2015

ELUCIDATING THE MECHANISM OF LIPL: A NON-HEME FE(II), α -KETOGLUTARATE: URIDINE-5'-MONOPHOSPHATE DIOXYGENASE

Anwesha Goswami
University of Kentucky, ago223@uky.edu

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Anwesha Goswami, Student

Dr. Steven G Van Lanen, Major Professor

Dr. Jim Pauly, Director of Graduate Studies

ELUCIDATING THE MECHANISM OF LIPL: A NON-HEME FE(II), α -KETOGLUTARATE: URIDINE-5'-MONOPHOSPHATE DIOXYGENASE

DISSERTATION

A dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in the
College of Pharmacy
at the University of Kentucky

By

Anwesha Goswami
Lexington, Kentucky

Director: Dr. Steven Van Lanen, Assistant Professor of Pharmaceutical Sciences
Lexington, Kentucky

2015

ABSTRACT OF DISSERTATION

ELUCIDATING THE MECHANISM OF LIPL: A NON-HEME FE(II), α -KETOGLUTARATE: URIDINE-5'-MONOPHOSPHATE DIOXYGENASE

Several nucleoside natural product antibiotics from *Streptomyces* sp. and actinomycetes have recently been shown to target bacterial peptidoglycan cell wall biosynthesis by inhibiting the bacterial translocase I (MraY). The biosynthetic gene clusters for A-90289, liposidomycins and caprazamycins revealed a protein with sequence similarity to proteins annotated as α-KG:taurine dioxygenases (TauD). This enzyme (LipL) is a mononuclear, non-heme, Fe(II) dependent α-keto glutarate (α-KG) :uridine monophosphate (UMP) dioxygenase responsible for the net dephosphorylation and two electron oxidation of UMP to uridine-5'-aldehyde. The postulated reaction coordinates involving the activation of the C-5' center in UMP and the corresponding formation of uridine-5'-aldehyde are modeled on extensive spectroscopic and structural characterizations of TauD. In this dissertation, the postulated radical mechanism for LipL involving the formation of an unstable hydroxylated intermediate is investigated via the characterization of a key product obtained from the reaction of LipL (and its homolog Cpr19) with a synthetically modified surrogate substrate where the bridging phosphoester oxygen in UMP is replaced with a 5' C-P bond. We further validate our hypothesis by analyzing the reactions of both LipL and Cpr19 with specifically $^{2}H_{1}$ – labeled UMP substrate and confirming the expected products via mass spectrometry. In addition, we explore substrate promiscuity of the enzymes and utilize a set of site specific mutants of Cpr19 as means of gaining better insight into the active site residues. Predictive models for Cpr19 and LipL structures are developed by the combination of experimental results and chemical logic.

KEYWORDS: nucleoside antibiotics, peptidoglycan cell wall, dioxygenases, iron-dependent enzymes

Anwesha Goswami

Student's signature

January 28, 2015

Date

ELUCIDATING THE MECHANISM OF LIPL: A NON-HEME FE(II), α -KETOGLUTARATE: URIDINE-5'-MONOPHOSPHATE DIOXYGENASE

By

Anwesha Goswami

Dr. Steven Van Lanen

Director of Dissertation

Dr. Jim Pauly

Director of Graduate Studies

January 28, 2015

Date

DEDICATION

This work is dedicated to my grandparents and my furrever friends, whom I hope to someday meet across the Rainbow Bridge.

ACKNOWLEDGEMENTS

This dissertation and body of research was an incredible learning experience for me, and a cumulation of all the support and guidance that I gained from a great number of people. I would like to begin by extending my heartfelt and sincerest gratitude to Dr. Steven Van Lanen for his continued support, guidance and encouragement and his willingness to provide a refuge for me in his lab despite initial difficulties. I am especially grateful for the amount of freedom I was given to come up with ideas for my projects, and the platform to validate those with smartly designed and effective experiments under his mentorship. A special thanks to Dr. Rohr for co-chairing my committee, for providing key suggestions for my research projects and presentations throughout graduate school, and for my overall professional development. I would like to thank the remaining members of my committee: Dr. Kyung-Bo Kim and Dr. Rolf Craven for your critical and helpful suggestions, and for taking the time to oversee the completion of my dissertation thesis. I want to extend my gratitude to Drs. Joe Chappell, Jon Thorson, Sylvie Garneau-Tsodikova and Oleg Tsodikov for becoming a part of the extended Natural products family at UKCOP and for expanding the breadth of our group by participating, providing constructive suggestions and overseeing our research on a regular basis at our meetings. Many thanks to Dr. Luke Moe for agreeing to serve as my external examiner.

I need to thank all my friends and coworkers who have been my teachers in different aspects of my life in graduate school. A very special thanks to Catina Rossoll for *always* patiently helping with any issues that I could come up with, and to Todd Sizemore for constantly preventing major disasters in the lab. I attribute everything I have learnt in the lab to Drs. Nidhi Tibrewal, Lalit Sharma, Pallab Pahari, Khaled Shaaban, Xiuling Chi, Wenlong Cai, Xiaodong Liu, Ashley Arlinghaus and Shanteri Singh. The work presented

here would not have been possible without help from each and every one of them. I am immensely thankful to Wenlong for his help with key parts of my research. And special thanks to Theresa, Stevi, Nilay, Jhong-Min, Vidya, Eleftharia, Ore, Ryan, Matt, Tyler, and Tyler (Bucci) for always being available for a friendly conversation for those times that warranted one!

Above all, I want to acknowledge my family: my parents, my brothers (Sourya and Samya) and Kakai for motivating me to follow my ambitions and expanding my horizons. Despite her conventional upbringing and her own limiting situations, my mother (Susmita) worked very hard through odds to make sure I worked towards becoming independent and have strong career goals. I want to take thank her wholeheartedly, and hopefully my small achievements would validate her sacrifices in some way.

My biggest thanks to my friends who provided a home away from home in Kentucky: my loving 'gazette folks' namely Suvid, Apurv, Sweta, Sumesh, Abhijit, Nikhil, Suraj, Prachi, Pratik, Sonam, Raghava, Rutooj, Saket, Sai, Shristi and Amit. Thank you guys for a ton of memories! That said, I am left to acknowledge the people who are the very essence of my support system: Devagnik, Chiquita, Debarati, Anwesha C, Sayan, Souvik, Soumya, Shaunak, Yokesh, Abhra, Alam, Rohan da and Buli pishi -- simply because I love you guys the most! A final thanks to my furry companion, Billi, for riding along with the most generic name on earth, for her fuzzy aloofness and her signature feline worldly indifference that helped smooth over many a difficult day.

Go Wildcats!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

amp ampicillin

ATP adenosine triphosphate

B. Subtilis Bacillus subtitles

BLAST basic local alignment search tool

bp base pairs

CIP calf intestinal phosphatase

CMP cytidine monophosphate

COSY correlation spectroscopy

CTP cytidine triphosphate

DAPdiaminopimelic acid

DNA deoxyribonucleic acid

2D-NMR two-dimensional nuclear magnetic resonance spectroscopy

DMF dimethylformamide

DMSO dimetylsulfoxide

EDTA ethylenediaminetetraacetic acid

E. coli Escherichia coli

GlcNAc *N*-acetylglucosamine

hr hour

His6 hexahistidine

HMBC heteronuclear multiple-bond correlation spectroscopy

HPLC high performance liquid chromatography

HSQC heteronuclear single-quantum correlation spectroscopy

IPTG isopropyl-β-D-thiogalactopyranoside

k kilo

Kan kanamycin

kb kilo base pairs

kcat turnover rate

kDa kilo Dalton

α-KG α -ketoglutarate

Ki inhibition constant

KmMichaelis-Menten constant

LB Luria broth

LC-MS Liquid chromatography–mass spectrometry

MDR multiple drug resistant

min minute

MraY phospho-MurNAc-pentapeptide translocase

MRSA Methicillin-resistant Staphylococcus aureus

MurNAc *N*-acetylmuramic acid

MW molecular weight

NADP+nicotinamide adenine dinucleotide phosphate

NADPH reduced form of NADP+

nm nanometer

NME new molecular entities

NMR nuclear magnetic resonance spectroscopy

NTP nucleoside triphosphate

NDP nucleoside diphosphate

OD optical density

ORF open reading fram

PAGE polyacrylamide gel electrophoresis

PBPs penicillin binding proteins

PCR polymerase chain reaction

PEP phosphoenolpyruvic acid or phosphoenolpyruvate

PRPP synthetase phosphoribosylpyrophosphate synthetase

S. aureus Staphylococcus aureus

S. lividans Streptomyces. lividans

sec second

SHMT Serine hydroxymethyltransferase

SDS sodium dodecyl sulfate

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

sp. Species

sp. Streptomyces

TB tuberculosis

TFA trifluoroacetic acid

UDP-Glc*N*Ac5-diphospho-*N*-acetylglucosamine

UDP-Mur*N*AcUDP-*N*-acetylmuramic acid

UMP uridine monophosphate

UV/Vis ultraviolet/visible

XMP xanthosine 5'-monophosphate

Chapter one: Introduction and Background

1.1. Natural Products – Significance

Natural products and their derivatives have historically been an invaluable source for the discovery of novel therapeutics[1]. When applied to drug discovery prior to the advent of high-throughput screening and the post-genomic era, more than 80% of drugs were either natural product derived or their semisynthetic versions [2]. In the last ~30 years, amongst all clinically approved drugs, almost half of the small-molecule new chemical entities introduced were either natural products, semi-synthetic natural product analogues or synthetic compounds based on natural products [1]. Even in the context of current therapeutics, the primary reasons why natural products continue to be important sources are: a) most of the currently available classes of drugs either contain natural products or have these as original leads, i.e. there are no known synthetic substitutes for these complex molecules [3] b) they continue to inspire synthetic, semisynthetic, and chemo-enzymatic efforts to replicate and diversify complex functional scaffolds [4], c) they are invaluable tools for deciphering complex metabolic pathways and the associated unique chemical machineries [5], and d) there is still immense potential for the discovery of novel therapeutics from unexplored sources [6, 7]. The use of natural products as sources of therapeutics have conferred unparalleled benefits to mankind, by significantly increasing the average lifespan of the population, nearly eliminating some infectious diseases and by exerting control over several neoplastic and viral diseases [8]. As such, natural products also continue to represent a significant share of the current drug market [9].

Despite their historical importance, natural product discovery has been subjected to a slow decline over the last two decades prompted by a number of scientific and economic challenges [8] calling into question why natural products still matter in the pharmaceutical landscape. The increasing resistance to antibiotics, coupled with decreased efficiency in research, failure in the productivity of classical screening methods, and decreasing profits to pharmaceutical companies due to regulatory obstacles and increasing research expenses, have cumulatively contributed to a shift in the research paradigm [10, 11]. Whilst most large pharma have shifted their research efforts from natural products discovery to the more profitable drug candidates meant to treat chronic diseases, research efforts in this field have greatly increased in academia and specialized pharmaceutical/biotechnological companies with renewed approaches to improved screening and greater emphasis on developing them to front-line drug candidates [12, 13]. As a result, emerging trends in addition to unrealized expectations from current research and development strategies are prompting a renewed interest in natural products as a source of biochemical diversity and lead generation.

1.2. Need for new antibiotics

Every antibiotic that is introduced for clinical use has a limited shelf life, due to innate or acquired mechanisms of resistance present in all bacteria. Consequently, the need for new antibacterial drugs for clinical use is a constant one [14]. Resistance has developed to all main classes of antibiotics, both natural and synthetic, varying in timeline between

different classes of drugs after their first clinical use, and many infectious diseases have emerged and reemerged worldwide with changes in environmental parameters as well as with migration in populations [15]. Overuse is a primary driver of antibiotic resistance, but it is not just in humans – "nontherapeutic" treatment of farm animals and livestock with low doses of these drugs to promote growth and prevent diseases has also been criticized as a controversial practice that can potentially introduce antibiotic-resistant bacteria into the human food supply [16].

Pathogens that are resistant to multiple drugs continue to emerge around the globe, leading to the ongoing, cyclical need for innovation to combat infectious diseases caused by multiple drug resistant (MDR) pathogens [17]. According to the Centers for Disease Control and Prevention's 2013 Threat Report [18] approximately 23,000 people are killed each year by infectious diseases in the United States alone. Many more die from complications from other conditions (patients undergoing chemotherapy, dialysis for renal failure, surgery, organ transplantation, etc.) exacerbated by infections from resistant pathogens. The CDC Report estimates an annual expenditure of more than $20 billion arising from these maladies. Of these, almost 11,000 deaths were from methicillin-resistant *Staphylococcus aureus* (MRSA) infections in the United States, while other diseases like multidrug-resistant and extensively drug-resistant tuberculosis (MDR and XDR TB) are an increasing threat outside of the United States. In 2012, there were 8.6 million new TB cases globally with a reported 1.3 million deaths, and of them, an estimated 450,000 people were reported to have acquired MDR-TB [19].

These statistics are grim, and they are made even more so by the fact that in the last 45 years only five new classes of FDA-approved antibiotics have been introduced to the market [12, 17, 20]. The dramatic decreases in FDA-approved systemic antibiotics, caused in part to due to bacterial resistance, regulatory disincentives, and a big innovation gap has led us to an age where no new molecular entities are currently in phase III [20]. To revitalize the discovery of new antibiotics is imperative, and the issues associated with drug resistances and the current drug pipeline means that new compounds with novel modes of action and/or new targets are of great importance in the continued fight against infectious diseases.

While traditional antibiotic discovery has been based on cell-growth inhibition assays followed by identification of the targets, contemporary drug discovery is based in large part on the screening of small molecules for their ability to bind or otherwise inhibit specific macromolecular targets [1]. Historically, targets for antibiotic action have been classified into four major groups: bacterial cell wall biosynthesis, protein biosynthesis (translation), DNA replication and storage, and folate coenzyme biosynthesis [15]. Research in our lab has been focused on the discovery and identification of potential inhibitors to a target enzyme via activity-based high throughput enzyme assays, more specifically, inhibitors to the enzyme bacterial translocase I, a key enzyme that participates in the biosynthesis of the peptidoglycan cell wall biosynthesis in bacteria. The

overarching goal of our research is to investigate the biosynthetic pathway of these new classes of inhibitors/antibiotics focusing on the combined chemical logic and enzymatic machineries involved.

1.3. Biosynthesis of the peptidoglycan cell wall

Peptidoglycan is the primary polymeric constituent of bacterial cell walls and is essential for the survival of all bacteria [21]. The backbone of peptidoglycan consists of an alternating β-1,4-linked glycan composed of *N*-acetylglucosamine (Glc*N*Ac) and N-acetylmuramic acid (Mur*N*Ac) (**Figure 1.3.1**). A short polypeptide of five amino acids is attached to the 3-position of the Mur*N*Ac sugar, which allows the cross linking between the 3-amino acid (lysine or D-aminopimelate) and the 4/5 peptide bond of another polymeric unit. These crosslinks provide the structural integrity that allows the cell wall to withstand the osmotic pressure of the cytoplasm. The assembly of the cell wall and peptidoglycan biosynthesis is a complex process that begins at the cytoplasmic side of the cell membrane. The entire assembly can be broken down into three distinct stages: 1) polymerization of the disaccharides and attachment with the polypeptide chains, 2) initiation of the lipid linked cycle wherein the sugar activated hydrophilic precursor is attached to a lipid carrier and consequently flipped outwards of the membrane, and 3) the cross-linking of the polymeric backbone to establish the final structure of the cell membrane.

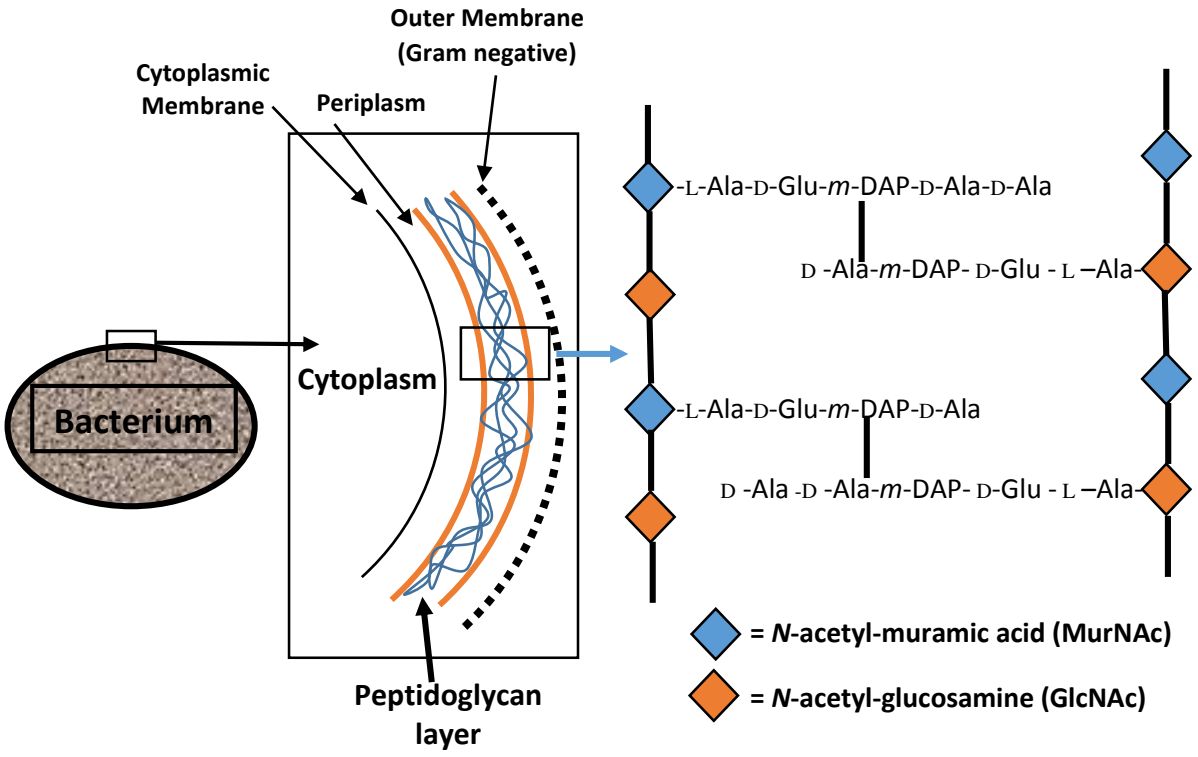


Figure 1.3.1. Structure of peptidoglycan. The backbone of peptidoglycan constructed by alternating β-1,4-linked glycan composed of *N*-acetylglucosamine (Glc*N*Ac) and N-acetylmuramic acid (Mur*N*Ac), which in turn are cross-linked via their polypeptide chains.

The biosynthesis is carried out by 12 ubiquitous enzymes found in both Gram-positive as well as Gram-negative bacteria, some of which still remain to be characterized in detail [22] (**Figure 1.3.2**). In the initial cytosolic stage, the uridine-5'-phosphate (UDP)-sugars are the biosynthetic starter molecules for all cell wall components and the process begins with the transformation of the activated UDP-Glc*N*Ac to UDP- Mur*N*Ac (catalyzed by MurA and MurB), followed by attachment of a series of amino acids leading to the polypeptide chain (L-Ala-γ-D-Glu-X-D-Ala-D-Ala where X is either L-Lys or D-aminopimelate). The addition of the polypeptide chain is cumulatively carried out by a set of ATP-dependent ligases Mur C-F. The resultant UDP-Mur*N*Ac-pentapeptide is then used

to provide phospho-Mur*N*Ac-pentapeptide that is transferred to a membrane bound undecaprenyl phosphate by the enzyme translocase I (phosphor-MurNAc-pentapeptide translocase) or MraY. Addition of another Glc*N*Ac sugar onto the 4'-OH of Mur*N*Ac (catalyzed by the glycosyltransferase MurG) creates lipid intermediate II, which in turn, is flipped outwards from the cytosolic side of the cell membrane presumably by a 'flippase' protein [23]. In some Gram-positive bacteria, additional amino acids are latched onto lipid intermediate II from amino-acyl-tRNA donors (for examples, five Gly residues are added on by FemABX in *Staphylococcus aureus* (26)). On the cell surface, lipid intermediate II is polymerized via transglycosylation to afford a glycan polysaccharide which is then cross-linked via transpeptidation by the penicillin binding proteins (PBPs) [24]. Transglycosylation relieves undecaprenyl pyrophosphate which can then be recycled via enzymatic dephosphorylation [23].

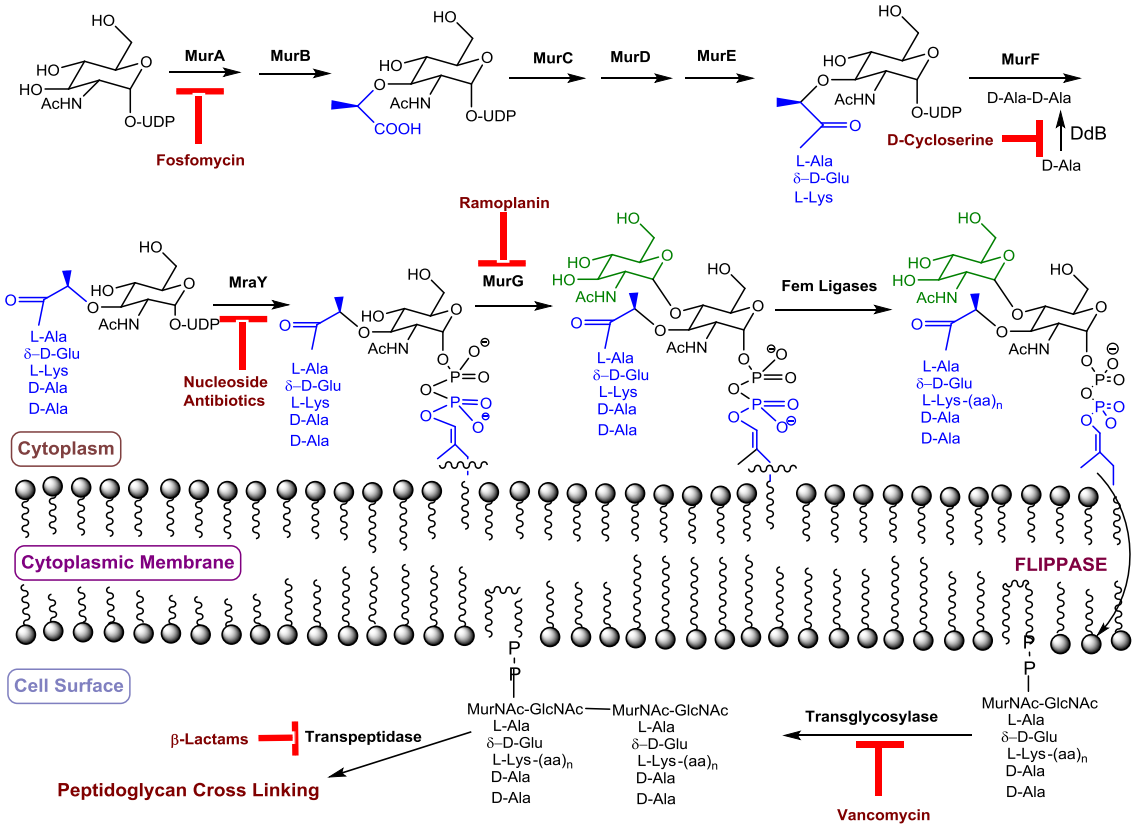


Figure 1.3.2. Peptidoglycan assembly. The reactions catalyzed by each enzyme is shown in **blue**, and the natural products inhibiting specific steps in the pathway are highlighted in **red**. Ramplanin can also bind to the product of MurG (Lipid II), while Vancomycin inhibits the activity of transglycosylases by binding to the substrate.

Since most of the enzymes that partake in the biosynthesis of peptidoglycan lack mammalian homologs, targeting cell wall biosynthesis has been proven successful as traditional means for inhibiting bacterial survival and growth. Of the cytosolic stages, only two of the enzymes are targeted by natural products fosfomycin and D-cycloserine [21], whereas MurG is inhibited by cyclic peptides ramoplanin and enduracin [25]. Some antibiotics like bacitracin target the lipid carrier itself, by binding irreversible to undecaprenyl phyrophosphate [26]. The vancomycin group of glycopeptide antibiotics inhibit the transglycosylation of lipid intermediate II, by binding to the substrate itself

thereby preventing its interaction with the transglycosylase [27], and the β-lactams irreversibly inhibit the final transpeptidation steps [24]. Natural product inhibitors have been found for six of the biosynthetic steps, of which five have been tested clinically. Although several have been successful as antibacterial antibiotics, many of them are plagued with issues of resistance within years of clinical introduction. The inhibition of the enzyme translocase I as a means of antibiotic activity, however, has not been fully realized for clinical purposes yet, and therefore represents a unique frontier for discovery of novel antibiotics.

1.4. MraY – Structure and Function

MraY (phosphor-Mur*N*Ac-pentapeptide translocase) is an integral membrane enzyme responsible for the second stage of peptidoglycan biosynthesis, in that, it catalyzes the transfer of phosphor-Mur*N*Ac- pentapeptide from UDP-Mur*N*Ac-pentapeptide to the lipid carrier undecaprenyl-pyrophosphate, generating undecaprenyl-pyrophosphoryl-Mur*N*Ac-pentapeptide or Lipid intermediate I [28] (**Figure 1.4.1**). This catalytic step is Mg^{2+}-dependent and essential for all bacterial viability, and therefore a promising pharmacological platform for the development of new classes of antibiotics [29]. The transferase activity of MraY was first discovered in 1965 by Neuhaus and coworkers [30], although the gene for MraY remained unidentified until 1991 [29] when it was overexpressed in *E. coli* and implicated for its putative transferase activity.

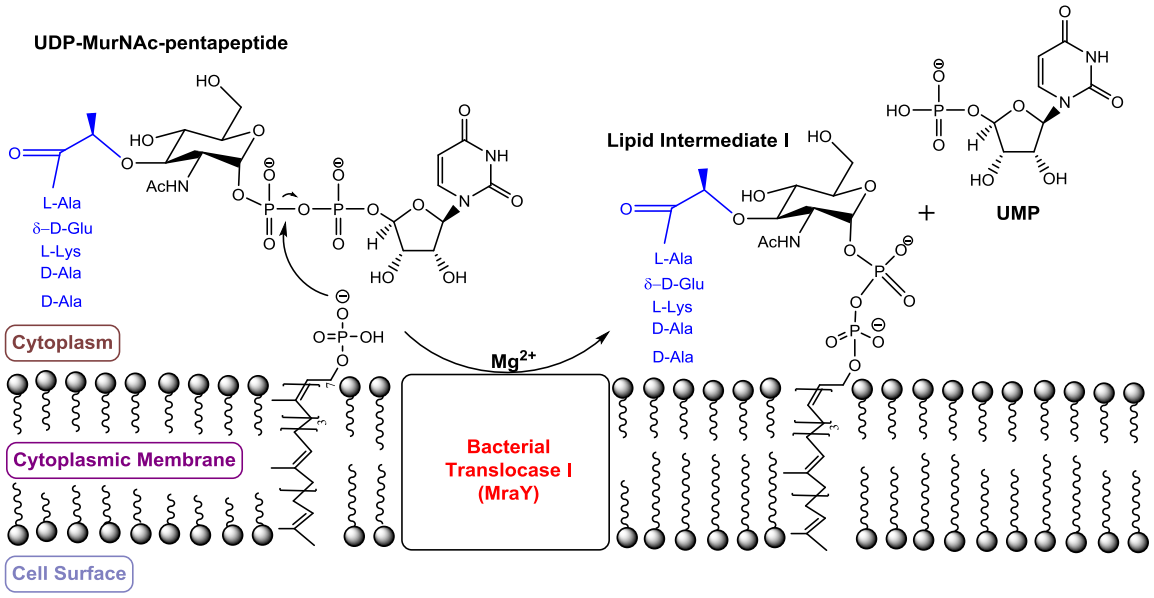


Figure 1.4.1. Reaction catalyzed by MraY. MraY is a transmembrane protein responsible for initiating the lipid-linked cycle of peptidoglycan biosynthesis.

MraY belongs to a subfamily of the polyprenyl-phosphate *N*-acetyl hexosamine 1-phosphate transferase (PNPT) superfamily of enzymes, that also includes enzymes responsible for the synthesis of cell envelope polymers like the O-antigen and teichoic acid in bacteria, and the GPT (UDP-Glc*N*Ac:dolichol-P Glc*N*Ac-1-P transferase) enzyme family responsible for N-linked glycosylation in eukaryotes [31]. *E. coli* MraY has been overexpressed, extracted into detergent micelles, and characterized kinetically using a continuous fluorescence assay, but could not be purified to homogeneity (35). *B. subtilis* MraY has also been purified in small quantities, although the specific activity is reduced substantially upon purification (36). Earlier studies conducted with MraY isolated from *M. luteus* [30] and *S. aureus* [32] in conjunction with the MraY from E. coli helped establish it as a transmembrane protein indispensable for cell survival [33]. Studies conducted in *E.*

coli involving mutational inactivation of the *mraY* gene led to growth inhibition and lethal phenotype, typically observed from similar inactivation of genes involved in peptidoglycan biosynthesis [34]. In addition to gram-negative *E. coli*, the *mraY* gene has also been shown to be essential for survival in gram-positive bacteria *Streptococcus pneumonia,* and bioinformatics studies have been conducted to suggest that only a single copy of the essential *mraY* gene is encoded per genome based on all currently available sequenced microbial genome [23]. Isotope enrichment experiments conducted with *S. aureus* translocase I suggest a two-step reaction involving a proposed S_N2 type nucleophilic substitution mechanism via the formation of an intermediate [30]. The structural basis of enzyme function was elusive for a long time in the absence of structural information, until recently Lee and coworkers published the crystal structure of MraY from *Aquifex aeolicus* at 3.3 Å resolution [35] (**Figure 1.4.2**).

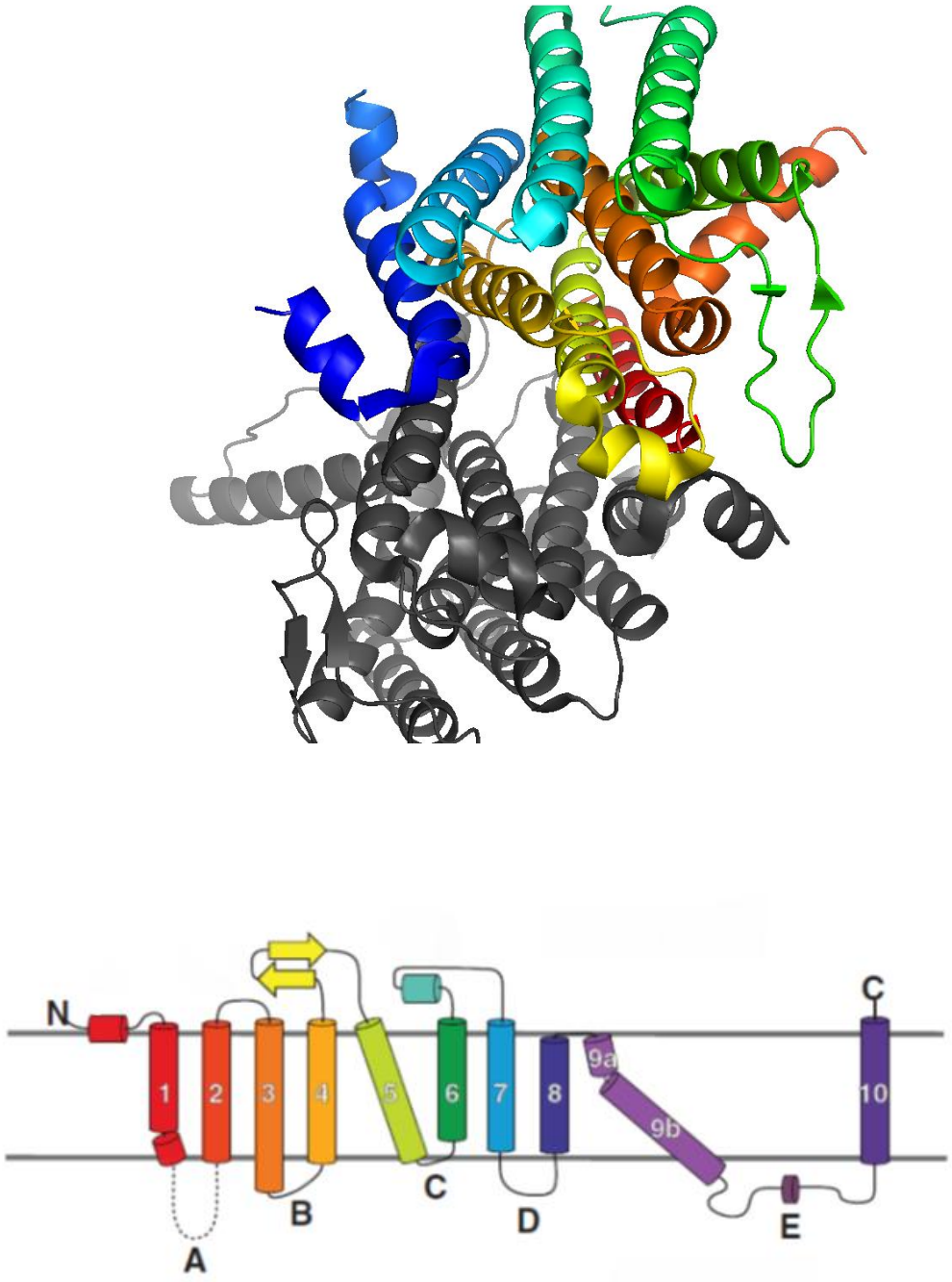


Figure 1.4.2. X-ray crystal structure of MraY. The recently resolved structure of MraY shows that it exists as a dimer with ten transmembrane helices with a visible tunnel at the center of the dimer. Schematic for lateral view (bottom)

MraY from *Aquifex aeolicus* ($MraY_{AA}$) has been shown to crystallize as a dimer, generating an oval-shaped tunnel at the center of the dimer [35]. The tunnel is predominantly surrounded by hydrophobic amino acid residues and is large enough to presumably accommodate lipids. Each subunit is comprised of 10 transmembrane helices, an interfacial helix, a periplasmic β hairpin, a periplasmic helix and five cytoplasmic loops. Mapping of conserved sequences onto the crystal structure revealed the highest conservation localization around a cleft formed by the cytosolic and inner-leaflet membrane regions of four of the transmembrane helices. Recent mutational studies conducted in *B*. *subtilis* MraY established 14 invariant charged amino acid residues, essential for enzyme activity [36] and most of these residues can be found in this 'cleft' region suggesting that this region serves as the active site. Three aspartate residues (Asp^{117}, Asp^{118} and Asp^{265}) each found on a cytoplasmic loop, are strictly conserved throughout the entire PNPT family and two invariant histidine residues (His^{324} and His^{325}) in the MraY family have been shown to be catalytically important by mutational inactivation, wherein, mutation of any of the three aspartate residues and one of the histidine residues (His^{324}) resulted in complete loss of activity. Based on these results, it is proposed that Asp^{117} and Asp^{118} in $MraY_{AA}$ may be involved in Mg^{2+} coordination, and Asp^{265} may be a possible active site nucleophile used in the formation of a covalent enzyme-phospho-Mur*N*Ac-pentapeptide intermediate (**Figure 1.4.3**).

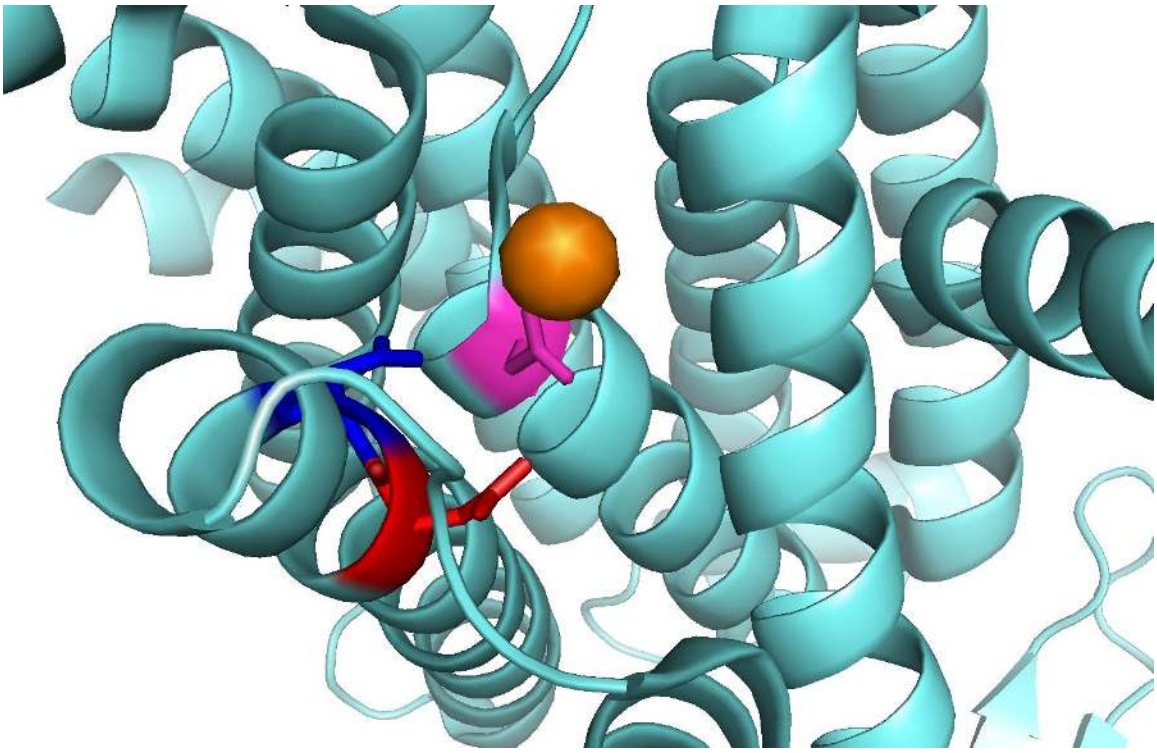


Figure 1.4.3. Proposed catalytic site of MraY: Asp 116 (**red**), Asp 117 (**blue**) and Asp 265 (**pink**) are shown here along with bound Mg^{2+} ion (**orange**).

1.5. Inhibitors of MraY

MraY has long been known as a promising target for the development of new antibiotics because it is the target of several different classes of natural product inhibitors with antibacterial activity as well as bacteriolytic lysis protein E from bacteriophage ϕX174 [36, 37]. Protein E (a 91-amino acid polypeptide) is another integral membrane protein encoded by DNA phage ϕ-174 that leads to bacteriolysis via an unclear mechanism. Mutational inactivation of Phe^{288}, found near the extracellular face of the membrane was shown to abolish protein activity [36]. An 18-residue polypeptide with the wild-type sequences from Protein E was shown to possess the minimal requisites for lysis of host

cells [38], whereas in another study, a 37-amino acid polypeptide containing the transmembrane domain of protein E was shown to specifically bind and inhibit MraY [39]. Another category of MraY inhibitors is the lipopeptides amphomycin, friulimcin and glycinocins [29]. These cyclic molecules are known to complex with undecaprenyl phosphates in the presence of Ca^{2+}, in a mechanism reminiscent of the glycopeptides like vancomycin. Amphomycin is, therefore, a very potent inhibitor of Gram-positive bacteria like *streptococci* and *enterococci* [40].

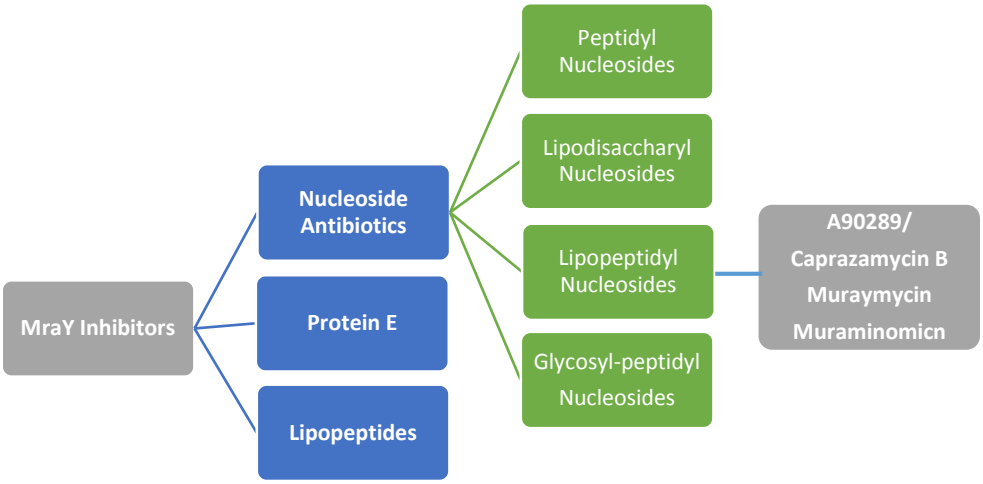


Figure 1.5.1. Types of MraY Inhibitors.

To date, the largest reported group of MraY inhibitors is the nucleoside antibiotics, and their activities are typically associated with the presence of key nucleoside components in these molecules (**Figure 1.5.1**). They can be classified into four structural groups: 1) the peptidyl nucleosides represented by pacidamycin from *Streptomyces coeruleorubidus* [41] and mureidomycin from *Streptomyces flavidovirens* [42], 2) the lipodisaccharyl nucleosides represented by tunicamycins from *Streptomyces lysosuperificus* [43], 3) the

lipopeptidyl nucleosides represented by A-90289 from *Streptomyces* sp. SANK 60405 [44] and caprazamycin from *Streptomyces* sp. MK730 -62F2 [45] and 4) the glycosyl-peptidyl nucleosides represented by the capuramycins A-500359s from *Streptomyces griseus* SANK 60196 [46], A-503083s from *Streptomyces* sp. SANK 62799 [47] and A-102395 from *Amycolatopsis* sp. SANK 60206 [48]. Despite several notable structural variations between the four groups, all of them share a key feature shown to be critical for their biological activities: the presence of a nucleoside component that consists of high-carbon furanoside, wherein the typical ribosyl component is replaced by a hexofuranoside(C6), a heptafuranoside (C7), or in the case of tunicamycins, an undecafuranoside (C11) (**Figure 1.5.2**) [49].

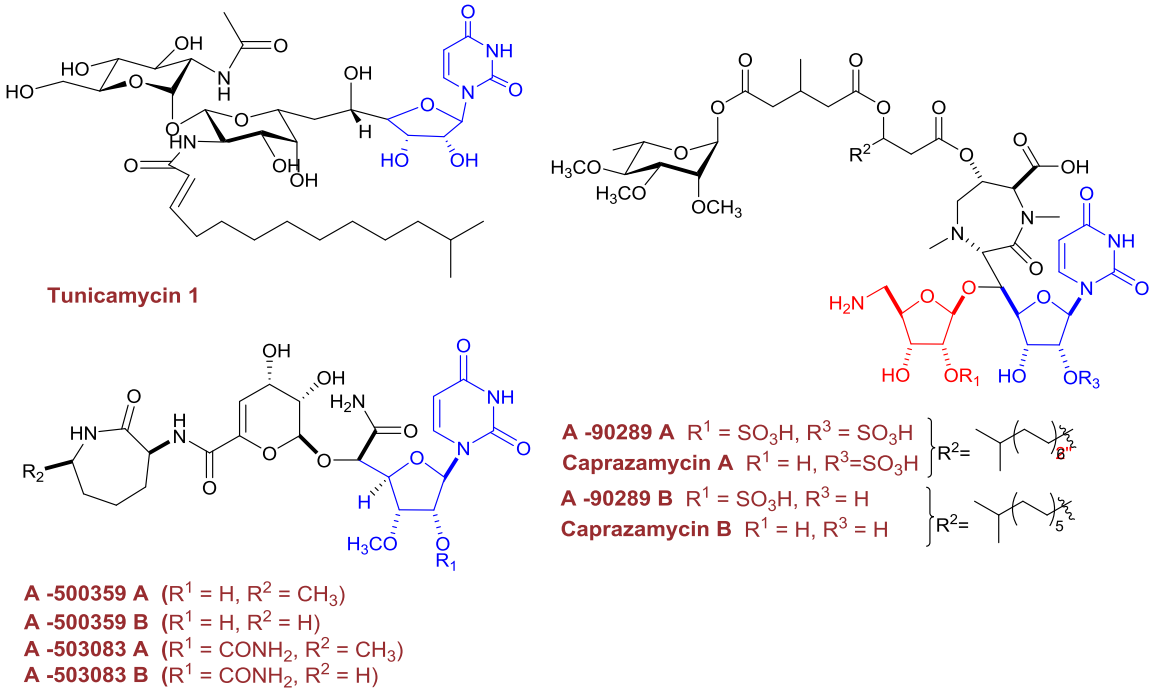


Figure 1.5.2. Representative nucleoside antibiotics. The key structural features are highlighted in these compounds. The uracil component found in all of them is highlighted in **blue**, and has been shown to be critical for activity in all of these inhibitors.

The variations in the structural components of each of these groups are presumably responsible for the slight differences in their specific mechanism of inhibitions. The peptidyl nucleosides share a typical 3'-deoxyuridine nucleoside attached via a 4',5'-enamide linkage to an *N*-methyl 2,3-diaminobutyric acid (DABA) residue, to which are attached the amino acid residues typically found in this group. Mureidomycin, a representative of this nucleoside type has been shown to be a competitive inhibitor against UDP-Mur*N*Ac-pentapeptide [42, 50]. The tunicamycins contain an additional GlcNAc moiety, a unique 11-carbon aminodialdose sugar or tunicamine [51] and an amide-linked fatty acid attached to the tunicamine sugar. As such, they can act as structural mimics of the diphospho-residue of UDP-Mur*N*Ac-pentapeptide leading to

competitive inhibition of MraY. Tunicamycins are additionally known for inhibiting the GPT (UDP-Glc*N*Ac:dolichol-P Glc*N*Ac-1-P transferase) enzyme family responsible for N-linked glycosylation in eukaryotes, and are therefore poor pharmacological candidates due to toxicity [52, 53]. On the other hand, the nucleoside inhibitors A-90289s, caprazamycins, and the muraminomicins, muraymycincs all contain uridine, aminoribose, diazepanone and fatty acyl moieties, with the major structural differences in the presence or absence of their sugar appendages (**Figure 1.5.3**). Structure-activity relationship studies using simplified synthetic analogues of these compounds have shown the uridyl and aminoribosyl moieties as critical for optimal antibiotic activity [54, 55]. The capuramycins A-500359s, A-500358s and A-102395 however lack the aminoribose, and instead contains an unique caprolactam moiety which has been shown to be critical to its activity [46, 56]. Despite these structural differences, all of these nucleoside antibiotics have been shown to selectively inhibit MraY-catalyzed transferase activity. The overarching aim in our lab has been the identification and characterization of the core structural components responsible for inhibitory activity.

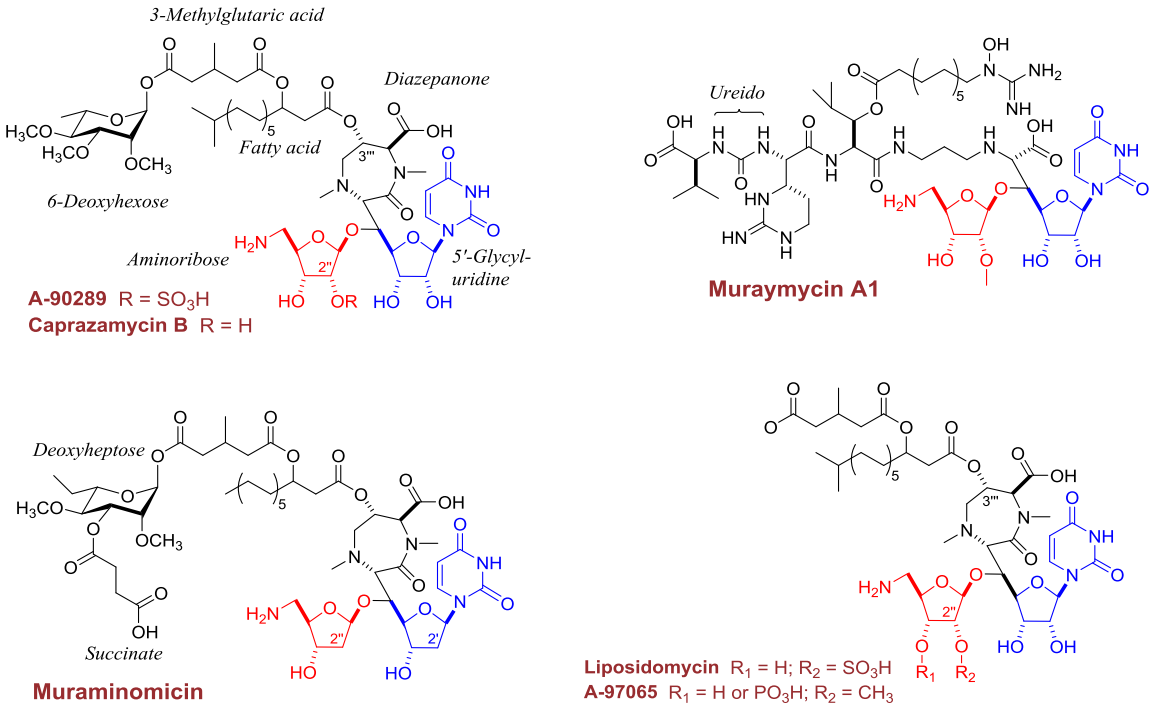


Figure 1.5.3. Structurally related nucleoside antibiotics. Structural variations within the liposidomycins are manifested by components in their side chains, while the core uracil and aminoribose components are present in all of them.

