Studies of 3-Deoxy-D-\textit{m}anno-\textit{O}ctulosonate 8-Phosphate Phosphatase: Mechanistic Insights and a Gene Fusion Example

by

Li Yi

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Chemistry) in The University of Michigan 2009

Doctoral Committee:

Professor Ronald W. Woodard, Chair
Professor Ayyalusamy Ramamoorthy
Assistant Professor Kristina I. Hakansson
Assistant Professor Bruce Allan Palfey
To my adoptive parents,
Shaorong Shi and Yisu Huang,
and my late mother,
Shaolan Shi
Acknowledgements

I would particularly like to thank my advisor Dr. Ronald Woodard for his time and efforts putting into my doctoral education. It is under his guidance that I become confident and capable in research and future career. I would like to thank my committee members for their helpful suggestions and encouragement. I would also like to thank my collaborators, Dr. Oleg Tsodikov and Dr. Tapan Biswas. They have given me a great vision of protein crystallography and inspired my interest in this field. Thank former and present group members in the Woodard group for their intellectual input as well as personal support. They have been great teachers, colleagues, and friends.

Special thanks to my adoptive parents who took care of me as their own child with all their heart, cultured me, and encouraged me to be a scientist. Thank my husband Song Ge who walked me through these years. Finally, thank all other family members and friends. Without your love and support, I can not be standing here today as whom I am.
# Table of Contents

Dedication........................................................................................................ii

Acknowledgements..........................................................................................iii

List of Figures...................................................................................................vi

List of Tables.....................................................................................................ix

List of Appendices.............................................................................................x

Chapter

1. Introduction.....................................................................................................1
   1.1 Bacterial Pathogenesis...........................................................................1
   1.2 KdsC.................................................................................................5
   1.3 KdsB.................................................................................................9
   1.4 Dissertation Rationale..........................................................................11
   1.5 Reference.........................................................................................12

2. Crystal Structures of KdsC and its Variant.................................................17
   2.1 Summary...........................................................................................17
   2.2 Introduction.......................................................................................18
   2.3 Experimental Procedures....................................................................22
   2.4 Results..............................................................................................29
   2.5 Discussion..........................................................................................37
   2.6 Future Directions...............................................................................40
   2.7 Acknowledgements............................................................................40
   2.8 Reference..........................................................................................42
List of Figures

Figure

1-1  Gram-negative envelope .................................................................2
1-2  KDO₂-Lipid A .................................................................................3
1-3  The KDO biosynthetic pathway .........................................................4
1-4  Sequence Alignment of HADSF enzymes ..........................................6
1-5  Proposed catalytic mechanism of KdsC ..............................................7
1-6  Anomeric isomers of KDO₈P, KDO₈P analogues and KDO .....................8
1-7  Crystal structure of KdsC from H. influenzae .....................................8
1-8  Crystal structure of KdsB from A. aeolicus ......................................10
2-1  Structure of HADSF family members ..............................................19
2-2  Sequence alignment of KdsC homologues .......................................29
2-3  SDS-PAGE of purified wild-type KdsC<sub>Ec</sub> and KdsCΔC₈ ................30
2-4  A schematic of KdsC structures ......................................................30
2-5  The tetrameric structure of KdsC in the top view and the side view .......31
2-6  The 2F₀-F<sub>c</sub> electron density map of the active site captured in the 1.4 Å KdsC structure, contoured at 1 σ .........................................................32
2-7  Conformational differences between the open state and the closed state of the active site .................................................................33
2-8  a) Specific contacts between the active sites residues and the tail residues of the Mg²⁺-bound KdsC<sub>Ec</sub>. b) Interactions between the active site residues and the catalytic products KDO and phosphate in KdsCΔC₈ .....................34
2-9  Scheme of specific interactions of KDO, P<sub>i</sub> and Mg²⁺ in the KdsC<sub>Ec</sub> active site ..35
2-10  Omit F_o-F_c electron density map of the active site contoured at 2.3σ………………36
3-1  Comparison of the C-terminal structure of KdsC from *Escherichia coli*,
*Haemophilus influenzae*, and *Aquifex aeolicus* .........................................46
3-2  Multiple sequence alignment of KdsC from various Gram-negative organisms...56
3-3  Phylogenetic tree of KdsC from selective microorganisms ..............................57
3-4  Saturation curve of KdsCΔC8 plotted with initial velocity vs. concentration of
KDO8P ........................................................................................................60
3-5  Salt effect of KdsC ......................................................................................62
3-6  Model of KdsC catalytic cycle .....................................................................66
4-1  Organization of *kds* and *kps* gene clusters ...............................................74
4-2  Unrooted phylogenetic tree of KdsB/KpsU and KdsD/KpsF ............................75
4-3  Scheme of sequential reactions catalyzed by KdsC and KdsB ......................77
4-4  Scheme of reactions in EK color development and the MG assay ..............79
4-5  Two proposed domains in PpBC and the location of artificial division of PpB
and PpC ........................................................................................................80
4-6  Calibration curves of malachite green assay and eikonogen assay in
conventional tube-based set-up..................................................................91
4-7  Calibration curves of malachite green assay and eikonogen assay in
96-well plate format .................................................................................93
4-8  MG/EK measurement of P_i and PP_i mixture at a ratio of 1:1 and deconvolution
of the signal ............................................................................................94
4-9  Sequence alignment of two bifunctional enzymes and several individual
KdsCs and KdsBs from G- bacteria strains by ClustalW...............................96
4-10 SDS-PAGE of purified PpBC, PpB and PpC.............................................97
4-11 pH dependence of KdsC and KdsB activities..............................................99
4-12 a) Temperature profiles of PpBC and PpB. b) Thermostability of PpBC......100
4-13 Magnesium dependence of PpB and PpBC............................................102
4-14  Metal requirement of PpC, PpB, and PpBC

4-15  Time course of coupled reactions catalyzed by PpBC or a mixture of PpB and PpC
List of Tables

Table
1-1 Summary of properties of recombinant KdsC…………………………………….5
1-2 Comparison of kinetic parameters of E. coli KdsB and KpsU…………………..10
2-1 Crystal growth and cryo-stabilization conditions………………………………….27
2-2 Data collection and refinement statistics ………………………………………28
3-1 Molecular weight of KdsCΔC8 and KdsCΔa determined from gel filtration……59
3-2 Michaelis-Menten parameters of KdsCs…………………………………………61
4-1 List of all microorganisms possessing the kds gene cluster…………………….72
4-2 Comparison of each kds gene from the kps/kds gene cluster in P. putida F1 to its homologue(s) in E. coli CFT073 via pairwise sequence alignment ………….75
4-3 Comparison of the two sets of KDO biosynthetic enzymes from Pseudomonas putida F1 via pairwise sequence alignment…………………………………….76
4-4 Extinction coefficients (ε) of P, and PP, for the MG or EK assay ………………..92
4-5 Comparison of the pyrophosphate reagent to the MG/EK measurements………95
4-6 Substrate specificity of PpBC……………………………………………………101
4-7 Steady state kinetic parameter measurements of PpBC, PpC, PpB……………..104
List of Appendices

Appendix

A. Attempts to Express S-Adenosyl-L-Methionine Hydrolase

B. Studies on Quorum Sensing Autoinducer Synthase:
   A S-Adenosyl-L-Methionine Lactonizing Enzyme
Chapter 1

Introduction

1.1 Bacterial Pathogenesis

Microbial infection continues to be one of the major problems of human health today, therefore efforts in antimicrobial drug development have continued unabated. One challenge of drug development is that some pathogenic bacterial strains mutate frequently to generate antimicrobial resistance, outdistancing drug discovery [1, 2]. The resistance may be generated under pressure of misuse of antibiotics or inappropriate operations in hospitals [3, 4]. The increase of resistance contributes to an incessant demand for new antibiotics [5].

Strategies combating microbial resistance include targeting the resistance mechanisms and finding new targets using genomics [6]. Genes that are essential for bacterial cell virulence or survival have long been drugable targets, and new candidates are under development. A classic example of a drugable target is the cell-wall biosynthetic pathway that produces the protective peptidoglycan layer. The transpeptidase in the pathway can be inhibited by the substrate-like β-lactam-containing molecules, thus preventing peptidoglycan from cross-linking to a protective wall in the bacteria [7]. However, these drugs are more effective against Gram-positive bacteria (G+)
than Gram-negative cells (G-), which differ from G+ by possessing an outer membrane outside the peptidoglycan wall [8]. Since many pathogenic G- strains cause more severe or even fatal diseases in human [9-11], antibiotics developed to specifically target the outer membrane biosynthesis are urgently needed.

Figure 1-1 Gram-negative envelope [8]

The asymmetric outer membrane of G- bacteria (Figure 1-1) consists of phospholipids on the inner leaflet and lipid A on the outer leaflet. Lipid A, polysaccharide core, and the outreaching O-antigen region constitute the lipopolysaccharide layer (LPS, also known as endotoxin). Lipid A is a phosphorylated N-acetylglucosamine dimer connected by a β(1→6) linkage with 6 or 7 fatty acids attached [12]. It is responsible for the toxicity of LPS [13]. The polysaccharide chains are linked to lipid A through an eight carbon carbohydrate molecule, 3-deoxy-D-manno-octulosonate (KDO). Two molecules of α(2→6)-linked KDO are required to incorporate
into LPS for proper *E. coli* cell growth (**Figure 1-2**) [14]. The number of KDO required in the LPS can vary from species to species. For instance, *Bordetella pertussis* [15] and *Haemophilus influenzae* [16] only incorporate one KDO molecule, while *Chlamydia trachomatis* [17] utilizes three KDO molecules as the linker between lipid A and polysaccharide. At the outreaching end of polysaccharide chains is a region termed O-antigen which consists of repeating oligosaccharides and determines the antigenic specificity of bacterial strains [8]. Although the polysaccharide chain and O-antigen region adopt a wide range of carbohydrate molecules, KDO is universally present between lipid A and the polysaccharide chains. Failure to produce KDO in bacterial cells may result in an accumulation of lipid A precursor and cell growth arrest [18]. Cells interrupted in KDO biosynthesis are also less pathogenic and more labile to antibiotics [19, 20]. Since KDO is only found in the outer membrane of G- bacteria, pectins of higher plants [21-24] and cell walls of some green algae [25], it is a promising potential antimicrobial target with high selectivity.

**Figure 1-2** KDO$_2$-Lipid A
Four sequential enzymatic steps (Figure 1-3) are involved in the KDO biosynthetic pathway: 1) the isomerization of D-ribulose 5-phosphate (Ru5P) to form D-arabinose 5-phosphate (A5P) by KdsD, 2) the condensation of A5P and phosphoenolpyruvate (PEP) to form 3-deoxy-D-manno-octulosonate 8-phosphate (KDO8P) by KdsA, 3) the hydrolysis of the phosphate ester KDO8P catalyzed by KdsC; and 4) the activation of KDO via addition of a CMP portion of CTP by KdsB. Finally the KDO moiety from the activated sugar nucleotide is transferred onto lipid A by a membrane-bound transferase, WaaA. These enzymes are all essential for *E. coli* cell survival (GenoBase, http://ecoli.aist-nara.ac.jp/). But in the cases where gene redundancy exists, for example, in *E. coli* where KdsD and GutQ share the same A5P isomerase activity, both the isozyme genes need to be knocked out to arrest cell growth [26].

![Figure 1-3](image-url)

**Figure 1-3.** The KDO biosynthetic pathway. (1) KdsD: D-Arabinose 5-phosphate isomerase [EC: 5.3.1.31], (2) KdsA: KDO 8-phosphate synthase [EC: 2.5.1.55], (3) KdsC: KDO 8-phosphate phosphatase [EC: 3.1.3.45], (4) KdsB: KDO cytidylyltransferase [EC: 2.7.7.38] (5) WaaA: KDO transferase [EC: 2.4... ]
1.2 KdsC

The 3-deoxy-d-manno-octulosonate 8-phosphate phosphatase, KdsC [EC: 3.1.3.45], is the third enzyme in the KDO biosynthetic pathway. This enzyme catalyzes the hydrolysis of KDO8P to form a molecule of KDO and an inorganic phosphate. Ghalambor and Heath first suggested the existence of a KDO8P phosphatase that was required in the KDO pathway, based on the inability of KdsB to produce activated KDO directly from KDO8P [27]. In 1975, the phosphatase activity specific to hydrolyze KDO8P but not A5P or p-nitrophenylphosphate was isolated from *E. coli* crude extract [28]. The wild-type KdsC was first purified and characterized by Ray and Benedict in 1980 [29, 30]. However, the gene encoding this enzyme was not identified until 2003 [31]. The recombinant enzyme was characterized and revealed to share similar properties with the wild-type enzyme, including metal requirement, substrate specificity, pH optimum, kinetic parameters, etc. (Table 1-1).

**Table 1-1.** Summary of properties of recombinant KdsC as reported by Wu [31] and wild-type KdsC as reported by Ray and Benedict [29].

<table>
<thead>
<tr>
<th>Property</th>
<th>Recombinant KdsC [31]</th>
<th>Wild-type KdsC [29]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (µM)</td>
<td>75</td>
<td>91</td>
</tr>
<tr>
<td>$V_{max}$ (units/mg)</td>
<td>500</td>
<td>480</td>
</tr>
<tr>
<td>Alternate substrate</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Metal requirement</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>pH optimum</td>
<td>5.5-7.0</td>
<td>5.5-6.5</td>
</tr>
<tr>
<td>pI</td>
<td>4.6</td>
<td>4.7</td>
</tr>
<tr>
<td>MW (kDa)</td>
<td>89 (gel filtration)</td>
<td>80 (gel filtration)</td>
</tr>
<tr>
<td></td>
<td>23 (SDS-PAGE)</td>
<td>40-43 (SDS-PAGE)</td>
</tr>
<tr>
<td></td>
<td>19.881 (mass spectroscopy)</td>
<td></td>
</tr>
</tbody>
</table>
Sequence analysis of KdsC revealed that it belongs to the Haloacid Dehalogenase Superfamily (HADSF) [31]. The HADSF is one of the largest and most ubiquitous superfamilies, including varieties of enzymes catalyzing mostly phosphoryl transfer reactions [32]. Four sequence motifs are conserved across the superfamily (Figure 1-4): I) Dx Dx V/T, II) T/Sxx, III) K, and IV) GDxxxD Motifs [32]. Besides these four motifs, KdsC also contains a highly conserved signature sequence near the C-terminal region, G G x G A x R E, which is not universally conserved in other HADSF enzymes [31]. This signature sequence was recently reported to be shared in a close KdsC homolog, 2-keto-3-deoxy-D-glycero-D-galacto-9-phospho-nononic acid (KDN9P) phosphatase [33].

Though the overall sequence homologies between the members of HADSF are low (<15%) [34], the structural similarities are high, since a mixed α/β Rossmannoid fold is commonly adopted in all the members [32]. In addition, most of the enzymes from this

![Sequence Alignment of HADSF enzymes. KdsC: 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase; PSP: phosphoserine phosphatase; BPGM: β-phosphoglucomutase; SPP: sucrose phosphate phosphatase; PMM: phosphomannomutase.](image_url)
superfamily require divalent metal ions as cofactors, in many cases Mg$^{2+}$ [35]. All these similarities suggest that enzymes from HADSF are possibly evolved from a common ancestor [36].

**Figure 1-5** Proposed catalytic mechanism of KdsC. Residues were assigned for *E. coli* KdsC. $\text{R} = \text{KDO}$ in the $\alpha$-pyranose form.

The generally accepted mechanism for phosphatases and ATPases in HADSF is proposed as two sequential $\text{S}_\text{N}2$-like reactions [32, 35]. Specifically for KdsC ([Figure 1-5](#)), upon KDO8P binding, an active site aspartate side chain attacks the phosphorus and releases the sugar moiety, forming an aspartylphosphate intermediate (mixed anhydride). Then a second nucleophilic attack on the mixed anhydride intermediate by an activated water molecule releases an inorganic phosphate and regenerates the native enzyme. The native substrate form of KDO8P is proposed to be $\alpha$-pyranose anomer of KDO ([Figure 1-6a](#)) by Baasov and Jakob [37]. The $\alpha$-anomeric 2-deoxy KDO8P analogue ([Figure 1-6b](#)) retains 50% of the activity of the natural KDO8P, while the $\beta$-anomeric 2-deoxy KDO8P analogue ([Figure 1-6c](#)) has no observable hydrolytic activity [37]. Two aspartates in sequence motif I were demonstrated to be essential for enzyme activity via
mutational studies by Aggarwal [38]. One of these two aspartates should play the role as nucleophile and the other as the water-activator.

**Figure 1-6** Anomeric isomers of KDO8P, KDO8P analogues and KDO. **a)** α-KDO8P-pyranose; **b)** 2-deoxy analogue of α-KDO8P-pyranose; **c)** 2-deoxy analogue of β-KDO8P-pyranose; **d)** β-KDO-pyranose.

**Figure 1-7** Crystal structure of KdsC from *H. influenzae*. The cobalt atoms are shown as blue spheres. [39]

Two crystal structures have been determined for KdsC to date, one from *Haemophilus influenzae* (PDB ID 1J8D) [39] and the second from a hyperthermal organism, *Aquifex aeolicus* (PDB ID 2P9J). Neither of these structures included substrate, products, or the physiological metal magnesium [29, 31] in the active site. KdsC in both
crystal structures formed tetramers through interactions between adjacent short β-hairpins, each donated from one monomer. Each monomer folded in a basic Rossmannoid core (Figure 1-7), without any cap domains as observed in some other HADSF members [34, 40, 41]. A hexacoordination of the active site divalent metal Co\(^{2+}\) was observed in the *H. influenzae* KdsC cocrystallized with Co\(^{2+}\).

1.3 KdsB

The enzyme immediately following KdsC in the KDO biosynthetic pathway is the CMP-KDO synthetase, KdsB [EC. 2.7.7.38], also known as KDO cytidylyltransferase. This enzyme catalyzes the reaction between KDO and CTP to transfers a CMP portion onto the KDO moiety, thus making the activated form of KDO that can be later transferred onto lipid A. It was initially isolated and studied by Ghalambor and Heath [27], and further characterized by Ray and Benedict [42, 43]. *E. coli* KdsB is specific to the β-pyranose form of KDO (Figure 1-6d), as reported in a \(^{13}\)C NMR study by Kohlbrenner and Fesik [44]. KDO analogues such as 2-deoxy and 5-deoxy-5-fluoro derivatives of KDO were synthesized to study the mechanism of KdsB [45, 46]. CMP-β-KDO is reported by Sugai et al. to be highly unstable with a half-life of 34 min at 25°C [46]. Lin et al. revealed the molecular basis of the fast hydrolysis of CMP-β-KDO which adopts a twist-boat conformation in the transition state [47]. Due to the surprisingly short half-life, this compound has never been synthesized chemically or isolated from the KdsB reaction mixture. Crystal structures of KdsB from *H. influenzae* [48] and *A. aeolicus* [49] (Figure 1-8) were solved, both revealing a dimeric organization.
A close homologue of KdsB that shares ~40% sequence identity with KdsB is a protein named KpsU. KpsU was reported to be correlated to capsular expression in encapsulated *E. coli* strains [50]. It catalyzes the same reaction as KdsB but with distinctly different kinetic properties (Table 1-2) [51, 52]. The crystallographic studies of *E. coli* KpsU by Jelakovic and Schulz also revealed a dimeric structure and a highly similar active site to that of KdsB [53, 54]. The capsule is composed of extracellular polysaccharide enclosing the bacterium but still attached to the cell. Having KDO biosynthetic genes in the capsular-related gene cluster is not surprising, since KDO contributes 40-60% of the capsular polysaccharide mass in some species as opposed to 2-5% of LPS mass [55]. However, to date, only homologues of KdsB (named KpsU) and KdsD (named KdsF) were found associated with capsular expression.

**Table 1-2** Comparison of kinetic parameters of *E. coli* KdsB and KpsU.

<table>
<thead>
<tr>
<th></th>
<th>$K_M^{KDO}$ (mM)</th>
<th>$K_M^{CTP}$ (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_M^{KDO}$ (mM$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KdsB$^{Ec}$ [43]</td>
<td>0.29 ± 0.02</td>
<td>0.20</td>
<td>279 ± 9</td>
<td>9.6 x 10$^2$</td>
</tr>
<tr>
<td>KpsU$^{Ec}$ [52]</td>
<td>2.0 ± 0.4</td>
<td>2.5 ± 0.5</td>
<td>1.35 ± 0.27</td>
<td>0.7</td>
</tr>
</tbody>
</table>

*Figure 1-8* Crystal structure of KdsB from *A. aeolicus*. [49]
1.4 Dissertation Rationale

The ultimate goals of this dissertation are to gain mechanistic insights into KdsC to guide antibiotic design, and to understand how KdsC may interact with downstream enzymes in the pathway. A wide range of techniques and methods were used to achieve these goals, including molecular biology, structural biology, enzymology, biochemistry, and analytical chemistry.

In Chapter 2, a set of seven crystal structures are reported for *E. coli* KdsC and its C-terminal “tail-less” variant. For the first time, products of KdsC, KDO and inorganic phosphate, are successfully trapped in the active site. Contacts formed between the products and the active site residues help define the substrate specificity. A structural element that had undergone dramatic conformational change is identified from these crystal structures and hypothesized to have an impact on KdsC’s catalytic efficiency. This hypothesis was tested by biochemical characterization of an *E. coli* KdsC C-terminal tail deletion mutant as well as KdsCs from other microorganisms which are naturally tail-less KdsCs, and these results are in Chapter 3. Based on both structural information and biochemical characterization, a model of KdsC catalytic cycle has finally been proposed.

A natively occurring gene fusion of *kdsC* and *kdsB* was discovered from database mining as the first example of gene fusion in the KDO biosynthetic pathway. The full characterization of this fusion gene product is reported in Chapter 4. A coupled colorimetric assay was developed to measure both phosphate and pyrophosphate amount independently and simultaneously in a reaction mixture to assist in the kinetic analysis of this unique fusion protein.
1.5 References


Chapter 2

Crystal Structures of KdsC and Its Variant

2.1 Summary

While all four enzymes in the 3-deoxy-D-manno-octulosonate (KDO) biosynthetic pathway have been studied for activities, only recently has the gene of 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (kdsC) been identified. Although the protein product of the kdsC gene has been previously characterized biochemically and structures of the protein from two microorganisms have been reported, the mechanistic details of KdsC remain unclear. KdsC is a member of the haloacid dehalogenase superfamily (HADSF). This phosphatase lacks an essential structural element, a cap domain, which is observed in many HADSF members. When present, the cap domain is responsible for specific substrate recognition, but nonetheless KdsC is highly specific for its substrate, KDO 8-phosphate. To better understand this unexpected substrate specificity, studies of the KdsC active site in the presence and absence of the substrate or products were pursued. From the crystal structures, a dramatic conformational change was observed for the C-terminal tail region of E. coli KdsC. Based on this observation, a tail deletion mutant was constructed and studied. The catalytic products of the KdsC reaction were observed in
the structures of the tail deletion mutant. In total, seven crystal structures of _E. coli_ KdsC and its tail-less variant with different metal cofactors, soaked or not with the substrate, were obtained. The detailed structural information revealed from my study demonstrates an important role for tetramerization in defining the substrate specificity, and predicts a potential regulatory role of the C-terminal tail region in the catalytic cycle.

2.2 Introduction

The haloacid dehalogenase superfamily (HADSF) is one of the largest superfamilies discovered to date [1]. Members in this superfamily share a similar catalytic scaffold that evolved to catalyze mainly the phosphoryl-transfer reactions, except for the namesake enzyme [2]. The interest in the HADSF is increasing as more enzymes from this superfamily with important biological functions are discovered, including the biosynthesis of essential metabolic molecules [3-5], transport [6], signal transduction [7, 8], transcription [9] and DNA repair [10]. However, most of the HADSF members remain uncharacterized.

