Functional Analysis of Early Core Region Modification in the LPS of Gram-Negative Bacteria

by

Andrew Conley Pratt

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Medicinal Chemistry)
in the University of Michigan
2017

Doctoral Committee:

Professor Ronald W. Woodard, Chair
Professor Henry I. Mosberg
Associate Professor Zaneta Nikolovska-Coleska
Associate Professor Patrick J. O'Brien

ACKNOWLEDGEMENTS

I would like to thank Professor Ronald Woodard for taking me in to his lab and keeping me thoroughly entertained with his stories. He has given me a unique insight into science that I will carry with myself for years. I would also like to thank the only other current Woodard laboratory member David Cech. Without him, I would be sitting in the lab all alone talking to myself. I owe a great debt to Dr. Uwe Mamat from Forshungszentrum Borstel for giving me the opportunity to learn new techniques, spend a month in Germany, and answering all of my dumb questions, both scientific and cultural. Without him, this work would not be at all possible.

I would also like to thank my family and friends. Without them I never would be where I am today, and no one would ever listen to me complain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

APPENDIX

CHAPTER I

Introduction

Lipopolysaccharide in Gram-negative bacteria and its functions

Most bacteria typically are characterized as being Gram-positive or Gram-negative depending on their susceptibility to staining by crystal violet stain. Gram-negative bacteria have a cell wall composed of two membranes, the inner membrane and the outer membrane, separated by the peptidoglycan containing periplasmic space which distinguishes them from Gram-positive bacteria. The cellular envelope structure can be seen in Figure 1.1. Gram-negative bacteria include a diverse set of many bacteria of interest including four of the six so-called "ESKAPE" pathogens, namely *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter spp.* which cause the majority of nosocomial infections and show resistance to many antibacterials.[1] Overall, Gram-negative bacteria are known to account for approximately 60% of all cases of sepsis.[2]
The outer leaflet of the outer membrane of Gram-negative bacteria is composed of an essential, unique glycolipid, lipopolysaccharide (LPS) which

is also known as endotoxin. This glycolipid is a major component of the outer membrane accounting for about 70% of the outer leaflet.[3]

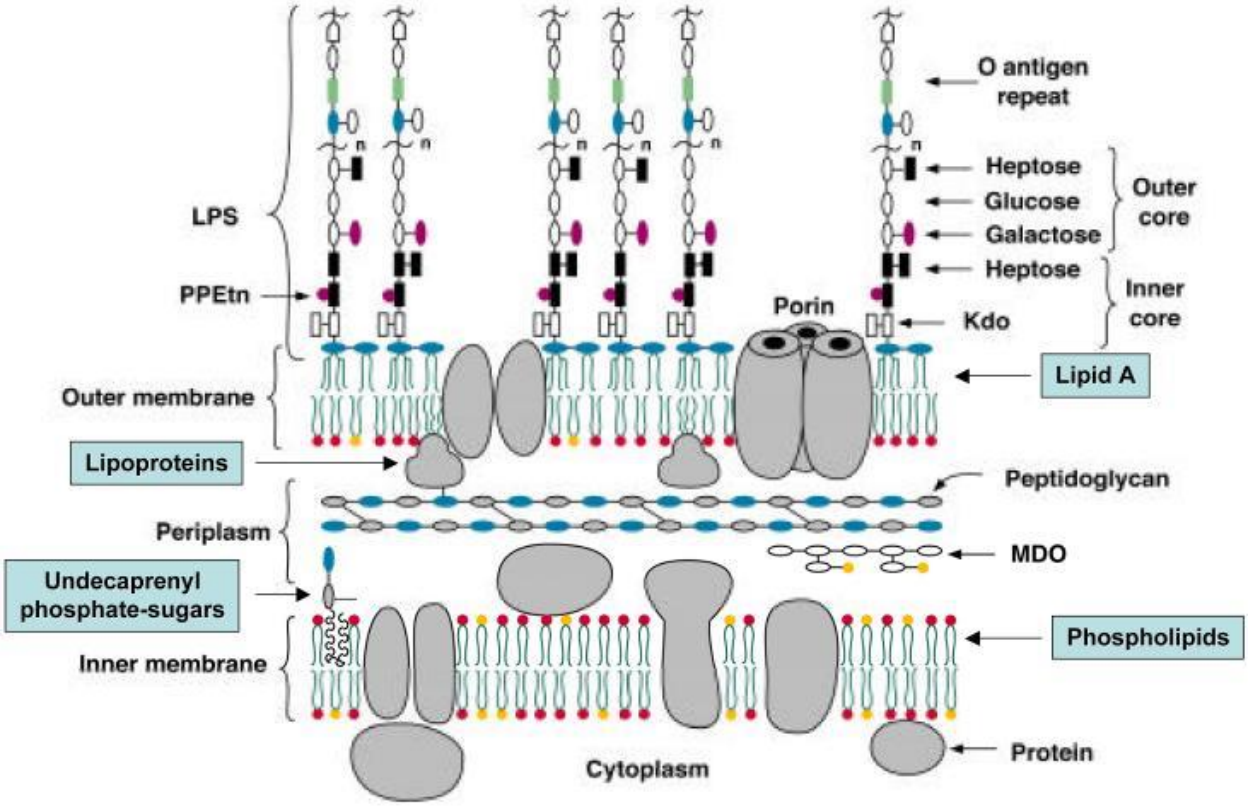


Figure 1.1. Structure of Gram negative cell envelope. Representative organization of the cellular envelope of the prototypical Gram-negative bacterium *Escherichia coli*. LPS can be seen in the outer leaflet of the outer membrane composing the entirety of the lipids of that leaflet. [Figure adapted from Ref. (10)]

LPS functions in the detection and evasion of host immune cells, stabilization of the outer membrane, and resistance to antibiotics [4] and being the most external portion of the membrane serves as a first line in identification and protection of the cell. Additionally, LPS is a pathogen-

associated molecular pattern which antagonizes mammalian Toll-like receptor 4 at very low concentrations leading to endotoxic shock characterized by fever, changes in white blood cell counts, disseminated intravascular coagulation, hypotension, shock, and death.[5] However, it is possible that in order to utilize these antagonistic effects for a positive end, and LPS is commonly used as an adjuvant in vaccine therapies.[6] Indeed, recent studies have been undertaken in order to biosynthetically engineer less toxic LPS adjuvants for vaccine therapies.[7] In these capacities, LPS plays a critical role both in bacterial survival, pathogenesis, and as an effective immune stimulating tool.

Escherichia coli, a Gram-negative bacteria used in research and recombinant technologies, is used to generate about 50% of all biopharmaceuticals[8] with total biopharmaceutical sales in the tens of billions of dollars per year.[9] The costs to remove endotoxin and methods to demonstrate acceptably low endotoxin levels can be exceedingly expensive and time consuming.[10] Due to the toxicity of LPS, the essentiality for survivability and proper function of the bacteria, and a high cost associated with LPS removal in biopharmaceuticals, LPS presents an interesting target for the study of its biosynthesis and inhibition.

Diversity in core oligosaccharide

LPS is an amphipathic glycolipid composed of three major components namely the lipid anchor called Lipid A; the core bridging oligosaccharide, and the repeating terminal oligosaccharide called the O-antigen. The Lipid A anchor is an acylated disaccharide which is essential to most Gram-negative bacteria. In many species the acylation varies from tetra to heptaacylated

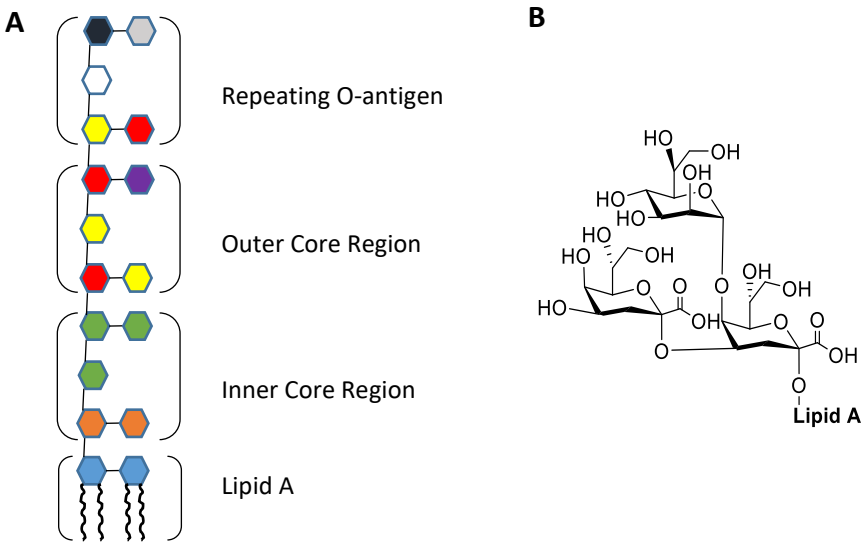


Figure 1.2. Structure of LPS glycolipid. (A) A representative diagram of *E. coli* K-12 MG1655 LPS with each component bracketed. Each color represents a unique monosaccharide showing the inner core region composed entirely of Kdo (orange) or heptose (green). (B) The molecular structure of the most conserved portion of LPS in *Enterobacteriaceae* composed of Hep-Kdo_2-Lipid A.

forms.[11] Attached to the Lipid A is the core oligosaccharide which is typically quite similar within closely related species. The final portion, the O-antigen, is highly variable even from subspecies to subspecies and is commonly not found in typical laboratory strains.[12] Endotoxicity of LPS is known to be associated strictly with the Lipid A portion of the molecule. Lipid A in most species cannot be fully acylated, and thereby made endotoxic, without the addition of the first sugar in the oligosaccharide core.[13] The core oligosaccharide typically starts with the addition of 2-keto-3-deoxy-D-*manno*-octulosonic acid (Kdo) which is a unique eight carbon sugar which is typically also essential for survivability.[14] In some species the Kdo is substituted by a closely related eight carbon sugar analog. The number of Kdo analogs added varies from species to species typically between one to four, with *E. coli* incorporating two Kdo moieties. The overall structure of LPS from *E. coli* can be seen in Figure 1.2. The addition of the Kdo moieties allows the final addition of the final acyl groups by LpxL and LpxM yielding the endotoxic species Kdo_2-Lipid A.[15] For wild-type *E. coli* this is the minimal species required for survival although mutants have been generated which can survive without this requirement.[16] As such, proteins involved in the biosynthesis of Kdo_2-Lipid A represent an attractive target for antibacterial development.

Core oligosaccharide is typically divided into the early core and the late core wherein the early core is highly conserved between many disparate

species and the late core is much more variable. In *E. coli* as well as most known *Enterobacteriaceae* and many Gram-negative species the addition of Kdo is followed by the addition of a number of L-*glycero*-D-*manno*-heptose moieties[17], comprising the early core, followed by a diverse set of other six carbon monosaccharides, making up the late core.[18] Cells lacking these sugars, while still viable, show increased susceptibility to hydrophobic antibiotics and increased susceptibility to phagocytosis by macrophages.[19, 20] Due in large part to the limited diversity seen in the early core region of LPS, its antigenic properties for vaccine studies as well as development of inhibitors of its biosynthesis have recently been undertaken.[21-24] Therefore, an understanding core oligosaccharide biosynthesis and the ability to generate and design unique analogs is of increasing interest today.

Dissertation Research Rationale

LPS core oligosaccharide is a diverse glycolipid with many known biological functions. Understanding the effects of diversity in the oligosaccharide has to date been a major challenge due to the complexity of the molecule and the many genes involved in its generation.[25] In order to address this problem, we developed a strain of Gram-negative bacteria which lacks the ability to elaborate the lipid IV_A precursor to LPS in order to generate various glycoforms, both natural and novel, within a common and

easy to use Gram-negative cell line. We hypothesized that deletion of the genes involved in LPS core biosynthesis would allow for heterologous expression of genes from other Gram-negative species. This would allow for the generation and study of many LPS glycoforms that could not be easily generated with any other modern system. Utilizing this new cell line, we were able to create multi-glycosylated Lipid IV_A species by heterologous expression of Kdo and heptosyltransferases. Using this system we were able to gain insights into functionalities of these proteins and mechanistic insights into the function of Kdo transferases.

In order to expand our understanding of early core biosynthesis, we decided to explore the selectivity and incorporation of the Kdo analog, Kd8N (8-carbon amino sugar 8-amino-3,8-dideoxy-D-*manno*-octulosonic acid) from *Shewanella oneidensis*. Utilizing a series of in vitro and in vivo experiments, we determined the cytidylyltransferase from S. oneidensis has selectivity for Kd8N over Kdo, but a cell line lacking Kd8N will incorporate Kdo into the LPS core. This leads us to conclude that the overall selectivity stems from the presence or absence of available Kd8N in the cell rather than absolute selectivity of the enzymes themselves. With this knowledge, we can move forward to study properties of the incorporation of this interesting sugar into the core of LPS.

References

[1] Rice, L. B. (2008) Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE, *Journal of Infectious Diseases 197*, 1079-1081.

[2] Cohen, J. (2002) The immunopathogenesis of sepsis, *Nature 420*, 885-891.

[3] van der Poll, T., and Opal, S. M. (2008) Host–pathogen interactions in sepsis, *The Lancet Infectious Diseases 8*, 32-43.

[4] Brandenburg, K., and Wiese, A. (2004) Endotoxins: relationships between structure, function, and activity, *Curr Top Med Chem 4*, 1127-1146.

[5] Hardie, E. M., and Kruse-Elliott, K. (1990) Endotoxic shock. Part I: A review of causes, *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 4*, 258-266.

[6] Arenas, J. (2014) Molecular Vaccines, *springer*, 527-536.

[7] Zariri, A., and van der Ley, P. (2015) Biosynthetically engineered lipopolysaccharide as vaccine adjuvant, *Expert Review of Vaccines 14*, 861-876.

[8] Walsh, G. (2010) Post-translational modifications of protein biopharmaceuticals, *Drug Discov Today 15*, 773-780.

[9] Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., and Villaverde, A. (2016) Recombinant pharmaceuticals from microbial cells: a 2015 update, *Microb Cell Fact 15*, 1-7.

[10] Saraswat, M., Musante, L., Ravidá, A., Shortt, B., Byrne, B., and Holthofer, H. (2013) Preparative Purification of Recombinant Proteins: Current Status and Future Trends, *BioMed Research International 2013*, 1-18.

[11] Raetz, C., Reynolds, C., Trent, M., and Bishop, R. (2007) Lipid A modification systems in gram-negative bacteria, *Annual review of biochemistry 76*, 295-329.

[12] Raetz, C. R. H., and Whitfield, C. (2002) Lipopolysaccharide Endotoxins, *Annual Review of Biochemistry 71*, 635-700.

[13] Raetz, C. R. H., Garrett, T. A., Reynolds, C. M., Shaw, W. A., Moore, J. D., Smith, D. C., Ribeiro, A. A., Murphy, R. C., Ulevitch, R. J., Fearns, C., Reichart, D., Glass, C. K., Benner, C., Subramaniam, S., Harkewicz, R., Bowers-Gentry, R. C., Buczynski, M. W., Cooper, J. A., Deems, R. A., and Dennis, E. A. (2006) Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4, *J. Lipid Res. 47*, 1097-1111.

[14] Lodowska, J., Wolny, D., and Weglarz, L. (2013) The sugar 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) as a characteristic component of bacterial endotoxin - a review of its biosynthesis, function, and placement in the lipopolysaccharide core, *Canadian journal of microbiology 59*, 645-655.

[15] Klein, G., Lindner, B., Brabetz, W., Brade, H., and Raina, S. (2009) Escherichia coli K-12 Suppressor-free Mutants Lacking Early Glycosyltransferases and Late Acyltransferases MINIMAL LIPOPOLYSACCHARIDE STRUCTURE AND INDUCTION OF ENVELOPE STRESS RESPONSE, *J. Biol. Chem. 284*, 15369-15389.

[16] Meredith, T. C., Aggarwal, P., Mamat, U., Lindner, B., and Woodard, R. W. (2006) Redefining the Requisite Lipopolysaccharide Structure in Escherichia coli, *ACS chemical biology 1*, 33-42.

[17] Gronow, S., Xia, G., and Brade, H. (2010) Glycosyltransferases involved in the biosynthesis of the inner core region of different lipopolysaccharides, *European journal of cell biology 89*, 3-10.

[18] Holst, O. (2007) The structures of core regions from enterobacterial lipopolysaccharides – an update, *FEMS Microbiology Letters 271*, 3-11.

[19] Coleman, W. G., and Leive, L. (1979) Two mutations which affect the barrier function of the Escherichia coli K-12 outer membrane, *J. Bacteriol. 139*, 899-910.

[20] Wang, Z., Wang, J., Ren, G., Li, Y., and Wang, X. (2015) Influence of Core Oligosaccharide of Lipopolysaccharide to Outer Membrane Behavior of Escherichia coli, *Marine Drugs 13*, 3325-3339.

