal W2 alignment of apicoplast protein with same protein from other species, as previously reported for *P. falciparum* DXR (Jomaa et al. 1999).
3.1.3 Targeting non-mevalonate pathway in Plasmodium species

Recent discovery of the Non-mevalonate pathway (NMVA) in eubacteria and parasites has opened a new platform for the development of antibiotics and antimalarial drugs (Goodman, Su and McFadden 2007, Mbaya et al. 1990, Sacchettini and Poulter 1997). Based on bioinformatics tool and in vitro studies it has been established that NMVA exists in all the plasmodium species and tuberculosis strains, thus making the enzymes involved in NMVA as an alluring target for drug discovery. In this study 1-deoxy-D-xylulose-5-phosphate synthase (DXS), the first enzyme of NMVA pathway in *P. vivax* has been explored. DXS has been cloned and overexpressed in *E. coli* using genomic DNA and codon optimized DNA as a source. Full length DXS with signal and transit peptide as well as mature protein without these peptides by serial deletion has also been overexpressed. Kinetic parameters have been calculated and compared it to the DXS from other species.
3.2 Materials and Methods:

3.2.1 Materials

Thiamine pyrophosphate, pyruvate, glyceraldehyde-3-phosphate, 1-deoxy-D-xylulose-5-phosphate sodium salt, bovine serum albumin and LB-broth were purchased from Sigma Aldrich. NADPH was from Alexis Biochemical’s, Ni-NTA resin was from Invitrogen, and β-mercaptoethanol (β-Me) was from Fisher. E. coli XL-10 cells, deoxynucleotide mix PCR grade, pfuUltra hotstart DNA polymerase and acetonitrile (HPLC grade) were purchase from Agilent. The DNA vectors pET28a(+) and pET21b(+), E. coli strains BL-21 B(DE3) and Rosetta B(DE3) cells were purchased from EMD Biosciences. DNA sequencing services and primers were purchased from MWG operon. All the other reagents were of the highest quality commercially available.

3.2.2 Cloning of recombinant \emph{P. vivax} 1-deoxy-D-xylulose-5-phosphate synthase (DXS) from \emph{P. vivax} genomic DNA

\emph{P. vivax} dxs gene was extracted from strain \emph{Sal-I} provided by Dr. John H. Adams Lab (Department of Global Health, University of South Florida). PCR primers used were: forward primer 5’- CCTAGGATCCGATGATAATGGGAACCTCCTCTG -3’ with \emph{BamHI} restriction site (underlined) and 5’- ATCCAAGCTTGGTGGACACCCTCCTTGCAG-
3’ as reverse primer with *HindIII* restriction site (underlined). PCR conditions were optimized by varying the concentration of DNA template, primers, concentration of DMSO, and annealing temperature. The final optimized reaction conditions used were 50 ng of the genomic DNA, 10 ng/µL of each primer and 10% DMSO. Thermocycler condition used for the amplification of 3333 base pair fragment were 95 °C for 4 min., followed by 30 cycles of denaturation at 95 °C for 30 sec., annealing at 60°C for 30 sec, elongation at 72 °C for 3 minutes 30 seconds and one final elongation step at 72 °C for 10 min.

The PCR product of *P. vivax* dxs gene with *BamHI* and *HindIII* restriction sites was used to clone in pET21b(+) with C-terminal His-tag to yield pET21b(+)-vivax-DXS plasmid. Briefly, pET21b(+) vector (*ampicillin* resistance) and PCR product were digested with *BamHI* and *HindIII* restriction enzyme for 1 hour each at 37 °C respectively, followed by digestion of pET21b(+) with Antarctic phosphatase for 30 mins at 37 °C. The digested samples were then run on gel (0.7%) and the appropriate bands were extracted for the overnight ligation reaction at 16 °C. An aliquot of the ligation reaction (4µL) was used to transform *E. coli* XL-10 cells and plated on LB plate supplemented with 100µg/mL *ampicillin*. Colonies with pET21b(+)vivax-DXS plasmid grew on the plate supplemented with *ampicillin*. These colonies were cultured and the plasmid was extracted for primary screening to confirm the insertion of dxs gene in pET21b(+) vector. Cloning was further confirmed by DNA sequencing at MWG operon.
3.2.3 Overexpression and purification of P. vivax DXS

pET21b(+)-vivax-DXS plasmid was used to transform E. coli strain BL-21 and Rosetta B(DE3) E. coli cells. An overnight culture of E. coli in LB broth containing 100 µg/ml ampicillin and 1% glucose was diluted 100-folds, cultured at 37 °C until the OD<sub>600</sub> reached ~ 0.6 and then cooled down to 20 °C. Expression was induced by the addition of 0.5mM isopropyl β-D-1-thiogalactopyranoside (IPTG), after shaking for 6 more hrs at 20 °C, the cells were harvested by centrifuged at 6,000g for 10 min. at 4 °C and finally stored at -80°C before purification. Cells were thawed and all the purification steps were performed at 4°C. Cells were resuspended in binding buffer (20 mM Tris, 500 mM NaCl, 5 mM Imidazole, 10 mM β-Me, pH = 7.5) supplemented with 1 mM phenylmethanesulfonylfluoride (PMSF), 4 µg/mL leupeptin, 2 µg/mL pepstatin, and sonicated using Heat systems W-380 ultrasonic processor, and centrifugation at 16,000g for 20 min. to remove cell debris. The supernatant from the Cell lysate was applied on Ni-NTA resin (1.5 cm x 5cm) which had been equilibrated with binding buffer at rate of 1 mL/min and non-bound proteins eluted from the column by first washing with 5 column volume of binding buffer and 20 column volume of wash buffer (20 mM Tris, 500 mM NaCl, 60 mM Imidazole, 10 mM β-Me, pH = 7.5). P. vivax DXS was eluted using elution buffer (20 mM Tris, 500 mM NaCl, 250 mM Imidazole, 10 mM β-Me, pH = 7.5). Fractions containing protein were dialyzed at 4°C overnight against 2L of 20 mM Tris, 100 mM NaCl, 10 mM β-Me, pH = 7.5 and concentrated using an Amicon ultra centrifugation filter. Final yield from the expression was ~40 µg /L of culture from BL-21 cells and ~80 µg /L of culture from Rosetta B(DE3) cells with ~ 80% purity as judged by SDS-PAGE.
3.2.4 Cloning of recombinant *P. vivax* 1-deoxy-D-xylulose-5-phosphate synthase (DXS) and constructs without signal and transit peptide from codon optimized *P. vivax* dxs gene

A codon optimized synthetic *P. vivax* dxs gene with 5’-*NdeI* and 3’-*HindIII* restriction sites in pUC57 vector was obtained from Genescript, New Jersey. pET21b(+) vector (*ampicillin* resistance) was used to clone in the full length DXS with C-terminal His-tag to yield pET21b(+) -DXS (SH-1) plasmid. Briefly, pET21b(+) vector (*ampicillin* resistance) and PCR product were digested with *NdeI* and *HindIII* restriction enzyme for 1 hour each at 37 °C respectively, followed by digestion of pET21b(+) with Antarctic phosphatase for 30 mins at 37 °C. The digested samples were then run on gel (0.7%) and the appropriate bands were extracted for the overnight ligation reaction at 16 °C. An aliquot of the ligation reaction (4µL) was used to transform *E. coli* XL-10 cells and plated on LB plate supplemented with 100µg/mL *ampicillin*. Colonies with pET21b(+) -DXS (SH-1) plasmid grew on the plate supplemented with *ampicillin*. These colonies were cultured and the plasmid was extracted for primary screening to confirm the insertion of dxs gene in pET21b(+) vector. Cloning was further confirmed by DNA sequencing at MWG operon.

Shorter fragmented construct from SH-2 to SH-9 was constructed by PCR extraction of the gene using the set of primer mentioned in Table 3.2 and cloned in pET28a(+) vector (*kanamycin* resistance). The PCR product and the plasmid were
digested with NdeI and XhoI restriction enzyme for 1 hour at 37\(^\circ\)C, followed by digestion of pET28a(+) with Antarctic phosphatase for 30 minutes at 37\(^\circ\)C respectively, followed by digestion of pET28a(+) with Antarctic phosphatase for 30 mins at 37 \(^\circ\)C. The digested samples were then run on gel (0.7\%) and the appropriate bands were extracted for the overnight ligation reaction at 16 \(^\circ\)C. An aliquot of the ligation reaction (4\(\mu\)L) was used to transform \textit{E. coli} XL-10 cells and plated on LB plate supplemented with 50 \(\mu\)g/mL \textit{kanamycin}. Colonies with respective plasmid grew on the plate supplemented with \textit{kanamycin}. These colonies were cultured and the plasmid was extracted for primary screening to confirm the insertion of dxs gene in pET28a(+) vector. Cloning was further confirmed by DNA sequencing at MWG operon.