1.6. Current understanding of the biosynthetic pathways

As touched upon in the earlier sections, most of the nucleoside antibiotics contain highly modified sugar nucleosides which are usually modified at the C-5' of the parent ribose to generate furanosides containing 6-11 carbons. Initial isotopic enrichment studies using different high-carbon sugar nucleosides as models led to the realization that the glycosidic bond with the nucleoside base is established prior to C-5' modification, i.e. the nucleoside is the starting precursor for direct C-C bond formation with other precursors (discussed in the following section) as the other carbon source(s) [57-60]. The biosynthetic gene

clusters for several of these compounds, including the capuramycin-type antibiotics A-500359s [61], A-503083s [62], liposidomycins [63], A-90289s [44], caprazamycins [64], muraymycins [65], muraminomicins [66] and tunicamycins [67, 68] have been cloned and sequenced. With the sole exception of tunicamycin, all the aforementioned gene clusters have been shown to contain a shared *orf* that encodes a protein with sequence similarity to proteins annotated as α-ketoglutarate: taurine dioxygenases (TauD) [69], which catalyzes the conversion of taurine to aminoacetaldehyde and sulfite in *E. coli*, as means of sulfite scavenging. Our current understanding of a *proposed* divergent pathway beginning with a common precursor uridine-5'-monophosphate is illustrated in **Figure 1.6.1** [70, 71].

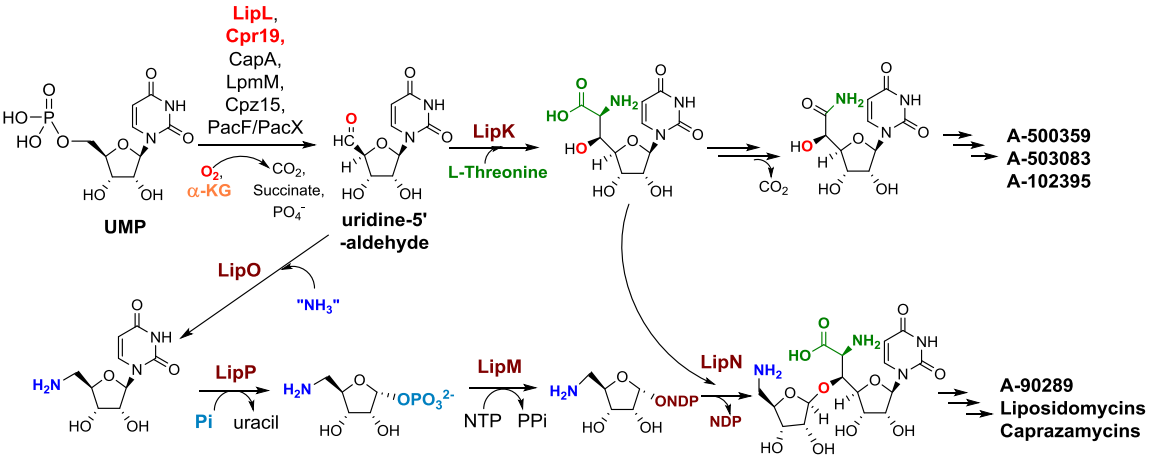


Figure 1.6.1. Biosynthetic pathway towards nucleoside antibiotics. Functional assignment of enzymes shown: (i) LipL: Fe(II): α-KG dioxygenase, (ii) LipK: L-threonine: uridine-5'-aldehyde transaldolase, (iii) LipO: L-methionine: uridine-5'-aldehyde aminotransferase, (iv) LipP: 5'-amino-5'-deoxyuridine phosphorylase, (v) LipM: UTP: 5-amino-5-deoxy-α-D-ribose-1-phosphate uridyltransferase, (vi) LipN: 5-amino-5-deoxyribosyltransferase.

Cumulative characterization and functional assignment of the genes from A-90289 led to the delineation of its biosynthetic pathway and elucidation of possible shared intermediates leading to the capuramycins, caprazamycins and all other related molecules. Following the generation of uridine-5'-aldehyde by LipL, the pathway bifurcates – in one route, enzyme LipK (a L-threonine: uridine-5'-aldehyde transaldolase) installs a 5'-C-glycyluridine unit via a pyridoxal-5'-phosphate (PLP) dependent reaction. The resultant intermediate can serve as a template for further modifications (involving a proposed decarboxylation event) leading to the capuramycin type antibiotics. The second route involves a cascade of reactions resulting in the generation of the functionally crucial aminoribose, which is then condensed with the intermediate produced by LipK by the enzyme LipN, a 5-amino-5-deoxyribosyltransferase. Kaysser et al. [64] conducted a series of gene deletion experiments resulting in the functional assignment of the enzyme Cpz21 as an acyltransferase responsible for installing the 3-methylglutaryl moiety in caprazamycins (**Figure 1.6.2**)

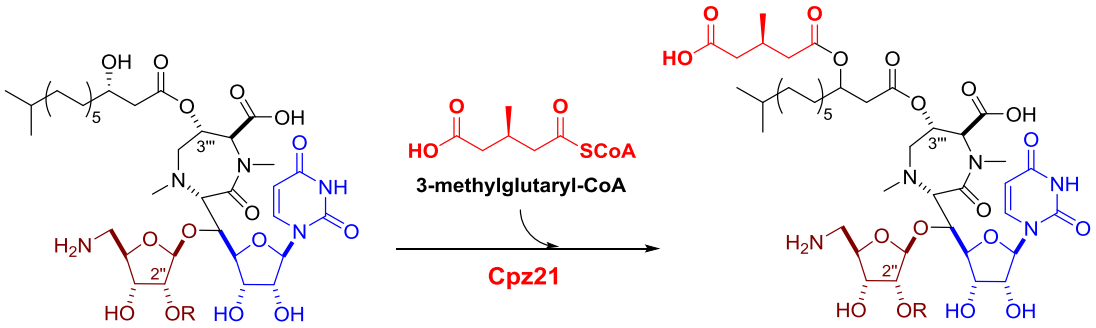


Figure 1.6.2. **Side chain modifications in the caprazamycin pathway**. The transfer of 3-methylglutaryl-CoA is catalyzed by Cpz21 in caprazamycin biosynthesis.

Clues towards amide bond formation in the capuramycin pathway can be gleaned from the enzyme CapW, identified as a putative Class C β-lactamase encoded within the biosynthetic gene cluster for the capuramycins. CapW was found to catalyze a transacylation resulting in the addition of an L-aminocaprolactam at the expense of the methyl ester (**Figure 1.6.3**) [62]. The methyl ester was shown to be produced by CapS, an *S*-adenosyl-L-methionine-dependent carboxylmethyltransferase that activates the carboxylic acid component of the capuramycin precursor to the methyl ester, thereby providing a kinetically competent substrate for the transacylase. The putative active site Ser of CapW was mutated to Ala resulting in loss of enzyme activity, lending support to the hypothesis that the reaction proceeds by a serine-dependent acylation/deacylation mechanism typical of Class C β-lactamases.

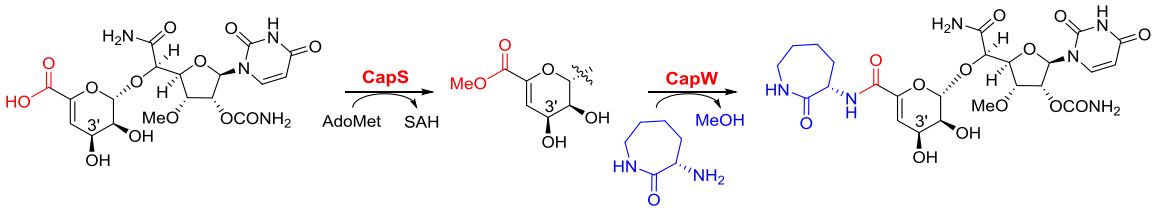


Figure 1.6.3. Amide-bond catalysis in capuramycin. Installation of the unique caprolactam in capuramycins is achieved by concerted CapS and CapW reactions and involved activation of the substrate by a methyl ester, followed by amide bond formation.

Another key observation is the predictive self-resistance mechanism in these antibiotics. A gene encoding a putative aminoglycoside 3-phosphotransferase (referred to as ORF21) within the gene cluster for the A-500359s was demonstrated to be highly expressed during the production of the A-500359s, and was implicated in conferring self-resistance when expressed in heterologous hosts *E. coli* and *Streptomyces albus* [61]. A similar gene (*capP*) from the A-503083 gene cluster was shown to encode the enzyme CapP, an ATP-dependent capuramycin phosphotransferase that regiospecifically transfers the γ-phosphate to the 3”-hydroxyl of the hexuronic acid moiety of A-503083 [72]. Kinetic characterization of CapP with three major A-503083 congeners established that CapP preferentially phosphorylates A-503083s containing an aminocaprolactam moiety attached to the hexuronic acid. Consistent with these observations, the product obtained from the CapP reaction lost its antibiotic activity against *Mycobacterium smegmatis*, and this loss in bioactivity is primarily due to a 272-fold increase in the IC_{50} in the bacterial translocase I-catalyzed reaction. Recent work in our group (Wenlong Cai) has established an identical mechanism for resistance to A-102395.

1.7. Functional assignment of LipL

The *orf* from A-90289 was shown to encode the protein LipL which was functionally assigned by our lab as a non-heme, Fe(II)-dependent α-ketoglutarate (α-KG): uridine-5'-monophosphate (UMP) dioxygenase catalyzing the conversion of UMP to uridine-5'-aldehyde during A-90289 biosynthesis [70]. The LipL reaction was modeled on the enzyme TauD, the best studied member of a large and diverse superfamily of mononuclear, non-heme Fe(II)-dependent enzymes that are generally agreed to follow similar reaction coordinates involving oxidative decarboxylation of α-ketoglutarate (hereby abbreviated as α-KG) to presumably generate an enzyme-bound Fe(IV)-oxo intermediate during the reaction. This strong oxidizing agent subsequently abstracts a hydrogen atom on the so-called prime substrate (UMP for LipL) to generate a carbon centered radical that leads to an unstable hydroxylated intermediate [73, 74] (**Figure 1.7.1**).

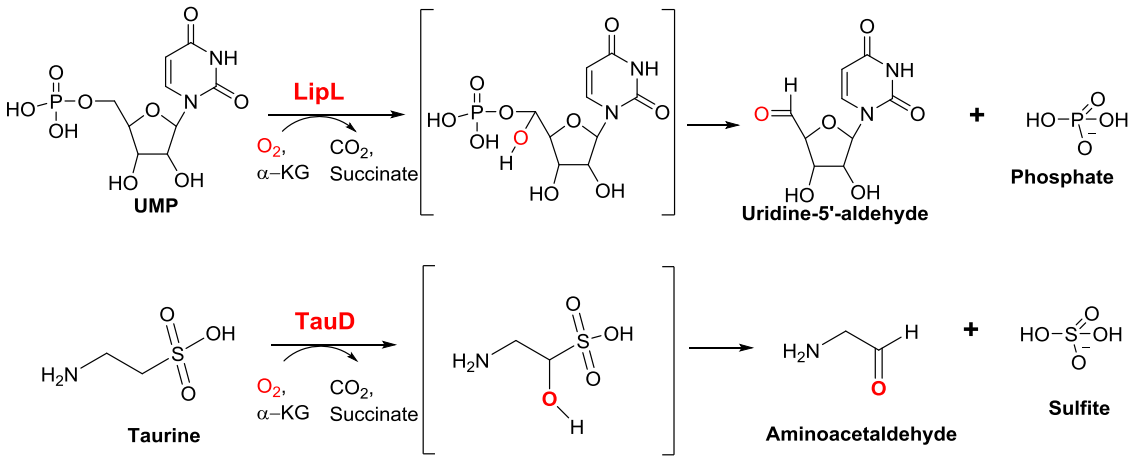


Figure 1.7.1. The reaction catalyzed by LipL. LipL catalyzes the first reaction in A-90289 biosynthetic pathway, and has been successfully characterized and modeled on the reaction catalyzed by E. coli TauD (below).

The Fe(II)/ α-KG dependent enzymes are a mechanistically and functionally diverse superfamily of enzymes responsible for a variety of reactions involved in protein side-chain modifications, repair of alkylated DNA/RNA, biosynthesis of antibiotics, lipid metabolism and biodegradation of a number of compounds [75]. The conserved structural element is a β-strand "jellyroll" fold typically containing three metal-binding ligands found in a His^1-X-Asp/Glu-X_n-His^2 motif [76, 77]. The only exception can be found in halogenases in this group, where the carboxylate ligand is absent [78] (**Figure 1.7.2**). The co-substrate α-KG, chelated with the Fe(II) center via its C-2 keto group and C-1 carboxylate, is oxidatively decarboxylated to succinate, leading to the formation of the Fe(IV)-oxo center responsible for hydroxylation of the prime substrate via a putative "oxygen rebound" mechanism (or in other instances results in desaturation, cyclization, ring closure/expansion etc.) [73, 74, 79, 80]. A few members show resemblances to this

group in terms of protein structures or chemical mechanisms but somewhat unexpectedly do not require α-ketoglutarate as a co-substrate [81].

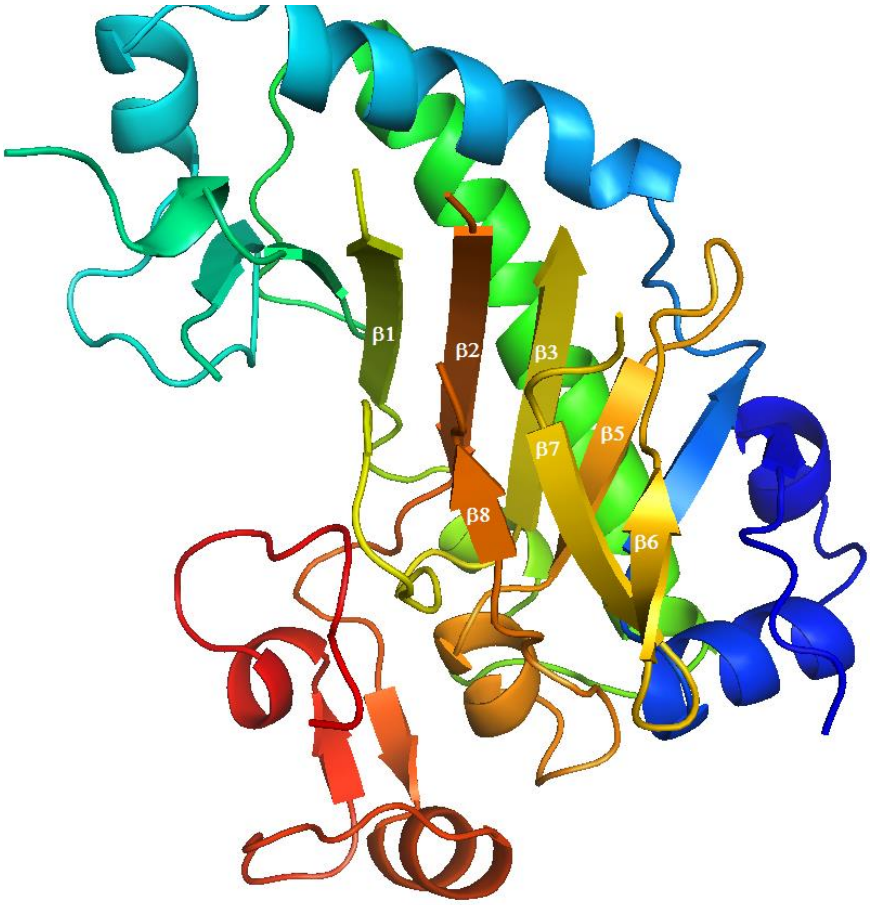


Figure 1.7.2. 'β- barrel jellyroll fold' in DAOCS (Deacetoxycephalosporin C-synthase). The eight parallel β-strands (indicated as β1-8) are conserved in all members of the enzyme superfamily, despite the broad spectrum of substrates recognized by the enzymes of this superfamily.

The putative reaction coordinates for representative enzymes from this group are elaborated in the following chapters, with special emphasis on those closest in analogy to the reactions catalyzed by LipL and Cpr19, the latter being the functional homolog Cpr19 from the A-102395 biosynthetic pathway. The ability to delve into intricate details of this

intriguing toolbox of enzymatic machinery represents a fantastic platform to gain improved understanding of their mode of action, to engineer these biocatalysts towards generating novel scaffolds and facilitating future redesign. The focus of this thesis is an in-depth characterization of LipL and its functional homolog Cpr19, with the major goal to identify the putative reaction intermediates involved in its unique chemical coordinates, along with additional efforts to elucidate the structural basis for the reactivity by studying site-specific mutant variants of Cpr19.

Chapter two: Elucidating the mechanism of LipL and Cpr19

2.1. Background: Fe(II)/ α-ketoglutarate-dependent enzymes

We briefly review the amazing diversity of reactions catalyzed by members of this enzyme superfamily and use specific examples to illustrate the same. Most representatives couple the oxidative decomposition of α-KG to succinate and CO_2 to the activation of their prime substrates via the generation of a reactive oxygen species at the non-heme iron center. The reactive oxygen species can then display alternative reactivity in related enzymes that may result in hydroxylations, stereoinversions, desaturations, ring closure or ring expansions (**Figure 2.1.1**).

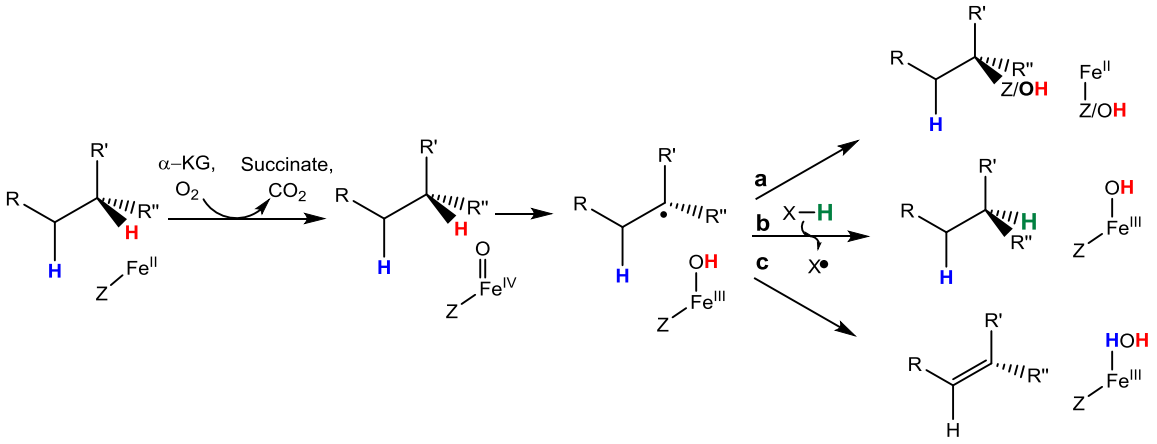


Figure 2.1.1. Variable outcomes of reactions catalyzed by Fe(II): α-KG dependent enzymes. Divergence of the proposed pathway resulting in (a) radical-group transfer, (b) stereoinversion, or (c) desaturation outcomes

The first identified hydroxylase from this group was the enzyme prolyl-4-hydroxylase [82] shown to be involved in the generation of *trans*-4-hydroxyprolyl products. In mammals, this reaction is essential for the formation of collagens, elastins and several other proteins

[83]. Several other enzymes from this group have since been shown to be required for protein side chain modifications including lysyl hydroxylase [84], prolyl-3-hydroxylase, aspartyl *β*-hydroxylase [85], etc (**Figure 2.1.2**).

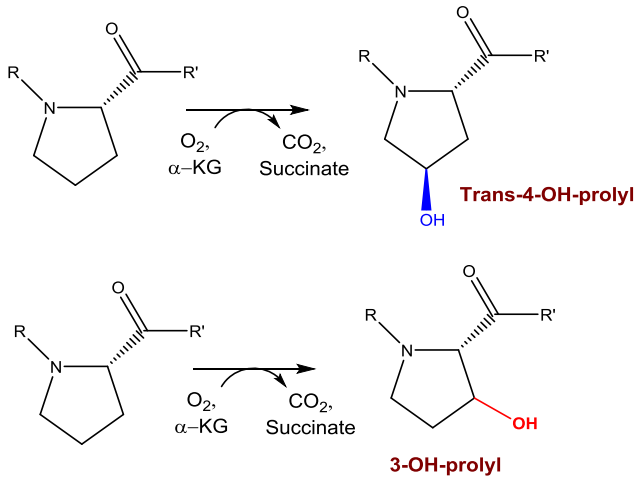


Figure 2.1.2. Representative hydroxylation reactions. Reactions catalyzed by Prolyl-4-hydroxylase (top) versus Prolyl-3-hydroxylases.

Some of these enzymes have been implicated in the repair of DNA/RNA – for example, *E. coli* possesses the enzyme *alkB* that has been implicated in the direct repair of methylated DNA and RNA lesions [86]. Expression of *E. coli alkB* in human cells was shown to confer resistance to high concentrations of S_N2 alkylating agents and in 2002, Trewick et al demonstrated the direct repair of methylated DNA by AlkB in an Fe(II)-dependent process that consumes oxygen plus α-KG and produces succinate and formaldehyde [87]. The single isozyme of clavaminate synthase (CAS) exemplifies the versatility of Fe(II)/α-KG dioxygenases by catalyzing three separate oxidative steps in the synthesis of clavulanic acid. First identified in 1999 by Lloyd et al [88], the enzyme has been shown to hydroxylate

the β-lactam precursor in the first step, followed by sequential cyclization and desaturation in three different steps to yield the precursor for the final product (**Figure 2.1.3**).

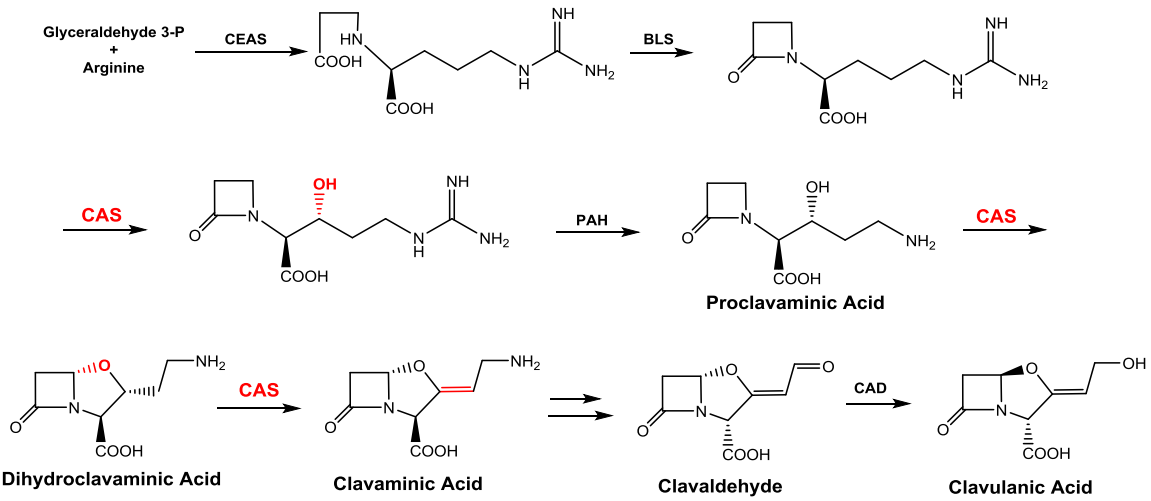


Figure 2.1.3. **Role of CAS**. CAS is responsible for carrying out three distinct oxidative reactions in this pathway. Abbreviations: CEAS: carboxyethylarginine synthase, BLS: β-lactam-synthetase, PAH: proclaviminic acid aminidino synthase, CAS: Clavaminate synthase, CAD: Clavaldehyde dehydrogenase

Another interesting member of this family is the enzyme CarC (carbapenem synthase) involved in a desaturation reaction similar to CAS, but also exhibiting an additional *epimerization* at the tertiary carbon atom that joins the two rings in the bicyclic product. A recent article published in Science in 2014 [89] illustrates the unique mechanism of CarC by which it carries out C5 stereoinversion in the biosynthesis of carbapenem antibiotics. Following the attachment of a carboxymethylene unit to the C5 of L-proline of the 2-pyrrolidine precursor unit, the Fe(IV)-oxo intermediate abstracts a hydrogen atom from

C5, and the key residue tyrosine165 in CarC then proceeds to donate the hydrogen atom to the opposite face of the resultant radical, an event that is sufficient to direct the nonredox stereoinversion outcome that distinguishes CarC from other Fe(II)/α-KG dioxygenases that typically proceed through a putative "oxygen-rebound" mechanism (**Figure 2.1.4**).

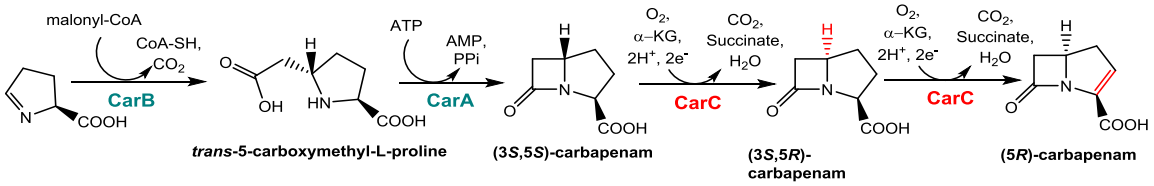


Figure 2.1.4. Mechanism of CarC. CarC catalyzes both epimerization and desaturation reactions in generating the (5R)-carbapenam scaffold. We elaborate more on this mechanism in chapter three, alluding to its recently resolved crystal structure.

Few members of this group are additionally involved in the biosynthesis of plant products. For example, the enzyme flavanone 3*β*-hydroxylase catalyzes a key step in flavonoid biosynthesis [90] in which flavanones (for example naringenin illustrated in **Figure 2.1.5**) are converted to the corresponding trans-dihydroflavonols. We have additionally illustrated the reaction catalyzed by flavone synthase I in the same figure, since it also utilizes flavones as substrates, but carries out a desaturation reaction resulting in the introduction of a double bond in the final product [91]. Fe(II)/α-KG hydroxylases also participate in lipid metabolism – two of the enzymes taking part in carnitine synthesis have been characterized from humans [92] and phytanoyl-CoA hydroxylase is required

for degrading the plant lipid phytanic acid [93]. The enzyme TauD from *E. coli* is involved in the decomposition of alkyl sulfonates [73] along with the enzyme AtsK, from *Pseudomonas sp*. that catalyzes the decomposition of alkyl sulfates, as opposed to sulfonates, in a reaction mechanism identical to TauD [94]. A related biotransformation has been observed for phosphorus scavenging in *Pseudomonas stutzeri* WM88, in which, the genes htxA encodes the Fe(II)/α-KG hydroxylase HtxA capable of hydroxylating reduced forms of phosphorus including hypophosphite (H_3PO_2) and phosphite (H_3PO_3) [95].

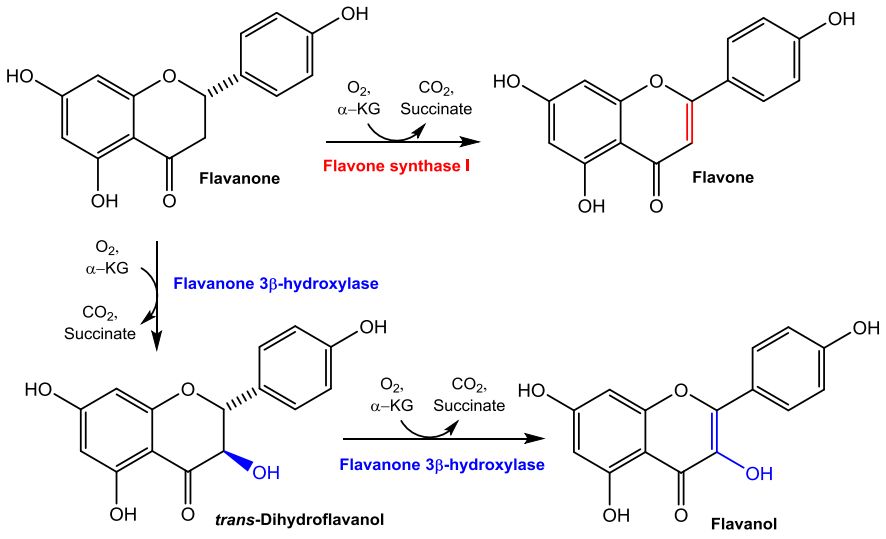


Figure 2.1.5. Enzymes involved in Flavonoid biosynthesis. Flavone synthase I (**red**) catalyzes desaturation of the substrate to introduce a double bond (top), while the enzyme flavonone 3β-hydroxylase exhibits tandem hydroxylation/oxidation with the same substrate.

Though the reactive Fe(IV)-oxo intermediate displays slight variations of the primary activity in catalyzing desaturations, ring expansions, ring closures or other oxidative

biotransformations within this group, all the enzymes described herein are agreed to follow a more or less conserved mechanism which was formulated more than 20 years ago by Hanauske-Abel and Günzler (solely on the basis of theoretical considerations) [96]. As more enzymes are added to this intriguing group and the spectrum of its functionalities expand, we continue to broaden our understanding of the exact chemical coordinates central to all of these reaction mechanisms.

2.2. Reaction mechanism of Fe(II)/α-KG – dependent dioxygenases

Dioxygenases are vital enzymes with key functional roles in nature, utilizing molecular oxygen and transferring both oxygen atoms to activated substrates [79, 97]. Monooxygenases, on the other hand, transfer a singular oxygen atom from molecular oxygen while the other oxygen atom leaves in the form of water. The most well-known monooxygenases, the P450s are some of the most versatile enzymes in the body involved in such diverse functions such as detoxification of organic substrates in the liver to the biosynthesis of hormones [98]. P450s are also a group of enzymes that utilize an oxo-iron species for abstracting a hydrogen atom from their substrates, albeit with a heme iron center involving a catalytic cycle that proceeds via two reduction and two protonation steps. Although elusive, the oxo-iron species has been implicated by indirect evidence from product distributions and kinetic isotope effects, and in case of peroxidases has been fully characterized by crystallography, EPR and vibrational spectroscopy [98, 99]. A similar oxo-iron species is central to the non-heme, Fe(II)/α-KG-dependent group of enzymes,

however it remains elusive in terms of in-depth characterization [100]. So the interesting question is, why are there two different classes of enzymes developed by nature for essentially similar tasks: the hydroxylation of substrates via the activation of unreactive carbon centers? This question can perhaps be addressed by the differences in the heme and non-heme iron centers in these two groups, and the absence of a required cofactor in case of the dioxygenases, which creates significant differences in the putative reaction coordinates for the Fe(II)/α-KG-dependent enzymes (**Figure 2.2.1**).

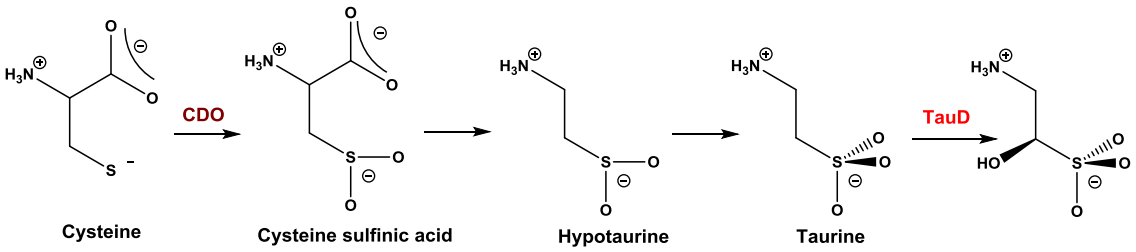


Figure 2.2.1. Role of TauD in the body. The metabolism of cysteine in the body involves both TauD and cysteine dioxygenase (CDO)

The most extensively studied member of non-heme Fe(II) dioxygenases is *E. coli* TauD, implicated in the transfer of one of the oxygen atoms from molecular oxygen to α-KG, and utilization of the second oxygen atom for putative hydroxylation of its prime substrate taurine. **Figure 2.2.2** shows the catalytic cycle of TauD based on cumulative theoretical and experimental observations [101]. The cycle starts from resting state (**A**) where the non-heme iron center is coordinated by two histidine (His_{99}, His_{205}) residues, a carboxylic acid group from Asp_{101} and three water molecules [73]. The first step is binding

of α-KG generating a bidentate ligand (**B**) by the displacement of two of the coordinating water molecules in the structure [102]. This intermediate complex has been isolated and characterized as a Fe(II)/α-KG chromophore with absorption at 530 nm [103]. This step is followed by sequential binding of the prime substrate taurine, which does not bind directly to the center but displaces the remaining water molecule (which leads to the complex **C** with absorption at 520 nm) allowing molecular oxygen to bind (**D**), initiating the radical mediated sequential steps and simultaneously limiting the scope for identification of the exact intermediates. It is however postulated that the dioxygen can attack the α-KG leading to the formation of a five-membered ring structure (**E**), followed by spontaneous decarboxylation of α-KG to succinate yielding the oxo-iron species in complex **F** [104]. The reactive oxo-iron species can putatively abstract a hydrogen from the prime substrate to afford a carbon centered radical on taurine (**G**), and rebound of the hydroxyl group ("oxygen rebound") to the reactive center forms the product complex **H**. Release of products (succinate and hydroxylated taurine) and rebinding of water can restore the catalytic center to its resting state (**A**) and initiate another catalytic cycle. This mechanism has been corroborated with stopped-flow experiments which provided evidence for at least two stable intermediates [102]: the first identified intermediate is the oxo-iron complex that absorbs at 318 nm and develops after 20 – 25 ms, but decays after 600 ms. Evidence for this intermediate can also be gleaned from oxygen isotope studies using $^{16}O/^{18}O$ labeled substrates [105], and from observations with hydrogen/deuterium kinetic isotope effects using taurine and taurine-d_2 which proved that this complex is responsible for hydrogen abstraction [106]. The second identified

intermediate corresponds to the complex **H**, formed after the oxo-iron species. Further evidence from isotope labeled studies indicate the retention of each atom of oxygen, one in succinate and the other in acetaldehyde [105]. However, this has never been detected in the latter due to rapid solvent exchange.

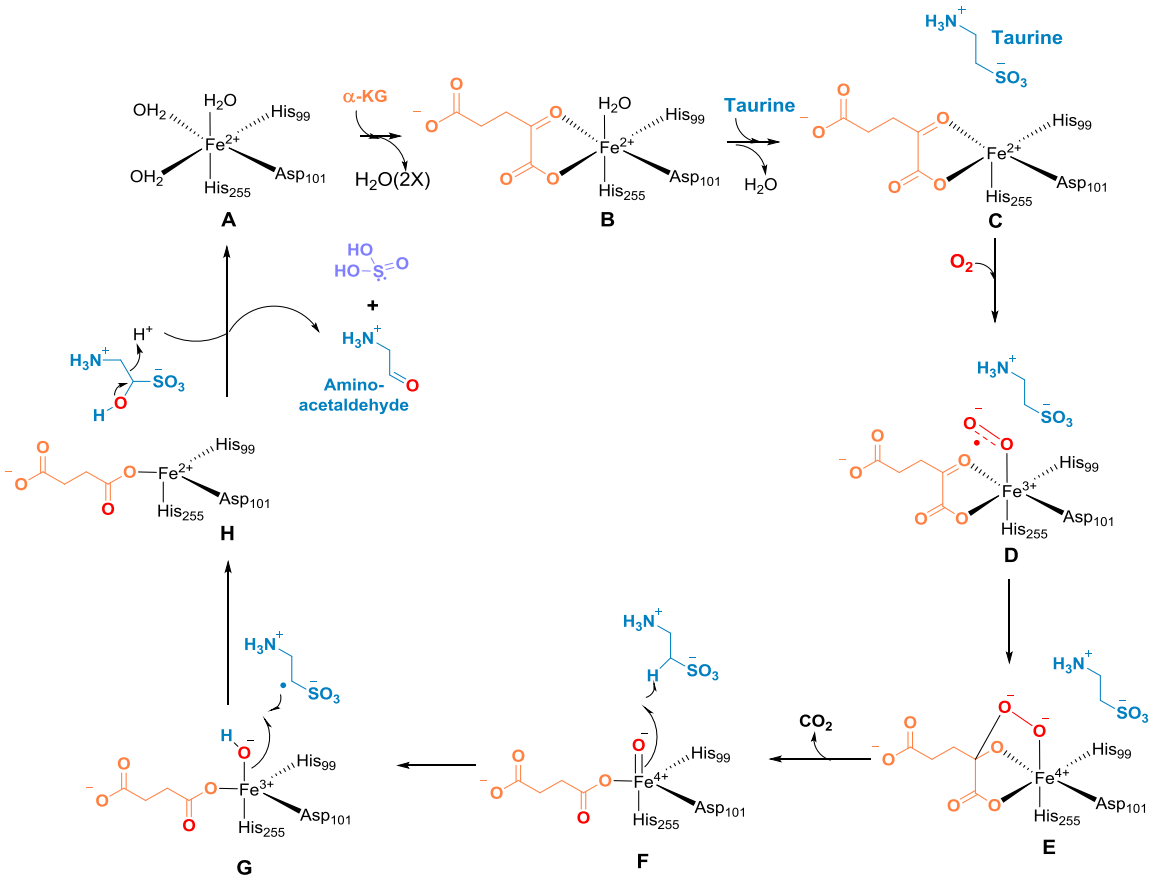


Figure 2.2.2. Putative intermediates in the catalytic cycle of TauD. Intermediates A-H have been identified by a combination of isotopic enrichment studies, stopped flow experiments, spectroscopic and computational analyses. The only intermediates directly identified are F and H.

For LipL and Cpr19, we begin by charting out a putative catalytic cycle modeled on TauD (**Figure 2.2.3**), wherein the prime substrate UMP (**1**) can be *putatively* hydroxylated at C-5 and thereafter dephosphorylated spontaneously to the final product uridine-5'-

aldehyde (**2**) and inorganic phosphate. The pathway is proposed to proceed via intermediates a-h, starting with the 'resting state' where the Fe(II) center is coordinated via the facial triad in the catalytic center, and three water molecules (a). Ordered sequential binding of α-KG and UMP is proposed to cause conformational changes that allow for binding of molecular oxygen to the catalytic center to form a Fe(III)-superoxo species (d). The reaction can then proceed with attack of the distal oxygen on the α-keto group of the bidentate coordinated α-KG, which in turn can lead to decarboxylation and O-O bond cleavage to generate byproducts succinate, CO_2 and and the high-spin Fe(IV)-oxo intermediate (f). This intermediate is proposed to abstract the C-5' hydrogen from UMP, leading to the trigonal pyramidal complex 'g', as well as a carbon-centered radical on UMP C-5'. 'Oxygen rebound' can lead to the hydroxylation at this position generating the understandably short-lived 5'-OH-5'-phosphouridine, which can spontaneously dephophorylate to the uridine-5'-aldehyde product following the TauD mechanism. Though the catalytic cycle (described above) for TauD is widely accepted as a consensus for this group of enzymes, in the context of the great spectrum of functional variability of this group, another rational alternative can be envisioned for some of the later steps in the catalytic cycle for LipL: a plausible alternative, in line with mechanisms demonstrated for CarC and ANS [89], is a desaturation/enol-tautomerization mechanism, or a non-redox conversion (CarC) that can take place following the first hydrogen abstraction step. An alternative route for LipL via desaturase/enol tautomerization towards the final uridine-5'-aldehyde product is additionally conceptualized and discussed in details in the

following section. We explore a number of strategies to provide a definitive distinction between the two mechanisms and gain insights into the structural basis for the same.

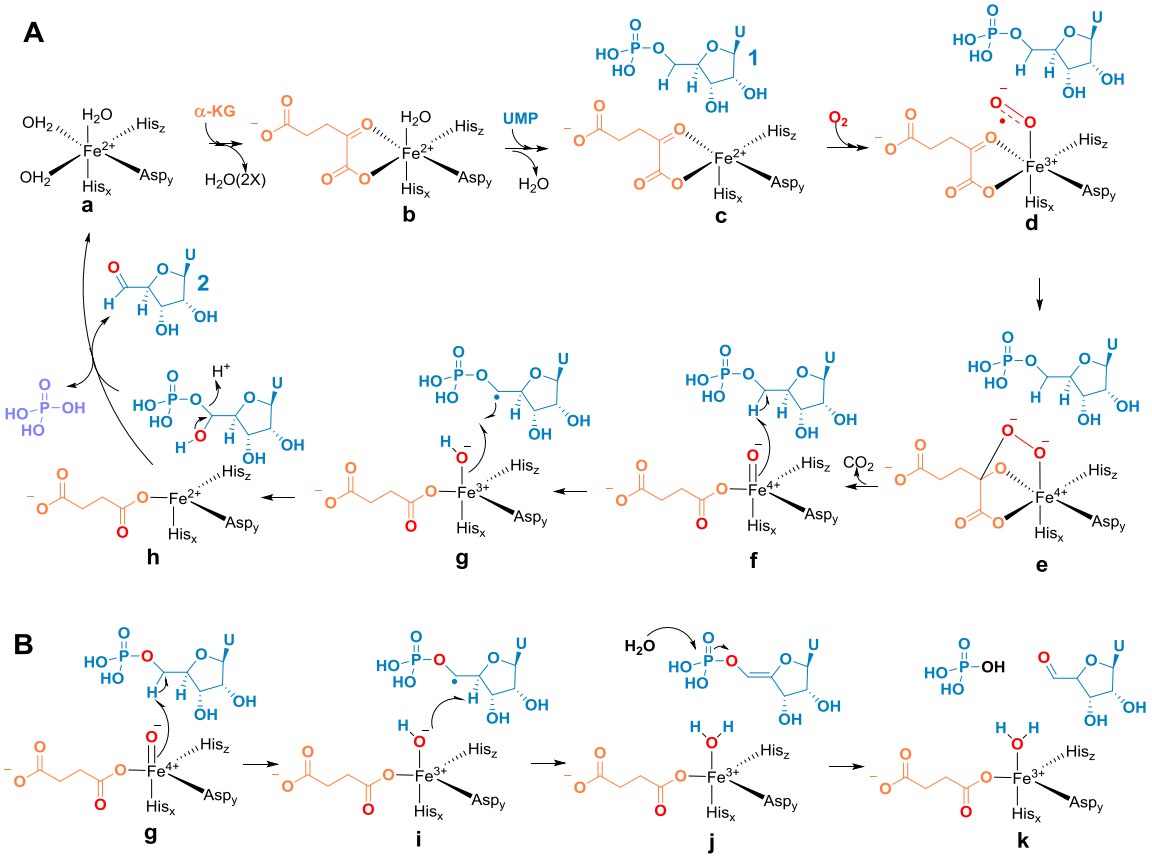


Figure 2.2.3. Putative reaction coordinates for LipL (modeled on TauD and CarC); (**A**) Proposed 'hydroxylation' mechanism proceeds through intermediates **a-h**, and cycles back to **a**; (**B**) Alternative proposal for 'desaturase' mechanism bifurcates from intermediate **g**, and proceeds through intermediates **g-k**, eventually cycling back to resting state **a**; U:Uracil.

2.3. Strategy for elucidating 'Desaturase' versus 'Hydroxlation' Hypothesis

For the desaturation hypothesis for LipL, in line with mechanisms demonstrated for CarC, ANS, favone synthase etc., starting from proposed intermediate **g,** the reaction can be

thought to proceed via sequential abstraction of both C-5' and C-4' hydrogens from UMP followed by phosphate hydrolysis and enol tautomerization to yield the final aldehyde product (uridine-5'-aldehyde). To be able to provide a clear distinction between the hydroxylation and desaturation hypotheses for LipL, therefore, the key is to track the fate of the C-4' and C-5' hydrogens. To monitor the fate of the C-4' and C-5' hydrogens, we strategized the utilization of a selectively deuterated UMP as the prime substrate that could be analyzed directly by mass spectrometric methods. As can be deduced from **Figure 2.2.3 (B)**, logic dictates that the retention of the deuterium label at the C-4' position would negate the possibility of a second hydrogen abstraction event from UMP in the original reaction, and thereby provide convincing evidence to *eliminate* a desaturation mechanism. In line with this argument, we formulated the synthesis of a selectively deuterated UMP analog ([2,3,4,5,5'-$^{2}H_1$]-uridine-5'-monophosphate) as the primary substrate to be tested in reaction with LipL (and its homolog Cpr19 from the strain A102395) and set out to track the fate of the C-4'-$^{2}H_1$ isotope label in the final aldehyde product (**Figure 2.3.1**).

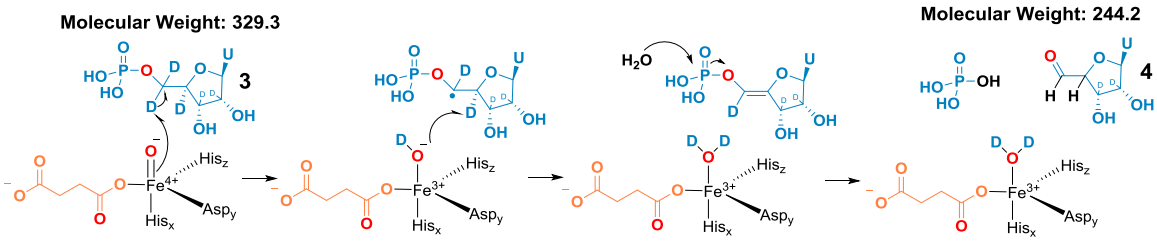


Figure 2.3.1. **Strategy with deuterated UMP substrate (3).** If the reaction were to follow sequential abstraction from both C-4' and C-5' positions of **3**, then the final aldehyde product (**4**) would lose both $^{2}H_1$ labels and have predicted MW 244.2.

For the hydroxylation hypothesis following the TauD model, we theorized that a direct evidence of hydroxylation can be observed from tracking the critical bridging phosphoester oxygen in UMP (**1**). Since the phosphate group in UMP is a good leaving group **Figure 2.2.3 (A)** it is difficult to establish the existence of the putative 5'-hydroxy-5'-phosphouridine (C-5'-OH-UMP) intermediate due to its innate instability.

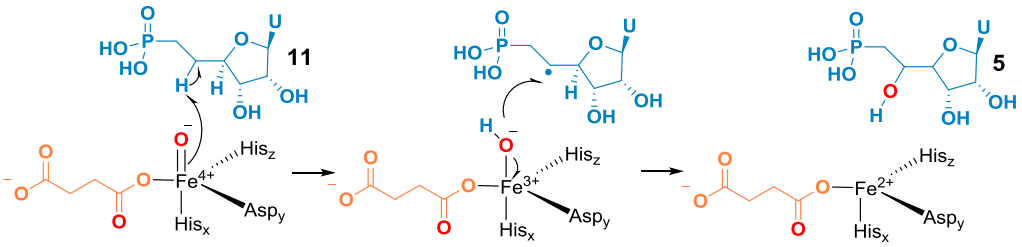
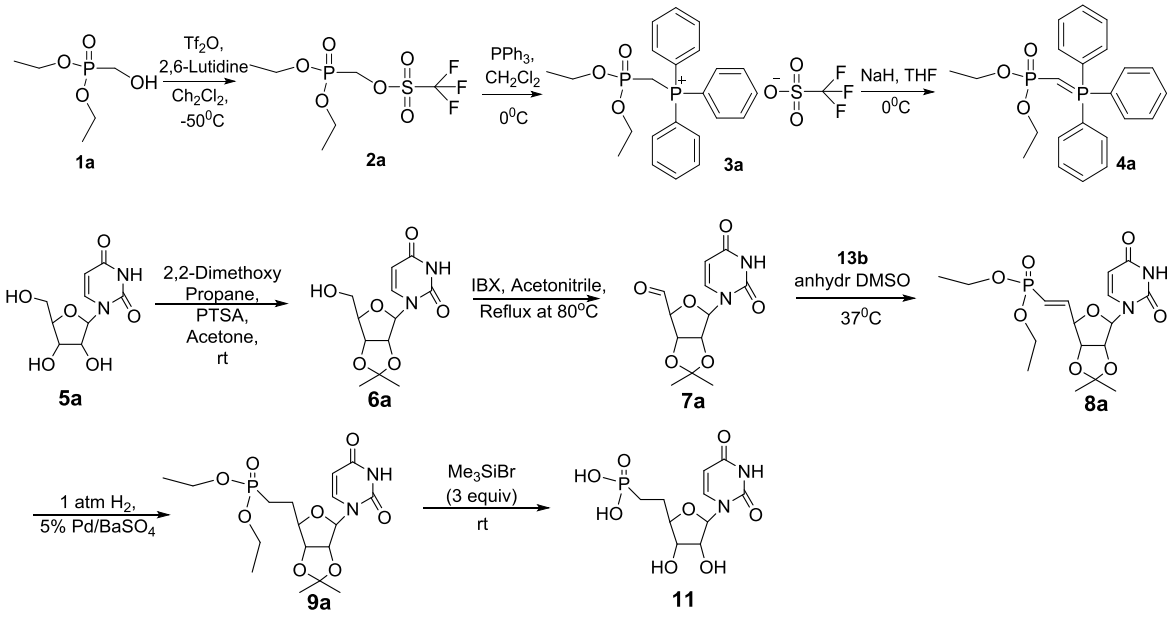


Figure 2.3.2. Strategy for utilization a modified substrate analog (11). We would attempt entrapment of the elusive intermediate hydroxylated at the functionalized C-5' position (**5**)

To circumvent this challenge, we strategized the synthesis of a *structural* analog of UMP (**Figure 2.3.2**) wherein the phosphoester oxygen bond O-P could be replaced by a C-P bond to make the phosphonate derivative, effectively transforming the phosphate to a poor leaving group. By doing so, and potentially entrapping the -OH at C-5' position from the reaction of LipL with the substrate analog (and preventing the intermediate from converting back to the uridine-5'-aldehyde product by spontaneous dephosphorylation) we could use it to mirror a similar hydroxylation step in the original reaction. To achieve

this, we have adopted a synthetic strategy following on the method described by Xu et al. [107] for the synthesis of phosphonates (**Scheme 2.3.3**).



Scheme 2.3.3. Synthesis of key substrate analog (11). Synthesis is planned via retrosynthetic wittig condensation of protected uridine-5'-aldehyde (**6a**) and corresponding ylide (**4a**).