Sequence alignment of the HADSF members suggests four universally conserved motifs: I) DxDxV/T, II) T/Sxx, III) K, and IV) GDxxxD [1]. Residues in these motifs line up in the active site to participate in catalysis or substrate binding [1, 4, 6, 11]. Sequence insertions are observed for some HADSF enzymes at either of two locations. These insertions form a separate domain extended from the basic α/β-hydrolase core (Rossmannoid fold), and are termed the cap domain. The cap
domain can move dynamically in the solvent to either regulate the access to the active site, or form a roof over the active site (Figure 2-1) [2]. The HADSF members are thereby classified into three groups based on the absence and presence, as well as the location of the cap domain [1]. The C0 group members with no sequence insertion fold into a single-domain Rossmannoid core. The C1 group, represented by phosphoserine phosphatase and β-phosphoglucomutase, has a sequence insertion between α/β strand 1 and strand 2 that forms a cap. The C2 group, represented by sucrose-6-phosphate phosphatase and phosphomannomutase, has a cap domain formed from the sequence insertion immediately after strand 3. The classification of KdsC as a HADSF was due to the sequence conservation at the four characteristic motifs [12]. The absence of sequence insertions as well as the basic Rossmannoid fold structure without cap [13] further assigned this enzyme to the C0 group of HADSF.

Figure 2-1 Structure of HADSF family members. C1: phosphoserine phosphatase (1F5S); C2: phosphatase TM0651 from Thermotoga maritime (1NF2); C0: KdsC (1J8D). [14]
Some members of HADSF display broad substrate specificities, while other members have narrow substrate specificities [15]. A correlation between substrate specificity of the enzyme and the presence of the cap domain has been derived and seems due to the substrate specific contacts provided by the cap domain [14]. C0 group enzymes usually utilize bulky substrates [10] or adopt a wide range of substrates [16] with the active sites exposed to the solvent. However, KdsC, the archetypical C0 group enzyme, displays a very narrow substrate specificity despite the absence of the cap domain [12]. The origin of KdsC’s substrate specificity remains elusive.

To date, two crystal structures of KdsC from *Haemophilus influenzae* (KdsC\(^{Hi}\), PDB ID 1J8D) [13] and *Aquifex aeolicus* (KdsC\(^{Aa}\), PDB file 2P9J) have been solved. Both structures revealed a tetramer of KdsC stabilized by interactions of short \(\beta\)-hairpins inserted between strand 1 and strand 2. Each monomer seems to structurally resemble a cap domain over the active site of its adjacent monomer. Specific contacts with the substrate, KDO8P, or the products, KDO and Pi, could not be deduced since neither reported structure contained the substrate nor the products. The location of active site was marked with a hexacoordinated Co\(^{2+}\) in the structure of KdsC\(^{Hi}\), where all the conserved residues in the four motifs are also present. For *E. coli* KdsC (KdsC\(^{Ec}\)), the two aspartates in motif I, D32 and D34, were demonstrated to be essential for activity through mutational studies [17]. But the details of the active site can not be obtained without cocrystallized substrates or products in the structure.
The high catalytic efficiency [12] and the lack of any strong-binding inhibitor of KdsC\textsubscript{Ec} make it challenging to capture the substrate KDO8P in the active site. Several strategies can be considered to slow down the reaction rate, therefore achieving substrate retention in the active site. One possible strategy involves using KdsC from \textit{A. aeolicus} (KdsC\textsubscript{Aa}), a hyperthermophilic bacteria isolated from a hot spring, to represent its mesophilic homolog. Catalytic rates of enzymes from this bacterial strain are significantly lower at 4°C given the optimum living temperature is 95°C [18]. However, KdsC\textsubscript{Aa} was found to be less thermally stable than other enzymes from \textit{A. aeolicus} since it degraded above 60°C, therefore indicating this strategy may not apply [17]. An alternate strategy is to use Ca\textsuperscript{2+} as the cofactor for KdsC\textsubscript{Ec} instead of Mg\textsuperscript{2+}, thereby lowering the activity 50 fold [12].

During our initial efforts to obtain a crystal structure of substrate- or products-bound KdsC\textsubscript{Ec}, we observed an unusual substrate-like insertion of the C-terminal eight-amino acid region of one monomer into the neighboring monomer’s active site thus forming several specific contacts with active site residues. This is a unique phenomenon that has not been reported for any HADSF protein structure. A deletion mutant removing the eight amino acids from the C-terminus (KdsC\Delta8) was constructed and utilized for crystallography. The catalytic products, KDO and inorganic phosphate (P\textsubscript{i}), were successfully captured in the active site, enabling the identification of the basis for specific substrate recognition. Overall, seven crystal structures of KdsC\textsubscript{Ec} and KdsC\Delta8 with different combinations of divalent metals and catalytic products were determined.
2.3 Experimental Procedures

Materials – Tris(hydroxymethyl)aminomethane was obtained from Research Organics. Phosphoenolpyruvate mono(cyclohexylammonium) salt, thiobarbituric acid, D-arabinose 5-phosphate, sodium arsenate, and bovine albumin serum (BSA) were purchased from Sigma-Aldrich Company. Bacto\textsuperscript{TM} tryptone was from Becton, Dickinson and Company. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Gold Biotechnology, Inc. Potassium chloride, sodium chloride, yeast extract, ampicillin sodium salt, and ammonium sulfate were from Fisher Scientific Inc. Plasmid extraction was performed using Wizard\textsuperscript{®} Plus SV Minipreps DNA purification kit from Promega Corporation. DNA sequencing was conducted by the University of Michigan Biomedical Resources Core Facility. BL21(DE3) E. coli strain was from Novagen, EMD Chemicals Inc. Bradford Protein Assay Reagent and AG-MPI resin were from Bio-Rad Laboratories, Inc. The Q-sepharose resin, Phenyl Superose (HR 10/10) column and FPLC\textsuperscript{®} system were from Pharmacia, GE Healthcare. The 3-deoxy-D-manno- octulosonate 8-phosphate synthase (KdsA) from E. coli was purified by Jingjing Li in the Woodard laboratory. High grade spectra/Por\textsuperscript{®} 7 dialysis tubing (10,000 Da molecular weight cut-off and metal free) was from VWR International, LLC. The Millex\textsuperscript{®} syringe driven filter units (0.22 μm) was obtained from Millipore.

Sequence analysis – Database searches of multiple microbial organisms were performed utilizing the BLAST program at the NCBI website (http://ncbi.nlm.nih.gov
Multiple sequence alignments were generated using Clustal W (http://www.ebi.ac.uk/clustalw).

**Protein Concentration Assay** – Protein concentration was determined using the Bradford assay by using BSA as a standard and mixing 1 mL dye to 10 µL proteins. Alternately, ultra-violet absorbance of a protein solution measured at 280 nm can be converted into protein concentration, based on the extinction coefficient calculated from ProtParam tool (ExPASy).

**Expression and purification of KdsC\(^{Ec}\)** – The pT7-7-\(kdsC^{Ec}\) plasmid was constructed previously. The protein was expressed and purified as described [12]. Briefly, the sequenced plasmid was transformed into chemically competent *E. coli* BL21(DE3) cells for expression. The cells were grown in 1L LB media supplemented with 100 mg/L ampicillin at 37°C and induced with IPTG at a final concentration of 0.2 mM when OD\(_{600}\) reached 0.8. The cells were allowed to grow at 17°C after induction for another 18 h before being harvested by centrifugation at 6000 rpm in a Beckman JA-10 rotor for 10 min. The cell paste was subjected to sonication in buffer A (20 mM Tris, pH 7.5) with four pulses of 30 s each with intervals of 2 min in between, and clarified by centrifuging at 29,000 g for 30 min. The resultant supernatant was passed through Q-sepharose column with a linear gradient elution of 0-2M KCl in 90 min at 1 mL/min. Fractions containing KdsC\(^{Ec}\) as identified by Bradford assay and SDS-PAGE were eluted at around 0.4-0.6 M KCl, combined, and ammonium sulfate was added to 20% (w/v). The protein solution was then loaded on a phenyl superpose column (HR 10/10), and eluted with a reverse gradient of
ammonium sulfate from 20% to 0% (w/v). Homogenous KdsC\textsuperscript{Ec} was collected in fractions with 0% ammonia sulfate as identified by SDS-PAGE. The pooled protein was dialyzed against buffer A, concentrated to 2.7 mg/mL, and stored at -80°C. The total yield was 20 mg/L of cell culture. Proteins were further purified on an S-200 gel filtration column equilibrated in 20 mM HEPES pH 7.2, and finally concentrated to 20 mg/mL using Ultra-15 centrifugal filter unit (Millipore) for crystallization.

**Expression and purification of KdsC\textDelta 8** – The coding sequence for KdsC\textDelta 8 was constructed by inserting a stop codon in pT7-7-kdsC\textsuperscript{Ec} after amino acid residue 180 using QuickChange Mutagenesis kit. The plasmid pT7-7-KdsC\textDelta 8 was transformed into chemically competent \textit{E. coli} BL21(DE3) cells for expression. Protein was overexpressed and purified according to the same procedure as described for KdsC\textsuperscript{Ec}. Cells were broken in buffer A (20 mM Tris pH 7.5) by sonication and centrifuged at 6000 rpm for 10 min in a Beckman JA-10.5 rotor. The supernatant was loaded on Q-sepharose column and eluted in fractions with 0.3 ~ 0.6 M KCl. Fractions containing KdsC were treated with 25% ammonium sulfate and applied on a phenyl-superose column (HR 10/10). Protein eluted at 4% of ammonia sulfate was pooled, dialyzed against buffer A, and concentrated to 2.2 mg/mL. A typical yield was 20 mg/L of cell culture. The resultant KdsC\textDelta C8 was also further purified on an S-200 gel filtration column and concentrated to 20 mg/mL for crystallization.

**Quantification of KDO8P** – The Aminoff assay was used to quantitate 2-oxo-3-deoxy sugars including KDO8P and KDO [19]. In a 10-mL glass culture tube, 100 µL of appropriately diluted sugar solution was subjected to complete oxidization
with 0.2 mL of 0.025 M NaIO₄ in 0.25 N H₂SO₄ at room temperature for 10 min. The excess oxidizing reagent was reduced by the addition of 0.4 mL of 2% (w/v) NaAsO₂ in 0.5 M HCl, indicated by the disappearance of yellow color in the liquid. This mixture was then treated with 1 mL thiobarbituric acid (0.36% w/v, pH 9.0) and heated at 100°C for 10 min to develop pink color. The amount of KDO8P was determined by measuring the absorbance of the reaction mixture at 549 nm (ε = 1.03×10⁵ M⁻¹ cm⁻¹ for the pink chromophore formed between α-formylpyruvate and thiobarbiturate).

**Enzymatic synthesis of KDO8P** – KDO8P was synthesized through enzymatic reaction catalyzed by 3-deoxy-D-manno-octulosonate 8-phosphate synthase (KdsA), and purified as described [12, 20]. Equal amounts of D-arabinose 5-phosphate and phosphoenolpyruvate (100 µmol each) were mixed in 100 mM HEPES pH 7 in a volume of 4 mL in the presence of 3.2 mg KdsA. The reaction mixture was incubated at 37°C for 16 h without stirring, and quenched with 4 mL of 10% trichloroacetic acid. Progression of the reaction was monitored by the Aminoff assay, and a yield of 82% was achieved. The clarified supernatant was adjusted to pH 7 with titration of NaOH solutions, and loaded on an AG-MPI anion exchange column equilibrated with water. The fractions containing KDO (chemical degradation product of KDO8P) and KDO8P were eluted sequentially in 500 mL of 0-400 mM LiCl gradient wash, as identified by the Aminoff assay. The latter fractions showing positive for Aminoff signal were combined, lyophilized, and reconstituted in 3 mL of water. The solution was then passed through a P-2 desalting column, and eluted with
water. Fractions positive for the Aminoff assay (indicating the presence of 3-deoxy monosaccharides), negative for malachite green assay (indicating the absence of inorganic phosphate), and negative for AgNO₃ (indicating the absence of Cl⁻) were pooled and lyophilized. The resultant pure KDO₈P was dissolved in water, adjusted to a concentration of 100 mM, and stored at -80°C for future use.

Crystallographic experiments and structure determination – All crystal forms of KdsC and KdsCΔ8 were grown by hanging-drop crystallizations at 21 °C (see Table 2-1). The data were processed with HKL2000 [21]. Initially, data for the 1.4 Å P1 crystal form were used to determine the structure of KdsC by molecular replacement program PHASER [22] using a monomer of KdsC[H] [13] as a search model. The structure was then iteratively built and refined at 1.4 Å using programs COOT [23] and REFMAC [24], respectively. Structures of the other crystal forms were determined using this refined structure as a molecular replacement search model and then refined analogously. The data and refinement statistics are summarized in Table 2-2. Solvent accessible surface area was calculated using program Surface Racer [25].
<table>
<thead>
<tr>
<th>Crystal forms</th>
<th>Protein solution</th>
<th>Reservoir solution</th>
<th>Cryoprotectant Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>KdsC-Mg\textsuperscript{2+}</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) MgCl\textsubscript{2} 20mM</td>
<td>PEG8K ~13.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM</td>
<td>PEG8K 17.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM MgCl\textsubscript{2} 10mM Glycerol 15%</td>
</tr>
<tr>
<td>KdsC-Mg\textsuperscript{2+}-Cl\textsuperscript{-} (Crystal grown with MgCl\textsubscript{2})</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) MgCl\textsubscript{2} 20mM</td>
<td>PEG8K ~13.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM</td>
<td>PEG8K 17.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM MgCl\textsubscript{2} 10mM Glycerol 15%</td>
</tr>
<tr>
<td>(Crystal grown with MgCl\textsubscript{2} &amp; KDO8P)</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) MgCl\textsubscript{2} 20mM</td>
<td>Ammonium Sulfate 1.875 M Na Cacodylate pH 6.5 100mM Glycerol 2.5%</td>
<td>Ammonium Sulfate 2.1 M Na Cacodylate pH 6.5 100mM Glycerol 20%</td>
</tr>
<tr>
<td>KdsC(\Delta)C\textsubscript{8}-Mg\textsuperscript{2+}-KDO\textsubscript{8}P (Crystal grown with MgCl\textsubscript{2} &amp; KDO8P)</td>
<td>KdsC(\Delta)C8(13mg/ml in 10mM HEPES pH 7.0) MgCl\textsubscript{2} 5mM KDO8P 2mM</td>
<td>Ammonium Sulfate 1.875 M Na Cacodylate pH 6.5 100mM Glycerol 2.5%</td>
<td>Ammonium Sulfate 2.1 M Na Cacodylate pH 6.5 100mM Glycerol 20%</td>
</tr>
<tr>
<td>KdsC-Ca\textsuperscript{2+}-PO\textsubscript{4} (Crystal grown with CaCl\textsubscript{2} and then soaked with KDO8P)</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) CaCl\textsubscript{2} 20mM</td>
<td>PEG8K ~12.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM</td>
<td>PEG8K 17.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM MgCl\textsubscript{2} 10mM Glycerol 15%</td>
</tr>
<tr>
<td>KdsC-Ca\textsuperscript{2+}-Cl\textsuperscript{-} (Crystal grown with CaCl\textsubscript{2})</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) CaCl\textsubscript{2} 20mM</td>
<td>PEG8K ~12.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM</td>
<td>PEG8K 17.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM MgCl\textsubscript{2} 10mM Glycerol 15%</td>
</tr>
<tr>
<td>KdsC-Cl\textsuperscript{-} (Crystal grown in absence of a divalent metal)</td>
<td>KdsC(12mg/ml in 10mM HEPES pH 7.0) CaCl\textsubscript{2} 20mM</td>
<td>PEG8K ~15.0% Bis Tris Propane pH 6.0 100mM NaCl 100mM Glycerol 5%</td>
<td>PEG8K 17.5% Bis Tris Propane pH 6.0 100mM NaCl 100mM MgCl\textsubscript{2} 10mM Glycerol 15%</td>
</tr>
<tr>
<td>Crystal composition</td>
<td>2R8E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KdsC-Mg(^{2+})</td>
<td>KdsC-Mg(^{2+})-Cl(^-) (Crystal grown with MgCl(_2) &amp; KDO8P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KdsCΔC8-Mg(^{2+})-KDO-PO(_4)</td>
<td>KdsC-Ca(^{2+})-PO(_4) (Crystal grown with CaCl(_2) &amp; then soaked with KDO8P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KdsC-Ca(^{2+})-Cl(^-) (Crystal grown with CaCl(_2))</td>
<td>KdsC-Cl(^-) (Crystal grown in absence of a divalent metal)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Space group</th>
<th>P1</th>
<th>P4(_{2}12_1)</th>
<th>C222(_1)</th>
<th>P2(_{1})</th>
<th>P2(_{1})</th>
<th>P2(_{1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (Å)</td>
<td>82.9</td>
<td>118.8</td>
<td>122.2</td>
<td>90</td>
<td>65.0</td>
<td>90</td>
</tr>
<tr>
<td>b (Å)</td>
<td>83.0</td>
<td>118.8</td>
<td>122.2</td>
<td>90</td>
<td>144.1</td>
<td>90</td>
</tr>
<tr>
<td>c (Å)</td>
<td>85.9</td>
<td>90.1</td>
<td>202.1</td>
<td>90</td>
<td>145.9</td>
<td>90</td>
</tr>
<tr>
<td>α (°)</td>
<td>118.8</td>
<td></td>
<td>118.8</td>
<td>90</td>
<td>156.9</td>
<td>90</td>
</tr>
<tr>
<td>β (°)</td>
<td>85.9</td>
<td></td>
<td>202.1</td>
<td>90</td>
<td>114.1</td>
<td>90</td>
</tr>
<tr>
<td>γ (°)</td>
<td>83.0</td>
<td></td>
<td>118.8</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resolution (Å)</th>
<th>50-1.4</th>
<th>50-2.85</th>
<th>50-2.5</th>
<th>50-2.1</th>
<th>50-1.9</th>
<th>2.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rmerge (%)</td>
<td>3.5 (57.9)</td>
<td>12.4 (70.0)</td>
<td>12.5 (46.2)</td>
<td>12.9 (40.9)</td>
<td>8.9 (57.0)</td>
<td>7.4 (46.6)</td>
</tr>
<tr>
<td>I/σI</td>
<td>39 (1.6)</td>
<td>22.4 (3.3)</td>
<td>8.9 (2.3)</td>
<td>13.5 (2.2)</td>
<td>19.6 (2.2)</td>
<td>23.0 (3.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95 (85)</td>
<td>100 (100)</td>
<td>92.7 (94.2)</td>
<td>98.4 (90.1)</td>
<td>99.2 (94.9)</td>
<td>99.1 (96.5)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8 (3.2)</td>
<td>27.7 (22.7)</td>
<td>4.1 (3.9)</td>
<td>5.3 (2.6)</td>
<td>6.8 (4.3)</td>
<td>6.6 (5.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>30-1.4</td>
<td>40-2.85</td>
<td>40-2.5</td>
<td>30-2.1</td>
<td>37-1.85</td>
<td>30-2.6</td>
</tr>
<tr>
<td>R/Rfree (%)</td>
<td>16.0/18.8</td>
<td>22.0/27.6</td>
<td>20.6/24.1</td>
<td>18.6/23.1</td>
<td>21.5/24.4</td>
<td>20.3/23.6</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>296631</td>
<td>36441</td>
<td>21494</td>
<td>162029</td>
<td>231079</td>
<td>86138</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td>0.010</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.010</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.009</td>
<td>0.006</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.368</td>
<td>1.104</td>
<td>1.042</td>
<td>1.045</td>
<td>1.090</td>
<td>0.864</td>
</tr>
<tr>
<td>No. of atoms</td>
<td>12808</td>
<td>10844</td>
<td>5442</td>
<td>24342</td>
<td>22566</td>
<td>22256</td>
</tr>
<tr>
<td>No. of monomers per asymm. unit</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>
2.4 Results

Sequence alignment – Homologues of KdsC from various G- organisms are compared in multiple sequence alignment (Figure 2-2). All the HADSF motifs are conserved. Secondary structures of the species with known crystal structure data are consistent with each other.

Figure 2-2 Sequence alignment of KdsC homologues. The secondary structure of *E. coli* KdsC is shown schematically above the alignment. Residues involved in the tetramerization interface are labeled with black bars (in the $\beta_2$-$\beta_3$-hairpin) and black circles (outside of the hairpin). Residues coordinating to $\text{Mg}^{2+}$ and contacting with $\text{P}_i$ are marked with open red circles. Residues involved in KDO recognition are shown as the solid red circles.

Preparation of KdsC<sup>Ec</sup> and KdsC$\Delta$C8 – Both KdsC<sup>Ec</sup> and KdsC$\Delta$C8 were overexpressed in *E. coli* BL21(DE3) cells and purified to homogeneity through two chromatographic steps. SDS-PAGE results showed a molecular weight of around 22 kDa for KdsC<sup>Ec</sup> and 21 kDa for KdsC$\Delta$C8 (Figure 2-3). Both proteins were
concentrated to 20 mg/mL for crystal growing and resulted in various crystal forms [17].

**Figure 2-3** SDS-PAGE of purified 1) wild-type KdsC\textsuperscript{Ec} and 2) KdsCΔC8 with eight amino acids DEAKGQSI truncated from the C-terminus.

**KdsC structures** – Seven crystal forms were obtained, with wild-type KdsC and its C-terminal tail deletion mutant KdsC cocrystallized with Mg\textsuperscript{2+} or Ca\textsuperscript{2+}, soaked or not with KDO8P (Table 2-2, Figure 2-4).

**Figure 2-4** A schematic of KdsC structures. Red dots represent Mg\textsuperscript{2+} ion, orange dots for Cl\textsuperscript{-}, and purple dots for Ca\textsuperscript{2+}. The phosphate portion is shown in green sticks and the KDO portion is shown in blue sticks. The eight-amino acid tail insertions are shown in lines in consistent colors consistent with each monomer.
The monomers of KdsC adopt a Rossmannoid fold, as observed in the initial structure at 1.4 Å of KdsCEc cocrystallized with its physiological metal Mg$^{2+}$ (Figures 2-5). Four monomers organize in a tetramer, similar to previously reported KdsC$^{Hi}$ [13] and KdsC$^{Aa}$ (PDB ID 2P9J) structures. The tetrameric structure is stabilized by interactions between β-hairpins from each monomer. The β-hairpins are formed by sequence inserts at the same location as the C1 caps but much shorter, and organized as a central 8-stranded intermolecular β-barrel mainly through the hydrophobic interactions. Conserved residues Leu41, Tyr43, Lue51, and Ala53 are involved in the hydrophobic contacts and bury 880 Å$^2$ of solvent accessible surface area per monomer in the β-barrel interface.

![Figure 2-5](image)

**Figure 2-5** The tetrameric structure of KdsC in a) the top view and b) the side view. The Mg$^{2+}$ ions are shown in red dots and Cl$^-$ shown in orange dots.

The active site cleft of KdsC is formed in the interface between two monomers, and is marked with the presence of Mg$^{2+}$. The metal ion is hexacoordinated (Figure 2-6) by the side chain of Asp32 and the backbone carbonyl of Asp34, both from conserved motif I of HADSF; the side chain of Asp125 from motif IV, and three
water molecules, similar to the Co\textsuperscript{2+} captured in KdsC\textsuperscript{H} structure [13]. The active site cleft is largely lined up with polar and charged residues, including Thr76 from motif II and Lys102 from motif III.

**Figure 2-6** The 2F\textsubscript{o}−F\textsubscript{c} electron density map of the active site captured in the 1.4 Å KdsC structure, contoured at 1 σ.

**Binary states of active site** – Two tetramers present in the same asymmetric unit of the Mg\textsuperscript{2+}-bound KdsC crystal have evident differences in the location of the C-terminal tail (**Figure 2-4** crystal a and b). In one tetramer, all monomers have the tail inserted into the active site of the adjacent monomer and occluded it, resulting in all four active sites in a closed states. Besides the Mg\textsuperscript{2+} ion, one Cl\textsuperscript{−} ion from the protein storage buffer is also present in the closed active site, 4.4 Å from the Mg\textsuperscript{2+}, indicating a possible location of the P\textsubscript{i} binding site. In the other tetramer, all C-terminal tails remain disordered, leaving the active sites exposed to the solvent, thus being termed the open state. No anion can be found bound in the open state of active sites near the metal cofactor Mg\textsuperscript{2+}. In the crystal of Mg\textsuperscript{2+}-bound KdsC soaked with KDO8P (**Figure 2-4** crystal c), the diagonal pair of active sites remains in closed
state while the other diagonal pair is in open state. The product-bound KdsCΔC8 active sites are also considered in closed states because solvent is excluded by the KDO portion (Figure 2-4 crystal d, Figure 2-8a). In all closed active sites, either a scissile phosphate or a Cl⁻ ion is bound, while in all open active sites, no anion is present except in the tail-deletion mutant. The tail-in/tail-out conformational change also induces corresponding conformational change of the active site loop I and loop II, characteristic in displacement of some highly conserved residues including Thr76, Lys102 and Arg78 (Figure 2-7).