Table 3.2: Primer used for the DXS constructs (restriction site underlined).

<table>
<thead>
<tr>
<th>Construct</th>
<th>Type</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH-2</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG TAT CCG ACC TCC GCG ACG 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
<tr>
<td>SH-3</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG GAC AAC CGT GAT CCG CCG 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
<tr>
<td>SH-4</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG CTG GAT GCC CTG TTC GAC3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
<tr>
<td>SH-5</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG TAC GGA GAA AATC TAC C 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
<tr>
<td>SH-6</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG GAA AAA TCG GTG GCC CTG 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
<tr>
<td>SH-7</td>
<td>Sense</td>
<td>5’ GTAG CAT ATG AAA AAA GTG CTG AAA GTT GTG 3’</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>5’ GTAG CTC GAG TTA GTT GCT AAC ACC CTG 3’</td>
</tr>
</tbody>
</table>
3.2.5 Overexpression of full length and truncated P. vivax DXS

Both full length and truncated fragment of P. vivax DXS were expressed in E. coli BL-21 B(DE3) cells. A fresh colony of E. coli containing the appropriate plasmid was cultured at 37 °C in LB broth containing 100 µg/ml ampicillin for SH-1 and 50 µg/ml kanamycin for SH-2 to SH-9 supplemented with 0.8% glucose and 25mM potassium phosphate (pH = 7.2), was grown for 2-3 hours until OD$_{600}$ reached ~ 0.3. Culture was then diluted 100- fold and cultured at 37 °C until OD$_{600}$ reached ~ 0.6 and then cooled down to 25 °C. Expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), after shaking for 5-6 hours cells were harvested by centrifuged at 6,000g for 10 min. at 4°C and finally stored at -80 °C before purification.

3.2.6 Ni-NTA Affinity chromatography purification

Cells were thawed and all the purification steps were performed at 4°C. Cells were resuspended in binding buffer (20 mM Tris, 500 mM NaCl, 20 mM Imidazole, 10
mM β-Me, pH = 7.5) supplemented with 1 mM phenylmethanesulfonylfluoride (PMSF), 4 µg/mL leupeptin, 2 µg/mL pepstatin, sonicated using Heat systems W-380 ultrasonic processor, and centrifuged at 16,000g for 20 min. to remove cell debris. The supernatant from the cell lysate was applied to Ni-NTA resin (1.5 cm x 5 cm) which had been equilibrated with binding buffer at rate of 1mL/min and non-bound proteins eluted from the column by first washing with 5 column volume of binding buffer followed by 20 column volume of wash buffer (20 mM Tris, 500 mM NaCl, 60 mM Imidazole, 10 mM β-Me, pH = 7.5). DXS SH-1 and SH-2 was eluted using elution buffer (20 mM Tris, 500 mM NaCl, 250 mM Imidazole, 10 mM β-Me, pH = 7.5) and fractions containing protein were concentrated for gel filtration chromatography.

3.2.7 Gel filtration chromatography purification of SH-2 construct

Gel filtration chromatography was performed on a 2.5 x 90 cm Pharmacia Sephadex G-200 column. Column was calibrated with Apoferritin (400 kDa), β-Amylase (200 kDa), Alcohol dehydrogenase (150 kDa), Bovine serum albumin (66 kDa), and Carbonic anhydrase (29 kDa) at the rate of 0.5 mL/min with buffer containing 50 mM Tris (pH = 7.5), 150 mM NaCl, and 10 mM β-ME. The void volume for the column was calculated using Blue Dextran 2000. Affinity purified construct SH-2 was loaded on the column and fractions were collected (recording A280). Fractions with positive DXS activity were pooled together and concentrated to 1 mg/mL using 10 kDa Amicon ultra centrifugation filter. The molecular mass of recombinant *P. vivax* DXS SH-2 construct
was determined from a plot of the molecular mass (log $M_r$) versus the ratio of elution volume to void volume ($V_e/V_o$). Final yield from the expression was 1 mg/L of culture. Enzyme was flash frozen in liquid nitrogen, stored at -80°C and the purity was judged by SDS-PAGE.

3.2.8 Determination of D-GLP concentration

D,L-GLP was obtained from Sigma Aldrich as a suspension in water. Concentration of D-GLP was measured spectrophotometrically glyceraldehyde-3-phosphate dehydrogenase (GAPDH). A reaction mixture (0.5mL) containing 30 mM Sodium pyrophosphate buffer (pH = 8.6), 5 mM L-cysteine, 0.6 mM NAD$^+$, 0.6 mM Sodium arsenate, and D,L-GLP at concentration less than 0.2 mM were incubated at 30 °C for 10 min, the reaction initiated with 2μg GAPDH (Novus Biotechnology) and the progress of the reaction monitored spectrophotometrically at 340 nm. The final concentration of D-GLP was calculated based on the production of NADPH ($\varepsilon_{340} = 6,220$ M$^{-1}$ cm$^{-1}$).

3.2.9 Preparation of Apo-DXS SH-2
Apo-DXS SH-2 was prepared using a modified procedure (Wang, Martin and Singleton 1997, Meshalkina et al. 2010). Briefly, 0.5 mg/mL of DXS SH-2 in buffer containing 50 mM Tris (pH=7.5), 150 mM NaCl, and 10 mM β-ME was resuspended with saturated ammonium sulfate, 5 mM TCEP, pH = 3.5 at ratio of 2:3, and kept on ice for 20 min. After incubation, the mixture was centrifuged at 12,000g for 10 minutes and the pellet was resuspended in buffer containing 50 mM Tris (pH = 7.5), 150 mM NaCl and 2mM TCEP. Removal of TPP and Mg$^{2+}$ was monitored by assaying for DXS activity using DXS-DXR coupled assay (section 3.2.10).

### 3.2.10 DXS-DXR coupled assay

The DXS-DXR coupled assay solution consists of 100 mM HEPES buffer (pH = 7.0), 100 mM NaCl, 1 mg/ml BSA, 1 mM thiamine pyrophosphate (TPP), 1.5 mM Magnesium chloride (MgCl$_2$), 2 mM β-Me or 2 mM TCEP, 0.15 mM NADPH, 0.2 mg/ml DXR, and varying concentration of pyruvate or D-GLP. Steady-state kinetics studies were performed by varying pyruvate and D-GLP at fixed concentration of the co-substrate. A DXS-DXR reaction solution was incubated at 37 °C for 5 min, the reaction was initiated by addition of 96.2 nM *P. vivax* DXS SH-2, and the progress of the reaction was monitored at 340 nm for the consumption of NADPH. NADPH consumption directly corresponds to the production of DXP. Assays were performed in triplicate and the data was fitted using equation 1 as discussed in section 3.2.12 to calculate the steady-state kinetic parameters.
3.2.11 DXS HPLC assay

Non-continuous DXS assay was performed in a 0.25 mL reaction mixture consisting of 100 mM HEPES buffer (pH = 7.0), 80 mM NaCl, 1 mg/ml BSA, 2 mM β-Me, 5 mM Magnesium chloride (MgCl₂), 1 mM thiamine pyrophosphate (TPP) and varying concentration of pyruvate, D-GLP and fixed concentration of co-substrate were incubated at 37 °C for 5 min. Reaction was initiated by addition of 192.6 nM of *P. vivax* DXS SH-2, and at the desired time, an aliquot (50 µL) was removed and quenched into equal volume of cold methanol, and incubated at 4°C for 20 min. Samples were centrifuged and a 50 µL aliquot of the supernatant was incubated with equal volume of 2,4-DNP reagent (100 mM 2,4-DNP in 2N H₂SO₄). After a 5 min. incubation, the pH was adjusted to 5-7 by addition of 36 µL 4M Tris buffer (pH = 10.0), and then centrifuged to remove excess 2,4-DNP. Samples were stored at -20°C before HPLC analysis. The DNP derivatives of pyruvate, D-GLP, and DXP were separated using a 5 µm Discovery® C₁₈ column (25 cm × 4.6 mm) developed using the following conditions: flow rate = 1.5 mL/min; solvent A, 100 mM ammonium acetate, 0.05% TFA (pH = 4.6); solvent B, acetonitrile, 0.05% TFA; 20% - 35% solvent B over 12 min and then 35 – 60% solvent B over 2 min. In case of pyruvate at fixed concentration of D-GLP concentration (0.25 mM), DXP hydrazone peak area (DXP production) was measured and the concentration was determined by standard curve of DXP hydrazone. Assay was performed in triplicate and the data was fitted and analyzed using equations discussed in section 3.2.12.
3.2.12 Equation for analyzing the kinetic data