2.4. Materials and methods

2.4.1. Chemicals and Reagents

UMP, (1,2,3,4,5,6,6-$^{2}H_{1}$)glucose, 2- ketoglutaric acid (α-KG), β-nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt (β–NADPH), adenosine 5'-triphosphate disodium salt (ATP), phosphor (enol) pyruvate trisodium salt (PEP), uracil, uridine-5'monophosphate, and ascorbic acid were purchased from Sigma-Aldrich (St. Louis, MO) or Promega (Madison, WI). Buffers, salts, organic solvents and media

components were purchased from Sigma-Aldrich (St. Louis, MO) and Fisher Scientific (Pittsburgh, PA). Synthetic oligonucleotides were purchased from Integrated DNA Technologies 21 (Coralville, IA). Wizard® *Plus* SV Minipreps DNA Purification Systems**,** Wizard® SV Gel and PCR Clean-Up System were purchased from Promega (Madison, WI, USA). pET-30 Xa/LIC Vector Kit was purchased from Calbiochem (San Diego. CA, USA). InstaGene Matrix was purchased from Bio-Rad (Hercules, CA). Ni-NTA agarose was purchased from Qiagen (Valencia, CA). Amicon Ultra 10000 MWCO centrifugal filter was purchased from Millipore (Billerica, MA). PD-10 desalting column was purchased from GE Healthcare (Pittsburgh, PA). DNA sequencing was performed using the BigDye™Terminator version 3.1 Cycle Sequencing kit from Applied Biosystems, Inc. (Foster City, CA) and analyzed at the University of Kentucky Advanced Genetic Technologies Center.

2.4.2. Instrumentation

UV/Vis spectroscopy was performed with a Bio-Tek µQuant microplate reader using Microtest™ 96-well plates (BD Biosciences) or a Shimadzu UV/Vis-1800 Spectrophotometer. HPLC was performed with a Waters Alliance 2695 separation module (Milford, MA) equipped with a Waters 2998 diode array detector and an analytical Apollo C-18 column (250 mm x 4.6 mm, 5 µm) or a semi-preparative Apollo C-18 column (250 mm x 10 mm, 5 µm) purchased from Grace (Deerfield, IL). Electrospray ionization-MS was performed using an Agilent 6120 Quadrupole MSD mass spectrometer (Agilent Technologies, Santa Clara, CA) equipped with an Agilent 1200 Series Quaternary LC

system and an Eclipse XDB-C18 column (150mm x 4.6 mm, 5 µm, 80Å). High-resolution MS was obtained from either University of Kentucky Mass Spectrometry Core Facility, or from University of Minnesota, Department of Chemistry Mass Spectrometry Facility. NMR data were collected using a Varian Unity Inova 400 or 500 MHz Spectrometer (Varian, Inc., Palo Alto, CA) at the University of Kentucky, and a Bruker Avance III 700 MHz spectrometer equipped with a 5 mm QCI probe at the University of Wisconsin, Madison. Malachite green binding assay was performed with a colorimetric-based Sensolyte MG Phosphatase Assay Kit from AnaSpec, Inc. (Fremont, CA).

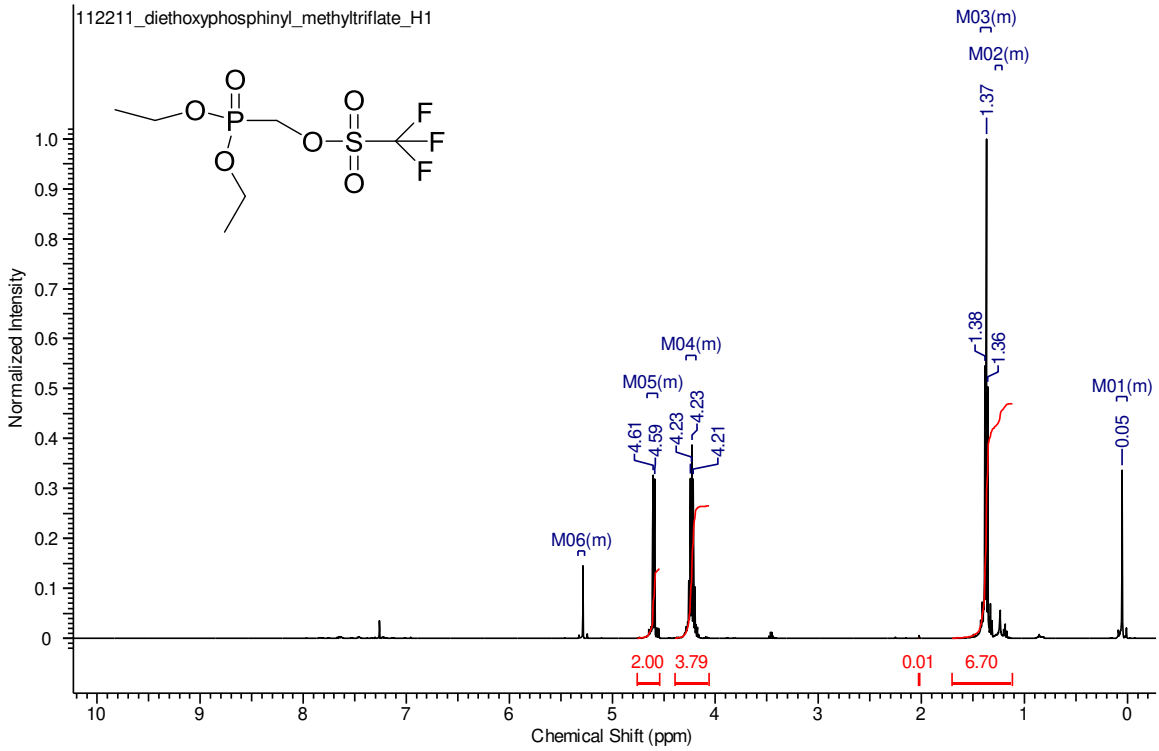
2.4.3. Bacterial strains and Enzymes

NovaBlue GigaSingles™ Competent Cells was purchased from Calbiochem (San Diego. CA, USA). One Shot® BL21 (DE3) Chemically Competent *E. coli* was purchased from Invitrogen (Camarillo, CA)**.** TaKaRa LA Taq® DNA polymerase with GC Buffer was purchased from Takara Bio Inc (Otsu, Shiga, Japan), T4 DNA ligase, NdeI, and Hind III were purchased from New England Biolabs (Ipswich, MA). Expand long template PCR system was purchased from Roche Applied Science (Indianapolis, IN). Commercial varieties of hexokinase, pyruvate kinase, glucose-6-phosphate dehydrogenase, glutamate dehydrogenase, phosphogluconate dehydrogenase, myokinase and inorganic pyrophosphatase were purchased from Sigma-Aldrich (St. Louis, MO).

2.4.4. Synthesis of 1-[5,6-Dideoxy-6-(dihydroxyphosphinyl)-β-D-*ribo*hexofuranosyl] uracil (11)

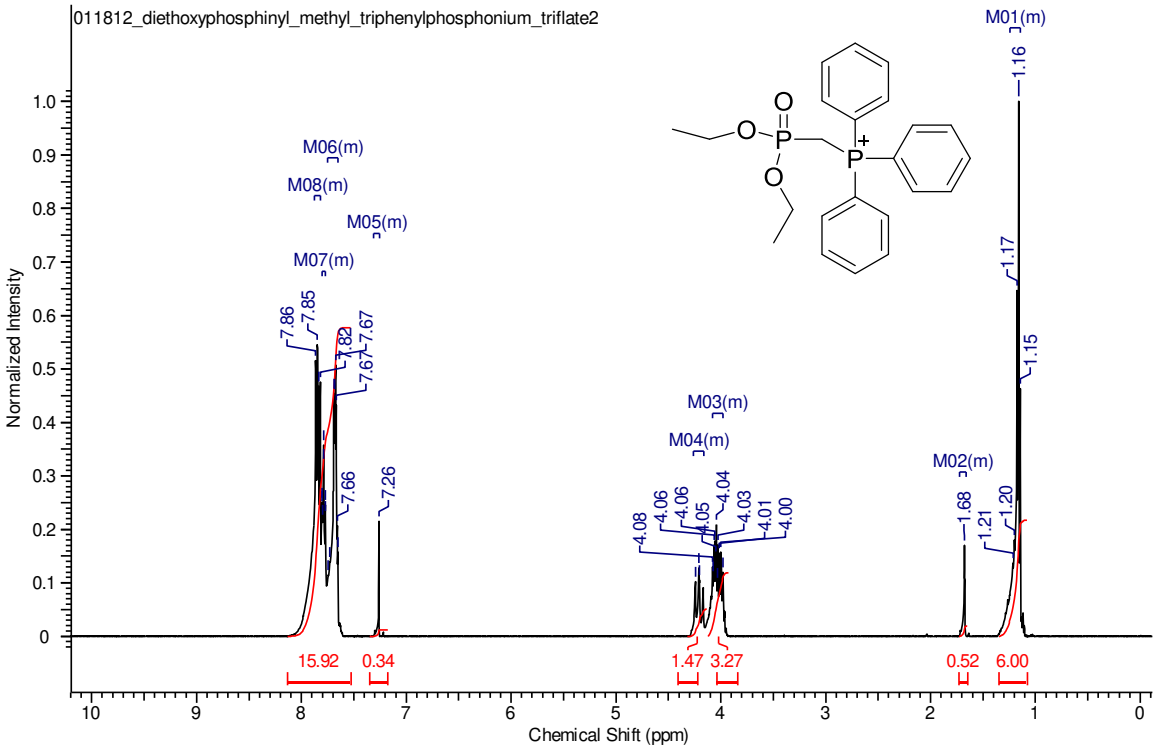
The synthetic strategy following on the method described by Xu et al. (122) for the synthesis of phosphonates (**Scheme 2.3.3**) is used to generate the final substrate analog (**11**). **Table 1** summarizes the spectral characterization of compound **11,** at the end of this section.

(Diethoxyphosphinyl)methyl Triflate (2a): Trifluoromethanesulfonic anhydride (35.5 mmol) was added dropwise to a stirred solution of commercially available diethyl (hydroxymethyl)phosphonate (30.6 mmol) and 2,6-lutidine (37.6 mmol) in anhydrous dicholoromethane (50 ml) at -50°C under N_2 atmosphere. The reaction mixture was then allowed to warm to 0˚C over a period of 1.5- 2 hrs, and diluted thereupon by with ether (300 ml). A precipitate formed upon addition of the diethyl ether, which was removed by filtration and the filtrate was successively washed with water, 1 N HCl, and brine and then dried over Mg_2SO_4. A yellowish oil was obtained upon concentration of the dried solution, and this was used for the next step without further purification.^{1}H NMR (500MHz, $CDCl_3$) δ 1.37 (t, 6 H), 4.21-4.23 (m, 4 H), 4.61 (d, 2 H).



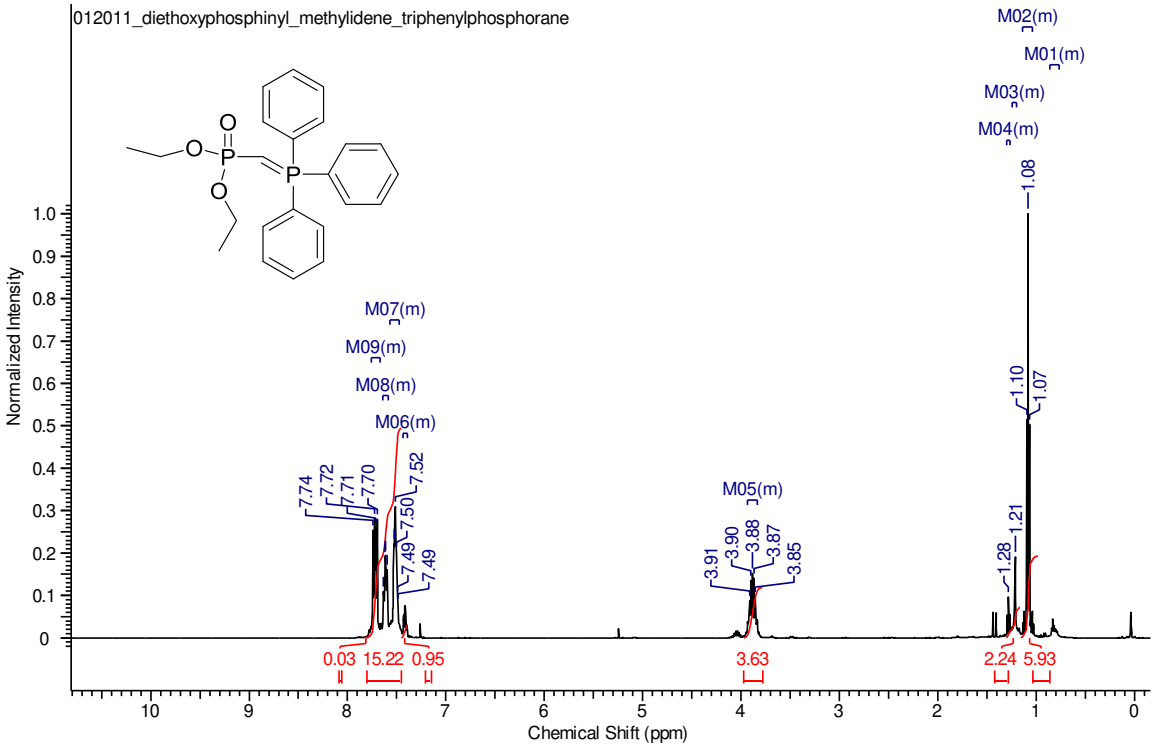
[(Diethoxyphosphinyl)methyl] triphenylphosphonium Triflate (3a): 30 mmols of [(diethoxyphosphinyl)methyl] triphenylphosphonium triflate (2) was added dropwise to a stirred solution of triphenylphosphine (34.4 mmol) in anhydrous CH_2Cl_2 (50 ml) at 0 ^oC under N_2. The solution was left to reach room temperature and then stirred at this temperature overnight. The CH_2Cl_2 was then evaporated under reduced pressure to about one third of the volume, and the resultant oil was triturated with ether (200 ml). This led to the formation of a white solid which was then collected by filtration. The final triflate (2) was obtained after being washed with ether twice (50 ml x 2) as a white solid. Owing to its relative instability, it was used directly in the next step. Analytical sample was

prepared by rescrystallization from ethyl acetate/hexane. ^{1}H NMR (500MHz, $CDCl_3$) δ 1.16 (t, 6h), 4.0-4. 06 (m, 4H), 4.20 (dd, 2H), 7.66-7.86 (m, 15H)



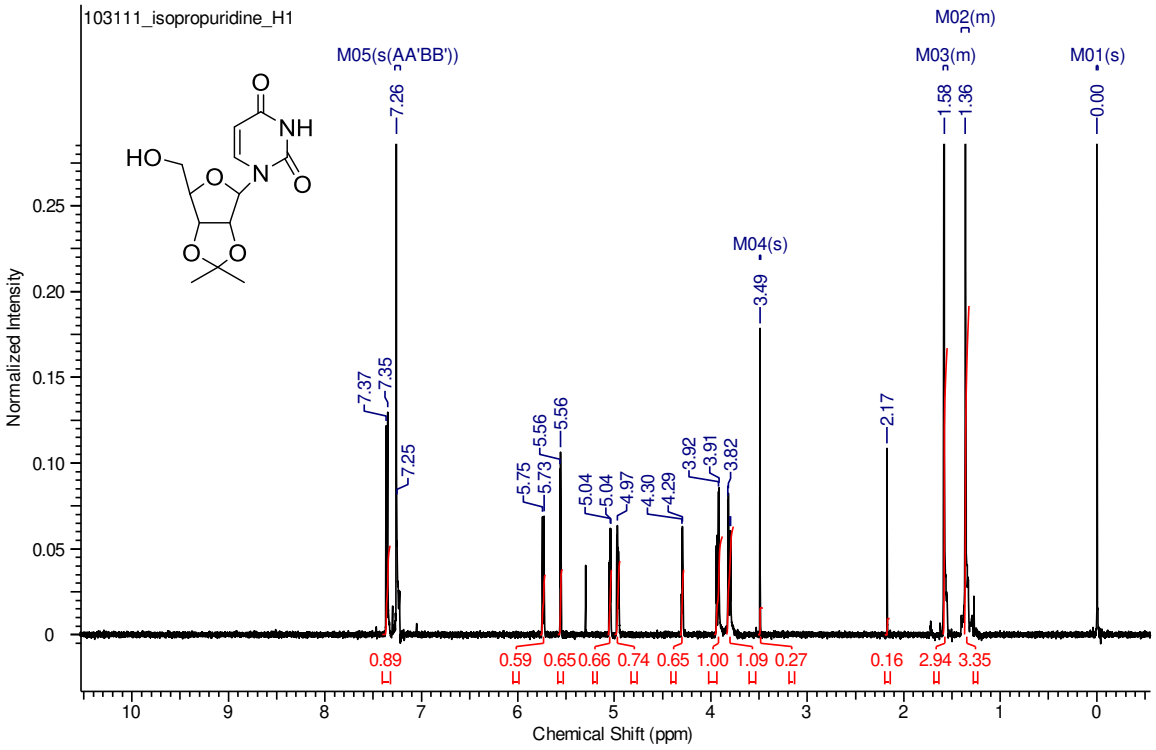
[(Diethoxyphosphinyl)methylidene] triphenylphosphorane (4a): The triphenylphosphonium triflate salt (3) (in anhydrous THF) was used directly for reaction with a stirred suspension of NaH (1.25 mmol) in anhydrous THF (2 ml) at 0 ^{o}C under N_2 and the reaction was allowed to proceed for about 0.5 h. The solvent was then removed under reduced pressure and the resultant residue was extracted using CH_2Cl_2. The colorless oil obtained from concentration of the extracts was triturated with hexane to yield another white solid. This solid too was unstable, and as such used directly for the

next step. ^{1}H NMR (500MHz, $CDCl_3$) δ 1.08 (t, 6H), 1.28 (d, 1H), 3.88 (apparent quintet, 4H), 7.49-7.74 (m, 15H)

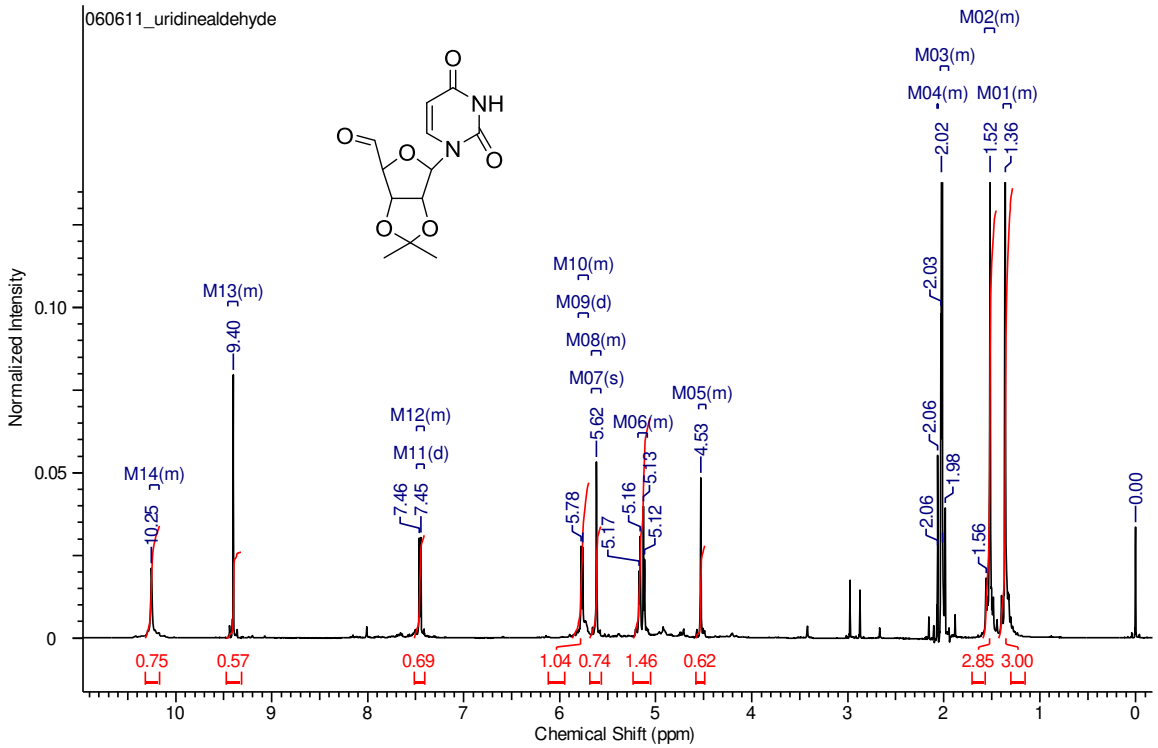


2',3'-O-isopropylidene-uridine (6a): 2,2-dimethoxypropane (270.68 mmol) was added to a mixture of uridine (**5**) (20.48 mmol) and *p*-toluenesulfonic acid (1.75 mmol) in dry acetone (166.5 mL, 0.123 M) and stirred at room temperature for 4 hrs. The color of the solution turned yellow, and after removal of the solvent under reduced pressure the color changed to deep purple. TLC analysis indicated completion of the reaction, and the dried residue was used directly for purification via silica gel chromatography and eluted with 5% ethyl acetate in hexane. The final product was a white powder. ^{1}H NMR (500MHz,

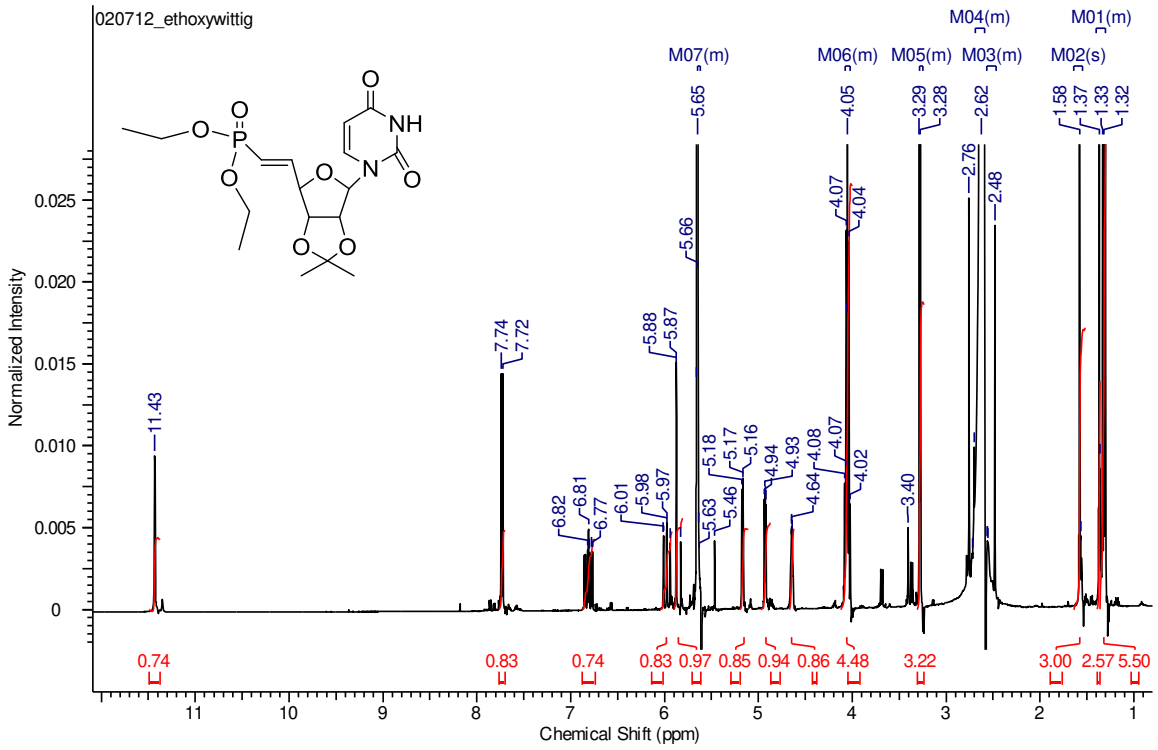
$CDCl_3$) δ 1.36 (s, 3H), 1.58 (s, 3H), 3.82 – 3.92 (m, 2H), 4.29 (m, 1H), 4.97 – 5.04 (m, 2H), 5.56 (d, 1H), 5.73 (m, 1H), 7.37 (d, 1H)



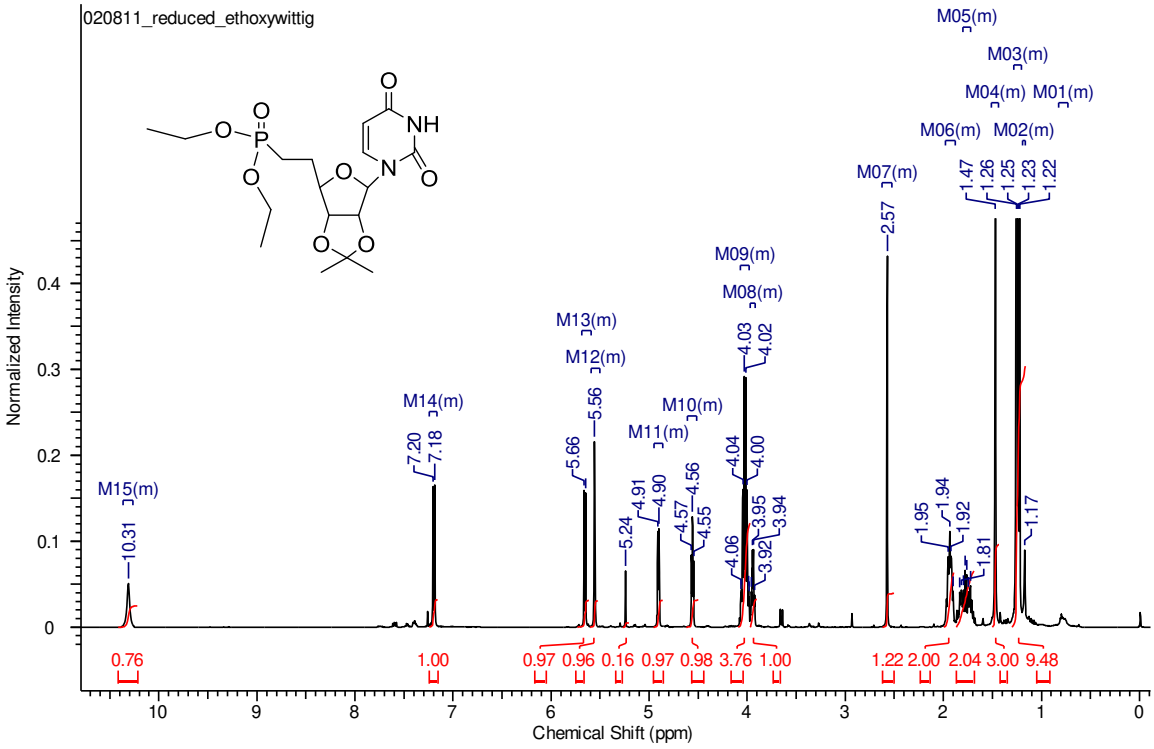
2',3'-O-isopropylidene-uridine-5'-aldehyde (7a): 2.476 mmols of IBX was added to a solution of compound **6** (1.478 mmol) in dry acetonitrile(2.956ml, 0.5M) and stirred under reflux conditions at 80˚C for 2 hrs. Following completion of the reaction, the IBX was removed by filtration through Celite followed by removal of the solvent under reduced pressure. The resultant white solid was used for analysis. ^{1}H NMR (500MHz, $CDCl_3$) δ 1.36 (s, 3H), 1.52 (s, 3H), 4.53 (m, 1H), 5.12 – 5.17 (m, 2H), 5.62 (d, 1H), 5.78 (d, 1H), 7.45 (d, 1H), 9.4 (s, 1H), 10.25 (d, 1H)



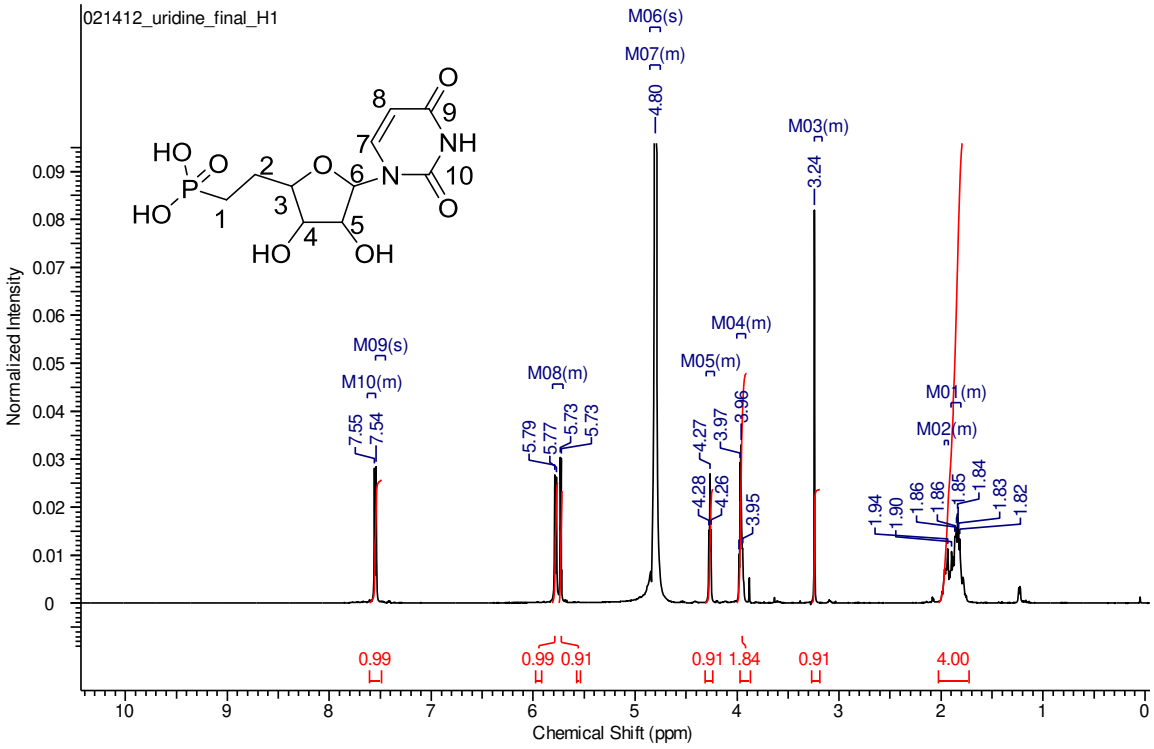
1-[5,6-Dideoxy-6-(diethoxyphosphinyl)-2,3-O-isopropylidene-D-ribo-hex-5-enofuranosyl] uracil (8a): The triphenylphosphorane (4)(1.63 mmol) obtained earlier was added to a stirred solution of 2',3'-*O*-isopropylideneuridine (1.63 mmol) in anhydrous DMSO (200 ml) at room temperature under N_2. The reaction was allowed to proceed overnight, and TLC analysis indicated completion of the reaction. The final ylide was extracted from the DMSO medium using CH_2Cl_2 and purified via silica gel chromatography and elution with 5% methanol in CH_2Cl_2. ^{1}H NMR (500MHz,$CDCl_3$) δ 1.22 (t, 6H), 1.33 (s, 3H), 1.58 (s, 3H), 4.02 – 4.08 (m, 5H), 4.64 (dd, 1H), 4.93 (dd, 1H), 5.16 (ddd, 2H), 5.87 (d, 1H), 5.97 – 6.01 (dt, 1H), 6.77-6.82 (ddd, 1H), 7.72 (d, 1H), 11.43 (s, 1H)

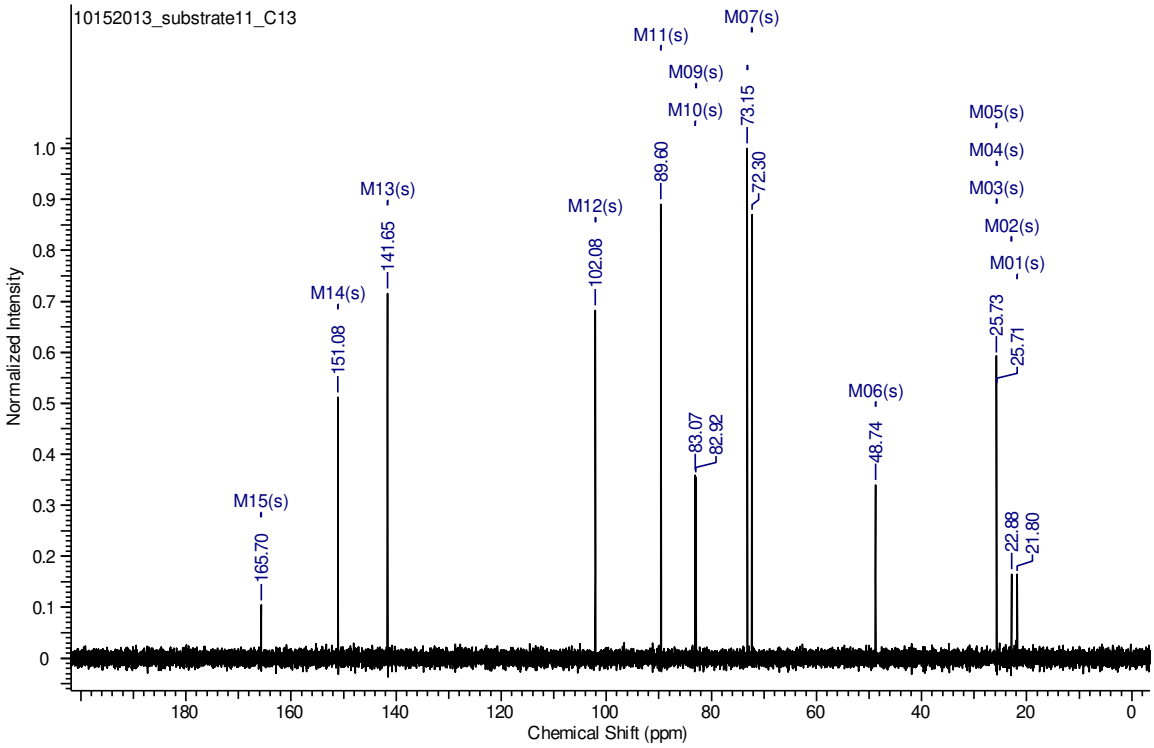


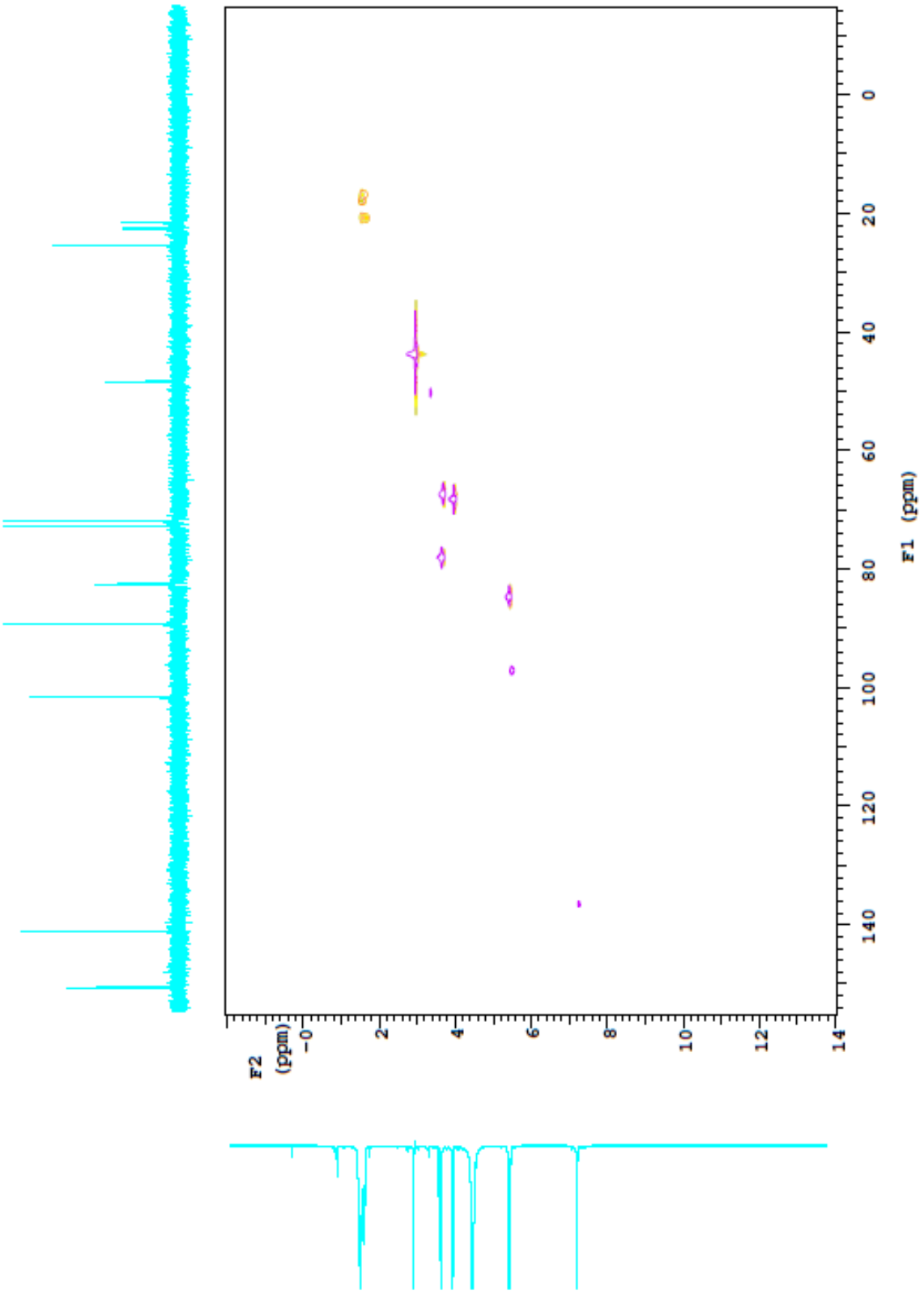
1-[5,6-Dideoxy-6-(diethoxyphosphinyl)-2,3-O-isopropylidene-β-D-ribo-hexofuranosyl] uracil (9a): 10% Pd/C (2 mg) was added to a solution of compound 8 (1 mmol) in methanol (10 ml) and the reaction was allowed to proceed under H_2 atmosphere overnight. The mixture was then filtered through Celite for removal of the catalyst and was shown by TLC to have been completely converted to the product. A yellowish oil was obtained upon concentration of the solution, and this was used for the next step without further purification. ^{1}H NMR (500MHz, $CDCl_3$) δ 1.25 (m, 9H), 1.47 (s, 3H), 1.72 – 1.95 (m, 4H), 3.94 (d, 1H), 4.03 (m, 4H), 4.56 (dd, 1H), 4.90 (dd, 1H), 5.56 (d, 1H), 5.66 (dd, 1H), 7.2 (d, 1H), 10.31 (s, 1H)



1-[5,6-Dideoxy-6-(dihydroxyphosphinyl)-β-D-ribohexofuranosyl] uracil (11): In the final step, TMSBr (1.626 mmol) was utilized for global deprotection of compound 9 (0.16 mmol) in CH_2Cl_2 (2 ml). After overnight stirring, the solvents were removed under reduced pressure and the residue was dissolved in water and lyophilized to give pure compound 11. ^{1}H NMR (500MHz, D_2O) δ 1.82 – 1.94 (m, 4H), 3.95 – 3.98 (m, 2H), 4.27 (app t, 1H), 5.73 (d, 1H), 5.79 (d, 1H), 7.55 (d, 1H); ^{13}C NMR (500MHz, D_2O) δ 21.8, 22.88, 72.3, 73.15, 82.92, 89.6, 102.08, 141.65, 151.08, 165.70. HRMS (ESI+) calcd. for $C_{10}H_{15}N_2O_8P$ $[M + H]^+$ 323.2098; found 323.0646.


021412_uridine_final_H1
M06(s)
M07(m)
4.80
M03(m)
3.24
M04(m)
M05(m)
M09(s)
M10(m)
M08(m)
M01(m)
M02(m)
7.55
7.54
5.79
5.77
5.73
5.73
4.27
3.97
3.96
4.28
4.26
3.95
1.94
1.90
1.86
1.86
1.85
1.84
1.83
1.82
0.99
0.99
0.91
0.91
1.84
0.91
4.00
Normalized Intensity
Chemical Shift (ppm)


10152013_substrate11_C13
M07(s)
M11(s)
M09(s)
M10(s)
73.15
89.60
72.30
M05(s)
M04(s)
M03(s)
M02(s)
M01(s)
M13(s)
141.65
M12(s)
102.08
M14(s)
151.08
25.73
25.71
M06(s)
48.74
83.07
82.92
M15(s)
165.70
22.88
21.80
Normalized Intensity
Chemical Shift (ppm)


F2 (ppm)
F1 (ppm)

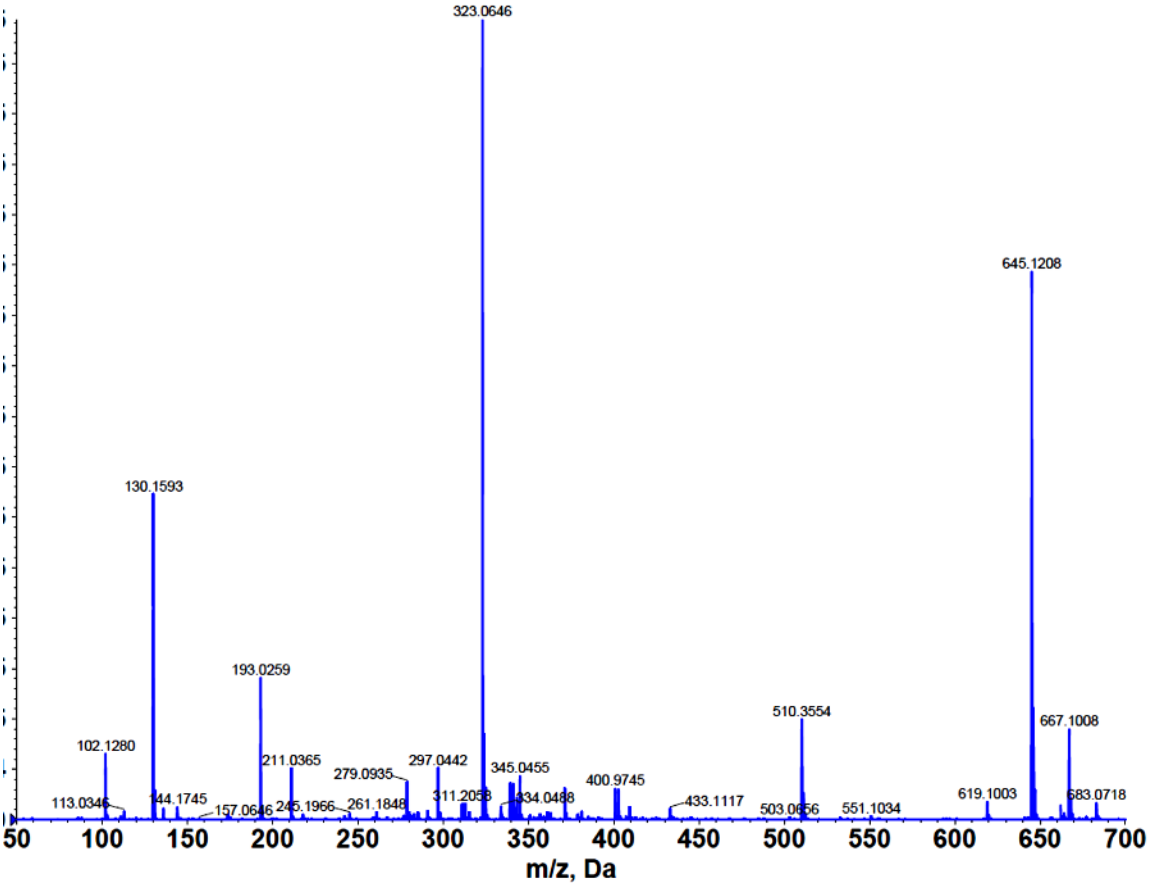


Figure 2.4.4.1. Spectroscopic characterization of S11. ^{1}H, ^{13}C, 2D ^{1}H – ^{13}C gHSQC (500MHz, D_2O) and HRMS for substrate analog **11**.

Table 1. Assignment of each peak of the final substrate analog **S11** to the corresponding ^{1}H and ^{13}C peaks obtained from the cumulative spectral data:

Position	δ_C, mult.	δ_H
1	21.73, CH_2	1.89, m
2	22.82, CH	1.96, m
3	82.86, CH	3.94
4	72.24, CH	3.98, m
5	89.54, CH	4.28, app t
6	73.09, CH	5.79, d
7	141.5, CH	7.55, d
8	102.2, CH	5.74, d
9	165.7, C	
10	151.03, C	

2.4.5. Cloning of genes for heterologous expression

The genes were amplified by PCR using the Expand Long Template PCR system from Roche with supplied buffer 2, 200mM dNTPs, 5% dimethyl sulfoxide, 10ng of DNA template, 5 units of DNA polymerase, and a 10mM concentration of each of the following primers (**Table 2**). DNA templates for PCR cloning were either *E. coli* DH5α genomic DNA (*EcRipA, EcUpp*), cosmid pN1 (*lipL* gene), pNCap02 (prepared using the genomic DNA of *Amycolatopsis* sp. SANK 60206) (*cpr19* gene) and Salmonella typhimirium plasmid pBRS11R (from Dr. Vern L. Schramm, Albert Einstein University, New York) (StPRPPsynthase). The thermocycler program included an initial hold at 94˚C for 10s, 56˚C for 15s, and 68˚C for 50s. The DNA fragment of the expected sizes were purified by 1% agarose gel and the purified PCR products were inserted into pET-30 Xa/LIC using ligation-independent cloning following the provided protocol to yield pET30- *Ecprpp,* pET30-*EcRipA,* pET30-*EcUpp,* pET30-*StPRPPsynthase,* pET30-*lipL* and pET30-*cpr19*. PCR-amplified DNA was then sequenced to confirm its identity.

Table 2. List of primers used

Primers	Oligonucleotide sequence
*StPRPPsynthase*_for	5'-GGTATTGAGGGTCGC ATGCCTGATATCAAGCTTTTTGCTGG-3'
StPRPPsynthase _rev	5'-AGAGGAGAGTTAGAGCCTCAATGCTCGAACATGGCGGAAATC-3'
*EcRpiA*_for	5'-GGTATTGAGGGTCGCATGACGCAGGATGAATTGAAAAAAG-3'
*EcRpiAi*_rev	5'-AGAGGAGAGTTAGAGCCTCATTTCACAATGGTTTTGACACC-3'
*EcUpp*_for	5'-GGTATTGAGGGTCGCATGAAGATCGTGGAAGTCAAAC-3'
*EcUpp*_rev	5'-AGAGGAGAGTTAGAGCCTTATTTCGTACCAAAGATTTTGTC-3'
*lipL*_for	5'- GTGATTGAGGGTCGCATGTCCGTGCTGGGGCGG - 3'
*lipL*_rev	5'- AGAGGAGAGTTAGAGCCTCATGAGGGCTTCTTGGTG – 3'
*cpr19*_for	5' - GGTATTGAGGGTCGCATGCAGCAGCTGCAAGCCG - 3'
*cpr19*_rev	5'- AGAGGAGAGTTAGAGCCTCAATTGGAGGCGCGGGG - 3'

The plasmids were introduced into *E. coli* BL21 (DE3) cells by transformation and the recombinant strains were grown in LB medium supplemented with 30 µg/ml kanamycin. Recombinant culture were then grown at 18˚C with 250 rpm, following inoculation with 500 ml of LB with 30 µg/ml kanamycin in a 2.5 L Erlenmeyer flask, until the cell density reached an OD_{600} = 0.5. Expression was induced with 0.1 mM isopropyl 1-thio-*β*-D-galactopyranoside (IPTG) and after overnight incubation at 18˚C, cells were harvested and lysed using a French press with one pass at 15,000 psi and immediately centrifuged at 18,000 rpm. The obtained supernatant was utilized for purification of the desired proteins using affinity chromatography with a nickel- nitriloacetic acid-agarose from followed by desalting of the recombinant proteins into 50mM Tris-HCl (pH 8), 100mM NaCl, and 5% glycerol using a PD-10 desalting column. The purified proteins were reconcentrated with an Amicon Ultra 10000 MWCO centrifugal filter prior to addition of glycerol (final 40%) for storage at -20˚C. Protein solubility and purity were assessed by 12% acrylamide SDS-PAGE; His_6-tagged proteins were utilized without further modifications.

Table 3. List of plasmids used

Strain/Plasmid	Characteristics and Relevance	References
E. coli Nova-blue	Host for routine cloning	Novagen
E. coli BL21 (DE3)	Host for protein expression	Novagen
pET30	Expression vector	Novagen
pET30- *StPRPPsynthase*	*StPRPPsynthase* gene cloned to pET30	This study
pET30- *EcRpiA*	*EcRpiA* gene cloned to pET30	This study
pET30- *EcUpp*	*EcUpp* gene cloned to pET30	This study
pET30- *lipL*	*lipL* gene cloned to pET30	This study
pET30- *cpr19*	*cpr19* gene cloned to pET30	This study

2.4.6. Enzymatic synthesis of 2',3',4',5',5''-$^2H_1$-uridine-5'-monophosphate(3)

Single reaction mixture (1 ml) consisted of 50mM Tris-HCl (pH 7.5), 10mM $MgCl_2$, 5mM uracil, 20 mM PEP, 1mM $NADP^+$, 1mM ATP, 2.5 mM α-KG, 1mM NH_4Cl, 1mM (1,2,3,4,5,6,6'-$^2H_1$)glucose, 80 U of hexokinase, 160 U of pyruvate kinase, 100 U of glucose-6-phosphate dehydrogenase, 160 U of glutamate dehydrogenase, 8 U of 6-phosphogluconate dehydrogenase, 25 µg of phosphoriboisomerase (EcRpiA), 100 µg of 5-phosphoribosyl-1-pyrophosphate synthetase (StPRPPsynthase), 25 µg of uracil phosphopribosyl transferase (EcUpp), 80 U of myokinase and 5 U of inorganic pyrophosphatase. The reaction was carried out at 30°C overnight and terminated by ultracentrifugation using a Microcon YM-3. Following removal of protein by centrifugation, the reactions were analyzed by HPLC using a C-18 reverse-phased column under ion-pairing conditions (monitored at 254 nm). A linear gradient of from 40 mM acetic acetic acid-triethylamine pH 6.5 (A) to 20% methanol (B) (0-4 min, 0% B; 4-24 min, 50% B; 24-26 min 100% B; 26-32 min, 100% B; 32-35 min, 0% B) with flow rate of 1 ml/min was used to analyze the reactions and elution was monitored at 260 nm. LC-MS was performed using a linear gradient from 0.1% formic acid in water to 0.1% formic acid in acetonitrile over 20 min. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm (**Figure 2.4.6.1**). Isolation of the deuterium labeled UMP product (**3**) starting from universally deuterated glucose was carried out with HPLC using a C-18 reverse-phase column using the aforementioned ion-pairing conditions. The peak corresponding to the product was collected and freeze-dried prior to mass-spectroscopic analysis (discussed in **section 2.5.2**).