[21] Tikad, A., Fu, H., Sevrain, C. M., Laurent, S., Nierengarten, J.-F. F., and Vincent, S. P. P. (2016) Mechanistic Insight into Heptosyltransferase Inhibition by using Kdo Multivalent Glycoclusters, *Chemistry (Weinheim an der Bergstrasse, Germany) 22*.

[22] Durka, M., Tikad, A., Perion, R., Bosco, M., Andaloussi, M., Floquet, S., Malacain, E., Moreau, F., Oxoby, M., Gerusz, V., and Vincent, S. P. (2011) Systematic synthesis of inhibitors of the two first enzymes of the bacterial heptose biosynthetic pathway: towards antivirulence molecules targeting lipopolysaccharide biosynthesis, *Chemistry 17*, 11305-11313.

[23] Kong, L., Vijayakrishnan, B., Kowarik, M., Park, J., Zakharova, A. N., Neiwert, L., Faridmoayer, A., and Davis, B. G. (2016) An antibacterial vaccination strategy based on a glycoconjugate containing the core lipopolysaccharide tetrasaccharide Hep_2Kdo_2, *Nature chemistry 8*, 242-249.

[24] Cross, A. S. (2013) Anti-endotoxin vaccines: back to the future, *Virulence 5*, 219-225.

[25] Schnaitman, C. A., and Klena, J. D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria, *Microbiol Rev 57*, 655-682.

CHAPTER II

Construction and applications of an *E. coli* strain lacking LPS core biosynthetic genes

Summary

Lipopolysaccharide (LPS) is an essential glycolipid found in Gram-negative bacteria that has many medical and biological functions and applications. Many of these key features are due to the presence of an oligosaccharide core attached directly to the main Lipid A anchor. The oligosaccharide core varies from species to species and can often be quite complex making it hard to study properties of each core as well as the activities and selectivities of the enzymes involved in LPS core biosynthesis. WOD02 is the first viable *E. coli* strain lacking all glycosyltransferases involved in LPS core biosynthesis that is capable of generating new non-native glycoforms of oligosaccharide core. This system allows for generation of various glycoforms using a simple, non-pathogenic *E. coli* background forgoing the need of slow-growing, pathogenic bacteria for many

downstream applications. Using this system, we now report characterizations of heptosyltransferases from four Gram-negative bacteria and insights into the mechanism of the 2-keto-3-deoxy-D-*manno*-octulosonic acid transferase. Additionally, use of this cell-line gives evidence to the presence of a Lipid A-3'-acyloxylacyl deacylase in *E. coli* K-12 with activity similar to that found in other Gram-negative bacteria.

Introduction

Gram-negative bacteria have an asymmetric outer membrane wherein the majority of the outer leaflet is composed of the essential and unique glycolipid lipopolysaccharide (LPS), or endotoxin.[1-3] LPS is the major surface antigen in Gram-negative bacteria and is composed of three major components, namely lipid A, core oligosaccharide, and O-antigen.[4] The lipid A portion is the region of least variability between different species and is often a hexaacylated disaccharide molecule. The core oligosaccharide is the linking region between the lipid A moiety and the O-antigen and typically shows little variability between closely related species. The most variable region of LPS is the often repeating oligosaccharide, O-antigen, often comprised of repeating oligosaccharide subunits. O-antigen can vary widely even within different strains of the same species and is often not present in

common laboratory strains of Gram-negative bacteria such as in *Escherichia coli* K-12 MG1655.[5, 6]

The unique eight carbon sugar 2-keto-3-deoxy-D-*manno*-octulosonic acid, Kdo, is found in the majority of LPS glycoforms[7] in Gram-negative bacteria and serves to bridge the Lipid A portion with the core oligosaccharide. For wild-type *Escherichia coli* a minimal LPS structure, Kdo_2-Lipid A or Re-LPS, is required for survivability of the cell line[4, 8]. The attachment of Kdo is catalyzed by the enzyme Kdo transferase, WaaA[#], utilizing the activated sugar nucleotide CMP-Kdo. WaaA depending on the species could be mono-[9, 10], bi-[11-13], tri-[14], or tetra-functional[14, 15] transferring either one, two, three, or four Kdo molecules, respectively, to the tetra-acylated Lipid A precursor, Lipid IV_A. The mechanism and selectivity by which multifunctional Kdo transferases function is currently not known. Work by Chung and Raetz has shown that the functionality of Kdo transferases can be changed by swapping a few amino acids between the Kdo transferases of *Haemophilus influenzae* and *E. coli* showing that the functionality is enzyme specific.[16] Additionally, previous research in our lab determined the first crystal structure of a WaaA from *Aquifex aeolicus* co-crystallized with the substrate analog CMP[17] giving insight into the binding of substrate. However, the mechanism dictating the number of Kdo subunits transferred is not yet known.

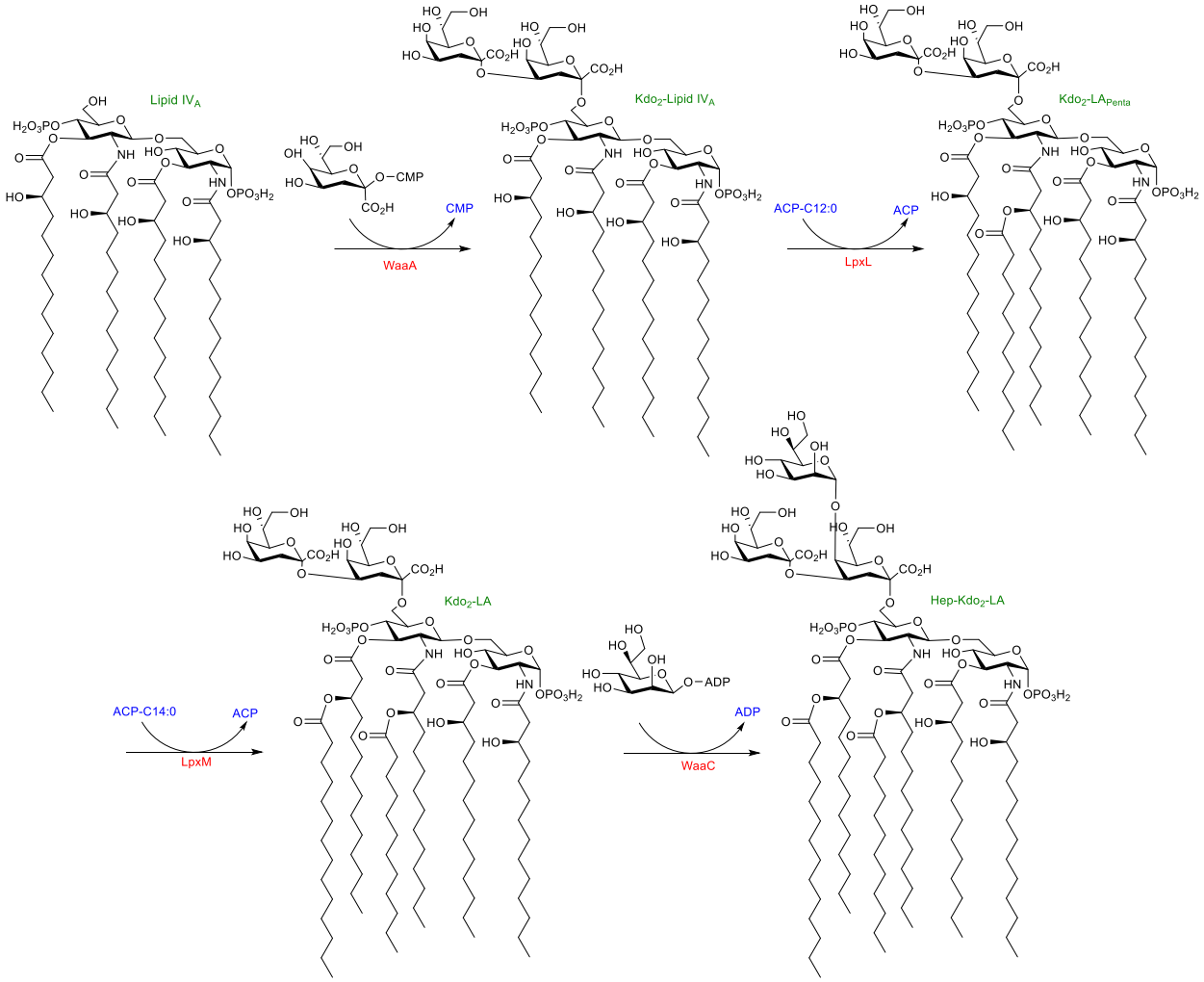


Figure 2.1. Biosynthesis of Early LPS core. Biosynthetic scheme of the early modification of Lipid IV_A in *E. coli* K-12 including the glycosylation by WaaA, fatty acid acylation by LpxL and LpxM to generate the endotoxic Lipid A anchor followed by heptosylation by WaaC.

After the addition of the Kdo moieties in *E. coli*, the late acyltransferases, LpxL and LpxM, transfer the final fatty acids to the Kdo_2-Lipid IV_A giving the hexaacylated Kdo_2-Lipid A.[18] For some species this simple Kdo glycosylated Lipid A is the complete LPS, such as with *Chlamydia trachomatis*[19], while most other species further elaborate of the core

oligosaccharide. In *E. coli* as in most *Enterobactericeae*, the next step is the addition of the seven carbon sugar L-*glycero*-D-*manno*-heptose catalyzed by the heptosyltransferase I, WaaC[20] The pathway from Lipid IV_A to the first heptosylation is shown in Figure 2.1. The inability to transfer this heptose in mutant *E. coli* strains leaves the cells with the ability to generate solely Kdo_2-Lipid A.[21] This phenotype is known to cause increased susceptibility to hydrophobic antibiotics and increased susceptibility to phagocytosis by macrophages.[22] This makes Lipid A biosynthesis and incorporation of these unusual sugars into LPS interesting targets for novel antibacterial development.[23-25] Additionally, vaccines of the highly conserved inner core of some species such as *Neisseria meningitidis*[26, 27] are under investigation for use as a vaccine antigen in addition to the longstanding use as an adjuvant.[28]

Unfortunately, to be able to study the selectivity of the enzymes involved in core biosynthesis, and to be able to generate these inner cores for downstream applications, it is often necessary to culture virulent, slow-growing bacteria. In order to develop a cell line capable of generating inner core oligosaccharides not found in wild type *E. coli*, we attempted to delete the genes involved in core oligosaccharide biosynthesis we adopted a strategy that took advantage of a unique mutant of *E. coli*, KPM56.[10] This cell line was chosen as it is capable of surviving with only Lipid IV_A due to the presence of the phenotypic suppressor mutant in a gene of unknown function, *yhjD400.*[29]

The cell line WOD02 which lacks all core oligosaccharide biosynthetic genes is reported herein and is capable of biosynthesizing early inner core structure using Kdo transferases and heptosyltransferases from a variety of Gram-negative bacteria.

Experimental Procedures

Bacterial Strains, Plasmids, and Growth Conditions. All strains and plasmids used in the present study are described in Table 2.1. Standard cultures of *E. coli* were routinely grown aerobically with shaking (250 rpm)

Table 2.1. List of strains, plasmids, and primers used in Chapter II.

Strain, plasmid, or primer	Description	Source
KPM56	*E. coli* K-12 MG1655 F^- *rph*$^+$ *fnr*$^+$ Δ*gutQ* Δ*kdsD yhjD400* $kdsD^+$ Δ*waaC* Δ*waaA*	[10]
WOD01	KPM56 Δ(*waaF-waaA*)::*kan*	This study
WOD02	KPM56 Δ(*waaF-waaA*)	This study
WOD03	WOD02 + pETduet1 (empty vector)	This study
WOD04	WOD02 + pHiA (carrying *H. influenzae waaA* gene)	This study
WOD05	WOD02 + pEcA (carrying *E. coli waaA* gene)	This study
WOD06	WOD02 + pCtA (carrying *C. trachomatis waaA* gene)	This study
WOD07	WOD02 + pCpA (carrying *C. psittaci waaA* gene)	This study
WOD08	WOD02 + pHpC (carrying *H. pylori waaC* gene)	This study
WOD09	WOD02 + pEcC (carrying *E. coli waaC* gene)	This study
WOD10	WOD02 + pCjC (carrying *C. jejuni waaC* gene)	This study
WOD11	WOD02 + pAaC (carrying *A. aeolicus rfaC1* gene)	This study
WOD12	WOD02 + pHiAHpC (carrying *H. influenzae waaA* and *H. pylori waaC* gene)	This study
WOD13	WOD02 + pHiAEcC (carrying *H. influenzae waaA* and *E.coli waaC* gene)	This study
WOD14	WOD02 + pHiACjC (carrying *H. influenzae waaA* and *C. jejuni waaC* gene)	This study

WOD15	WOD02 + pHiAAaC (carrying *H. influenzae waaA* and *A. aeolicus rfaC1* gene)	This study
WOD16	WOD02 + pEcAHpC (carrying *E. coli waaA* and *H. pylori waaC* gene)	This study
WOD17	WOD02 + pEcAEcC (carrying *E. coli waaA* and *E. coli waaC* gene)	This study
WOD18	WOD02 + pEcACjC (carrying *E. coli waaA* and *C. jejuni waaC* gene)	This study
WOD19	WOD02 + pEcAAaC (carrying *E. coli waaA* and *A. aeolicus rfaC1* gene)	This study
WOD20	WOD02 + pCtAHpC (carrying *C. thracomatis waaA* and *H. pylori waaC* gene)	This study
WOD21	WOD02 + pCtAEcC (carrying *C. thracomatis waaA* and *E. coli waaC* gene)	This study
WOD22	WOD02 + pCtACjC (carrying *C. thracomatis waaA* and *C. jejuni waaC* gene)	This study
WOD23	WOD02 + pCtAAaC (carrying *C. thracomatis waaA* and *A. aeolicus rfaC1* gene)	This study
WOD24	WOD02 + pCpAHpC (carrying *C. psittaci waaA* and *H. pylori waaC* gene)	This study
WOD25	WOD02 + pCpAEcC (carrying *C. psittaci waaA* and *E. coli waaC* gene)	This study
WOD26	WOD02 + pCpACjC (carrying *C. psittaci waaA* and *C. jejuni waaC* gene)	This study
WOD27	WOD02 + pCpAAaC (carrying *C. psittaci waaA* and *A. aeolicus rfaC1* gene)	This study
pET26b	Kan^R; T7 expression vector	Novagen
p26HpC	*Helicobacter pylori* J99 *waaC* inserted into NdeI/BamHI sites of pET26b	This study
p26EcC	*E. coli* K-12 *waaC* inserted into NdeI/BamHI sites of pET26b	This study
p26CjC	*Campylobacter jejuni* NCTC 11168 *waaC* inserted into NdeI/BamHI sites of pET26b	This study
p26AaC	*Aquifex aeolicus rfaC1* inserted into NdeI/BamHI sites of pET26b	This study
pUM211	pET16b containing the *waaA* gene from *H. influenzae* I69	*10*
pUM212	pET16b containing the *waaA* gene from *E. coli* K-12	*10*
pFEN207	Amp^R; pUC8 plasmid containing the *waaA* gene from *C. trachomatis* L2	*30*
pUM210	pET16b containing the *waaA* gene from *C. psittaci* 6BC	*10*
pETduet1	Amp^R; T7 expression vector for dual gene expression	Novagen
pHiA	*Haemophilus influenzae* I69 *waaA* inserted into NdeI/XhoI of pETduet1	This study
pEcA	*E. coli* K-12 *waaA* inserted into NdeI/XhoI of pETduet1	This study
pCtA	*Chlamydia trachomatis* L2 *waaA* inserted into NdeI/XhoI of pETduet1	This study
pCpA	*Chlamydophila psittaci* 6BC *waaA* inserted into NdeI/XhoI of pETduet1	This study
pHpC	*H. pylori* J99 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pEcC	*E. coli* K-12 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pCjC	*C. jejuni* NCTC 11168 *waaC* inserted into XbaI/BamHI sites of pETduet 1	This study
pAaC	*A. aeolicus rfaC1* inserted into XbaI/BamHI sites of pETduet1	This study