**Figure 2-7** Conformational differences between the open state and the closed state of the active site. An open-state monomer from Mg²⁺-bound KdsC crystal is shown in red line and a closed-state monomer from product-bound KdsCΔC8 is shown in blue lines. Mg²⁺ from both monomers are overlapped and shown in a red ball.

**Insertion of the C-terminal tail into the active site** – In many crystal forms of wild-type KdsC<sup>Ec</sup>, the eight-amino acid C-terminal tails are found inserted into the active site of the adjacent monomer in *trans*, resulting in closed states of the active
sites. The tail residues make extensive specific contacts with the active site residues, especially the three amino acids at the C-terminal end (Figure 2-8a). Generally, the carbonyl on the side chain of Gln186 forms a hydrogen bond with Arg63 from the same monomer. The hydroxyl group of Ser187 interacts with the backbone carbonyl of Val56 from the same monomer. Finally, the terminal carboxylate of Ile188 forms salt bridges with Arg86 from the same monomer and Arg78 from the monomer possessing the active site. In the structure of Ca$^{2+}$-bound KdsC$^{Ec}$ soaked with KDO8P (Figure 2-4 crystal e), a phosphate ion was captured in the tail-in closed active site, hydrogen bonding to the backbone carbonyl of Ser187. The active site is stabilized through these interactions between the tail and the polar or charged residues. The occlusion of the tail at the entrance of the active site completely excludes solvent.

Figure 2-8 a) Specific contacts between the active sites residues and the tail residues of the Mg$^{2+}$-bound KdsC$^{Ec}$. b) Interactions between the active site residues and the catalytic products KDO and phosphate in KdsCΔC8. Two adjacent monomers are shown in orange and green sticks, respectively. The tail residues shown as pink sticks are from the same monomer as the green residues. KDO is in blue sticks and the inorganic phosphate is in cyan sticks. Mg$^{2+}$ is designated as a red dot and Cl$^{-}$ is an orange dot. Pictures were made using PyMol.
Presence of catalytic products in the active site – Initially, Ca$^{2+}$ was used as the cofactor to slow down the reaction of KdsC$^{Ec}$ in the hope of capturing KDO8P in the active site. However, only the scissile phosphate was trapped in the active site, while the tail residues occupied the rest of the active site cavity. Very interestingly, when the tail-deletion mutant KdsCΔC8 was crystallized and soaked with KDO8P, a substrate-shaped electron density was obtained in the active sites. After structural refinement, the electron density turned out to be the cleaved products, KDO and P$_i$. The tetramer obtained in this structure contains two open-state active sites organized diagonally, with P$_i$ and Mg$^{2+}$ bound, and two closed-state active sites occupied by KDO, P$_i$, and Mg$^{2+}$ (Figure 2-4 crystal d). P$_i$ in these crystal forms is at a similar site as Cl$^-$ ions in other crystal forms.

Figure 2-9 Scheme of specific interactions of KDO, P$_i$ and Mg$^{2+}$ in the KdsC$^{Ec}$ active site. Residues labeled in green are from the adjacent monomer.
The products-occupied active site has the same Mg$^{2+}$ coordination as observed in the tail-occupied active site or the open-state active sites, except that an oxygen on P$_i$ coordinates to Mg$^{2+}$ instead of a water oxygen (Figure 2-9). The phosphate ion also makes specific contacts to the main chain amide group of Gly77 and the side chains of Lys102 and Thr76, all within 3 Å, and to the backbone amide group of Val33 through a water-mediated hydrogen bond.

![Figure 2-10 Omit F$_o$-F$_c$ electron density map of the active site contoured at 2.3σ.](image)

The presence of the product KDO in the active site was confirmed by an omit electron density map (F$_o$-F$_c$) when excluding KDO and P$_i$ densities (Figure 2-10). The conformation of KDO was determined by both the electron density map and knowledge of chemical bonding and catalytic mechanism. The hydroxyl group on C8 of KDO points to the P$_i$ with 2.3 Å distance away from the oxygen on P$_i$, indicating cleavage of the phosphoester bond (Figure 2-9). The KDO sugar is locked in position by forming extensive hydrogen-bonding, electrostatic and hydrophobic interactions.
with enzyme residues, many of which are similar to the contacts made by the tail residues (Figure 2-8). Arg78 from the active site and Arg86 and Val56 from the adjacent monomer are almost located at the same positions as observed in the tail-occupied active sites. The universally conserved Arg78 interacts with O4 and O5 of KDO and is stabilized by the charge of Glu50 at the same time. Arg86 forms a strong salt-bridge with the C1 carboxylate of KDO in the active site of an adjacent monomer (in trans), locking the sugar in α-anomeric configuration. Val56 and Leu90 make hydrophobic interactions with methylene groups at C8 and C3, respectively. An evident conformational change can be observed for Arg63, which goes from interacting with the side-chain carbonyl of Gln186 on the tail to hydrogen bonding with the C2 hydroxyl group on KDO in trans, in both cases through its secondary amine group. All the surface-accessible portions of KDO formed specific contacts with residues from either monomer, defining a high degree of substrate specificity.

2.5 Discussion

The position of Mg$^{2+}$ both herein and in previously reports indicates the general location of the active site. However, the location of the substrate or catalytic products in the active site was not clear. The Ca$^{2+}$-bound crystal was originally hypothesized to allow substrate trapping by minimizing the catalytic turnover as compared to Mg$^{2+}$-assisted reactions, but contrary to the hypothesis, KDO8P was hydrolyzed into products and only P$_i$ was invariably captured in all the active sites. When soaking the tail-deletion mutant KdsCΔC8 with KDO8P, catalytic products
KDO and P$_i$ were successfully captured in two of the diagonal active sites, and P$_i$ alone was observed in the other two diagonal active sites. The structural analysis on the products-bound crystal elucidated the details of the active site binding environment. Conserved residues Lys102, Gly77 and Thr76 line up in the flexible loops to position the P$_i$ molecule, and this position is consistent with an in-line nucleophilic attack by an aspartate side chain as observed in other HADSF enzymes [26]. Asp32 is highly likely to be the nucleophilic attacking residue, because it coordinates to Mg$^{2+}$ with the side chain carboxylate, and undergoes rotamer change in the presence or absence of P$_i$. This may reflect destabilization of the catalytic center to allow product dissociation.

Multiple crystal forms were obtained in this study, representing various stages in the KdsC catalytic cycle. Dramatic conformational changes were observed in two regions of the enzyme: 1) the C-terminal tail region and 2) the loop I and loop II in the active site. The presence of both tail-in and tail-out monomers in the same tetramer (Figure 2-4 crystal c) indicates that the active site of free enzyme exists in an equilibrium of open and closed state, achieved by the movement of the tail. In the active sites closed by the tail, a negative ion is always present, either a Cl$^-$ from the protein storage buffer or a scissile phosphate from the cleavage of KDO8P, suggesting the tail completely excludes solvent from the active site and enables locking of a counter ion near the metal cofactor (Figure 2-4 crystals a, c, e, f and g). The fact that both products are present in the KdsCΔC8 active site strongly suggests that the C-terminal tail is quite possibly involved in regulation of the catalytic rate of
KdsC. Conformational changes of loops at the active site corresponds to the tail movement. In the tail-in monomer, both loop I (residues 75-80) and loop II (residues 95-102) were driven closer to the tail, forming a bulge in the active site cavity. The same conformational change was observed in the KDO-bound structure of KdsCΔC8, in which the KDO portion was locked precisely in the active site by specifically interacting with residues on the loops as well as residues on the adjacent monomer.

Instead of possessing a cap domain as the C1 and C2 group enzymes of HADSF, each monomer in the tetrameric KdsC forms a cap-like structure over the active site of the adjacent monomer. The tetramerization results in only a narrow passage into each active site, and successfully reduces the solvent exposure of the active site to a large extent. This is critical to the enzyme activity since solvent exclusion is required for formation of the aspartylphosphate intermediate [2]. Another role the cap domain plays in HADSF enzymes is to assist in substrate recognition [14], and this role is also played by the monomer covering the adjacent active site. Polar and nonpolar residues on the neighboring monomer interact with the substrate in trans via extensive hydrogen bonding, salt-bridge or hydrophobic interactions, and define the substrate specificity. The residues forming direct contacts with KDO include Arg63, Arg86, Val56 and Leu90. A very similar organization of polar and nonpolar residues was reported for the cap domain of β-phosphoglucomutase, in which Arg49 forms a charge interaction with a phosphate moiety and backbone of Val47 hydrogen bonds to a hydroxyl group of the β-D-glucose 1, 6-phosphate intermediate of similar size as KDO8P [14]. Among the KdsC “cap-like” residues, Arg86 interacting with the
C1 carboxylate of KDO through electrostatic force is likely the essential residue performing the anomeric selectivity. According to studies of Baasov and Jakob using 2-deoxy analogues of KDO8P, KdsC retains 50% activity when utilizing the α-anomer of the KDO8P analogue which has the C1 carboxylate in the equatorial position, but has no activity with the β-anomer [27].

As a conclusion, the structural analysis revealed the atomic basis of substrate specificity supported by both the active site residues and the residues from the adjacent cap-like monomer. The tetramerization also facilitates the participation of the C-terminal tail in the neighboring active site through conformational change.

2.6 Future Directions

The residue Arg86 interacting with the C1 carboxylate of KDO in trans is likely to play the critical role of defining the anomeric specificity of KDO8P. This hypothesis can be further tested through mutational studies. An Asn mutant of conserved Arg86 can be constructed and assayed with Baasov’s KDO8P analogues if available. Removal of the charge from Arg86 may weaken the selectivity.

2.7 Acknowledgements

The tail deletion mutant KdsCA8 was cloned by former postdoctoral fellow Dr. Jing Wu in the Woodard group. Wild-type KdsCEc was purified by previous graduate student Parag Aggarwal for crystallization. KdsA used for enzymatic synthesis of KDO8P was prepared by previous graduate student Jingjing Li. Crystallization
screening was performed by Hauptman Woodward Medical Research Institute (Buffalo). Dr. Jeanne Stuckey and John Rubin from the Life Science Institute assisted the data collection of the Mg$^{2+}$-bound KdsC$^{Ec}$. All other data collections and structural refinement were performed by Professor Oleg Tsodikov and Dr. Tapan Biswas from the Department of Medicinal Chemistry.
2.8 References


Chapter 3

The Role of the C-terminal Tail of KdsC

3.1 Summary

An eight-amino acid random coil tail at the C-terminus of *E. coli* KdsC (KdsC\textsuperscript{Ec}) undergoes a dramatic conformational change as revealed by crystal structures. The transition between the solvent-exposed state and the active site-occluded state indicates a possible role for the C-terminal tail in the catalytic cycle. Multiple sequence alignments of KdsCs from various microorganisms suggest a possible classification of KdsCs into tail-less, random-coiled tailed, and α-helical tailed groups. This classification is further supported by the crystal structures of a representative from each group. To investigate the function of the tail, a deletion mutant of KdsC\textsuperscript{Ec} with eight amino acids removed from the C-terminus was constructed and annotated as KdsCΔC8. In addition, two enzymes belonging to the tail-less KdsC group were prepared for kinetic characterization, one from *Aquifex aeolicus* and the other from *Coxiella burnetii*. Compared to the full-length KdsC\textsuperscript{Ec}, the tail-absent KdsCs have similar $K_M$ values but $10^2$-$10^3$ fold lower catalytic rates ($k_{cat}$), suggesting a critical role of the tail in the catalytic cycle. Combined with structural information discussed in Chapter 2, a model of the KdsC catalytic cycle is
proposed, highlighted with the active participation of the tail in product release through a conformational change. This product release step is the rate-limiting step.

3.2 Introduction

Structural studies on *E. coli* KdsC (KdsC\(^{E_{c}}\)) revealed a dramatic conformational change in the C-terminal region. This random coiled region, formed by the eight amino acids from the C-terminus, is flexible in moving in and out of the active site of the neighboring monomer, and is therefore termed the “tail” (Figure 3-1a). It has been reported in other enzymes that small C-terminal domains may play regulatory roles through conformational changes between ordered and disordered states. One example is the subunit ε in ATP synthase that has a C-terminal α-helical tail undergoing conformational change between the hairpin and extended structures, which is necessary for ATPase inhibition [1]. Another example is a DNA-mediated conformational switch of the C-terminal tail in bacteriophage λ integrase between an

Figure 3-1 Comparison of the C-terminal structure of KdsC from a) *Escherichia coli* (2R8E), b) *Haemophilus influenzae* (1J8D), and c) *Aquifex aeolicus* (2P9J). All C-terminals are colored in red.
activated trans-packing state and an inactivated cis-packing state [2]. However, this phenomenon has never been reported for any haloacid dehalogenase superfamily (HADSF) enzyme.

Interestingly, the two previously solved structures of KdsC from *Haemophilus influenzae* (KdsC\(^{Hi}\)) [3] and *Aquifex aeolicus* (KdsC\(^{Aa}\), PDB ID 2P9J) have conspicuously different C-terminal regions compared to the random coiled extension in KdsC\(^{Ec}\), although the remaining part of the enzymes, i.e. the Rossmannoid core, are almost identical. KdsC\(^{Hi}\) has a longer C-terminal tail than KdsC\(^{Ec}\), and forms a short \(\alpha\)-helix with residue 162-180 (Figure 3-1b). The significance of this bulky tail has not been addressed in any studies. KdsC\(^{Aa}\), however, has a sequence ending at the 6\(^{th}\) \(\alpha\)-helix in the Rossmannoid core without an observable C-terminal extension (Figure 3-1c).

KdsC\(^{Aa}\) was studied in the Woodard lab previously [4]. It requires divalent metal ion as cofactors, and displays an acidic pH preference (optimum at pH 5), a high substrate specificity for KDO8P, and a \(K_m\) value of 65 \(\mu\)M. All of these properties are similar to KdsC\(^{Ec}\) [5], except that the \(k_{cat}\) value of KdsC\(^{Aa}\) (5.6 s\(^{-1}\)) is 30-fold smaller than that of KdsC\(^{Ec}\) (175 s\(^{-1}\)). Whether the presence or absence of a tail plays any role in affecting the catalytic rate remains unclear.

To further investigate the role of the C-terminal tail in KdsC, the tail-deletion mutant KdsC\(\Delta C8\) used for crystallization was characterized with steady-state kinetics in parallel with the full-length KdsC\(^{Ec}\) and two naturally tail-less forms of KdsC. In addition to *A. aeolicus*, another Gram-negative bacterium, *Coxiella burnetii*, was
found to carry a tail-less \textit{kdsC} gene using the NCBI database search. This highly pathogenic strain is responsible for causing Q fever, a notoriously infectious disease occurring in both humans and livestock [6]. A comparison of kinetic behavior of these assorted enzymes, \textit{KdsC}^{Ec}, \textit{KdsC}_{ΔC8}, \textit{KdsC}^{Aa}, and \textit{KdsC}^{Cb}, combined with the structural information suggest that the C-terminal tail is likely to be involved in the catalytic cycle of \textit{KdsC} by facilitating the second step of product release. This regulatory function is achieved by conformational changes of the C-terminal tail.

3.3 Experimental Procedures

\textbf{Materials} – The genomic DNA of \textit{Coxiella burnetii} RSA 493 was a generous gift from Professor Jim Samuel at Texas A&M University. Malachite green carbinol hydrochloride, ammonium molybdate tetrahydrate, molecular weight marker kit, and bovine serum albumin (BSA) were purchased from Sigma-Aldrich Company. Tris(hydroxymethyl)aminomethane was from Research Organics. Bacto™ tryptone was from Becton, Dickinson and Company. Anhydrous sodium phosphate monobasic was obtained from J. T. Baker Company. Isopropyl β-D-1- thiogalactopyranoside (IPTG) was from Gold Biotechnology, Inc. Sulfuric acid, hydrochloric acid, yeast extract, sodium chloride, potassium chloride, ammonium sulfate, ampicillin sodium salt, Nunc™ polystyrene 96-well flat bottom assay microplates, and Fisherbrand polystyrene disposable plastic cuvettes were from Fisher Scientific Inc. Polypropylene 96-well half skirted reaction PCR plates were obtained from DOT scientific, Inc. PCR primers were synthesized by Invitrogen corporation. FailSafe™ pre-mix from
Epicenter was used as PCR buffer. Plasmid extraction was performed using Wizard® Plus SV Minipreps DNA purification kit from Promega Corporation. Restriction enzymes, calf alkaline phosphatase and T4 ligase were purchased from New England Biolabs, Inc. DNA sequencing was conducted by the University of Michigan Biomedical Resources Core Facility. Bradford Protein Assay Reagent, Mini-PROTEAN II electrophoresis unit, and microplate reader Model 550 were from Bio-Rad Laboratories, Inc. The Q-sepharose resin, Phenyl Superose (HR 10/10) column, prepacked HiPrep Sephacryl S-100 (26/60), and FPLC® system were from Pharmacia, GE Healthcare. XL1-Blue *E. coli* strain was from Stratagene, Agilent Technologies. BL21(DE3) *E. coli* strain, BL21-Codon Plus™(DE3)RIL *E. coli* strain, and HisBind® Ni-NTA resin were from Novagen, EMD Chemicals Inc. High grade spectra/Por® 7 dialysis tubing (10,000 Da molecular weight cut-off and metal free) was from VWR International, LLC. The Millex® syringe driven filter units (0.22 µm) was obtained from Millipore. The 8453 UV-vis diode-array spectrophotometer was from Hewlett Packard. All the programmed temperature controls were performed on a MJ research PTC-200 Peltier thermal cycler.

**Sequence analysis** – Database searches of multiple microbial organisms was performed utilizing the BLAST program at the NCBI website (http://ncbi.nlm.nih.gov/BLAST). Multiple sequence alignments were generated using ClustalW (http://www.ebi.ac.uk/clustalw).

**Protein Concentration Assay** – Protein concentration was determined using the Bradford assay by mixing 1 mL dye with up to 10 µL of protein. BSA served as a
standard. Alternatively, ultra-violet absorbance of a protein solution measured at 280 nm was converted into protein concentration, based on the extinction coefficient calculated from ProtParam tool (ExPASy).

**Expression and Purification of KdsC** – The *A. aeolicus kdsC* gene was cloned into the pT7-7 vector previously in the Woodard lab [4]. The plasmid was transformed into chemically competent *E. coli* BL21-Codon Plus™(DE3)RIL cells for expression. Cells were grown in 1 L 2×YT culture, and induced with a final concentration of 0.2 mM IPTG when OD<sub>600</sub> reached 1.6. After growing at 17°C for 18 h after induction, cells were harvested by centrifuging at 6000 rpm for 10 min on a Beckman JA-10.5 rotor, and sonicated in 15 mL of buffer A (20 mM Tris pH 7.5). The supernatant was clarified by centrifuging at 15000 rpm for 30 min on a Beckman JA-25.5, and divided into two equal portions. Half of the crude extract was subjected to a heating procedure according to a previous protocol [7]. Briefly, after addition of NaCl to a final concentration of 0.1 M, the crude extract was heated in a boiling water bath for 2 min and then heated at 80°C for 8 min with continuous stirring. The suspension was separated into supernatant and precipitate, and subjected to SDS-PAGE analysis. The unheated portion of crude extract was applied on a Q-sepharose column equilibrated with buffer A, and eluted with a 0-2 M KCl gradient in 90 min. Fractions containing KdsC<sup>4a</sup> as judged by SDS-PAGE were pooled and treated with ammonium sulfate to a final concentration of 20% (w/v). The protein solution was loaded on a phenyl superpose (HR 10/10) column with a reverse elution from 20% to 0% (w/v) of ammonium sulfate in buffer A. The fractions containing
KdsC$^{da}$ were eluted around 4% of ammonium sulfate and combined for dialysis against buffer A. The resultant protein solution was 1.36 mg/mL and was stored at -80°C for future use.

**Cloning and expression of KdsC$^{Cb}$** – Genomic DNA of *C. burnetii* RSA 493 was quantified by absorbance readings at 260 nm and prepared to a concentration of 25 ng/µL. The PCR was carried out using the forward primer GATTCTAGAATTC ATATGCAAACAACGGCTGAC and the reverse primer GATTTTATTTTATCTTTG CCTTTGACTGCAGTTGGATCCGAATTC. The reaction mixture included 25 µL FailSafe™ pre-mix K, 1 µL of genomic DNA solution (25 ng/µL), 1 µL each of the primers (100 ng/µL), and sterile water up to 49 µL. After heating the reaction mixture at 95°C for 2.5 min, 1 µL Vent DNA polymerase was added. The program was set with 30 cycles each consisting of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 45 s. The PCR product was purified from 1% (w/v) low-melt agarose gel and digested with *NdeI* and *BamHI*. The expression vector pT7-LOH was similarly double digested and treated with calf intestinal alkaline phosphatase. The PCR product and the vector are ligated by T4 ligase at room temperature overnight. The ligation was transformed into chemically competent XL1-blue *E. coli* cells, which were then spread on an agar plate of LB supplemented with 100 mg/L ampicillin (LB/amp). A single colony containing pT7-LOH-KdsC$^{Cb}$ plasmid, as proved by DNA sequencing, was amplified and stored at -80°C in 25% glycerol.
The pT7-LOH-\textit{Kds}^{Cb} plasmid was transformed into chemically competent BL21(DE3) for expression. Cells were grown in 2 L of LB/amp media, and induced with IPTG to a final concentration of 0.2 mM when OD_{600} reached 0.9. After growing at 17°C for 18 h after induction, the cells were harvested by centrifugation at 6000 rpm in a Beckman JA-10 rotor for 10 min. The cell paste was lysed in 20 mL HisBind® (Novagen) binding buffer (20mM Tris pH8.0, 500mM NaCl, 5mM imidazole) by sonication for 30 s for 4 times. The crude extract was then separated from the cell debris by centrifuging at 15000 rpm for 30 min in a Beckman JA-25.5 rotor.

The supernatant obtained was treated with 10 µL of 10 mg/mL DNase I and RNase A each at room temperature for 10 min. After filtration through a 0.22 µm Millex® membrane, the crude extract was mixed with 4 mL HisBind® resin (Novagen) in a 50 mL centrifuge tube. The resin was saturated with the his-tagged protein by rotating the mixture on Labquake® rotator (Barnstead International) at a low speed at 4°C overnight. The resin/protein mixture was then transferred to a Poly-Prep® column (Bio-Rad) for washing, and eluted using column chromatographic techniques. The target protein was eluted with 250 mM imidazole in the binding buffer. To avoid formation of precipitate after elution, 1 mL of the binding buffer containing 1 mM EDTA was added to the collection tube prior to protein elution, and 1 mL of protein was collected in each tube. Finally, the pooled protein solution was subjected immediately to dialysis against 20 mM Tris pH 7.5 with two buffer changes, initially with 1 mM EDTA and later without EDTA. A small amount of precipitate was
removed via centrifugation and the solution was tested for protein concentration using the Bradford assay. The yield was 10 mg/liter of cell culture. The homogenous KdsC\textit{Ch} at 1 mg/mL was stored at -80°C for future use.

**Gel filtration** – A Sephacryl S-100 (26/60) column was pre-equilibrated by running 20 mM Tris pH 7.5, 150 mM NaCl at 0.4 mL/min for 24 h. Blue dextran, beta amylase, alcohol dehydrogenase, albumin, cytochrome C, and carbonic anhydrase from the molecular weight marker kit were prepared in solutions of 2 mg/mL and each was applied in 1 mL volume on the column by injection into the sample loop. The column was washed using the same buffer at 1 mL/min. Retention time of each protein was recorded and a calibration curve was plotted. Purified KdsC\textit{AaC8} and KdsC\textit{Aa} were applied to the column in the same manner as described for the molecular weight markers. The molecular weight of the target proteins was then calculated based on their retention time and the calibration curve.