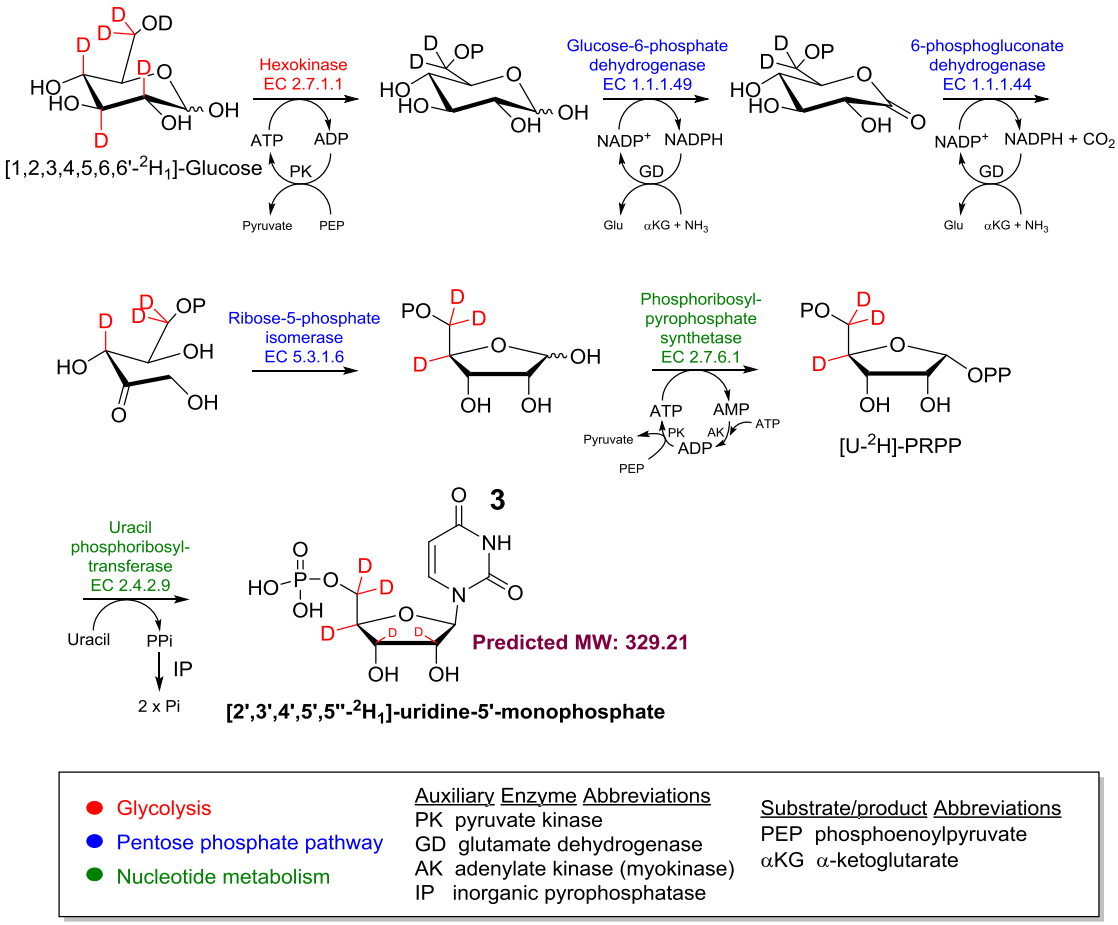


Figure 2.4.6.1. Schematic for the synthesis of deuterated UMP analog (3). The pathway is derived from an amalgamation of enzymatic steps from the glycolytic (**red**) and pentose phosphate pathways (**blue**), as well as from nucleotide metabolism (**green**). Three of the enzymes (His_6-phosphoriboisomerase (His_6-EcRpiA), His_6-uracil phosphopribosyl transferase (His_6-EcUpp), and His_6-5-phosphoribosyl-1-pyrophosphate synthetase (His_6-StPRPPsynthase) were synthesized in our lab; rest of the enzymes were available commercially.

2.4.7. *In-vitro* reactions with LipL and Cpr19:

Reactions with LipL typically consisted of 50mM Tris-HCl (pH 7.5), 1mM UMP (or 2',3',4',5',5''-$^2H_1$-UMP), 1.25 mM α-KG, 200 µM ascorbate, 100 µM $FeCl_2$, and 100 nM LipL

at 30°C. And reactions with Cpr19 typically consisted of 50mM Tris-HCl (pH 7.5), 1mM UMP, 1.25 mM α-KG, 1 mM ascorbate, 500 µM $FeCl_2$, and 100 nM Cpr19 at 30°C. Reactions were terminated by ultracentrifugation using a Microcon YM-3. Following removal of protein by centrifugation, the reactions were analyzed by HPLC using a C-18 reverse-phased column under ion-pairing conditions (monitored at 254 nm). A linear gradient of from 40 mM acetic acetic acid-triethylamine pH 6.5 (A) to 20% methanol (B) (0-4 min, 0% B; 4-24 min, 50% B; 24-26 min 100% B; 26-32 min, 100% B; 32-35 min, 0% B) with flow rate of 1 ml/min was used to analyze the reactions and elution was monitored at 260 nm. LC-MS was performed using a linear gradient from 0.1% formic acid in water to 0.1% formic acid in acetonitrile over 20 min. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm.

2.4.8. Kinetic characterization of LipL and Cpr19:

The activities of LipL and Cpr19 were detected by monitoring the formation of inorganic phosphate with the malachite green binding assay [108]. Detection of phosphate released as an enzyme product is a well-established technique for assaying phosphatase activity. For standardizing and quantifying the phosphate detected, the general procedure involves removal of 60µL of the reaction mixture and adding 20 µL of the MG reagent which is previously dispensed into individual wells in a 96-well format. The two were mixed well and left to incubate at room temperature for 5 min. We used 20mM (20 µL) EDTA (Ethylenediaminetetraacetic acid chelates Fe(II) from the enzyme) to terminate the reactions with LipL and Cpr19. Absorbance was measured at 620 nm using a microplate

reader. A standard curve was developed using phosphate standards of 0.8, 1.6, 3.2, 6.7, 12.5, 25, and 50 μM (provided in the kit). For single-substrate kinetic analyses involving LipL and Cpr19, reactions consisted of 50mM Tris-HCl (pH 7.5), 1 mM ascorbate, 500 μM $FeCl_2$, 100 nM LipL, near saturating α-KG (1mM) and variable UMP (50 μM– 1 mM). The reactions were initiated at 30°C by adding LipL (100 nM) or Cpr19 (100 nM) using the pipettor for mixing. Reactions were terminated after 3 min (<10% product formation) by addition of EDTA, using a sample without enzyme as a blank/control. Each data point represents triplicate end point assays. Kinetic constants were obtained by nonlinear regression analysis using GraphPad Prism.

For kinetic analysis of inhibition of LipL activity by synthesized substrate analog **11**, inhibition parameters were obtained by addition of variable concentrations of **11** (5 μM– 1 mM) to assays consisting of 50mM Tris-HCl (pH 7.5), 1 mM ascorbate, 500 μM $FeCl_2$, 100 nM LipL, near saturating α-KG (1mM) and variable UMP (100 μM– 10 mM). For each individual inhibition curve, we conducted the reactions by varying the concentration of **11** with a constant UMP concentration, and then repeating each study with a different concentration of UMP. The reactions were performed at 30°C for 3 min and analyzed under initial velocity conditions. Each data point represents a minimum of three replicate end point assays. For analyzing the inhibition constants, we generated cumulative data points from a total of five inhibition curves. We plotted a Lineweaver-Burke plot from the cumulative data using GraphPad Prism to arrive at the inhibition constants.

2.4.9. NMR and mass spectrometric characterization of synthetic standards

We received two synthetic standards from our collaborator Dr. Suzanne Peyrottes (University Montpellier, France) [109] that were used to test the identity of product obtained from the reaction of Cpr19 with synthetic phosphonate substrate analog **11**. Compound UA1768 was received as a 91:9 mixture of the sodium salts of 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'S)-hexofuranosyl]uracil and 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'R)-hexofuranosyl]uracil. In compound UA1923, the ratio was reversed. HPLC analysis (under ion pairing conditions) of the two reflect the elution profile and relative ratio of the standards **(Figure 2.4.9.1).**

1-[6'-Deoxy-6'-phosphono-β-D-ribo-(5'S)-hexofuranosyl]uracil (Disodium salt) (UA1768): Obtained as synthetic standard. ^{1}H NMR (300MHz, D_2O) δ 1.7 – 1.95 (m, 2H), 4.03 (t, 1H), 4.10 (m, 1H), 4.2 – 4.3 (m, 2H), 5.83 (d, 1H), 5.87 (d, 1H), 7.83 (d, 1H); ^{13}C NMR (300MHz, D_2O) δ 31.6, 67.2, 68.7, 73.5, 87.5, 87.9, 102.6, 141.9, 151.9, 166.1. HRMS (ESI+) calcd. for $C_{10}H_{16}N_2O_9P$ [M – Na + 2H]+ 339.0593; found 339.0592.

1-[6'-Deoxy-6'-phosphono-β-D-ribo-(5'R)-hexofuranosyl]uracil (Disodium salt) (UA1923): Obtained as synthetic standard. ^{1}H NMR (300MHz, D_2O) δ 1.70 – 1.90 (m, 2H), 4.01 (dd, 1H), 4.02 – 4.18 (m, 1H), 4.21 (dd, 1H), 4.3 (dd, 1H), 5.84 (d, 1H), 5.89 (d, 1H), 7.95 (d, 1H); ^{13}C NMR (300MHz, D_2O) δ 32.0, 67.3, 70.3, 73.7, 87.0, 88.6, 102.4, 142.0, 151.8, 166.2. HRMS (ESI+) calcd. for $C_{10}H_{16}N_2O_9P$ [M – Na + 2H]+ 339.0593; found 339.0592.

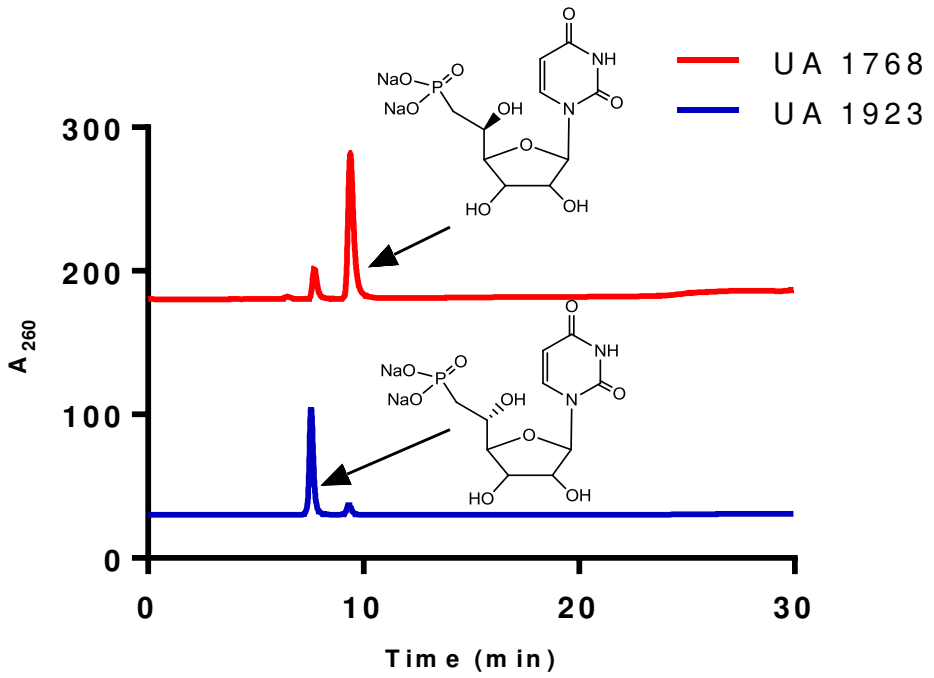


Figure 2.4.9.1. Synthetic standards. UA1768 consists of the major peak corresponding to the *S*-hydroxy-phosphonate, and UA1923 contains *R*-hydroxy-phosphonate as the major peak.

2.5. Results

2.5.1. *In vitro* characterization of LipL and Cpr19:

In previous studies, the *orf* from A-90289 was shown to encode the protein LipL which was functionally assigned by our lab as a non-heme, Fe(II)-dependent α-ketoglutarate (α-KG): uridine-5'-monophosphate (UMP) dioxygenase catalyzing the conversion of UMP to uridine-5'-aldehyde during A-90289 biosynthesis [70]. Cpr19 is a homolog of LipL from the

strain A-102395 [110], and has been characterized similarly to carry out a net two-electron oxidation of UMP to the final aldehyde product (**Figure 2.5.1.1**).

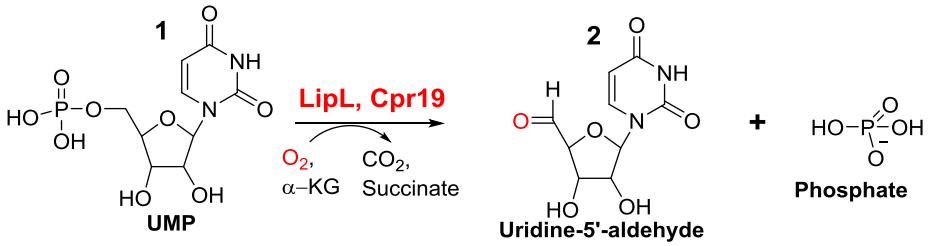


Figure 2.5.1.1. Reaction mechanism of LipL and Cpr19. Net two-electron oxidation of UMP (**1**) to uridine-5'-aldehyde (**2**) catalyzed by homologs LipL (A-90289) and Cpr19 (A-102395).

The *lipL* and *cpr19* genes were cloned and expressed in E. coli to yield soluble protein with the expected sizes (**Figure 2.5.1.2**). HPLC analysis of both reactions catalyzed by LipL and Cpr19 revealed a peak corresponding to the product uridine-5'-aldehyde (**2**) and the tris-adducts of the aldehyde eluting later (red and blue spectra) (**Figure 2.5.1.3**). Negative control consists of the reaction mixture sans the enzymes, wherein substrate UMP (**1**) elutes at t =12 min (black trace). LC-MS analyses of both reactions revealed $(M-H)^-$ ions at *m/z* = 240.8, that confirms the identity of the aldehyde product from the reaction of both enzymes (**Figure 2.5.1.4**).

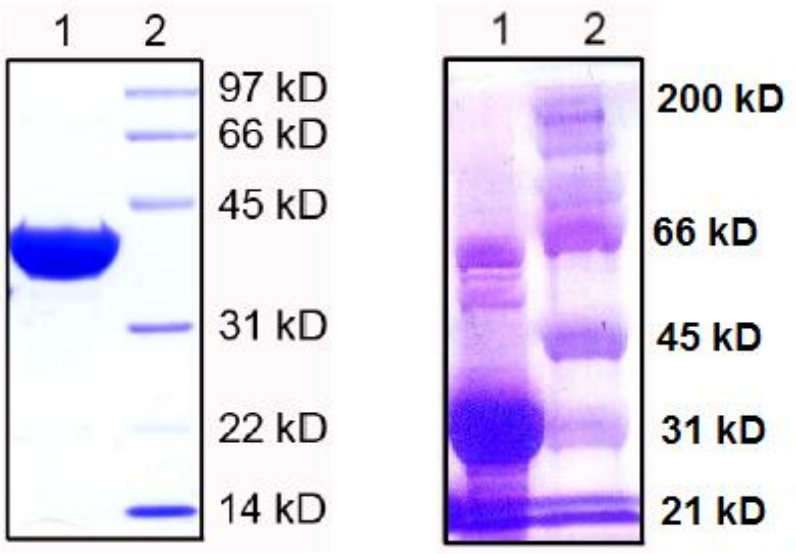


Figure 2.5.1.2. SDS-PAGE analysis of purified proteins (A) His_6-LipL (expected 38.2 kD) and **(B)** His_6-Cpr19 (expected 31.7 kD)

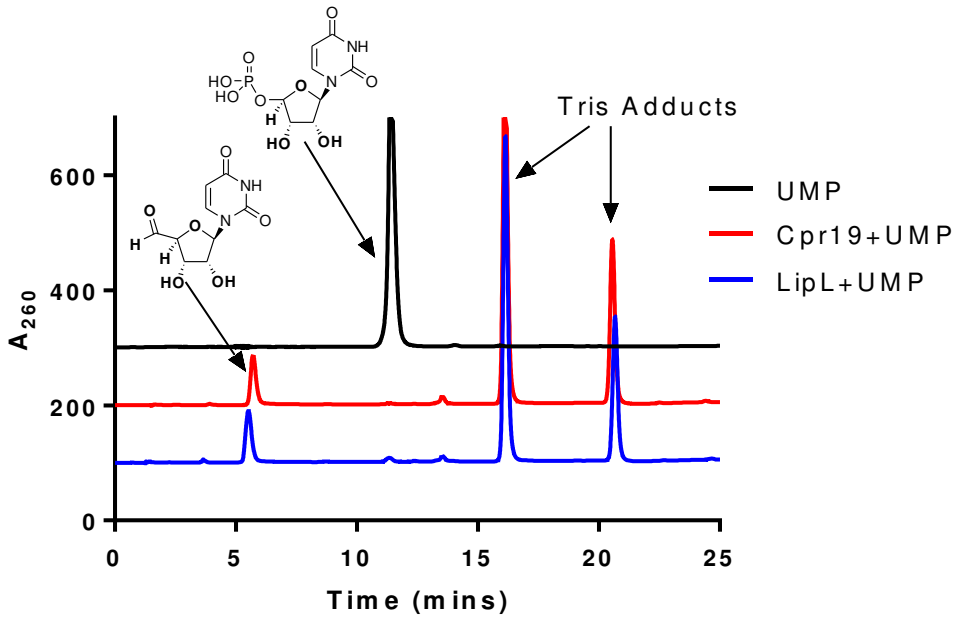


Figure 2.5.1.3. HPLC analyses of *in-vitro* reactions catalyzed by LipL and Cpr19. HPLC trace for negative control (in the absence of enzymes) is depicted in **black**, and for the reactions catalyzed by Cpr19 (**red**) and LipL (**blue**) complete conversion of the UMP (**1**) peak (eluting at t =12 min) leads to the uridine-5'-aldehyde product (**2**) eluting at t = 5 min (along with its corresponding mono- and bi- tris-adducts, as labeled). A_{260}: absorbance at 260 nm.

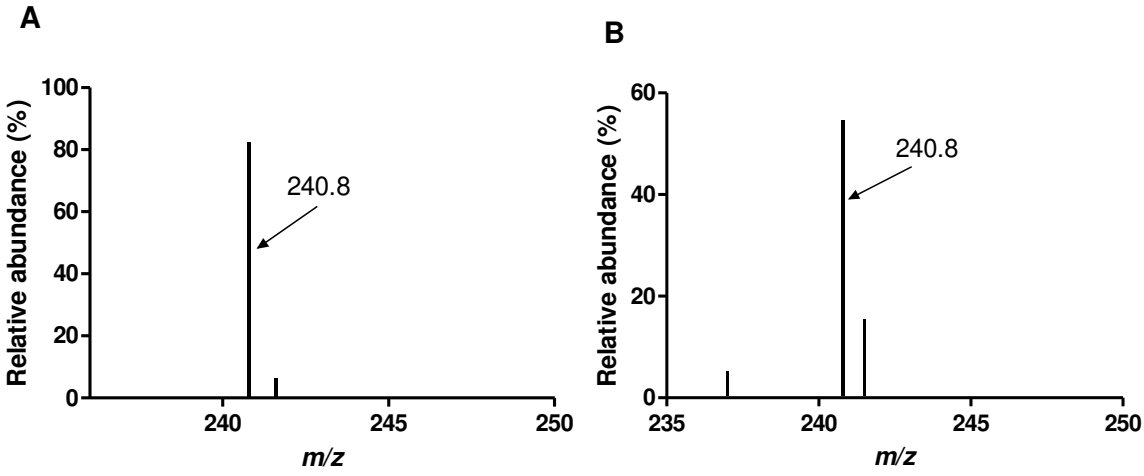


Figure 2.5.1.4. LC-MS analysis of the dioxygenase reactions. (A) Mass spectrum for the ion peak eluting at t =3.9 min for the reaction catalyzed by LipL, (B) Mass spectrum for the same peak, obtained from Cpr19 reaction.

2.5.2. *In-vitro* synthesis of 2',3',4',5',5''-$^{2}H_{1}$-uridine-5'-monophosphate(3)

The genes for phosphoriboisomerase (EcRpiA), uracil phosphopribosyl transferase (EcUpp) and 5-phosphoribosyl-1-pyrophosphate synthetase (StPRPPsynthase) were cloned and expressed in E. coli to yield soluble protein with the expected sizes (**Figure 2.5.1.1**). HPLC analysis of the one-pot synthesis revealed a peak with retention time t = 12 min, as is expected for UMP (**Figure 2.5.2.1**). The peak was collected (as described in **section 2.4.6**) and LC-MS was utilized for confirming the identity of the product, and to establish the retention of four deuterium labels (**Figure 2.5.2.3**).

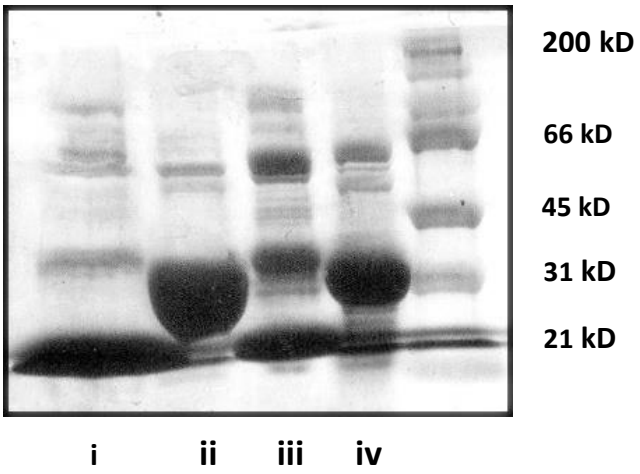


Figure 2.5.2.1. SDS-PAGE analysis of purified proteins (i) His_6- phosphoriboisomerase or His_6-EcRpiA (expected 22.8 kD), (ii) His_6-Cpr19 (expected 31.7 kD), (iii) His_6-uracil phosphoporibosyl transferase or His_6-EcUpp (expected 22.5 kD) and (iv) His_6-5-phosphoribosyl-1-pyrophosphate synthetase or His_6- StPRPPsynthase (expected 34 kD).

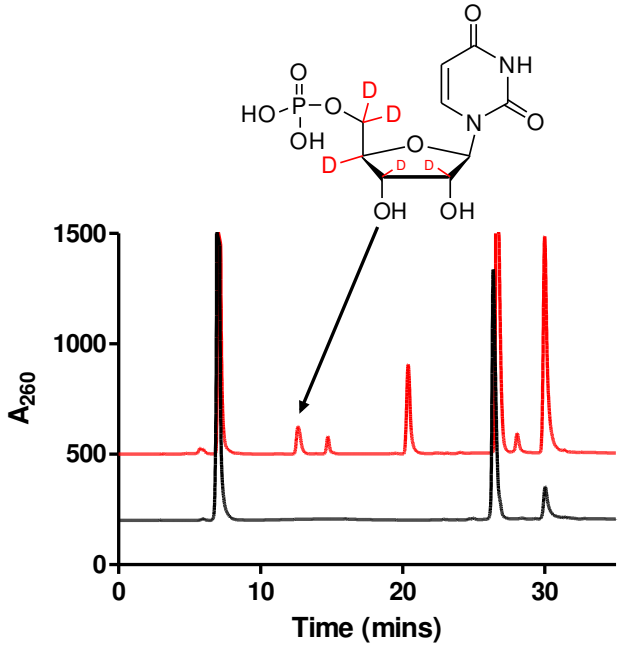


Figure 2.5.2.2. HPLC trace for the one-pot reaction. The pawthay is engineered from a combination of reactions adopted from the glycolytic and pentose phosphate pathways, and from nucleotide metabolism. The observed peak at t = 12 min corresponds to the

desired deuterium labeled UMP product (**3**) from this reaction. A_{260}: absorbance at 260 nm.

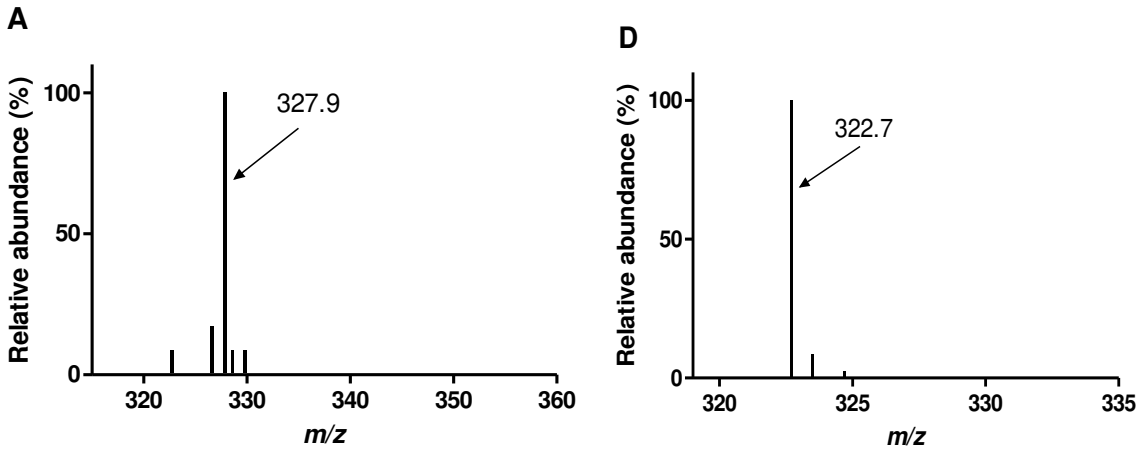


Figure 2.5.2.3. LC-MS analysis of the peak collected from the one-pot reaction. (A) Mass spectrum for the ion peak eluting at t = 4.08 min, corresponding to 2',3',4',5',5''-$^{2}H_1$-UMP (**3**) (calculated 329.3) (B) Mass spectrum for the peak eluting at t = 4.09 min, obtained from injecting standard UMP (calculated 324.7). Comparison of the two LC-MS results help in establishing the retention of five $^{2}H_1$ labels in the synthesized product.

2.5.3. *In-vitro* utilization of 2',3',4',5',5''-$^{2}H_1$-uridine-5'-monophosphate(3) by LipL and Cpr19

The synthesized 2',3',4',5',5''-$^{2}H_1$-uridine-5'-monophosphate(**3**) was tested in reaction *in vitro* with both LipL and Cpr19. Initial activity tests with HPLC revealed that UMP (**1**) is converted to uridine-5'-aldehyde (**2**), in line with previous observed spectra (**Figure 2.5.1.2**). Keeping with our earlier experiences with the inherent instability associated with the aldehyde product, we did not try to collect the product peak from HPLC. Instead, we directly injected the reaction mixture into the LC-MS and observed a peak with $(M-H)^-$ ion at *m/z*=244.8 corresponding to 2',3',4',5',-$^{2}H_1$-uridine-5'-aldehyde (calculated 246.8). As

negative control, we injected a reaction mixture obtained from the reaction of LipL (and/or Cpr19) with standard UMP (**1**). This reaction mixture revealed a peak with $(M-H)^-$ ion at *m/z*=322.7, corresponding to unlabeled uridine-5'-aldehyde (**2**) (**Figure 2.5.3.1**). This result led to the conclusion that four $^2H_1$ labels from (**3)** are retained in the corresponding aldehyde product.

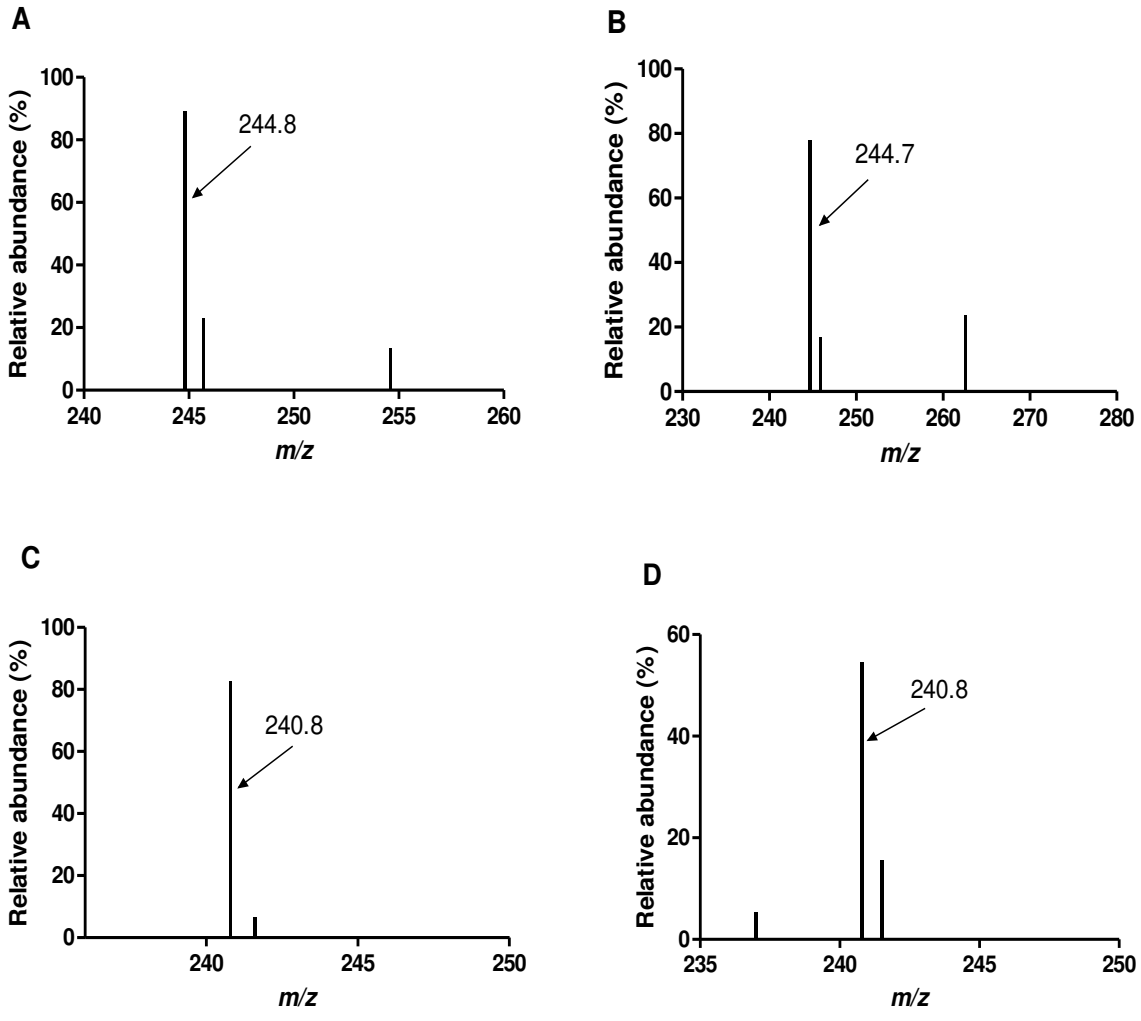


Figure 2.5.3.1. LC-MS analysis of the peak collected from the reactions of LipL (A) and Cpr19 (B) with 3. Mass spectrum for the ion peaks eluting at t = 4.08 min, corresponding to 2',3',4',5',-$^2H_1$-uridine-5'-aldehyde (*m/z*=246.8) (C) Mass spectrum for the peak eluting

at t = 4.09 min, obtained from injecting reaction mixture consisting of standard UMP with LipL and (D) with Cpr19. The $(M-H)^-$ ion at *m/z*=240.8 corresponds to unlabeled uridine-5'-aldehyde. These results confirm the retention of $^2H_1$ labels at the C-4' and C-5' positions in UMP.

2.5.4. *In-vitro* utilization of synthetic phosphonate substrate analog (11) by LipL and Cpr19

Initial activity tests using HPLC revealed that LipL did not recognize synthesized analog **11** as a substrate. Variation of both substrate concentration and enzyme concentrations proved futile (**Figure 2.5.4.1**). However, HPLC tests with Cpr19 and **11** were more productive, with the observation of an (albeit small) product peak eluting directly ahead of the substrate **11** peak (**Figure 2.5.4.2)**.

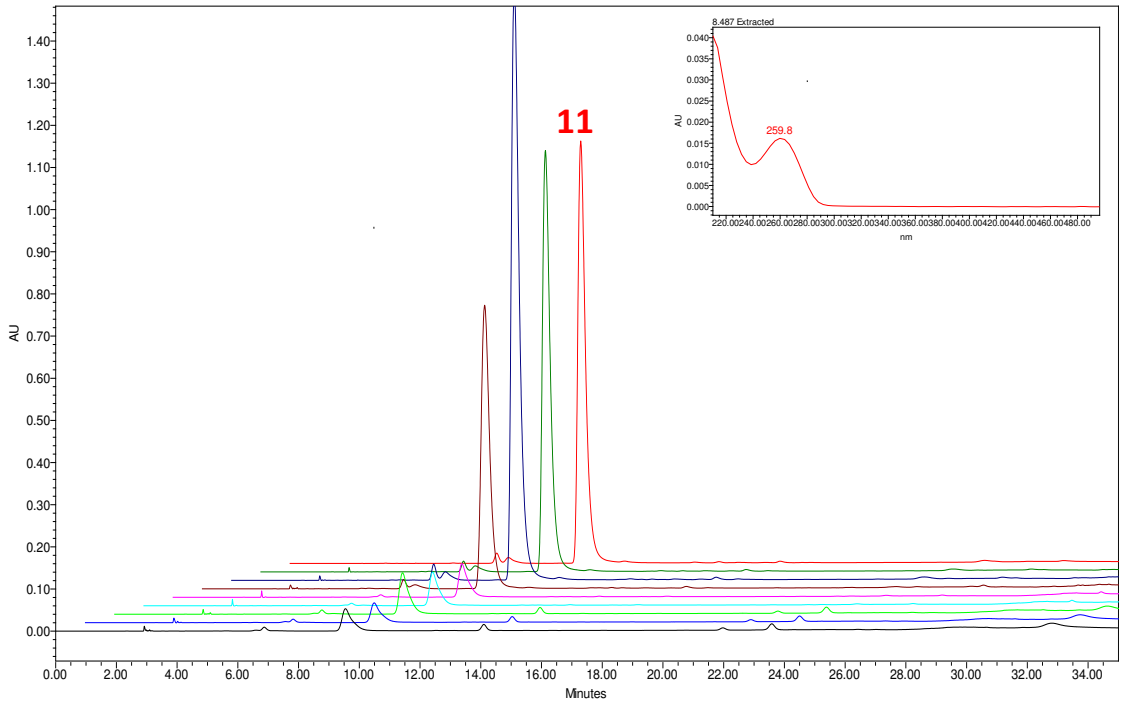


Figure 2.5.4.1. Reaction of LipL with S11. Comparative HPCL analyses of reactions carried out with phosphonate substrate analog **11** with LipL, under variable concentrations of **11**. The observed peak corresponds to unutilized **11** in reaction. None of these studies yielded a product peak. Inset: Absorbance maxima at 260 nm.

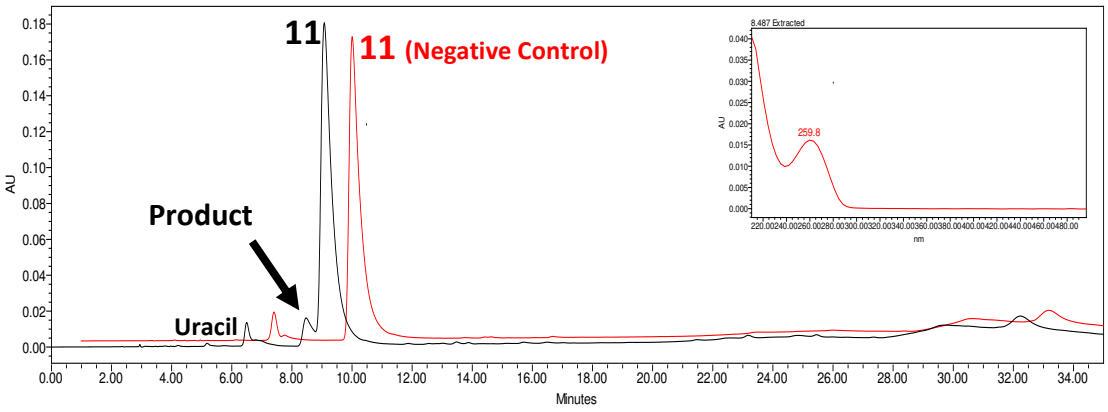


Figure 2.5.4.2. Reaction of Cpr19 with S11. Comparative HPLC analyses of the reaction of 11 conducted with Cpr19. The major peak is 11, and the small peak at t =7 min is residual uracil (decomposition product from reactions). (A) Negative control, which consists of the reaction mixture *without* Cpr19. (B) Addition of Cpr19 to the reaction mixture yields a product peak with absorbance at 260 nm (inset).

2.5.5. Spectroscopic characterization of Cpr19 product

We used the synthetic standards received from our collaborators as a means for confirming the identity of the product observed in the reaction of Cpr19 with substrate analog **11**. Comparative HPLC analyses of the negative control (reaction mixture in the absence of Cpr19), synthetic standards, reaction mixture with Cpr19, and reaction mixture co-eluting with synthetic standard UA1768 confirmed the identity of the observed product peak as 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'S)-hexofuranosyl]uracil (**5**) (**Figure 2.5.5.1**).

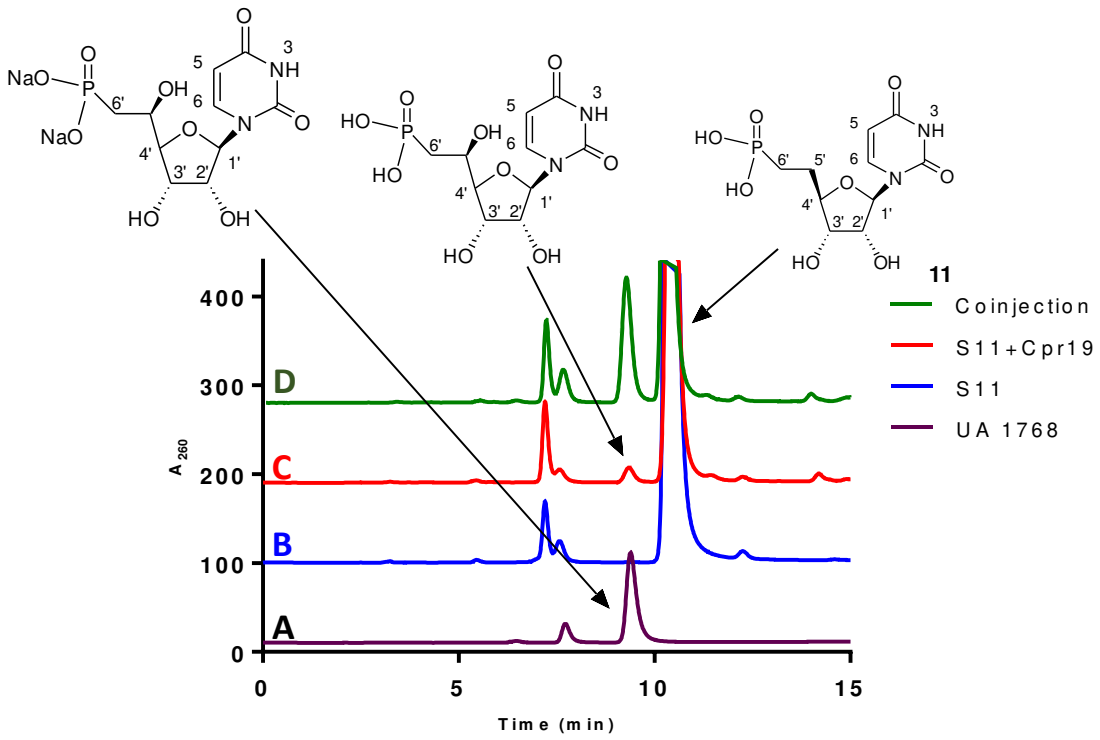


Figure 2.5.5.1. Comparative HPLC traces for confirming the identity of the Cpr19 product. (A) UA1768 synthetic standard, (B) Negative control consisting of **11** in reaction mixture, in the absence of enzyme, (C) Formation of product peak from **11**, catalyzed by Cpr19, (D) Co-elution of reaction mixture with UA1768 indicates enrichment of the peak.

2.5.6. Production of the Cpr19 product for NMR and mass spectroscopic analyses:

Large scale isolation of the Cpr19 product starting from synthesized UMP analog (**11**) was carried out with HPLC using a C-18 reverse-phase semipreparative column using previously described iron-pairing conditions. The peak corresponding to the product 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'*S*)-hexofuranosyl]uracil (**5**) was collected and freeze-dried prior to HRMS, 1D and 2D NMR spectroscopic analysis (**Table 4, Figures 2.5.6.1-2.5.6.5**). ^{1}H NMR (600MHz, D_2O) δ 1.89 (m, 2H), 4.09 (app t,1H), 4.14 (m,1H), 4.29 – 4.30 (m, 2H), 5.86 (d, 1H), 5.93 (d, 1H), 7.90 (d, 1H); ^{13}C NMR (600MHz, D_2O) δ 27.5, 67.5, 68.7, 73.5, 87.5, 102.2, 141.5, 151.9, 165.7. HRMS (ESI+) calcd. for $C_{10}H_{15}N_2O_9P$ $[M-H]^-$:337.0463; found 337.04652.

Table 4. Assignment of each peak of the Cpr19 product (**5**) to the corresponding ^{1}H and ^{13}C peaks obtained from the cumulative spectral data:

Position	δ_C, mult.	δ_H (*J* in Hz)	COSY	HMBC
1	27.5, CH_2	1.89, m	2	
2	67.5, CH	4.14, m	1	
3	87.5, CH	4.09, app t	4	
4	68.7, CH	4.30, m	3	
5	73.5, CH	4.29, m	6	
6	87.5, CH	5.93, d (5.7)	5	
7	141.5, CH	7.90, d (8.1)	8	9, 10
8	102.2, CH	5.86, d (8.1)	7	
9	165.7, C			
10	151.9, C			

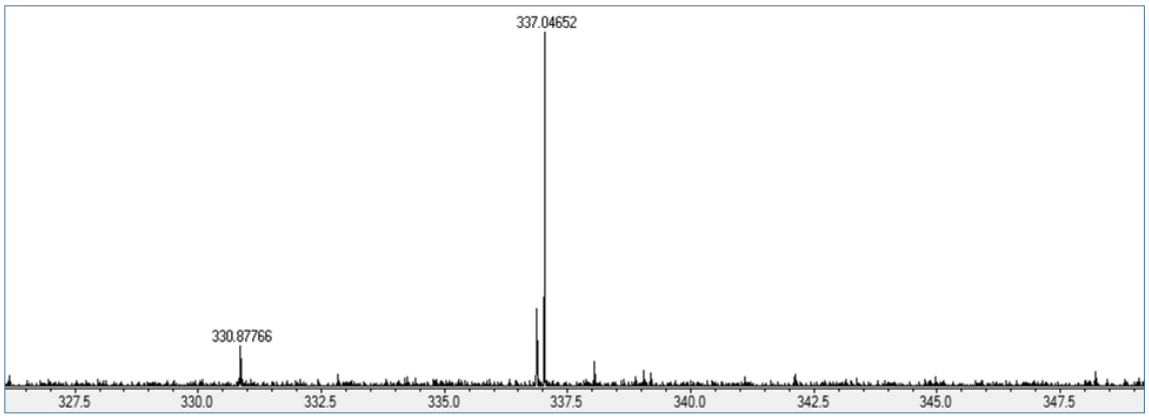


Figure 2.5.6.1. HRMS of the collected Cpr19 product peak with S11. Expected $[M-H]^-$:337.0463; found 337.04652.

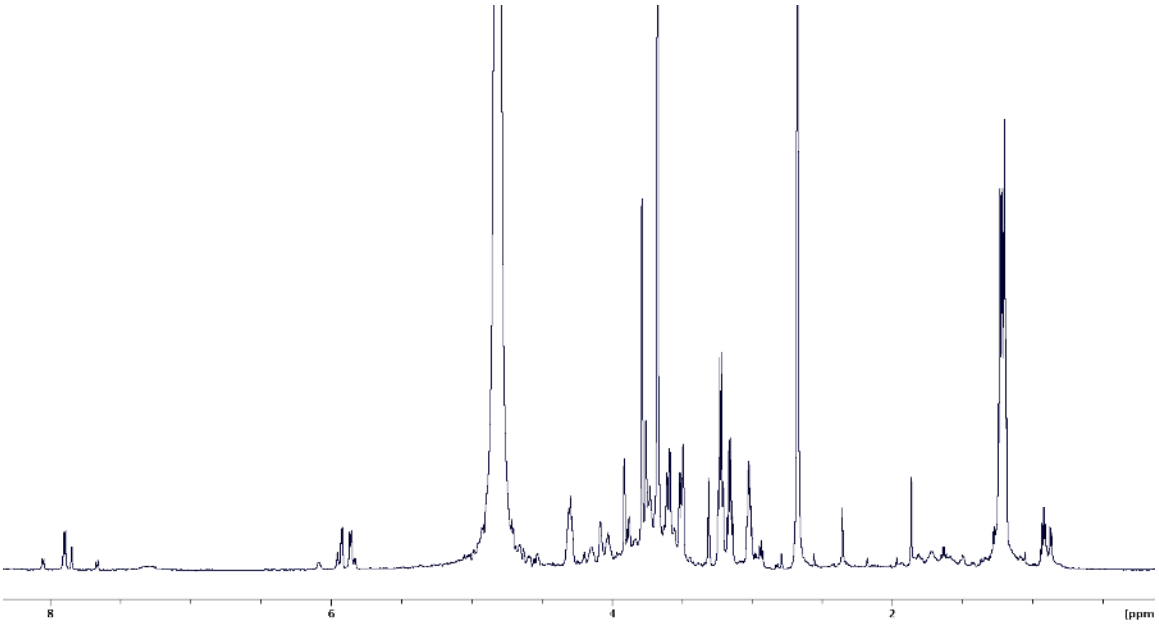


Figure 2.5.6.2. ^1H NMR (600MHz, D_2O) for Cpr19 product peak (**5**).

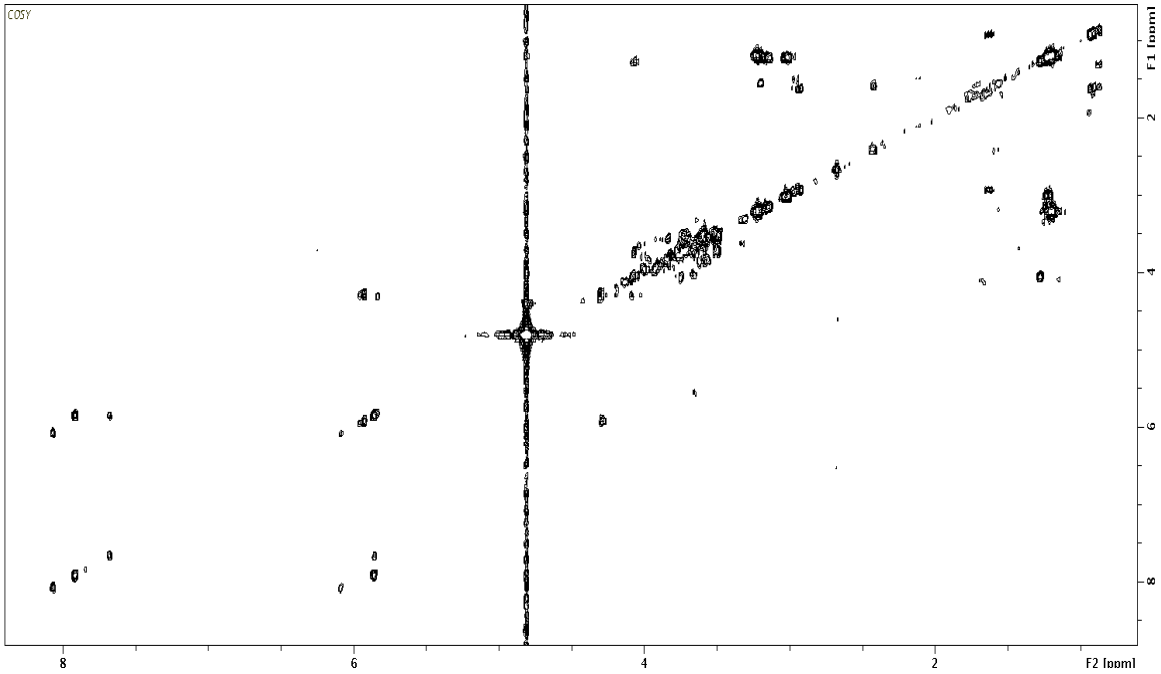


Figure 2.5.6.3. 2D ^{1}H - ^{1}H gCOSY (600MHz, D_2O) for Cpr19 product peak (**5**).