pHiAHpC	*H. influenzae* I69 *waaA* inserted into NdeI/XhoI and *H. pylori* J99 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pHiAEcC	*H. influenzae* I69 *waaA* inserted into NdeI/XhoI and *E. coli* K-12 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pHiACjC	*H. influenzae* I69 *waaA* inserted into NdeI/XhoI and *C. jejuni* NCTC 11168 *waaC* inserted into XbaI/BamHI sites of pETduet 1	This study
pHiAAaC	*H. influenzae* I69 *waaA* inserted into NdeI/XhoI and *A. aeolicus rfaC1* inserted into XbaI/BamHI sites of pETduet1	This study
pEcAHpC	*E. coli* K-12 *waaA* inserted into NdeI/XhoI and *H. pylori* J99 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pEcAEcC	*E. coli* K-12 *waaA* inserted into NdeI/XhoI and *E. coli* K-12 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pEcACjC	*E. coli* K-12 *waaA* inserted into NdeI/XhoI and *C. jejuni* NCTC 11168 *waaC* inserted into XbaI/BamHI sites of pETduet 1	This study
pEcAAaC	*E. coli* K-12 *waaA* inserted into NdeI/XhoI and *A. aeolicus rfaC1* inserted into XbaI/BamHI sites of pETduet1	This study
pCtAHpC	*C. trachomatis* L2 *waaA* inserted into NdeI/XhoI and *H. pylori* J99 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pCtAEcC	*C. trachomatis* L2 *waaA* inserted into NdeI/XhoI and *E. coli* K-12 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pCtACjC	*C. trachomatis* L2 *waaA* inserted into NdeI/XhoI and *C. jejuni* NCTC 11168 *waaC* inserted into XbaI/BamHI sites of pETduet 1	This study
pCtAAaC	*C. trachomatis* L2 *waaA* inserted into NdeI/XhoI and *A. aeolicus rfaC1* inserted into XbaI/BamHI sites of pETduet1	This study
pCpAHpC	*C. psittaci* 6BC *waaA* inserted into NdeI/XhoI and *H. pylori* J99 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pCpAEcC	*C. psittaci* 6BC *waaA* inserted into NdeI/XhoI and *E. coli* K-12 *waaC* inserted into XbaI/BamHI sites of pETduet1	This study
pCpACjC	*C. psittaci* 6BC *waaA* inserted into NdeI/XhoI and *C. jejuni* NCTC 11168 *waaC* inserted into XbaI/BamHI sites of pETduet 1	This study
pCpAAaC	*C. psittaci* 6BC *waaA* inserted into NdeI/XhoI and *A. aeolicus rfaC1* inserted into XbaI/BamHI sites of pETduet1	This study
pKD46	Amp^R; λ Red recombinase expression plasmid	[31]
pKD4	Amp^R, Kan^R; template plasmid for kanamycin resistance cassette	[31]
pCP20	Amp^R, Cm^R; FLP recombinase expression plasmid	[31]
HpWaaCfor	gattctagaattcatATGAAAATAGCGATTGTCAGGCTTTCAG	This study
HpWaaCrev	gaattcggatccTCATTCTTTTTCCTTTAAAACGTTTAAAACGC	This study
EcWaaCfor	gattctagaattcatATGCGGGTTTTGATCGTTAAAAC	This study
EcWaaCrev	gaattcggatccTTATAATGATGATAACTTTTCCAAAACTGC	This study
CjWaaCfor	tatacatATGATCTTTTTTTATTATTTTTTAACTTGGACGGC	This study
CjWaaCrev	atatggatccTTATAAGCTTTTTCTTGCATCAATTCCC	This study
AaRfaC1for	gattctagaattcatATGAAGAAGGCGTTAATAGTGAGG	This study
AaRfaC1rev	gaattcggatccTTACGGCTTGGTATTCAAAATGTTTATAG	This study
HiWaaAfor	atatatatcatATGTGGCGTTTTTTTTATACCAGCT	This study
HiWaaArev	atatctcgagTCATACATTGCGCTCCAAATAAGG	This study
EcWaaAfor	actccatATGCTCGAATTGCTTTACACCG	This study
EcWaaArev	cagtctcgagTCAATGCGTTTTCGGTGGC	This study
CtWaaAfor	gattctagaattcatATGATAAGACGTTGGTTAACATCTCG	This study
CtWaaArev	gaattcctcgagTTAGATTTTCATGCAAGTAATTTGGCTC	This study
CpWaaAfor	actgcatATGGTGGGGCTTCCTAGGATT	This study

CpWaaArev	cactctcgagCTATATTTTTACACAAGGGATATATCTTTTAAAAG	This study
WaaFKOfor	GTGTAACGGAATACATGGCCTGGCTGAATCGCGACGCATAAGAGC TCTGCgtgtaggctggagctgcttc	This study
WaaAKOrev	TAATGGGATCGAAAGTACCCGGATAAATCGCCCGTTTTTGCATAAC AACCcatatgaatatcctccttag	This study
WaaFKOforseq	CCGTTCAAAACCGTTGCTGAAG	This study
WaaAKOrevseq	CGC GTC ACG ATA TCG ATA TGA CC	This study

at 37 °C in LB medium (10 g/liter Bacto-tryptone, 5 g/liter yeast extract, 10 g/liter NaCl). When necessary, kanamycin (30 μg/ml), chloramphenicol (30 μg/ml), or ampicillin (100 μg/ml), was added to the media.

DNA Methods. Standard recombinant DNA methods were used for nucleic acid preparation and analysis. [32] Primer sequences are listed in Table 2.1. To construct gene deletions, the phage λ Red recombinase procedure of Datsenko and Wanner was used as described[31], except plasmids pKD46 and pCP20 were cured at 37 °C to accommodate the temperature sensitive phenotype of lipid IV_A-expressing *E. coli* derivatives.[33] For generation of a cell line lacking all genes involved in core oligosaccharide biosynthesis, WaaFfor and WaaArev were used to generate the KAN resistance cassette from pKD4 which targeted the gene region from the nucleotides 3794929-3809817 which includes the genes *waaF*, *waaC*, *waaL*, *waaU*, *waaZ*, *waaY*, *waaJ*, *waaR*, *waaB*, *waaS*, *waaP*, *waaG*, *waaQ*, and *waaA* within three separate operons. The gene clusters are seen in Figure 2.2. Transformation of this linear cassette into KPM56/pKD46 afforded the kanamycin-resistant

strain, WOD01. Removal of the kanamycin resistance gene by curing with pCP20 afforded the strain WOD02.

Figure 2.2. Waa gene cluster. The gene clusters containing the *waa* genes involved in core oligosaccharide biosynthesis

To clone the heptosyltransferases, *HpwaaC* (jhp_0264), *EcwaaC* (b3621), *CjwaaC* (cj1133), and *AarfaC1* (aq_1542) from *H. pylori* J99, *E. coli* K-12 MG1655, *C. jejuni* Strain NCTC 11168, and *A. aeolicus* VF5, respectively, into pET26b and the Kdo transferases *HiwaaA*, *EcwaaA*, *CtwaaA*, and *CpwaaA* from *H. influenzae* I69, *E. coli* K-12 MG1655, *C. trachomatis* L2, and *C. psittaci* 6BC, respectively, into pETduet1, the appropriate primers were chosen. For the heptosyltransferases the appropriate genomic DNAs were used as templates, while for the Kdo transferases pUM211[10], pUM212[10], pFEN207[30], or pUM210[10] were used as the appropriate templates. The heptosyltransferases were subcloned into the multiple cloning site of pET26b using the restriction sites for NdeI and BamHI while the Kdo transferases were subcloned into pETduet1 using the restriction sites NdeI and XhoI. In order to use pETduet1 to coexpress both a heptosyltransferase and a Kdo transferase each of the four pETduet1 plasmids containing each Kdo transferase gene were cut with XbaI and

BamHI removing the promoter and ribosomal binding site of the empty multiple cloning site. The pET26b plasmids containing the heptosyltransferases were similarly restricted giving a fragment carrying the promoter region, ribosomal binding site, and heptosyltransferase gene. These fragments were subcloned into the Kdo transferase containing pETduet1 or a similarly digested empty pETduet1 plasmid giving a total of 24 pETduet1 plasmids generated. These plasmids underwent DNA sequencing to verify sequences of the desired genes. Each pETduet1 plasmid were transformed into WOD02 by electroporation.

Cell growth for LPS Production. Transformed cells were grown on LB agar plates supplemented with 100 μg/mL ampicillin. A 2 mL culture of liquid LB medium, also supplemented with ampicillin, was inoculated from single colonies at 37°C overnight with shaking at 200 rpm. This culture was diluted 1:100 in 200 mL fresh LB medium and incubated at 37°C, 200 rpm overnight. Cells were then sedimented by centrifugation at 4000 x g, 4°C, 30 min (Eppendorf Centrifuge 5804R). The resulting pellet was resuspended in 30 mL of phosphate buffered saline before being recentrifuged. The cells were then resuspended in 20 mL of ethanol and stirred at 4°C overnight. Cells were once again centrifuged before being resuspended in 20 mL of acetone which was left at 4^{o}C stirring for 2 h. This process of centrifuging, resuspending in acetone, and stirring was repeated once more. The suspension was centrifuged a final time. To the resulting biomass was

added a small volume of diethyl ether to make a slurry which was subjected to filtration by indirect vacuum. The biomass was left to dry overnight and then pulverized to a homogenous fine particulate.

LPS purification. LPS samples were purified by a modified version of the phenol-chloroform-petroleum ether procedure[34] as reported previously.[35] The dried samples were weighed and resuspended into LPS-free water (Braun water) to a concentration of 1 mg/mL for mass spectrometry analysis and TLR4 activation assays.

Electrospray-Ionization Fourier-transformed Ion Cyclotron Mass Spectrometry. LPS samples (1 mg/ml) were diluted 15-fold in a solution consisting of 50% 2-propanol supplemented with 4 mM triethylamine and 0.35 mM acetic acid resulting in a pH of approximately 8.5. All samples were analyzed in the negative ion mode using a Q Exactive Plus instrument (Thermo, Bremen, Germany) with Electrospray Ionization (ESI). Samples were delivered with a syringe pump at a flow rate of 5uL/min using a sheath gas flow of 5L/min, transfer capillary temperature of 250C and a S-lens level of 100. Spectra were recorded at highest resolution (280,000 FWHM defined at m/z 200) with no further collisional activation or all ion fragmentation activated with NCE set to 25V. All mass spectra can be found in Appendix B.

TLR4 stimulation assays. Using the diluted LPS samples, TLR4 activation was assayed using the HEK-Blue™ hTLR4 and HEK-Blue™ Null2 cells in accordance with the specifications of the supplier of the cell lines (InvivoGen) as reported previously.[36]

Results

Construction of chromosomal deletion of LPS core oligosaccharide biosynthesis. The cell line KPM56 was chosen for the construction of an *E. coli* mutant lacking the genes involved in core oligosaccharide biosynthesis based on its ability, as well as that of its progenitor KPM22[33], to survive chromosomal deletions of genes required for LPS biosynthesis such as *kdsD*, *gutQ*, *waaA*, and *waaC*. Additionally, unlike in many species of Gram-negative bacteria, *E. coli* K-12 MG1655 contains all genes in core oligosaccharide biosynthesis in a single gene cluster[37] thereby facilitating removal in this cell line. Two of these genes, namely, *waaA* and *waaC* were removed from KPM56 previously. In total, 14 genes have been removed by the Datsenko-Wanner method from this gene cluster leaving *hldD*, the gene encoding ADP-L-glycero-D-*manno*heptose epimerase and *coaD,* the gene encoding pantethine-phosphate adenylyltransferase intact, as confirmed by DNA sequencing of this region. This generates a cell line,

named WOD02, capable of generating the appropriate sugar precursors of core biosynthesis, but lacking the capability to transfer these sugars to the Lipid IV_A molecule. Unlike the KPM56 progenitor, WOD02 forms highly mucoid colonies on standard LB-miller agar plates. This mucoid morphology is likely related to colonic acid production due to lack of WaaL as reported previously.[38]

WOD02 is capable of expressing non-native inner core. With a blank template in hand, we turned our efforts toward verifying the number of Kdo glycosylation events catalyzed by the WaaA of various Gram-negative bacteria and toward probing the substrate specificities of WaaC, the next enzyme in the biosynthetic pathway of a number of diverse microorganisms. For example, one would assume that a WaaC from an organism that displays two Kdos would not glycosylate a mono-Kdo glycosylated Lipid A species since no such species have been observed in their native organism. This approach, therefore, would determine if a microorganism in which WaaC normally glycosylates a single Kdo-Lipid A would transfer a heptose to a di- or tri-Kdo glycosylated species. Novel, as well as natural, LPS molecules would likely result from this study.

The approach was to first introduce a WaaA from an organism which has been seen to have mono-, bi-, tri-, or tetrafunctional activity with *H. influenzae*[9], *E. coli*[11], *C. trachomatis*[14], and *Chlamydophila psittaci*[15], respectively chosen to represent each group and analyze the number of

A

Kdo transferase	*Kdo-Lipid IV_A*	*Kdo_2-Lipid A*	*Kdo_3-Lipid A*
HiWaaA	+	-	-
EcWaaA	+	+	+
CtWaaA	+	+	+
CpWaaA	+	+	+

B

Heptosyltransferase	*Hep-Kdo-Lipid IV_A*	*Hep-Kdo_2-Lipid A*	*Hep-Kdo_3-Lipid A*
HpWaaC	+	+	-
EcWaaC	+	+	+
CjWaaC	-	+	-
AaRfaC1	+	-	-

Table 2.2. Functionalities of Kdo and heptosyltransferases. (A) Table showing functionalities observed by each Kdo transferase expressed in the WOD02 system. A + represents glycoform was seen and a – represents no glycoform of that type seen. (B) Table showing modification of the Kdo containing glycoform by the addition of a single heptose by each heptosyltransferase. A + represents that the glycoform was detected in a WOD02 cell line expressing the corresponding heptosyltransferase and a – represents no corresponding glycoform detected.

Kdos displayed in this cell line. It should be noted that various atypical *E. coli* glycoforms with one, two or three Kdos have been previously observed[39, 40] and as such in the samples unique Kdo glycosylation is not necessarily expected.

To create a matrix as well as probe the substrate specificities, four different WaaCs was co-expressed in tandem with the WaaA in each of the above stains containing the various Kdo derivatives. The heptosyltransferases were chosen based on predicted specificity and possible enhanced promiscuity. The results from the various combinations and permutations are listed in Table 2.2a and 2.2b and discussed below.

In order to verify the ability of WOD02 cells to generate glycoforms from other Gram-negative bacteria, the LPS of WOD02 was purified and analyzed by electrospray ionization Fourier transform ion cyclotron mass spectrometry in negative ion mode (see Appendix B for all mass spectra). As expected, this cell line lacking the *waa* gene cluster was unable to glycosylate LPS giving major peaks of 1404.85 u and 1527.86 u for only Lipid IV_A (1,4'-bisphosphorylated tetraacylated Lipid A, calculated mass 1404.854 u) and the phosphoethanolamine (P-EtN) modified Lipid IV_A (calculated mass 1527.862 u), respectively which is similar to the LPS found in the related cell lines KPM22,[33] and KPM56.[10] The cell lines WOD03 (Empty Vector), WOD08 (HpWaaC), WOD09 (EcWaaC), WOD10 (CjWaaC), and WOD11 (AaRfaC1) which all lack Kdo transferases all show a similar production of Lipid IV_A and P-EtN-Lipid IV_A which demonstrates that neither the vector alone nor any of the heptosyltransferases can cause modification of the truncated LPS.

WOD02 cells heterologously expressing Kdo transferases, namely WOD04 (HiWaaA), WOD05 (EcWaaA), WOD06 (CtWaaA) and WOD07 (CpWaaA), all have restored ability to transfer Kdo to Lipid IV_A wherein WOD04, WOD05, WOD06, and WOD07 all show a peak of 1624.91 u representing a single Kdo attached to Lipid IV_A (Kdo-Lipid IV_A, calculated mass 1624.912 u). Additionally, WOD04, WOD05, and WOD06 show each contain an M+1 peak, 2018.28 u, for the diacylated Kdo-Lipid IV_A, Kdo-Lipid

A (calculated mass 2017.277 u). WOD05 and WOD06 showed an M+1 peak at 2238.34 u which corresponds to Kdo_2-Lipid A (calculated mass 2237.336 u), and WOD06 also shows the presence of the triglycosylated Kdo_3-Lipid A (M+1 exact mass 2458.40 u, calculated mass 2457.39 u). Although CpWaaA has been shown previously to have tetrafunctional Kdo transferase activity[39] no samples were seen to contain the peak for Kdo_4-Lipid A (calculated mass 2677.45 u). Other Kdo transferase activities could be more readily seen in other samples, namely, the presence of Kdo_3-Lipid A in WOD17 (EcWaaA and EcWaaC) and the presence of Kdo_2-Lipid A in WOD24 (CpWaaA and HpWaaC), WOD26 (CpWaaA and CjWaaC), and WOD27 (CpWaaA and AaRfaC1) and Kdo_3-Lipid A in WOD24. The summary of the activities of the Kdo transferases used are found in Table 2.2a. In all, it was seen that with the heterologous expression of Kdo transferases in the WOD02 cell line, it is possible to generate either mono-, bi-, or tri-functionalized LPS analog.