The steady state initial velocity for DXS measured at various concentrations of pyruvate and D-GLP were fit to eq 7 using nonlinear regression analysis in Sigma-Plot 12.0

\[ v = \frac{V_{\text{max}} [S]}{K_m + [S]} \]

3.2.13 Inductively coupled plasma mass spectrometry (ICP-MS) for Mg\(^{2+}\) estimation

Magnesium ion content in *P. vivax* DXS SH-2 was determined using Perkin Elmer Elan II quadrupole inductively coupled plasma-mass spectrometer (ICP-MS). Instrument was calibrated using 1, 20, 60, and 100 ppb Mg\(^{2+}\) standard solution. Enzyme sample was diluted to 0.05 mg/mL corresponding to a concentration of 414 nM and was used further with no modification. Experiment was repeated in duplicate.
3.3 Results and Discussion

3.3.1 Cloning and expression of full length P.vivax DXS using genomic DNA

Malaria is the most neglected of all human diseases and it is estimated that approximately 50% of the world population were at the risk of malaria in 2010 (WHO 2010). Global problem in malaria and other neglected diseases such as tuberculosis, leishmaniasis, and trypanosomiasis can be addressed by research in the development of drugs targeting novel pathways. But, due to low return on the investment all the pharmaceutical companies haven’t been keen to address this issue. A recent amendment act of 2007 by the Food and Drug Administration (FDA) is giving an opportunity to companies to apply for a “voucher” to expedite the FDA review of any drug made by the company, given the condition of promotion to develop pharmaceutical products for neglected diseases (Kesselheim 2008). All the current antimalarial chemotherapeutics targets the asexual stages of parasite and more specifically the pathways that exists in cytosol, food vacuole, mitochondria, parasite membrane and apicoplast (Rosenthal 2003, Sahu, Sahu and Kohli 2008).

It has been reported previously by Cassera et al., that the non mevalonate pathway is functionally active in all intraerythrocytic stages of P.falciparum (Cassera et al. 2004), and none of the enzymes from this pathway has homologues enzymes in mammals, thus
making this pathway as an unexplored territory for the development of antimalarial drug. The complete genome sequencing of *P. vivax* has further confirmed the presence of proteome of the NMVA, with the presence of bipartite N-terminal sequence for apicoplast targeting (Carlton et al. 2008). So, far there have been two reports on the expression of enzymes involved in NMVA in *P. falciparum*. In the first report, DXR the second enzyme of the NMVA pathway without the signaling and transit peptide has been expressed in *E. coli* as inclusion bodies and further refolded to regain the active enzyme (Jomaa et al. 1999). The other report involves the expression of *P. falciparum* IspF, which express as inclusion bodies when expressed as full length enzyme. When IspF was expressed as truncated protein (without signaling and transit peptide), it was expressed as soluble homogenous protein with only 5% maximum catalytic activity as compared to *E. coli* homolog (Rohdich et al. 2001). Study done in this report presents a first novel example of expression of soluble, homogeneous and active *P. vivax* DXS.

Open reading frame on the reverse complement of nucleotide sequence from 693,715 to 697,050 on chromosome 12 of *P. vivax* Sal-1 was PCR out using genomic DNA (Figure 3.3). The *P. vivax* dxs gene flanked by the engineered in *BamHI* and *HindIII* restriction sites was cloned into the pET21b(+) vector with a C-terminal His<sub>6</sub> tag. Occurrence of rare codon in high frequency is one of the challenging problems of protein expression in *E. coli* and in *P. vivax* dxs gene it accounts for about 9.1%. The rare codon frequency and occurrence of tandem repeat of rare doublet and triplet codon (total 102 out of 1112 codons) could makes the expression of this protein very difficult. In order to overcome this barrier Rosetta B(DE3) cell line was used. This cell line provides the
tRNA for the rare codon (Arg-AGG, AGA, CGA, Leu-CTA, Pro-CCC, Ile-ATA) which are in low abundance in *E. coli*.

Figure 3.3: (A) *P. vivax* dxs gene encoded on chromosome-12 (highlighted in yellow color), (B) DNA gel showing- Marker (lane 1) and *P. vivax* dxs gene isolated using genomic DNA as template (lane 2).

After, confirming the DNA sequence, the pET21b(+)-vivax-DXS vector was used to transform *E. coli* strains, BL-21 B(DE3) and Rosetta B(DE3). Recombinant DXS was purified using Ni-NTA resin as described in section 3.2.6. The final DXS produced as described herein was ~70-80 % pure as judged by SDS-PAGE (Figure 3.4). The final yield of 40µg of partially purified enzyme / L of culture from BL-21 cells and 80µg using Rosetta B(DE3) cell line was obtained. This full length *P. vivax* DXS was active and
specific activity of the 70% pure protein (0.9 ± 0.1 µmoles/min/mg) was comparable to the value with *D. rad* DXS (5.6 µmoles/min/mg).

Figure 3.4: SDS-PAGE analysis of *P. vivax* DXS in *E. coli*. MW standard (lane 1), cell lysate (lane 2), Flow through from Ni-NTA resin (lane 3), binding buffer wash (lane 4), wash buffer wash (lane 5), and purified *P. vivax* DXS (lane 6).

### 3.3.2 Cloning and expression of full length *P. vivax* DXS using codon optimized DNA

Expressions of protein/enzyme from plasmodium species in *E. coli* have always been hard, and most of the time it ends up in an inclusion bodies. But in case of *P. vivax* DXS, we can use the result from the expression of this enzyme from genomic DNA as a proof of concept that this enzyme can be expressed functionally active in *E. coli* with no formation of inclusion bodies. In order to overcome the effect of high rare codon frequency in genomic DNA and low expression yield, gene for the *P. vivax* DXS was
codon optimized (Genescript) for expression in *E.coli*. Initial analysis of the *dxs* gene revealed other factor such as cis-acting element (ribosome binding site, poly A and poly T) apart from high rare codon frequency which could have hindered the expression of *dxs* gene (Sharp and Li 1987). In total there were total of six *E. coli* ribosome binding site (AGGAGG), two poly-A (AAAAAAA), and two poly-T(TTTTTT) were found in the *dxs* gene. Synthetic *P.vivax* *dxs* gene with high codon adaptive index (0.89) and without the cis-acting element was optimized for expression in *E.coli* and supplied in pUC 57 vector with gene flanked by 5’-*NdeI* and 3’-*HindIII* restriction sites (Figure 3.5)

![Figure 3.5: pUC 57 vector with codon optimized *P. vivax* *dxs* gene.](image)

Codon optimized *P. vivax* *dxs* gene in pUC 57 vector obtained from Genescript was cloned in pET21b(+) vector using *NdeI* and *HindIII* restriction site. After sequence
confirmation the pET21b(+) vector was transformed into BL-21 B(DE3) cell line, and *P. vivax* DXS was expressed and purified as discussed in section 3.2.3. Expression of full length *P. vivax* DXS which includes both signal and transit peptide always yielded a truncated protein along with full length protein (Figure 3.6) with an overall yield of ~1.5 mg per L of culture (~18 fold more than the vector pET21b(+) vector). Various conditions such as induction temperature, protease inhibitor, induction time, and IPTG concentration were tried to avoid this truncation of full length protein, but it cannot be avoided.

![Image of SDS-PAGE analysis](image)

Figure 3.6: SDS-PAGE analysis of *P. vivax* DXS SH-1 in *E. coli*. MW standard (lane 1), cell lysate (lane 2), Flow through from Ni-NTA resin (lane 3), binding buffer wash (lane 4), wash buffer wash (lane 5), and Elution *P. vivax* DXS (lane 6-9).

### 3.3.3 Cloning and expression of truncated *P. vivax* DXS

To overcome the problem of full length protein truncation, the next step used was to express the *P. vivax* DXS without the nuclear encoding region (signaling and transit).
All the enzymes involved in the non mevalonate pathway are nuclear encoded (Figure 3.2), but transported to the apicoplast for the expression of the enzyme (Hunter 2007). Recent discovery of the apicoplast targeting enzymes and the role of signaling and transit peptide in transportation and folding of gene from nucleus to apicoplast are still not clear. Based on the sequence of protein functionally active in apicoplast, set of rules have been defined to characterize the apicoplast targeting proteins: (a) starts with signal peptide, (b) a stretch of 40 amino acids with at least nine asparagines and/or lysines residues after the signal cleavage site, (c) ratio of basic residues to acidic residues in aspargine/lysine-enriched region to be at least 5 to 3 (Foth et al. 2003). Over the year many software has been developed to predict the apicoplast targeting protein and cleavage site between signal and transit peptide, and using these web based tools predicts similar results with size ranging approximately from 25-30 amino acids (Table 3.3).