**Enzyme assay** – KdsC activity was monitored by the malachite green (MG) assay for inorganic phosphate (P\textsubscript{i}) release [5, 8]. A 0.045% malachite green hydrochloride aqueous solution was mixed with 4.2% ammonium molybdate in 4N HCl to a ratio of 3:1, stirred at room temperature for 1 h, and passed through Whatman No.5 filter paper. To every 100 µL sample, 900 µL MG reagent was added. The mixture was transferred into a cuvette and subjected to a UV-vis measurement at 660 nm after a 5-min color development. The assay can also be adapted to a 96-well plate format with modifications in reagent preparation. The reagent was prepared by combining 0.09% malachite green hydrochloride aqueous solution and 8.4%
ammonium molybdate in 11N HCl to a ratio of 3:1. The mixture was stirred and filtered as described for the conventional tube-based measurement. In each well of a 96-well reading plate, 200 µL of MG reagent was added to 20 µL of sample. After incubated at room temperature for 5 min, the plate was read at 655 nm.

A typical enzyme reaction for KdsC<sub>Ec</sub>, KdsCΔC8 or KdsC<sub>Cb</sub> contained 100 mM HEPES pH 7, 1 mM MgCl<sub>2</sub>, 10 µL enzyme, and deionized water up to 50 µL. After incubated at 37°C for 1 min, the reaction was quenched with 50 µL 10% trichloroacetic acid and subjected to either a conventional tube-based or a 96-well plate format MG assay. KdsC<sup>Ad</sup> was assayed the same way except with 2 mM MgCl<sub>2</sub> and an incubation at 60°C.

**Steady-state kinetic measurements** – The P<sub>i</sub> released in one minute was quantified in each reaction containing KDO8P varying from 0 to 2 mM. The initial reaction rate (v<sub>i</sub>) was derived as µmol of P<sub>i</sub> production per minute per mg of enzyme. The plots of v<sub>i</sub> vs. KDO8P concentration showed hyperbolic Michaelis-Menten curves, and were fit to 

\[ v_i = E_0 \cdot \frac{k_{cat} \cdot S}{K_M + S} \]

using the Origin program.

**Salt dependence** – Specific activity of KdsC<sub>Ec</sub>, KdsCΔC8 or KdsC<sub>Cb</sub> was measured in the presence of NaCl at various concentrations (0-500 mM). Other components included in the reaction mixture are as described above. Steady state kinetic parameters were also measured at 200 mM NaCl.
3.4 Results

**Multi sequence alignment** – The hits obtained querying the NCBI database against the KdsC\textsuperscript{Ec} sequence are identified as putative KdsCs on the basis of the conservation of the four HADSF motifs as well as the signature GGxGxxRE sequence [5]. Recent work revealed several sequence markers of 2-keto-3-deoxy-D-glycero-D-galacto-9-phospho-nononic acid (KDN 9-P) phosphatase, a HADSF enzyme sharing >25% sequence identity with KdsC including all the four motifs and the signature sequence [9]. These sequence markers in KDN 9-P phosphatase, Glu56 (aligned to Arg78 in KdsC\textsuperscript{Ec} in multiple sequence alignment) and Lys67 (aligned to Thr89 in KdsC\textsuperscript{Ec}), are also taken into consideration when identifying possible KdsC from the database. The multiple sequence alignment reveals differences in the C-terminal region of the KdsC amino acid sequence, thereby leading to the categorization of three groups of KdsC (Figure 3-2). The majority of the entries fall into Group 1, which is characterized by a short C-terminal extension consisting of eight amino acids and is proved to be a random coil based on the structure of KdsC\textsuperscript{Ec}. Proteins in this group vary in total length, but the C-termini are all aligned on the partially conserved eight tail residues. These tails bear abundant hydrophilic side chains, revealing a basis for solvent exposure. About 10 species, including *Pasteurella multocida*, *Pseudomonas putida* F1, *Neisseria meningitides* 29E, some *Haemophilus* strains, some *Mannheimia* strains, and some *Actinobacillus* strains, have KdsC with a long C-terminal extension, and are assigned to Group 2 KdsC. This partially conserved long tail adopts an α-helical secondary structure as observed in the structure of
Figure 3-2 Multiple sequence alignments of KdsC from various Gram-negative organisms.

KdsC<sup>Hi</sup> (PDB ID 1J8D). Only a few KdsCs belong to Group 3, which have no C-terminal extension compared to KdsC<sup>Ec</sup> or KdsC<sup>Hi</sup>. This group includes only *Aquifex aelius* and *Coxiella burnetii* so far from the database. Some other “tail-less” KdsC identified from the database, such as those from *Methanosarcina acetivorans*, *Acaryochloris marina*, and *Chlorobium tepidum*, are more likely to be KDN 9-P phosphtase based on the sequence markers mentioned above, but the sequences of...
these strains are still shown in the alignment for comparison. The KDN 9-P phosphatases seem to be universally tail-less.

**Phylogenetic tree of KdsC** – To examine the relationship between the KdsCs from the evolutionary perspective, a Phylip-format dendrogram was generated (Figure 3-3). The Group 1 and Group 3 KdsCs seem to bifurcate early from a co-ancestor not yet discovered. The Group 2 α-helical-tailed KdsCs possibly originated from the tail-less KdsCs in Group 3, and then evolved completely independently from the Group 1 random-coil-tailed KdsCs. The only exception is the KdsC from *Helicobacter pylori* that is included in the Group 2/3 lobe of the phylogenetic tree. To avoid sequence bias contributed by the differences in the C-termini, all KdsC sequences were truncated from the N-terminus until one amino acid before the Motif I and C-terminus for the tail sequences.

![Figure 3-3 Phylogenetic tree of KdsC from selective microorganisms. The figure was generated by ArboDraw software. Shown in pink are Group 2 KdsCs with α-helical tails, and in blue are Group 3 tail-less KdsCs. The rest of black are Group 1 random-coiled tails.](image-url)
**Preparation of KdsC\textsuperscript{Aa}** – Enzymes from *A. aeolicus* are usually thermostable at high temperatures, therefore heating at 85-100°C can be the first step in the purification [7, 10]. However, KdsC\textsuperscript{Aa} only remained soluble below 60°C [4]. This was confirmed in purification trials when crude extracts containing overexpressed KdsC\textsuperscript{Aa} were heated and KdsC\textsuperscript{Aa} co-precipitated with the *E. coli* proteins at high temperature. Purified KdsC\textsuperscript{Aa} was therefore obtained through two column chromatographies performed at room temperature. The yield of homogenous protein was 30 mg/L of culture.

**Preparation of KdsC\textsuperscript{Cb}** – The gene of KdsC\textsuperscript{Cb} was cloned into the his-tagged vector pT7-LOH and overexpressed as soluble protein. A limited amount of pure protein was obtained from the one-step Ni-NTA column purification, because the protein precipitated right after elution from the column. This is possibly due to the co-elution of Ni\textsuperscript{2+} ions at high concentration of imidazole that can form complexes with several proteins, therefore causing aggregation. Adding EDTA in the collecting tube efficiently reduced the precipitation, and reducing the imidazole concentration from 500 mM to 250 mM helped enhancing the solubility of the protein without affecting eluting efficiency. Finally, a yield of 10 mg/liter of cell culture was achieved.

**Oligomerization of KdsC** – The Sephacryl S-100 column was first calibrated with several proteins of known molecular weight. A linear correlation between relative retention time (\(t_r/t_0\), \(t_0 = \) retention time for blue dextran) and log of molecular weight was derived. The retention time was measured for samples and molecular
weight was calculated (Table 3-1). KdsCΔC8 in solution showed a molecular weight of 102.5 kDa, indicating a tetrameric structure, since the monomer is calculated to be ~18 kDa. KdsC<sup>4a</sup> exhibited a molecular weight of 14.9 kDa, suggesting that although the crystal structure revealed a tetrameric construction, the protein in solution remains as a monomer. This result agrees with a previous trial [4].

Table 3-1 Molecular weight of KdsCΔC8 and KdsC<sup>4a</sup> determined from gel filtration.

<table>
<thead>
<tr>
<th>Protein/Marker</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome C</td>
<td>12.4</td>
</tr>
<tr>
<td>KdsC&lt;sup&gt;4a&lt;/sup&gt;</td>
<td>14.9</td>
</tr>
<tr>
<td>Carbonic anhydrase</td>
<td>29</td>
</tr>
<tr>
<td>Bovine Serum Albumin</td>
<td>66</td>
</tr>
<tr>
<td>KdsCΔC8</td>
<td>102.5</td>
</tr>
<tr>
<td>Alcohol Dehydrogenase</td>
<td>150</td>
</tr>
<tr>
<td>β-Amylase</td>
<td>200</td>
</tr>
<tr>
<td>Blue Dextran</td>
<td>2,000</td>
</tr>
</tbody>
</table>

Steady-state kinetics of KdsC – A saturation curve for each reaction was measured and non-linearly fit to the Michaelis-Menten equation (Figure 3-4 shows an example). Three sets of data were recorded and each hyperbolic curve was fit to derive $k_{cat}$ and $K_M$. Mean and standard deviation were calculated for the parameters (Table 3-2).
When the basic kinetic model is assumed (Equation 3-1), $K_M$ value obtained from the Michaelis-Menten equation fitting can be expressed as Equation 3-2. Since both products but not the substrate were captured in the crystal structure of KdsCΔC8, the product release must be the rate-limiting step in the catalytic cycle. Therefore it is logic to deduce that $k_{cat}$ is much smaller than $k_1$, thus $K_M$ should be close to the dissociation constant of the enzyme-substrate complex, $K_D$ (Equation 3-3).

$$
\begin{align*}
E + S & \underset{k_1}{\overset{k_{cat}}{\rightleftharpoons}} E \cdot S \quad \overset{k_{cat}}{\longrightarrow} E + P \\
\end{align*}
$$  

(3-1)

$$
\nu = \frac{k_{cat} E_0 S}{S + \frac{k_1 + k_{cat}}{k_1}} 
$$  

(3-2)

$$
K_M = \frac{k_1 + k_{cat}}{k_1} \approx \frac{k_1}{k_1} = K_D
$$  

(3-3)
The $K_M$ of $\text{KdsC}^{\text{Ec}}$ is close to the one previously reported [5] (72 vs. 75 µM), but the $k_{cat}$ value is 2.6 fold higher (456 vs. 175 s$^{-1}$), probably due to the inaccurate estimation of the enzyme concentration. In the absence of the C-terminal tail sequence, $\text{KdsC}^{\Delta \text{C8}}$ has a similar $K_M$ (99 µM) but ~$10^2$-fold decrease in $k_{cat}$ compared with the full-length $\text{KdsC}^{\text{Ec}}$. Similar $K_M$ values and weak catalytic turnover numbers are also obtained for the native tail-less $\text{KdsC}$, $\text{KdsC}^{\text{Aa}}$, ~$10^3$-fold lower than $\text{KdsC}^{\text{Ec}}$. $\text{KdsC}^{\text{Cb}}$, another native tail-less enzyme, has a slightly higher $K_M$ (227 µM) but also a low $k_{cat}$.

Among all the tested KdsCs, $\text{KdsC}^{\text{Ec}}$ with a tail sequence at the C-terminus has the highest catalytic efficiency, while all the tail-less KdsCs have a consistently low catalytic turnover number although they remain relatively close in $K_M$.

Table 3-2 Michaelis-Menten parameters of KdsCs. All reactions were carried out in 100 mM HEPES pH 7 buffer without salt unless marked with asterisk sign. $\text{KdsC}^{\text{Ec}}$, $\text{KdsC}^{\Delta \text{C8}}$, and $\text{KdsC}^{\text{Cb}}$ were measured at 37°C, and $\text{KdsC}^{\text{Aa}}$ was measured at 60°C. (*$\text{KdsC}^{\text{Ec}}$ measured with addition of 200 mM NaCl.)

<table>
<thead>
<tr>
<th></th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}$ / $K_M$ (µM$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{KdsC}^{\text{Ec}}$</td>
<td>72 ± 9</td>
<td>456 ± 17</td>
<td>6.4 ± 0.8</td>
</tr>
<tr>
<td>$\text{KdsC}^{\text{Ec}}_{\text{NaCl}*}$</td>
<td>148 ± 5</td>
<td>145 ± 2</td>
<td>0.98 ± 0.04</td>
</tr>
<tr>
<td>$\text{KdsC}^{\Delta \text{C8}}$</td>
<td>99 ± 9</td>
<td>2.66 ± 0.08</td>
<td>0.027 ± 0.003</td>
</tr>
<tr>
<td>$\text{KdsC}^{\text{Aa}}$</td>
<td>88 ± 7</td>
<td>0.38 ± 0.02</td>
<td>0.0043 ± 0.0004</td>
</tr>
<tr>
<td>$\text{KdsC}^{\text{Cb}}$</td>
<td>230 ± 30</td>
<td>1.2 ± 0.1</td>
<td>0.0052 ± 0.0008</td>
</tr>
</tbody>
</table>

**Salt effect on KdsC** – To test the hypothesis that the hydrophilic tail regulates the activity of KdsC through a conformational change, $\text{KdsC}^{\text{Ec}}$-catalyzed reactions were incubated at NaCl concentration varying from 0 to 500 mM and measured for...
specific activity (Figure 3-5). Activities of the full-length KdsC<sup>Ec</sup> decrease as the salt concentration increases, retaining only ¼ the activity when salt concentration is 500 mM. Indeed, a saturation curve measurement at 200 mM NaCl revealed a 2-fold increase in $K_M$ and 3-fold decrease in $k_{cat}$ (Table 3-2). However, neither KdsC<sup>ΔC8</sup> nor KdsC<sup>Aa</sup> activities are salt-dependent, suggesting that the ionic strength of the buffer only affects the tail sequence.

**Figure 3-5** Salt effect on KdsC. KdsC activities measured in various concentration of NaCl. Specific activity is defined as µmol of P<sub>i</sub> produced per min per mg of protein.

### 3.5 Discussion

KdsC proteins from different microorganisms align with few gaps in the middle region, from the conserved motif I containing the catalytic aspartate to the
signature sequence near the C-termini. The N-terminal and C-terminal sequences vary from species to species. However, according to gene sequences available in the current database, the C-terminal regions fall into three patterns: Group 1 the random-coiled tail, Group 2 the $\alpha$-helical tail, and Group 3 without a tail, each represented by a crystal structure. The structural comparison reinforces the conspicuous difference in the C-terminal region (Figure 3-1). The phylogenetic tree generated from sequences of assorted KdsCs also supports this classification. To avoid the bias introduced by the varying N-terminal and C-terminal sequence of KdsC, only the middle regions of the sequences were submitted for analysis. The phylogenetic tree clearly revealed a close relationship between enzymes from the same groups. The tree structure also suggests that Group 1 and Group 3 KdsC occur relatively independently in evolution, and the Group 1 random-coiled KdsC thrived to be the largest group while the tail-less KdsC kept evolving to generate Group 2 long-tailed KdsC at some point. Advantages of Group 1 over Group 3 KdsC can be supported by kinetic data.

KdsC$^{Ec}$, a Group 1 KdsC, is an efficient enzyme characterized by a high turnover number ($k_{cat} = 456 \text{ s}^{-1}$). The Group 2 representative, KdsC$^{Hi}$, is also a robust enzyme with a $V_{\text{max}}$ value (420 units/mg) comparable to that of KdsC$^{Ec}$ (480 units/mg) [5]. However, the Group 3 enzymes, KdsC$^{Aa}$ and KdsC$^{Cb}$, retain only $10^{-3}$-fold of the catalytic efficiency ($k_{cat}/K_{M}$, Table 3-2). Based on the structural model discussed in Chapter 2, the extremely low catalytic efficiency of Group 3 KdsC can possibly be attributed to the absence of the solvent-exposed and dynamically movable C-terminal tail.
The tail deletion mutant of KdsC\textsuperscript{Ec}, KdsCΔC8, was constructed to probe the function of the tail. The $K_M$ value of KdsCΔC8 is comparable to that of the full length enzyme, which excludes the possibility of tail involvement in the substrate-binding step. On the other hand, the $k_{cat}$ value of KdsCΔC8 is 200-fold lower than KdsC\textsuperscript{Ec}, suggesting that the tail is critical to either the chemistry step or the following products releasing step. The latter possibility is supported by the crystal structure of KdsCΔC8 that captured both KDO and P$_i$ in the active site with the phosphoester bond cleaved, suggesting that in the absence of a tail, the product release rate declines. The C-terminal amino acids on the tail occupy the same space as the KDO molecule, indicating that the tail possibly plays an active role in competing the products out of the active site to renew the enzyme. The agreement on the low $k_{cat}$ values of both KdsCΔC8 and the native tail-less KdsC\textsuperscript{Aa} strongly indicate that the C-terminal tail is the major factor affecting catalytic efficiency of KdsC.

The hypothesis of the tail affecting the catalytic rate by helping both KDO and P$_i$ release is based on the assumption of the feasibility of conformational change. It is demonstrated in the crystal structures of Mg$^{2+}$-bound KdsC\textsuperscript{Ec} that the tail can exist either in a completely disordered conformation exposed to the solvent or in an ordered conformation locked in the active site of the adjacent monomer via specific interactions with the active site residues, especially two salt-bridges between Arg86 and Ile188 (Chapter 2). The driving force for this conformational change is unknown, but it is possible that the tail distributes between these two states at a certain probability ratio. The ionic strength of the buffer may result in a shift of the
probability distribution between the two states if electrostatic interactions participate in the rate-limiting step. An increase in salt concentration will generate more ions to compete with the tail in the active site binding pocket, making the hydrophilic tail more favorably exposed in the buffer. If the transition of the tail from solvent-exposed to active site-bound state is rate-limiting, higher salt concentration will slow down the product release step. Oppositely, if the conformational change from ordered to disordered state is rate-limiting, concentrated salt in the buffer will accelerate the reaction. Indeed, the activity of KdsCEc decreases as NaCl concentration increases (Figure 3-5), suggesting that the tail insertion is the rate-limiting step. Consistent with this speculation, for KdsCs without a tail, including both the KdsCΔC and KdsCΔa, activities are independent of salt concentration.

For tails to interact in trans with the neighboring active site, a tetrameric structure of KdsC has to be maintained (Chapter 2). KdsC Ec was proven to be a tetramer in solution [5], and the removal of tail did not change the oligomerization state (Table 3-1). However the molecular weight of KdsCΔa determined from gel filtration chromatography suggests a monomeric structure in solution. Although the crystal structure reveals a tetramer, this could possibly be due to the intensive packing of molecules. It is natural to deduce that the substrate binding in KdsCΔa would be lowered in the absence of a neighboring monomer which is known to provide substrate contacts from the structural studies in Chapter 2. However, the KM value of KdsCΔa is within the same level as that of KdsC Ec, indicating that either the KdsCΔa monomers self-assemble upon substrate binding in solution or the specific contacts
provided by the cap monomer do not lower the binding energy significantly. Further mutational studies on the cap monomer are required to clarify these points.

Figure 3-6 Model of KdsC catalytic cycle. Closed monomers are in red while open monomers are in green. Mg$^{2+}$ ions are represented by red dots. KDO8P molecules are represented by long blue squiggles and KDO molecules are short blue squiggles. P$_i$ molecules are green dots.

Based on all the biochemical characterization discussed in this chapter and the structural information revealed in Chapter 2, a model of KdsC catalytic cycle can be proposed as follows (Figure 3-6). Empty monomers of KdsC exist in the equilibrium of open states and closed states, according to the tail-out and tail-in conformation. In a tetramer, the four active sites can adopt all-in, all-out, or diagonal pair-wise in/out states. The substrate KDO8P can only bind to the cleft between the open state monomer and a cap monomer, when the tail from the cap monomer is exposed to the solvent thus generating a passage for the substrate to enter the active site. Substrate binding can induce a conformational change in the active site by rearrangement of two loops to fit the shape of the substrate. A similar induced-fit change also occurs when the tail binds to the active site. The substrate is cleaved in the completely
solvent-exclusive active site, and the KDO sugar molecule is released first. A tail from the cap monomer inserts into the substrate binding cleft, occupies the vacancy left by KDO, and forms similar specific contacts as the KDO does with the residues from both monomers. Finally, the transition from the tail-closed active site to open state is accompanied by P\textsubscript{i} release to generate the reactivated enzyme. Release of KDO and P\textsubscript{i} are possibly rate-limiting, since both products are captured in the active site of the substrate-soaked KdsC\textDelta C8 crystal. In addition, in the absence of tail, the proposed product-release facilitator, the catalytic rate is reduced to a large extent. Tetramers labeled with letters a-g in Figure 3-6 are supported by the corresponding structure snapshots in Figure 2-4.

In conclusion, the C-terminal extension in Group 1 KdsC is confirmed to play a critical role in the catalytic cycle through a conformational change, very possibly by facilitating product release. The discovery of this regulatory mechanism is novel in HADSF systems.

3.6 Future Directions

Tail residues form specific interactions with the active site residues, especially the last three amino acids on the C-terminus, Gln186, Ser187 and Ile188. Removal of the hydrophilic side chain of Gln186 and/or Ser187 may raise the energy of the tail-in state, and cause a decrease in catalytic efficiency. If this statement is true, it will support the speculation that the movement of tail is driven by the distribution of probability into all possible states, otherwise it will suggest another driving force.
Shortening the tail may alter the flexibility of the tail and also reduce the catalytic rate. These hypotheses can be tested by mutagenesis studies. The function of tail can also be verified by adding a tail of KdsC\textsuperscript{Ec} to a tail-less enzyme.

3.7 Acknowledgements

Genomic DNA of \textit{C. burnetii} RSA 493 was a generous gift from Professor Jim Samuel at Texas A&M University. KdsC from \textit{Aquifex aeolicus} was cloned by former graduate student Parag Aggarwal in the Woodard group. I thank Professor Oleg Tsodikov, Dr. Tappan Biswas, and members in the Woodard group for their helpful discussions.
3.8 References


Chapter 4

Native Bifunctional KdsB-KdsC Fusion Protein

4.1 Summary

The carbohydrate KDO (3-deoxy-d-manno-octulosonate) is not only an essential linker molecule that anchors repeating oligosaccharides to lipid A to form the lipopolysaccharide (LPS) layer in Gram-negative bacteria but also a major component in the capsular polysaccharide region. Four enzymes, KdsD, KdsA, KdsC and KdsB, are involved in the LPS KDO biosynthetic pathway. The genes encoding these enzymes are usually dispersed throughout the bacterial genome as opposed to being localized on a common operon. On the other hand, the *kps* genes involved in capsular biosynthesis and assembly commonly reside in a cluster. In the *kps* cluster, only the *kdsD* and *kdsB* homologues of the KDO biosynthetic pathway are present. To date, no gene fusion in either of the pathways has been reported. Here we report a fusion of *kdsB* and *kdsC* genes in *Pseudomonas putida* F1, annotated as *kdsBC*, discovered through a BLAST search of the NCBI database. This gene fusion is unique, and more interestingly, it resides in a gene cluster comprised of both *kds* and *kps* genes. The putative bifunctional enzyme encoded by this naturally occurring fusion gene was cloned, expressed and characterized. The separate KdsB and KdsC domains of KdsBC were also cloned, expressed and
characterized. A combined colorimetric method was established to study the coupled activity of KdsBC by monitoring phosphate released by KdsC and pyrophosphate released by KdsB, independently and simultaneously. The fusion protein displayed both KdsC and KdsB activities with the KdsB displaying the rate-limiting activity. Overall, the covalently linked fusion protein is more efficient than the mixture of individually cloned KdsB and KdsC domains.