In order to further modify these initial glycosylated strains, the Kdo transferases were coexpressed with putative heptosyltransferase I enzymes from various species. The WaaC of *Helicobacter pylori* was chosen because while the Kdo transferase is bifunctional giving two Kdo moieties attached to the Lipid IV_A, the distal Kdo is removed by a Kdo hydrolase. This modification is believed to help H. pylori evade the host immune system.[12] Due to the presence of both Kdo-Lipid IV_A and Kdo_2-Lipid IV_A it was

speculated that the heptosyltransferase I of *H. pylori* may have enhanced promiscuity. The heptosyltransferase I from *E. coli* was chosen both a control due to it being the best characterized heptosyltransferase I to date as well as the only WaaC whose structure has been determined.[41] *E. coli* WaaC is also known to transfer a heptose to Kdo_2-Lipid A, Kdo_3-Lipid A,[42] or a fully deacylated Lipid A analog.[43] The third heptosyltransferase I chosen was from the *Campylobacter jejuni* whose truncated LPS has been seen to have been composed of Hep-Kdo-Lipid A[44] or Hep-Kdo_2-Lipid A[45] leading to the possibility of enhanced WaaC promiscuity. The last enzymes chosen were RfaC1 (gene: aq_1543) and RfaC2 (gene: aq_145) from the hyperthermophilic *Aquifex aeolicus* as a putative heptosyltransferases. Both genes in *A. aeolicus* are annotated as ADP-heptose: LPS heptosyltransferases. Although the structure of full LPS of *A. aeolicus* is unknown, analysis of the LPS from the related species, *A. pyrophilus*, gave evidence for the presence of two heptose molecules in the LPS core.[46] This would lead one to believe that A. aeolicus *rfaC1* and *rfaC2* encode not for two heptosyltransferase I enzymes but instead one being a heptosyltransferase I and the other a heptosyltransferase II. The Kdo transferase of *A. aeolicus* is known to transfer only a single Kdo[10], and as such it is speculated that the heptosyltransferase I of *A. aeolicus* will only transfer to a single Kdo.

Using the aforementioned Kdo transferases coexpressed with HpWaaC yielded heptose modification as follows. WOD12 (HiWaaA and HpWaaC) and WOD24 show the presence of Hep-Kdo-Lipid IV_A (exact mass 1816.98 u, calculated mass 1816.975 u). WOD20 (CtWaaA and HpWaaC) on the other hand show the presence of Hep-Kdo_2-Lipid A (M+1 exact mass 2430.40 u, calculated mass 2429.399 u).

For EcWaaC the presence of Hep-Kdo-Lipid IV_A can be seen in WOD13 (HiWaaA and EcWaaC), WOD17 (EcWaaA and EcWaaC), and WOD25 (CpWaaA and EcWaaC). WOD17 and WOD21 (CtWaaA and EcWaaC) both show the presence of Hep-Kdo_2-Lipid A. Additionally, WOD21 also shows the presence of the larger Hep-Kdo_3-Lipid A (M+1 exact mass 2650.46 u, calculated mass 2649.458 u).

CjWaaC can only generate Hep-Kdo_2-Lipid A at any sufficient quantities as in WOD22 (CtWaaA and CjWaaC).

On the contrary AaRfaC1 shows the least promiscuity generating only Hep-Kdo-Lipid IV_A in WOD15 (HiWaaA and AaRfaC1), WOD19 (EcWaaA and AaRfaC1), WOD23 (CtWaaA and AaRfaC1), and WOD27. The summary of the activities of the heptosyltransferases used in these experiments are shown in Table 2.2b. With these genes it is now possible to generate Hep-Kdo-Lipid IV_A, Hep-Kdo_2-Lipid A, and Hep-Kdo_3-Lipid A *in vivo* in WOD02.

WOD02 shows a putative 3'-*O*-Deacylase activity. LPS samples from a number of the Kdo transferase and heptosyltransferase expressing WOD02 cell lines contained peaks which correspond to Kdo_2-LA_{Tetra} (exact mass 1800.94 u, calculated mass 1800.944 u), which is the 3'-O-deacylated Kdo_2-Lipid A wherein the 3'-acyloxylacyl moiety has been removed. This peak is seen in all of the tested EcWaaA expressing cell lines (WOD05, WOD17, WOD18, and WOD19) in all CtWaaA expressing cell lines (WOD06, WOD20, WOD21, WOD22, and WOD23) as well as two of the CtWaaA expressing cell lines (WOD26 and WOD27). Additionally, a peak which corresponds to the heptosylated Kdo_2-LA_{Tetra}, Hep-Kdo_2-LA_{Tetra} (M+1 exact mass 1994.01 u, calculated mass 1993.007 u) was seen in three samples (WOD17, WOD20, and WOD21). Kdo_2-LA_{Tetra} was previously reported[47] in the mutant *E. coli* F515 which is known to be deficient in heptosyltransferase activity giving Kdo_2-Lipid A as a maximal LPS structure.[48]

Endotoxicity. In order to verify the lack of endotoxic product in the WOD02 cell line as well as to show the ability to return endotoxicity with the addition of Kdo transferases hTLR4/MD-2 activation was assayed by subjecting HEK-Blue™ hTLR4 and HEK-Blue™ Null2 cells to purified LPS samples. With HEK- Blue™ Null2 being the parent strain of HEK-Blue™ hTLR4 lacking the TLR4, MD-12, and CD14 needed to elicit the NF-κB mediated response. Activation of the TLR4 complex by LPS induces production of the reporter protein, secreted embryonic alkaline phosphatase

(SEAP). The phosphatase activity was measured using a colorimetric substrate with an absorbance reading at 655 nm. Percent activation was determined based on the highest reading of the *E. coli* K-12 LPS at 1000 ng/mL.

Activation of hTLR4 for LPS samples purified from WOD03 which lacks all glycosyltransferases for core region biosynthesis is seen in Figure 2.3, to be the lowest of all samples tested with lower than 2% activity compared to full *E. coli* K-12 LPS at the same concentration. It is also shown that the addition of Kdo transferases help return hTLR4 activation to these cells. The addition of only a single Kdo (seen in WOD04) returns low hTLR4 activation while all other cell lines containing Kdo transferases show a higher return of hTLR4 activation. A comparison of each Kdo transferase with each heptosyltransferase (see graphs in Appendix C, Figure C.1A-D) show that in general, the addition of a heptosyltransferase for EcWaaA, CtWaaA, and to a lesser extent CpWaaA seems to have little to no difference on the overall hTLR4 activation. The cell lines containing HiWaaA on the other hand seem to vary more widely based on heptosyltransferase. While this change in activation is quite clear, the underlying reason for this is unknown, and may be an artefact of change in overall lipid yield or activation of late acyltransferases. It is sufficient to say from this data though that the LPS produced inherently by the WOD background has very low hTLR4 activation

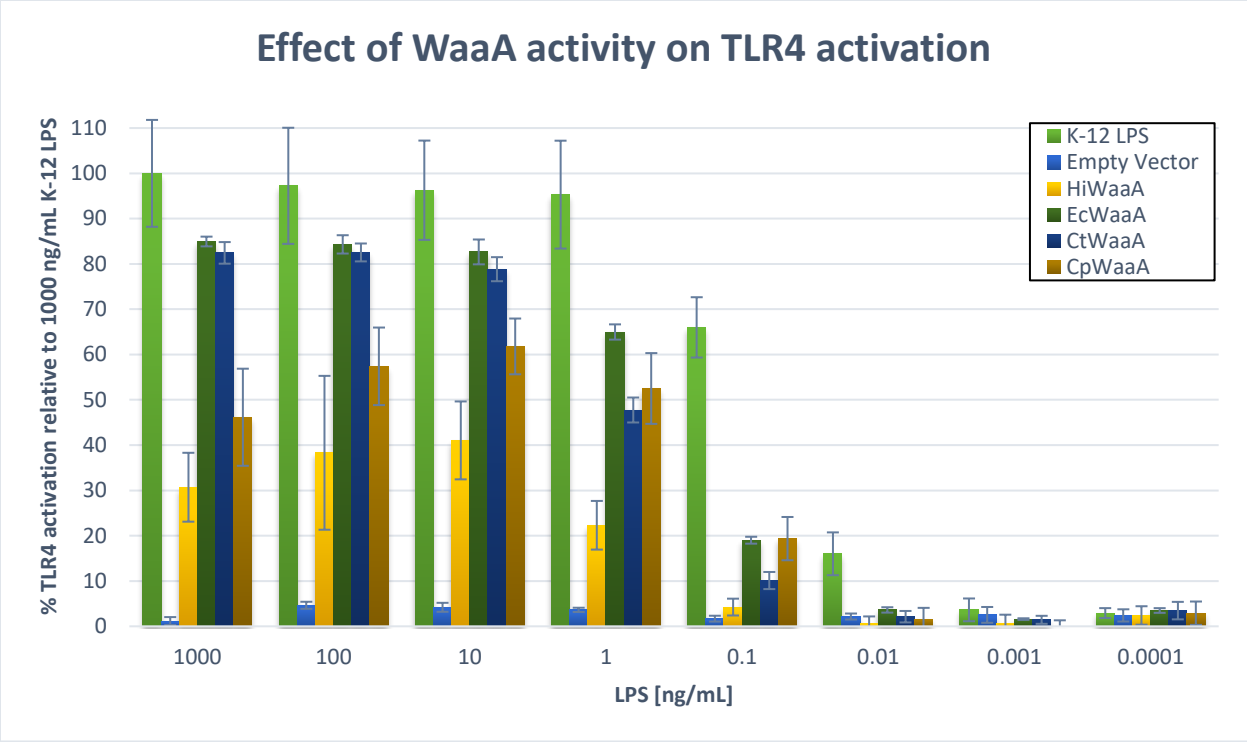


Figure 2.3. Effect of WaaA activity on hTLR4 activation. Measure of hTLR4 activation by SEAP secretion. Total activities are compared to commercial full-length *E. coli* K-12 LPS (100% activation). Samples include WOD02 with empty vector (WOD03), or expressing HiWaaA (WOD04), EcWaaA (WOD05), CtWaaA (WOD06), or CpWaaA (WOD07)

acyltransferases. It is sufficient to say from this data though that the LPS produced inherently by the WOD background has very low hTLR4 activation when compared to wild type *E. coli* LPS, and that in general, the addition of a Kdo transferase allows a return of hTLR4 activation of the purified LPS from this cell line.

Discussion

The ability to quickly and easily generate a series of glycoforms of LPS has to date been a challenge. Many tools exist to study Kdo transferases[10, 49] but to study biosynthetic genes beyond Kdo is much less straight forward. Due to the importance of LPS both for the bacteria containing it as well as for its endotoxic properties, a cell line lacking all core oligosaccharide biosynthetic genes was desired. By removing all *waa* genes within the same gene cluster in a unique strain previously produced in our laboratory wherein Lipid A was no longer the minimal LPS structure required for *E. coli* K-12 survival[29, 33] the new WOD02 cell line was generated. With the ability of WOD02 to continue to produce Kdo, it was possible to reintroduce Kdo transferase genes from other species into this cell line in order to study their functionalities. Indeed, when the four Kdo transferases used in this study were introduced into WOD02, they were able to return Kdo glycosylation activity to the cells. Each Kdo transferase was able to modify the Lipid IV_A to produce the expected glycoforms with the exception of the Kdo transferase from *C. psittaci* which in this system was unable to generate the tetraglycosylated Lipid IV_A glycoform. The lack of full glycosylation by this particular enzyme is potentially due to lack of WaaZ activity in this cell line as WaaZ has been shown to be required for addition of more than two Kdo moieties in wild-type *E. coli*.[40, 50] Although WOD02 was unable to tetra-

glycosylate Lipid IV_A, the cell line was able to generate mono-, bi-, or tri-Kdo glycosylated analogs. Through hTLR4 activation, it is shown that WOD02 would have very low endotoxic activity alone, but the addition of a Kdo transferase greatly increases endotoxicity due to the activities of the acyltransferases, LpxL and LpxM, which are known to typically require Kdo glycosylation of Lipid IV_A.

As the study of heptosyltransferases has to date typically been limited to the native WaaC[42] or studying the enzyme *in vitro*[43], the generation of the WOD02 cell line is a useful tool to study different combinations of Kdo transferases with heptosyltransferases in a Gram-negative *in vivo* environment. With a pETduet system, combinations of both very important early core biosynthetic genes were generated and analyzed. With this system, it has been shown that typically the substrate specificities of these enzymes are quite broad with all tested heptosyltransferases modifying a single Kdo glycosylated Lipid IV_A species. Three of the tested heptosyltransferases, HpWaaC, EcWaaC, and CjWaaC further would glycosylate Kdo_2-Lipid A, and EcWaaC continuing on to glycosylate Kdo_3-Lipid A. For HpWaaC and CjWaaC, these studies are the first reports to our knowledge of their substrate specificities. *H. pylori* LPS is particularly of interest due to the presence of the Kdo hydrolase not found in most Gram-negative bacteria which allows removal of a Kdo moiety in biglycosylated Kdo_2-Lipid A. This study shows that HpWaaC is able to heptosylated the

growing LPS core potentially before or after the cleavage of the distal Kdo by the Kdo hydrolase. The ability to do this could be a remnant of evolution or could mean that H*. pylori* may at times maintain bi- Kdo glycosylated LPS.

AaRfaC1 is seen for the first time to be a heptosyltransferase I is unique among those tested in having specificity only for Kdo-Lipid IV_A as a substrate. This would leave one to believe that AaRfaC2 would most likely then have heptosyltransferase II activity, but this is yet to be seen. As the full structure of the LPS of *A. aeolicus* is not currently know, this gives us novel insight into the potential core oligosaccharide of this unique bacterium.

The presence of AaRfaC1 in samples containing Kdo_2-Lipid A or Kdo_3-Lipid A were not further glycosylated. Due to the activity of AaRfaC1, we gain an indirect view of Kdo transferase activity as multi-functional Kdo transferases must go through either an iterative or processive mechanism. In an iterative mechanism, the WaaA would bind to the lipid acceptor molecule, glycosylate, and then release only to rebind later for further glycosylation. In a processive mechanism, the WaaA would remain bound to the lipid acceptor until the correct number of Kdo glycosylation events occur where upon the glycosylated lipid is then released. As the WaaA from *A. aeolicus* is monofunctional, it is quite likely that AaRfaC1 would only be able to detect a single Kdo glycosylated Lipid IV_A. In order for heptosylation to occur, free Kdo-Lipid IV_A must be available to AaRfaC1. For this to occur in

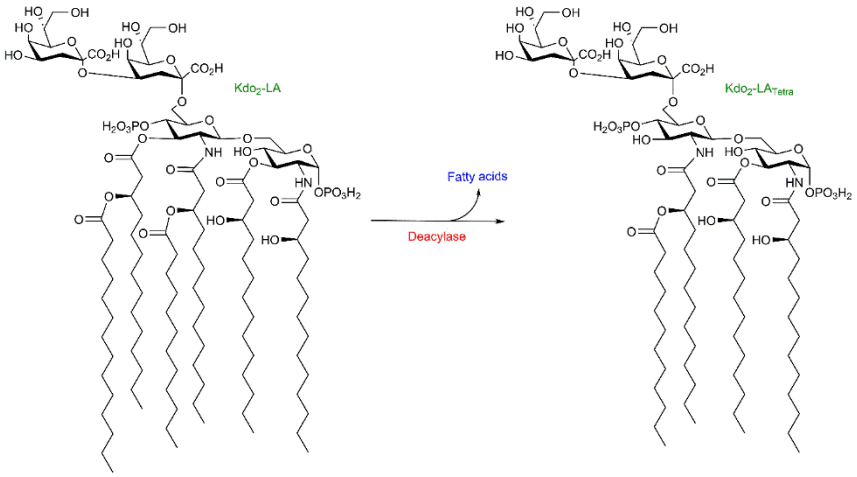


Figure 2.4. LpxR-like deacetylation. Deacylase activity such as that caused by *S. typhimurium* LpxR leads to loss of the 3'-acyloxoacyl group of Lipid A.

the WOD02 cell line, the Kdo transferases must be undergoing an iterative glycosylation mechanism whereby each glycosylation event is followed by a release of the acceptor allowing for the heptosyltransferase to bind and transfer the heptose.

An additional feature seen in the LPS analysis of WOD02 expressing Kdo and heptosyltransferases is the presence of a 3'-O-deacylated Kdo_2-LA_{tetra}. LpxR from *Salmonella typhimurium* has been found to have 3'-O-deacylation activity although no homologues have been found in *E. coli* K-12. Nevertheless, heterologous expression in *E. coli* elicits 3'-O-deacylation (Figure 2.4).[51] The activity of LpxR activity is known to cause reduced TLR4 activation, and is likely a method by which cells can evade a host immune system.[52] This activity has been seen before[47] in another *E. coli* LPS mutant. Additionally, this activity is only seen in Kdo_2-LA_{tetra} or Hep-Kdo_2-LA_{tetra} both

of which are native *E. coli* LPS core structures. It is quite possible that this putative deacylase activity is a response to stress on the outer membrane caused by truncated LPS. Further studies are needed to identify and genes involved and to study how this modification is activated in *E. coli*.

In summary, WOD02 represents a new tool for the study of LPS core biosynthesis. This cell line allows for the characterization of biosynthetic genes, generation of various LPS core oligosaccharides, and potentially serves as a tool to study LPS peripheral modifications. This cell line could be used further to generate large oligosaccharides by continuation of the methods herein of heterologous gene expression. Additionally, the ability to generate these novel glycoforms could help with the generation of complex glycolipids as antigens for future vaccine research and aid in our efforts to eliminate bacterial infections.