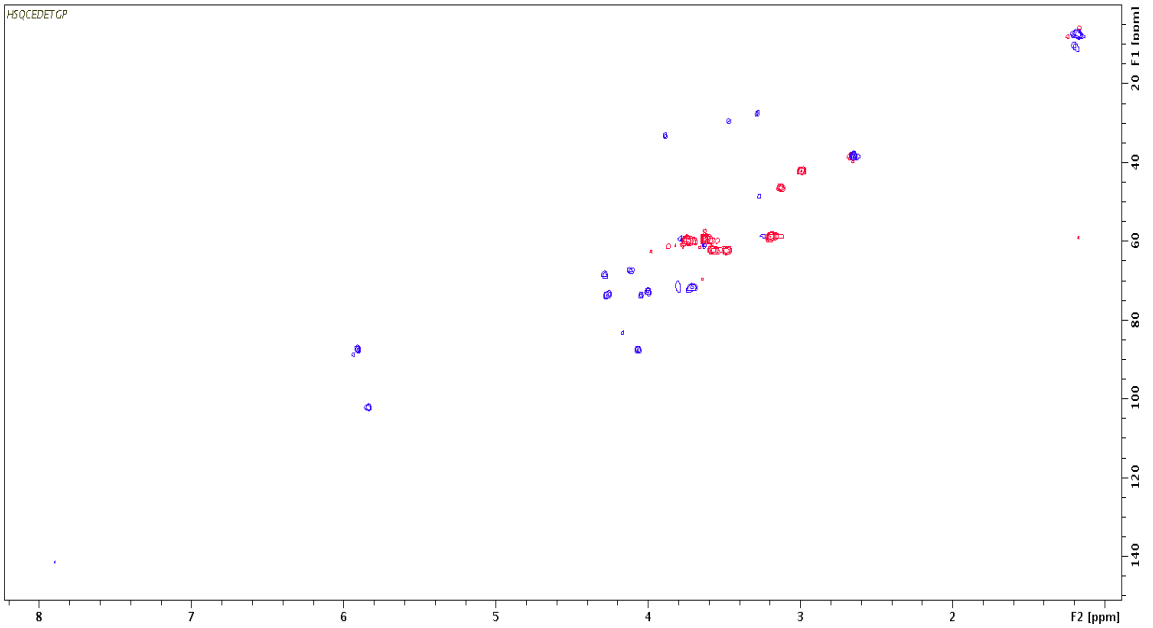


Figure 2.5.6.4. 2D ^{1}H – ^{13}C gHSQC (600MHz, D_2O) for Cpr19 product peak (**5**).

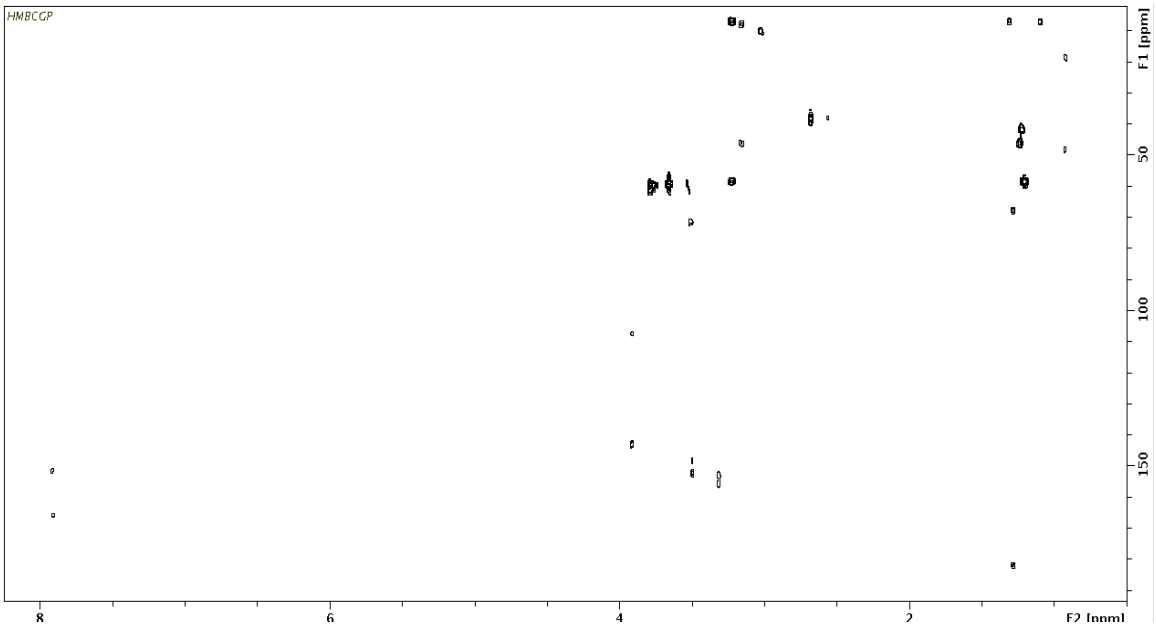


Figure 2.5.6.5. 2D ^{1}H – ^{13}C gHMBC (600MHz, D_2O) for Cpr19 product peak (**5**).

2.5.7. Kinetic characterization of LipL and Cpr19, with respect to substrate UMP

For single-substrate kinetic analyses involving LipL and Cpr19, reactions were conducted at 30°C at pH=7.5. Reactions were terminated after 3 min (<10% product formation) by addition of EDTA, using a sample without enzyme as a blank/control. Each data point represents triplicate end point assays. Kinetic constants were obtained by nonlinear regression analysis using GraphPad Prism (**Figure 2.5.7.1**). The extracted kinetic constants are listed in **Table 5**.

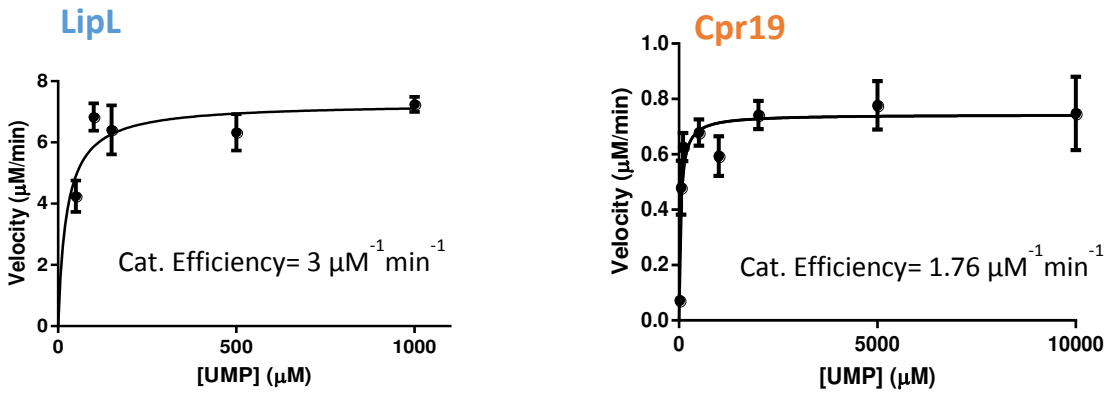


Figure 2.5.7.1. Kinetic analyses of LipL and Cpr19. Single-substrate kinetic analysis with variable UMP (1) at pH=7.5 for LipL and Cpr19 obtained from malachite-green assay.

For kinetic analysis of inhibition of LipL activity by synthesized substrate analog **11**, inhibition parameters were obtained by addition of variable concentrations of **11** (5 μM, 10 μM , 50 μM , 500 μM and 1 mM) to assays consisting of 50mM Tris-HCl (pH 7.5), 1 mM ascorbate, 500 μM $FeCl_2$, 100 nM LipL, near saturating α-KG (1mM) and variable UMP (100 μM– 10 mM). For each individual inhibition curve, we conducted the reactions by varying the concentration of **11** with a constant UMP concentration, and then repeating each study with a different concentration of UMP. The reactions were performed at 30°C for 3 min and analyzed under initial velocity conditions. Each data point represents a minimum of three replicate end point assays. For analyzing the inhibition constants, we generated cumulative data points from a total of five inhibition curves. We plotted a Lineweaver-Burke plot (**Figure 2.5.7.2**) from the cumulative data using Graphpad Prism to arrive at the inhibition constants.

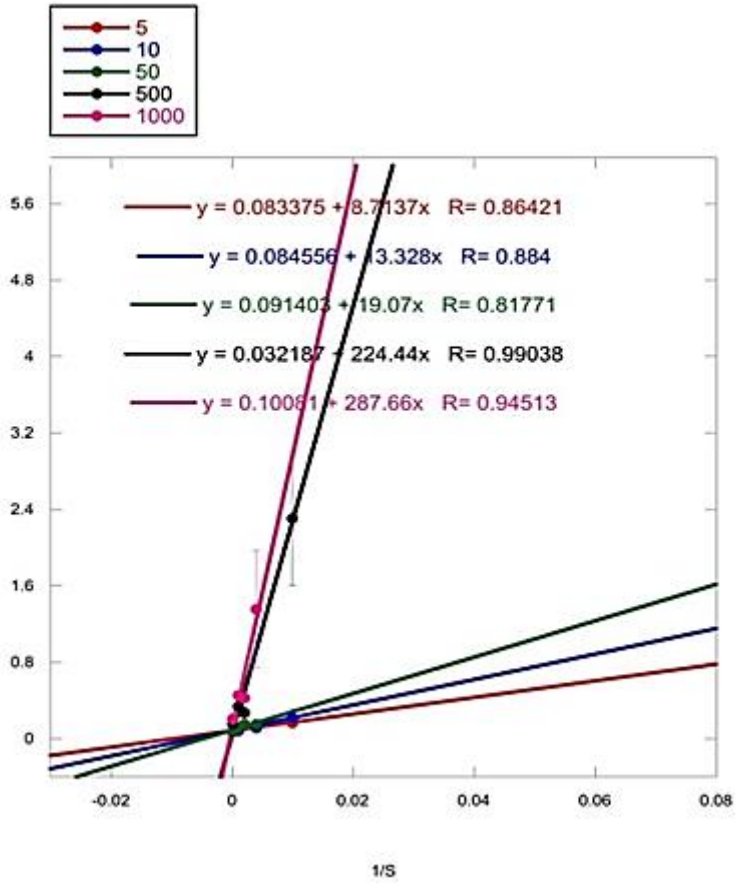


Figure 2.5.7.2. Lineweaver-Burke plot for competitive inhibition of LipL by synthesized substrate analog 11. The plot is obtained by plotting inverse substrate (UMP) concentration (1/S) (in μM^{-1}) on the X-axis versus the inverse of velocity (1/V) (in μM^{-1}min) on the Y-axis, and then calculating K_i from the intercept on the Y-axis. The extracted inhibition constant is K_i= 800 nM. Legend on the top represents the different concentrations of **11** (in µM) used to generate each reciprocal graph.

Table 5. Kinetic constants for LipL and Cpr19

Enzyme	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat} / K_m ($\mu M^{-1} min^{-1}$)	Ki (nM)
LipL	UMP	24 ± 10	72 ± 20	3	--
LipL	UMP + (**11**) (competitive inhibitor)	--	--	--	800
Cpr19	UMP	42 ± 14	74 ± 10	1.76	--

2.6. Conclusion

In summary, we have established that the dixoygenases LipL (from A-90289 pathway) and Cpr19 (A-102395) proceed via a 'hydroxylation' mechanism to generate an unstable hydroxylated intermediate by functionalizing prime substrate UMP (**1**) at C-5'. We can rationally predict that this intermediate can spontaneously dephosphorylate to yield free phosphate and the uridine-5'-aldehye (**2**) product that is utilized variously further down the biosynthetic pathways for nucleoside antibiotics (**Figure 1.6.1**). Mass spectrometric characterization of the product obtained from the utilization of selectively deuterated UMP (**3**) by both LipL and Cpr19 in reaction provides indubitable evidence for rejecting the 'desaturase' hypothesis, as was predicted earlier as an alternative pathway. The surrogate hydroxylated intermediate (**5**) characterized as the Cpr19 product by employing synthesized surrogate substrate (**11**) in reaction mirrors a hydroxylated UMP intermediate in the original reaction. The end result is a pathway highlighted in **Figure 2.6.1**. Our observations from this pathway can be rationally extended to the reaction

catalyzed by TauD in predicting the existence of the hydroxylated taurine intermediate, which has remained elusive to date (67).

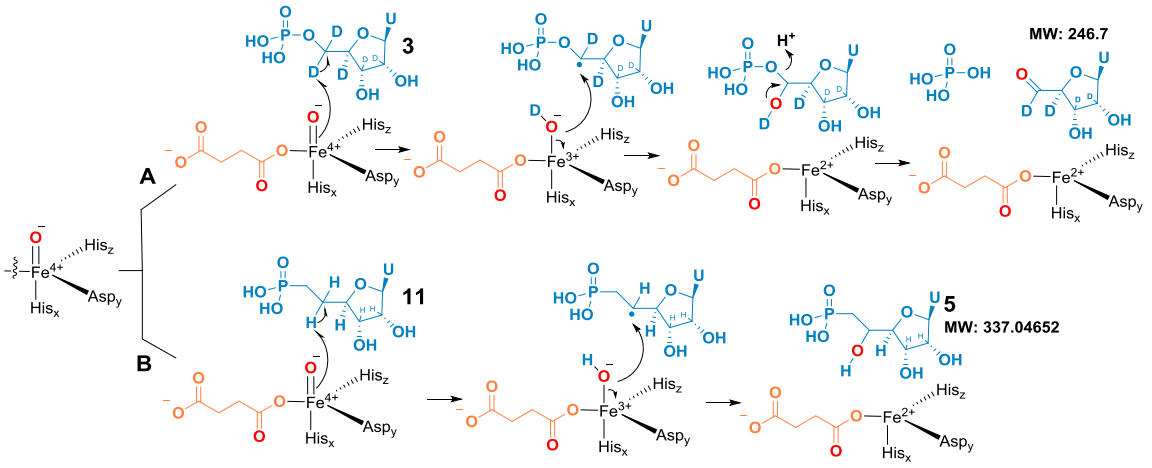


Figure 2.6.1. **Revised biosynthetic pathway for dioxygenases LipL and Cpr19 via a 'hydroxylation' mechanism.** (A) Employment of deuterated substrate UMP (**3**) resulting in deuterated aldehyde product, (B) Utilization of synthetic substrate analog (**11**) to 'trap' hydroxylated intermediate (**5**).

2.7. Discussion

Enzymes from the Fe(II): α-KG-dependent superfamily catalyze a diverse array of biotransformations ranging from hydroxylations, stereoinversions, desaturations to ring closure and ring expansions [75]. Bioinformatic analyses of several recently discovered gene clusters for minimally two related families of nucleoside antibiotics revealed a shared open reading frame encoding a protein with sequence similarity to TauD, the best characterized member of this superfamily. We have characterized representative

enzymes from the A-90289 and A-503083 gene clusters to reveal yet another diverse biotransformation catalyzed by members of this superfamily. With this, we have thus established uridine-5'-aldehyde as the starting intermediate for the biosynthesis of both C7 high carbon nucleosides, as well as for the C6 high carbon nucleosides **(Figure 1.6.1**). It is also reasonable to predict that other high carbon nucleoside antibiotics such as the uracil-containing tunicamycin [68] and A-9464 [111], the adenine containing griseolic acid [112], and cytosine-containing ezomycin A_1 [113] may employ a similar enzymatic strategy to initiate the organization of their respective high carbon sugar moieties (**Figure 2.7.1**). We can therefore use the information from our gene clusters as a platform to rationally predict starting metabolites and similar biosynthetic assembly in related high carbon sugar nucleosides. Interestingly, phosphate release by LipL has been prognosticated as a potential additional function for this dioxygenase family, analogous to sulfate and sulfite scavenging by TauD and AtsK [74], and we have demonstrated that this is indeed plausible by providing the enzyme precedence for this chemistry.

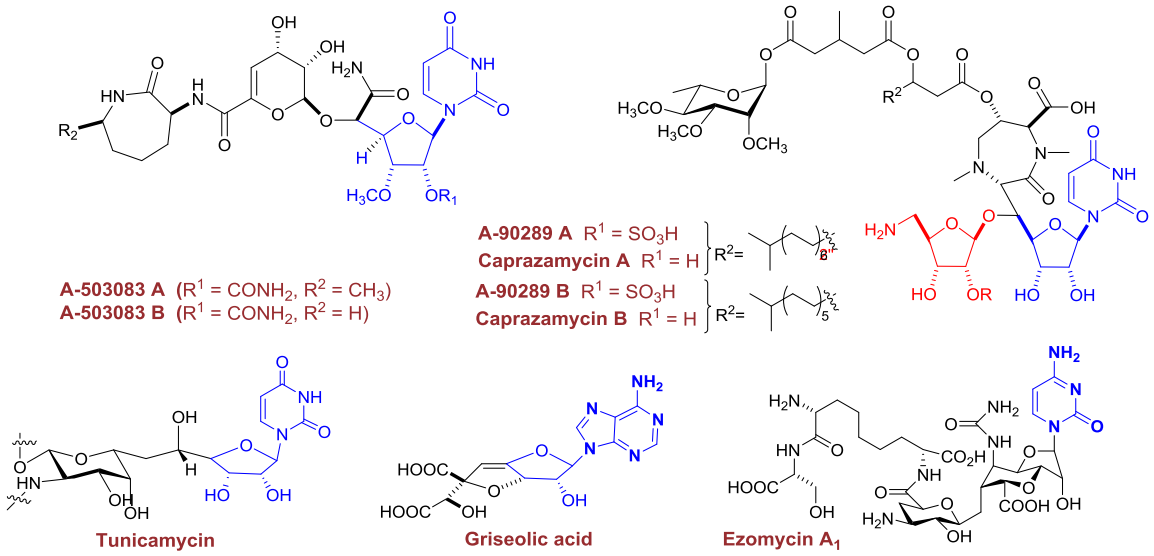


Figure 2.7.1. Platform for pathway prediction. High carbon sugar nucleosides with predicted biosynthetic pathways that may include an activity similar to LipL and Cpr19.

Most of the biochemical characteristics for LipL and Cpr19 have been reported for other enzymes of this family of enzymes. For example, activity for both enzymes is stimulated by ascorbate, certain divalent metals inhibit activity, oxidative decarboxylation of α-KG can still occur in the absence of prime substrate UMP [70]. For LipL, activity was shown to be absolutely dependent upon the presence of α-KG and molecular O_2, and isotope enrichment studies conducted with $^{18}O_2$ and H_2O^{18} established the incorporation of one of the oxygen atoms into the byproduct succinate. Unfortunately, the rapid and reversible formation of a germinal diol in water did not allow the analysis of putative oxygen incorporation into uridine-5'-aldehyde. While LipL is specific for both prime substrate and α-KG and does *not* recognize substrate analog **S11,** Cpr19 recognizes both the substrate analog as well as some other α-keto acids as cosubstrate (Wenlong Cai, unpublished data). Keeping with our observations, we characterized **S11** as a *competitive* inhibitor

based on a lineweaver burke plot derived for its inhibitory effect on LipL activity (Ki=800 nM).

Clearly, the unique feature of LipL and Cpr19 is the specific utilization of UMP and the mechanism by which net dephosphorylation and oxidative decarboxylation of α-KG takes place. In our studies, we have provided convincing proof that, instead of tandem phosphatase and oxidoreductase mechanisms as was postulated in earlier studies, the reaction occurs via the generation of a cryptic hydroxyl germinal to a good leaving group (phosphate) reminiscent of TauD and AtsK. A clear distinction between the hydroxylation *versus* desaturation mechanisms can be gleaned by tracking the fate of labeled $^{2}H_{1}$ atoms from utilization of deuterated UMP substrate, as well from the key bridging C-6' in the synthesized phosphonate analog (**S11**). To the best of our knowledge, this is the first direct identification of the hydroxylated intermediate (**5**) as the transient state in this type of reaction, while also demonstrating the specific abstraction of the *pro-S* hydrogen atom from C-5' of UMP. The total enzymatic synthesis of deuterated UMP (**3**) via a combinatory route derived from glycolytic, pentose phosphate and nucleotide metabolic pathways is another distinctive platform that we developed as a substitute to chemical synthesis of a deuterated analog. The most obvious advantage of this methodology is the convenience of a one-pot setup and the stereospecific utilization of substrates eliminating the need for stepwise purification, while the yield of the final product was a clear disadvantage. The retention of five of the six deuterium labels starting from the universally deuterated glucose substrate is possibly a result of recycled cofactors from within the one-pot reaction.

Chapter three: Exploring the structural basis of Cpr19 mechanism

3.1. Background: Structural basis for mechanism of Fe(II)/α-KG dependent enzymes

The stereoselective oxidation of an unactivated alkyl C-H bond is possibly one of the most difficult common functional group transformations in chemistry. Nature has tuned metal-dependent oxygenases or oxidases to carry out such reactions. The most well characterized enzymes that do so are the cytochrome P450 monooxygenases, and they have been studied in explicit details both mechanistically and structurally [98, 99]. Non-heme oxidizing enzymes that require an α-keto acid as a cosubstrate constitute the largest and most diverse family of *mononuclear* enzymes catalyzing many pivotal metabolic transformations, some of which were highlighted in previous chapters. Despite the vast array of distinct transformations these enzymes carry out, crystal structures of a number of different members from this group show the double-stranded β-helix (or jellyroll) as a common architecture for this superfamily [76, 77] (**Figure 3.1.1**) suggesting a close evolutionary relationship between the three branches of the family identified by sequence comparisons (diiron-using enzymes like methane monooxygenases, and monoiron-using enzymes also including Fe(III) dependent enzymes like lipooxygenases) [114, 115].

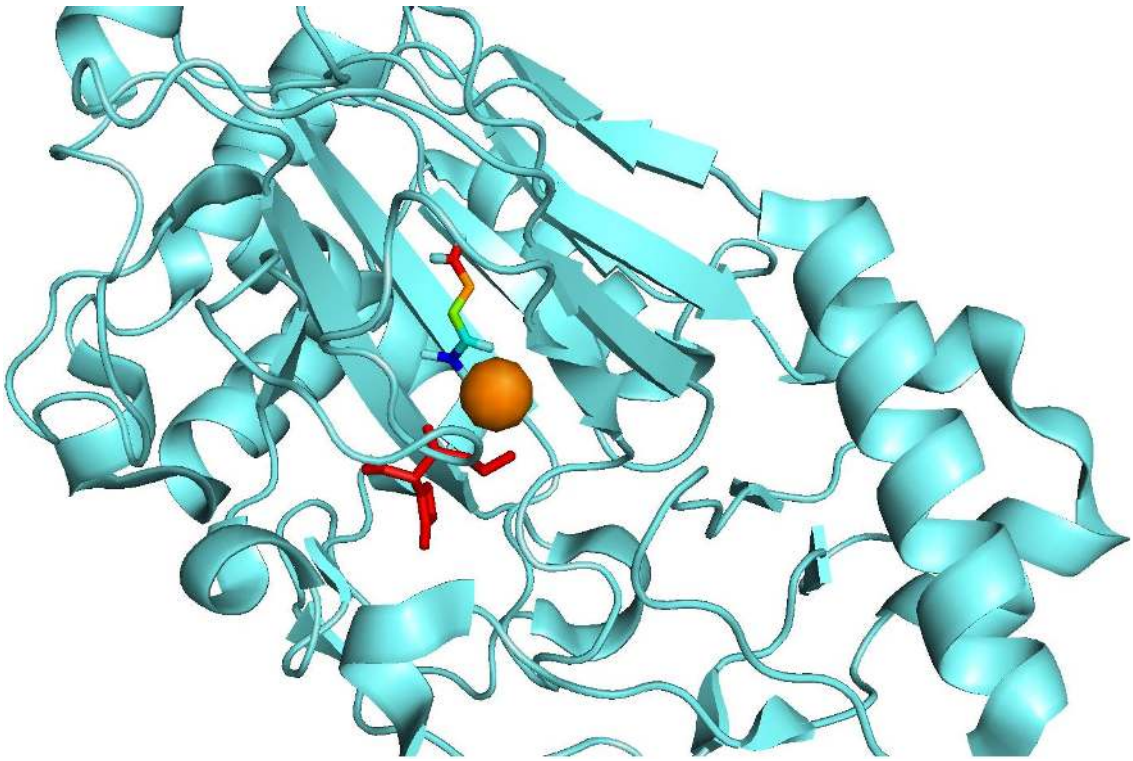


Figure 3.1.1. 'Jelly-roll' fold in Clavaminic Acid Synthase (CAS). Crystal structure of CAS complexed with mononuclear Fe(II) center (shown here in **orange**), complexed with substrate PCV (5-amino-3-hydroxy-2-(2-oxo-azetidin-1-yl)-pentanoic acid) (shown in **red**) and cosubstrate α-KG (in **rainbow** colors).

The highly conserved His^1-X-Asp/Glu-X_n-His^2 motif (facial triad) constitutes the facial metal-binding ligands that binds the high-spin Fe(II) center. One of the oxygen atoms from the dioxygen molecule has been shown to be incorporated into succinate, generated by oxidative decarboxylation of α-KG [105]. Sequence analyses of members in this group reveals little overall similarity, leading to the proposal that convergent evolution to a common mechanism and active site chemistry occurred within the wider family of Fe(II)/ α-KG dependent and related oxygenases. The presence of such a metal binding motif is not limited to iron-dependent enzymes – it is also known to occur in Mn(II)-dependent

extradiol oxygenases and Zn (II)-dependent hydrolase enzymes like thermolysin [116]. In Zn(II)-binding groups, the presence of an HEXXH…E motif provides the metal binding glutamate and histidine ligands, very similar to the Fe(II)/ α-KG facial triad. The enzyme peptide formylase from *E. coli*, originally characterized as a Zn(II)-dependent hydroxylase has in fact been shown to utilize Fe(II) coordinated by two histidines and a cysteine residue [117]. Cysteine residues are uncommon as ligands in the active site, predictably since it would lead to reaction with molecular oxygen generating reactive oxidizing species. However, the oxidation of an iron-linked thiol is elegantly exploited in the enzyme isopenicillin N synthase (IPNS) [79].

Several enzymes in this class have substrates with a built-in α-keto acid function and therefore do not require α-KG as a cosubstrate. Naturally, these enzymes appear to have slightly dissimilar topologies from enzymes utilizing α-KG. Examples include HPP dioxygenases, 4-hydromandelate synthase, 1-amino-1-cyclopropanecarboxylic acid synthase (ACCO), isopenicillin N synthase (IPNS), etc. [79]. IPNS catalyzes the four-electron oxidation of the tripeptide δ- (L-α-aminoadipoyl)- L- cysteinyl-D-valine (ACV) to produce the penicillin nucleus and two water molecules [118]. The structure of IPNS was the first reported structure for a member of this group [119] and provided an impressive picture for the mechanism of this enzyme (**Figure 3.1.2).**

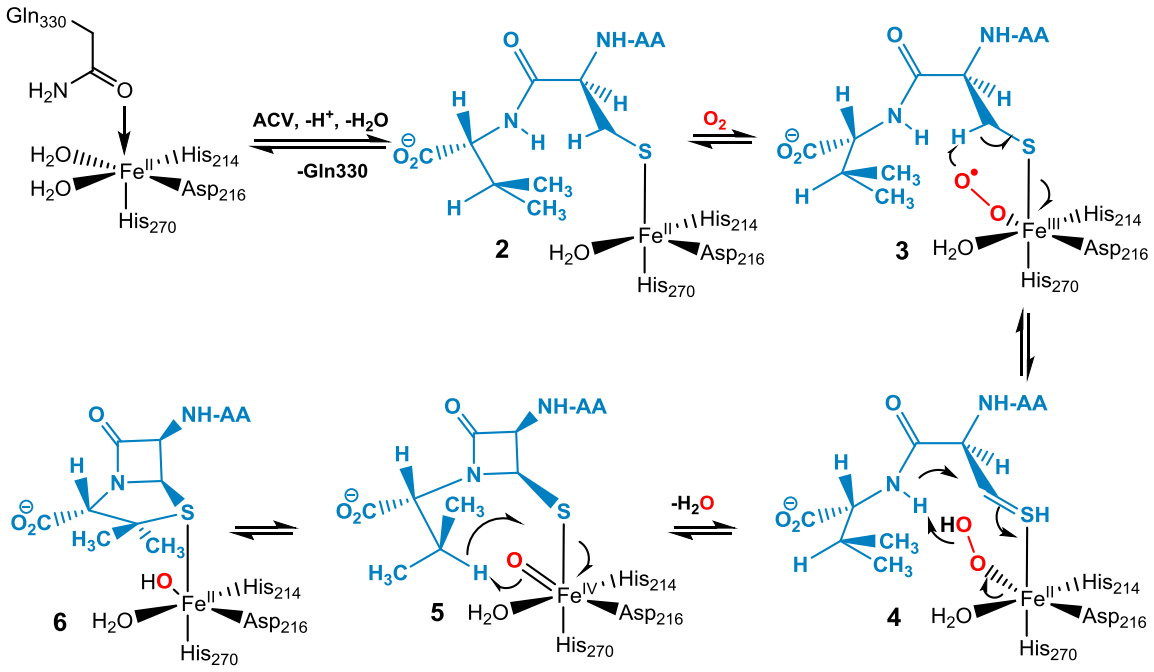


Figure 3.1.2. Reaction mechanism of IPNS. IPNS catalysis involves forging a unique thiol bond between the tripeptide substrate and the Fe(II) center of the enzyme. Dioxygen binding in intermediate **3** leads to the abstraction of a hydrogen bond from the cysteinyl residue, which paves the way to sequential deprotonation and β-lactam formation (**6**) by IPNS.

The crystal structure of IPNS complexed with ferrous iron at resting state revealed a metal ion octahedrally coordinated by His 214, Asp 216, His 270, Gln 330 and two water molecules. Addition of the tripeptide ACV did not distort the 'jelly-roll' core, however the side chain of Gln 330 which is coordinated to Fe(II) in the resting state is replaced by the ACV thiolate. Additionally, one of the water molecules ligating the metal ion is displaced, changing the metal coordination geometry from octahedral to square pyramidal (**2** in **Figure 3.1.2**). The valine isopropyl group of ACV is held in van der Waals contact with the

iron by interactions with Leu231, Val 272, Pro 283 and Leu 223. Interestingly, the valine βC-H bond, which is cleaved during the formation of the thiazolidine ring, is directed away from the iron center in the initial substrate complex, implying that rotation of the valine Cα-Cβ bond must occur during catalysis. The carboxylate of the valine residue of ACV is prevented from coordinating directly to the Fe(II) center via a hydrogen-bonding network in the resting enzyme complex. The aminoadipoyl residue of ACV was shown to be in an extended conformation with its carboxylate group participating in forming a salt bridge with Arg 87. Binding of substrate to enzyme complex initiates the reaction cycle and allows for the coordination of molecular oxygen to the Fe(II) center leading to intermediate **3**, which presumably leads to the abstraction of the pro-3-S hydrogen of the ACV cysteinyl residue. Subsequent deprotonation of the amide N-H by the hydroxo-superoxide species presumably allows for simultaneous β-lactam ring closure with concomitant generation of the elusive Fe(IV)=O intermediate (**5**). It is reasonable to assume that the isopropyl group must undergo a rotation parallel to β-lactam formation to relive its steric interactions with the sulfur ligand, which also facilitates the interaction of the valine β-hydrogen with the ferryl iron center (**Figure 3.1.3**).

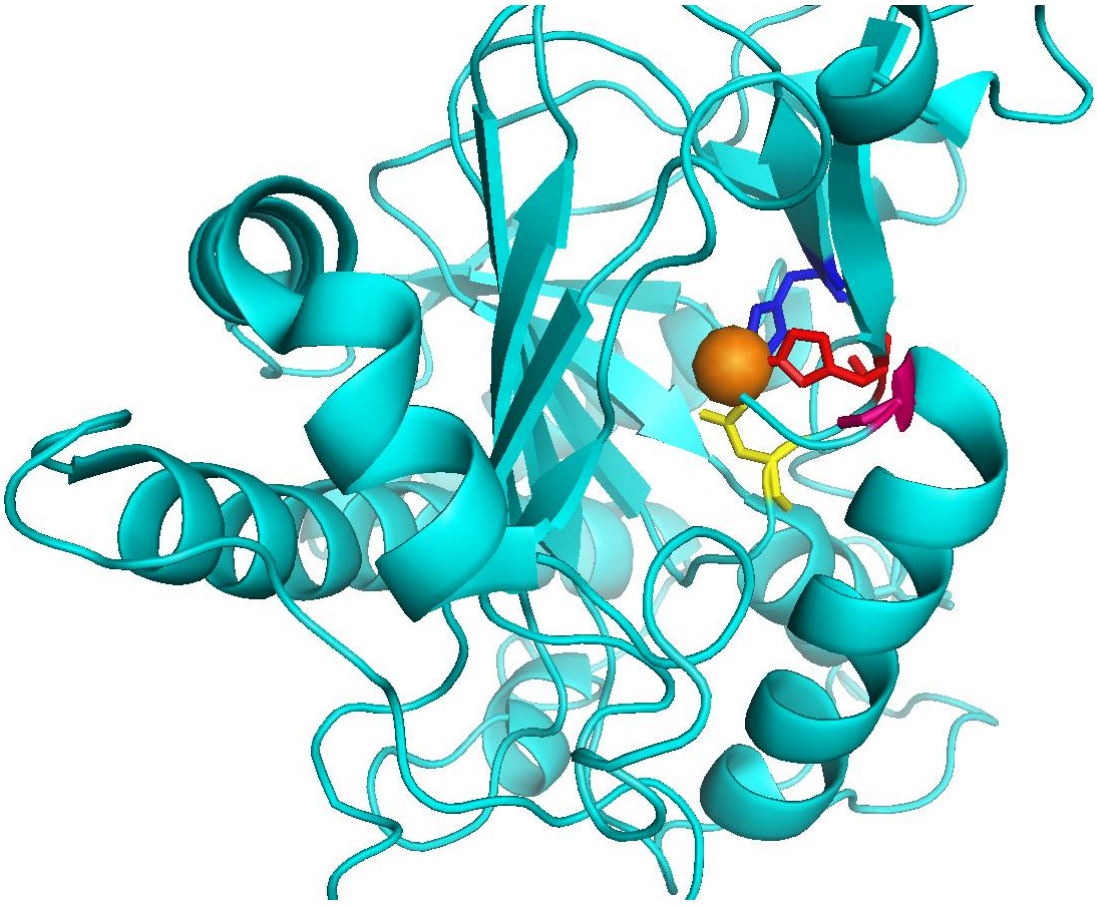


Figure 3.1.3. Crystal structure of IPNS depicting the 'jelly roll' topology. The Fe(II) center (shown here in **orange**) is facially coordinated by His 214 (**red**), Asp 216 (**yellow**) and His 270 (**blue**). The residue shown in **pink** is Gln 330 which forms the initial coordination bond with the ferrous iron center in the absence of ACV.

Since the reaction mechanism of both LipL and Cpr19 were shown to be similar to the reaction catalyzed by TauD, the X-ray crystal structure of TauD complexed to Fe(II) and both substrates (taurine and α-KG) provides a convincing model to predict the amino acid residues responsible for the hydroxylation mechanism catalyzed by the enzymes under investigation [120] (**Figure 3.1.3**). In fact, resolution of the TauD crystal structure led to a predictive model of the enzyme TfdA, responsible for the first step of biodegradation of

the herbicide 2,4-dichlorophenoxyacetate (2,4-D) [120] which in turn helped in identification of the predicted residues in TauD and TfdA that undergo a self-hydroxylation mechanism observed in these enzymes, in the absence of prime substrate.

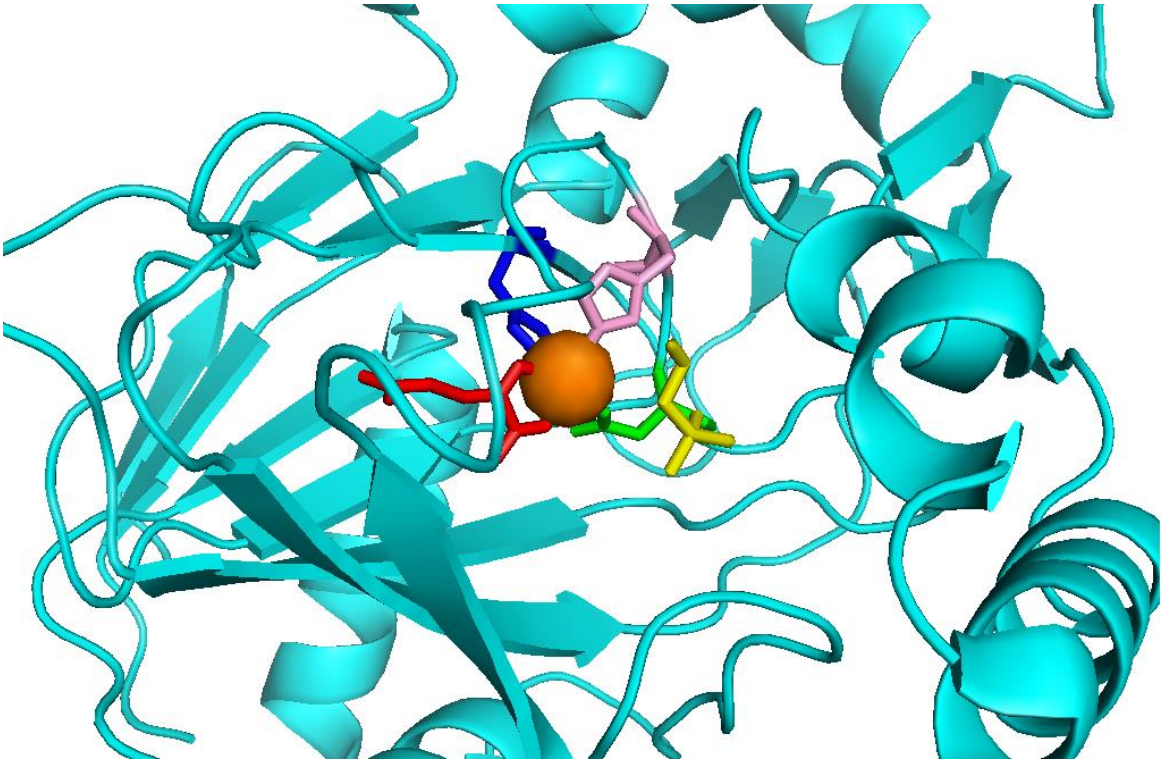


Figure 3.1.3. **Crystal structure of a monomer of E. coli TauD.** The Fe(II) center (shown here in **orange**) is facially coordinated by His 99 (**pink**), Asp 101 (**green**) and His 255 (**blue**). Primary substrate taurine (**yellow**) and cosubstrate α-KG (**red**) are shown bound to Fe(II) center as well.

In TauD, the pentacoordinate Fe(II) is bound to the enzyme via the conserved facial triad, composed of His 99, Asp 101 and His 255. Cosubstrate α-KG is bound via its C-1 carboxylate and C-2 keto group to the non-heme Fe(II) center, and via its C-5 carboxylate which forms a salt bridge with Arg 266 and a hydrogen-bond with Thr 126. The amine end of taurine can be stabilized by three hydrogen bonds with the phenolic side chain of Tyr

73, the hydroxyl of Ser 158 and the amide oxygen of Asn 95. The taurine sulfonate is similarly involved in hydrogen bond formation with Arg 270, His 70 and the backbone N-H of Val102. His 70 is responsible for interacting with taurine, as well as form a hydrogen bond with the backbone N-H of Val72. This lead to the realization that His 70 is possibly uncharged and is most likely to form a hydrogen bond with the sulfonate group of taurine (**Figure 3.1.4)**.

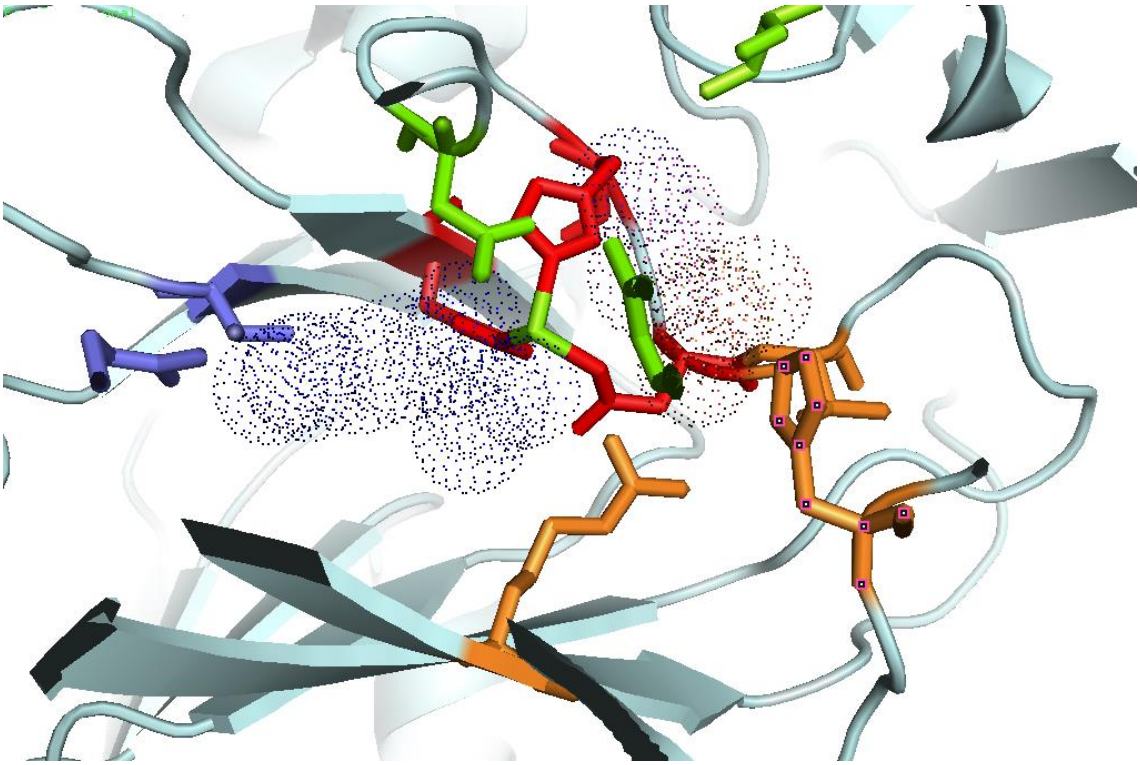


Figure 3.1.4. Catalytic pocket of TauD. Stereodiagram of TauD showing the protein ligands around the Fe(II) center. Shown here are the residues that form the facial triad (His 99, Asp 101, His 255) (**red**), alongside the other amino acid residues (Thr 126, Arg 266 in **blue**) involved in stabilizing α-KG (depicted in dots). The residues interacting with the amine end of primary substrate taurine are Tyr 73, Asn 95 and Ser 158 (**green**) while those involved in interacting with the sulfonate moiety are Arg 270, His 70 and the N-H backbone of Val 102 (**orange**).

It is interesting to note that while Arg270 is relatively conserved for several members of this enzyme family, the other two hydrogen bonds forged with Val72 and His70 help to select for a tetrahedral sulfate anion. This leads to a key observation that His 70 is conserved for the dioxygenases which specifically utilize sulfate or sulfonate substrates (like AtsK) but is absent in other enzymes from this superfamily that bind substrates containing a carboxylate moiety (like CarC, TfdA or CAS), and is therefore responsible for imparting substrate selectivity to TauD.

Parallel to ongoing efforts to crystallize both LipL and Cpr19 in our lab, in this chapter we explore some site specific mutants of Cpr19 developed either based on TauD as a model or to provide mechanistic details. We explore the effects of these mutants on the activity of Cpr19 to gain insights on the *putative* active site residues. Along the same lines, we investigate substrate promiscuity of Cpr19 (as well as LipL) to investigate flexibility of substrate recognition by testing its activity with four substrate analogs (xanthosine-5'-monophosphate, S2, S3 and S4) that retain the ribose-5'-monophosphate moiety but differ structurally from the 'base' component of the UMP nucleotide (**Figure 3.1.5**). Bioinformatic analyses of homologous dixoygenases, as well as the development of *in silico* models with docked substrates helps in providing a *predictive* model for these enzymes, assimilating the information obtained from site specific mutants.

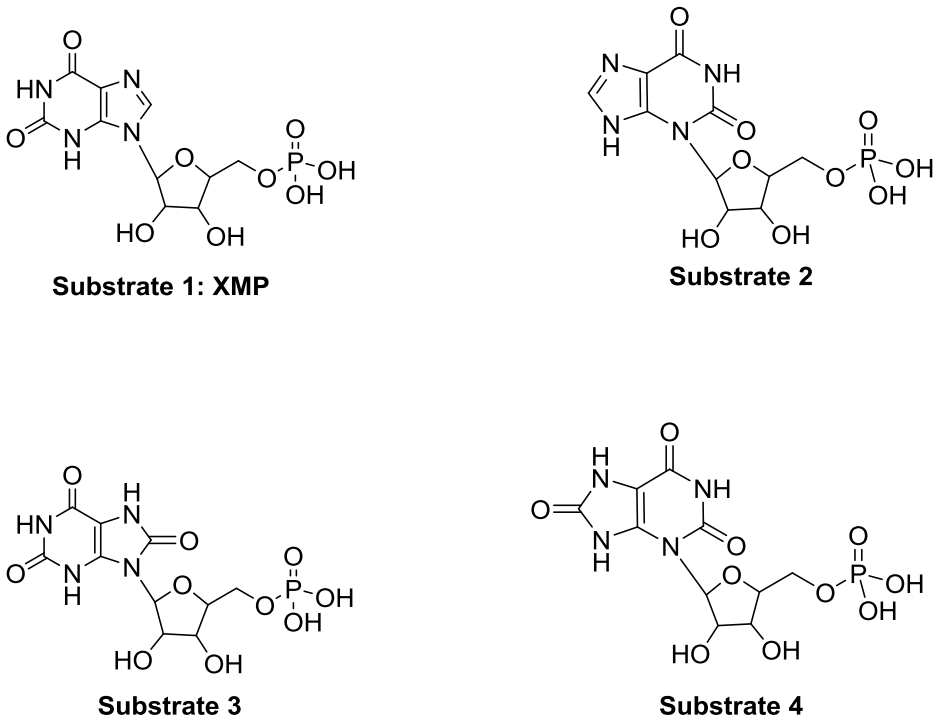


Figure 3.1.4. Structural substrate analogs of UMP. These analogs were generated by enzymatic synthesis for testing substrate promiscuity of Cpr19 and LipL with alternative substrates. They were generated previously in our lab from uptake of xanthine and uric acid in reaction, and were characterized preliminarily as isomeric pairs differing in their connections by mass spectrometric data.

3.2. Materials and Methods

3.2.1. Chemicals and Instrumentation

Uridine-5'-monophosphate (UMP), 2- ketoglutaric acid (α-KG), xanthine, uric acid, PRPP, Nucleoside bases, nucleosides, and nucleotides were purchased from Sigma or Promega. Buffers, salts, and media components were purchased from Fisher Scientific. Synthetic oligonucleotides were purchased from Integrated DNA Technologies. DNA sequencing was performed using the BigDye™Terminator version 3.1 Cycle Sequencing kit from Applied Biosystems, Inc. and analyzed at the University of Kentucky Advanced Genetic

Technologies Center. UV/Vis spectroscopy was performed with a Bio-Tek µQuant microplate reader using Microtest™ 96-well plates or a Shimadzu UV/Vis-1800 Spectrophotometer. HPLC was performed with a Waters Alliance 2695 separation module (Milford, MA) equipped with a Waters 2998 diode array detector and an analytical Apollo C-18 column (250 mm x 4.6 mm, 5 µm) or a semi-preparative Apollo C-18 column (250 mm x 10 mm, 5 µm) purchased from Grace. Electrospray ionization-MS was performed using an Agilent 6120 Quadrupole MSD mass spectrometer equipped with an Agilent 1200 Series Quaternary LC system and an Eclipse XDB-C18 column (150mm x 4.6 mm, 5 µm, 80Å). NMR data were collected using a Varian Unity Inova 500 MHz Spectrometer. High Resolution Mass Spectrometric data were obtained from University of Kentucky Mass Spectrometry Core Facility.

3.2.2. Site directed mutagenesis

Point mutations of Cpr19 were generated by PCR amplification using pET30-cpr19 as a template and the Expand Long Template PCR system (Roche Applied Science (Indianapolis, IN)). Reactions were conducted as described in the provided protocol with supplied buffer 2, 5% dimethyl sulfoxide, each of the following primers (**Table 6**) and the reverse compliment (the engineered Ala component is underlined below for the mutant variants). The thermocycler program included an initial hold at 94°C for 10s, 56°C for 15s, and 68°C for 50s. The template DNA was digested with *DpnI* for 1h at 68°C and transformed into E. coli DH5α competent cells. Introduction of the desired point mutations and the sequence of the entire gene including 200bp upstream and

downstream were confirmed by DNA sequencing to yield the desired set of Cpr19 mutants as listed in the table below (**Table 7**). The *nikR* gene was amplified by PCR cloning from genomic DNA isolated from Streptomyces tendae, following the protocol described in Section 2.4.5.

Table 6. List of primers used for site-directed mutagenesis

Primers	**Oligonucleotide Sequence**
*cpr19(S79A)*_for	5' – CGGAGAGACCAAGGACTACGCCGCGCCCTACAGCGAC – 3'
*cpr19(T119A)*_for	5' – CATCGGCGAAATCAAGACCGCGATCCTGTACTGCGTCCG – 3'
*cpr19(W104A)*_for	5' – CACGACGGCCGGTCAGATTGCGCACGTGGACGGTCTCCT – 3'
*cpr19(H105A)*_for	5' – GACGGCCGGTCAGATTTGGGCCGTGGACGGTCTCCTCGATG -3'
*cpr19(D107A)*_for	5' – GGTCAGATTTGGCACGTGGCCGGTCTCCTCGATGACATCG – 3'
*cpr19(C123A)*_for	5' – CAAGACCACGATCCTGTACGCCGTCCGGGCCGCTCACC – 3'
*cpr19(C198A)*_for	5' – ACACCGACAACGAGACGGCCACGTGGGACTACTCGGCCGAT– 3'
*cpr19(R260A)*_for	5' – CAAGCCGGACGCCCGCGCCCACCTGGTCAGGGCGCTC – 3'
*nikR*_for	5'- AGAGGAGAGTTAGAGCCTCAGTCGTCCGTGCCGAAG - 3'
*nikR*_rev	5'- GGTATTGAGGGTCGCATGACGCAGGATGAATTGAAAAAAG – 3'

3.2.3. Cloning and heterologous expression of genes

Plasmids were introduced into *E. coli* BL21 (DE3) cells by transformation and the recombinant strains were grown in LB medium supplemented with 30 µg/ml kanamycin.