4.2 Introduction

The C-terminal tail of *E. coli* KdsC was found to be a structural element that actively participates in the catalytic cycle. The *A. aeolicus* KdsC, which natively lacks the C-terminal tail, displays a lower catalytic activity. I hypothesized that the absence and presence as well as the length of the tail may allow categorization of KdsC into groups. To test this hypothesis, I searched the NCBI database to examine various KdsC sequences. The amino acid sequence alignment revealed three distinctive groups of tails: short tail, long tail, and no tail. Interestingly, in a BLAST search for conserved domains using *E. coli* KdsC as a query, two examples of a gene fusion of KDO8P phosphatase (*kdsC*) and CMP-KDO synthetase (also known as KDO cytidylyltransferase, *kdsB*) were discovered, both annotated as *kdsBC*. One fusion gene was found in *Pseudomonas putida* F1 (locus tag: Pput_3909), and the other was found in *Neisseria meningitidis* serogroup 29E (gene name: *cap29eG*). The protein product of these fusion genes, if unmodified by post-translational cleavage, should lead to a covalently linked KdsB and KdsC on a single polypeptide chain.
Although it is common for genes from the same pathway to form a cluster, the kdsDACB genes in the KDO biosynthetic pathway are usually dispersed throughout the genome, with the occasional exception of kdsC and kdsD which locate sequentially in the genome map of some microorganisms. However, no gene fusion has been previously reported in the pathway. Results from a NCBI database search revealed that out of the 300 entries of microorganisms that carry kds genes, only 10 species have the kdsDACB genes residing closely or forming a gene cluster (Table 4-1, Figure 4-1acd). None of these examples have been studied. It is quite exciting to discover that both the kdsBC fusion genes in my study are found in a kds cluster.

**Table 4-1** List of all microorganisms possessing the kds gene cluster. Data were collected from the NCBI database as of 9/30/08. Locus tag prefixes are shown in parentheses in the first column, with the asterisk marks (*) standing for the numbers listed in column 2-5. The dagger signs (†) mark the fusion genes.

<table>
<thead>
<tr>
<th>Microorganism (Locus tag)</th>
<th>kdsA</th>
<th>kdsB</th>
<th>kdsC</th>
<th>kdsD</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pseudomonas putida F1</em> (Pput_*)</td>
<td>3908</td>
<td>3909†</td>
<td>3909†</td>
<td>3910</td>
</tr>
<tr>
<td><em>Neisseria meningitidis Serogroup 29E (Cap_</em>)*</td>
<td>29eF</td>
<td>29eG†</td>
<td>29eG†</td>
<td>29eH</td>
</tr>
<tr>
<td><em>Ralstonia metallidurans</em> (Rmet_*)</td>
<td>5733</td>
<td>5734</td>
<td>--</td>
<td>5735</td>
</tr>
<tr>
<td><em>Synechococcus elongatus</em> PCC7942 (Synpcc7942_*)</td>
<td>2289</td>
<td>2290</td>
<td>2288</td>
<td>2291</td>
</tr>
<tr>
<td><em>Synechococcus elongatus</em> PCC6301 (syc*_d)</td>
<td>1811</td>
<td>1810</td>
<td>1812</td>
<td>1809</td>
</tr>
<tr>
<td><em>Synechococcus sp. WH8102</em> (SYNW_*)</td>
<td>0180</td>
<td>0184</td>
<td>0186</td>
<td>0187</td>
</tr>
<tr>
<td><em>Synechococcus sp. CC9902</em> (Syncc9902_*)</td>
<td>0206</td>
<td>0209</td>
<td>0210</td>
<td>0211</td>
</tr>
<tr>
<td><em>Synechococcus sp. WH7803</em> (SynWH7803_*)</td>
<td>0231</td>
<td>0233</td>
<td>0237</td>
<td>0238</td>
</tr>
<tr>
<td><em>Methylacidiphilum infernorum</em> (Minf_*)</td>
<td>0954</td>
<td>0956</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><em>Prochlorococcus marinus</em> MIT 9303 (P9303_*)</td>
<td>2556</td>
<td>2553</td>
<td>2551</td>
<td>2550</td>
</tr>
</tbody>
</table>
The eight-carbon carbohydrate molecule, KDO, serves as a linker molecule in the
lipopolysaccharide (LPS) layer in Gram-negative bacterial outer membrane, and
comprises 2% of the LPS [1]. Many G- strains, such as *E. coli*, and some G+ strains, such
as *Bacillus megaterium*, contain an extracellular envelope or capsule attached outside of
outer membrane. The capsule is mainly composed of the K-antigen polysaccharide (KPS)
which is responsible for protecting bacteria against phagocytosis [2, 3]. In some
encapsulated *E. coli* [4-9] and other Gram-negative bacteria [10-15] such as
*Actinobacillus pleuropneumoniae*, *Burkholderia pseudomallei*, *Moraxella
nonliquefaciens*, etc., KDO is found as a common building block in repeating
oligosaccharides and makes up 40-60% of the KPS content. Among the different types of
capsules in *E. coli* strains, those containing KDO in the KPS are classified as Group 2
capsule. Genes responsible for Group 2 capsular biosynthesis and assembly were
annotated as *kps* and cluster in a fixed pattern of *kpsFEDUCS* [16], represented by *E. coli*
CFT073 genes shown in Figure 4-1b. These genes not only cluster together, but also
form enzyme complexes in some examples [17-19], but no gene fusion has been reported.
The functions of the *kpsEDCS* genes remain unclear but are correlated to capsule
expression [17]. The other two genes, *kpsF* and *kpsU*, are homologues to *kdsD* and *kdsB*,
respectively. Each homologous pair shares high sequence (~40%) and functional identity
[16, 20]. KdsB and KpsU are two isoymes of CMP-KDO synthetase (CKS) and
designated as L- (LPS) and K- (KPS) CKS, respectively. Accordingly, KdsD and KpsF
are annotated as L- and K-D-arabinose 5-phosphate isomerase (API), respectively [20]. In
the *kps* cluster, the absence of homologues of the other two essential KDO pathway genes,
*kdsA* and *kdsC*, is quite counterintuitive because it suggests that the only copy of *kdsA*
which is outside the kps cluster needs to produce enough KDO 8-phosphate to supply both the LPS and the KPS KDO biosynthesis, when both the outer membrane and the capsule are present. Interestingly, the heavy metal-dependent G- bacterium, Ralstonia

a) S. elongatus PCC6301

\[ \text{kdsD kdsB kdsA kdsC} \]

b) E. coli CFT073

\[ \text{kpsF kpsE kpsD kpsU kpsC kpsS} \]

c) R. metallidurans

\[ \text{kpsD kpsE kdsA kpsU kpsF kpsC kpsS} \]

\[ \text{Fusion gene} \]

d) P. putida F1

\[ \text{kpsE kpsD} \]

\[ \text{kdsA kdsBC kpsF kpsC kpsS} \]

Figure 4-1 Organization of kds and kps gene clusters. a) kds cluster of Synechococcus elongatus PCC6301 is shown as an example of Synechococcus strains that possess kds cluster without insertions of other genes. b) kps cluster of E. coli CFT073 is shown as an example of group 2 capsule gene organization. c) A kdsA gene inserts into the kps cluster on genome of Ralstonia metallidurans. The blue box highlights the genes that have KDO biosynthetic pathway activities. d) A kdsA and a kdsC insert into the kps cluster on the genome of Pseudomonas putida F1 with the latter fused with a kpsU gene. The fusion gene is annotated as kdsBC. The blue box highlights a complete set of genes that have all four KDO biosynthetic pathway activities.

metallidurans [21], has a kdsA inserted into the kps cluster (Figure 4-1c). This insertion is unique. This kdsA gene is the only copy of kdsA in the genome and has yet to be characterized. Also, the level of KDO expression in this strain is unknown. Therefore, it is not clear how this enzyme works to produce enough KDO for both polysaccarides. Even more intriguing, a complete set of kds genes seems to incorporate into a kps cluster
in *P. putida* F1, organized as *kdsA* and *kdsBC* fusion in the forward orientation, and *kdsD* in the reverse orientation (Figure 4-1d). After comparing these genes to their homologues in *E. coli* CFT073 (Table 4-2), the *kdsB* portion of the *kdsBC* is shown to share more sequence identity with *E. coli* *kpsU* than *kdsB* (Figure 4-2a). Similarly, the *P. putida* F1 *kdsD* in the cluster should instead be annotated as *kpsF* (Figure 4-2b). Therefore, the overall gene organization in this region is a full set of *kps* genes into which a *kdsA* and a *kdsC* gene has been inserted, with the latter fused to a *kpsU*. The fusion gene will still be annotated as *kdsBC* for convenience.

**Table 4-2** Comparision of each *kds* gene from the *kps/kds* gene cluster in *P. putida* F1 to its homologue(s) in *E. coli* CFT073 via pairwise sequence alignment.

<table>
<thead>
<tr>
<th></th>
<th><em>P. putida</em> F1 Copy 1</th>
<th><em>E. coli</em> CFT073</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>KdsD</td>
<td>Pput_3910</td>
<td>c3957 (KdsD)</td>
<td>46%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c3686 (KpsF)</td>
<td>52%</td>
</tr>
<tr>
<td>KdsA</td>
<td>Pput_3908</td>
<td>c1674</td>
<td>72%</td>
</tr>
<tr>
<td>KdsC</td>
<td>Pput_3909</td>
<td>c3958</td>
<td>38%</td>
</tr>
<tr>
<td>KdsB</td>
<td>Pput_3909</td>
<td>c1058 (KdsB)</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c3687 (KpsU)</td>
<td>61%</td>
</tr>
</tbody>
</table>

Figure 4-2 Unrooted phylogenetic tree of KdsB/KpsU and KdsD/KpsF. **a)** Comparison of KpsU and KdsB from *E. coli* CFT073 to KdsBs from *P. putida* F1; KdsB1 is part of the sequence of KdsBC fusion protein. **b)** Comparison of KpsF and KdsD from *E. coli* CFT073 to KdsDs from *P. putida* F1; KdsD1 is the copy in the *kds/kps* cluster. The phylip-format dendrograms are generated by Biology WorkBench 3.2.
P. putida F1, a Gram-negative bacterium originally isolated from a polluted creek in Urbana, IL, is capable of growing on a wide range of aromatic hydrocarbons such as toluene, benzene and trichloroethylene, thus detoxifying these environmental pollutants [22-24]. Besides the kdsA and kdsC inserted into the kps cluster, the F1 substrain carries a second set of all four kds genes which are dispersed throughout the genome. Each homologous pair of genes shares high degree of sequence identity with each other (Table 4-3). Possessing two complete sets of kds genes is unique. All other substrains of P. putida with complete and available genomic data possess only the dispersed set of kds genes, and have no fusion gene. The clustered kds genes are likely to be the capsular polysaccharide KDO biosynthetic genes since they reside in a kps cluster. On the other hand, the dispersed kds genes can be assigned as the LPS KDO biosynthetic genes, because both kdsB and kdsD from this set share more sequence identity with the kds genes instead of the kps homologues in E. coli CFT073 (Figure 4-2).

Table 4-3 Comparison of the two sets of KDO biosynthetic enzymes from Pseudomonas putida F1 via pairwise sequence alignment. Genome locus tag and length of peptide chain are listed for each enzyme.

<table>
<thead>
<tr>
<th></th>
<th>P. putida F1 Copy 1</th>
<th>P. putida F1 Copy 2</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>KdsD/KpsF</td>
<td>Pput_3910</td>
<td>Pput_0996</td>
<td>49%</td>
</tr>
<tr>
<td>KdsA</td>
<td>Pput_3908</td>
<td>Pput_4166</td>
<td>70%</td>
</tr>
<tr>
<td>KdsC</td>
<td>Pput_3909</td>
<td>Pput_0995</td>
<td>39%</td>
</tr>
<tr>
<td>KdsB/KpsU</td>
<td>Pput_3909</td>
<td>Pput_3812</td>
<td>48%</td>
</tr>
</tbody>
</table>
The *N. meningitidis* serogroup 29E is known to be relatively low in pathogenicity compared to other serogroups of *N. meningitidis* [25], however it is less susceptible to penicillin than other *N. meningitidis* serogroups [26]. The 29E strain is the only serogroup in *N. meningitidis* species that contains KDO in the capsular polysaccharides [27]. The four *kds* genes lay tandem on the genome, similar to the gene organization in the *kds/kps* cluster in *P. putida* F1. However, in *N. meningitidis* serogroup 29E, *kdsD* and *kdsB* genes are present instead of *kdsF* and *kpsU*, and no *kps* cluster is observed in the region near these *kds* genes in *N. meningitidis* serogroup 29E. Since this strain has not been completely sequenced, there is no evidence for the existence of a set of dispersive *kds* genes.

The *kdsBC* from *P. putida* F1 was chosen to study because of the uniqueness of the potential fusion of two *kds* genes, as well as the special location of the gene as described above. This gene may represent a critical evolutionary stage of the KDO biosynthetic pathway. The protein product of this gene may display fused KdsC and KdsB activities with intriguing catalytic properties. In addition, this putative bifunctional enzyme could potentially serve as a model for studying protein-protein interaction in the KDO biosynthetic pathway.

![Figure 4-3 Scheme of sequential reactions catalyzed by KdsC and KdsB.](image-url)
The putative bifunctional protein should catalyze first the hydrolysis of KDO8P to release one equivalent of phosphate (P$_i$) and KDO, followed by reaction of the KDO with CTP to allow the addition of CMP onto KDO. The latter transformation should be accompanied by the release of one equivalent of pyrophosphate (PP$_i$) (Figure 4-3). Phosphatase activities are usually measured by monitoring P$_i$ release using various end point colorimetric [28, 29] and continuous methods [30, 31]. The CMP-KDO synthetase activities can be measured by following CMP-KDO production via a modified Aminoff assay ([32]). This multi-step method eliminates the contamination of the substrate KDO by reducing it with sodium borohydride, leaving only the KDO portion of CMP-KDO (the KDO in this molecule is stable to borohydride treatment) to generate a colorimetric signal through periodate oxidation followed by thiobarbituric acid addition. To avoid this time-consuming method, I chose to monitor PP$_i$ production in the reaction. Generally, PP$_i$ can be converted into two molecules of P$_i$ by inorganic pyrophosphatase and analyzed as total P$_i$ [33-35]. Alternatively, the PP$_i$ producing enzyme can be coupled to PP$_i$-dependent fructose-6-phosphate kinase, fructose-1,6-diphosphate aldolase, triosephosphate isomerase, and glycerophosphate dehydrogenase consecutively, and finally monitored by the absorbance change at 340 nm for the oxidation of NADH to NAD$^+$ [36, 37]. A molecule of P$_i$ is generated as a by-product in the reaction catalyzed by the fructose-6-phosphate kinase. To assay KdsBC, I wanted to monitor both P$_i$ and PP$_i$ production to acquire information for both activities simultaneously. The methods of PP$_i$ detection mentioned above all generate extra P$_i$, and therefore would interfere with the P$_i$ assay for the KdsC portion of the reaction, thus losing the information about P$_i$ generated from the KdsC
reaction. Developing an assay capable of determining the amounts of Pₐ and PPₐ from a reaction mixture was a major experimental challenge.

The strategy chosen was to combine two colorimetric assays: 1) the commonly used malachite green assay (MG) that detects only Pₐ [38, 39], and 2) the colorimetric assay based on ammonium molybdate and eikonogen reagent (EK) that detects both Pₐ and PPₐ independently and “simultaneously” [29, 40] (Figure 4-4). If a sample is to be analyzed for both Pₐ and PPₐ, Pₐ can be directly determined using the MG assay. The PPₐ can be calculated by subtracting the concentration of Pₐ determined by MG from the EK signal which is the combined signal of both Pₐ and PPₐ. However, it was unclear whether Pₐ and PPₐ could compete for any reagent in the color development mixture and thus interfere with each other in chromophore formation. Only when the chromophore formations of Pₐ and PPₐ are independent can both species correlate linearly to absorbance. Also, since the combination of two assays, MG and EK, may result in error accumulation, the accuracy of PPₐ determination should be confirmed.

![Scheme of reactions in a) EK color development and b) the MG assay](image)

**Figure 4-4** Scheme of reactions in a) EK color development and b) the MG assay
With the enzyme of interest and assay method identified, the bifunctional KdsBC from *Pseudomonas putida* F1 (annotated as PpBC) was cloned, overexpressed and fully characterized. Each individual domain (annotated as PpB for the KdsB domain and PpC for the KdsC domain, Figure 4-5) was also cloned and isolated, and characterized to compare with the fusion protein.

**Figure 4-5** Two proposed domains in PpBC and the location of artificial division of PpB and PpC.

### 4.3 Experimental Procedures

**Materials** – The *Pseudomonas putida* F1 strain (ATCC 700007) was purchased from American Type Culture Collection. Malachite green carbinol hydrochloride, ammonium molybdate tetrahydrate, sodium sulfite, sodium metabisulfite, Trizma® base, sodium pyrophosphate tetrabasic decahydrate, pyrophosphate reagent and bovine serum albumin (BSA) were purchased from Sigma-Aldrich company. Tris(hydroxymethyl) aminomethane was from Research Organics. Bacto™ tryptone was from Becton, Dickinson and Company. Nutrient broth powder was from Difco Laboratories, Inc. Anhydrous sodium phosphate monobasic and 2-mercaptoethanol were obtained from J. T.
Baker Company. Isopropyl β-D-1-thiogalactopyranoside was from Gold Biotechnology, Inc. Sulfuric acid, hydrochloric acid, yeast extract, sodium chloride, potassium chloride, ammonium sulfate, ampicillin sodium salt, Nunc™ polystyrene 96-well flat bottom assay microplates, and Fisherbrand polystyrene disposable plastic cuvettes were from Fisher Scientific Inc. Polypropylene 96-well half-skirted reaction PCR plates were obtained from DOT scientific, Inc. The 1-amino-2-naphthol-4-sulfonic acid was purchased from Mallinckrodt Chemical Works. PCR primers were synthesized by Invitrogen Corporation. FailSafe™ pre-mix from Epicenter was used as PCR buffer. Plasmid extraction was performed using Wizard® Plus SV Minipreps DNA purification kit from Promega Corporation. Restriction enzymes, calf alkaline phosphatase, Vent DNA polymerase, Pfu DNA polymerase and T4 ligase were purchased from New England Biolabs, Inc. DNA sequencing was conducted by the University of Michigan Biomedical Resources Core Facility. Bradford Protein Assay Reagent, Mini-PROTEAN II electrophoresis unit, and microplate reader Model 550 were from Bio-Rad Laboratories, Inc. The Q-sepharose resin, Phenyl Superose (HR 10/10) column and FPLC® system were from Pharmacia, GE Healthcare. XL1-Blue E. coli strain was from Stratagene, Agilent Technologies. BL21(DE3) E. coli strain and HisBind® Ni-NTA resin were from Novagen, EMD Chemicals Inc. High grade spectra/Por® 7 dialysis tubing (10,000 and 15,000 Da molecular weight cut-off and metal free) was from VWR International, LLC. The Millex® syringe driven filter units (0.22 µm) was obtained from Millipore. The 8453 UV-vis diode-array spectrophotometer was from Hewlett Packard. All the programmed temperature controls were performed on a MJ research PTC-200 Peltier thermal cycler.
**Sequence analysis** – Database searches of multiple microbial organisms were performed utilizing the BLAST program at the NCBI website (http://ncbi.nlm.nih.gov/BLAST). Multiple sequence alignments were generated using Clustal W (http://www.ebi.ac.uk/clustalw).

**Protein Concentration Assay** – Protein concentration was determined using the Bradford assay by mixing 1 mL dye with up to 10 µL of protein samples. BSA served as a standard. Alternately, ultra-violet absorbance of a protein solution measured at 280 nm was converted into protein concentration, based on the extinction coefficient calculated utilizing the ProtParam tool (ExPASy).

**Conventional tube-based MG/EK assay** – Reagents for the malachite green assay were prepared according to Lanzetta et al. [40]. Briefly, 270 mg malachite green powder was dissolved in 600 mL deionized water (ddH₂O), and mixed with 200 mL of an aqueous solution containing 8.4 g ammonium molybdate and 67 mL 12N HCl. This final mixture (MG reagent) was stirred for 1 h at room temperature and passed through Whatman No.5 filter paper before being stored at 4°C for future use. MG reagent (900 µL) was added to an eppendorf tube containing 100 µL of sample solution and allowed to react for 5 min for color development. The absorbance was measured at 660 nm.

The three reagents used in the EK assay were prepared as previously reported [38, 39]. Briefly, the ammonium molybdate (AM) solution was prepared by dissolving 2.5 g ammonium molybdate in 80 mL ddH₂O, followed by addition of 13.6 mL 98% H₂SO₄. The volume was taken to 100 mL with ddH₂O. The 0.5 M β-mercaptoethanol (β-ME) solution was prepared by diluting 350 µL β-ME to 10 mL with ddH₂O. Both the AM and β-ME solutions were stored at 4°C. Finally, the eikonogen reagent was prepared by
dissolving 0.25 g of sodium sulfite, 14.65 g of meta-bisulfite, and 0.25 g of 1-amino-2-naphthol-4-sulfonic acid in 100 mL ddH₂O and stored in an amber bottle at room temperature. The eikonogen reagent was stirred and warmed on a hot plate each time before using to redisolve the salt crystals. To an eppendorf tube containing 380 µL of sample solutions was added 50 µL AM, 50 µL β-ME, and 20 µL eikonogen reagent sequentially. Color was developed by incubating the tube in a 37°C water bath for 10 min. The absorbance was measured at 580 nm.

96-well plate format MG/EK combined assay – The MG assay reagent for the 96-well plate assay is more concentrated than the MG reagent for the conventional tube-based assay. First, 162 mg malachite green powder was dissolved in 180 mL ddH₂O. Second, 5.04 g ammonium molybdate was mixed with 55 mL of 12N HCl and diluted to 60 mL with ddH₂O. These two solutions were mixed by stirring in a glass graduated cylinder for 1 h and filtered as described above. The MG reagent (200 µL) was added to each well of a 96-well plate containing 20 µL of sample. The plate was incubated at room temperature for 5 min and read at 655 nm.

The plate-based EK assay reagents were prepared as described for the conventional tube-based EK assay. To 50 µL of the enzyme reaction mixture in an individual well of a 96-well reaction PCR plate was added sequentially 50 µL AM, 50 µL β-ME, and 20 µL eikonogen reagent. The mixture was incubated in a thermal cycler at 37°C for 10 min and then transferred to a reading plate for absorbance measurement at 540 nm.

A standard 1 mM phosphate solution was prepared utilizing anhydrous NaH₂PO₄. A standard 1 mM pyrophosphate solution was prepared from NaP₂O₇·10H₂O. Both
reagents were stored at 4°C. Serial dilutions of each standard were prepared to contain 0-40 nmol P\textsubscript{i} or PP\textsubscript{i} in 100 µL aqueous solution. These standards were divided into two separate portions (40 µL each), one for the malachite green assay and one for the eikonogen assay.

The 96-well plate format determination of P\textsubscript{i} and PP\textsubscript{i} in a mixture – From a 100 µL sample containing both P\textsubscript{i} and PP\textsubscript{i}, 20 µL was transferred for the MG assay and 50 µL was transferred for the EK assay. The difference in sample size is due to the different sensitivities of the two assays. The concentration of P\textsubscript{i} was directly determined from the MG assay according to equation 4-1. The concentration of PP\textsubscript{i} was solved from equation 4-2 using the P\textsubscript{i} concentration calculated from equation 4-1. \(A_{655}\) was the absorbance reading of the MG assay and \(A_{540}\) was the absorbance reading of the EK assay.

All extinction coefficients (\(\varepsilon\)) were determined from calibration curves.

\[
A_{655} = \varepsilon_{\text{Pi-655}} [P_i] \tag{4-1}
\]

\[
A_{540} = \varepsilon_{\text{Pi-540}} [P_i] + \varepsilon_{\text{PPi-540}} [PP_i] \tag{4-2}
\]

Pyrophosphate assay – A commercial kit from the Sigma-Aldrich company was used to determine the PP\textsubscript{i} concentration in a mixture of P\textsubscript{i} and PP\textsubscript{i} [41]. The kit consists of a series of enzymes that correlate the PP\textsubscript{i} concentration to the disappearance of the NADH signal at 340 nm. The result from this method was used to verify the accuracy of the PP\textsubscript{i} concentration measured by MG/EK method. Three mixtures of P\textsubscript{i} and PP\textsubscript{i} at different ratios were prepared. A vial of lyophilized pyrophosphate reagent as supplied was reconstituted with ddH\textsubscript{2}O to 6 mL. To 50 µL of sample, 950 µL of pyrophosphate reagent solution was added and incubated at room temperature for 10 min. Absorbance readings at 340 nm were converted into concentration of NADH (\(\varepsilon = 6220 \text{ M}^{-1}\text{cm}^{-1}\)) first
and next into concentration of PP\textsubscript{i}, according to 1:1 stoichiometric ratio of NADH and PP\textsubscript{i}. The same samples were tested using the MG/EK assay. From each 100 µL sample, 20 µL was transferred for the MG assay and 50 µL was transferred for the EK assay. P\textsubscript{i} and PP\textsubscript{i} concentrations were calculated as described above.