Chapter contributions/acknowledgements

We thank Brigitte Kunz of Forschungzentrum Borstel for her assistance in performing mass spectrometry experiments and David Cech for useful discussions. All other work and writing was done by ACP.

References

[1] Nikaido, H. (1996) Outer Membrane, In *Escherichia coli and Salmonella typhimurium : cellular and molecular biology* (Neidhardt, F. C., Ed.), pp 29-47, American Society for Microbiology, Washington, D.C.

[2] Wiese, A., Brandenburg, K., Ulmer, A. J., Seydel, U., and Muller-Loennies, S. (1999) The dual role of lipopolysaccharide as effector and target molecule, *Biol Chem 380*, 767-784.

[3] Heine, H., Rietschel, E. T., and Ulmer, A. J. (2001) The biology of endotoxin, *Mol Biotechnol 19*, 279-296.

[4] Raetz, C. R. H., and Whitfield, C. (2002) Lipopolysaccharide Endotoxins, *Annual Review of Biochemistry 71*, 635-700.

[5] Hobman, J. L., Penn, C. W., and Pallen, M. J. (2007) Laboratory strains of Escherichia coli: model citizens of deceitful delinquents growing old disgracefully?, *Molecular Microbiology 64*, 881-885.

[6] Liu, D., and Reeves, P. R. (1994) Escherichia coli K12 regains its O antigen, *Microbiology 140*, 49-57.

[7] Lodowska, J., Wolny, D., and Weglarz, L. (2013) The sugar 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) as a characteristic component of bacterial endotoxin - a review of its biosynthesis, function, and placement in the lipopolysaccharide core, *Canadian journal of microbiology 59*, 645-655.

[8] Gronow, S., and Brade, H. (2001) Lipopolysaccharide biosynthesis: which steps do bacteria need to survive?, *J Endotoxin Res 7*, 3-23.

[9] Brabetz, W., Mueller-Loennies, S., and Brade, H. (2000) 3-Deoxy-D-manno-oct-2-ulosonic Acid (Kdo) transferase (WaaA) and Kdo Kinase (KdkA) of Haemophilus influenzae are both required to complements a

waaA knockout mutation of Escherichia coli, *J. Biol. Chem. 275*, 34954-34962.

[10] Mamat, U., Schmidt, H., Munoz, E., Lindner, B., Fukase, K., Hanuszkiewicz, A., Wu, J., Meredith, T., Woodard, R., Hilgenfeld, R., Mesters, J., and Holst, O. (2009) WaaA of the hyperthermophilic bacterium Aquifex aeolicus is a monofunctional 3-deoxy-D-manno-oct-2-ulosonic acid transferase involved in lipopolysaccharide biosynthesis, *The Journal of biological chemistry 284*, 22248-22262.

[11] Belunis, C. J., and Raetz, C. R. (1992) Biosynthesis of endotoxins. Purification and catalytic properties of 3-deoxy-D-manno-octulosonic acid transferase from Escherichia coli, *J Biol Chem 267*, 9988-9997.

[12] Stead, C., Tran, A., Ferguson, D. J., McGrath, S., Cotter, R., and Trent, S. (2005) A Novel 3-Deoxy-D-manno-octulosonic Acid (Kdo) Hydrolase that removes the outer Kdo sugar of Helicobacter pylori Lipopolysaccharide, *J. Bacteriol. 187*, 3374-3383.

[13] Noah, C., Brabetz, W., Gronow, S., and Brade, H. (2000) Cloning, sequencing, and functional analysis of three glycosyltransferases involved in the biosynthesis of the inner core region of Klebsiella pneumoniae lipopolysaccharide, *Journal of Endotoxin Research 7*, 25-33.

[14] Heine, H., Mueller-Loennies, S., Brade, L., Lindner, B., and Brade, H. (2003) Endotoxic activity and chemical structure of lipopolysaccharides from Chlamydia trachomatis serotypes E and L2 and Chlamydophila psittaci 6BC, *European Journal of Biochemistry 270*, 440-450.

[15] Rund, S., Lindner, B., Brade, H., and Holst, O. (2000) Structural analysis of the lipopolysaccharide from Chlamydophila psittaci strain 6BC, *European Journal of Biochemistry 267*, 5717-5726.

[16] Chung, H. S., and Raetz, C. R. H. (2010) Interchangeable domains in the Kdo transferases of Escherichia coli and Haemophilus influenzae, *Biochemistry 49*, 4126-4137.

[17] Schmidt, H., Hansen, G., Singh, S., Hanuszkiewicz, A., Lindner, B., Fukase, K., Woodard, R. W., Holst, O., Hilgenfeld, R., Mamat, U., and Mesters, J. R. (2012) Structural and mechanistic analysis of the membrane-embedded glycosyltransferase WaaA required for lipopolysaccharide synthesis, *Proceedings of the National Academy of Sciences of the United States of America 109*, 6253-6258.

[18] Raetz, C., Reynolds, C., Trent, M., and Bishop, R. (2007) Lipid A modification systems in gram-negative bacteria, *Annual review of biochemistry 76*, 295-329.

[19] Rund, S., Lindner, B., Brade, H., and Holst, O. (1999) Structural analysis of the lipopolysaccharide from Chlamydia trachomatis Serotype L2, *J. Biol. Chem. 274*, 16819-16824.

[20] Kadrmas, J. L., and Raetz, C. R. (1998) Enzymatic synthesis of lipopolysaccharide in Escherichia coli. Purification and properties of heptosyltransferase i, *J Biol Chem 273*, 2799-2807.

[21] Rietschel, E. T., Kirikae, T., Schade, F. U., Mamat, U., Schmidt, G., Loppnow, H., Ulmer, A. J., Zahringer, U., Seydel, U., Di Padova, F., Schreier, M., and Brade, H. (1994) Bacterial endotoxin: molecular relationships of structure to activity and function, *Faseb J 8*, 217-225.

[22] Coleman, W. G., and Leive, L. (1979) Two mutations which affect the barrier function of the Escherichia coli K-12 outer membrane, *J. Bacteriol. 139*, 899-910.

[23] Bodewits, K., Raetz, C. R. H., Govan, J. R., and Campopiano, D. J. (2010) Antimicrobial Activity of CHIR-090, an inhibitor of Lipopolysaccharide biosynthesis, against the Burkholderia cepacia Complex, *Antimicrob Agents Chemother 54*, 3531-3533.

[24] Moreau, F., Desroy, N., Genevard, J. M., Vongsouthi, V., Le Fralliec, G., Oliveira, C., Floquet, S., Denis, A., Escaich, S., Wolf, K., Busemann, M., and Aschenbrenner, A. (2008) Discovery of new Gram-negative antivirulence drugs: Structure and properties of novel E. coli WaaC inhibitors, *Bioorganic & Medicinal Chemistry Letters 18*, 4022-4026.

[25] Durka, M., Tikad, A., Perion, R., Bosco, M., Andaloussi, M., Floquet, S., Malacain, E., Moreau, F., Oxoby, M., Gerusz, V., and Vincent, S. P. (2011) Systematic synthesis of inhibitors of the two first enzymes of the bacterial heptose biosynthetic pathway: towards antivirulence molecules targeting lipopolysaccharide biosynthesis, *Chemistry 17*, 11305-11313.

[26] Kong, L., Vijayakrishnan, B., Kowarik, M., Park, J., Zakharova, A. N., Neiwert, L., Faridmoayer, A., and Davis, B. G. (2016) An antibacterial vaccination strategy based on a glycoconjugate containing the core lipopolysaccharide tetrasaccharide Hep_2Kdo_2, *Nature chemistry 8*, 242-249.

[27] Nagy, G., and Pal, T. (2008) Lipopolysaccharide: a tool and target in enterobacterial vaccine development, *Biol. Chem. 389*, 513-520.

[28] Zariri, A., and van der Ley, P. (2015) Biosynthetically engineered lipopolysaccharide as vaccine adjuvant, *Expert Review of Vaccines 14*, 861-876.

[29] Mamat, U., Meredith, T. C., Aggarwal, P., Kuhl, A., Kirchhoff, P., Lindner, B., Hanuszkiewicz, A., Sun, J., Holst, O., and Woodard, R. W. (2008) Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-D-manno-oct-2-ulosonic acid-depleted Escherichia coli, *Molecular Microbiology 67*, 633-648.

[30] Nano, F., and Caldwell, H. (1985) Expression of the chlamydial genus-specific lipopolysaccharide epitope in Escherichia coli, *Science 228*, 742-744.

[31] Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products, *PNAS 97*, 6640-6645.

[32] Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[33] Meredith, T. C., Aggarwal, P., Mamal, U., Lindner, B., and Woodard, R. W. (2006) Redefining the requisite lipopolysaccharide structure in Escherichia coli, *ACS chemical biology 1*, 33-42.

[34] Galanos, C., Luderitz, O., and Westphal, O. (1969) A new method for the extraction of R lipopolysaccharides, *Eur J Biochem 9*, 245-249.

[35] Mamat, U., Meredith, T. C., Aggarwal, P., Kühl, A., Kirchhoff, P., Lindner, B., Hanuszkiewicz, A., Sun, J., Holst, O., and Woodard, R. W. (2008) Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-d-manno-oct-2-ulosonic acid-depleted Escherichia coli, *Molecular Microbiology 67*, 633-648.

[36] Mamat, U., Wilke, K., Bramhill, D., Schromm, A. B., Lindner, B., Kohl, T. A., Corchero, J. L., Villaverde, A., Schaffer, L., Head, S. R., Souvignier, C., Meredith, T. C., and Woodard, R. W. (2015) Detoxifying Escherichia coli for endotoxin-free production of recombinant proteins, *Microb Cell Fact 14*.

[37] Roncero, C., and Casadaban, M. J. (1992) Genetic analysis of the genes involved in synthesis of the lipopolysaccharide core in Escherichia coli K-12: Three operons in the rfa locus, *J. Bacteriol. 174*, 3250-3260.

[38] Meredith, T. C., Mamat, U., Kaczynski, Z., Lindner, B., Holst, O., and Woodard, R. W. (2007) Modification of lipopolysaccharide with colanic acid (M-antigen) repeats in Escherichia coli, *J. Biol. Chem. 282*, 7790-7798.

[39] Brabetz, W., Lindner, B., and Brade, H. (2000) Comparative analyses of secondary gene products of 3-deoxy-D-manno-oct-2-ulosonic acid transferases from Chlamydiaceae in Escherichia coli K-12, *Eur J Biochem 267*, 5458-5465.

[40] Klein, G., Lindner, B., Brade, H., and Raina, S. (2011) Molecular basis of lipopolysaccharide heterogeneity in *Escherichia coli*. Envelope stress-responsive regulators control the incorporation of glycoforms with a third 3-deoxy-α-D-manno-oct-2-ulosonic acind and rhamnose, *J. Biol. Chem. 286*, 42787-42807.

[41] Grizot, S., Salem, M., Vongsouthi, V., Durand, L., Moreau, F., Dohi, H., Vincent, S., Escaich, S., and Ducruix, A. (2006) Structure of the Escherichia coli heptosyltransferase WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose, *Journal of molecular biology 363*, 383-394.

[42] Gronow, S., Lindner, B., Brade, H., and Muller-Loennies, S. (2009) Kdo-(2 --> 8)-Kdo-(2 --> 4)-Kdo but not Kdo-(2 --> 4)-Kdo-(2 --> 4)-Kdo is an acceptor for transfer of L-glycero-alpha-D-manno-heptose by Escherichia coli heptosyltransferase I (WaaC), *Innate Immun 15*, 13-23.

[43] Czyzyk, D., Liu, C., and Taylor, E. (2011) Lipopolysaccharide biosynthesis without the lipids: recognition promiscuity of Escherichia coli heptosyltransferase I, *Biochemistry 50*, 10570-10572.

[44] Naito, M., Frirdich, E., Fields, J. A., Pryjma, M., Li, J., Cameron, A., Gilbert, M., Thompson, S. A., and Gaynor, E. C. (2010) Effects of sequential *Campylobacter jejuni* 81-176 lipooligosaccharide core truncations on biofilm formation, stress survival, and pathogenesis, *J. Bacteriol. 192*, 2182-2192.

[45] Cullen, T. W., O'Brien, J. P., Hendrixson, D. R., Giles, D. K., Hobb, R. I., Thompson, S. A., Brodbelt, J. S., and Trent, M. S. (2013) EptC of

Campylobacter jejuni mediates phenotypes in volved in host interaction and virulence, *Infect Immun 81*, 430-440.

[46] Plotz, B. M., Lindner, B., Stetter, K. O., and Holst, O. (2000) Characterization of a Novel Lipid A Containing D-Galaturonic acid that replaces phophate residues, *J. Biol. Chem. 275*, 11222-11228.

[47] Zahringer, U., Salvetzki, R., Wagner, F., Lindner, B., and Ulmer, A. J. (2001) Structural and biological characterisation of a novel tetra-acyl lipid A from Escherichia coli F515 lipopolysaccharide acting as endotoxin antagonist in human monocytes, *Journal of Endotoxin Research 7*, 133-146.

[48] Schmidt, G., Jann, B., and Jann, K. (1970) Studies on R Mutants with an incomplete core, derived from E. coli O8:K27, *Eur J Biochem 16*, 382-392.

[49] Reynolds, M. C., and Raetz, C. R. H. (2009) Replacement of lipopolysaccharide with free lipid A molecules in Escherichia coli mutants lacking all core sugars, *Biochemistry 48*, 9627-9640.

[50] Frirdich, E., Lindner, B., Holst, O., and Whitfield, C. (2003) Overexpression of the waaZ Gene Leads to Modification of the Structure of the Inner Core Region of Escherichia coli Lipopolysaccharide, Truncation of the Outer Core, and Reduction of the Amount of O Polysaccharide on the Cell Surface, *J. Bacteriol. 185*, 1659-1671.

[51] Reynolds, M. C., Ribeiro, A. A., McGrath, S. C., Cotter, R. J., Raetz, C. R. H., and Trent, S. M. (2006) An Outer Membrane Enzyme Encoded by Salmonella typhimurium lpxR That Removes the 3′-Acyloxyacyl Moiety of Lipid A, *J. Biol. Chem. 281*, 21974-21987.

[52] Kawasaki, K., Teramoto, M., Tatsui, R., and Amamoto, S. (2012) Lipid A 3′-O-deacylation by Salmonella outer membrane enzyme LpxR

modulates the ability of lipid A to stimulate Toll-like receptor 4, *Biochemical and Biophysical Research Communications 428*, 343-347.

CHAPTER III

Insights into the selective incorporation of 8-amino-3,8-dideoxy-D-*manno*-oct-2-ulosonic acid in *Shewanella oneidensis*

Summary

The LPS of *Shewanella* species are unique among Gram-negative bacteria in that they contain the 8-carbon amino sugar 8-amino-3,8-dideoxy-D-*manno*-octulosonic acid (Kd8N) in their core lipopolysaccharide. This sugar is biosynthesized from the 8-hydroxy sugar (Kdo) commonly found in the LPS core of most other Gram-negative bacteria. The method by which *S. oneidensis* selectively incorporates Kd8N into its core oligosaccharide over Kdo is not known. Herein we test the substrate specificities of both the activating enzyme (KdsB or Kdo cytidylyltranferase) and the glycosyltransferase (WaaA or Kdo transferase). KdsB is shown *in vitro* to prefer its native substrate, Kd8N, while WaaA *in vivo* does not have a substrate preference. A mutant strain of *S. oneidensis* lacking the ability to generate Kd8N can survive without Kd8N and generates fully elaborated core oligosaccharide.

These observations suggest that KdnA and KdnB expression may ultimately be the deciding factors for Kd8N selectivity.