Recombinant culture were then grown at 18°C with 250 rpm, following inoculation with 500 ml of LB with 30 µg/ml kanamycin in a 2.5 L Erlenmeyer flask, until the cell density reached an OD_{600} = 0.5. Expression was induced with 0.1 mM isopropyl 1-thio-*β*-D-galactopyranoside (IPTG) and after overnight incubation at 18°C, cells were harvested and lysed using a French press with one pass at 15,000 psi and immediately centrifuged at 18,000 rpm. Obtained supernatant was utilized for purification of the desired proteins using affinity chromatography with a nickel- nitriloacetic acid-agarose from followed by desalting of the recombinant proteins into 50mM Tris-HCl (pH 8), 100mM NaCl, and 5% glycerol using a PD-10 desalting column. Purified proteins were concentrated with an Amicon Ultra 10000 MWCO centrifugal filter prior to addition of glycerol (final 40%) for storage at -20^{0}°C. Protein solubility and purity were assessed by 12% acrylamide SDS-PAGE; His_6-tagged proteins were utilized without further modifications.

Table 7. List of plasmids used for site-directed mutagenesis

Strain/Plasmid	**Characteristics and Relevance**	**References**
E. coli Nova-blue	Host for routine cloning	Novagen
E. coli BL21 (DE3)	Host for protein expression	Novagen
pET30	Expression vector	Novagen
pET30- *cpr19(S79A)*	*cpr19(S79A)* gene cloned to pET30	This study
pET30- *cpr19(T119A)*	*cpr19(T119A)* gene cloned to pET30	This study
pET30- *cpr19(W104A)*	*cpr19(W104A)* gene cloned to pET30	This study
pET30- *cpr19(H105A)*	*cpr19(H105A)* gene cloned to pET30	This study
pET30- *cpr19(D107A)*	*cpr19(D107A)* gene cloned to pET30	This study
pET30- *cpr19(C123A)*	*cpr19(C123A)* gene cloned to pET30	This study
pET30- *cpr19(C198A)*	*cpr19(C198A)* gene cloned to pET30	This study
pET30- *cpr19(R260A)*	*cpr19(R260A)* gene cloned to pET30	This study
pET30- *nikR*	*nikR* gene cloned to pET30	This study

3.2.4. *In-vitro* characterization of Cpr19 mutant variants

Reactions with each Cpr19 mutant typically consisted of 50mM Tris-HCl (pH 7.5), 1mM UMP (or substrate analog **S11**), 1.25 mM α-KG, 1 mM ascorbate, 500 µM $FeCl_2$, and 100 nM enzyme at 30˚C. Reactions were terminated by ultracentrifugation using a Microcon YM-3. Following removal of protein by centrifugation, the reactions were analyzed by HPLC using a C-18 reverse-phased column under ion-pairing conditions (monitored at 254 nm). A linear gradient of from 40 mM acetic acetic acid-triethylamine pH 6.5 (A) to 20% methanol (B) (0-4 min, 0% B; 4-24 min, 50% B; 24-26 min 100% B; 26-32 min, 100% B; 32-35 min, 0% B) with flow rate of 1 ml/min was used to analyze the reactions and elution was monitored at 260 nm.

3.2.5. Kinetic characterization of Cpr19 mutants

The activities of the generated Cpr19 mutants were detected by monitoring the formation of inorganic phosphate with the malachite green binding assay as described in the previous protocol [108]. As before, we used 20mM (20 µL) EDTA to terminate the reactions of the Cpr19 mutants with substrate UMP. Absorbance was measured at 620 nm using a microplate reader. A standard curve was developed using phosphate standards of 0.8, 1.6, 3.2, 6.7, 12.5, 25, and 50 µM (provided in the kit). For single-substrate kinetic analyses involving Cpr19 mutants, reactions consisted of 50mM Tris-HCl (pH 7.5), 1 mM ascorbate, 500 µM $FeCl_2$, 100 nM enzyme, near saturating α-KG (1mM) and variable UMP (50 µM– 1 mM). The reactions were initiated at 30˚C by adding each

Cpr19 mutant (100 nM) using the pipettor for mixing. Reactions were terminated after 3 min (<10% product formation) by addition of EDTA, using a sample without enzyme as a blank/control. Each data point represents triplicate end point assays. Kinetic constants were obtained by nonlinear regression analysis using GraphPad Prism.

3.2.6. Enzymatic synthesis of structural substrate analogues of UMP

Reactions with NikR and EcUpp typically consisted of 50mM HEPES buffer (pH 7.5), 2mM xanthine (or uric acid), 2.5mM PRPP, 20mM $MgCl_2$ and NikR/EcUpp (100 nM) at 30˚C. Following removal of protein by centrifugation, the reactions were analyzed by HPLC using a C-18 reverse-phased column (monitored at 254 nm). A series of linear gradients was developed from 0.1 % TFA in 5 % acetonitrile (A) to 0.1 % TFA in 90% acetonitrile (B) in the following manner (beginning time and ending time with linear increase to % B): 0-4 min, 100% B; 4-24 min, 50% B; 24-26 min, 100% B; 26-32 min, 100% B; and 32-35 min, 0% B. LC-MS was performed using a linear gradient from 0.1% formic acid in water to 0.1% formic acid in acetonitrile over 20 min. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm.

3.2.7. In vitro utilization of substrate analogs by LipL and Cpr19

Reactions typically consisted of 50mM HEPES buffer (pH 7.5), 1mM each of the substrate analogs (xanthosine 5'-monophosphate (XMP), S2, S3 or S4), 1.25 mM α-KG, 200 µM ascorbate, 100 µM $FeCl_2$, and 100 nM LipL or Cpr19 at 30˚C. Reactions were terminated

by ultracentrifugation using a Microcon YM-3. Following removal of protein by centrifugation, the reactions were analyzed by HPLC using a C-18 reverse-phased column under ion-pairing conditions (monitored at 254 nm). A linear gradient of from 40 mM acetic acetic acid-triethylamine pH 6.5 (A) to 20% methanol (B) (0-4 min, 0% B; 4-24 min, 50% B; 24-26 min 100% B; 26-32 min, 100% B; 32-35 min, 0% B) with flow rate of 1 ml/min was used to analyze the reactions and elution was monitored at 260 nm. Large scale isolation of the respective reaction products starting from one of four substrate analogs (XMP, S2, S3 or S4) were carried out with HPLC using a C-18 reverse-phase semipreparative column. The linear gradient utilized was exactly identical to the gradient used for analytical characterizations. However, the flow rate was kept constant at 3.5 mL/min, and elution was monitored at 254 nm. LC-MS was performed using a linear gradient from 0.1% formic acid in water to 0.1% formic acid in acetonitrile over 20 min. The flow rate was kept constant at 0.4 mL/min, and elution was monitored at 254 nm.

3.2.8. Development of *in-silico* models for Cpr19 and LipL structures

Cpr19 model: The SWISS-MODEL [121] predicted Cpr19 structure elements were aligned with the crystallographically determined structure of E. coli TauD (PDB ID: 1GQW) template. The Cpr19 sequence was modeled onto the TauD chain A ignoring Fe(II), α-KG and primary substrates using PyMOL [122]. Cpr19 residues for which there were no equivalent residue in TauD were masked. From the initial model, we encountered a gap in the TauD structure (between residues Tyr162 and His171) and as a result the predicted

structure for Cpr19 did not align well with this region. The predicted model suggested an additional α-helix in Cpr19 for this region, and in the absence of a convincingly superior model we decided to omit residues 160-182 from the template to gain a better model. Fe(II) and α-KG were inserted into the Cpr19 model based on the alignment of its 2 His – 1 carboxylate facial triad with that of TauD, with direct substitution of the α-KG and Fe(II) positions. The residues interacting with the carboxylate of α-KG (Thr 133 and Arg 260) were in perfect alignment with the equivalent residues in TauD. Primary substrate UMP (**1**) was manually positioned onto the modeled Cpr19 active site using AutoDock Vina [123]. As an initial constraint, we rationalized that the phosphate moiety of UMP must align with the analogous sulfate moiety of taurine in TauD, and superimposed UMP accordingly. Since we showed in the preceding chapter that Cpr19 selectively oxidizes the *pro*-S hydrogen of UMP, this hydrogen was positioned to face the active site Fe(II). We proceeded to model the other four substrate analogs (XMP, S2, S3, S4) in the active site, following the same protocol.

LipL model: For this model, the previously predicted Cpr19 homology model was used as a template to generate the structural elements for LipL via SWISS MODEL, as before. The LipL sequence was modeled onto the Cpr19 model with Fe(II) and co-substrate α-KG using PyMOL, resulting in nearly perfect alignment of the active site residues as expected. AutoDock Vina was used to manually dock UMP into the active site following on the same rationale that was used for the Cpr19 model.

3.3. Results

3.3.1. Bioinformatic analyses of LipL and Cpr19 with homologous dioxygenases

Since iron binding by the mononuclear non-heme, α-KG binding dioxygenases from this superfamily of enzymes is mediated by the facial binding triad, the conserved essential motif is HX_1D/EX_nH, wherein X_n can vary between 40 to 153 amino acids. For aligning the sequences of LipL and Cpr19 with the sequence of TauD we used the ClustalW Omega online program. An example of a sequence alignment of TauD (accession number AAB18091.1) with LipL (accession number BAJ05888) and Cpr19 is shown in **Figure 3.3.1.1**. We confirm the presence of the Fe(II) binding motif essential for enzyme catalysis (shown in the black rectangular boxes, with the three residues highlighted in **pink**). Residues that are *similar* are highlighted in **grey**, and the conserved residues are highlighted in **light blue**. As is evident from this alignment, LipL and Cpr19 share very little sequence identity with TauD (less than 20%) which is expected since most other members from this superfamily usually have very little sequence similarities given the great diversity in their substrates [120].

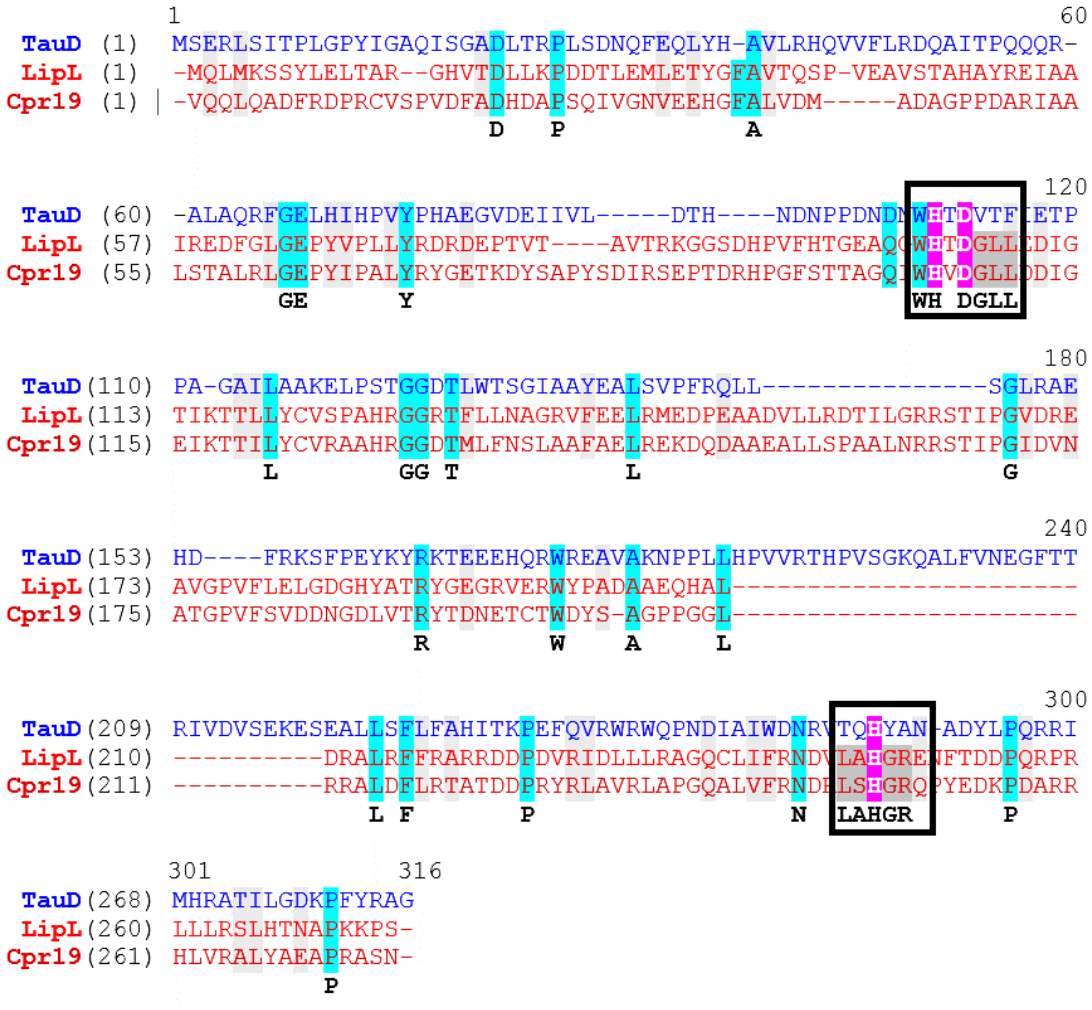


Figure 3.3.1.1. Sequence analyses of TauD (blue), LipL (orange) and Cpr19 (red) using ClustalW. Conserved regions are boxed and contain the residues critical for binding the non-heme Fe(II) center. While TauD has 153 residues between the 2nd and 3rd ligands, LipL and Cpr19 both have 140 residues between those two.

The enzyme alkylsulfutase AtsK from *Pseudomonas putida* [124] is a close homolog of TauD that catalyzes the hydroxylation of an alkyl sulfate ester which similarly decomposes to sulfate and an aliphatic aldehyde. We used this as a second template (accession number YP_008111264) to conduct sequence analyses of LipL and Cpr19 using the ClustalW program. We observed a similar set of aligned residues (highlighted as before)

with the Fe(II)- binding motif shown in the conserved boxed regions, confirming that the sulfate releasing enzymes share very low sequence identity with LipL and Cpr19 (**Figure 3.3.1.2)**.

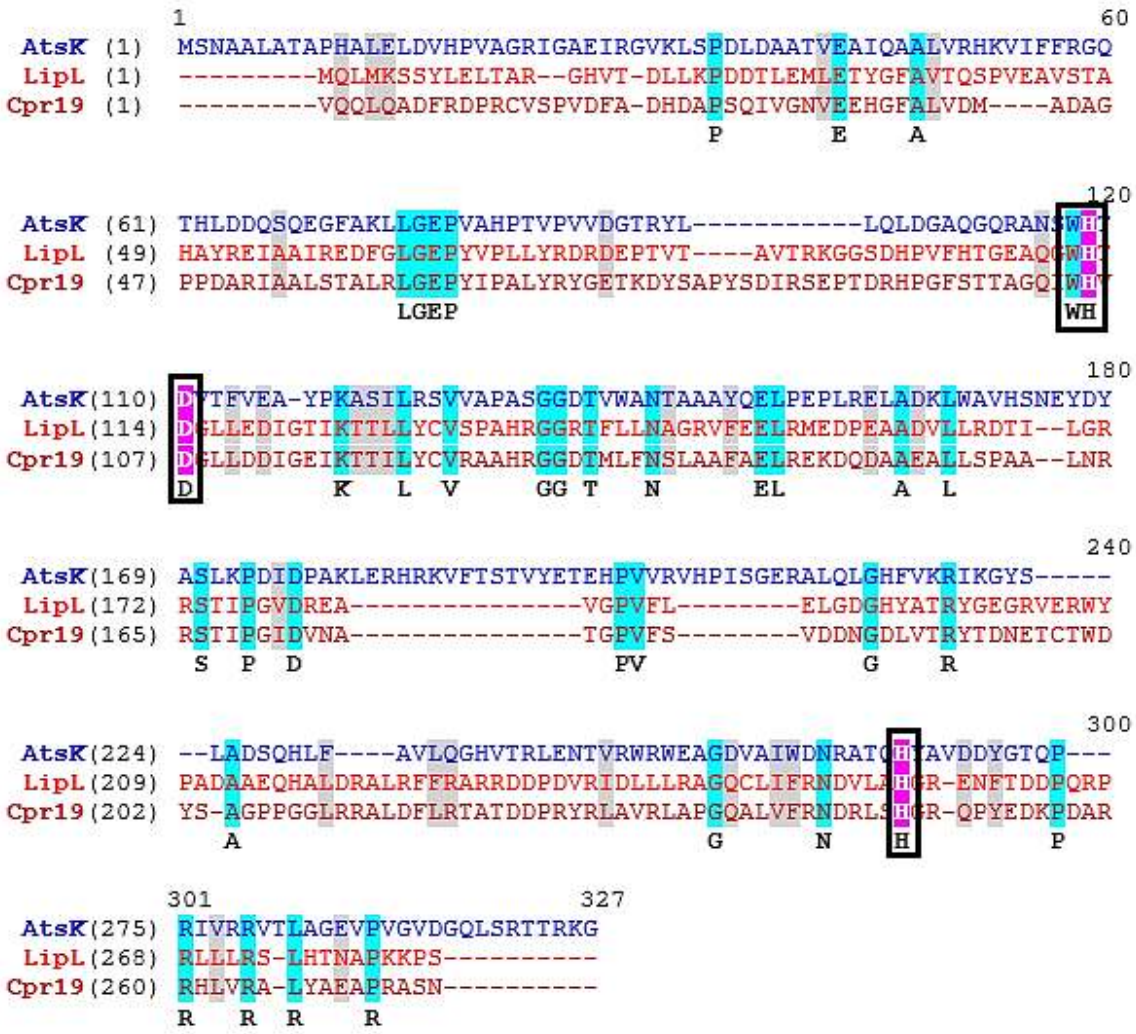


Figure 3.3.1.2. Sequence analyses of AtsK (blue), LipL (orange) and Cpr19 (red) using ClustalW. As before, the conserved regions are boxed and represent the residues critical for binding Fe(II). AtsK has a 155 residue spacer between the 2nd and 3rd ligands.

3.3.2. In vitro characterization of Cpr19 mutant variants:

The genes for the set of Cpr19 mutants (**Table 7**) were cloned and expressed in E. coli to yield soluble proteins. HPLC analysis of the reactions catalyzed by each of the mutant variants revealed either a peak consistent with the product uridine-5'-aldehyde (**2**) (and the tris-adducts of the aldehyde eluting later) if the site-specific mutant did not exhibit altered activity, *or* a peak corresponding to unutilized UMP (**1**) if the corresponding point mutation was responsible for abolishing activity in the mutant (**Figure 3.3.2.1**). The subset of mutants that exhibited abolished or partial activity are depicted separately in **Figure 3.3.2.2**.

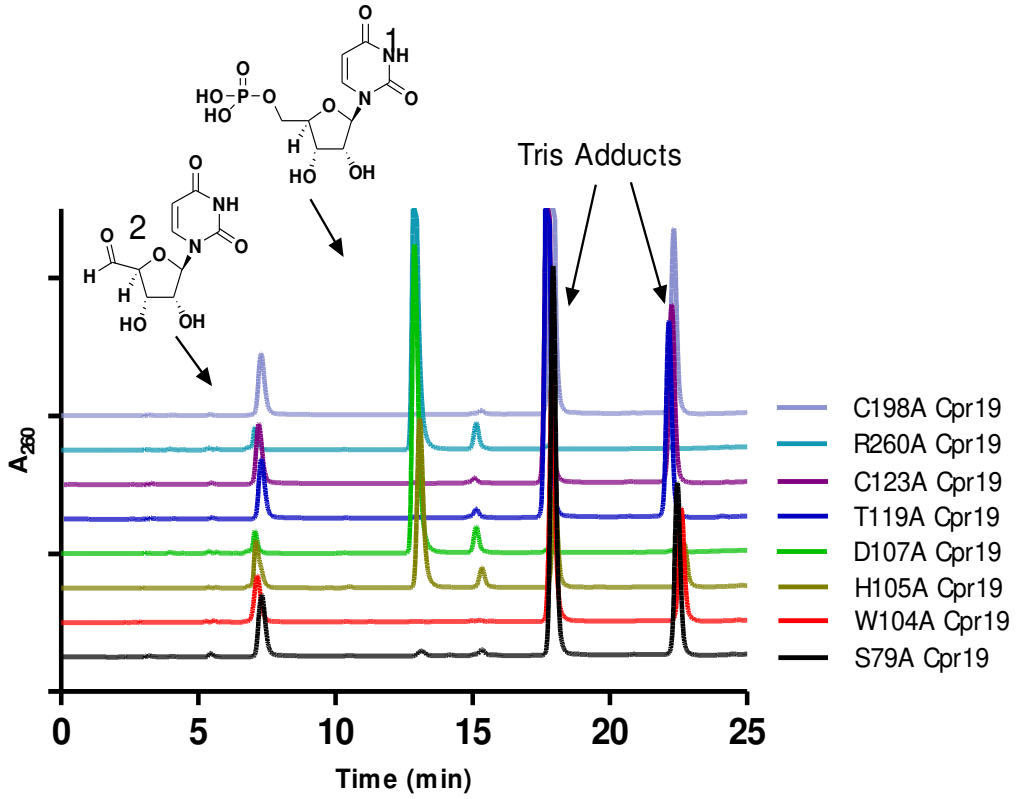


Figure 3.3.2.1. Color coded HPLC traces for reactions catalyzed by generated set of Cpr19 site specific mutants. Observed peaks for uridine-5'-aldehyde (**2**) and related tris-adducts indicate retention of activity for mutants S79A, W104A, T119A, C123A, C198A and partial conversion for H105A. Abolishment of activity in D107A and R260A mutants are inferred from the unutilized substrate UMP (**1**) peak eluting at t = 12 min. A_{260}: absorbance at 260 nm.

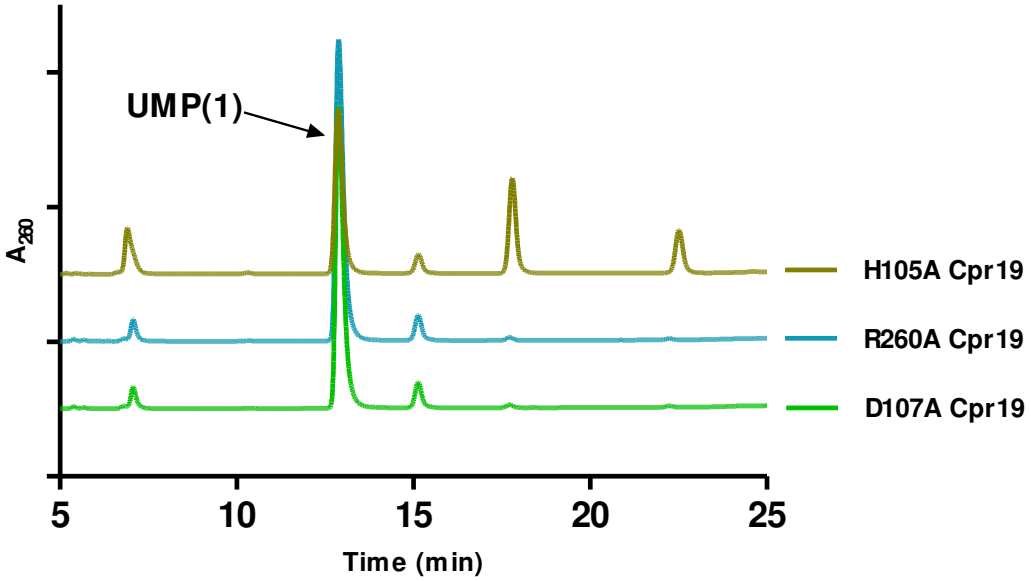


Figure 3.3.2.2. Subset of mutant enzymes that exhibited altered activity. H105A was shown to partially convert UMP (**1**) to uridine-5'-aldehyde (**2**), while the activities for both D107A and R260A were completely abolished. A_{260}: absorbance at 260 nm.

3.3.3. In vitro utilization of substrate analog S11 by Cpr19 mutant variants

As an additional characterization of the Cpr19 mutants, we wanted to test if the mutants could recognize synthesized analog **S11** as a substrate (with regards to substrate analog recognition in wild-type Cpr19). HPLC analysis of the reactions catalyzed by each of the mutant variants revealed a either a peak identical to the product 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'S)-hexofuranosyl]uracil (**5**) as was previously observed for wild-type Cpr19, *or* a peak consistent with unutilized **S11** if the corresponding point mutation was responsible for abolishing substrate utilization in the mutant (**Figure 3.3.3.1**). These

results were in agreement with the results with primary substrate UMP, with the exception of C123A which recognized UMP as substrate but failed to do so with **S11**.

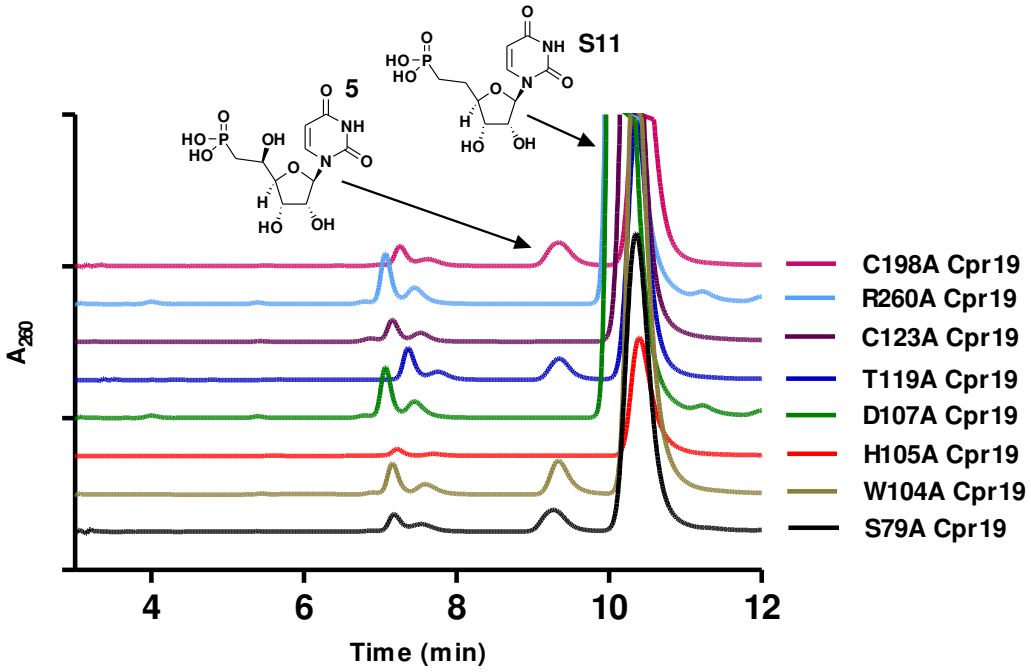


Figure 3.3.3.1. Color coded HPLC traces for reactions catalyzed by Cpr19 mutants when tested in reaction with substrate analog S11. Peaks corresponding to product 1-[6'-Deoxy-6'-phosphono-β-D-*ribo*-(5'S)-hexofuranosyl]uracil (**5**) indicate recognition of **S11** as a substrate by mutants S79A, W104A, H105A, T119A and C198A. Mutants C123A, D107A and R260A, however, did not utilize **S11** as a substrate. A_{260}: absorbance at 260 nm.

3.3.4. Kinetic characterization of Cpr19 mutants with respect to substrate UMP

For single-substrate kinetic analyses involving the generated set of Cpr19 mutants, we conducted reactions at 30°C at pH=7.5. Reactions were terminated after 3 min (<10% product formation) by addition of EDTA, using a sample without enzyme as a blank/control. Each data point represents triplicate end point assays. Kinetic constants

were obtained by nonlinear regression analysis using GraphPad Prism (**Figure 3.3.4.1**). The extracted kinetic constants are listed in **Table 8**.

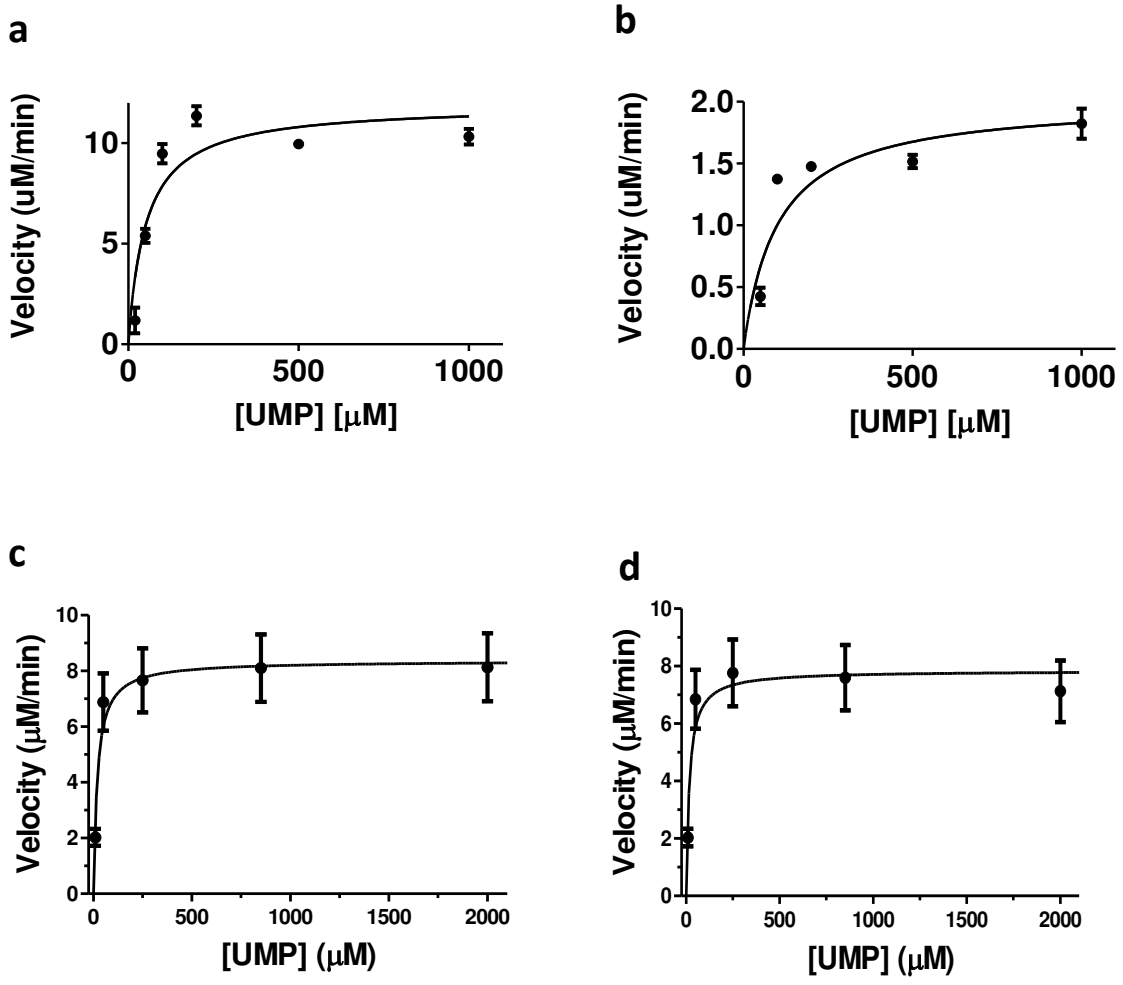


Figure 3.3.4.1. Single-substrate kinetic analysis of Cpr19 mutants with variable UMP at pH=7.5: (a) W104A, (b) H105A, (c) C123A, (d) C198A.

Table 8. Kinetic constants for Cpr19 mutants

Enzyme	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat} / K_m ($\mu M^{-1}min^{-1}$)
W104A	UMP	51 ± 14	119 ± 18	2.33
H105A	UMP	100 ± 25	2 ± 0.1	0.02
C123A	UMP	19.87 ± 5	35 ± 1.38	1.76
C198A	UMP	16.96 ± 6	65.35 ± 3.4	3.85

3.3.5. Enzymatic synthesis and spectroscopic characterization of structural substrate analogues of UMP

The genes for nikR and uracil phosphopribosyl transferase (EcUpp) were cloned and expressed in E. coli to yield soluble proteins with the expected sizes (**Figure 3.3.5.1**). Previous HPLC analysis of the reactions catalyzed by both these enzymes when tested in reaction with substrates xanthine and uric acid led to the identification of four different products in our lab (unpublished data). HPLC analyses revealed the uptake of xanthine and uric acid by both nikR and EcUpp, wherein it led to the observation of two distinct product peaks in each reaction leading to predicted isomeric product pairs. A representative HPLC trace is shown below (**Figure 3.3.5.2** and unpublished data provided by Dr. Van Lanen). Each of the peaks were collected from a semi-preparative C-18 HPLC column and LC-MS was utilized for preliminary confirmation the identity of the products (data not shown). In the present study, we provide NMR spectroscopic analyses of the products further to discern the 'connections' in obtained products that were all substrates analogs of UMP (differing in the 'base' component of the 5'-monophosphate) (**Figures 3.3.5.3- 6**). Peaks are assigned to each product in the following tables (**Tables 9-12**).

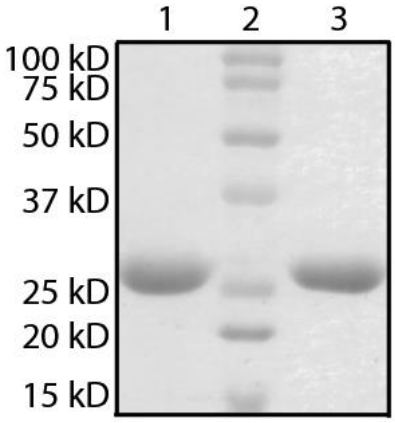


Figure 3.3.5.1. SDS-PAGE analysis of purified proteins (1) His_6-NikR (expected 24.99 kD) and **(2)** SDS-PAGE analysis of purified His_6-EcUPP (expected 22.5 kD)

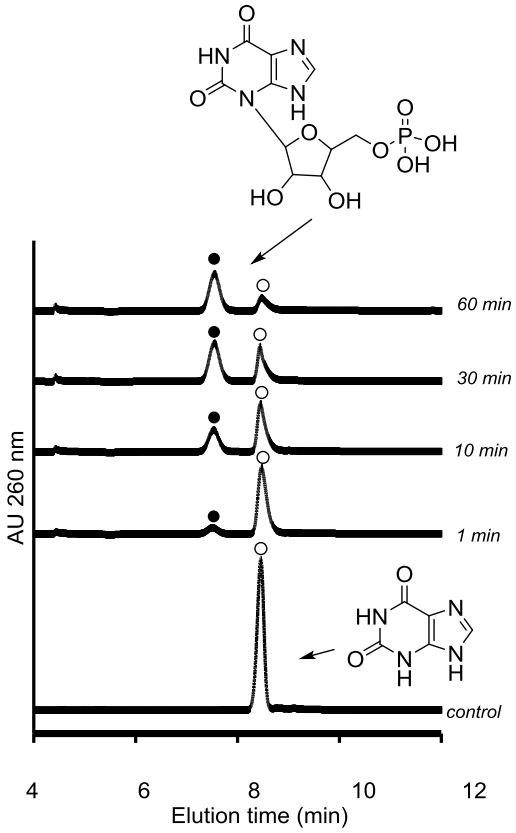


Figure 3.3.5.2. HPLC traces for the reaction catalyzed by NikR.

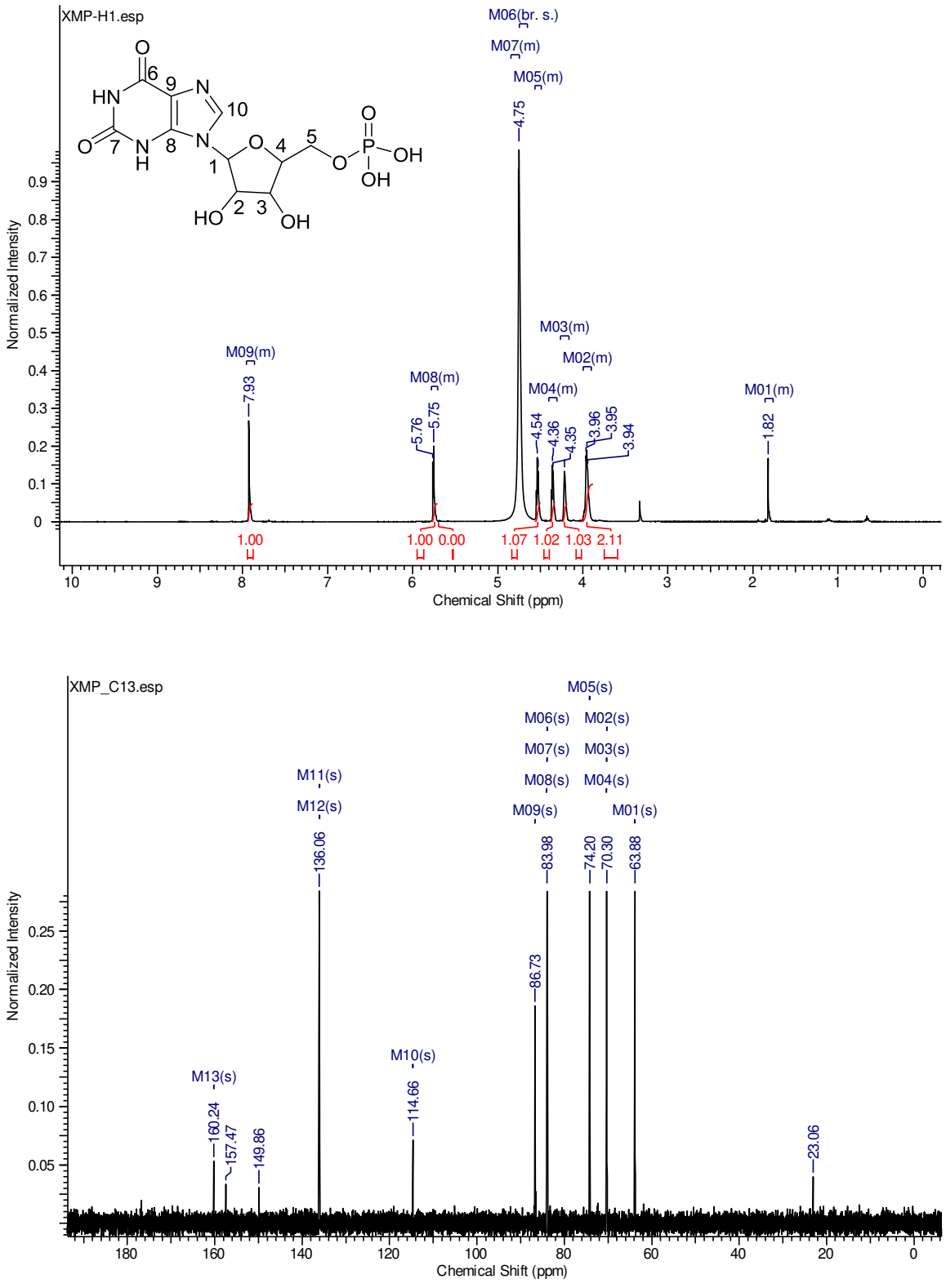
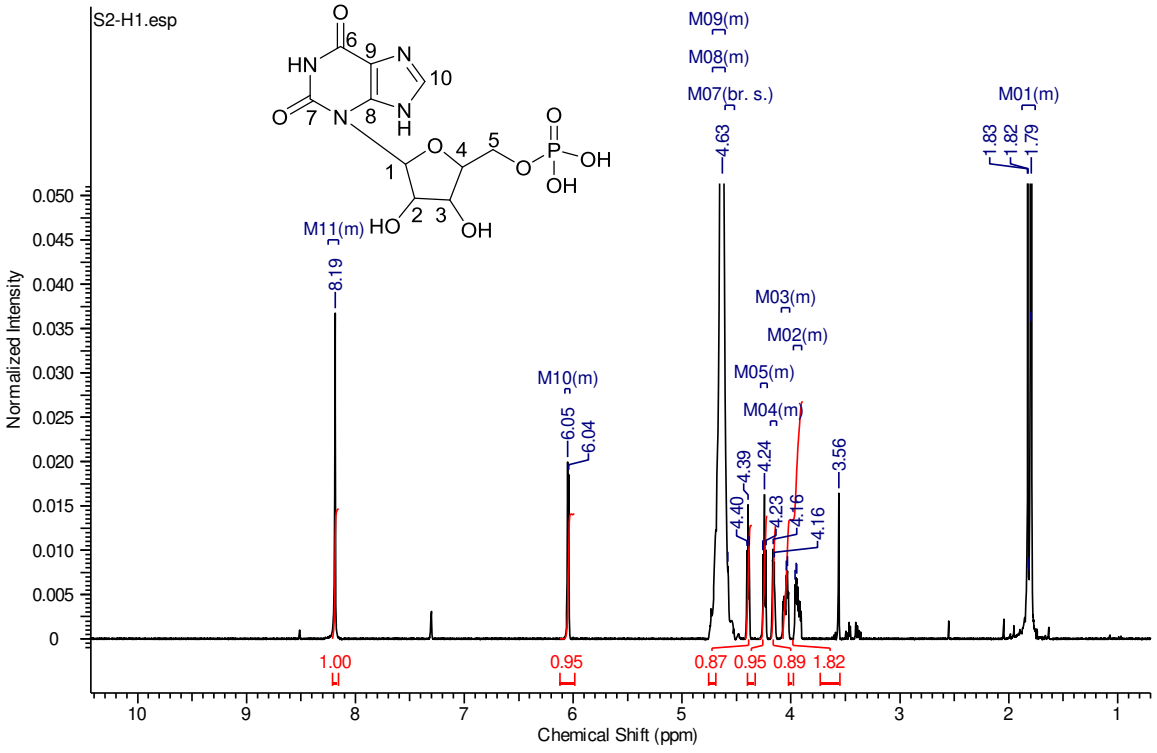


Figure 3.3.5.3. ^{1}H and ^{13}C (500MHz, D_2O) for xanthosine 5'-monophosphate (XMP).

Table 9. Peak assignment for xanthosine 5'-monophosphate (XMP)

Position	δ_C, mult.	δ_H (J in Hz)
1	5.76 (d, 1H)	86.73
2	4.54 (m, 1H)	74.20
3	4.36 (m, 1H)	70.30
4	4.21 (m, 1H)	83.79
5	3.96-3.94 (m, 2H)	63.88
6	--	160.24
7	--	157.47
8	--	149.86
9	--	114.66
10	7.93 (s, 1H)	136.06



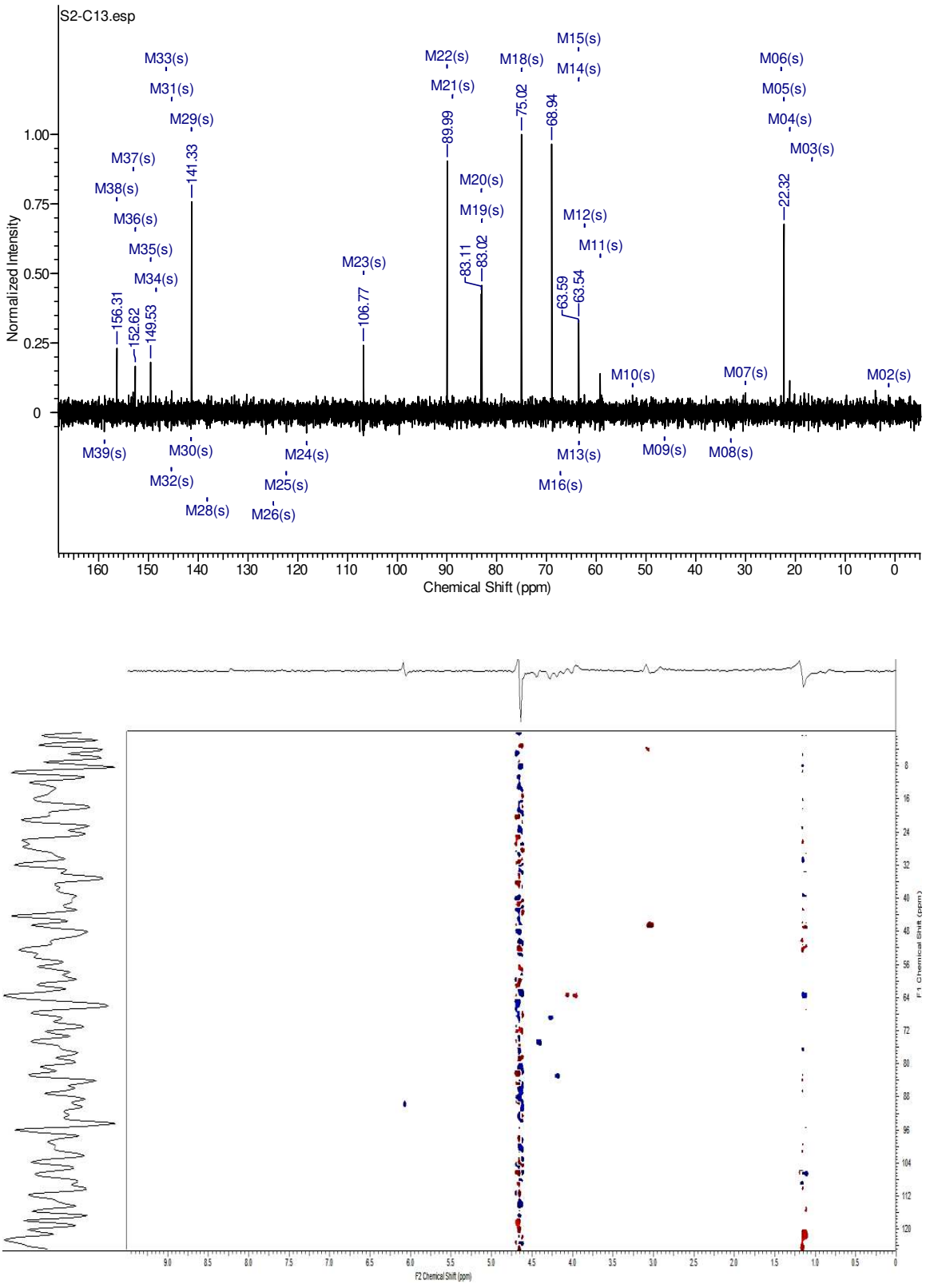
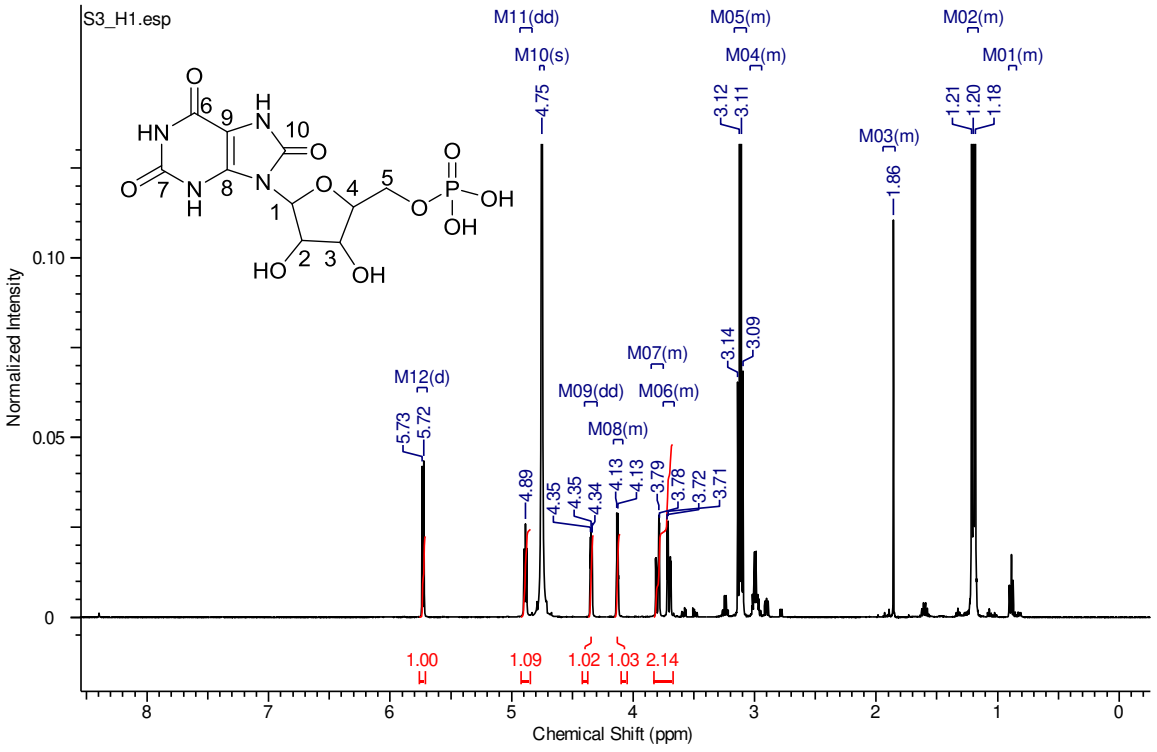
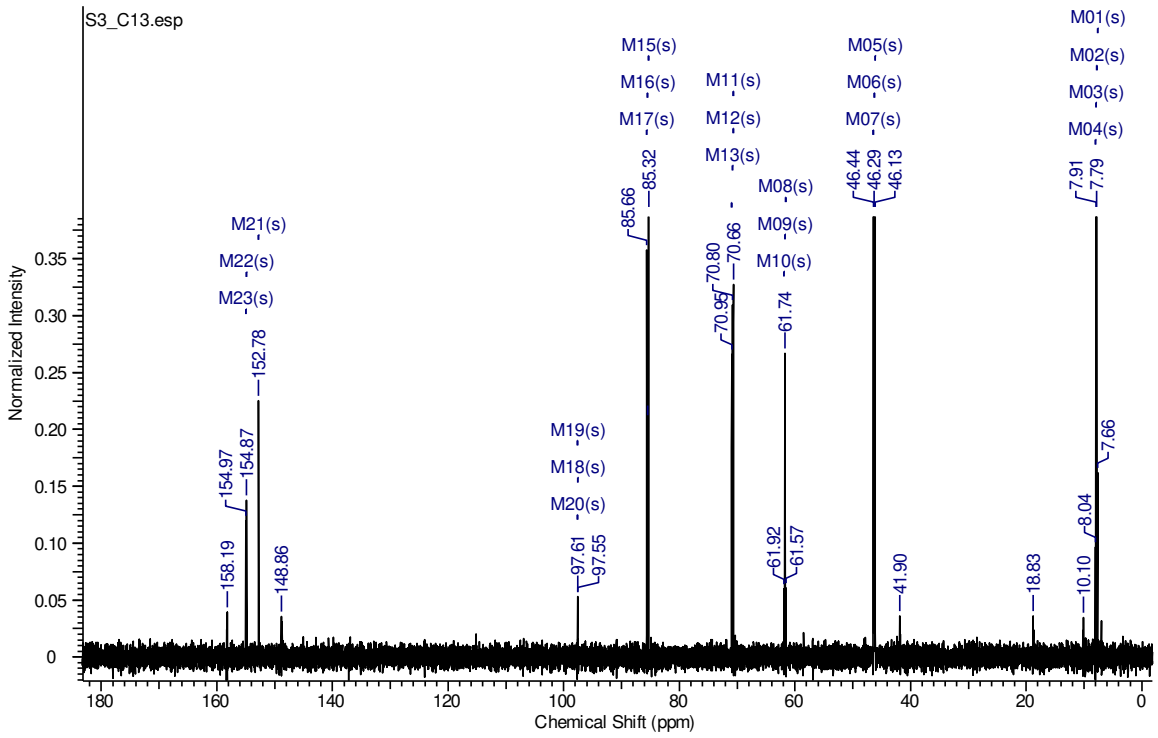


Figure 3.3.5.4. ^{1}H, ^{13}C, 2D ^{1}H – ^{13}C gHSQC (500MHz, D_2O) for substrate analog S2.