Table 3.3: Predicted cleavage site between signal and transit peptide.

<table>
<thead>
<tr>
<th>Software used</th>
<th>N-terminal Sequence (cleavage site highlighted)</th>
<th>Expected Size</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SignalP</td>
<td>MIMGTSSLLLLLALIITSMHV SLNAAASGKCEV</td>
<td>25</td>
<td>(Petersen et al. 2011b)</td>
</tr>
<tr>
<td>PSORT</td>
<td>MIMGTSSLLLLLALIITSMHV SLNAASGKCEV</td>
<td>30</td>
<td>(Nakai and Kanehisa 1992)</td>
</tr>
<tr>
<td>PATS</td>
<td>MIMGTSSLLLLLALIITSMHV SLNAASGKCEV</td>
<td>23</td>
<td>(Zuegge et al. 2001)</td>
</tr>
<tr>
<td>PlasmoAP</td>
<td>MIMGTSSLLLLLALIITSMHV SLNAAASGKCEV</td>
<td>25</td>
<td>(Foth et al. 2003)</td>
</tr>
</tbody>
</table>
Transit peptide cleavage site from mature protein upon transportation to the apicoplast is still unanswered, and it’s difficult to develop software to predict the cleavage site because the transit peptide can vary from around 24 to over few hundreds amino acids and have no conserved sequence motifs when compared to the apicoplast targeting proteins. As a control, sequence of DXS from *Plasmodium* species were compared and aligned to get a better picture of transit peptide (Figure 3.7) and the comparison of 200 amino acids from N-terminal, showed no sequence homology and identity. Showing that even the same enzyme from different plasmodium species shares no correlation in transit peptide sequence. A recent study using transit peptide of acyl carrier protein from *P. falciparum*, it has been proposed that transit peptide is either low occupancy helix or disordered both *in-vivo* and *in-vitro* (Gallagher, Matthews and Prigge 2011), and upon introduction of ordered structure through mutation, the ability of peptide to translocate to apicoplast was lost.
Figure 3.7: N-terminal Sequence (~400 amino acid) comparison of DXS from *P. berghei*, *P. yoelii*, *P. falciparum*, *P. vivax*, and *P. knowlesi*.
In order to express the soluble and active *P. vivax* DXS without the signaling and transit peptide, sequence of *E. coli*, *D. rad* and *P. vivax* DXS were compared to understand the cleavage site between transit peptide and mature protein. Clone SH-7 to SH-9 (Table 3.4) were designed first based on high sequence similarity with *E. coli* and *D. rad* DXS using the same analogy as used for the expression of DXR from *P. falciparum* (Jomaa et al. 1999).

Table 3.4: N-terminal amino acid truncation for SH-2 to SH-9.

<table>
<thead>
<tr>
<th>Clone Name</th>
<th>N-terminal amino acids truncated</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH-2</td>
<td>210</td>
</tr>
<tr>
<td>SH-3</td>
<td>245</td>
</tr>
<tr>
<td>SH-4</td>
<td>252</td>
</tr>
<tr>
<td>SH-5</td>
<td>272</td>
</tr>
<tr>
<td>SH-6</td>
<td>303</td>
</tr>
<tr>
<td>SH-7</td>
<td>347</td>
</tr>
<tr>
<td>SH-8</td>
<td>357</td>
</tr>
<tr>
<td>SH-9</td>
<td>395</td>
</tr>
</tbody>
</table>
Figure 3.8: Clustal W2 alignment of *E. coli*, *D. rad.* and *P. vivax* DXS, and showing the truncation site for *P. vivax* DXS.
Figure 3.8: Continued

Deletion of 347 to 395 amino acids from N-terminal to generate clone SH-7 to SH-9 respectively, yield a soluble but in-active enzyme. Therefore amino acids from N-
terminal were sequentially removed to achieve homogenous and active enzyme (Figure 3.9). Alignment of DXS from different species and between the plasmodium species shows that sequence between amino acid 200 and 395 shares a high homology between plasmodium species (Figure 3.7 and 3.8). Sequential deletion led to the generation of clone SH-2, which was expressing soluble, homogenous and active protein in good yields.

![DNA gel showing cloned vector digested with NdeI and XhoI restriction enzymes. Marker (lane-1), SH-1 (lane-2), SH-2 (lane-3), SH-3 (lane-4), SH-4 (lane-5), SH-5 (lane-6), SH-6 (lane-7), SH-7 (lane-8), SH-8 (lane-9), and SH-9 (lane-10).](image)

**3.3.4 Purification of P. vivax DXS SH-2**
*P. vivax* DXS clone SH-2 was overexpressed in *E. coli* BL-21 B(DE3) cell line. Due to the instability of plasmid at high cell density, cells were always kept OD$_{600} < 0.8$ before expression. Along with this, cells were supplemented with 0.8% glucose and 25 mM potassium phosphate buffer (pH = 7.2) to overcome the toxicity of plasmid in cells. Affinity chromatography was used as a first step of purification and final yield observed after this step was 2.5 mg per liter of culture (Figure 3.10). All the active fractions from this step were pooled together, concentrated (8mL with 5% glycerol) and loaded on to gel filtration column (Sephadex G-200), which was previously been calibrated with

![Figure 3.10: SDS-PAGE analysis of *P. vivax* DXS SH-2 affinity purification. MW standard (lane 1), uninduced control (lane 2), cell lysate (lane 3), flow through the Ni-NTA resin (lane 4), wash (lane 5), elution of *P. vivax* DXS SH-2 (lane 6-9).](image)
apoferitin, β-amylase, alcohol dehydrogenase, bovine serum albumin, and carbonic anhydrase. Fractions from gel filtration column were checked for DXS activity using DXS-DXR coupled assay. The first peak with higher $A_{280}$ value was found to have no or very little activity and based on standard calibration this peak might corresponds to either inactive tetramer of DXS or high molecular weight aggregate. Second peak with relatively lower A80 value gave the highest activity of DXS and probably exists as dimer (Figure 3.11 and 3.12). DXS from *R. capsulatus* and *D. rad* have also been reported previously to exist as dimers (Hahn et al. 2001, Xiang et al. 2007)

![Figure 3.11: Gel filtration chromatography of *P. vivax* DXS SH-2 on Sephadex G-200](image-url)
Active fractions peak from gel filtration column were pooled (1/3 of the peak height) and concentrated to 1mg/mL and used for further characterization. SDS-PAGE and western blot analysis of the affinity column and gel filtration column showed that \( P.\ vivax \) DXS SH-2 was >95% pure and homogenous (Figure 3.13). After separation of high molecular weight aggregate or inactive tetramer, the final yield of purified enzyme was around 1 mg per L of \( E.\ coli \) cultures. This yield was lower as compared to the expression yield of \( D.\ rad \) DXS (7-8 mg per L of culture), where also codon optimized gene was used for the expression. This relatively low yield could be due to relatively high molecular weight of the \( P.\ vivax \) DXS (~104 kDa) as compared to \( D.\ rad \) DXS (69.9 kDa) and around 60% of the expressed protein was found out to be inactive after gel filtration chromatography.
3.3.5 Characterization of *P. vivax* DXS SH-2

Homogeneously purified *P. vivax* DXS SH-2 was further used for characterization. DXS-HLPC assay was used for the determination of pH optimum of the reaction catalyzed by DXS at saturating concentration of pyruvate and D-GLP. Optimum pH for the *P. vivax* DXS SH-2 was found out to be between 7-7.5 (Figure 3.14), which is comparable to DXS from *R. capsulatus* (7.0) and *E. coli* (8.0) (Hahn et al. 2001, Brammer and Meyers 2009).
DXS need TPP and divalent metal ion (Mg\(^{2+}\) or Mn\(^{2+}\)) for catalytic activity. In order to understand the binding of divalent metal ion and the concentration required to achieve maximum activity, enzyme activity was checked in absence of divalent metal ion. We observed that the enzyme was completely active even in absence of any metal ion and same observation was seen for TPP. Even though no external source of TPP and Mg\(^{2+}\) was used during culturing and purification, suggesting that *P. vivax* DXS SH-2 used TPP and Mg\(^{2+}\) cofactors from the media (LB broth) used for culturing and are very tightly bound.
To remove the bound TPP and Mg\textsuperscript{2+} from enzyme, it was dialyzed in presence of EDTA, and was found that even after dialyzing the enzyme in presence of EDTA, we found that enzyme was completely active in absence of TPP and Mg\textsuperscript{2+}. Protocol used to remove the tightly bound cofactors was adopted from the method used to remove TPP from transketolase (Wang et al. 1997, Meshalkina et al. 2010). Incubation of enzyme with saturated ammonium sulfate at lower pH (3.5), lead to complete removal of TPP with high recovery upon reconstitution, but Mg\textsuperscript{2+} was still bound to the enzyme as observed by active enzyme in absence of Mg\textsuperscript{2+} using HPLC assay. Activity of the enzyme was reconstituted back upon addition of TPP (Figure 3.15). Inclusion of 5mM EDTA in this method however, led to the degradation of enzyme with only 25 % recovery, indicating that Mg\textsuperscript{2+} plays crucial role in folding and absence of it causes the enzyme to be insoluble.