Introduction

Shewanella oneidensis and related species in the genus are common marine Gram-negative bacteria that are of interest due to their unique bioremediation properties including the ability to reduce heavy metals[1] as well as the ability to reduce azo dyes.[2] Bacteria in the genus are also known to be opportunistic pathogens[3] and are known to be involved in the spoilage of fish.[4] *Shewanella spp*. represent an interesting genus of bacteria with very unique biological properties and an important set of metabolic functions.[5]

Shewanella oneidensis as with other Gram-negative bacteria contain an amphipathic glycolipid, lipopolysaccharide or LPS, found in the outer-leaflet of the outer membrane. These glycolipids are composed of the Lipid A anchor, the bridging oligosaccharide core, and the repeating O-antigen.[6] The *Shewanella* genus is unique in that it is the only known genus to incorporate 8-amino-3,8-dideoxy-D-*manno*-octulosonic acid (Kd8N) rather than 3-deoxy-D-*manno*-octulosonic acid (Kdo) into the oligosaccharide core of lipopolysaccharide.[7] To date *Shewanella putrefaciens* is the only naturally

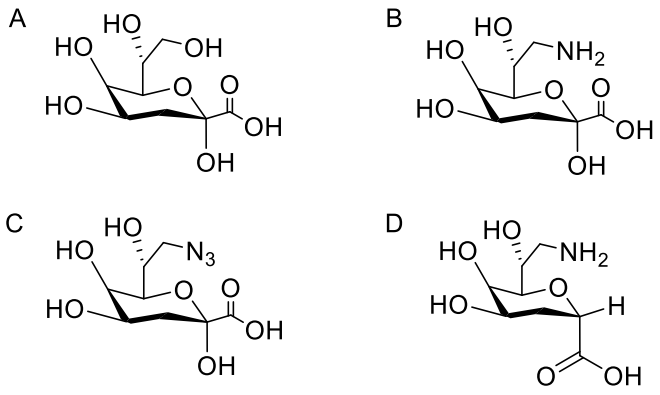


Figure 3.1. Structures of Kdo analogs. Structures of Kdo analogs. (A) Kdo, (B) Kd8N, (C) Kd8Z, (D) the KdsB inhibitor 2,3,8-trideoxy-8-amino-β-D-*manno*-octulosonic acid

occurring date *Shewanella putrefaciens* is the only naturally occurring species in this genus which has been shown to incorporate Kdo into the LPS.[8] Kdo is of interest due to being essential in most known Gram-negative strains with the minimal LPS structure, Kdo_2-Lipid A, being necessary for most wild-type *E. coli* strains. The presence of the 8-amine group is striking due to the inhibitory properties seen in the Kd8N analog 2,3,8-trideoxy-8-amino-β-D-*manno*-octulosonic acid, a KdsB inhibitor and one of the few known Kdo biosynthetic pathway inhibitors.[9-11] Kdo and some pertinent analogs are shown in Figure 3.1. Understanding the selectivity of Kd8N over Kdo could yield interesting insights for the development of novel Kdo analog inhibitors for future antibacterial development.

Recent work by Gattis et al. showed the biosynthetic genes necessary for generation of Kd8N in *S. oneidensis*, namely the alcohol dehydrogenase (*kdnB*) and the aminotransferase (*kdnA*).[12]

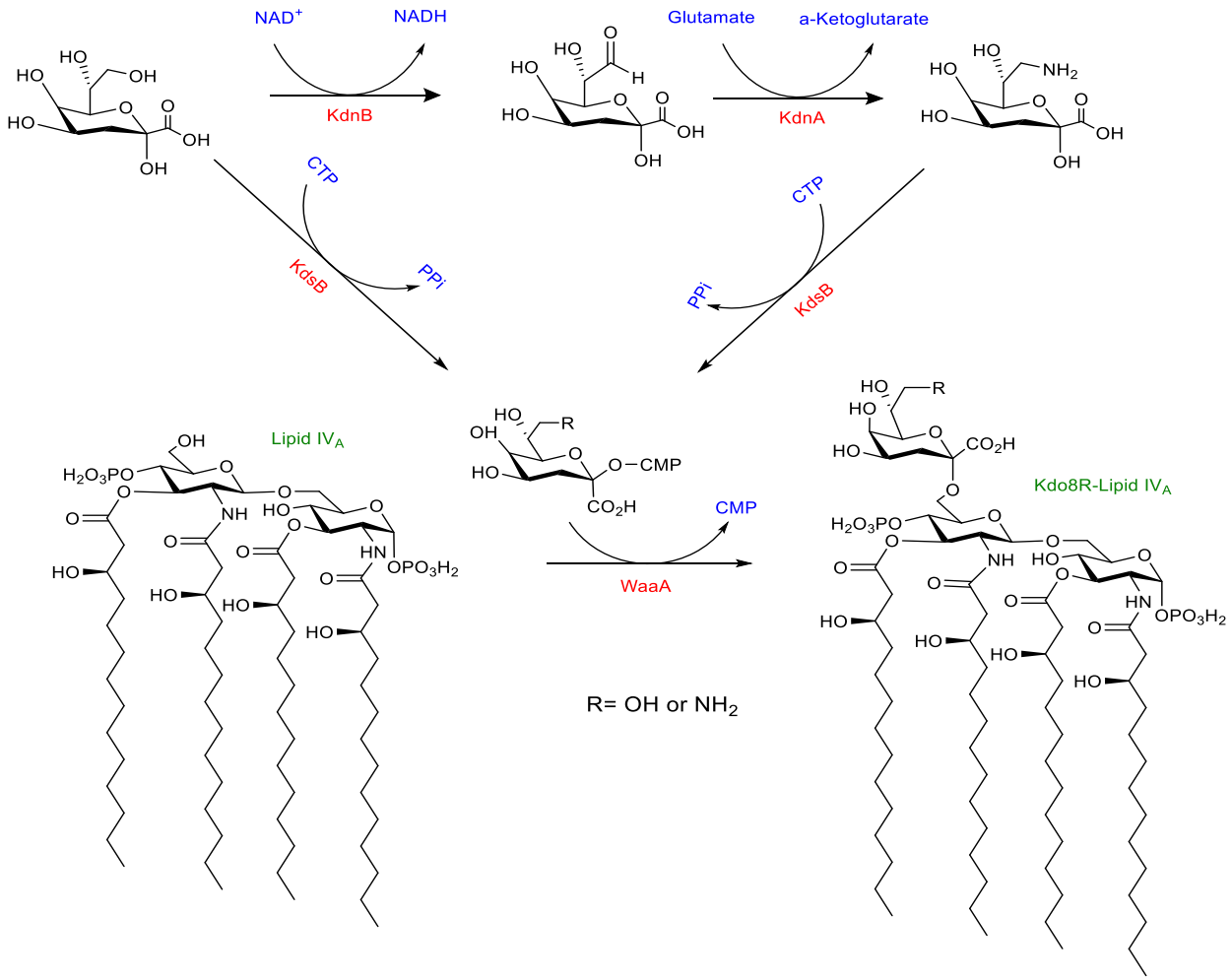


Figure 3.2. Kd8N biosynthesis and potential selectivity. Biosynthetic scheme for generation of Kd8N by the alcohol dehydrogenase, KdnB and the aminotransferase, KdnA. In *S. oneidensis* due to the potential presence of both Kdo and Kd8N, either Kdo or Kd8N should be able to be activated by the cytidyltransferase, KdsB generating either CMP-Kd8N or CMP-Kdo. These activated sugar nucleotides can then be transferred to Lipid IV_A (fatty acid content of *E. coli* Lipid IV_A shown for simplicity) generating either Kd8N-Lipid IV_A or Kdo-Lipid IV_A, respectively.

The crystal structures of the corresponding proteins were recently determined.[13] When a *S. oneidensis* Δ*kdnA* Δ*kdnB* strain was constructed it remained viable but showed a slight increase in susceptibility to polymyxin B

and bile salts[12] suggesting that Kd8N may not be essential for survivability. The function of Kd8N in the LPS of *Shewanella* is not currently known, but it is suspected to play a role in membrane integrity and resistance to polycationic peptides similar to that found by the addition other cationic or zwitterionic LPS modifications such as 4-amino-4-deoxy-L-arabinose[14] or phosphoethanolamine.[15, 16]

When *S. oneidensis* was subjected to click-mediated fluorescent labeling experiments utilizing the Kdo analog 8-azido-3,8-dideoxy-D-*manno*-octulosonic acid (Kd8Z), it was the only Gram-negative strain tested which didn't show any labeling of the outer membrane.[17] The authors speculated that the lack of fluorescence was due to the use of Kd8N rather than Kdo in the LPS of *S. oneidensis*. Knowing now that Kd8N is biosynthesized from Kdo, it would seem that *S. oneidensis* must have some mechanism by which to selectively incorporate Kd8N rather than Kdo, or even Kd8Z, into the core of the growing LPS oligosaccharide. The selectively in *S. oneidensis* could come from complete turnover of Kdo into Kd8N by KdnB and KdnA, substrate specificity of the activating enzyme (Kdo cytidylyltranferase or KdsB), or substrate specificity of the glycosyltransferase which modifies the Lipid IV_A precursor (Kdo transferase or WaaA). The pathway of Kd8N biosynthesis and routes of selectivity are show in Figure 3.2. The KdsB of *S. oneidensis* MR-1 (gene: SO_2478) shares a sequence identity of 45% with the KdsB of *E. coli* (see alignment, Figure 3.3). While the WaaA of *S.*

oneidensis MR-1 (gene: SO_4676) shares a sequence identity of 47% with the WaaA of *E. coli* (see alignment, Figure 3.4). Neither the KdsB nor the WaaA show any obvious variation that would cause a selective preference for one sugar over another. In order to elucidate the substrate selectivity of KdsB and WaaA from *S. oneidensis in vivo* and *in vitro* studies were undertaken.

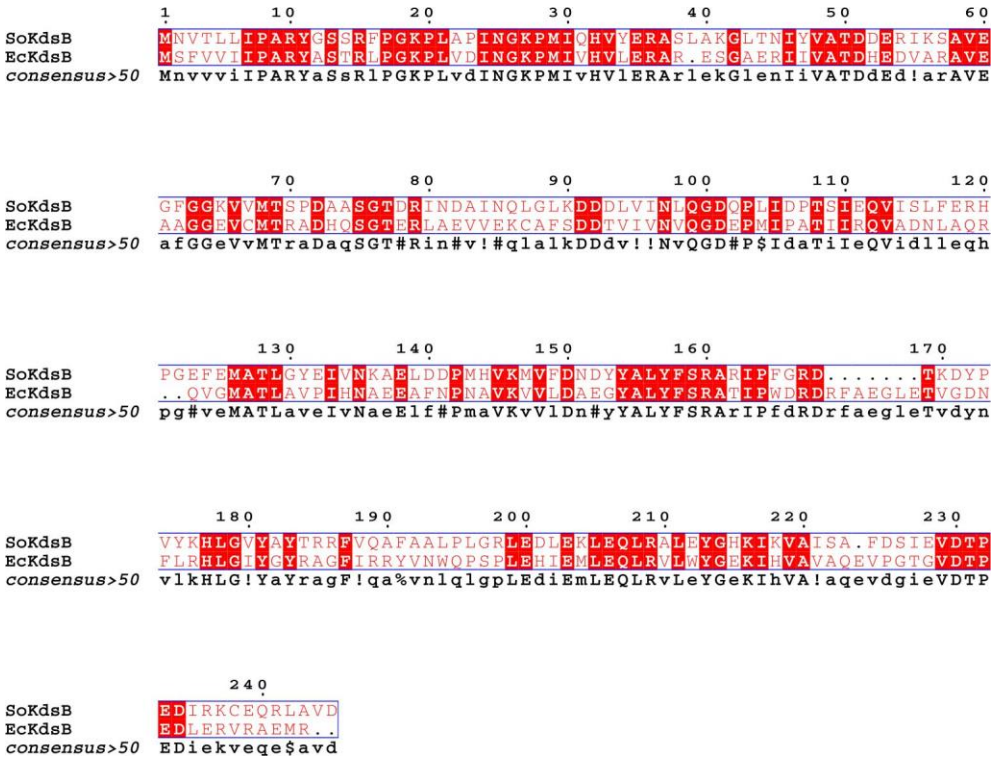


Figure 3.3. **Alignment of KdsB from *S. oneidensis* and *E. coli*.**
Alignment of KdsB from *S. oneidensis* and *E. coli* with identities shown in red boxes. The enzymes share 45% sequence identity.

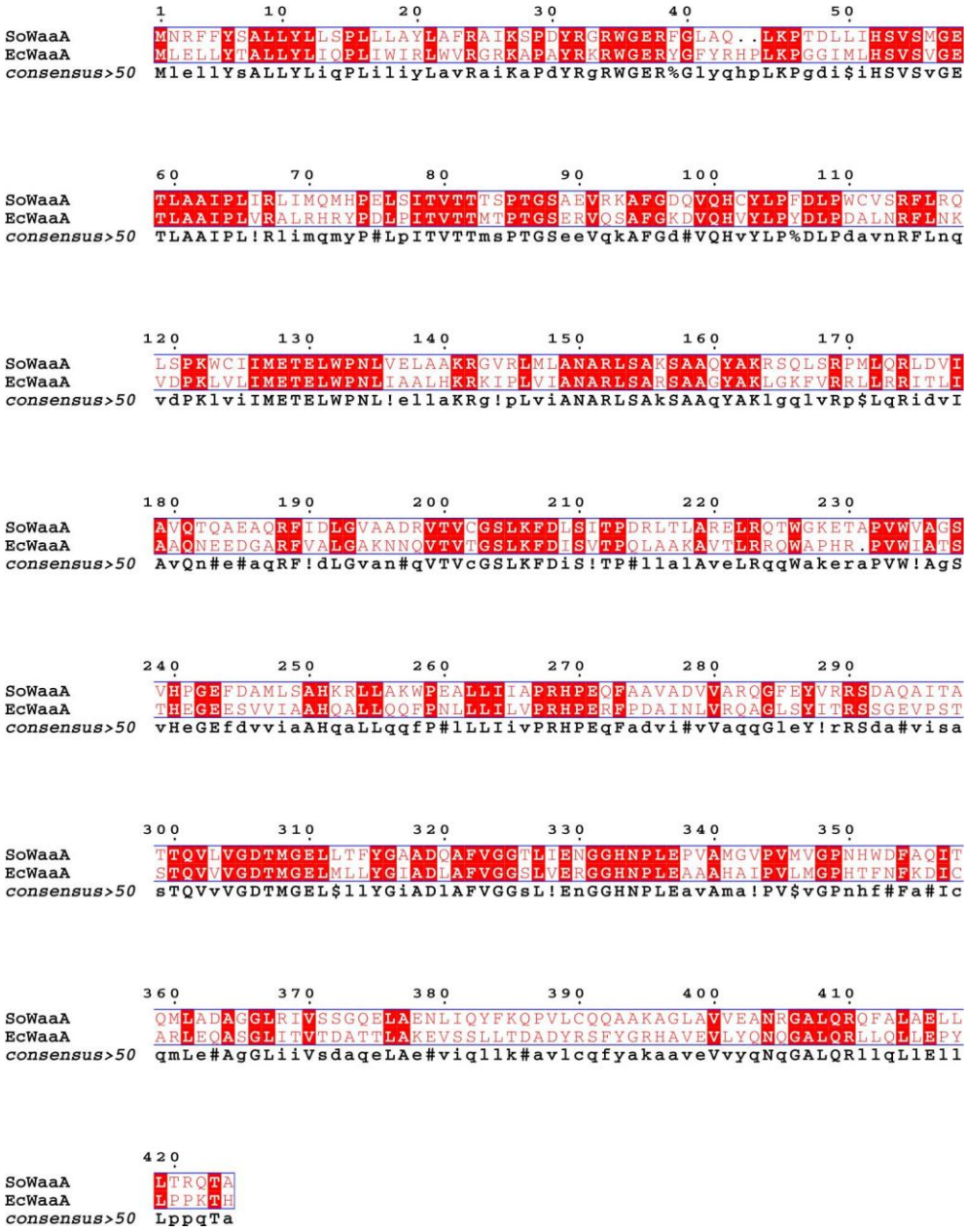


Figure 3.4. **Alignment of KdsB from *S. oneidensis* and *E. coli*.** Alignment of WaaA from *S. oneidensis* and *E. coli* with identities shown in red boxes. The enzymes share 47% identitiy.

Experimental Procedures

Bacterial Strains, Plasmids, and Growth Conditions. All strains and plasmids used in the present study are described in Table 3.1. Standard cultures of *E. coli* were routinely grown aerobically with shaking (250 rpm)

Table 3.1. List of strains, plasmids, and primers used in Chapter III.