Table 10. Peak assignment for substrate analog S2

Position	δ_C, mult.	δ_H (J in Hz)
1	6.05 (d, 1H)	89.99
2	4.39 (m, 1H)	75.02
3	4.24 (m, 1H)	68.94
4	4.16 (m, 1H)	83.11
5	4.04-3.94 (m, 2H)	63.59
6	--	158.61
7	--	152.62
8	--	149.53
9	--	106.77
10	8.19 (s, 1H)	141.33




S3_C13.esp
Normalized Intensity
Chemical Shift (ppm)
M01(s)
M02(s)
M03(s)
M04(s)
M05(s)
M06(s)
M07(s)
M08(s)
M09(s)
M10(s)
M11(s)
M12(s)
M13(s)
M15(s)
M16(s)
M17(s)
M18(s)
M19(s)
M20(s)
M21(s)
M22(s)
M23(s)
158.19
154.97
154.87
152.78
148.86
97.61
97.55
85.66
85.32
70.95
70.80
70.66
61.92
61.74
61.57
46.44
46.29
46.13
41.90
18.83
10.10
8.04
7.91
7.79
7.66

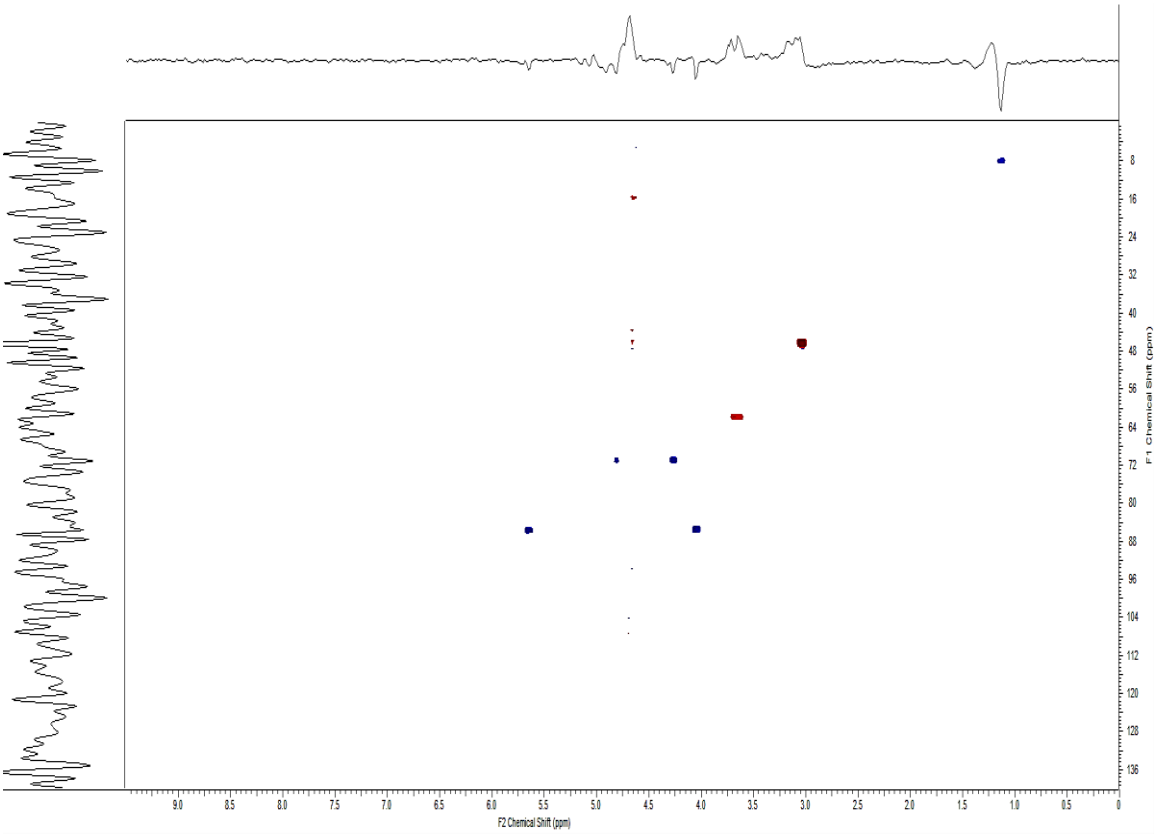
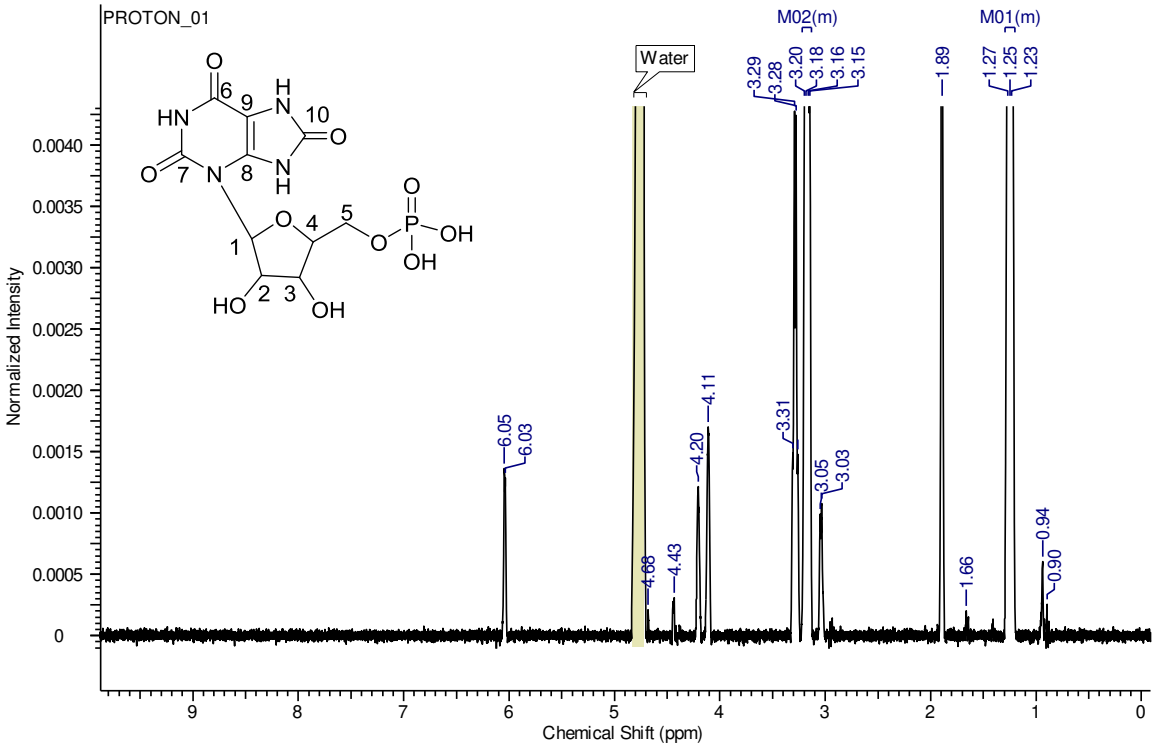

F2 Chemical Shift (ppm)
F1 Chemical Shift (ppm)

Figure 3.3.5.5. ^{1}H, ^{13}C, 2D ^{1}H – ^{13}C gHSQC (500MHz, D_2O) for substrate analog S3.

Table 11. Peak assignment for substrate analog S3

Position	δ_C, mult.	δ_H (J in Hz)
1	5.73 (d, 1H)	85.66
2	4.89 (m, 1H)	70.95
3	4.35 (app t, 1H)	70.66
4	4.13 (m, 1H)	85.32
5	3.81-3.69 (m, 2H)	61.74
6	--	154.97
7	--	148.86
8	--	97.61
9	--	152.78
10	--	158.19



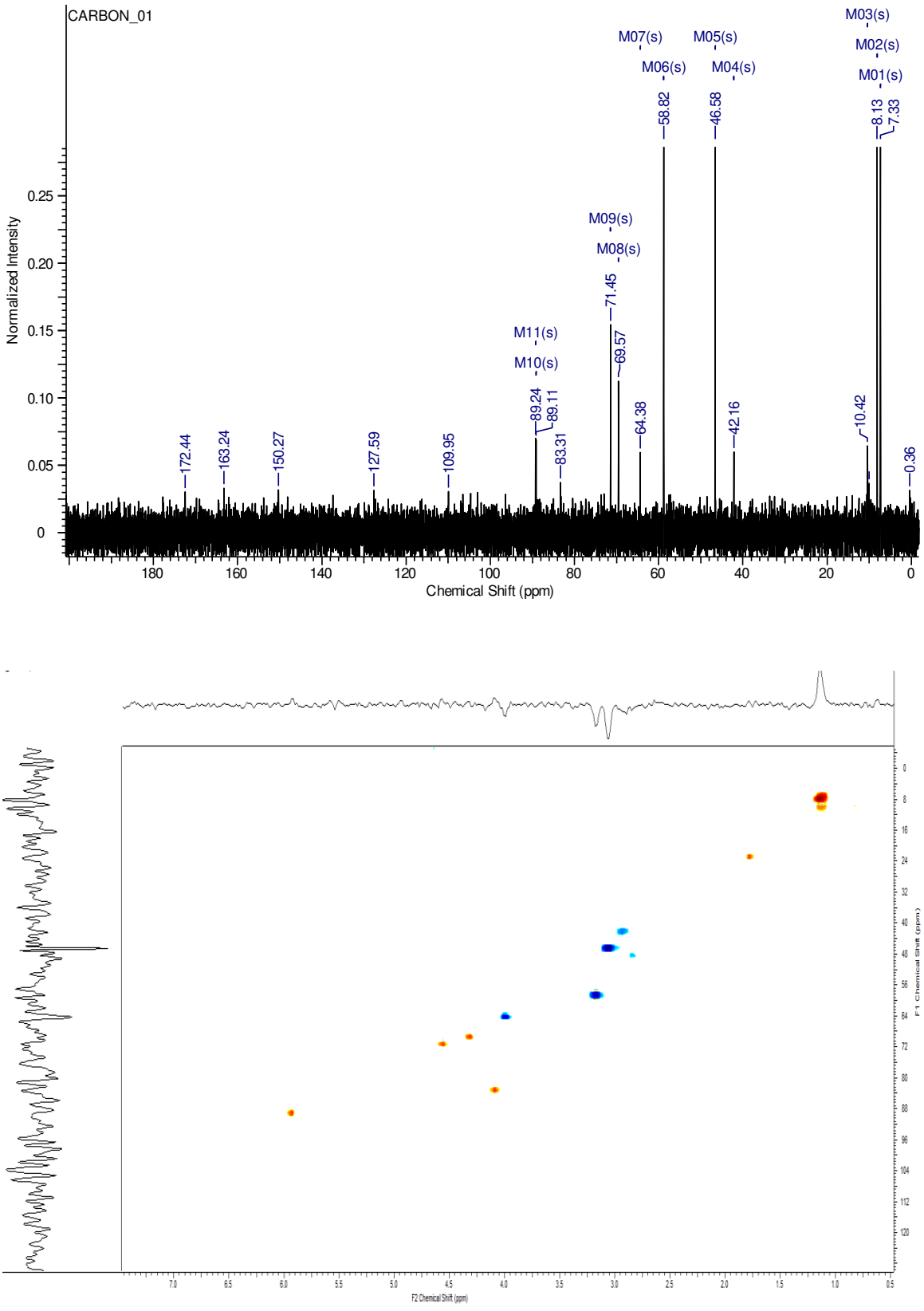


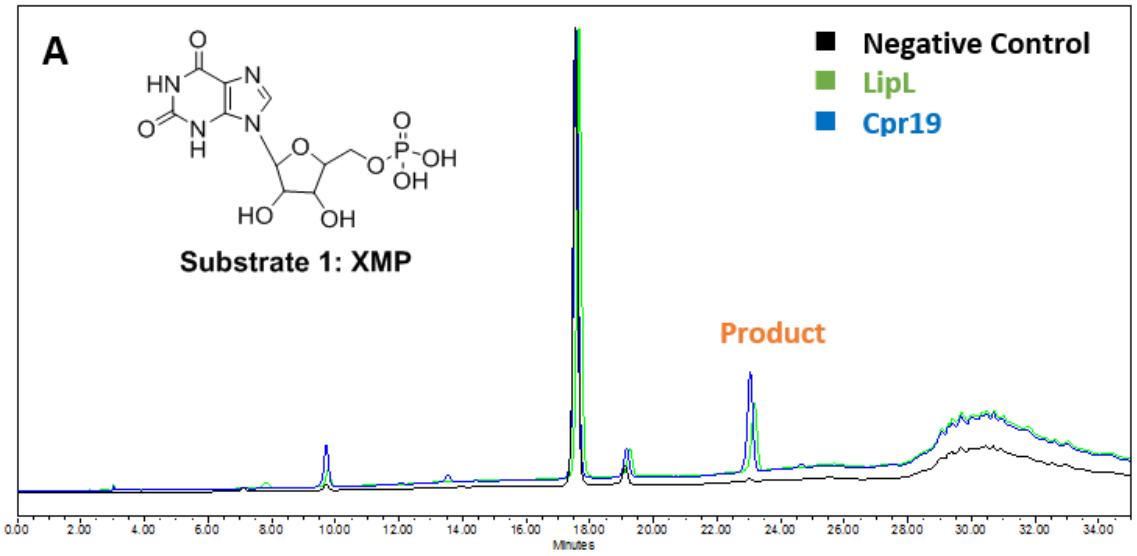
Figure 3.3.5.6. ^{1}H, ^{13}C, 2D ^{1}H – ^{13}C gHSQC (500MHz, D_2O) for substrate analog S4.

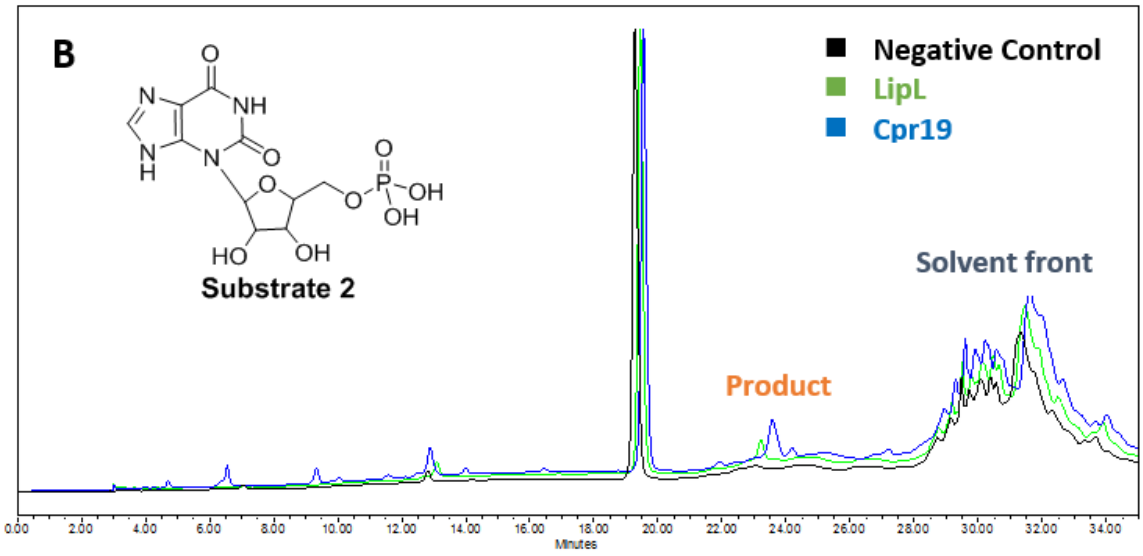
Table 12. Peak assignment for substrate analog S4

Position	δ_C, mult.	δ_H (*J* in Hz)
1	6.04 (d, 1H)	89.11
2	4.67 (m, 1H)	71.45
3	4.42 (app t, 1H)	69.57
4	4.2 (m, 1H)	83.31
5	4.10 (m, 2H)	64.38
6	--	163.24
7	--	127.59
8	--	109.95
9	--	150.27
10	--	172.44

3.3.6. *In- vitro* utilization of substrate analogs (XMP, S2, S3, S4) by LipL and Cpr19:

The previously synthesized substrate analogs (XMP, S2, S3 and S4) were tested in reaction *in vitro* with both LipL and Cpr19, with the initial aim of probing substrate promiscuity for both enzymes. HPLC traces obtained from these studies revealed product peaks generated from all four substrate analogs by *both* LipL and Cpr19 (**Figure 3.3.6.1**). It is important to note here that we conducted these reaction in HEPES buffer (as opposed to Tris-HCl) since we did not want the *predicted* aldehyde products to be converted to their corresponding tris-adducts. The generated product peaks were collected for subsequent chromatographic analyses.


A
Negative Control
LipL
Cpr19
Substrate 1: XMP
Product
Minutes


B
Negative Control
LipL
Cpr19
Substrate 2
Solvent front
Product
Minutes

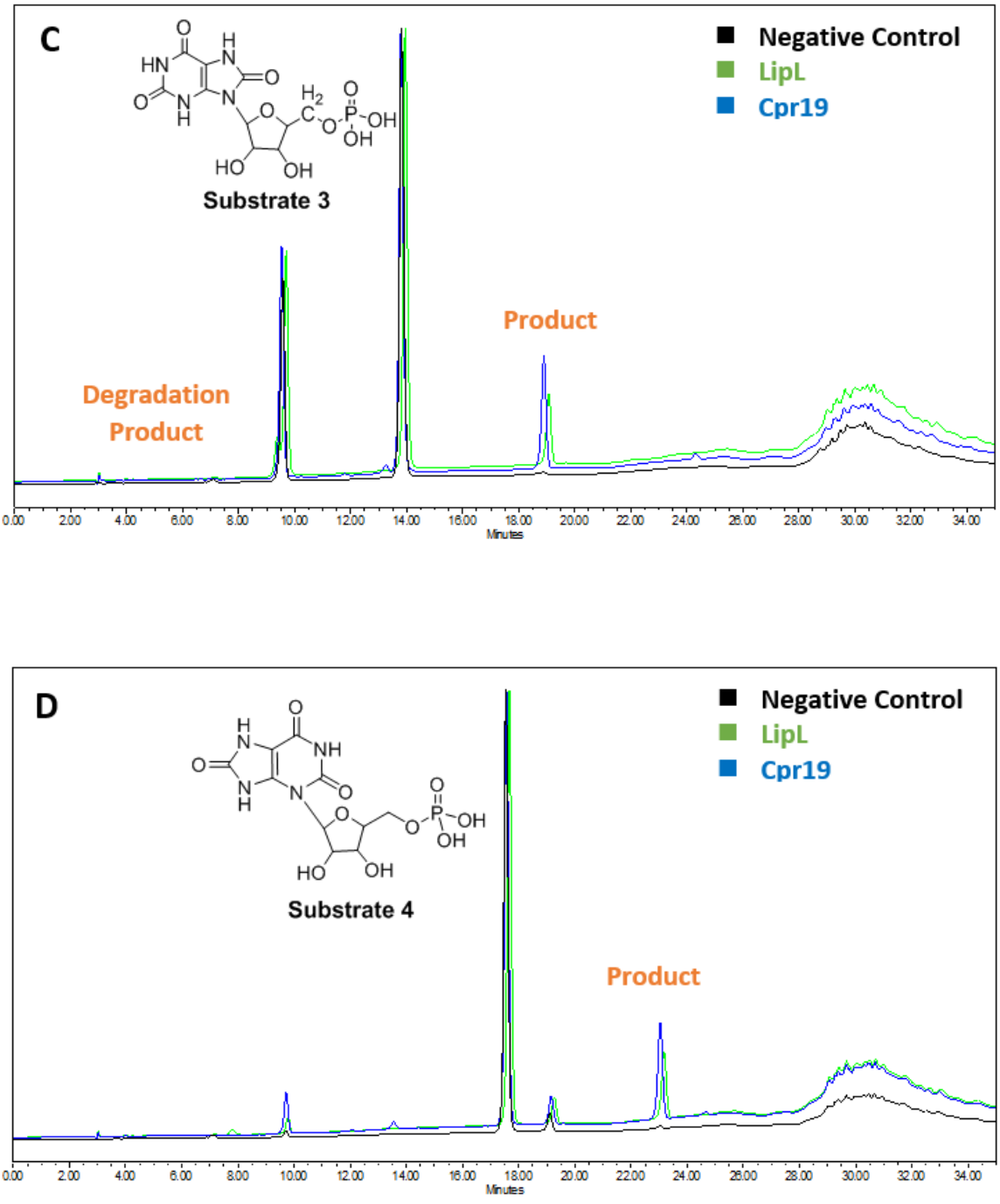


Figure 3.3.6.1. HPLC analyses of reactions catalyzed by Cpr19 and LipL when tested with substrate analogs XMP (A), S2 (B), S3 (C) and S4 (D). In all four chromatograms negative controls are shown in **black**, while reactions catalyzed by LipL and Cpr19 are indicated in **green** and **blue** respectively. In all four sets of reactions, the major peaks represent the respective substrate peaks as labeled, and the product peaks (indicated in **orange** in each set of reactions) were collected for further identification.

3.3.7. Scale-up of product for NMR and mass spectroscopic analyses

Large scale isolation of the *predicted* aldehyde products starting from substrate analogs (XMP, S2, S3 and S4) was carried out with HPLC using a C-18 reverse-phase semi-preparative column using previously described iron-pairing conditions (flowrate of 3.5 mL/min). We selected the reaction with the highest yield of a product peak (starting from **S3**) specifically to generate enough of the product to be able to gain both high resolution mass spectrometric as well as ^{1}H and 2D ^{1}H-^{13}C gHSQC NMR analyses (**Figure 3.3.7.1, Figure 3.3.7.2(c), Figure 3.3.7.3**) . The innate instability of the aldehyde product limited the scalability to obtain enough of the compound for assigning ^{13}C spectra to the peaks. Identity of the other three products were confirmed by high resolution mass spectrometric analyses, since a low yield of product combined with the inherent instability of aldehydes when subjected to column purification conditions made it difficult to generate enough of the product for NMR analyses.

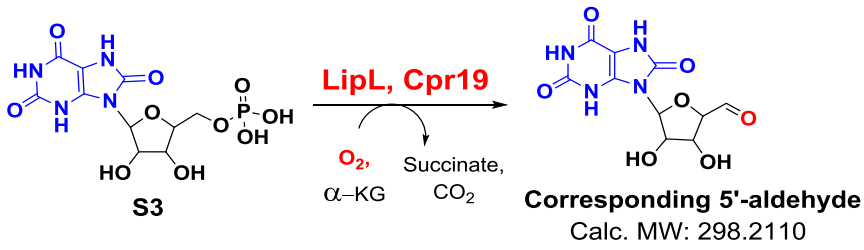
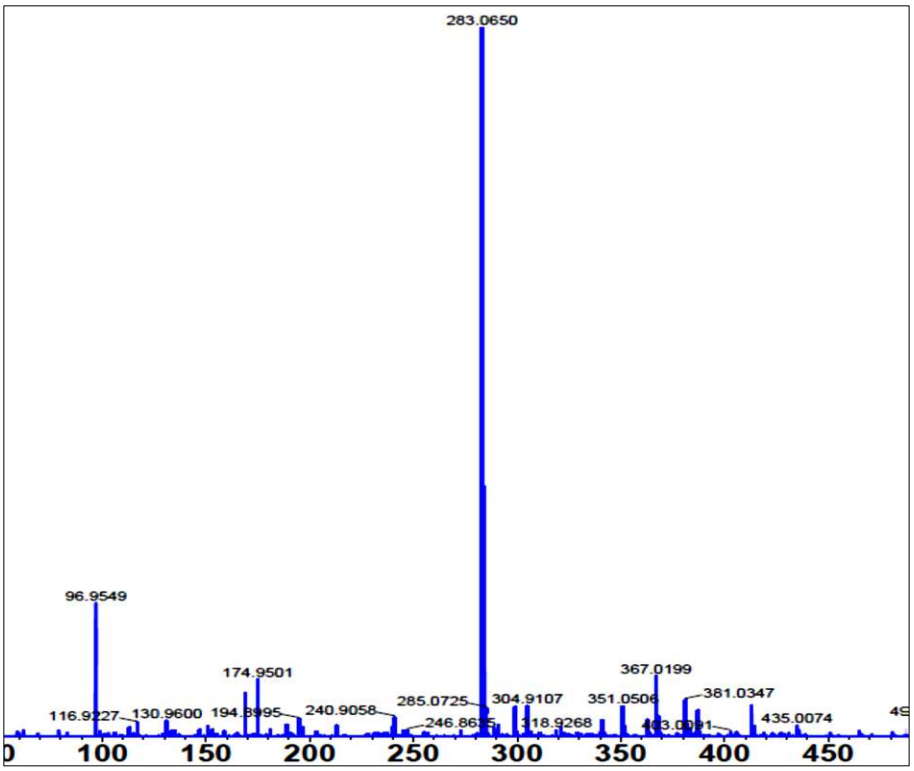
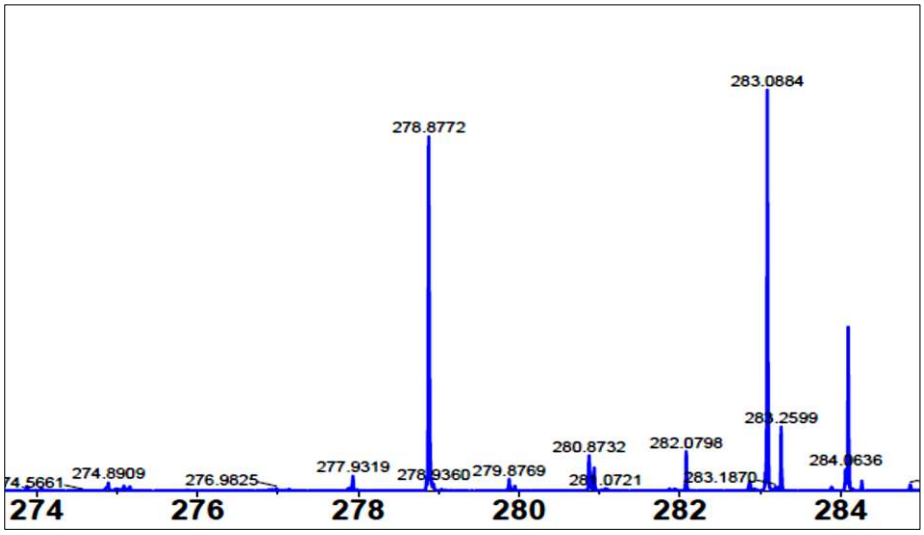
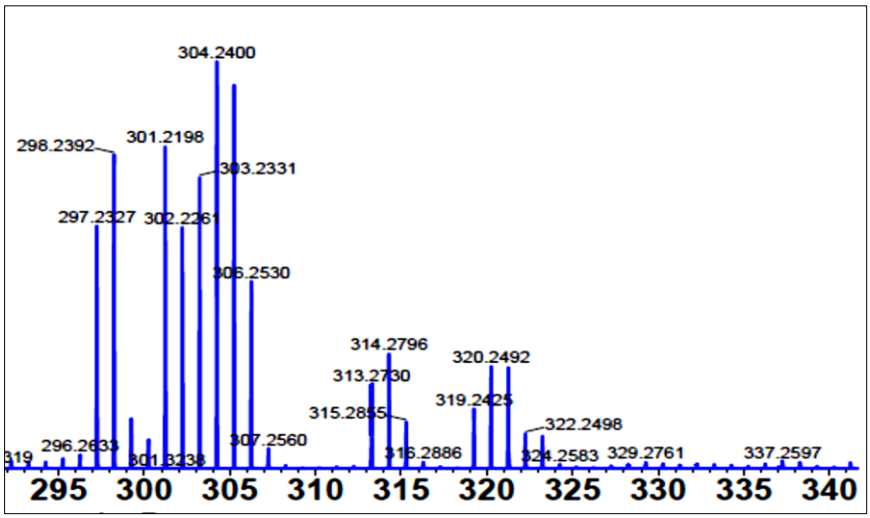


Figure 3.3.7.1. Projected net two-electron oxidation of substrate analog S3 to corresponding 5'-aldehyde catalyzed by both LipL and Cpr19. This conversion would imply flexibility of substrate recognition for both enzymes with respect to the 'base' component of its nucleotide substrate.


283.0650
96.9549
174.9501
367.0199
381.0347
116.9227
130.9600
194.8995
240.9058
285.0725
304.9107
351.0506
435.0074
100
150
200
250
300
350
400
450


283.0884
278.8772
280.8732
282.0798
274.8909
276.9825
277.9319
279.8769
283.1870
284.0636
274
276
278
280
282
284



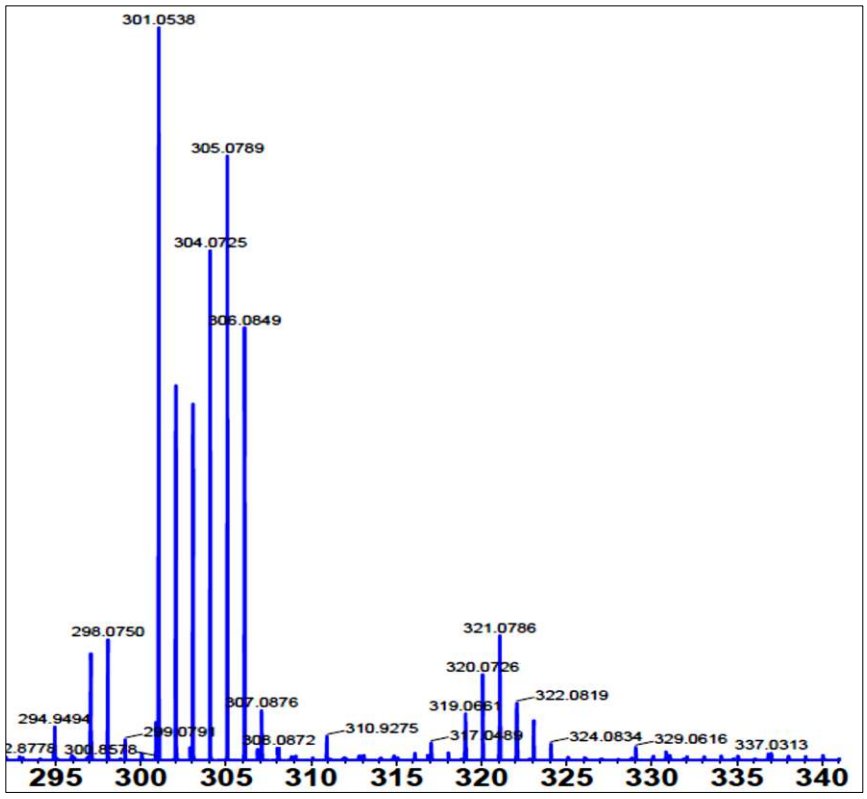


Figure 3.3.7.2. HRMS analyses for product peaks collected from reactions starting from XMP (A), S2 (B), S3 (C) and S4 (D). (**A**) HRMS (ESI+) calcd. for $C_{10}H_{10}N_4O_6$ $[M + H]^+$ 283.0634; found 283.0650; (**B**) HRMS (ESI+) calcd. for $C_{10}H_{10}N_4O_6$ $[M + H]^+$ 283.0634; found 283.0884; (**C**) HRMS (ESI+) calcd. for $C_{10}H_{10}N_4O_7$ $[M + H]^+$ 298.0549; found 298.2392; (**D**) HRMS (ESI+) calcd. for $C_{10}H_{10}N_4O_7$ $[M + H]^+$ 298.0549; found 298.0750.

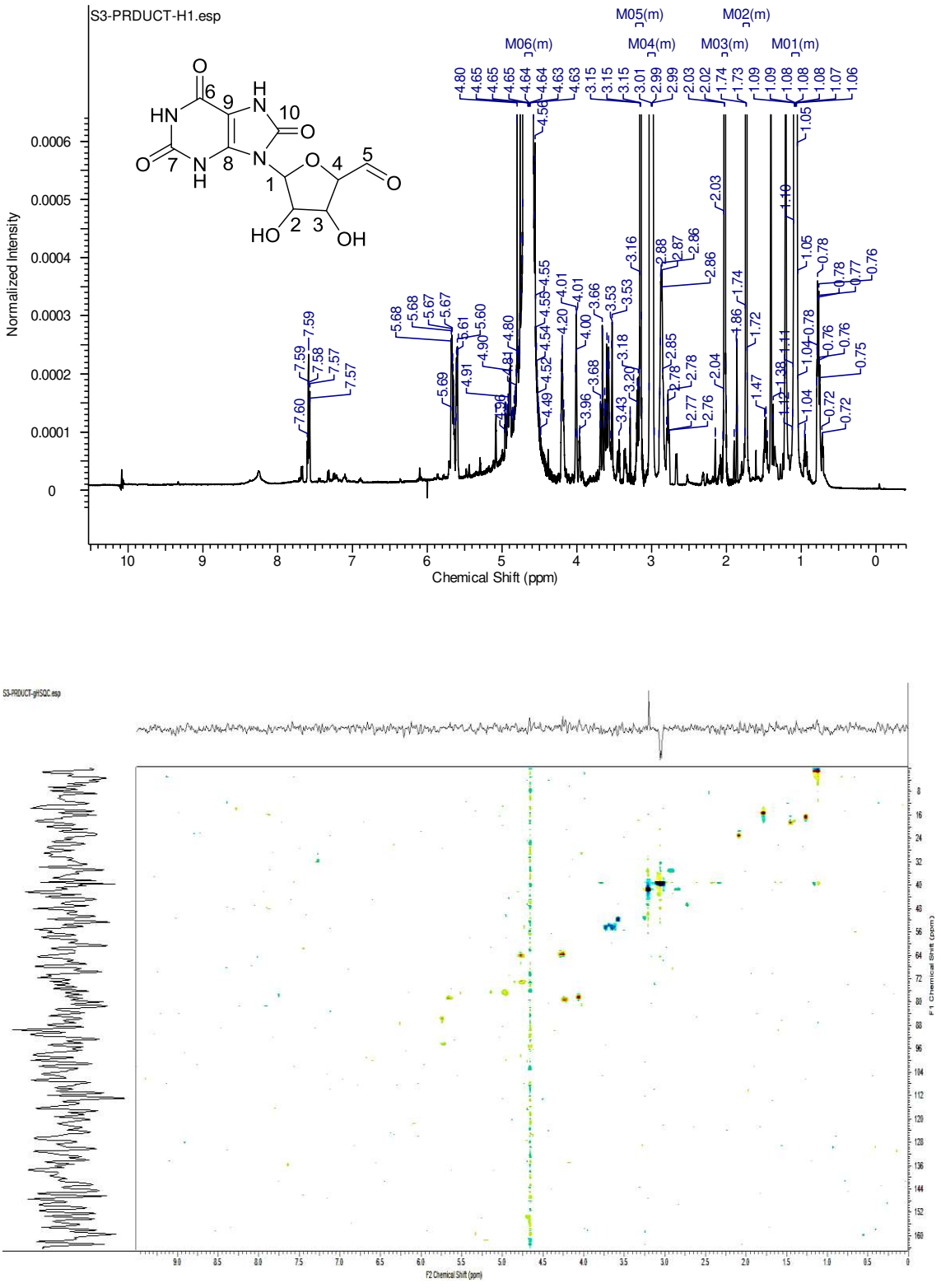


Figure 3.3.7.3. ^{1}H, 2D ^{1}H – ^{13}C gHSQC (500MHz, D_2O) for the predicted aldehyde product from substrate analog S3.

Table 13. Peak assignment for proposed aldehyde product from S3

Position	δ_C, mult.	δ_H (*J* in Hz)
1	5.72 (d, 1H)	85.99
2	4.77 (m, 1H)	64.13
3	4.27 (app t, 1H)	63.55
4	4.08 (m, 1H)	78.12
5	7.67 (s, 1H)	135.84
6	--	Could not be assigned
7	--	Could not be assigned
8	--	Could not be assigned
9	--	Could not be assigned
10	--	Could not be assigned

3.3.8. Homology Modeling of Cpr19 and LipL structures

The predicted overall fold of Cpr19 is significantly similar to that of TauD, with ~25% sequence identity. Interactions with Fe(II) in the catalytic center are proposed to involve the residues His 105, Asp 107 and His 247 which form the facial triad, consistent with our mutational studies. The predicted residues involved in stabilizing α-KG in the active site are Thr 133 and Arg 260, based on alignment with TauD residues as well as our observation of abolished enzyme activity with the Cpr19 R260A mutant (**Figure 3.3.8.1-2**).

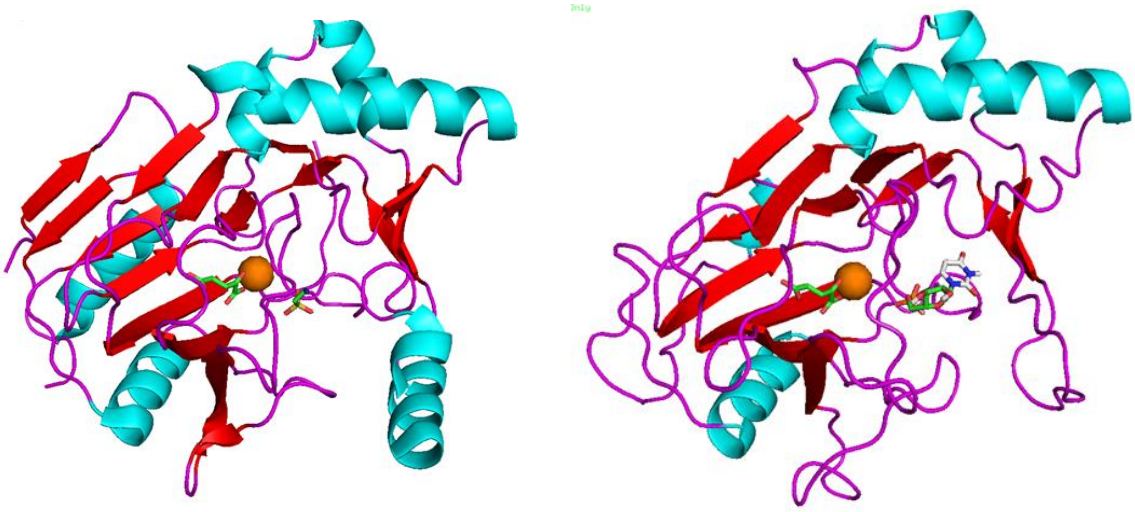


Figure 3.3.8.1: **Comparison of the structures of TauD (on the left) and the predicted structure for Cpr19 (on the right).** The beta-strands of the conserved jellyroll motif are colored in separately in **red**. Similarity of the core structures are visible, while the main differences involve the extended loops over the active sites and helix in TauD (residues 160-182) that could not be modeled onto our structure. Substrates of each enzymes are shown as sticks (α -KG, taurine and UMP) or the Fe(II) center is depicted as a sphere.

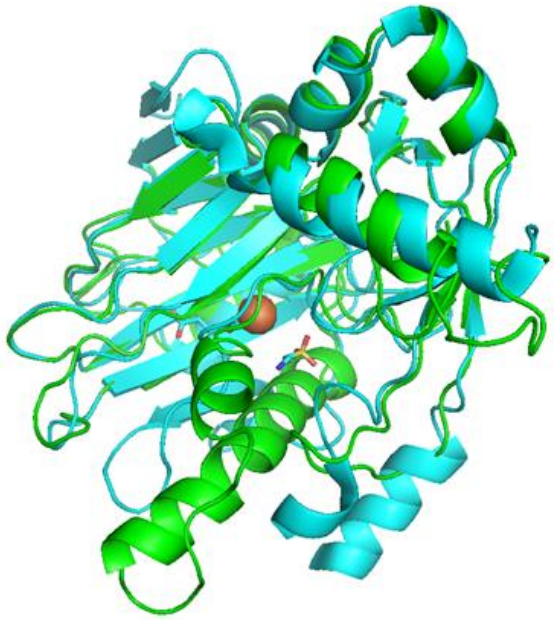


Figure 3.3.8.2. The predicted structure of Cpr19 (in blue) superimposed on the TauD crystallographic structure (in green) indicates the regions that align smoothly, barring the region extending from residues 160-182 in TauD.

Primary substrate UMP (**1**) was docked onto the active site manually (**Figure 3.3.8.3**). As an initial constraint, we superimposed the phosphate moiety of UMP with the analogous sulfate moiety of taurine in TauD to position UMP in the active site. Since we showed in the preceding chapter that Cpr19 selectively oxidizes the *pro*-S hydrogen of UMP, we docked the substrate in a position so that this hydrogen would face the active site Fe(II) center (**Figure 3.3.8.4**). In TauD, the amine of taurine is hydrogen bonded to Tyr 73, Asn 95 and Ser 158, whereas the sulfonate interacts with Arg 270, His 70 and the backbone N-H of Val 102. While we could not implicate the residues that could be involved in binding the uracil component of UMP, we could implicate Arg 264 (as the putative equivalent of Arg 270 in TauD) for stabilizing the phosphate moiety of the substrate.

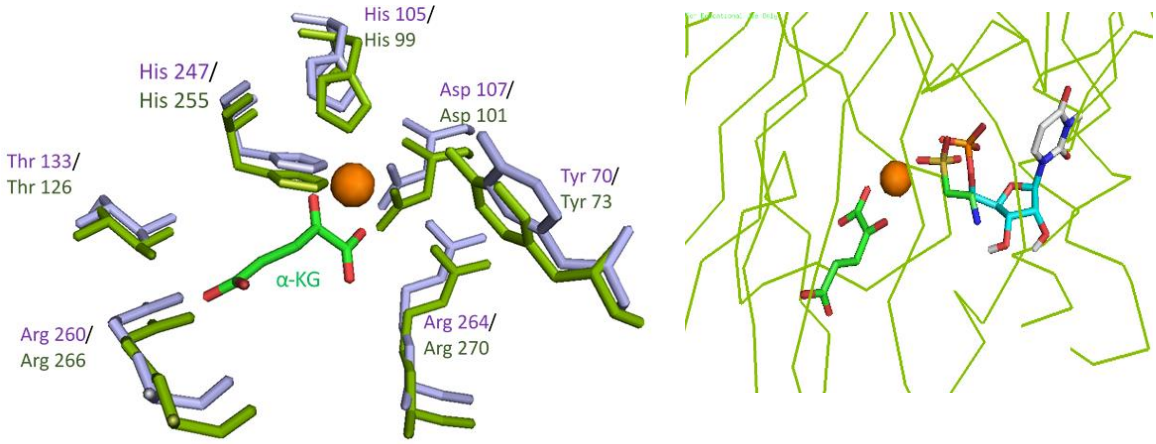


Figure 3.3.8.3. Overlay of the binding pocket of TauD (green) and Cpr19 (light blue). The active site amino acids that are conserved between the two enzymes are shown in green (TauD) and blue (Cpr19) (left). Docked UMP resides close to taurine in the TauD active site, and its phosphate moeity has a conformation similar to the sulfate group in taurine. Background ribbon shows surrounding protein structure (right).

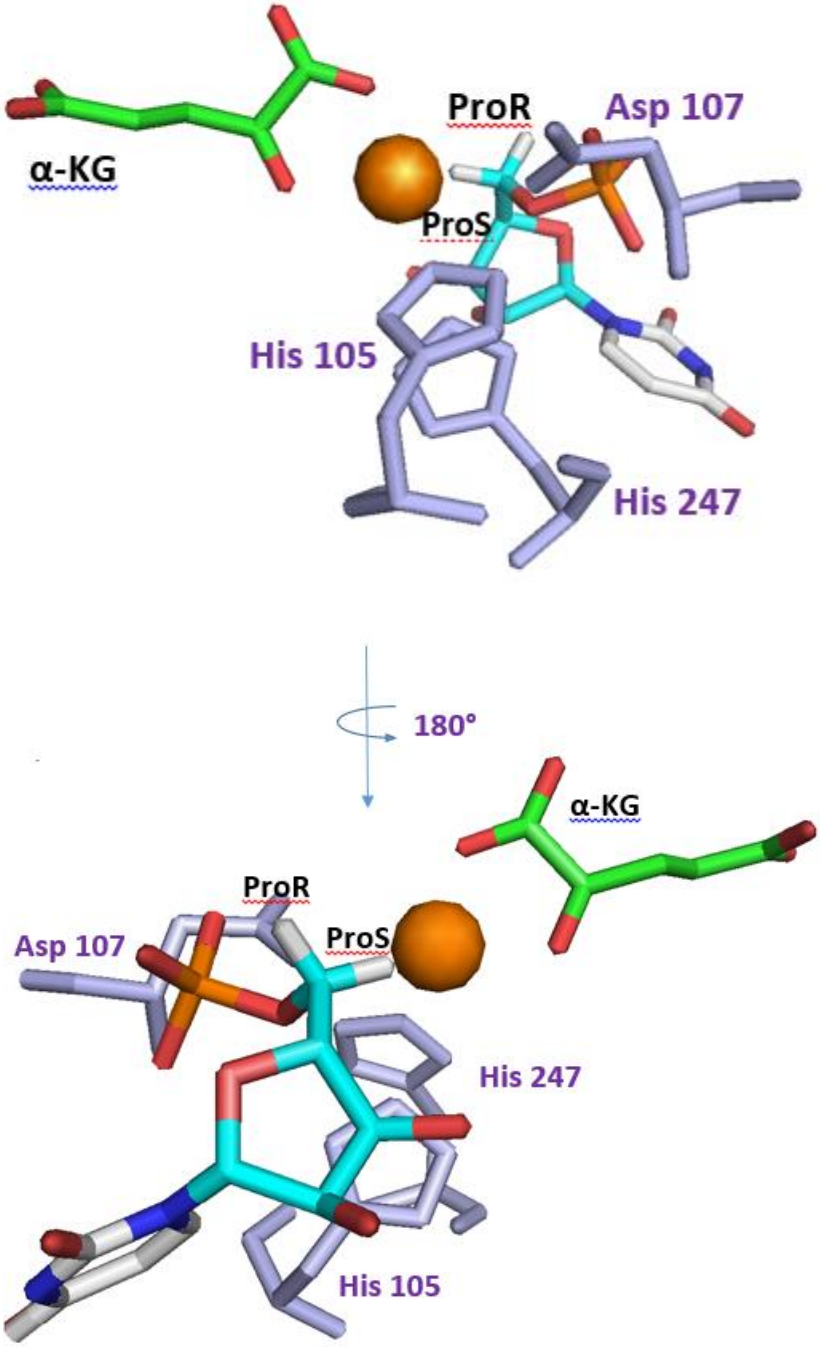


Figure 3.3.8.4. Docked UMP in the predicted active site of Cpr19. UMP was positioned such that the pro-S hydrogen at C-5' is directed towards the Fe(II) center (shown in **orange**). Amino acid residues in the facial triad and co-substrate α-KG is additionally depicted (Top). A 180^{o} rotation of the same is viewed in the bottom panel.

It is interesting to note here, in TauD the residues binding the amine part are less conserved than the residues that bind the sulfonate component of substrate taurine. Additionally, the amine component is relatively solvent exposed with a sizeable cavity adjacent to its position which confers relative substrate flexibility to TauD wherein it has been shown to accommodate alternative larger sulfonate substrates such as 3-(N-morpholino)propanesulfonic acid (MOPS) [69], which can be understood in relation to the available substrate-binding cavity and presumably the higher importance of the sulfonate-binding ligands over those binding the amine group of taurine. In our predicted Cpr19 model, we observe a similar cavity where the uracil component of UMP was docked. We borrow on this logic to dock the alternative substrates that were shown to be recognized as substrates (as well as being converted to their corresponding 5'-aldehydes), by both Cpr19 and LipL in the preceding sections. We aligned the monophosphate moieties of each of these substrates with the taurine sulfonate group to arrive at predictive active site positions for each of these substrates as shown in **Figures 3.3.8.5-6.**

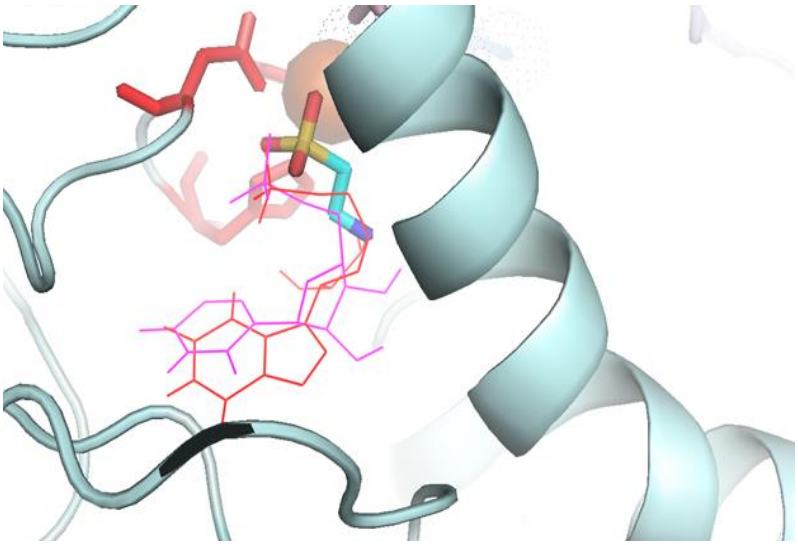


Figure 3.3.8.5. Docked XMP in the predicted active site. XMP (in **red**) overlaps primary substrate UMP (shown in **pink**), which in turn resides close to taurine (in **rainbow** colors) in the active site. Amino acid residues involved in the facial triad are depicted as sticks that interact with the Fe(II) center in the background.