![Figure 3.15: Steady state kinetics of P. vivax DXS SH-2 upon reconstitution with TPP](image)

Inductively coupled plasma resonance mass spectrometry (ICP-MS) was then used to quantify the amount of Mg\textsuperscript{2+} in enzyme. Molar ratio of Mg\textsuperscript{2+} to DXS was found out to be 0.97 ± 0.01, which is approximately 1:1. These results suggests that Mg\textsuperscript{2+} is
bound more tightly than TPP and replacement of this tightly bound Mg$^{2+}$ with other divalent metal ion (Mn$^{2+}$, Co$^{2+}$ and Ca$^{2+}$) could only be achieved by culturing the cells in minimal media with the respective divalent metal ion.

Steady state kinetic parameters for pyruvate and D-GLP for *P. vivax* DXS SH-2 were calculated using DXS-DXR coupled assay and are comparable with value from *D. rad* and *E. coli* DXS (Table 3.5 and 3.6). Apparent K$_m$ value of pyruvate for *P. vivax* DXS SH-2 were slightly higher than the value from *D. rad* and *E. coli*, one of the reason for this behavior could be due to different pH optima of DXS in these organisms.

<table>
<thead>
<tr>
<th>Organism</th>
<th>pH</th>
<th>Pyruvate$^a$</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K$_m$ (mM)</td>
<td>K$_{cat}$ (s$^{-1}$)</td>
</tr>
<tr>
<td><em>P. vivax</em></td>
<td>7.0</td>
<td>0.87±0.11</td>
<td>9.5 ± 0.5</td>
</tr>
<tr>
<td><em>D. rad</em></td>
<td>8.0</td>
<td>0.28 ± 0.03</td>
<td>7.4 ± 0.3</td>
</tr>
<tr>
<td><em>R. capsulatus</em></td>
<td>8.0</td>
<td>0.61 ± 0.05</td>
<td>1.9 ± 0.1</td>
</tr>
</tbody>
</table>

$^a$Apparent kinetic constants were measured by varying the initial concentration of pyruvate at 0.1 mM D-GLP.
Table 3.6: Comparison of steady state kinetic constants for D-GLP

<table>
<thead>
<tr>
<th>Organism</th>
<th>pH</th>
<th>D-GLP&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$K_m$ (mM)</td>
<td>$K_{cat}$ (s&lt;sup&gt;-1&lt;/sup&gt;)</td>
</tr>
<tr>
<td><em>P. vivax</em></td>
<td>7.0</td>
<td>0.019±0.004</td>
<td>10.9 ± 0.9</td>
</tr>
<tr>
<td><em>D. rad</em></td>
<td>8.0</td>
<td>0.05 ± 0.01</td>
<td>7.9 ± 0.4</td>
</tr>
<tr>
<td><em>R. capsulatus</em></td>
<td>8.0</td>
<td>0.068 ± 0.001</td>
<td>1.9 ± 0.1</td>
</tr>
</tbody>
</table>

<sup>b</sup>Apparent kinetic constants were measured by varying the initial concentration of D-GLP at 4.0 mM pyruvate.

3.4 Summary

Expression of *plasmodium* protein in *E. coli* has always been a challenging problem specially for the apicoplast targeting protein (Vedadi et al. 2007). There have been only two reports on the expression of enzyme involved in NMVA pathway from *P. falciparum* (Jomaa et al. 1999, Rohdich et al. 2001). This study presents the first report on the expression of *P. vivax* DXS in *E. coli*, where it is expressed as soluble and homogenous enzyme. This enzyme has been characterized in great detail and it behaves similar to DXS from other organisms.
3.5 References:


Apicomplexan organisms. *Molecular and Biochemical Parasitology*, 151, 100-110.


Chapter 4: Conclusion and Future Directions

4.1 Conclusion

In this study, mechanism of reaction catalyzed by TPP and Mg$^{2+}$ dependent DXS was studied, along with its potential role in development of novel antibiotics and antimalarial. DXS is the first and also the rate limiting enzyme in NMVA pathway. Presence of NMVA pathway in many human pathogens and parasites; and non-existent of any homologues enzymes of NMVA pathway in human opens a unique opportunity to develop inhibitors against the rising incidents of neglected human diseases.

DXS reaction mechanism and substrate binding was studied using steady state kinetics, product inhibition, dead end inhibitor, and through rational mutant design. Family of TPP dependent enzyme mostly catalyzed the reaction through ping-pong mechanism. Previous reports using *R. capsulatus* and *E. coli* DXS, the reaction is predicted to be ordered sequential and random sequential mechanism respectively. In this report, using *D. radiodurans* DXS we were able to differentiate the mechanism to be random sequential mechanism. We observed that DXP inhibits the reaction catalyzed by DXS with inhibition constant in high micro molar range ($K_i = 128 \ \mu M$ for pyruvate and $K_i = 828 \ \mu M$ for D-GLP) with competitive against pyruvate and non-competitive with respect to D-GLP. DXP bind to the same enzyme pocket as does the pyruvate even though resembling more with D-GLP. Similar trend was observed with F-Pyr. Summing up all these results with a novel domain organization in DXS as discussed in section
2.1.4, it can be concluded that the enzyme exhibits striking difference in substrate binding as compared to other TPP dependent enzymes and follows random sequential mechanism.

Expression of protein from apicomplexan such as *Plasmodium*, *Cryptosporidium* and *Toxoplasma* in *E. coli* has always been a challenging problem especially for apicoplast targeting protein (Vedadi et al. 2007). It has been predicted that 10% of the *plasmodium* species protein are targeting to the apicoplast through secretory pathway. Translocation and folding of protein from nucleus to the apicoplast is through use of a bipartite N-terminal signaling and transit sequence is still unclear (Tonkin, Kalanon and McFadden 2008). All the enzymes required for the synthesis of isoprenoids have been shown to translocate to apicoplast, and opens up a new platform for the development of novel antimalarial (Carlton et al. 2008, Gardner et al. 2002).

Present study deals with the expression of *P. vivax* DXS in *E. coli*. To overcome the expression problem due to codon bias and cis-acting element, dxs gene was codon optimized. Expression of *P. vivax* DXS without the signaling and transit sequence was the next challenging question. To answer this question, step wise deletion approach was used (section 3.3.3). Deletion of 210 amino acids from N-terminal led to the expression of soluble and active enzyme. Affinity and gel filtration chromatography was used to purify *P. vivax* DXS SH-2 clone to homogeneity. The purified enzyme was characterized and compared with DXS from other organisms.
4.2 Future Directions

4.2.1 Lead compound for P. vivax DXS inhibition

DXS catalyzes the rate-determining step of the NMVA pathway, and is required for the erythrocytic development of Plasmodium, and is positioned at the interface of two other critical metabolic pathways (Cassera et al. 2004). The product of the DXS reaction, is also required for biosynthesis of pyridoxal and thiamin (Müller, Hyde and Wrenger 2010). As a consequence, DXS is an attractive drug target. A DXS inhibitor given in combination with an inhibitor developed specifically to inhibit fatty acid synthesis in the apicoplast, a metabolic pathway crucial to hepatic development of Plasmodium (Tarun et al. 2008, Vaughan et al. 2009) could be a used as a powerful drug combination for the treatment of malaria. Such a drug combination would be lethal to the parasite and could exhibit limited human toxicity and would be less likely to evoke resistance in the parasite. A modern and interdisciplinary approach to identify lead compounds for the inhibition of P. vivax DXS could prove useful for novel antimalarial.

Antimalarial compound libraries from GSK, Novartis, and St. Jude's Children's Research Hospital (encompassing over 300,000 compounds) can be used for virtually screening against the P. vivax DXS homology model build in collaboration with Dr. Guida’s lab. In addition, other commercially available chemical libraries like the ZINC
database (http://zinc.docking.org/) can also be virtually screened. Top hits for the virtual screen will then be screened for *P. vivax* DXS inhibition using DXS-DXR coupled assay.