**Cloning, overexpression and purification of PpBC** – The *P. putida F1* cells were recovered by suspending approximately 50 mg of cell paste into 5 mL nutrient broth media, and incubating the mixture in a 37°C shaker overnight. A 200 µL aliquot from the overnight cell culture was spread on a nutrient broth agar plate and then incubated at 37°C for 16 h. The Pput_3909 gene (gi148549115) was amplified by PCR using a single colony from the plate as the template. The forward primer was GATTCTAGAATTCATATGGCAAATTCTGAACAGCGATTG, and the reverse primer was AATTCGGATCCAAGCTGCAGTCATTGCGCCATATTC. The PCR mixture contained 25 µL buffer K, 1 µL of each primers (OD\textsubscript{260}~15), the colony as a template, and sterile ddH\textsubscript{2}O up to 49 µL. After heating the reaction mixture at 95°C for 2.5 min, 1 µL Vent DNA polymerase was added, followed by 30 cycles consisting of denaturation at 94°C for 1 min, annealing at 50°C for 1 min and extension at 72°C for 1.5 min. After size verification by horizontal electrophoresis on 1% (w/v) agarose gel, the PCR product was digested with *NdeI* and *BamHI*. The expression vector pT7-LOH was similarly double digested and treated with calf intestinal alkaline phosphatase. The PCR product and the vector were ligated by T4 ligase at room temperature overnight. The ligation was transformed into chemically competent XL1-blue *E. coli* cells, which were then spread on an agar plate of LB supplemented with 100 mg/L ampicillin (LB/amp). A single colony containing the pT7-LOH-*PpBC* plasmid, verified by DNA sequencing, was
amplified and stored at -80°C in 25% glycerol. The pT7-7-PpBC plasmid was constructed in the same manner.

The pT7-LOH-PpBC plasmid was transformed into chemically competent BL21(DE3) E. coli cells for expression. The cells were grown in 1 L LB/amp at 37°C with shaking at 230 rpm. Isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 0.2 mM when the cell culture reached OD600 of 0.5. The cells were allowed to grow at 17°C after induction for another 18 h, before being harvested by centrifugation at 6000 rpm in a Beckman JA-10 rotor for 10 min. The cell paste was resuspended in HisBind® (Novagen) binding buffer (20 mM Tris pH 8.0, 500 mM NaCl, 5 mM imidazole), and sonicated in an ice-water bath for 30 s for 4 pulses, with 2 min intervals. The crude extract was then separated from the cell debris by centrifuging at 29000 g for 20 min.

The supernatant obtained from above (15 mL, 150 mg protein in total) was treated with 15 µL of DNase I and RNase A, each 10 mg/mL, at room temperature for 10 min. After filtration through a 0.22 µm Millex® membrane, the crude extract was mixed with 4 mL HisBind® resin (Novagen) in a 50 mL centrifuge tube. The resin was saturated with his-tagged protein by rotating the mixture on Labquake® rotator (Barnstead International) at a low speed in 4°C overnight. The resin/protein mixture was then transferred in a Poly-Prep® column (Bio-Rad) for washing, and eluted using column chromatographic techniques. The target protein was eluted in the binding buffer with 250 mM imidazole, and immediately dialyzed against 20 mM Tris pH 7.5. The resultant protein was homogenous as determined by SDS-PAGE with a total yield of 20 mg/liter of cell culture, and concentrated to 5.5 mg/mL before storage at -80°C.
Cloning, Overexpression, and Purification of PpB and PpC – A short nucleotide containing a stop codon and an Ndel site (TAAACATATG) was inserted into pT7-LOH-PpBC after the kdsB gene. The mutagenesis PCR was performed using pT7-LOH-PpBC plasmid as the template, and the primer pair GCCTCTGACTTCATAAACATATGAAACCGAAACTGG and CCAGTTTCTGTTTCATATGTTTATGAAGTCAGAGGC. Pfu ultra DNA polymerase was used because of its high fidelity and efficiency in DNA replication. The PCR product was subjected to *DpnI* digestion at 37°C for 1 h to degrade the original template, and then transformed into XL1-blue *E. coli* cells for plasmid amplification. The plasmid obtained is designated pT7-LOH-PpB. Similarly, a pT7-7-PpB plasmid was also constructed based on pT7-7-PpBC. The pT7-7-PpB plasmid was digested by *Ndel* and self-ligated between two *Ndel* sites, one before the start codon and the other after the stop codon, resulting in the excision of PpB gene and formation of a pT7-PpC plasmid. All plasmids were sequenced.

The pT7-LOH-PpB and pT7-PpC plasmids were each transformed into BL21(DE3) cells respectively for protein expression. Both PpB and PpC proteins were overexpressed the same way as described for PpBC. PpB bearing a his-tag was purified according to the same protocol as that of PpBC. A yield of 60 mg/L of cell culture was obtained. The homogenous PpB solution was concentrated to 7.1 mg/mL and stored at -80°C.

Cells with overexpressed PpC were sonicated in buffer A (20 mM Tris pH 7.5). The clarified supernatant was applied to Q-sepharose equilibrated with buffer A. A linear gradient elution was performed from 0 to 2M KCl in 90 min. PpC was eluted in fractions containing approximately 0.4 M KCl, and verified by SDS-PAGE. The pooled fractions
were taken to 20% (w/v) ammonium sulfate by adding ammonium sulfate crystals before loading to a phenyl superpose column (HR 10/10). Protein was eluted using a reverse gradient from 20% to 0% (w/v) ammonium sulfate. The target protein was collected in fractions with around 3% ammonium sulfate. The pooled homogenous PpC was dialyzed against buffer A and concentrated to 3.6 mg/mL before stored at -80°C. The total yield was 30 mg/L of cell culture.

pH dependence – Buffers used were sodium acetate pH 4-5, MES pH 5.5-6.5, HEPES pH 7-8, Tris pH 8.5-9.5, and glycine-NaOH pH 10-11.5 with pH values adjusted at 37°C. All reaction mixtures contained 100 mM buffer, 5 mM MgCl\(_2\), substrate(s), and enzyme in a final volume of 50 µL. The reaction was initiated by adding 10 µL enzyme solution as will be specified, incubated at certain temperature for certain amount of time as will be specified, and stopped by quenching with 50 µL 100% ice-cold ethanol. The phosphatase activity reactions catalyzed by PpBC and PpC were annotated as PpBC(KDO8P) and PpC(KDO8P), respectively. Specifically, the phosphatase assay included 0.8 mM KDO8P as the substrate and 5 µg PpBC or 7 µg PpC as the enzyme. The reactions were incubated at 37°C for 1 min before quenching. The cytidyltransferase reactions catalyzed by PpBC and PpB were accordingly marked as PpBC(KDO/CTP) and PpB(KDO/CTP), respectively. Typically, two substrates, 0.8 mM KDO and 5 mM CTP, and 5 µg PpBC or 0.7 µg PpB were included. The reaction mixtures were incubated at 50°C for 5 min. The overall reaction is referred to as PpBC(KDO8P/CTP), with 0.8 mM KDO8P and 5 mM CTP supplied as substrates and 5 µg PpBC included. The reaction mixtures were also incubated at 50°C for 5 min. Out of each 100 µL quenched reaction, 20 µL was utilized for MG assay and 50 µL was transferred for EK assay. The amounts
of P\textsubscript{i} and PP\textsubscript{i} were derived as described in the 96-well plate format determination of P\textsubscript{i} and PP\textsubscript{i}. The signal was expressed as specific activity in unit/mg, which was defined as µmol of product (P\textsubscript{i} or PP\textsubscript{i}) produced per minute per mg of enzyme.

**Temperature profile and thermostability** – Each reaction included 100 mM glycine-NaOH pH 10, 5 mM MgCl\textsubscript{2}, 0.8 mM KDO8P or KDO, 5 mM CTP if needed, and enzyme as specified in the “pH dependence” section in a total volume of 50 µL. To measure temperature profile, reactions were carried out at a fixed temperature in the range of 20-90°C, with 10°C intervals. The quenched reaction mixtures were divided for color development as stated above. For thermostability measurements, enzymes were first incubated for 5 min at a single temperature ranging of 20-90°C (with 10°C intervals) and then added to pre-warmed reaction mixtures for 1 min incubation at 50°C.

**Substrate specificity** – Various phosphosugars or aromatic compounds were used as substituent for KDO8P. Reactions were carried out with 100 mM glycine-NaOH pH 10, 5 mM MgCl\textsubscript{2}, 0.8 mM of the corresponding substrate, 5 mM CTP, and 5 µg of PpBC in a total volume of 50 µL. After incubated at 50°C for 5 min, the reaction mixtures were quenched and analyzed as stated in the pH-dependence measurement section.

**Magnesium dependence** – The Mg\textsuperscript{2+} dependence of the PpBC(KDO8P) activity was measured by mixing 100 mM glycine-NaOH pH buffer, 1 mM KDO8P, and 0-10 mM Mg\textsuperscript{2+} and ddH\textsubscript{2}O up to 50 µL. The reaction mixtures were incubated at 50°C for 1 min before subject to colorimetric assays. For PpBC(KDO/CTP) activity, the reaction mixtures contained 3 mM KDO, 5 mM CTP, and 0-20 mM Mg\textsuperscript{2+} at pH 10, and were incubated at 50°C for 5 min before analysis.
**Metal analysis** – The isolated PpBC, PpB and PpC proteins were treated with 10 mM EDTA at 25°C for 2 h and then dialyzed against 20 mM Tris pH 7.5 for 24 h with two changes of buffer. The metal-depleted apo-enzymes were then treated with various divalent metal ions, including Mg$^{2+}$, Mn$^{2+}$, Co$^{2+}$, Zn$^{2+}$, Ba$^{2+}$, Ni$^{2+}$, Hg$^{2+}$, Cd$^{2+}$, and Ca$^{2+}$, and assayed for activity. All reactions were carried out in 100 mM glycine-NaOH buffer at pH 10. The PpC was assayed with 0.8 mM KDO8P and 1 mM metal ion, and measured for P$_i$ release. The PpB was supplemented with 0.8 mM KDO, 5 mM CTP and 5 mM metal ion, and measured for PP$_i$ release. The reaction mixture for the bifunctional enzyme PpBC included 0.8 mM KDO8P, 5 mM CTP and 5mM metal ion, and was measured for both P$_i$ and PP$_i$ production. All buffers were made with Trizma, metal-free HCl and ddH$_2$O.

**Kinetic measurements** – Steady-state kinetics were measured under optimum conditions for each enzyme, with one substrate concentration varying at each time. The PpBC(KDO8P), PpC(KDO8P) were assayed at pH 10 with an incubation at 50°C for 1 min. The PpBC(KDO/CTP), PpB(KDO/CTP) and PpBC(KDO8P/CTP) were assayed at pH 10 with an incubation at 50 °C for 5 min. Amounts of enzymes used were specified in the “pH dependence” section. The resultant saturation curves were fit nonlinearly to the Michaelis-Menten equation using Origin software to derive values of $V_{max}$, $K_M$ and $k_{cat}$. All data points are triplicates.

**Time course** – The fusion enzyme PpBC was subjected to a time course measurement. A mixture of PpB and PpC at a 1:1 molarity ratio was also measured for time dependence. The reactions contained 100 mM Na-glycine pH 10, 5 mM MgCl$_2$, 5 mM CTP, 0.8 mM KDO8P, and 10 µL enzyme solution containing 3 µg of PpBC or
PpB/PpC in the mixture. Each 50 µL reaction was initiated by the addition of enzyme solution, and incubated at 50 °C. The reactions were quenched after 1, 2, 5, 10, 20, 30, 40 and 60 min, and divided for separate MG and EK color development.

4.4 Results

**Assay development** – A combination of malachite green and eikonogen (MG/EK) colorimetric measurement was used for detecting $P_i$ and $PP_i$ in a reaction mixture.

![Figure 4-6](image)

**Figure 4-6** Calibration curves of the malachite green assay and eikonogen assay in the conventional tube-based format. For a mixture of $P_i$ and $PP_i$, total eikonogen signal is the sum of both $P_i$ and $PP_i$ signal, shown as a hypothetic line in the eikonogen assay plot.

Calibration curves of MG/EK were first obtained in a conventional tube-based format. Samples of $P_i$ and $PP_i$ were prepared in a range between 0-40 nmol in a volume of 100 µL, and equally divided for the MG and EK assays. All data points were measured in triplicate. Linear curves were observed for all calibrations (**Figure 4-6**). In the MG calibration curve, $PP_i$ showed negligible absorbance at 660 nm, whereas in the EK assay both $P_i$ and $PP_i$ exhibit measurable signals. The slope of each curve was derived from a
linear fit equation and converted to an extinction coefficient with units of M⁻¹cm⁻¹. The extinction coefficients are listed in Table 4-4.

Table 4-4 Extinction coefficients (ε) of Pᵢ and PPᵢ for the MG or EK assay.

<table>
<thead>
<tr>
<th></th>
<th>εᵢ (M⁻¹cm⁻¹)</th>
<th>εPPᵢ (M⁻¹cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube-based</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG 660 nm</td>
<td>2760</td>
<td>2</td>
</tr>
<tr>
<td>EK 580 nm</td>
<td>765</td>
<td>3225</td>
</tr>
<tr>
<td>96-well plate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG 655 nm</td>
<td>6700</td>
<td>44</td>
</tr>
<tr>
<td>EK 540 nm</td>
<td>880</td>
<td>3470</td>
</tr>
</tbody>
</table>

Assays were also carried out in a 96-well plate format to more easily include systematic controls and to create a format for future high throughput screening. Calibration curves of the MG and EK assays in a 96-well plate format were obtained. Individual samples of Pᵢ and PPᵢ were prepared at concentrations between 0-0.4 mM, and each divided equally for the MG (40 µL) and EK (40 µL) assays. All data points were collected in triplicate. Linear curves obtained followed the same pattern as measured with the conventional tube-based method (Figure 4-7). Extinction coefficients are reported in M⁻¹ units because the light path in the absorbance measurement differs with sample volume in the wells (Table 4-4). Since the MG is two-fold more sensitive than EK assay in 96-well plate measurements (εᵢ₆₅₅ ≈ 2εPPᵢ₅₄₀), the reaction mixtures to be analyzed were divided unevenly in measurements: typically 20 µL for MG and 50 µL for EK from every 100 µL total reaction mixture. The extinction coefficient was calculated from a four-point standard curve on each plate to reduce error from buffer absorbance, plate absorbance, duration of color development, and light path variation.
Figure 4-7 Calibration curves of malachite green assay and eikonogen assay in the 96-well plate format. For a mixture of \( P_i \) and \( PP_i \), total eikonogen signal is the sum of both \( P_i \) and \( PP_i \) signal, shown as a hypothetic line in the eikonogen assay plot.

The determination of \( P_i \) and \( PP_i \) concentrations in a mixture is illustrated in Figure 4-8. A binary mixture of \( P_i \) and \( PP_i \) at 1:1 was prepared and serially diluted to samples containing 0-0.2 mM \( P_i \) and \( PP_i \) each. The absorbance is linearly correlated to sample concentrations. Since the MG absorbance reading is only contributed by \( P_i \) (Figure 4-8a), the concentration of \( P_i \) can be directly calculated from the MG signal using an extinction coefficient deduced from a standard curve (Figure 4-8b). The total EK absorbance reading contributed by both \( P_i \) and \( PP_i \) is supposed to be a linear combination of \( P_i \) and \( PP_i \) concentrations, each weighted by an extinction coefficient (Figure 4-8c). With the amount of \( P_i \) determined from the MG assay, the portion of the EK signal derived from the \( PP_i \) can be deduced by subtracting the absorbance due to the \( P_i \) from the total EK signal. Finally, the amount of \( PP_i \) can be calculated from this portion of the EK signal contributed by \( PP_i \) alone (Figure 4-8d). The amounts of \( P_i \) and \( PP_i \) in 50 \( \mu \)L volume determined from the experimental data match well with the theoretically added values (Figure 4-8bd), suggesting that the assay is accurate.
The accuracy of the PP<sub>i</sub> measurement was further confirmed using a commercially available pyrophosphate reagent kit from Sigma. This kit detects PP<sub>i</sub> generated from the reaction mixture by reacting the PP<sub>i</sub> with fructose 6-phosphate catalyzed by PP<sub>i</sub>-fructose-6-phosphate kinase. The resultant fructose-1,6-diphosphate is further converted to dihydroxyacetone phosphate through two enzymatic reactions catalyzed by aldolase and triosephosphate isomerase. Finally, the reduction of
dihydroxyacetone phosphate is coupled to the conversion from NADH to NAD⁺, causing a decrease of NADH signal at 340 nm. All the coupling enzymes mentioned above and substrates except PPᵢ are included in the pyrophosphate reagent as supplied. Samples with random amount of Pᵢ and PPᵢ were “blindly” prepared and subjected to both the MG/EK coupled assay and the Sigma pyrophosphate assay kit. Results listed in Table 4-5 show that these two methods agree on pyrophosphate detection with only small errors, and are both close to the true values. These results indicate that the coupled MG/EK assay is an accurate and reliable method for Pᵢ and PPᵢ measurement.

**Table 4-5** Comparison of the Sigma pyrophosphate reagent to the coupled MG/EK assay. Data from MG/EK methods are the average of duplicate measurements. Percentages in parentheses are errors of the experimental values compared to the true values (shown as “Added” in the table).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Added (nmol)</th>
<th>Pyrophosphate Reagent (nmol)</th>
<th>MG/EK (nmol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Pᵢ 35</td>
<td>33.1 (5.4%)</td>
<td>34.13 ± 0.05 (2.5%)</td>
</tr>
<tr>
<td></td>
<td>Pᵢ 20</td>
<td>20.36 ± 0.65 (1.8%)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Pᵢ 10</td>
<td>9.9 (1.0%)</td>
<td>10.54 ± 0.68 (5.3%)</td>
</tr>
<tr>
<td></td>
<td>Pᵢ 45</td>
<td>39.94 ± 0.31 (11.2%)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Pᵢ 45</td>
<td>42.1 (6.4%)</td>
<td>43.46 ± 0.37 (3.4%)</td>
</tr>
<tr>
<td></td>
<td>Pᵢ 12.5</td>
<td>12.26 ± 0.26 (1.9%)</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 4-9** Sequence alignment of two bifunctional enzymes (the first two entries) and several individual KdsCs (conservative residues showing in red background) and KdsBs (conservative residues showing in blue background) from G- bacteria strains by ClustalW.

* fully conserved residue : conservation of strong groups . conservation of weak groups
Preparation of PpBC, PpB and PpC – Based on the sequence alignment with known KdsBs and KdsCs (Figure 4-9), we should be able to divide PpBC into two individually active peptides, PpB and PpC. The division site chosen was between Pro152 and Lys153 to allow the integrity of each part and to hopefully avoid secondary structure interruption in each resultant peptide.

The pT7-LOH-PpBC plasmid was constructed and transformed into *E. coli* BL21(DE3) for expression. The resulting recombinant protein was purified using a Ni-NTA affinity column and resulted in a yield of 20 mg/L of cell culture. The pT7-LOH-PpB and pT7-7-PpC plasmids were also constructed and transformed into *E. coli* BL21(DE3) for expression. The his-tagged PpB was purified with a Ni-NTA column and a yield of 60 mg/L of cell culture was achieved. PpC was purified using a combination of a Q-sepharose anion exchange column followed by a phenyl superose column. The final yield of PpC was 30 mg/L of cell culture. The homogeneity of the proteins was verified by SDS-PAGE (Figure 4-10). The observed molecular weights are 52 kDa for PpBC, 32 kDa for PpB, and 23 kDa for PpC, which are all slightly higher than the calculated molecular weight from the amino acid sequence.

![Figure 4-10 SDS-PAGE of purified PpBC, PpB and PpC.](image)
**pH dependence** – The phosphatase activity was measured at pH 4-11 for PpBC, PpC and *E. coli* KdsC (KdsC\textsuperscript{Ec}) using KDO\textsubscript{8}P as the only substrate (Figure 4-11a). The amount of P\textsubscript{i} released for each reaction after 1-min incubation was quantified by the MG assay. The relative activity was calculated by dividing the specific activity at each pH by the activity at the optimum pH. The pH dependent curves for PpBC and PpC almost overlap. The pH dependent curve of KdsC\textsuperscript{Ec} measured in this study has an identical shape to that reported by Wu and Woodard [42]. All pH dependence curves are bell-shaped and peak around pH 6. Although all these enzymes favor an acidic pH, PpBC and PpC seem to have a broader pH adaptation, spanning pH 5-8, than does the KdsC\textsuperscript{Ec}. Especially at pH 10, PpBC and PpC retain 60% of the optimum activity, while KdsC\textsuperscript{Ec} only retains 10%.

The cytidylyltransferase activity was measured for PpB with KDO and CTP as substrates by monitoring the amount of PP\textsubscript{i} released after a 5-min incubation. PpB reaches a maximum activity at pH 10, which agrees with the previously reported pH 9.5 value for both the *E. coli* KdsB (KdsB\textsuperscript{Ec}) [43] and a maize KdsB homologue [44]. The complete pH dependence curve, reported the first time, shows that the enzyme is only active at an alkaline pH.

Finally, the combined activities were measured for PpBC (KDO\textsubscript{8}P/CTP). Coupled MG/EK was used to determine both P\textsubscript{i} and PP\textsubscript{i} for 5-min enzyme reactions. The overall specific activity is represented by the unit time production of PP\textsubscript{i}. The pH dependence curve of PpBC almost overlaps with that of PpB, and also reaches a maximum activity at pH 10. However, the coupled activity for of the phosphatase and the cytidylyltransferase retains only 10% at pH 7, which is the optimum pH for PpC. The
KDO8P was incubated at pH 10 for 5 min and measured for free P_i, and no chemical degradation of KDO8P was observed.

**Figure 4-11** The pH dependence of KdsC and KdsB activities. 

**a)** *E. coli* KdsC (labeled as EcC), PpBC and PpC were assayed with KDO8P as substrate and activities were measured as P_i release. 

**b)** PpBC was assayed with KDO8P and CTP as substrates, and PpB was assayed with KDO and CTP as substrates. Both reactions are measured for PPI release.

**Temperature profile and thermostability** – Both phosphatase and cytidylyltransferase activities were measured at various temperatures. Although the living environment of *P. putida* F1 is at ambient temperature, both activities favor high temperatures (>50°C). The phosphatase activity assayed with PpBC utilizing KDO8P as a substrate is high between 50-80°C. The cytidylyltransferase activity measured with PpB utilizing KDO and CTP as substrates reaches a maximum at 50°C. In order to maximize both activities, the temperature of the coupled enzymatic reaction was chosen to be 50°C.

To investigate the thermostability of the enzymes, PpBC was incubated at various temperatures for 5 min and then assayed at 50°C. The enzyme activity is stable below 70°C, suggesting a good thermostability.
When assaying the enzymes at high temperatures such as 50°C, CTP and KDO8P may degrade to generate P_i and/or PP_i thus interfering with the determination of the P_i and PP_i released from the reaction mixture. Control wells containing only CTP or KDO8P were therefore included in the 96-well plates. Absorbance readings generated by CTP or KDO8P were subtracted from the enzymatic reaction mixture. CTP generated small amount of P_i and PP_i during the reaction, but KDO8P was quite stable at all tested temperatures.

**Figure 4-12 a)** Temperature profiles of PpBC and PpB. PpBC was measured for P_i release with KDO8P as the only substrate, whereas PpB was measured for PP_i release with KDO and CTP as substrates. **b)** Thermostability of PpBC, measured for PP_i release with KDO8P and CTP as substrates.