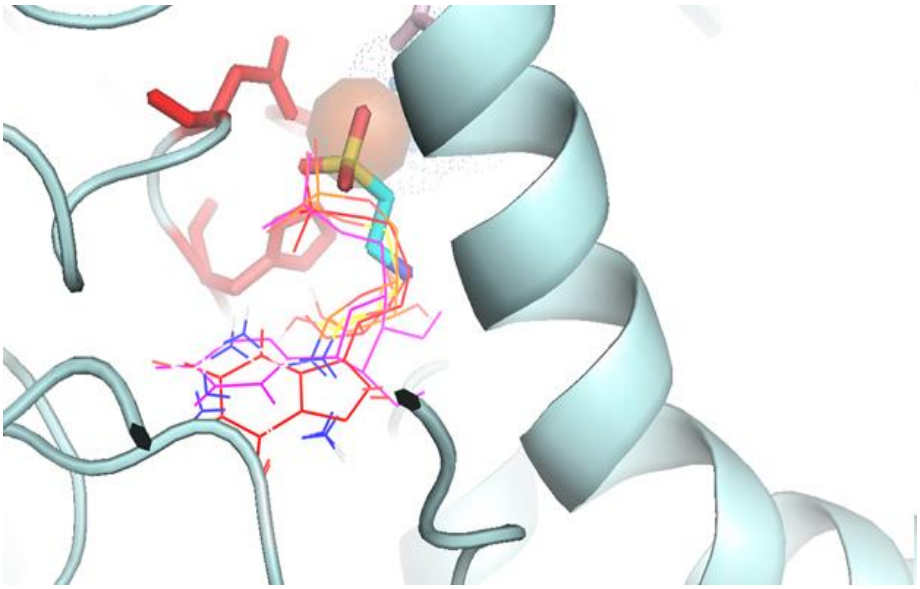


Figure 3.3.8.6. Docking all four substrate analogs into Cpr19 model. All four substrate analogs (XMP, S2, S3 and S4) are docked to overlap primary substrate UMP (shown in **pink**), docked onto the active site. The predicted Cpr19 model can accommodate these relatively larger substrate analogs in a proposed cavity adjacent to the position of the taurine amine in TauD.

Analyses of the other mutated residues that did *not* abolish the activity of Cpr19 in the predicted structure established that they are not directly involved in the active site binding pocket or in stabilizing the substrates. One of the cysteine residues (Cys 198) resides on a loop (**Figure 3.3.8.7**) and could by hypothesized to be involved in dimer formation. A native PAGE of Cpr19 (data not shown) adds weight to this hypothesis wherein we observed protein bands additional to the monomer band, which could theoretically correspond to a functional dimer or other polymeric chains of Cpr19. Both C198A and W104A mutants had higher activities than the wild type Cpr19 (**Table 8**), which could be indicative of indirect of indirect influences of both these residues on substrate access to the active site of the enzyme. Especially with reference to the Trp 104 residue, this could be a plausible explanation since it is located adjacent to His 105 and Asp 107, which are involved in forming the facial catalytic triad that coordinate the Fe(II) in the catalytic center.

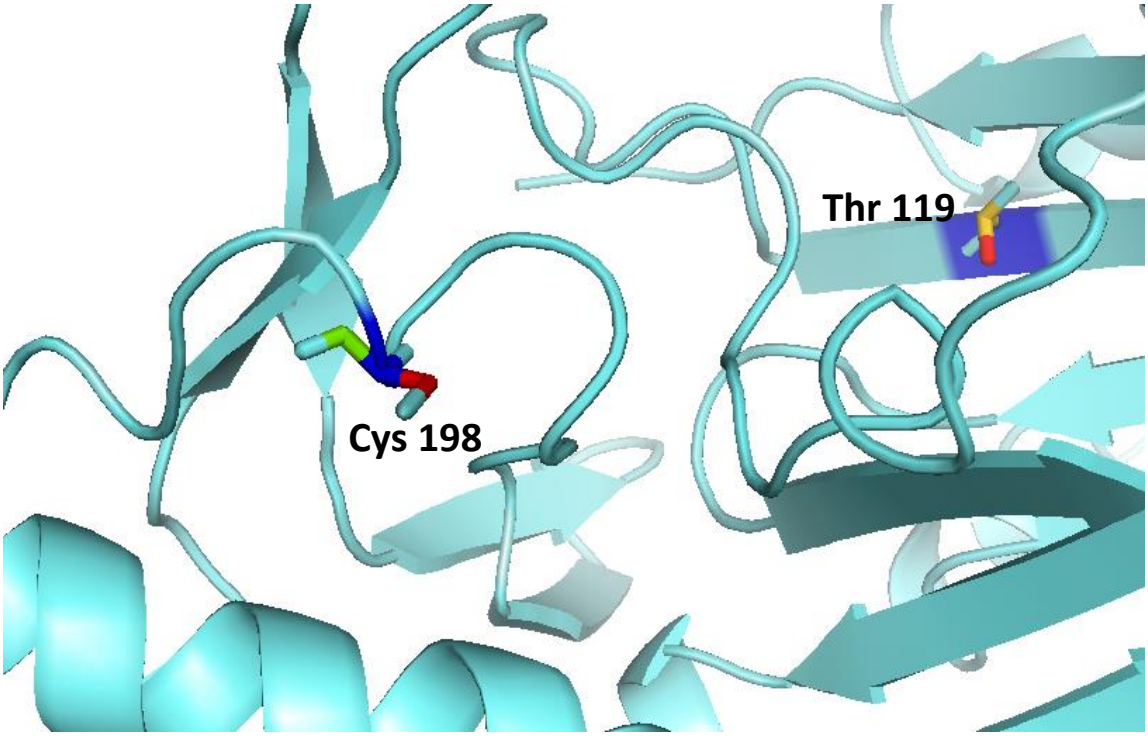


Figure 3.3.8.7. Mutated residues in Cpr19 model. Mutated residue Cys 198 is located on a loop and could be theorized to take part in disulfide formation in chain polymerization. Residue Thr 199 is located on one of the β-strands that constitute the jellyroll fold of Cpr19.

For developing the LipL model, the previously predicted Cpr19 homology model was used as a template given the high degree of homology between LipL and Cpr19. We then inserted Fe(II) and co-substrate α-KG using PyMOL onto the predicted LipL active site, resulting in nearly perfect alignment of the active site residues as expected (**Figures 3.3.8.8-9**).

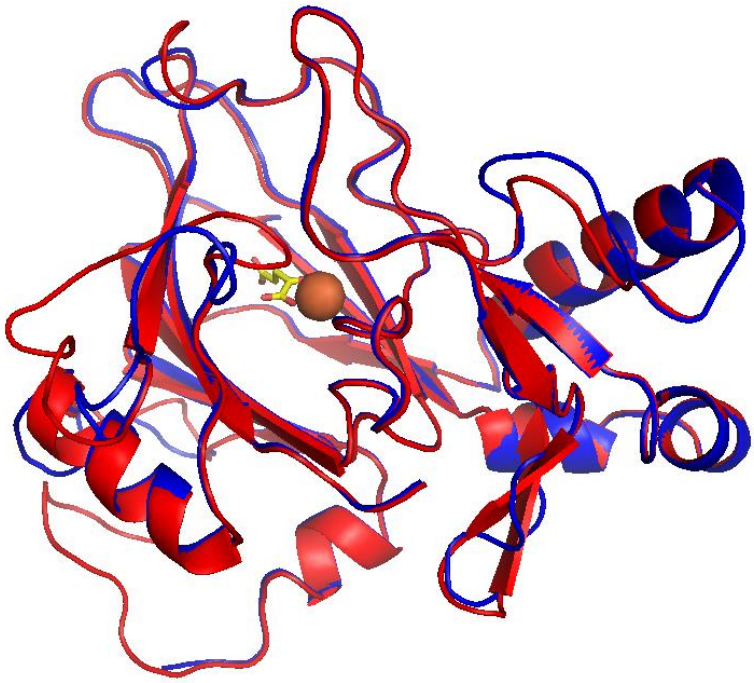


Figure 3.3.8.8. The predicted homology model of LipL. Predicted LipL structure (in **blue**) superimposed on the Cpr19 model (in **red**) indicates the regions that align smoothly, with minor differences in the extended regions with loops. Cosubstrate α-KG and Fe(II) are modeled onto the predicted active site.

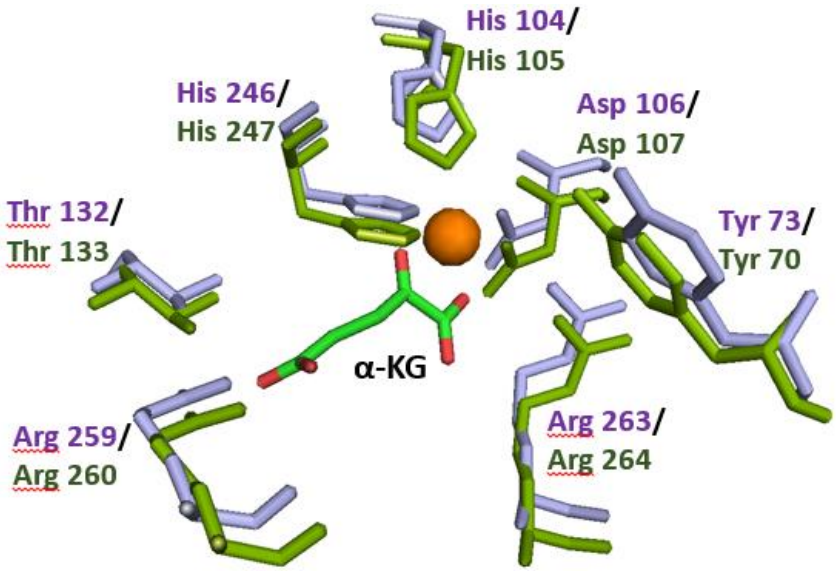


Figure 3.3.8.9. Overlay of the binding pocket of Cpr19 and LipL. Cpr19 residues (green) and LipL residues (light blue) are aligned to show conserved regions.

3.4. Conclusion

In summary, we have identified the amino acid residues His 104, Asp 107 and His 247 as the key residues involved in forming the 2His-1 carboxylate facial triad in Cpr19 via a combination of bioinformatics analyses, mutational and computational studies. By using TauD as a template, we first identified residues that were conserved with Cpr19 and LipL. This afforded us the platform to target certain residues in Cpr19 that could potentially be implicated in playing vital roles in enzyme catalysis. Initial hypothesis for a tandem phosphatase/oxidoreductase mechanism for Cpr19 (and LipL) led to the identification of the cysteine residues (Cys 123 and Cys 198) as plausible sites for a phosphatase type mechanism. However, characterization of the enzymes as dioxygenases, accompanied by the results from mutating these residues effectively negated the phosphatase hypothesis. Abolishment of activity in the Cpr19 R260A mutant as well as its alignment with the equivalent Arg 266 in TauD helped to establish its role in stabilizing co-substrate α-KG in the active site. Prediction and alignment of the Cpr19 model utilizing TauD crystallographic structure as a template furnished two more conserved residues: Thr 133, which is proposed to interact with α-KG, and Arg 264 which could be involved in stabilizing the phosphate moiety of UMP. Prime substrate UMP, as well as substrate analogs XMP, S2, S3 and S4 were docked into the predicted active site drawing on combinatory chemical logic and alignment with taurine sulfate moiety. A predictive model for LipL was also generated by using the predicted Cpr19 as a template.

3.5. Discussion

Enzymes belonging to the Fe(II): α-KG dependent superfamily are some of the most versatile enzymes that can stereoselectively activate alkyl C-H bonds in a variety of substrates resulting in many pivotal metabolic transformations. Despite the broad spectrum of distinct transformations catalyzed by these enzymes, crystal structures of a number of different members from this group show a conserved double-stranded β-helix (or jellyroll) as a common architecture for this superfamily . The highly conserved His^1-X-Asp/Glu-X_n-His^2 motif (facial triad) constitutes the facial metal-binding ligands that binds the high-spin Fe(II) center. Sequence analyses of members in this group reveals little overall similarity, leading to the proposal that convergent evolution to a common mechanism and active site chemistry occurred within the wider family of Fe(II)/ α-KG dependent and related oxygenases.

While a significant number of enzymes from this superfamily have been crystallized, the enzyme TauD has proven to be the most suitable model for the dioxygenases LipL and Cpr19 characterized in our lab, given the high degree of structural similarities between primary substrates taurine and UMP. In the absence of crystallographic structures of these enzymes, we have designed mutational variants of Cpr19 aimed at identifying key amino acid residues that impact enzyme activity. Either retention or abolishment of enzyme activity in individual mutants helped in establishing proposed roles for some of these residues in influencing catalysis. An in-silico model of Cpr19 was generated as an

attempt to locate these residues in the predicted enzyme structure. We assimilated the information gathered from the mutational studies with the obtained model to assign theoretical roles to a few more residues based on the model. Additionally, we docked the primary substrate UMP based on its alignment with taurine in the TauD active site, incorporating our observation from the preceding sections that the *pro*-S hydrogen at C-5' of UMP is preferentially abstracted. While these studies are purely theoretical, we believe we have obtained a suitable *model* for Cpr19 and LipL to further build upon, while parallel efforts to crystallize the enzymes are ongoing in our lab. We have had limited success with LipL, wherein 5Å resolution crystals were obtained from a high throughput screening (**Figure 3.5.1**). However, this resolution is too high to be able to visualize its tertiary structure accurately.

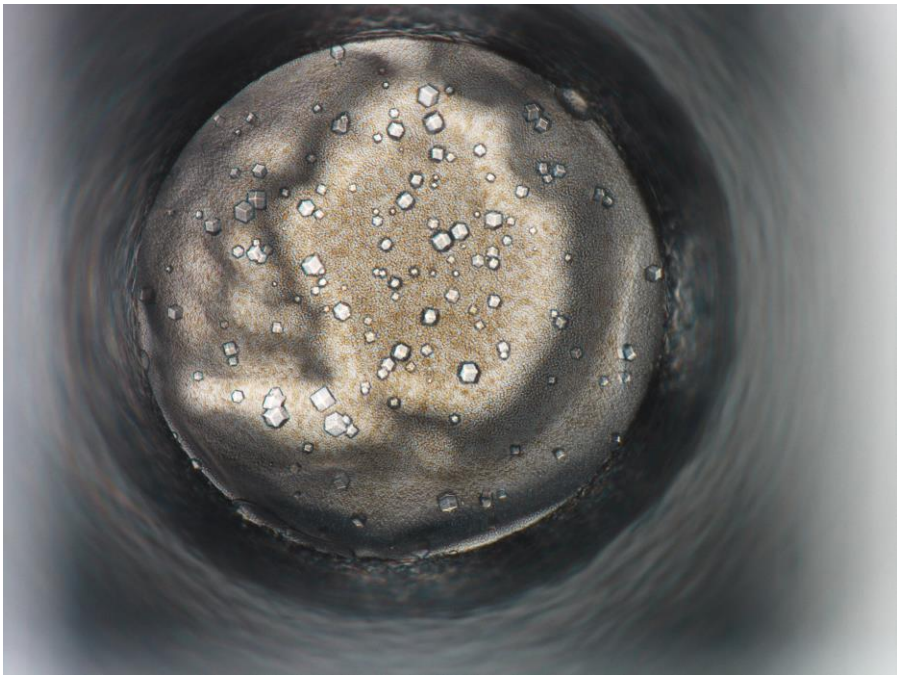


Figure 3.5.1. Crystals of LipL obtained from high throughput screens

Chapter four: Summary

MraY (translocase I) is one of the twelve ubiquitous enzymes involved in peptidoglycan biosynthesis, and viable target for targeting and antibiotic development. Of the several classes of nucleoside inhibitors, we emphasize on the pathways leading to the liposidomycins and capuramycins, and more specifically on the enzymes responsible for initiating these pathways. The dioxygenases LipL and Cpr19 were both characterized as members of an intriguing and diverse superfamily of Fe(II)/α-KG dependent enzymes and shown to possess the capability to utilize UMP as the starting precursor for high-carbon nucleoside biosynthesis. In this thesis, we delve into the mechanistic details of these enzyme machineries, highlighting the unique reaction coordinates of both these enzymes by overcoming the limitations of its proposed radical-mediated kinetics. By engineering substrate analogs for specific purposes (deuterium labeled UMP and chemically synthesized phosphonate analog) we provide distinctive methodologies to be able to differentiate between a 'hydroxylation' versus 'desaturation' hypothesis, establishing that the enzyme in fact hydroxylates UMP at C-5' and it does so by first preferentially abstracting the *pro-S* hydrogen from this position. This key observation was used down the line to obtain an *in-silico* model for Cpr19 with UMP docked onto its catalytic pocket. We additionally explored, and unexpectedly discovered substrate promiscuity of both enzymes resulting in novel products, which in theory can be further utilized to generate novel scaffolds from these (and related) pathways. Ongoing efforts to crystallize both enzymes would hopefully lend weight to our hypothesized models, and once completed

would afford the opportunity to engineer related enzymes from other strains as well to generate analogs by combinatorial biosynthesis.

Bibliography

[1] F.E. Koehn, G.T. Carter, The evolving role of natural products in drug discovery, Nat Rev Drug Discov, 4 (2005) 206-220.

[2] A.L. Harvey, Natural products in drug discovery, Drug Discovery Today, 13 (2008) 894-901.

[3] J.W.-H. Li, J.C. Vederas, Drug Discovery and Natural Products: End of an Era or an Endless Frontier?, Science, 325 (2009) 161-165.

[4] H.M.L. Davies, Organic chemistry: Synthetic lessons from nature, Nature, 459 (2009) 786-787.

[5] C.T. Walsh, M.A. Fischbach, Natural Products Version 2.0: Connecting Genes to Molecules, Journal of the American Chemical Society, 132 (2010) 2469-2493.

[6] M.G. Watve, R. Tickoo, M.M. Jog, B.D. Bhole, How many antibiotics are produced by the genus Streptomyces?, Archives of microbiology, 176 (2001) 386-390.

[7] R.H. Baltz, Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?, Journal of Industrial Microbiology and Biotechnology, 33 (2006) 507-513.

[8] J. Berdy, Thoughts and facts about antibiotics: Where we are now and where we are heading, J Antibiot, 65 (2012) 385-395.

[9] M. Tulp, L. Bohlin, Functional versus chemical diversity: is biodiversity important for drug discovery?, Trends in Pharmacological Sciences, 23 (2002) 225-231.

[10] A.L. Demain, S. Sanchez, Microbial drug discovery: 80 years of progress, J Antibiot, 62 (2009) 5-16.

[11] P. Fernandes, Antibacterial discovery and development[mdash]the failure of success?, Nat Biotech, 24 (2006) 1497-1503.

[12] M.S. Butler, M.A. Cooper, Antibiotics in the clinical pipeline in 2011, J Antibiot, 64 (2011) 413-425.

[13] S. Donadio, S. Maffioli, P. Monciardini, M. Sosio, D. Jabes, Antibiotic discovery in the twenty-first century: current trends and future perspectives, J Antibiot, 63 (2010) 423-430.

[14] A. Coates, Y. Hu, R. Bax, C. Page, The future challenges facing the development of new antimicrobial drugs, Nat Rev Drug Discov, 1 (2002) 895-910.

[15] H. Von Döhren, Antibiotics: Actions, origins, resistance, by C. Walsh. 2003. Washington, DC: ASM Press. 345 pp. $99.95 (hardcover), Protein Science : A Publication of the Protein Society, 13 (2004) 3059-3060.

[16] I. Phillips, M. Casewell, T. Cox, B. De Groot, C. Friis, R. Jones, C. Nightingale, R. Preston, J. Waddell, Does the use of antibiotics in food animals pose a risk to human health? A critical review of published data, Journal of Antimicrobial Chemotherapy, 53 (2004) 28-52.

[17] C. Walsh, Where will new antibiotics come from?, Nat Rev Micro, 1 (2003) 65-70.

[18] Threat Report CDC 2013.

[19] WHO Global Tuberculosis Report 2014.

[20] M.A. Fischbach, C.T. Walsh, Antibiotics for Emerging Pathogens, Science, 325 (2009) 1089-1093.

[21] T.D.H. Bugg, C.T. Walsh, Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance, Natural Product Reports, 9 (1992) 199-215.

[22] J.v. Heijenoort, Recent advances in the formation of the bacterial peptidoglycan monomer unit, Natural Product Reports, 18 (2001) 503-519.

[23] A. Bouhss, A.E. Trunkfield, T.D. Bugg, D. Mengin-Lecreulx, The biosynthesis of peptidoglycan lipid-linked intermediates, FEMS microbiology reviews, 32 (2008) 208-233.

[24] E. Sauvage, F. Kerff, M. Terrak, J.A. Ayala, P. Charlier, The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis, FEMS microbiology reviews, 32 (2008) 234-258.

[25] S. Walker, L. Chen, Y. Hu, Y. Rew, D. Shin, D.L. Boger, Chemistry and Biology of Ramoplanin: A Lipoglycodepsipeptide with Potent Antibiotic Activity, Chemical Reviews, 105 (2005) 449-476.

[26] T.D.H. Bugg, 3.10 - Bacterial Peptidoglycan Biosynthesis and its Inhibition, in: S.D.B.N. Meth-Cohn (Ed.) Comprehensive Natural Products Chemistry, Pergamon, Oxford, 1999, pp. 241-294.

[27] D. Kahne, C. Leimkuhler, W. Lu, C. Walsh, Glycopeptide and Lipoglycopeptide Antibiotics, Chemical Reviews, 105 (2005) 425-448.

[28] M. Ikeda, M. Wachi, H.K. Jung, F. Ishino, M. Matsuhashi, The Escherichia coli mraY gene encoding UDP-N-acetylmuramoyl-pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-pentapeptide transferase, Journal of Bacteriology, 173 (1991) 1021-1026.

[29] C. Dini, MraY Inhibitors as Novel Antibacterial Agents, Current Topics in Medicinal Chemistry, 5 (2005) 1221-1236.

[30] D.D. Pless, F.C. Neuhaus, Initial Membrane Reaction in Peptidoglycan Synthesis: LIPID DEPENDENCE OF PHOSPHO-N-ACETYLMURAMYL-PENTAPEPTIDE TRANSLOCASE (EXCHANGE REACTION), Journal of Biological Chemistry, 248 (1973) 1568-1576.

[31] M.A. Lehrman, Commentary: A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases, Glycobiology, 4 (1994) 768-771.

[32] A. Geis, R. Plapp, Phospho-N-acetylmuramoyl-pentapeptide-transferase of Escherichia coli K12 Properties of the membrane-bound and the extracted and partially purified enzyme, Biochimica et Biophysica Acta (BBA)-Enzymology, 527 (1978) 414-424.

[33] A. Bouhss, D. Mengin-Lecreulx, D. Le Beller, J. Van Heijenoort, Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis, Molecular Microbiology, 34 (1999) 576-585.

[34] D.S. Boyle, W.D. Donachie, mraY Is an Essential Gene for Cell Growth in Escherichia coli, Journal of Bacteriology, 180 (1998) 6429-6432.

[35] B.C. Chung, J. Zhao, R.A. Gillespie, D.-Y. Kwon, Z. Guan, J. Hong, P. Zhou, S.-Y. Lee, Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis, Science, 341 (2013) 1012-1016.

[36] T.G. Bernhardt, W.D. Roof, R. Young, Genetic evidence that the bacteriophage ϕX174 lysis protein inhibits cell wall synthesis, Proceedings of the National Academy of Sciences, 97 (2000) 4297-4302.

[37] T.G. Bernhardt, D.K. Struck, R. Young, The Lysis Protein E of ϕX174 Is a Specific Inhibitor of the MraY-catalyzed Step in Peptidoglycan Synthesis, Journal of Biological Chemistry, 276 (2001) 6093-6097.

[38] S. Tanaka, W.M. Clemons Jr, Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174, Molecular microbiology, 85 (2012) 975-985.

[39] S. Mendel, J.M. Holbourn, J.A. Schouten, T.D.H. Bugg, Interaction of the transmembrane domain of lysis protein E from bacteriophage φX174 with bacterial translocase MraY and peptidyl-prolyl isomerase SlyD, Microbiology, 152 (2006) 2959-2967.

[40] K.-i. Kimura, T.D.H. Bugg, Recent advances in antimicrobial nucleoside antibiotics targeting cell wall biosynthesis, Natural Product Reports, 20 (2003) 252-273.

[41] B.A. Chen RH, Whittern DN, McAlpine JB., Pacidamycins, a novel series of antibiotics with anti-Pseudomonas aeruginosa activity. II. Isolation and structural elucidation, The Journal of Antibiotics (Tokyo), 42 (1989) 512-612.

[42] F. Isono, M. Inukai, Mureidomycin A, a new inhibitor of bacterial peptidoglycan synthesis, Antimicrobial Agents and Chemotherapy, 35 (1991) 234-236.

[43] G. Tamura, T. Sasaki, M. Matsuhashi, A. Takatsuki, M. Yamasaki, Tunicamycin Inhibits the Formation of Lipid Intermediate in Cell-free Peptidoglycan Synthesis of Bacteria, Agricultural and Biological Chemistry, 40 (1976) 447-449.

[44] M. Funabashi, S. Baba, K. Nonaka, M. Hosobuchi, Y. Fujita, T. Shibata, S.G. Van Lanen, The Biosynthesis of Liposidomycin-like A-90289 Antibiotics Featuring a New Type of Sulfotransferase, ChemBioChem, 11 (2010) 184-190.

[45] M. Igarashi, Y. Takahashi, T. Shitara, H. Nakamura, H. Naganawa, T. Miyake, Y. Akamatsu, Caprazamycins, Novel Lipo-nucleoside Antibiotics, from Streptomyces sp, J Antibiot, 58 (2005) 327-337.

[46] M.A. Muramatsu Y, Ohnuki T, Ishii MM, Kizuka M, Enokita R, Tsutsumi S, Arai M, Ogawa Y, Suzuki T, Takatsu T, Inukai M., Studies on novel bacterial translocase I inhibitors, A-500359s. I. Taxonomy, fermentation, isolation, physico-chemical properties and structure elucidation of A-500359 A, C, D and G., The Journal of Antibiotics (Tokyo), 56 (2003) 243-342.

[47] O.T. Muramatsu Y, Ishii MM, Kizuka M, Enokita R, Miyakoshi S, Takatsu T, Inukai M., A-503083 A, B, E and F, novel inhibitors of bacterial translocase I, produced by Streptomyces sp. SANK 62799., The Journal of Antibiotics (Tokyo), 57 (2004) 639-646.

[48] F.Y. Murakami R, Kizuka M, Kagawa T, Muramatsu Y, Miyakoshi S, Takatsu T, Inukai M., A-102395, a new inhibitor of bacterial translocase I, produced by Amycolatopsis sp. SANK 60206., The Journal of Antibiotics (Tokyo), 60 (2007) 690-695.

[49] N.P.J. Price, B. Tsvetanova, Biosynthesis of the Tunicamycins: A Review, J Antibiot, 60 (2007) 485-491.

[50] M. Inukai, F. Isono, A. Takatsuki, Selective inhibition of the bacterial translocase reaction in peptidoglycan synthesis by mureidomycins, Antimicrobial Agents and Chemotherapy, 37 (1993) 980-983.

[51] A. Takatsuki, Y. Fukui, G. Tamura, Inhibition of Heparan Sulfate Biosynthesis in Cultures of Chick Embryo Fibroblasts by Tunicamycin and Formation of Lipid-linked Glucuronic Acid in Chick Embryo Microsomes, Agricultural and Biological Chemistry, 42 (1978) 1621-1623.

[52] A.D. Elbein, The tunicamycins — useful tools for studies on glycoproteins, Trends in Biochemical Sciences, 6 (1981) 219-221.

[53] A. Heifetz, R.W. Keenan, A.D. Elbein, Mechanism of action of tunicamycin on the UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase, Biochemistry, 18 (1979) 2186-2192.

[54] S. Hirano, S. Ichikawa, A. Matsuda, Structure–activity relationship of truncated analogs of caprazamycins as potential anti-tuberculosis agents, Bioorganic & Medicinal Chemistry, 16 (2008) 5123-5133.

[55] C. Dini, S. Didier-Laurent, N. Drochon, S. Feteanu, J.C. Guillot, F. Monti, E. Uridat, J. Zhang, J. Aszodi, Synthesis of sub-micromolar inhibitors of MraY by exploring the region originally occupied by the diazepanone ring in the liposidomycin structure, Bioorganic & Medicinal Chemistry Letters, 12 (2002) 1209-1213.

[56] S.M. Y Muramatsu, Y Ogawa, T Ohnuki, MM Ishii, M Arai, T Takatsu, M Inukai, Studies on Novel Bacterial Translocase I Inhibitors, A-500359s III. Deaminocaprolactam Derivatives of Capuramycin: A-500359 E, F, H, M-1 and M-2, The Journal of Antibiotics (Tokyo), 56 (2003) 259-267.

[57] B.J.A. DENNIS R. BERRY, INCORPORATION OF 14C-LABELED COMPOUNDS INTO SINEFUNGIN (A9145), A NUCLEOSIDE ANTIFUNGAL ANTIBIOTIC, The Journal of Antibiotics (Tokyo), 31 (1978) 185-191.

[58] K. Isono, R.J. Suhadolnik, The biosynthesis of natural and unnatural polyoxins by< i> Streptomyces cacaoi</i>, Archives of biochemistry and biophysics, 173 (1976) 141-153.

[59] H.H. Miyakoshi S, Shioiri T, Takahashi S, Torikata A, Yamazaki M., Biosynthesis of griseolic acids: incorporation of 13C-labeled compounds into griseolic acid A., The Journal of Antibiotics (Tokyo), 45 (1992) 394-399.

[60] M.Y. Ohnuki T, Miyakoshi S, Takatsu T, Inukai M., Studies on novel bacterial translocase I inhibitors, A-500359s. IV. Biosynthesis of A-500359s., The Journal of Antibiotics (Tokyo), 56 (2003) 268-279.

[61] M. Funabashi, K. Nonaka, C. Yada, M. Hosobuchi, N. Masuda, T. Shibata, S.G. Van Lanen, Identification of the biosynthetic gene cluster of A-500359s in Streptomyces griseus SANK60196, J Antibiot, 62 (2009) 325-332.

[62] M. Funabashi, Z. Yang, K. Nonaka, M. Hosobuchi, Y. Fujita, T. Shibata, X. Chi, S.G. Van Lanen, An ATP-independent strategy for amide bond formation in antibiotic biosynthesis, Nat Chem Biol, 6 (2010) 581-586.

[63] L. Kaysser, S. Siebenberg, B. Kammerer, B. Gust, Analysis of the Liposidomycin Gene Cluster Leads to the Identification of New Caprazamycin Derivatives, ChemBioChem, 11 (2010) 191-196.

[64] L. Kaysser, L. Lutsch, S. Siebenberg, E. Wemakor, B. Kammerer, B. Gust, Identification and Manipulation of the Caprazamycin Gene Cluster Lead to New Simplified Liponucleoside Antibiotics and Give Insights into the Biosynthetic Pathway, Journal of Biological Chemistry, 284 (2009) 14987-14996.

[65] L. Cheng, W. Chen, L. Zhai, D. Xu, T. Huang, S. Lin, X. Zhou, Z. Deng, Identification of the gene cluster involved in muraymycin biosynthesis from Streptomyces sp. NRRL 30471, Molecular BioSystems, 7 (2011) 920-927.

[66] X. Chi, S. Baba, N. Tibrewal, M. Funabashi, K. Nonaka, S.G. Van Lanen, The muraminomicin biosynthetic gene cluster and enzymatic formation of the 2-deoxyaminoribosyl appendage, MedChemComm, 4 (2013) 239-243.

[67] W. Chen, D. Qu, L. Zhai, M. Tao, Y. Wang, S. Lin, N.J. Price, Z. Deng, Characterization of the tunicamycin gene cluster unveiling unique steps involved in its biosynthesis, Protein Cell, 1 (2010) 1093-1105.

[68] F.J. Wyszynski, A.R. Hesketh, M.J. Bibb, B.G. Davis, Dissecting tunicamycin biosynthesis by genome mining: cloning and heterologous expression of a minimal gene cluster, Chemical Science, 1 (2010) 581-589.

[69] E. Eichhorn, J.R. van der Ploeg, M.A. Kertesz, T. Leisinger, Characterization of α-Ketoglutarate-dependent Taurine Dioxygenase from Escherichia coli, Journal of Biological Chemistry, 272 (1997) 23031-23036.

[70] Z. Yang, X. Chi, M. Funabashi, S. Baba, K. Nonaka, P. Pahari, J. Unrine, J.M. Jacobsen, G.I. Elliott, J. Rohr, S.G. Van Lanen, Characterization of LipL as a Non-heme, Fe(II)-dependent α-Ketoglutarate:UMP Dioxygenase That Generates Uridine-5'-aldehyde during A-90289 Biosynthesis, Journal of Biological Chemistry, 286 (2011) 7885-7892.

[71] S. Barnard-Britson, X. Chi, K. Nonaka, A.P. Spork, N. Tibrewal, A. Goswami, P. Pahari, C. Ducho, J. Rohr, S.G. Van Lanen, Amalgamation of Nucleosides and Amino Acids in Antibiotic Biosynthesis: Discovery of an l-Threonine:Uridine-5'-Aldehyde Transaldolase, Journal of the American Chemical Society, 134 (2012) 18514-18517.

[72] Z. Yang, M. Funabashi, K. Nonaka, M. Hosobuchi, T. Shibata, P. Pahari, S.G. Van Lanen, Functional and Kinetic Analysis of the Phosphotransferase CapP Conferring Selective Self-

resistance to Capuramycin Antibiotics, Journal of Biological Chemistry, 285 (2010) 12899-12905.

[73] J.M. Bollinger, J.C. Price, L.M. Hoffart, E.W. Barr, C. Krebs, Mechanism of Taurine: α-Ketoglutarate Dioxygenase (TauD) from Escherichia coli, European journal of inorganic chemistry, 2005 (2005) 4245-4254.

[74] R.P. Hausinger, Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes, Critical Reviews in Biochemistry and Molecular Biology, 39 (2004) 21-68.

[75] L. Que, One motif [mdash] many different reactions, Nat Struct Mol Biol, 7 (2000) 182-184.

[76] K. Koehntop, J. Emerson, L. Que, Jr., The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes, J Biol Inorg Chem, 10 (2005) 87-93.

[77] L.C. Blasiak, F.H. Vaillancourt, C.T. Walsh, C.L. Drennan, Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis, Nature, 440 (2006) 368-371.

[78] E.I. Solomon, T.C. Brunold, M.I. Davis, J.N. Kemsley, S.-K. Lee, N. Lehnert, F. Neese, A.J. Skulan, Y.-S. Yang, J. Zhou, Geometric and Electronic Structure/Function Correlations in Non-Heme Iron Enzymes, Chemical Reviews, 100 (1999) 235-350.

[79] M. Costas, M.P. Mehn, M.P. Jensen, L. Que, Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates, Chemical Reviews, 104 (2004) 939-986.

[80] P.L. Roach, I.J. Clifton, V. Fulop, K. Harlos, G.J. Barton, J. Hajdu, I. Andersson, C.J. Schofield, J.E. Baldwin, Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes, Nature, 375 (1995) 700-704.

[81] T.A. Hutton JJ Jr, Udenfriend S., Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase., Biochem Biophys Res Commun, 24 (1966) 179-184.

[82] P. Annunen, T. Helaakoski, J. Myllyharju, J. Veijola, T. Pihlajaniemi, K.I. Kivirikko, Cloning of the Human Prolyl 4-Hydroxylase α Subunit Isoform α(II) and Characterization

of the Type II Enzyme Tetramer: THE α(I) AND α(II) SUBUNITS DO NOT FORM A MIXED α(I)α(II)β2 TETRAMER, Journal of Biological Chemistry, 272 (1997) 17342-17348.

[83] K. Passoja, K. Rautavuoma, L. Ala-Kokko, T. Kosonen, K.I. Kivirikko, Cloning and characterization of a third human lysyl hydroxylase isoform, Proceedings of the National Academy of Sciences, 95 (1998) 10482-10486.

[84] R.S. Gronke, W.J. VanDusen, V.M. Garsky, J.W. Jacobs, M.K. Sardana, A.M. Stern, P.A. Friedman, Aspartyl beta-hydroxylase: in vitro hydroxylation of a synthetic peptide based on the structure of the first growth factor-like domain of human factor IX, Proceedings of the National Academy of Sciences, 86 (1989) 3609-3613.

[85] R.J. Fram, S.L. Mack, M. George, M.G. Marinus, DNA repair mechanisms affecting cytotoxicity by streptozotocin in E. coli, Mutation Research/DNA Repair, 218 (1989) 125-133.

[86] B.J. Chen, P. Carroll, L. Samson, The Escherichia coli AlkB protein protects human cells against alkylation-induced toxicity, Journal of Bacteriology, 176 (1994) 6255-6261.

[87] S.C. Trewick, T.F. Henshaw, R.P. Hausinger, T. Lindahl, B. Sedgwick, Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage, Nature, 419 (2002) 174-178.

[88] M.D. Lloyd, K.D. Merritt, V. Lee, T.J. Sewell, B. Wha-Son, J.E. Baldwin, C.J. Schofield, S.W. Elson, K.H. Baggaley, N.H. Nicholson, Product-substrate engineering by bacteria: Studies on clavaminate synthase, a trifunctional dioxygenase, Tetrahedron, 55 (1999) 10201-10220.

[89] W.-c. Chang, Y. Guo, C. Wang, S.E. Butch, A.C. Rosenzweig, A.K. Boal, C. Krebs, J.M. Bollinger, Mechanism of the C5 Stereoinversion Reaction in the Biosynthesis of Carbapenem Antibiotics, Science, 343 (2014) 1140-1144.

[90] L. Britsch, H. Grisebach, Purification and characterization of (2S)-flavanone 3-hydroxylase from Petunia hybrida, European Journal of Biochemistry, 156 (1986) 569-577.

[91] L. Britsch, Purification and characterization of flavone synthase I, a 2-oxoglutarate-dependent desaturase, Archives of Biochemistry and Biophysics, 282 (1990) 152-160.

[92] F.M. Vaz, R.J.A. Wanders, Carnitine biosynthesis in mammals, Biochemical Journal, 361 (2002) 417-429.

[93] M. Mukherji, C.J. Schofield, A.S. Wierzbicki, G.A. Jansen, R.J.A. Wanders, M.D. Lloyd, The chemical biology of branched-chain lipid metabolism, Progress in Lipid Research, 42 (2003) 359-376.

[94] I. Müller, A. Kahnert, T. Pape, G.M. Sheldrick, W. Meyer-Klaucke, T. Dierks, M. Kertesz, I. Usón, Crystal Structure of the Alkylsulfatase AtsK: Insights into the Catalytic Mechanism of the Fe(II) α-Ketoglutarate-Dependent Dioxygenase Superfamily†,‡, Biochemistry, 43 (2004) 3075-3088.

[95] A. Kahnert, M.A. Kertesz, Characterization of a Sulfur-regulated Oxygenative Alkylsulfatase from Pseudomonas putida S-313, Journal of Biological Chemistry, 275 (2000) 31661-31667.

[96] H.M. Hanauske-Abel, V. Günzler, A stereochemical concept for the catalytic mechanism of prolylhydroxylase: Applicability to classification and design of inhibitors, Journal of Theoretical Biology, 94 (1982) 421-455.

[97] T.D.H. Bugg, Dioxygenase enzymes: catalytic mechanisms and chemical models, Tetrahedron, 59 (2003) 7075-7101.

[98] B.C. Finzel, T.L. Poulos, J. Kraut, Crystal structure of yeast cytochrome c peroxidase refined at 1.7-A resolution, Journal of Biological Chemistry, 259 (1984) 13027-13036.

[99] M. Sivaraja, D.B. Goodin, M. Smith, B.M. Hoffman, Identification by ENDOR of Trp191 as the free-radical site in cytochrome c peroxidase compound ES, Science, 245 (1989) 738-740.

[100] P.J. Riggs-Gelasco, J.C. Price, R.B. Guyer, J.H. Brehm, E.W. Barr, J.M. Bollinger, C. Krebs, EXAFS Spectroscopic Evidence for an FeO Unit in the Fe(IV) Intermediate Observed during Oxygen Activation by Taurine:α-Ketoglutarate Dioxygenase, Journal of the American Chemical Society, 126 (2004) 8108-8109.

[101] S.P. de Visser, Elucidating enzyme mechanism and intrinsic chemical properties of short-lived intermediates in the catalytic cycles of cysteine dioxygenase and taurine/α-ketoglutarate dioxygenase, Coordination Chemistry Reviews, 253 (2009) 754-768.

[102] J.C. Price, E.W. Barr, B. Tirupati, J.M. Bollinger, C. Krebs, The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an α-Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe(IV) Complex in Taurine/α-Ketoglutarate Dioxygenase (TauD) from Escherichia coli†, Biochemistry, 42 (2003) 7497-7508.

[103] M.J. Ryle, R. Padmakumar, R.P. Hausinger, Stopped-Flow Kinetic Analysis of Escherichia coli Taurine/α-Ketoglutarate Dioxygenase: Interactions with α-Ketoglutarate, Taurine, and Oxygen, Biochemistry, 38 (1999) 15278-15286.

[104] M.L. Neidig, C.D. Brown, K.M. Light, D.G. Fujimori, E.M. Nolan, J.C. Price, E.W. Barr, J.M. Bollinger, C. Krebs, C.T. Walsh, E.I. Solomon, CD and MCD of CytC3 and Taurine Dioxygenase: Role of the Facial Triad in α-KG-Dependent Oxygenases, Journal of the American Chemical Society, 129 (2007) 14224-14231.

[105] D.A. Proshlyakov, T.F. Henshaw, G.R. Monterosso, M.J. Ryle, R.P. Hausinger, Direct Detection of Oxygen Intermediates in the Non-Heme Fe Enzyme Taurine/α-Ketoglutarate Dioxygenase, Journal of the American Chemical Society, 126 (2004) 1022-1023.

[106] J.C. Price, E.W. Barr, T.E. Glass, C. Krebs, J.M. Bollinger, Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:α-Ketoglutarate Dioxygenase (TauD), Journal of the American Chemical Society, 125 (2003) 13008-13009.

[107] Y. Xu, M.T. Flavin, Z.-Q. Xu, Preparation of New Wittig Reagents and Their Application to the Synthesis of α,β-Unsaturated Phosphonates, The Journal of Organic Chemistry, 61 (1996) 7697-7701.

[108] D.K. Fisher, T.J. Higgins, A sensitive, high-volume, colorimetric assay for protein phosphatases, Pharmaceutical research, 11 (1994) 759-763.

[109] M. Meurillon, L. Chaloin, C. Périgaud, S. Peyrottes, Synthesis of Pyrimidine-Containing Nucleoside β-(R/S)-Hydroxyphosphonate Analogues, European Journal of Organic Chemistry, 2011 (2011) 3794-3802.

[110] R. Murakami, Y. Fujita, M. Kizuka, T. Kagawa, Y. Muramatsu, S. Miyakoshi, T. Takatsu, M. Inukai, A-102395, a New Inhibitor of Bacterial Translocase I, Produced by Amycolatopsis sp. SANK 60206, J Antibiot, 60 (2007) 690-695.

[111] R. Murakami, Y. Fujita, M. Kizuka, T. Kagawa, Y. Muramatsu, S. Miyakoshi, T. Takatsu, M. Inukai, A-94964, a Novel Inhibitor of Bacterial Translocase I, Produced by Streptomyces sp. SANK 60404, J Antibiot, 61 (2008) 537-544.

[112] N.F. Takahashi S, Kawazoe K, Furukawa Y, Sato S, Tamura C, Naito A., Griseolic acid, an inhibitor of cyclic adenosine 3',5'-monophosphate phosphodiesterase. II. The structure of griseolic acid., The Journal of Antibiotics (Tokyo), 38 (1985) 830-834.

[113] S. Knapp, W.-C. Shieh, C. Jaramillo, R.V. Trilles, S.R. Nandan, Synthesis of the Ezomycin Octosyl Nucleoside, The Journal of Organic Chemistry, 59 (1994) 946-948.

[114] D.A. Hogan, S.R. Smith, E.A. Saari, J. McCracken, R.P. Hausinger, Site-directed Mutagenesis of 2,4-Dichlorophenoxyacetic Acid/α-Ketoglutarate Dioxygenase: IDENTIFICATION OF RESIDUES INVOLVED IN METALLOCENTER FORMATION AND SUBSTRATE BINDING, Journal of Biological Chemistry, 275 (2000) 12400-12409.

[115] R. Lukačin, L. Britsch, Identification of Strictly Conserved Histidine and Arginine Residues as Part of the Active Site in Petunia hybrida Flavanone 3β-Hydroxylase, European Journal of Biochemistry, 249 (1997) 748-757.

[116] J.E. Coleman, Zinc enzymes, Current Opinion in Chemical Biology, 2 (1998) 222-234.

[117] P.R. Rajagopalan, X.C. Yu, D. Pei, Peptide deformylase: a new type of mononuclear iron protein, Journal of the American Chemical Society, 119 (1997) 12418-12419.

[118] C.P. Pang, B. Chakravarti, R.M. Adlington, H.H. Ting, R.L. White, G.S. Jayatilake, J.E. Baldwin, E.P. Abraham, Purification of isopenicillin N synthetase, Biochemical Journal, 222 (1984) 789-795.

[119] P.L. Roach, I.J. Clifton, C.M.H. Hensgens, N. Shibata, C.J. Schofield, J. Hajdu, J.E. Baldwin, Structure of isopenicillinN synthase complexed with substrate and the mechanism ofpenicillin formation, Nature, 387 (1997) 827-830.

[120] J.M. Elkins, M.J. Ryle, I.J. Clifton, J.C. Dunning Hotopp, J.S. Lloyd, N.I. Burzlaff, J.E. Baldwin, R.P. Hausinger, P.L. Roach, X-ray Crystal Structure of Escherichia coli Taurine/α-Ketoglutarate Dioxygenase Complexed to Ferrous Iron and Substrates†,‡, Biochemistry, 41 (2002) 5185-5192.

[121] L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey, T. Schwede, Protein structure homology modeling using SWISS-MODEL workspace, Nature protocols, 4 (2008) 1-13.
[122] Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.3r1, in, 2010.
[123] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, Journal of computational chemistry, 31 (2010) 455-461.
[124] A. Kahnert, M.A. Kertesz, Characterization of a Sulfur-regulated Oxygenative Alkylsulfatase from Pseudomonas putida S-313, Journal of Biological Chemistry, 275 (2000) 31661-31667.

VITA

EDUCATION

Bachelor of Technology, Department of Biotechnology (2004-08)
Heritage Institute of Technology, WBUT, Kolkata, India
Cumulative GPA: 9.02/10

PROFESSIONAL EXPERIENCE

Research Assistant, University of Kentucky, KY (2011 – Current)

Teaching Assistant, PHS 914 Fall 2008; Accommodations coordinator, Spring 2009; Department of Pharmaceutical Sciences, University of Kentucky

Research intern, Indian Institute of Chemical Biology, Kolkata, India (Feb 2009- Aug 2009)

HONORS

August 2013, Awarded the Graduate Student Academic Year Fellowship, University of Kentucky

PUBLICATIONS

1. S. Barnard-Britson, X. Chi, K. Nonaka, AP Spork, N. Tibrewal, **A. Goswami**, P. Pahari, C. Ducho, J. Rohr and SG Van Lanen; "Amalgamation of Nucleosides and Amino Acids in Antibiotic Biosynthesis: Discovery of an L-Threonine:Uridine-5'-Aldehyde Transaldolase", *J. Am. Chem. Soc.,* 2012, **134** (45), pp 18514–18517

2. **Anwesha Goswami** and Steven G. Van Lanen, "Enzymatic strategies and biocatalysts for amide bond formation: Tricks of the trade outside of the ribosome"**,** *Molecular BioSystems*, 2014, DOI: 10.1039/C4MB00627E

3. **Anwesha Goswami**, Wenlong Cai, Xiaodong Liu, Thomas P. Wyche, Maïa Meurillon, Suzanne Peyrottes**,** Tim S. Bugni, Jurgen Rohr, Steven G. Van Lanen, "Mechanism of

the non-heme Fe(II), α-ketoglutarate:uridine-5'-monophosphate dioxygenase involved in nucleoside antibiotic biosynthesis", *J. Am. Chem. Soc., Submitted*

4. Wenlong Cai, **Anwesha Goswami**, Zhaoyong Yang, Sandra Barnard-Britson, Koichi Nonaka, Manjula Sunkara, Andrew J. Morris, Keith Green, Sylvie Garneau-Tsodikova, Jon S. Thorson, and Steven G. Van Lanen, "The biosynthesis of capuramycin antibiotics: insight into uridine-5'-carboxamide assembly and characterization of the A-102395 biosynthetic gene cluster", *ACS Chemical Biology, In preparation*

5. Xiaodong Liu**, Anwesha Goswami**, Wenlong Cai, Bertolt Gust, and Steven G. Van Lanen, **"**Biosynthesis of Nucleoside Antibiotics: Characterization of Gene Clusters and Functional and Mechanistic Insight of Key Enzymes", *Nat. Prod. Rep., In preparation*

Typed name of student on final copy:

Anwesha Goswami