### 4.2.2 Crystal structure of *P. vivax* DXS

Expression of proteins from *Plasmodium* species has always been challenging problem because of codon mismatch and toxicity of the protein with respect to the choice of host used for expression. According to protein data bank (pdb), till date there has been only 371 reports for *P. falciparum* and 52 reports for *P. vivax* on protein crystallization. These numbers are increasing at great rate because of the advancement in molecular biology. So, far there is only one report on the crystallization of an enzyme (DXR) of *P. falciparum* from NMVA pathway (Umeda et al. 2010). Study done in this report has successfully solved the first problem of expression of soluble and homogenous protein from *P. vivax*. The next step is to start the crystallization trial for this enzyme. A request has been submitted to “Seattle Structure Genomics for Infectious Disease” and *P. vivax* DXS is currently being screened to find the condition needed to obtain crystal. X-ray crystal structure of this enzyme will be helpful in rational drug design against these drug resistant malaria parasites.
4.3 References:


Part A: Abbreviations

IPP                      Isopentenyl pyrophosphate
DMAPP              Dimethylallyl pyrophosphate
MVA                    Mevalonate
NMVA                Non Mevalonate
HMGR             HMG-CoA reductase
GLP                    Glyceraldehyde-3-phosphate
DXS                    1-deoxy-D-xylulose-5-phosphate synthase
DXR                    1-deoxy-D-xylulose-5-phosphate reductoisomerase
IspD                  4-Diphosphocytidyl-2C-methyl-D-erythritol 4-phosphate synthase
IspE                  4-Diphosphocytidyl-2C-methyl-D-erythritol Kinase
IspF                   2C-Methyl-D-erythritol 2,4 diphosphate synthase
IspG                  2C-Methyl-D-erythritol 2, 4-cyclodiphosphate reductase
IspH                  1-Hydroxy-2-methyl-butenyl 4-diphosphate reductase
NADPH               Nicotinamide adeninedinucleotide phosphate reduced form
DXP                    1-deoxy-D-xylulose-5-phosphate
MEP                    2-C-methyl-D-erythritol 4-phosphate
CMP                    Cytidine monophosphate
HPLC              High-performance liquid chromatography
TPP                    Thiamine pyrophosphate
TK                    Transketolase
PDH                    Pyruvate dehydrogenase
F-Pyr                Fluoropyruvate
ICP-MS Inductively coupled plasma mass spectrometry

Part B: Figures

Figure S-1: *D. rad* DXS wild type pyruvate Kinetics

Figure S-2: *D. rad* DXS wild type D-GLP Kinetics
Figure S-3: SDS-PAGE analysis of DXS WT and mutants: Marker (lane 1), DXS WT (lane 2), H82A (lane 3), Y395F (lane 4), Y395A (lane 5), N181A (lane 6), N183A (lane 7), R423A (lane 8), H434A (lane 9), D430A (lane 10), R423K (lane 11), and H304A (lane 12).

Figure S-4: *D. rad* H82A DXS pyruvate Kinetics
Figure S-5: *D. rad* H82A DXS D-GLP Kinetics

Figure S-6: *D. rad* Y395A DXS pyruvate Kinetics
Figure S-7: *D. rad* Y395A DXS D-GLP Kinetics

Figure S-8: *D. rad* Y395F DXS pyruvate Kinetics
Figure S-9: *D. rad* Y395F DXS D-GLP Kinetics

Figure S-10: *D. rad* D430A DXS pyruvate Kinetics
Figure S-11: *D. rad* D430A DXS D-GLP Kinetics

Figure S-12: *D. rad* H434A DXS pyruvate Kinetics
Figure S-13: *D. rad* H434A DXS D-GLP Kinetics

Figure S-14: *D. rad* R423A DXS pyruvate Kinetics
Figure S-15: *D. rad* R423A DXS D-GLP Kinetics

Figure S-16: *D. rad* R423K DXS pyruvate Kinetics
Figure S-17: *D. rad* R423K DXS D-GLP Kinetics

Figure S-18: *D. rad* N181A DXS pyruvate Kinetics
Figure S-19: *D. rad* N181A DXS D-GLP Kinetics

Figure S-20: *D. rad* DXS wild type thiamine pyrophosphate binding kinetics
Figure S-21: *E. coli* DXR steady state kinetics with DXP

![DXR DXP Kinetics](image)

\[ y = 75.444 \times x, R^2=0.998 \]

Figure S-22: D-GLP standard curve using DNP assay

![D-GLP Standard curve using DNP assay](image)
Figure S-23: DXP standard curve using dinitrophenyl hydrazine

Figure S-24: *D. rad* DXS wild type pyruvate kinetics using DNP assay
Figure S-25: Lineweaver-Burk plot of *D. rad* DXS wild type pyruvate kinetics using DNP assay

Figure S-26: *D. rad* DXS wild type D-GLP kinetics using DNP assay
Figure S-27: Lineweaver-Burk plot of *D. rad* DXS wild type D-GLP kinetics using DNP assay.

---

Figure S-28: Mutagenesis by overlap extension method employs the polymerase chain reaction (PCR) as a mean of altering the DNA. This technique can be used for the generation of site specific mutants and insertion or deletion in the cloned gene with greater efficiency in few steps.
Appendix-2

Production of the Catalytic Core of Human Peptidylglycine α-Hydroxylating Monooxygenase (hPHMcc) in *Escherichia coli*

1. Introduction

Approximately 50% of all mammalian peptide hormones possess an α-amidated C-terminus, which is required for bioactivity [1,2]. The α-amidated peptides are produced by the oxidative cleavage of C-terminal glycine-extended precursors to generate the α-amidated peptide and glyoxylate [3,4]. The enzyme catalyzing this reaction is bifunctional peptidylglycine α-amidating monooxygenase (PAM). Bifunctional PAM is comprised of two separate catalytic units, peptidylglycine α-hydroxylating monooxygenase (PHM, EC 1.14.17.3) and peptidylamidoglycolate lyase (PAL, EC 4.3.2.5) (Figure. 1) [5-7]. The PAM/PHM/PAL enzyme system represents a significant impediment to the manufacture of therapeutically important α-amidated peptides. The large-scale, inexpensive production of recombinant α-amidated peptides has been thwarted by difficulties in the bacterial expression of PAM, PHM, or PAL. The only reported success in bacteria has been the co-expression of rat PAM with the glycine-extended calcitonin or glucagon in *Streptomyces* to yield the desired α-amidated peptide [8-10]. Methods to circumvent this limitation for α-amidated peptide production include the use of PAM *in vitro* to amidate recombinantly produced glycine-extended peptides [11-15], the overexpression of a glycine-extended peptide in mammalian cells that
express PAM [16-19], the expression of a suitable peptide precursor that is chemically converted to the α-amidated peptide in vitro [20-23], transacylation/transpeptide methods [24-27], production in the milk of transgenic rabbits [28], and an intein based method [29].

The expression of mammalian proteins in *E. coli* is often limited by (a) large differences in codon frequencies between mammalian and *E. coli* proteins, (b) the inability of *E. coli* to properly carry out the post-translational modifications found in many mammalian proteins, and (c) the formation of insoluble inclusion bodies. In order to overcome such problems for the *E. coli* expression of human PHM, we utilized a synthetic gene coding for the catalytic core of human PHM (hPHMcc), which was optimized for *E. coli* expression. Kolehkar *et al*. [30] have shown that amino acids 42-356 of rat PHM retained the catalytic properties of the full length enzyme. The two putative N-linked glycosylation sites, Asn-411 and Asn-762, within are absent in hPHMcc [31], which should aid in the expression of active enzyme. We evaluated a series of fusion partners, glutathione-S-transferase (GST), thioredoxin (Trx), N utilization substance A (NusA), maltose binding protein (MBP), and a His6 tag, for the expression of active, soluble hPHMcc. We report herein on the successful expression of a Trx-hPHMcc-His6 fusion protein that exhibits steady-state kinetic parameters similar to that of wildtype rat PHMcc.
2. Materials and methods

2.1 Materials

BamHI, HindIII, Antarctic Phosphatase, pMAL-c2X, and T4 DNA ligase were purchased from New England Biolab. Goat polyclonal anti-PAM monoclonal antibody (sc-17393), sheep anti-goat IgGs, and horseradish peroxidase were purchased from Santa Cruz Biotechnology. Oligonucleotides were obtained from Integrated DNA Technologies. Ni-NTA resin, pET21a(+), pET32a(+), pET42b(+), pET43.1a(+) vectors, and the E.coli strains BL 21 (DE3) and Origami B(DE3) were purchased from Novagen (Table S1, supplementary material). Protease inhibitor cocktail and N-dansyl-Tyr-Val-Gly were purchased from Sigma-Aldrich. Ampicillin sodium salt, carbenicillin, IPTG was purchased from Fisher Scientific. All other reagents were of the highest quality available from commercial suppliers.