**Substrate specificity** – In the coupled reaction of PpBC, KDO8P was substituted by substrate analogs, mostly phosphosugars. Each reaction mixture contained enzyme, 0.8 mM KDO8P or analog, 5 mM CTP, 5 mM magnesium ion, and 100 mM glycine-NaOH buffer at pH 10 and was incubated at 50 °C for 5 min. P_i and PP_i were measured by the MG/EK assay. Not surprisingly, all compounds except KDO8P and KDO release little P_i and negligible PP_i, which is consistent with the previously reported results for
KdsC\textsuperscript{Ec} [42], suggesting that PpBC has a high substrate specificity for the first enzymatic step, the phosphatase reaction. Besides utilizing the KDO\textsubscript{8}P/CTP pair as substrates for the first enzymatic reaction and producing KDO for the next step, PpBC can also uptake KDO directly from the buffer for the second enzymatic reaction. The specific activity for KDO/CTP is determined to be slightly higher than KDO\textsubscript{8}P/CTP (1.00 vs. 0.60 \(\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}\)), possibly due to the incompletion of the first step. Stated differently, KDO\textsubscript{8}P is not 100\% converted into KDO when processed by the phosphatase active site of PpBC. Excluding this lagging effect, the specific activities of these two reactions appear to be very close to each other.

**Table 4-6** Substrate specificity of PpBC.

<table>
<thead>
<tr>
<th>Compound</th>
<th>(P\text{_}{\text{production}}) (\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1})</th>
<th>(\text{PP}\text{_}{\text{production}}) (\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDO\textsubscript{8}P</td>
<td>1.34</td>
<td>0.60</td>
</tr>
<tr>
<td>KDO</td>
<td>&lt;0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>(p)-Nitrophenylphosphate</td>
<td>0.19</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D-Ribose 5-phosphate</td>
<td>0.17</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Phosphoenolpyruvate</td>
<td>0.13</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D-Arabinose 5-phosphate</td>
<td>0.12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D-Glucose 6-phosphate</td>
<td>0.12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D-Mannose 6-phosphate</td>
<td>0.12</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>D-Erythrose 4-phosphate</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

**Metal requirement** – All previously reported KdsC\textsuperscript{Ec} [42] and KdsB\textsuperscript{Ec} [43] require Mg\textsuperscript{2+} for activity. For the fusion protein activities, PpBC(KDO\textsubscript{8}P) representing
the KdsC activity is Mg$^{2+}$ dependent and PpB(KDO/CTP) representing the KdsB activity requires Mg$^{2+}$ to function (Figure 4-13). Reaction mixtures omitting MgCl$_2$ served as blanks. KdsC activity is very sensitive to metal ion concentration. It reaches maximum activity at 0.1 mM Mg$^{2+}$, and only drops to 60% at 10 mM. However, the KdsB activity is less sensitive to the metal ion and reaches a plateau above 5 mM Mg$^{2+}$. Based on these results, 5 mM was chosen to be the appropriate MgCl$_2$ concentration for the bifunctional enzyme activity.

![Figure 4-13 Magnesium dependence of PpB and PpBC. PpB was assayed for PP$_i$ release with KDO and CTP as substrate, whereas PpBC was assayed for P$_i$ release with KDO8P as the only substrate.](image)

Other metal ions were also tested. According to the activity assay results of PpC(KDO8P), PpB(KDO/CTP), and PpBC(KDO8P/CTP) supplemented with various metal ions (Figure 4-14), Mn$^{2+}$-assisted reactions also produce significant amounts of P$_i$ and PP$_i$, suggesting Mn$^{2+}$ as an alternate metal cofactor. The Co$^{2+}$ ion is a weaker cofactor for the KdsC activity based on the observable P$_i$ release, but not a cofactor for the KdsB activity. These results are again consistent with previous reports for *E. coli*
enzymes [42, 43], and prove Mg$^{2+}$ to be the best cofactor tested. Due to the unbeatable high activities in the presence of Mg$^{2+}$, this metal ion is possibly the physiological metal of this bifunctional enzyme.

Figure 4-14 Metal requirement of a) PpC, measured for P$_i$ release with KDO8P as substrate, b) PpB, measured for PP$_i$ release with KDO and CTP as substrates, and c) PpBC, measured for PP$_i$ release with KDO8P and CTP as substrates.

**Steady-state kinetics** – Saturation curves were obtained from each enzyme assay by varying the concentration of one substrate and fixing the concentration of the other. All the curves were fit nonlinearly to the Michaelis-Menten equation using Origin software. Parameters derived from the resultant fitting equations are listed in Table 4-7.

\[
\delta\left(\frac{k_{cat}}{K_M}\right) = \frac{k_{cat}}{K_M} \sqrt{\left(\frac{\delta k_{cat}}{k_{cat}}\right)^2 + \left(\frac{\delta K_M}{K_M}\right)^2} \quad (4-3)
\]

Errors for $k_{cat}/K_M$ were calculated according to error propagation (Equation 4-3).

The phosphatase activities represented by PpC(KDO8P) and PpBC(KDO8P) are
compared in the top panel of the table. When supplemented with KDO8P in the absence of CTP, PpBC displays only the phosphatase activity and cannot process the cytidylyl transfer reaction. PpC, as a truncated KdsC domain from PpBC, has a similar affinity for KDO8P ($K_M$) but a 2-fold higher catalytic efficiency ($k_{cat}$) compared to ppBC, suggesting that the substrate binding event of KdsC domain in the fusion protein is not affected by the covalently linked KdsB domain, but the catalytic turnover is somehow assisted by the KdsB domain. The slightly higher $K_M$ and 2-fold lower $k_{cat}$ make PpC a KdsC 3-fold less efficient than PpBC ($k_{cat}/K_M$). Overall, the KdsC activity in the fusion protein is weaker than KdsC$^{Ec}$ (Table 3-2, $K_M = 71$ µM, $k_{cat} = 456$ s$^{-1}$), with a 30-fold increase in $K_M$ and a 30-fold drop in $k_{cat}$.

**Table 4-7** Steady state kinetic parameter measurements of PpBC, PpC, PpB. Data in the top section were measured for $P_i$ release with KDO8P as the only substrate; data in the middle section were measured for $P_p$ release with KDO and CTP as substrates; data in the bottom section were also measured for $P_p$ release but with KDO8P and CTP as substrates. Apparent $K_M$ values are reported.

<table>
<thead>
<tr>
<th></th>
<th>$K_M$ KDO8P or KDO (mM)</th>
<th>$K_M$ CTP (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_M$ KDO8P or KDO (mM$^{-1}$s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PpBC KDO8P</td>
<td>2.1 ± 0.3</td>
<td>--</td>
<td>18 ± 2</td>
<td>8.3 ± 1.5</td>
</tr>
<tr>
<td>PpC KDO8P</td>
<td>2.4 ± 0.4</td>
<td>--</td>
<td>6.4 ± 0.6</td>
<td>2.7 ± 0.5</td>
</tr>
<tr>
<td>PpBC KDO/CTP</td>
<td>0.78 ± 0.06</td>
<td>0.67 ± 0.09</td>
<td>2.5 ± 0.1</td>
<td>3.3 ± 0.3</td>
</tr>
<tr>
<td>PpB KDO/CTP</td>
<td>1.7 ± 0.1</td>
<td>0.69 ± 0.01</td>
<td>13.7 ± 0.2</td>
<td>8.2 ± 0.7</td>
</tr>
<tr>
<td>PpBC KDO8P/CTP</td>
<td>0.66 ± 0.08</td>
<td>0.42 ± 0.12</td>
<td>1.07 ± 0.06</td>
<td>1.6 ± 0.2</td>
</tr>
</tbody>
</table>
When supplementing KDO and CTP in the reaction mixture, PpBC displays only the cytidylyltransferase activity. The middle panel in Table 4-7 lists the measured parameters of KdsB activities represented by PpB(KDO/CTP) and PpBC(KDO/CTP). The separate KdsB domain, PpB, behaves quite differently than the full-length enzyme by displaying a catalytic rate \(k_{\text{cat}}\) about 5-fold higher and an affinity for KDO \(K_M^{\text{KDO}}\) 2-fold lower. Comparing the CMP-KDO synthetase activity in \(P. \ putida\) F1 to \(E. \ coli\) enzymes, the \(K_M\) for KDO of PpBC and PpB is on the same order of magnitude to that reported for K-CKS (KpsU\(^{E_c}\)) [16], but 10-time higher than for L-CKS (KdsB\(^{E_c}\)) [43]. Both PpB and PpBC show a similar affinity for CTP, indicating a relative independent behavior of CTP binding.

The overall activity of PpBC (KDO8P/CTP) is quite comparable to PpBC (KDO/CTP) in both \(K_M\) and \(k_{\text{cat}}\). This information points out that the KdsB activity in the fusion protein is the rate-limiting step. This bifunctional enzyme can utilize both KDO and KDO8P to produce the activated CMP-KDO in similar rates.

**Time course** – PpBC was mixed with KDO8P and CTP, and measured for \(P_i\) or \(PP_i\) release after 1, 2, 5, 10, 20, 30, 40, 60 min incubation. \(P_i\) builds up within one minute in the coupled reaction catalyzed by PpBC, while \(PP_i\) accumulates slower and reaches the maximum in 30 min (Figure 4-15). These results agree with the steady-state kinetic measurements which reveal that the CMP-KDO synthetase activity is the rate-limiting step in the fusion enzyme (Table 4-7 PpBC\(_{\text{KDO8P}}\) and PpBC\(_{\text{KDO8P/CTP}}\)). The mixture of PpC/PpB was also measured for time course and compared to the covalently linked bifunctional enzyme. The indistinguishable rates of \(P_i\) and \(PP_i\) formation in the PpB/PpC catalyzed reactions suggest that the KDO produced from PpC is further consumed by
PpB immediately. This fact is coincident with the prediction based on the steady-state kinetic parameters (Table 4-7 PpC\textsubscript{KDO8P} and PpB\textsubscript{KDO/CTP}), which indicates that PpC is the rate-limiting activity in this enzyme pair. Overall, PpBC is more efficient than the PpC/PpB pair in converting KDO8P into CMP-KDO.

Figure 4-15 Time course of coupled reactions catalyzed by PpBC or a mixture of PpB and PpC.

4.5 Discussion

Assay of P\textsubscript{i} and PP\textsubscript{i} – A combined colorimetric assay was developed capable of measuring both P\textsubscript{i} and PP\textsubscript{i} simultaneously and independently. Decent sensitivities for P\textsubscript{i} and PP\textsubscript{i} detection are achieved, each in the range of 0-0.4 mM, as suggested by the linear calibration curves. In a mixture of P\textsubscript{i} and PP\textsubscript{i}, the amount of P\textsubscript{i} is directly derived from the malachite green signal. In contrast, absorbance of the eikonogen assay comprises signals of two chromophores which are formed independently by P\textsubscript{i} and PP\textsubscript{i}. Since the P\textsubscript{i}...
amount is solved from the MG signal, PP\textsubscript{i} can be calculated by subtracting the P\textsubscript{i} contribution from the total EK signal. Both assay reagents contain strong acid that results in a low pH in the detection mixture, and have the colors develop in the same time scale (minute range), therefore allowing the sample solutions for different assays being maintained in the same chemical environment. Only when the samples subjected to MG and EK assays are identical can the P\textsubscript{i} amount determined from MG be used in the deconvolution of EK signals. The accuracy of PP\textsubscript{i} measurement was confirmed by the pyrophosphate reagent assay from Sigma. Unlike other reported pyrophosphate determination assays, the combined MG/EK assay does not introduce any P\textsubscript{i} source into the reaction mixture, and therefore can monitor the P\textsubscript{i} purely produced by the target reaction. Although PP\textsubscript{i} is the final product, it is interesting to follow the product of the first enzymatic reaction, P\textsubscript{i}, at the same time. The combination of two colorimetric assays reduces complexity by avoiding using coupled enzyme, and has a relatively low cost by utilizing only commercially available chemicals.

The simplicity of this method enables the application of a 96-well plate format. After quenching with one volume of 100% ethanol, the reaction mixture on a reaction plate was immediately divided into two portions and subjected to MG and EK respectively, providing exactly the same sample volumes for both detection methods. The uneven division of a reaction mixture, 20 µL for MG and 50 µL for EK, is to achieve relatively the same level of absorbance in both assays, since the extinction coefficients differs about 2-fold. Using the 96-well plate format, systematic blanks can easily be set up for each test. For example, CTP self-degradation in acid can be subtracted from the
reaction by setting up various amounts of CTP in the reference wells. This method can be extended to any coupled reactions generating P; and PP; as products,

**PpBC is a bifunctional enzyme** – The putative bifunctional enzyme PpBC was first discovered at the primary structural level from a conserved domain BLAST search of the NCBI database utilizing KdsC Ec as the probe. The sequence alignment (Figure 4-9) reveals that the N-terminal amino acids 1-252 share many conserved residues with KdsBs, especially those residues proposed to be involved in Mg 2+ coordination (D106, D246), CTP binding (R15, S18, T19, R20, K24, R84) and KDO binding (R164, H192, Y196, E221, Q222)[45, 46]. Meanwhile, residues 253-428 in the C-terminal part show KdsC characteristics including not only the conservation of all the haloacid dehalogenase superfamily (HADSF) motifs (discussed in Chapter 1), but also the presence of the KdsC-specific signature sequence GGxGAXRE[42]. The sequence conservation explicitly divides PpBC into two parts. Each part should hopefully fold as an individual domain, and function as a KdsB (i.e. CMP-KDO synthetase or CKS) for the N-terminal part and a KdsC (i.e. KDO8P phosphatase) for the C-terminal part. Based on this prediction, the fusion enzyme was divided to two putative monofunctional peptide domains, which are individually cloned and expressed as PpB and PpC. The resultant soluble PpB and PpC display the expected activities, and are similar to their previously reported counterparts from other microorganisms in terms of metal requirement (Figure 4-13, 4-14) and substrate specificity (Table 4-5). All these facts prove that the division of the PpBC does result in two separate monofunctional enzymes. The full-length fusion protein was also prepared as a soluble protein, and tested for both separate activities as well as combined
activity. When incubated with KDO8P and CTP, PpBC released both \( P_i \) and \( PP_i \), indicating that it is indeed a bifunctional enzyme.

**The CKS activity** – When provided with KDO and CTP, PpBC solely displays CKS activity. The \( K_M \) values for KDO and CTP measured for PpBC are in the mM range, which is on the same order as the reported \( K_M \) for K-CKS (KpsU\textsuperscript{Ec}) [16], but 10-fold higher than that reported for L-CKS (KdsB\textsuperscript{Ec}) [43], further confirming that the activating activity is most like the K-CKS. Compared to the individual CKS domain, PpB (1.67 mM), ppBC has a 2-fold lower \( K_M \) for KDO (0.78 mM). On the other hand, PpBC (KDO/CTP) shows a \( k_{cat} \) on the same order of magnitude as that of K-CKS [16], 10-fold lower than PpB, and 100-fold lower than the reported L-CKS [43]. This difference between PpBC (KDO/CTP) and PpB is possibly related to the ability of dimer formation. A dimeric structure for both K-CKS [45] and L-CKS [47, 48] has been previously reported, however no mechanistic significance of dimerization was reported. PpBC may not be able to form a dimer in the presence of the covalently linked KDO8P phosphatase domains.

The 10-fold lower catalytic rate of the CKS activity, compared to the phosphatase activity (\( k_{cat} = 1.65 \) vs. \( 17.7 \) s\(^{-1}\)), indicates that CKS is the rate limiting step in this fusion enzyme, therefore the CKS activity should dominate the overall coupled activity of PpBC in kinetic properties. Based on this conclusion, it is not surprising that PpBC (KDO/CTP) and PpBC (KDO8P/CTP) have comparable catalytic rates (\( k_{cat} = 1.65 \) vs. \( 0.9 \) s\(^{-1}\)). In fact, the CKS activity is limiting the overall coupled activity at all tested pH values, since the pH dependence curves of PpB and PpBC(KDO8P/CTP) are almost overlapping (Figure
Consistent with these result is the fact that KdsB has been previously reported to be the rate limiting step in the overall LPS biosynthesis [43].

**The phosphatase activity** – The $K_M$ values for KDO8P of the phosphatase activities appears to be not affected by the covalently linked CKS domain, since the values are quite close for both PpBC and PpC (Table 4-6). In addition, activities of the fused phosphatase (PpBC) and the individual phosphatase domain (PpC) have almost overlapping pH dependence curves (Figure 4-11a). These results suggest that PpBC is likely to have an independent phosphatase active site that is not sterically hindered by the CKS domain, neither statically nor dynamically.

The turnover rate of the phosphatase activity in the fusion protein is about 30-fold lower than that of KdsC$^{Ec}$. A group 2 tail (represented by the $\alpha$-helical tail in *Haemophilus influenzae*) is present at the C-terminal of this fused KdsC, which locates in the C-terminal part of the fusion protein. Therefore, it is possible for the C-terminal tail to act as a movable region, as observed in KdsC$^{Ec}$ (Chapter 2). The movable tail may insert into the active site of another monomer, for example an adjacent monomer in the tetrameric KdsC$^{Ec}$, and assist in product release (Chapter 3). Tail-less KdsCs, either truncated KdsC$^{Ec}$ or natural tail-less KdsC$^{Ad}$, generally have a $k_{cat}$ $10^2$-$10^3$ fold lower than the full-length KdsC$^{Ec}$. As a tailed KdsC, PpBC (KDO8P) has a catalytic rate between other characterized tailed and tail-less KdsCs. On the other hand, the $K_M$ value for KDO8P of the fusion protein is 30-fold higher than that of KdsC$^{Ec}$. Therefore the catalytic efficiency ($k_{cat}/K_M$) of the phosphatase activity in the fusion protein is 900-fold lower than that of KdsC$^{Ec}$. A structural parallel between the KdsC domain of PpBC and KdsC$^{Ec}$ needs to be drawn to reveal molecular basis for the extraordinary low catalytic
efficiency of the former. Whether a tetrameric structure can be formed and whether the C-terminal tail can insert into the active site of another monomer may help understanding the mechanism of PpBC.

Though the turnover rate of the phosphatase in the fusion protein is low, the pH profile (Table 4-11a) may suggest an advantage of this phosphatase in the fusion protein. When comparing the pH dependence of the phosphatase activity of PpBC and PpC to that of KdsC{Ec}, an obvious relative activity enhancement in the alkaline pH range can be observed for PpC and PpBC. The activity of KdsC{Ec} drops immediately above pH 8, and retains only 10% activity at pH 10, which is the optimum pH of KdsBs. However for the fusion protein, the phosphatase activity retains 50% activity at pH 10. The optimum pH for the overall coupled activity is at pH 10 (Figure 4-11b), limited by the CKS activity. Therefore, the expansion of pH adaption to the alkaline range in the PpC domain may have enhanced the efficiency of coupling with the PpB domain.

**Enzyme coupling in the fusion** – To enable efficient enzyme coupling and avoid intermediate accumulation, comparable activities for both enzymes are required. In the case of this bifunctional enzyme, the $k_{cat}$ of PpBC (KDO8P) is about 10-fold higher than that of PpBC (KDO/CTP), which inevitably leads to production of more KDO than can be consumed. However, the advantage is that the lagging phase is shortened by a fast build-up of the KDO intermediate. An even larger difference in catalytic rate is observed for the KdsC{Ec} ($k_{cat} = 456 \text{ s}^{-1}$) and KpsU{Ec} ($k_{cat} = 1.35 \text{ s}^{-1}$ [16]) pair, making these two enzymes extremely inappropriate to couple at a 1:1 stoichiometric ratio, as occurs in PpBC where one molecule of PpB is covalently linked to one molecule of Ppc. The *E. coli* enzyme pair, unless having $10^3$-fold more KdsC{Ec} than KpsU{Ec}, will produce excess
KDO. In this sense, PpBC is much more efficient in unitizing KDO. The activity coupling is also facilitated by other similar properties shared between the two activities, including a close temperature optimum (Figure 4-12a), the overlapping Mg$^{2+}$ utilizing range (Figure 4-13), and the coincident pH adaption in the alkaline range (Figure 4-11a).

In PpBC, the phosphatase activity is 2.5 fold higher than the activating activity in catalytic efficiency ($k_{cat}/K_M$). In contrast, the individual domains have an inverse trend with PpB being 3-fold more active than PpC (Table 4-6). The time course of the PpBC reaction and the PpB/PpC mixture reaction (Figure 4-15) shows that the fusion enzyme achieves the exhaustion of the reactant faster than the mixture of two individual domains when utilizing KDO8P and CTP as substrates, suggesting that PpBC is indeed a better enzyme. The spatial proximity of two domains must be a critical factor contributing to the efficiency of the fusion protein.

Overall, the coupling of two activities in this bifunctional enzyme seems not to improve the catalytic rate for the formation of final product (monitored by PP$_i$), but the apparent catalytic rate of PpBC is higher than that of PpB/PpC mixture. However, other advantages of enzyme fusion, for example the sharing of a single regulatory site [49], substrate channeling [50], and synergic effects [51, 52], will require additional experiments to prove.

From the standpoint of evolution, possession of two complete set of KDO biosynthetic genes in P. putida F1 is possibly caused by gene duplication under environmental stress. In particular, the incorporation of a full set of KDO biosynthetic genes into the kps gene cluster may enhance the capsular expression in P. putida F1, which can live on otherwise toxic aromatic compounds. The occurrence of KDO
biosynthetic gene cluster (\textit{kpsF}, \textit{kdsA}, \textit{kdsC}, and \textit{kpsU}) may be a positive evolutionary product that possibly improves the efficiency of the entire pathway by shortening spatial distances between enzymes. The first two discovered fusion gene examples that encode the bifunctional KdsBC may be an evolutionary error-try intermediate, which is not greatly improved in enzyme activities but still needs perfection by nature. As more microbial genomes are sequenced, these speculations can be further tested.

4.6 Future Directions

A crystal structure of PpBC in the presence and/or absence of KDO8P and CTP should be obtained. A structural comparison between the bifunctional enzyme and its individual parts, PpB and PpC, could possibly reveal details of how KDO moves between domains.

The assay method developed for the bifunctional enzyme can be further applied to high throughput screening of the \textit{E. coli} KdsD, KdsA, KdsC, and KdsB biosynthetic pathway. If all these enzymes can be efficiently coupled, a high throughput assay can be performed on small molecular libraries to simultaneously screen for inhibitors of the entire pathway.

4.7 Acknowledgements

I thank Dr. Ted Huston in the University of Michigan, Department of Geology for helping to determine the metal constitution of PpBC, PpB and PpC using high-resolution inductively coupled plasma-mass spectrometry (ICP-MS). I would also like to thank members of the Woodard group for their helpful discussions.
4.8 References


47. Mamat U, e.a., Crystal structure of KdsB from Aquifex aeolicus in complex with CTP. unpublished data, 2008.


Chapter 5

Conclusions

Mechanistic insights into 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KdsC) were studied and presented in this dissertation. The combination of X-ray crystallography structural analysis and enzymatic characterization revealed the molecular basis of the regulatory mechanism and of substrate specificity of KdsC. A gene fusion of \textit{kdsC} and its consecutive gene in the pathway, \textit{kdsB}, was identified and expressed as a bifunctional enzyme. The characterization of this enzyme \textit{kdsBC} contributed to understanding the activity coupling in the pathway.

Seven sets of crystal structures for \textit{E. coli} KdsC and its variants were solved. The deletion of the last eight amino acids at the C-terminal end enabled products to be captured in the active site. Residues Lys102, Gly77, Thr76, etc. interact with KDO or P\textsubscript{i} portion in the active site. Tetramerization also plays a critical role in achieving high substrate specificity since each active site is partially covered by a cap-like adjacent monomer, with the cap monomer residues Arg86, Val56 and Leu90 forming salt bridges or hydrophobic interactions with the KDO portion. Mutagenesis studies may lead to a further clarification of the significance of these residues.