Strain, plasmid, or primer	Description	Source
E. coli TOP10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$	Invitrogen
E. coli BL21(DE3)	*E. coli* str. B F$^-$ *ompT gal dcm lon hsdS$_B$*($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	*18*
S. oneidensis	*Shewanella oneidensis* Venkateswaran et al. ATCC® 700550™	ATCC
SoΔAB	*S. oneidensis ΔkdnA-kdnB*	*12*
KPM56	*E. coli* K-12 MG1655 F$^-$ *rph*$^+$ *fnr*$^+$ Δ*gutQ* Δ*kdsD yhjD400* kdsD$^+$ Δ*waaC* Δ*waaA*	*19*
pET16b	pET expression vector carrying an N-terminal His•Tag® sequence followed by a Factor Xa site	Novagen
pET19b	pET expression vector carrying an N-terminal His•Tag® sequence followed by an enterokinase cleavage site	Novagen
pET21b	pET expression vector carrying an N-terminal T7•Tag® sequence plus an optional C-terminal His•Tag® sequence, Amp^R	Novagen
pET16b-SoWaaA	pET16b containing SoWaaA inserted in the NdeI and BamHI	This study
pET19b-SoKdsB	pET19b containing SoKdsB inserted in the NdeI and BamHI sites	This study
pET19b-EcKdsB	pET19b containing EcKdsB inserted in the NdeI and BamHI sites	This study
pCR2.1-TOPO	Parental bacterial expression vector for TOPO® TA cloning of PCR products	ThermoFisher Scientific
pCR2.1-*kdsD*	pCR2.1-TOPO plasmid containing *S. oneidensis kdsD*	This study
pCR2.1-*kdsA*	pCR2.1-TOPO plasmid containing *S. oneidensis kdsA*	This study
pCR2.1-*kdsC*	pCR2.1-TOPO plasmid containing *S. oneidensis kdsC*	This study
pCR2.1-*kdnB*	pCR2.1-TOPO plasmid containing *S. oneidensis kdnB*	This study
pCR2.1-*kdnA*	pCR2.1-TOPO plasmid containing *S. oneidensis kdnA*	This study
pCR2.1-*kdnA*-ndeIX	pCR2.1-TOPO plasmid containing *S. oneidensis kdnA with internal NdeI sites removed*	This study

pK8N-1	pET21b with AvrII inserted into the multiple cloning site	This study
pK8N-2	pK8N-2 containing *kdsD* inserted in NdeI and BamHI sites	This study
pK8N-3	pK8N-2 containing *kdsA* inserted in NdeI and BamHI sites	This study
pK8N-4	pK8N-2 containing *kdsC* inserted in NdeI and BamHI sites	This study
pK8N-5	pK8N-2 containing *kdnA* with internal NdeI sites removed inserted in NdeI and BamHI sites	This study
pK8N-6	pK8N-2 containing *kdnB* inserted in NdeI and BamHI sites	This study
PK8N-7	pK8N-2 containing kdsA fragment from pK8N-3 inserted in the XbaI and HindIII sites	This study
pK8N-8	pK8N-7 containing kdsC fragment from pK8N-4 inserted in the XbaI and HindIII sites	This study
pK8N-9	pK8N-8 containing kdnB fragment from pK8N-5 inserted in the XbaI and HindIII sites	This study
pK8N-10	pK8N-8 containing kdnA fragment from pK8N-5 inserted in the XbaI and AvrII sites	This study
pUC57 SoK8N	pUC57-SoK8N containing *kdsD, kdsA, kdsC, kdnA, and kdnB* codon optimized for *E. coli*, Kan^R	Genscript
EcKdsBfor	GAATTCTAGAATTCATATGAGTTTTGTGGTCATTATTCC	This study
EcKdsBrev	GAATTCGGATCCAAGCTGCAGTTAGCGCATTTCAGCGC	This study
SoKdsBfor	GATTCTAGAATTCATATGAATGTCACTCTGTTAATCCCGGCGCGTTAC	This study
SoKdsBrev	GAATTCGGATCCAAGCTGCAGTTAATCAACGGCTAAACGTTG	This study
SowaaAfor	GATTCTAGAATTCATATGAATCGATTTTTTTACAGTGCGCTGC	This study
SowaaArev	GAATTCGGATCCTTAGGCTGTTTGGCGCGTGAGTAAC	This study
AvrIIfor	GACAGCAAATGGGTCGGGATCCTAGGAATTCGAGCTCCGTCGACAAGCTTGC	This study
AvrIIrev	GCAAGCTTGTCGACGGAGCTCGAATTCCTAGGATCCCGACCCATTTGCTGTC	This study
kdsDfor	GATTCTAGAATTCATATGGTAGACCAATCACAATTGCGCCAATG	This study
kdsDrev	GAATTCGGATCCAAGCTGCAGTTAAATGACTCCCGCTTTAAC	This study
kdsAfor	GATTCTAGAATTCATATGAGTAATAAAATCATAAAGTTAG	This study
kdsArev	GAATTCGGATCCAAGCTGCAGTTATTTGCTGGTATCAATTG	This study
kdsCfor	GATTCTAGAATTCATATGCCACAGCAAGGATTTTACG	This study
kdsCrev	GAATTCGGATCCAAGCTGCAGTTATATGCTCATCCCATGG	This study
kdnAfor	GCGATTCTAGAATTCATATGCCCGGTTTTGAATTATTTGGTC	This study
kdnArev	GAATTCGGATCCAAGCTGCAGTTATTGGGCAAATGCTTTCTTAATG	This study
kdnBfor	GATTCTAGAATTCATATGAGTTTTAAAAATTTTAAAGTAGTTG	This study
kdnBrev	GAATTCGGATCCAAGCAGTTAAATGCGACGGAACAGGGCAG	This study
SokdnAmut1for	CCTTCCGTTATAACTTCGACCACATGCGTAATGATCGCTGGAAGA	This study
SokdnAmut1arev	TTGAAGCTGGTGTACCGTAATGA	This study
SokdnAmut1brev	AGTAATGCCATGTGGTCGAAGTT	This study
SokdnAmut2for	CTGAAATCTACGATAACGCCCACATGTTCTCCGATCATGGCCATG	This study
SokdnAmut2arev	GATAACGCCCACATGTTCTCCGAT	This study
SokdnAmut2brev	ATCGGAGAACATGTGGGCGTTATC	This study

at 37 °C in LB medium (10 g/liter Bacto-tryptone, 5 g/liter yeast extract, 10 g/liter NaCl). Cultures of *S. oneidensis* were also grown aerobically with shaking (250 rpm) at 30 °C in LB medium. When necessary ampicillin (100 μg/ml) was added to the media.

DNA Methods. Standard recombinant DNA methods were used for nucleic acid preparation and analysis.[20] Genomic DNA was prepared from *S. oneidensis* MR-1 (ATCC 700550) using a Qiagen DNeasy Blood and Tissue kit as suggested by the manufacturer. To clone *waaA* from *S. oneidensis* MR-1 into pET16b the primers So*waaA*for/So*waaA*rev were used to amplify the *waaA* gene from genomic DNA. To clone EcKdsB and SoKdsB into pET19b EcKdsBfor/EcKdsBRev or SoKdsBfor/SoKdsBrev respectively were used to amplify the corresponding kdsB gene from genomic DNA. The purified amplified gene products were digested with NdeI and BamHI and were ligated into similarly digested pET19b and transformed into *E. coli* TOP10. Isolated plasmid was verified to be correct by sequencing.

To generate Kd8N in large quantities we adopted a method based on that of Camci-Unal et al.[21] wherein genes involved in Kd8N synthesis were added step-wise. This strategy allows for genes to be placed into the plasmid sequentially using compatible cohesive ends (namely XbaI and AvrII) with a ribosomal binding site preceding each gene. Using site-directed mutagenesis with the primers AvrIIfor/AvrIIrev, an AvrII restriction site was introduced between the BamHI and EcoRI restriction site in the

multiple cloning site of pET21b. This new plasmid, pK8N-1, was transformed into *E. coli* TOP10, isolated, and verified by sequencing. The genes *kdsD* (So_3856)*, kdsA* (So_3827)*, kdsC* (So_3957)*, kdnA* (So_2476)*,* and *kdnB* (So_2477) were PCR amplified from genomic DNA using the primer combinations kdsDfor/kdsDrev, kdsAfor/kdsArev, kdsCfor/kdsCrev, kdnAfor/kdnArev, and kdnBfor/kdnBrev respectively and TOPO cloned into pCR2.1-TOPO. The *kdsC* gene fragment was generated with a silent point mutation changing the stop codon from TGA to TAA in order to remove an internal NdeI restriction site. The *kdnA* gene contained two internal *Nde*I sites, which were removed using the central overlapping primer (COP) site directed mutagenesis protocol[22] as described by Wang *et al.* to yield pCR2.1-*kdnA*-*Nde*IX using the primers SokdnAmut1for/SokdnAmut1arev/SokdnAmut1brev for the first mutation and SokdnAmut2for/SokdnAmut2arev/SokdnAmut2brev for the second mutation. The *kdsD*, *kdsA*, *kdsC*, *kdnB,* and *kdnA* containing pCR2.1-TOPO plasmids were then digested with NdeI and BamHI and the gene fragments were purified and ligated into similarly digested pK8N-1 generating pK8N-2, pK8N-3, pK8N-4, pK8N-5, and pK8N-6, respectively.

To generate pK8N-7, pK8N-2 was digested with *Avr*II and *Hind*III, while pK8N-3 was digested with *Xba*I and *Hind*III. Ligation of digested pK8N-2 with the insert, containing *kdsA*, from pK8N-3 yielded pK8N-7. pK8N-7 was then carried forward and this method was used to generate pK8N-8 and

pK8N-9, from the carried forward plasmid along with pK8N-4 and pK8N-5, respectively. The *kdn*A gene (pK8N-6) contains an internal *Hind*III site. Our cloning system allows for genes to be cloned in by digestion of the growing plasmid (pK8N-9) with *Avr*II and digestion of pK8N-6 with *Xba*I and *Avr*II, as *Xba*I and *Avr*II form compatible cohesive ends. Ligation of the insert into the plasmid destroys the upstream *Avr*II site, while maintaining the downstream *Avr*II site for further gene insertions. This method can be used in the generation of pK8N-10.

A synthetic construct, pUC57-SoK8N (complete sequence of insert given in Appendix C) was designed and purchased from Genscript in their CloneEZ method with *E. coli* codon optimization and the removal of internal restriction sites by silent mutation.

KdsB purification. Decahistidine-tagged EcKdsB and SoKdsB were similarly overexpressed in BL21(DE3) cells from pET19b overnight at 17 ^{o}C. The cells resuspended in 50 mM Tris pH 8.0, 500 mM NaCl and were then lysed by sonication. The enzymes were purified in a single step purification by Ni^{2+} affinity chromatography with a 5 mL Ni-IMAC HisTrap FF column (GE Healthcare) with washing consisting of resuspension buffer with 50 mM imidazole and elution with 500 mM imidazole. The proteins were purified to >95% purity as judged by SDS-PAGE analysis. The proteins were dialyzed against 10 mM Tris pH 8.0 with final protein concentrations determined by

Bradford assay with bovine serum albumin as a standard. Protein samples were stored at -80 °C.

KdsB assay. The activity of EcKdsB and SoKdsB was determined using the continuous EnzChek pyrophosphate assay kit (Invitrogen) as reported previously.[11] In brief, 200 µL reaction mixtures containing the appropriate enzymes and substrates for the EnzChek kit as specified by the manufacturer and the KdsB substrates, 100 mM CTP, and between 25-1000 µM Kdo analog, either Kdo (Sigma), Kd8N (synthesized by Vahlteich Medicinal Chemistry Core, University of Michigan), or Kd8Z (synthesized by Vahlteich Medicinal Chemistry Core, University of Michigan) were initiated by the addition of 20 nM EcKdsB or SoKdsB in 10 mM Tris pH 8.0, 5 mM $MgCl_2$ and incubated at 37 °C in 96-well plates in triplicate with appropriate controls. Initial rates were measured at 360 nm for 10 min.

LPS production and purification. KPM56 cells containing pET16b (EV) or pET16b-SoWaaA, *S. oneidensis* MR-1 and *S. oneidensis* Δ*kdnA* Δ*kdnB*[12] (a generous gift from Pei Zhou) were grown and prepared as described in Chapter II. LPS samples were purified by a modified version of the phenol-chloroform-petroleum ether procedure[23] as reported previously.[24] Samples were resuspended into LPS-free water (Braun water) to a concentration of 1 mg/mL for mass spectrometry analysis.

Electrospray-Ionization Fourier-transformed Ion Cyclotron Mass Spectrometry. LPS samples were analyzed by mass spectrometry as described in Chapter II.

Kd8Z incorporation and Microscopy. The *E. coli* cell line KPM56 which is capable of surviving without core oligosaccharide in the LPS due to lack of WaaA or WaaC was transformed with pET16b or pET16b-SoWaaA. Cells were grown with or without the presence of 4 mM Kd8Z, treated, and visualized by fluorescence microscopy as previously.[17]

Results

Generation of Kd8N from in vivo expression. Initial plasmids containing the *S. oneidensis* Kd8N biosynthetic genes were successfully generated and added iteratively by the method of Camci-Unal et al.[21] as described excluding the addition of the *kdnA* gene fragment. Each clone generated upon addition of the *kdnA* fragment contained a mutation in said gene. After many attempts, it was decided to purchase a preconstructed *E. coli* codon optimized gene construct constructed similarly to what would have been produced by the Camci-Unal method (see Appendix C for sequence). Transformants of the constructed plasmid were successful in *E. coli* TOP10 cells and the plasmid could be propagated and purified by Qiagen

miniprep. Attempts to transform the pUC57-SoK8N plasmid into *E. coli* BL21(DE3) for overexpression and Kd8N production were wholly unsuccessful, and we were unable to generate any Kd8N by *in vivo* expression of these genes in our *E. coli* system. This is likely due to toxicity from overproduction of Kd8N in *E. coli* cells either through depletion of Kdo or NAD^+ or glutamate used in the production of Kd8N.

KdsB substrate selectivity. Using the enzyme-coupled Enzchek pyrophosphate kit, it is possible to measure continuous enzyme activity of *E. coli* KdsB.[11] With this method pyrophosphate generated during the reaction of Kdo and CTP is cleaved by inorganic pyrophosphatase. The resulting phosphate is used by purine nucleoside phosphorylase to cleave the guanosine analog 2-amino-6-mercapto-7-methylpurine ribonucleoside giving a signal at 360 nm. Using this assay, it is possible to similarly test substrate analogs of Kdo, namely Kd8N or Kd8Z in this case, as well of other KdsBs. This methodology was used in order to probe the substrate specificities of EcKdsB and SoKdsB with these substrates. Each enzyme was tested with a range of Kdo analog concentrations at fixed CTP concentration. The dependence of the initial rate on Kdo analog concentration follows Michaelis-Menten kinetics and is show in Figure 3.5 with calculated kinetic parameters in Table 3.2. These kinetic data show that EcKdsB has strong preference for its native substrate Kdo over the other analogs Kd8N or Kd8Z. Conversely, SoKdsB has a strong preference for its native substrate Kd8N over the

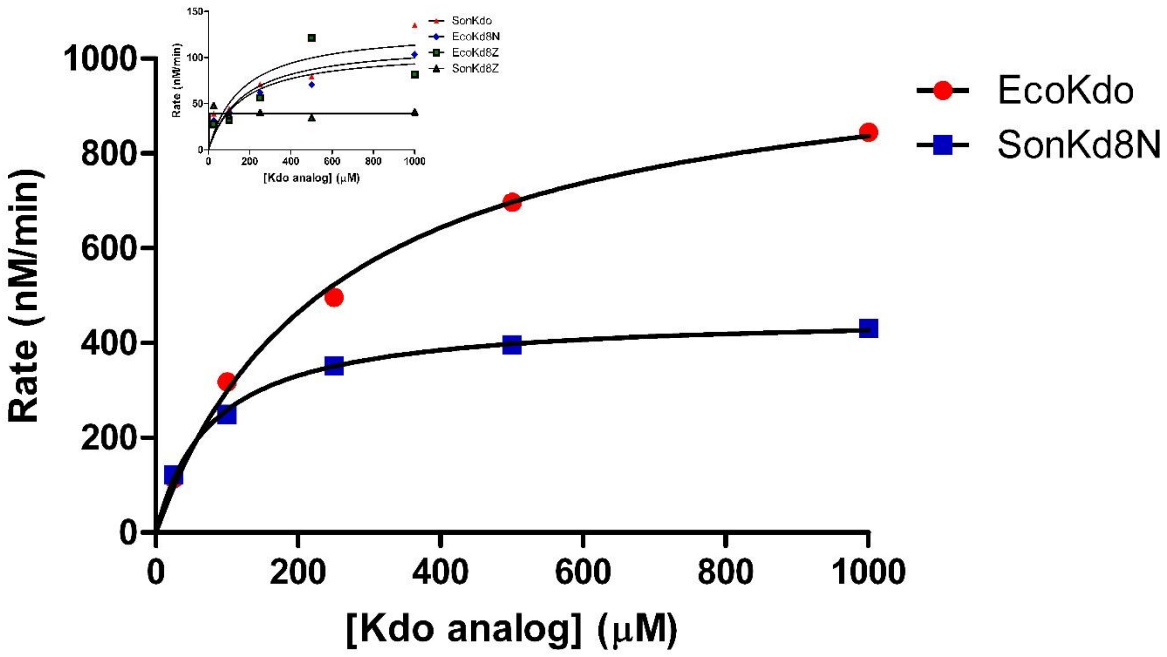


Figure 3.5. Kinetic curves for KdsB and various substrates. Initial rate curves upon varying Kdo analog concentration. Red circles- EcKdsB with Kdo, Red triangle- SoKdsB with Kdo, Blue diamond, EcKdsB with Kd8N, Blue square- SoKdsB with Kd8N, Green square- EcKdsB with Kd8Z, Green triangle- SoKdsB with Kd8Z. With the exception of the natural substrates with their endogenous enzymes most combinations show little to no activity. EcKdsB and SoKdsB show highest activities with Kdo and Kd8N respectively.

Enzyme and substrate	**K_m**	**V_{max}**	**k_{cat}**	**k_{cat}/K_m**
EcoKdo	250 ± 14 μM	1046 ± 21 nM/min	1.74 ± 0.04 s^{-1}	0.00696 μM^{-1} s^{-1}
SonKd8N	77.76 ± 2.25 μM	459.9 ± 6.0 nM/min	0.77 ± 0.01 s^{-1}	0.0099 μM^{-1} s^{-1}

Table 3.2. Kinetic parameters of EcKdsB and SoKdsB with their best substrates. Kinetic values of the best substrate for each EcKdsB with Kdo (EcoKdo) and SoKdsB with Kd8N (SonKd8N).

others Kdo or Kd8Z. Neither EcKdsB nor SoKdsB showed much activity with the non-native substrate analogs under the conditions used in this assay preventing kinetic parameters from being measured.