2.2 Molecular Cloning

The DNA sequence for the hPHMcc gene was codon optimized for expression in E. coli and commercially synthesized by the GeneArt (www.geneart.com). Based on the sequence of human PAM (accession no. P19021), the coding region from amino acid 40 to 454 was supplied in the pGA4 vector. This coding sequence was amplified by PCR with pGA4 as a template (primers, forward 5’-CATGGGATCCATGAACGAATGCCTGGGTAC-3’, reverse 5’-
ATGAAGCTTTTAAATCGGAATATTCGCTTCCG-3’) and cloned into the pET vectors using the \textit{BamHI} and \textit{HindIII} restrictions site. Sequences of the cloned vectors were confirmed by DNA sequencing from MWG Operon. The recombinant plasmid was transformed into the different \textit{E. coli} strains (Novagen) for expression. Co-expression experiments were performed using the chaperonin Plasmid Set (Takara Bio USA) containing the plasmids pG-KJE8, pGro7, pKJE7, pG-Tf2 and pTf16. \textit{E. coli} Origami B(DE3) cells were transformed with pET32a-hPHMcc vector, made competent chemically, and the competent \textit{E. coli} Origami B(DE3)-hPHMcc cells were transformed with each plasmid possessing a chaperonin gene.

2.3 Culture Cultivation

An overnight culture of \textit{E. coli} bearing pET32a-hPHMcc was diluted 100 fold in LB medium supplemented with 100 \( \mu \text{g/ml} \) carbenicillin and 1\% glucose and cultured at 37 \(^\circ\text{C}\). The culture was cooled down to 10 \(^\circ\text{C}\), was induced with IPTG (final concentration = 100 \( \mu \text{M} \)) when the \( \text{OD}_{600} \) reached 0.6, and incubated for another 4 hrs. The cells were collected by centrifugation at 10,000 rpm for 10 min and stored at -80 \(^\circ\text{C}\).

2.4 Protein Purification

All the purification steps were performed at 4 \(^\circ\text{C}\). The cell pellet was resuspended in Buffer A (20 mM Tris, 500 mM NaCl, and 20 mM imidazole, pH 7.8) at to a concentration of 0.1 g/ml, with 1 mg/ml lysozyme and protease inhibitor cocktail
added according to the manufacturer's instruction. Resuspended cells were sonicated and the cell lysate collected by centrifugation at 16,000 \( \times \) g for 20 min. The pelleted cell lysate was resuspended with Ni-NTA resin equilibrated in buffer A for 2 hr. Resin was then loaded into a column and washed with 10 column volumes of buffer A. Next, the Ni-NTA column was washed with 10 column volumes of buffer B (20 mM Tris, 500 mM NaCl, 1M guanidine hydrochloride, and 60 mM imidazole, pH 7.8), and then with 20 column volumes of buffer C (20 mM Tris, 500 mM NaCl, and 60 mM imidazole, pH 7.8). Recombinant hPHMcc was eluted from the Ni-NTA column using a linear gradient of 60 to 140 mM imidazole in buffer C. The fractions containing hPHMcc activity was pooled and concentrated using YM-10 centriprep ultrafilter. The active fractions were concentrated to 0.5 mg/ml and stored at -20 °C.

2.5 Expression and Analysis by Mass Spectrometry

We evaluated the *E. coli* constructs for hPHMcc expression by Western blot analysis. After induction, 5 ml was withdrawn from the culture at 1 hr interval, centrifuged to collect the cells, lysed using lysozyme, and evaluated using SDS-PAGE. Expression of hPHMcc was confirmed by Western blot using an anti-PAM antibody and compared it to an uninduced control.

After SDS-PAGE, the Comassie stained protein band corresponding to the correct mass for hPHMcc or the hPHMcc fusion proteins was extracted from the gel and the protein digested using restriction grade trypsin. A Hybrid linear ion trap-Orbitrap
2.6 Activity Assay

hPHMcc activity was assayed using a modification of a published procedures [32,33]. Briefly, the reactions at 37 °C were performed in 100 mM MES pH 6.5, 5 mM ascorbate, 2 µM CuSO₄, 5000 U bovine catalase, 0.01 (v/v) % Tween 20, 1% (v/v) ethanol, and 20 µM N-dansyl-Tyr-Val-Gly. The reaction was quenched at desired time with addition of trifluoroacetic acid to a final concentration of 2% (v/v). The product of the reaction N-dansyl-Tyr-Val-α-hydroxy-Gly was converted to N-dansyl-Tyr-Val-NH₂ by incubating with potassium hydroxide and 50 mM EDTA to a final pH = 8.5. Kinetic studies were carried out in triplicate. The product was analyzed on HPLC and the initial velocity kinetic data analyzed using SigmaPlot.

3. Results and discussion

The hPHMcc gene was inserted into the pET vectors using BamHI and HindIII restriction site (Table S2, supplementary material) to generate tagged hPHMcc with a C-terminal His₆-tag. Soluble hPHMcc-His₆ was not achieved by using the GST, Nus A and DscB tag in BL 21(DE3), C-41(DE3), and C-43(DE3) E. coli strains even after varying parameters like induction temperature, induction time, and addition of Cu(II) and/or PHMcc.
substrates to the growth media. In all cases, (except the Trx and MBP tag) hPHMcc was produced as inactive and insoluble inclusion bodies. We were completely unsuccessful in generating active PHMcc-His$_6$ after solubilizing and refolding the protein, which had been purified from the inclusion bodies.

Soluble and active Trx-hPHMcc-His$_6$ was generated by inserting the hPHMcc gene into pET32a(+) vector to yield hPHMcc with a N-terminal Trx fusion and a C-terminal His$_6$-tag. Recombinant Trx-hPHMcc-His$_6$ was produced as insoluble inclusion bodies after transformation into *E. coli* BL 21(DE3) with the pET32a(+) derived vector despite many experiments to optimize growth conditions for the generation of soluble protein. Our attempts to solubilize and refold Trx-hPHMcc-His$_6$ or hPHMcc after proteolytic separation from Trx and the His$_6$ tag were unsuccessful.

One likely reason for our inability to express active hPHMcc-His$_6$ or Trx-hPHMcc-His$_6$ in the *E. coli* BL 21(DE3) cells is relatively high reducing environment of cytoplasm. The high cytoplasmic reducing environment inhibits disulfide bond formation, a critical factor for the proper folding and activity of PHMcc [30]. We overcame this problem with the expression of Trx-hPHMcc-His$_6$ in the Origami B(DE3) strain of *E. coli*. The cytoplasmic environment of the Origami B(DE3) cells is more oxidizing than that of the BL-21 cells due to mutations which inactivate glutathione reductase and thioredoxin reductase [34,35]. After optimization (Figure S1, supplementary material), we found that soluble Trx-hPHMcc-His$_6$ was expressed at low temperature (10 °C) and a relatively low concentration of 100 µM IPTG (Figure 2). Optimal expression required a short period of
IPTG induction (4 hrs) to avoid hPHMcc degradation to lower molecular weight Western positive fragments of low activity (data not shown).

During the optimization of expression and purification of Trx-hPHMcc-His$_6$, we observed a 65 kDa protein that co-purified with that 52 kDa Trx-hPHMcc-His$_6$. This unknown 65 kDa protein represented 25-30% of the total protein in our samples and remained constant using a number of different purification strategies, including size exclusion chromatography, ion-exchange chromatography, and affinity chromatography. The unknown 65 kDa protein was extracted from the SDS-PAGE gel, digested with trypsin, and the tryptic peptides analyzed by LC-MS/MS on a LTQ-orbitrap. The 65 kDa protein was identified as the GroEL-GroES complex, an ATP-dependent molecular chaperonin which assists protein folding in bacteria (Horwich et al., 2009). It has been proposed that GroEL/GroES/ADP complex binds with non-native protein through hydrophobic contacts, followed by protein folding within the GroEL/GroES/ATP cis folding chamber [36]. In addition to GroEL-GroES, other chaperonins, such as dnaK, dnaJ, grpE, and tig, have been reported to increase the solubility and yield of proteins which are difficult to express in bacteria [37]. In order to determine if coexpression of hPHMcc with these chaperonins increased the yield of soluble, active enzyme, plasmids containing the gene for these chaperonins were inserted in Origami B(DE3) cells which already possess the pET32a(+)–hPHMcc-His$_6$ plasmid. Unfortunately, we found was no appreciable increase in the amount of soluble hPHMcc upon coexpression with a chaperonin.
The GroEL-GroES complex was removed from hPHMcc using a nickel affinity resin. The partially purified Trx-hPHMcc-His_6 was loaded onto a column packed with Ni-NTA resin, the column was washed with guanidine hydrochloride to overcome the interactions between Trx-hPHMcc-His_6 and the GroEL-GroES complex. The bound Trx-hPHMcc-His_6 was then eluted from the Ni-NTA resin using a linear, increasing gradient of [imidazole]. Trx-hPHMcc-His_6 was ≥90% homogeneous by SDS-PAGE (Figure 3) and Western analysis using an anti-His_6 antibody confirmed the presence of the His_6 on the 52.4 kDa Trx-hPHMcc-His_6. The final yield of purified Trx-hPHMcc-His_6 was 0.1-0.2 mg per liter of culture and the purified Trx-hPHMcc-His_6 was active. The steady-state kinetic parameters for the Trx-PHMcc-His_6 catalyzed oxidation of N-dansyl-Tyr-Val-Gly to N-dansyl-Tyr-Val-hydroxyglycine are as follows: $K_M = 7.7 \pm 0.8 \mu M$, $V_{MAX} = 0.40 \pm 0.02 \mu moles/min/mg$, $k_{cat} = 0.35 \pm 0.02 \text{ s}^{-1}$, and $k_{cat}/K_M = 4.68 \pm 0.2 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ (Figure 4). The $K_M$ and $V_{MAX}$ values of N-dansyl-Tyr-Val-Gly for rat PHMcc, under similar experimental conditions, are reported as 7 µM and 0.5-6 µmoles/min/mg [30], which are similar to the values we measured for Trx-hPHMcc-His_6.