The crystal structures illustrated two major conformational changes in KdsC. One conformational change occurs in the C-terminal tail, which undergoes transition between
solvent exposed state and active site insertion, and the other occurs in the two loops at the active site, which move to form a bulge coincident with the tail insertion or substrate binding. A tetrameric structure must be formed to enable the tail insertion into the active site of the adjacent monomer. The C-terminal end of the tail occupies the same space as the KDO portion, and forms similar specific contacts with the active site residues. The presence of the tail is essential for achieving a high catalytic efficiency for KdsC from *E. coli*, because slower turnover rates are observed for both the native tail-less KdsCs and the tail deletion mutant of *E. coli* KdsC. The tail possibly functions as an active facilitator of the releasing products by competing with KDO for active site occupancy and destabilizing P$_i$ binding in the active site. A catalytic cycle model for the tail was thereby proposed. This is a unique mechanism discovered in the Haloacid Dehalogenase Superfamily to date. The differences in the tail region lead to the classification of KdsC into three groups which are likely related evolutionarily. Characterization of KdsC from the different groups will be helpful to understand how the tail evolves to improve the function of the enzyme.

During the database search of KdsC, a fused gene was identified with *kdsB* gene on the N-terminal end and *kdsC* gene on the C-terminal end. This fusion gene is likely involved in the capsular polysaccharide biosynthesis because it resides within the *kps* cluster which encodes capsular synthesis and assembly enzymes. The recombinant protein expressed from the fused gene is proven to be a bifunctional enzyme with both KDO 8-P phosphatase activity and CMP-KDO synthetase activity. Characterization of the bifunctional enzyme reveals several special properties enabling the coupling of the two enzymes whose *E. coli* counterparts are difficult to couple in 1:1 ratio. The KdsB
activity is the rate-limiting step in this coupled reaction, and shows similar kinetic parameters with other capsular CMP-KDO synthetases. The KdsC activity is much slower than the similarly tailed KdsCs from *E. coli* and *H. influenzae*, making the fused gene KdsC activity only slightly faster than the KdsB activity to avoid over accumulation of the intermediate. The pH adaption of the KdsC activity expands to the alkaline range to overlap with the optimum pH of KdsB. A combination of Malachite green assay and Eikonogen assay allows measurement of P$_i$ and PP$_i$ releases in the same reaction mixture, therefore opens the possibility of obtaining kinetic information for both steps in the coupled reaction, and paves the way for assaying the whole KDO biosynthetic pathway for inhibitor screening.
APPENDICES
APPENDIX A

ATTEMPTS TO EXPRESS S-ADENOSYL-L-METHIONINE HYDROLASE

SAMase Catalyzes SAM Degradation

S-adenosyl-L-methionine (SAM) is one of the most ubiquitous and important naturally-occurring molecules [1]. The ability to donate various moieties to selected nucleophiles makes SAM an important substrate in the pharmacological activation and deactivation of various biological nucleophiles [2, 3]. Research tools are needed to regulate the pool size of SAM in vivo in order to elucidate the mechanism of important biological processes such as E. coli’s defense [4, 5]. Though chemical degradation of SAM has been studied [6], a biocompatible enzymatic reaction is required for down-regulating SAM in biological systems. The only known enzymatic degradation reaction of SAM is the hydrolysis catalyzed by SAM hydrolase (SAMase, EC 3.3.1.2), a 17 kDa protein. SAM is degraded to homoserine (HS) and methylthioladenosine (MTA) in this reaction (Figure A-1) [7]. SAMase was used to down-regulate the level of SAM by either adding a T3 phage infected E. coli extract[8] or introducing a pBAD plasmid containing the SAMase gene [5]. However, both methods cannot quantitatively control the pool size of SAM due to the lack of characterization of SAMase. Purified SAMase should be obtained and enzymatic characterization of this enzyme should be carried out.
Attempts to Clone Recombinant SAMase

SAMase is cytotoxic to *E. coli* host cells due to its unregulated hydrolytic activity to degrade SAM, a molecule required for cell survival. Therefore, cloning of SAMase gene into unrestricted vectors such as a pT7-7 vector failed. In the pT7-7 vector, the T7 promotor is driven by T7 RNA polymerase, which is expressed from an IPTG inducible gene on the host strain [9]. In the absence of IPTG, the target gene cloned into the pT7-7 still expresses small amount of protein. The “leak” of SAMase protein from the *SAMase*/pT7-7 plasmid expression resulted in a halt of *E. coli* host cell growth, therefore no protein could be obtained from fermentation. To avoid the leak of SAMase expression before IPTG induction, tightly regulated vectors were used. Posnick reported the recombination of the SAMase gene into a L-arabinose controlled pBAD plasmid (Invitrogen) and observed enzyme activity [5]. However, our group discovered that Posnick’s SAMase gene was actually a C75Y mutant, and the mutant enzyme has very weak enzymatic activity.
More attempts were made to express the full-length SAMase in a tightly regulated plasmid-based system in my study. First I utilized the pLEX vector (Invitrogen) which has a P_L promoter. A λ repressor upstream to the P_L promoter is subject to restriction by tryptophan. Therefore, only in the presence of tryptophan can this promoter be turned on. A tryptophan-absent media was used in the cloning experiments, however, only a couple of single colonies were obtained after ligation, and all clones sequenced were either frame-shifted products or multiple-site mutants. Another tightly regulated vector utilized in my study is a double-operator system, pAu21a (Expression Technologies Inc.). It has two lacI repressors upstream and another lacI repressor downstream to the multiple-cloning site. This design can increase the repression of background leakage 20 to 70 fold. Yet the cloning efforts resulted in no colonies.

Finally, an optimized SAMase gene using all E. coli-friendly codons were constructed by DNA 2.0 Inc. Still no positive clone was obtained after trials of cloning the optimized SAMase gene into any available circular plasmid. All these attempts strongly suggest that trace amount of SAMase is fatal to E. coli cells.

**Mutant C75Y of SAMase**

The plasmid encoding SAMase C75Y/pBAD was obtained from Posnick [5]. This mutant is a potential close representative of SAMase, therefore worth studying when the wild-type protein could not be expressed. Attempts were made to transfer the mutant gene from pBAD cloning vector into expression vectors such as pT7-7, pT7-LOH (an N-terminal his-tagged vector based on pT7-7), and pLEX. Only C75Y/pT7-LOH was successfully constructed. However, although colonies of XL1-Blue E. coli containing
C75Y/pT7-LOH could be amplified in liquid media, the expression cell lines such as BL21(DE3) and ER2566 containing the plasmid failed to grow in liquid media. A pLEX system compatible E. coli strain, GI724, was able to grow in liquid media after transformed with the C75Y/pT7-LOH plasmid. However, no overexpression of the C75Y protein was obtained. These results suggest that the C75Y mutant is less cytotoxic than the wild-type SAMase, but can only exist in small concentration before destroying the host cells.

**Intein-mediated Protein Ligation**

Intein-mediated protein ligation (IPL) allows separate expression of two sequential SAMase fragments with self-cleavable intein tags. Subsequent ligation of the two parts was used to form a full-length protein through a thiol substitution reaction [10, 11]. This strategy was adopted to avoid expression of the full-length active SAMase in E. coli host cells, because of the aforementioned cytotoxicity. The IMPACT™-TWIN system (New England Biolabs) was used to clone, express and purify the two fragments of SAMase [12]. SAMase I, the N-terminal fragment of SAMase including residue 1-74, was fused with an Mxe GyrA intein and a chitin binding domain (CBD) at its C-terminus. After immobilizing the fusion protein onto chitin columns through the interaction between its CBD and chitin resins, the Mxe GyrA intein tag can be cleaved under treatment of small thiol molecules such as dithiothreitol (DTT) or 2-mercaptopoethane sulfonic acid (MESNA). The cleavage of intein results in a thioester, a functional group required for the subsequential ligation reaction. The second peptide termed SAMase II consists of residue 75-152, starting with Cys75. The N-terminal residue of SAMaseII is
required to be a Cys, whose thiol side chain can serve as a nucleophile in the subsequent thiol substitution ligation reaction. SAMaseII was fused with a CBD and an Ssp DnaB intein at its N-terminus. The self-cleavage of the intein tag upon swift pH change from 8.5 to 7 allows regeneration of the N-terminal Cys. With both SAMase I-thioester and SAMase II available, a ligation reaction can be performed by incubating both pieces at 37°C overnight, and finally results in a native amide bond between the two halves (Figure A-2) [13].

**Figure A-2** Scheme of intein-mediated ligation reaction. SAMase I has a thiolester adduct generated by MESNA assisted cleavage of the *Mxe GyrA* intein. The thiol group on the N-terminal cysteine of SAMase II acts as the nucleophile, and finally results in a native peptide bond in between.

The clones of both halves fused with intein were obtained in pTWIN1 vector. The SAMase I/Mxe GyrA fusion protein was overexpressed in soluble fractions. Significant cleavage of the intein was observed when the cell extract was treated with DTT or MESNA (Figure A-3). The detection of small molecule weight peptides was enabled using tricine SDS-PAGE [14]. Purification of SAMase I was then carried out on a chitin column in a batch manner. The column was quickly flushed with buffer A (20 mM Tris
pH 8.5, 500 mM NaCl, 1 mM EDTA) containing 60mM MESNA, and incubated at room temperature for 24 h. The 8 kDa SAMase I peptide was eluted using buffer A, and the column was finally stripped to detect the fusion precursor consisting of the *Mxe GyrA* intein and CBD. The on-column cleavage efficiency was high, since large amount of fusion precursor was obtained. But the yield of SAMase I was low, possibly due to the low solubility of the peptide. N-terminal sequencing was performed on the resultant SAMase I from chitin column, and the residues identified matched the first amino acids in the sequence of SAMase I.

![Figure A-3 Cleavage tests of SAMase I/Mxe GyrA/chitin binding domain (CBD). Lane 1: molecular weight marker; 2: crude extract incubated with 40 mM DTT for 24 h; 3: crude extract; 4-7: crude extract incubated with 40, 60, 80, and 100 mM MESNA for 24 h.](image)

SAMase II was cloned into pTWIN1 vector downstream to a CBD and an *Ssp DnaB* intein. The fusion protein was overexpressed as inclusion bodies. Directly treating the inclusion body with a pH shift from 8.5 to 7 along with a temperature jump from 4 °C to 25 °C did not release any SAMase II, likely due to incorrect folding of the fusion protein. In order to achieve correct folding, the insoluble fusion protein was subjected to a denature-renature procedure starting from denaturing in 7 M guanidine chloride, and followed by stepwise dialysis against 8 M, 6 M, 4 M, and 2 M urea in the presence of the
glutathione redox pair at 4°C. The renatured fraction was finally dialyzed into a pH 7 buffer, but no cleavage of intein was observed. Due to the failure of generating a cys at the N-terminus of SAMase II through intein cleavage, the subsequential ligation reaction was not performed.

**Enzyme Assay**

SAMase was coupled with adenosine deaminase (ADA, EC 3.5.4.4), which turns MTA into methylthioinosine. Methylthioinosine has a significant lower extinction coefficient than that of MTA at 259 nm. Therefore, a decrease of absorbance at 259 nm can be monitored for reaction progression [15]. Before mixed in a reaction, SAMase solution as will be specified was treated with DNase I and RNase A, incubated at 37°C for 30 min, and ultrafiltered to get rid of any NTPs or dNTPs. The 70 µL reaction mixture contained 50 mM Tris pH7, 3.5-7 µM SAM chloride, 225 ng ADA, and 3 µL SAMase solution as will be specified. The reaction was incubated at 37 °C for 20 min and continuously monitored at 259 nm.

**Expression of SAMase through T3 Phage infection**

The only organism carrying a SAMase gene is bacteriophage T3, according to the genomic data in the NCBI database to date. The T3 phage expresses SAMase within minutes after infecting the *E. coli* host cells, and digests SAM which provides methyl groups for *E. coli* defensive modifications [16]. Besides the full length SAMase, a possible “truncated” SAMase containing 125 amino acids with molecular weight of 14 kDa could also be expressed from an internal start codon coding for Met28.
bacteriophage T3 is inactivated by ultraviolet radiation, it can no longer destroy the host cell, and therefore the recipient cells are able to replicate themselves.[17] The infection of *E. coli* with heavily UV-irradiated phage T3 leads to a 3 to 6 fold accumulation of SAMase activity over the non-irradiated phage. However, it was still difficult to obtain significant amounts of purified SAMase with this method.

Small scale phage infection experiments were set up in order to obtain small amounts of SAMase to validate the assay. Experiments were performed as described [18]. Briefly, phage titer was first calculated by plating phage dilutions and the host cell (*E. coli* ATCC 11303) together on soft LB agar, and counting the number of plaques. The mid-log phase host strain was infected with T3 phage at a multiplicity of infection (MOI) of 3-5, and harvested after 6-7 min of growth at 37°C. The cells were lysed, and the crude extract was precipitated with 30-60% ammonium sulfate. The target protein was found in the reconstituted precipitate by SDS-PAGE and the concentration of SAMase was estimated to be 0.05 mg/mL. This enzyme solution (3 µL) was subjected to enzyme assay and proved to be active.

**In vitro synthesis can produce trace amounts of active SAMase**

Another strategy to circumvent the cytotoxic full-length protein expression in cells was to synthesize the protein by using cell-free *in vitro* coupled transcription/translation. A linear DNA template was first constructed according to requirements [19] through three-steps of PCR. The finally resulted 701 bp DNA consists of components as following (from 5’ to 3’): a bacteriophage T7 promoter, a prokaryotic Shine-Dalgarno ribosomal binding site (RBS), an ATG initiation codon, the SAMase
gene, a stop codon, and a T7 termination region. This DNA template was verified by sequencing.

An in vitro transcription reaction (MAXIscript®, Ambion) was first performed to validate the production of mRNA at correct size. The reaction mixture contained 2 µL buffer as supplied, 1 µg linear DNA template, 1 mM each NTP, 2 µL enzyme mix as supplied, and nuclease-free water as supplied up to 10 µL volume. The reaction was incubated at 37 °C for 2 h and loaded on an agarose gel for detection. The agarose gel showed a 400 bp band when DNA ladder was used, indicating a ~800 base single-stranded mRNA production.

The coupled transcription/translation reaction was then carried out using an Expressway™ kit from Invitrogen. The reaction mixture contained 20 µL buffer as supplied, 20 µL E. coli cell extract as supplied, 1.25 µL 50 mM amino acids mixture, 1 µL 75 mM Met, 1 µL enzyme mix as supplied, 4 µg linear DNA template, and nuclease-free water up to 100 µL. The reaction was first incubated at 37°C for 30 min, and added with 25 µL feeding solution containing 12.5 µL feeding buffer as supplied, 0.6 µL 50 mM amino acids mixture, 0.5 µL 75 mM Met, and nuclease-free water. After 2 h incubation, counting from the beginning, another 25 µL feeding solution was added. The reaction was quenched by frozen at -20 °C after 6 h of total incubation. Microgram quantities of SAMase were produced in the form of insoluble protein whereas only small fraction of soluble SAMase was obtained (Figure A-4). The putative product of insoluble SAMase was cut from the SDS-PAGE gel, and subject to an in-gel protease digestion and MS/MS analysis. The resultant mass of peptide fragments was searched against the NCBI database, and the sequence hit agreed well with the amino acids sequence of SAMase.
Reaction temperatures and DNA template concentrations were altered in the hope to achieve better solubility, but none of the conditions resulted in significantly higher yield of soluble SAMase.

**Figure A-4** The *in vitro* transcription/translation synthesis of SAMase. Various concentrations of DNA template were used in the reaction mixture, and both soluble fractions and insoluble fractions were tested. The band showing in bold red box was subject to in-gel digestion and tandem MS analysis.

To refold the insoluble SAMase product of the *in vitro* synthesis, products from three reactions (450 µL total) were combined and treated with 7 M guanidine hydrochloride, followed by stepwise dialysis against 8 M, 6 M, 4 M urea with DTT, and 2 M, 0 M urea with glutathione redox pair. The resultant soluble fraction was 24.7 µg/mL, containing SAMase and impurities. This fraction (3 µL) was proved to be positive for SAMase activity (**Figure A-5**). However, when scaling up the refolding procedure with 1700 µL *in vitro* reaction mixture, all proteins were precipitated at 2 M urea.
Figure A-5 Time course of the ADA coupled SAMase reaction. Included in the legend, from up to down, are 1) self-degradation of SAM in the absence of any enzyme source; 2) the in vitro transcription/translation reaction mixture without a SAMase template; 3) the in vitro reaction mixture catalyzed SAM degradation; 4) the refolded SAMase catalyzed SAM degradation.

The production of insoluble SAMase from in vitro transcription/translation could possibly have resulted from the quick production of proteins, which causes the protein products to aggregate before finishing folding. Based on this hypothesis, genes of Phe3 and Lys5 in SAMase were modified to be less favored codon of E. coli, in order to reduce translation rate. However, the change of codon did not result in more soluble SAMase production from the in vitro synthesis.

Conclusion and Future Directions

The cytotoxicity of SAMase makes it difficult to clone SAMase gene into any circular plasmid, since trace background expression of the enzyme results in E. coli cell death. Other cell strains such as yeast, insect or mammalian cells could be used instead of
*E. coli*, but since SAM is critical to all these organisms, the problem of cytotoxicity still exists. In future studies, extremely tight-controlled vectors other than those utilized in this study should be used to achieve minimal leak of SAMase expression before induction.

The cell-free *in vitro* transcription/translation method produced μg levels of soluble SAMase at a high cost of purchasing the kit. The DNA template should be further optimized to synthesize soluble SAMase, with a goal of obtaining milligram levels of SAMase. The optimized SAMase gene sequence made by DNA 2.0 Inc. can be used as the linear DNA template in *in vitro* synthesis in order to enhance both solubility and yield.

The most promising method to produce large quantity of soluble SAMase is the intein-mediated protein ligation, since each half of SAMase can be overexpressed. The soluble and thioester-activated SAMase I was obtained in my study, while SAMase II was obtained in an insoluble and uncleavable fusion protein. Other cleavable affinity tags that can generate an N-terminal cysteine can be used to express and purify SAMase II. Once both halves are prepared in high concentration, a ligation reaction can be performed to generate native SAMase.
Reference


APPENDIX B

STUDIES ON QUORUM SENSING AUTOINDUCER SYNTHASE: A S-ADENOSYL-L-METHIONINE LACTONIZING ENZYME

Autoinducer biosynthesis in Quorum Sensing

A group of enzymes in bacteria catalyzes biosynthesis of acylated homoserine lactone (AHL) from acyl-acyl carrier protein (ACP) and S-adenosyl-L-methionine (SAM) (Figure B-1) [1]. The cleavage of SAM between C4 and the positively charged sulfur ion is also observed in S-adenosyl-L-methionine hydrolase, therefore mechanistic analogy can be drawn between these two types of enzyme.

\[ \text{3-oxo-octanoyl-acyl carrier protein} + \text{S-adenosyl-L-methionine} \rightarrow \text{Tral} \rightarrow \text{3-oxo-octanoyl-homoserinelactone} \]

Figure B-1 Scheme of Tral catalyzed reaction.
Bacteria release AHLs as messenger molecules in the environment to communicate with other bacteria of the same kind, and thereby regulating cell-population density. This process is termed quorum sensing [2]. The AHL molecules are termed autoinducer (AI), and enzymes catalyzing the synthesis of AHL are termed autoinducer synthetase [2]. Orthologs of AI synthase from various microorganisms have been studied and reported to transfer different acyl chain onto the cleaved homoserine from SAM [3-6]. Two crystal structures of AI synthetase have been solved revealing the substrate binding pocket and a hypothetic channel that could possibly fit the length of the acyl chains and therefore determining the specificity for the acyl chains [7, 8]. An AI synthases from *Agrobacterium tumefaciens* (TraI) was cloned and expressed for mechanistic studies.

**Expression and Purification of Soluble TraI**

TraI was previously cloned into pT7-7 vector in the lab and was moved into a his-tagged pT7-LOH vector for convenience of purification. The his-tagged TraI expressed as inclusion body after fermentation, which is consistent with a previous report [9]. The insoluble TraI failed to refold through a denaturation/renaturation procedure. Permutations of IPTG concentration, temperature and duration of growth were used in the fermentation to optimize the expressing condition of soluble proteins. Finally, when using a rare codon repairing cell line, BL21-CodonPlus® -RIL (Stratagene), and growing the cell under 37°C overnight after IPTG induction, soluble TraI was obtained (Figure B-2).
Figure B-2 Expression of TraI under different cell-growth conditions. 1. Cells grew at 17°C overnight after induction; 2. Cells grew at 37°C overnight after induction. CE = crude extract. P = cell pellet.

TraI was then purified from the crude extract through two chromatographic steps, a HisBind® (Novagen) Ni-NTA column followed by a phenyl superose column, and achieved above 90% homogeneity. Pure protein was subject to crystallization condition screening but no positive hit was obtained.

Purification of holo-ACP

One of the substrates of TraI is 3-oxo-octanoyl-ACP. ACPs are small proteins (~10 kDa) that serve as acyl donor in many biosynthetic processes. The functional form of ACP includes a 4-phosphopantetheine cofactor attached to a conserved serine residue, termed holo-ACP. If an acyl chain is attached to the thiol group on the 4-phosphopantetheine portion, the protein is termed loaded ACP. It is difficult to express recombinant E. coli ACP, because without the 4-phosphopantetheine cofactor, the apo-ACP is cytotoxic [10]. In this study, a spinach holo-ACP was expressed with modifications to the protocol reported by Broadwater and Fox [11]. A co-expression vector containing tandem ACP and E. coli holo-ACP synthase (ACPS) genes, pBHF-7A,
was generously donated by Fox. Instead of using the special fermenter mentioned in Fox’s report that can control pH and oxygen supply of the cell culture and grow cells at 17°C overnight, the cells were grown in a normal fermenter at 37°C for 4 h after induction. Holo-ACP expressed accounted for 90% of the ACP pool, while apo-ACP accounted for the other 10%. If the cells were grown in a normal fermenter at 17°C overnight, only 50% of the ACP pool was expressed as holo-ACP. The co-expressed holo-ACP and ACPS were dissociated by purifying the crude extract on a Mono-Q column in the presence of 20% isopropanol. The fractions containing holo-ACP were further purified using a HisBind® (Novagen) Ni-NTA column followed by a phenyl superose to achieve high homogeneity. The final homogenous fraction was subjected to intact mass measurement using MALDI-MS, and two species were detected: 9685 Da (10%) coincident with molecular weight of apo-ACP and 10022 Da (90%) coincident with molecular weight of holo-ACP. The difference between the two species (337 Da) is close to the molecular weight of the 4’-phosphopantetheine group (357.349).

**Loading of the Acyl Chain**

ACP can be loaded with the desired acyl chain in several ways: 1) (the *in vivo* way) forming holo-ACP from phosphopentetheinyl transferase (Sfp) catalyzed reaction between apo-ACP with CoA, followed by addition of AMP activated acyl chain [12]; 2) reacting apo-ACP with acyl-CoA catalyzed by Sfp [13]; 3) loading holo-ACP with acyl chains by acyl-ACP synthetase [14]. Involvement of enzymes in the reaction increases complexity of purification, and may limit the possibility of adding acyl chains that are not
favored by these enzymes. Since acyl-ACP is widely used in many biosynthetic pathways, we decided to develop a chemical method for acyl loading onto the holo-ACP.

The general scheme is to first make a reactive thiol addictive on the acyl chain, and then attach the acyl portion to the 4-phosphopentethiene through a thiol substitution reaction. The thiol addictive can be generated by attacking a carboxylate or anhydride form of the acyl portion with a thiol reagent. This reagent should be not only strong enough to nucleophilically attack the carboxylate or anhydride, but also a good leaving group when the thiol group on phosphopentethiene of holo-ACP attacks to substitute it. A compound reported by Johnson and Kent, (4-carboxylmethyl) thiophenol, may be a good candidate [15]. The special 3-oxo-octanoyl group required for TraI can be synthesized as described [9], by reacting hexanoyl chloride with Meldrum’s acid, followed by the thiol reagent attack to activate the acyl chain. These works are to be followed by organic synthesis chemists.

**Conclusion and Future Directions**

In this study, soluble TraI and holo-ACP were successfully obtained at high homogeneity. With the desired acyl chain synthesized, the holo-ACP can be loaded with an acyl group and serves as a substrate in TraI reaction. The other substrate SAM is lactonized during the reaction after cleavage between C4 and the sulfur ion. TraI and SAM hydrolase (Appendix A) possibly share similar mechanism. Therefore, mechanistic studies on TraI will help understanding both autoinducer synthases and the SAM hydrolase.
Reference