In other experiments (not shown), we found that both hPHMcc and MBP-PHMcc-His_6 were active when expressed in the E. coli Orgami B (DE3). The specific activity of MBP-PHMcc-His_6 was low, (0.12 µmoles/min/mg), and we did no further work with this hPHMcc fusion. We had engineered in thrombin cleavage site C-terminal to Trx and N-terminal to the His_6 tag to enable the thrombinolytic generation of hPHMcc from Trx-hPHMcc-His_6. Unfortunately, thrombin cleavage of the Trx-hPHMcc-His_6 unexpected cleaved within hPHMcc generating smaller PHMc-derived fragments. Experiments
varying the conditions for the thrombin cleavage of Trx-hPHMcc-His$_6$ always lead to the cleavage of hPHMcc and we were unable to completely purify full length hPHMcc away from the thrombin-derived fragments. Because the specific activity of the Trx-hPHMcc-His$_6$ fusion was in agreement with wildtype rat PHMcc for N-dansyl-Tyr-Val-Gly substrate, we did not pursue the expression of hPHMcc in *E. coli* any further.

4. Conclusion

We report here on the expression of a soluble, active form of the catalytic core of hPHMcc in *E. coli* [30]. The successful production of hPHMcc in *E. coli* will foster future mutagenesis studies to elaborate mechanistic details of PHM catalysis and new strategies for the inexpensive production of α-amidated peptides in *E. coli*. Co-expression of glycine-extended peptides and hPHMcc in *E. coli*, in combination with the methods to incorporate non-natural amino acids into α-amidated peptides [38], could rapidly lead to novel peptide analogs useful as therapeutics and research tools.

5. References


Figures:

**Figure 1.** The peptide amidation reaction. Glycine hydroxylation requires O\(_2\), Cu(II), and a reductant and is catalyzed by peptidylglycine \(\alpha\)-hydroxylating monooxygenase (PHM). Dealkylation of the \(\alpha\)-hydroxyglycine-extended peptide requires Zn(II) and Ca(II) and is catalyzed by peptidylamidoglycolate lyase (PAL).

**Figure 2.** Effect of induction temperature (24 °C and 10 °C) on expression of Trx-hPHMcc. Cells were collected at different time interval (0 - 5hrs) after induction and the cell lysate was used for the western blot analysis with the anti-PAM antibody.
Figure 3. SDS-PAGE Analysis hPHMcc Expression. Lane-1, molecular weight markers (with the molecular weight highlighted inside the bands); lane-2, purified hPHMcc inclusion bodies from affinity column; lane-3 purified Trx-hPHMcc with the GroEL-GroES complex; and lane-4, the purified Trx-PHMcc.

Figure 4. Steady-State Kinetic Analysis of N-Dansyl-Tyr-Val-Gly Hydroxylation.
### Supplementary Material:

Table S1.

Properties of the *E. coli* Strains and Plasmids Employed in Our Studies

<table>
<thead>
<tr>
<th><em>E. coli</em> Strains</th>
<th>Properties</th>
</tr>
</thead>
</table>
| BL 21 (DE3)       | The most common host for protein expression<br>
|                   | $F^{\text{ompT}}hsdSB(r_{B}^{-}m_{B}^{-})gal dcm$ (DE3) |
| C-41 (DE3)        | Genetically mutated to increase the tolerance against toxic protein<br>
|                   | $F-ompT hsdSB (rB- mB-) gal dcm$ (DE3) |
| C-43 (DE3)        | Derived from C-41 (DE3) to express different set of toxic proteins<br>
|                   | $F-ompT hsdSB (rB- mB-) gal dcm$ (DE3) |
| Shuttle T7 (DE3)  | Engineered *E. coli* strain to promote disulfide bond formation<br>
|                   | $fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal $att::pNEB3-r1-cDsbC (Spec<sup>R</sup>, lacI<sup>q</sup>) ΔtrxB sulA11 R(mcr-73::miniTn10--Tet<sup>S</sup>)<sup>2</sup> [dcm] R(zgb-210::Tn10 --Tet<sup>S</sup>) endA1 Δgor Δ(mcr-C-mrr)114::IS10 |
| Origami B (DE3)   | *E. coli* strain carries trxB/gor mutation to enhance disulfide bond formation<br>
|                   | $F^{\text{ompT}}hsdSB(r_{B}^{-}m_{B}^{-})gal dcm lacY1 aphC$ (DE3) gor522::Tn10 trxB (Kan<sup>R</sup>, Tet<sup>R</sup>) |

<table>
<thead>
<tr>
<th>Plasmid Vectors</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET21a(+)</td>
<td>C-terminal histidine tag</td>
</tr>
<tr>
<td>pET32a(+)</td>
<td>N-terminal thioredoxin tag, helps in catalyzing disulfide bond formation in the cytoplasm of trxB mutant</td>
</tr>
<tr>
<td>pET42a(+)</td>
<td>N-terminal glutathione-S-transferase (GST) tag to enhance the solubility of a target protein</td>
</tr>
<tr>
<td>pET43.1a(+)</td>
<td>Incorporates a solubility promoting Nus A tag</td>
</tr>
<tr>
<td>pmal-c2x</td>
<td>Maltose binding protein (MBP) fused to the N-terminal of the target protein expression</td>
</tr>
</tbody>
</table>
Table S2.
The Expression of hPHMcc in Different Strains of *E. coli* and Using Different Vectors

<table>
<thead>
<tr>
<th>Vector</th>
<th>BL 21(DE3)</th>
<th>C-41 (DE3)</th>
<th>C-43 (DE3)</th>
<th>Origami B(DE3)</th>
<th>T7 Shuttle</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pET21b(+)</strong></td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
</tr>
<tr>
<td><strong>pET32a(+)</strong> (Trx tag)</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Soluble Active Enzyme</td>
<td>Soluble Active Enzyme</td>
</tr>
<tr>
<td><strong>pET42a(+)</strong> (GST tag)</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>No Expression b</td>
<td>No Expression b</td>
</tr>
<tr>
<td><strong>pET43.1a(+)</strong> (Nus A tag)</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>No Expression b</td>
<td>No Expression b</td>
</tr>
<tr>
<td><strong>pET39b(+)</strong> (DscB tag)</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>Inclusion Bodies</td>
<td>No Expression b</td>
<td>No Expression b</td>
</tr>
<tr>
<td><strong>pMAL-c2X</strong> (MBP tag)</td>
<td>Inclusion Bodies</td>
<td>ND c</td>
<td>ND</td>
<td>Soluble Active Enzyme</td>
<td>ND</td>
</tr>
</tbody>
</table>

*E.coli* strains were obtained from the following commercial suppliers: T7 Shuttle (DE3) from New England Biolabs, BL 21 (DE3) and Origami B(DE3) from Novagen, (DE3) from New England Biolabs, and C-41 (DE3) and C-43 (DE3) from Lucigen.

bNo western blot positive bands were detected using the anti-PAM antibody.

cND = not done
Figure S1.

Optimization of hPMcc Expression

Protein expression was induced at 10 °C under these conditions (1) 0.1mM IPTG, (2) 0.1mM IPTG and 4 µM Ndansyl-Tyr-Val-Gly, (3) 0.1mM IPTG and 2µM CuSO₄, and (4) 0.1mM IPTG, 4µM N-dansyl-Tyr-Val-Gly, and 2µM CuSO₄.