Incorporation of azido-Kdo into the LPS of KPM56. KPM56 is a useful tool for the study of Kdo transferases due to its suppressor mutation in *yhjD*, a gene of unknown function.[25] Heterologous studies of the monofunctional Kdo transferase from *A. aeolicus* have previously been performed[19] showing its usefulness for these studies. A recent click-based fluorescence chemistry study showed the ability to label the outer membrane of a number of Gram-negative bacteria including *E. coli* K-12, the progenitor of KPM56, with *S. oneidensis* being the only species attempted which could not be labeled.[17] Using KPM56, it is possible to determine if the Kdo transferase from *S. oneidensis* is capable of glycosylating Lipid IV_A with Kd8Z. Figure 3.5 shows that when KPM56 expressing heterologous SoWaaA without Kd8Z is not fluorescently labeled while the same cells when grown with 4 mM Kd8Z shows fluorescent labeling demonstrating incorporation of Kd8Z into the early core of the LPS.

Mass spectrometric analysis of isolated LPS samples. LPS samples were generated, purified, and analyzed from the cell line KPM56[19] which is an *E. coli* cell line lacking the genes w*aaA* or *waaC*. Due to the absence of these genes waaA non-native Kdo transferases can be studied *in trans* allowing for characterization of these corresponding proteins in an *E. coli* system.

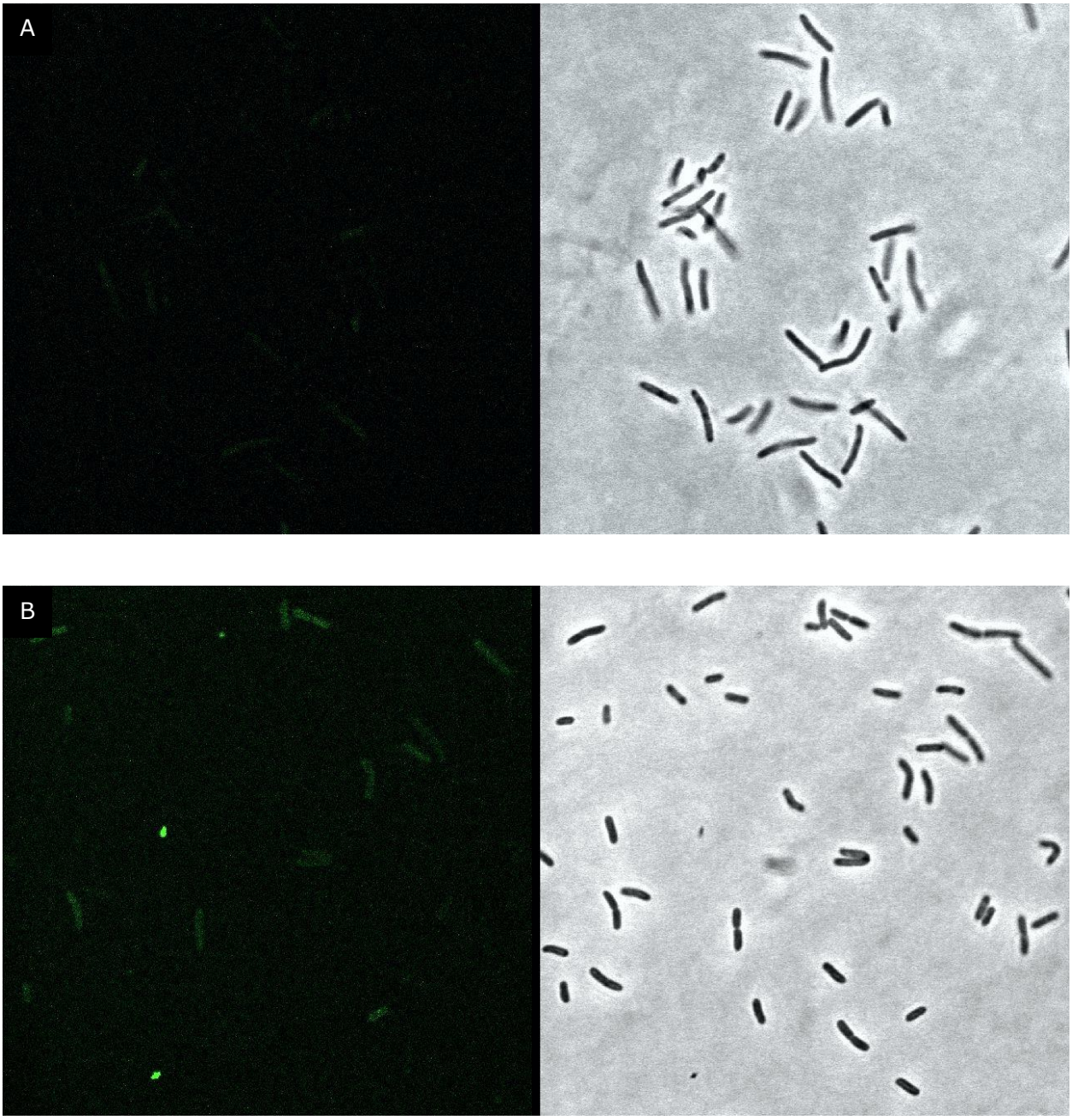


Figure 3.6. Fluorescent images of Kd8Z incorporation. Fluorescence images of *E. coli* KPM56 grown without (A) or with (B) the presence of 4 mM Kd8Z showing metabolic incorporation labeled with Alexa Fluor 488 fluorophore.

The lack of the WaaC protein prevents further heptosylation of the oligosaccharide which allows for simplification of the LPS analysis. When pET16b (EV) is transformed into the KPM56 and the LPS purified and analyzed by mass spectrometry only Lipid IV_A (1,4'-bisphosphorylated tetraacylated Lipid A, observed mass 1404.861 u, calculated mass 1404.854 u) and the phosphoethanolamine-modified Lipid IV_A (observed mass 1527.860 u, calculated mass 1527.862 u) is detected. Upon the addition of the pET16b-SoWaaA into KPM 56, The mass spectrum of the LPS shows the addition of peaks representing Kdo modification. Kdo-Lipid IV_A appears (observed mass 1624.907 u, calculated mass 1624.912 u) as well as the same species modified with phosphoethanolamine (observed mass 1747.917 u, calculated mass 1747.921 u) and then further modified with lauric acid (observed mass 1930.087 u, calculated mass 1930.088 u). The mass spectra can be seen in Figure 3.6.

Additional samples were purified and analyzed from *S. oneidensis* MR-1 and *S. oneidensis* Δ*kdnA* Δ*kdnB*. These samples show a pattern consistent with a full-length oligosaccharide core LPS. For the WT *S. oneidensis* strain the m/z peak of 2908.39 u seems to correspond of unmodified full LPS. Unfortunately, the exact mass of these species cannot be calculated as the full structure of *S. oneidensis* is unknown, although the structure of the core oligosaccharide has previously been determined.[7] The corresponding m/z peak for the mutant strain lacking the ability to generate Kd8N is 2909.37 u.

Both sets of peaks show Δm/z= ±14.01 u characteristic of longer or shorter fatty acid addition on the Lipid A portion of LPS. Additionally, both sets of spectra show the addition of Δm/z= 123 u characteristic of the addition of a single phosphoethanolamine adduct and Δm/z=80.07 u characteristic of the addition of a phosphate. These modifications are verification that these peaks correspond to those of LPS molecules purified from these cells. The spectra of the full-length LPS samples from *S. oneidensis* and the mutant can be seen in Figure 3.7.

The difference between the unmodified LPS masses for the WT and mutant samples of Δm/z=0.98 u. This mass difference most likely corresponds with a difference of incorporation of a single Kdo moiety into the LPS compared to a single Kd8Z. Kdo has an exact mass of 238.07 u while Kd8Z has an exact mass of 237.08 u giving a mass difference of 0.99 u. This shows that the mutant strain lacking Kd8Z is able to substitute Kdo into the core oligosaccharide of the LPS.

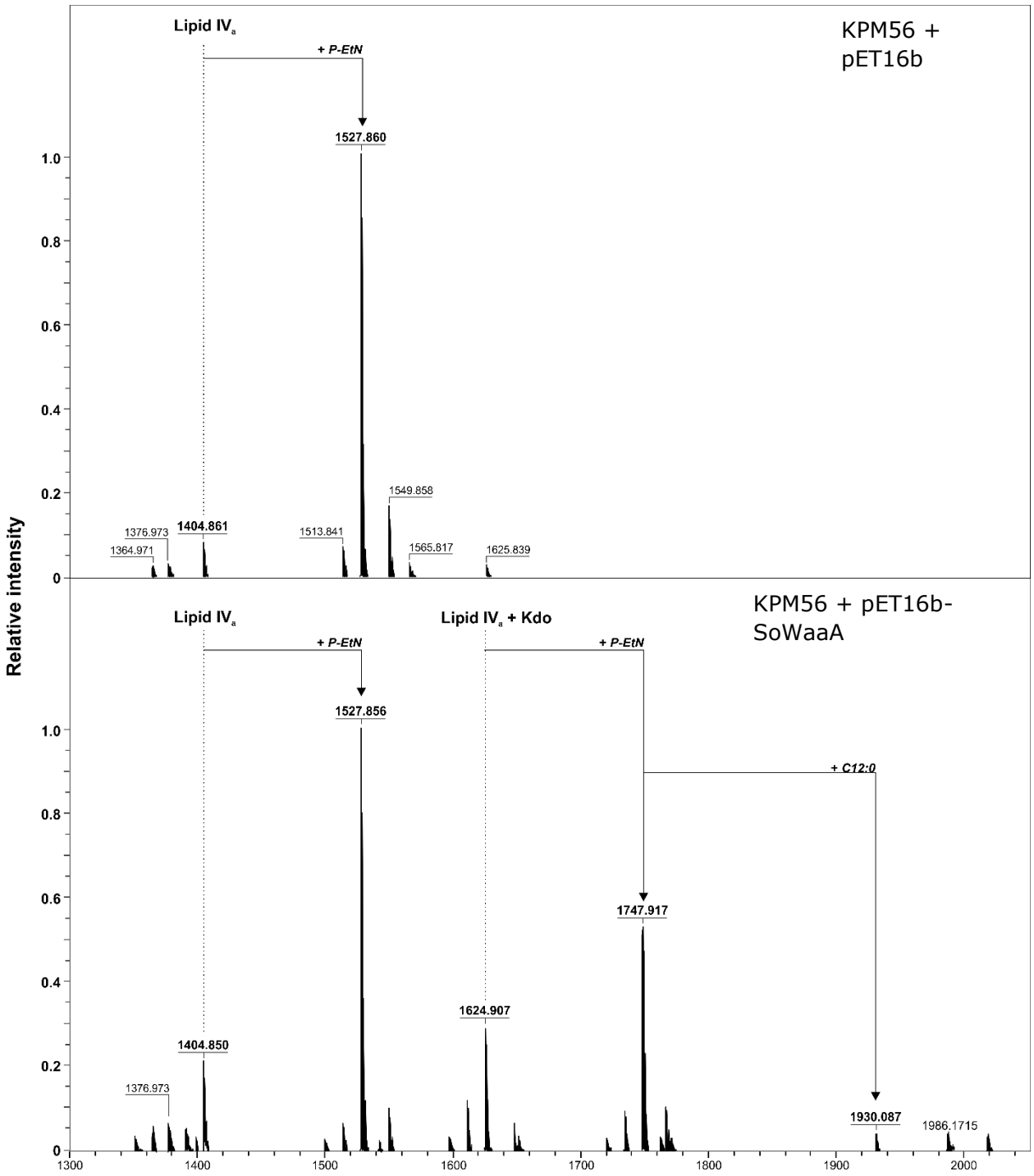


Figure 3.7. **Mass spectra for incorporation of Kdo by SoWaaA.** Mass spectra of the LPS purified from KPM56 containing either pET16b (A) or pET16b-SoWaaA (B). The addition of SoWaaA generates a glycosylated Kdo-Lipid IV_A molecule showing that SoWaaA is capable of utilizing Kdo.

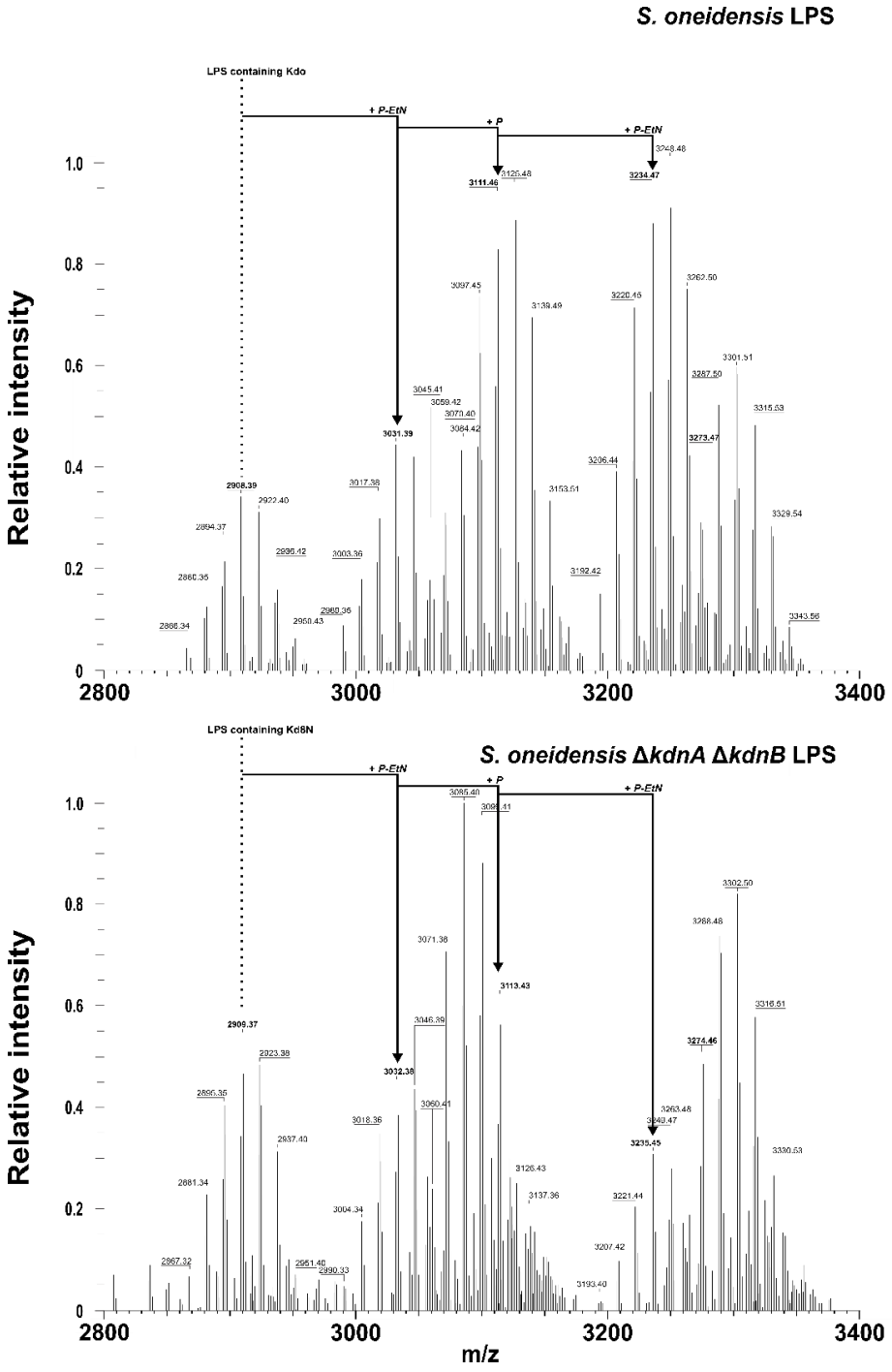


Figure 3.8. ***S. oneidensis*** **and Kd8N knockout full length LPS spectra**. Mass spectra of the LPS purified *S. oneidensis* of the Δ*kdnA* Δ*kdnB* mutant. Typical fatty acid variation pattern of Δm/z=±14.01 u can be seen in the samples as well as common LPS modifications. A Δm/z=0.98 u can be seen between equivalent peaks in the spectra representing a difference of Kdo vs Kd8N.

Discussion

Shewanella spp. with the only known exception being *S. putrefaciens* are unique among known Gram-negative bacteria for their ability to incorporate Kd8N into the core oligosaccharide of their LPS. The pathway by which *S. oneidensis* biosynthesizes Kd8N was recently discovered.[12] To begin understanding the method by which *S. oneidensis* can selectively incorporate Kd8N over Kdo, we attempted to generate a large quantity of Kd8N using the method of Camci-Unal et al.[21] Unfortunately this proved intractable due to the inability of *E. coli* cell lines tested to grow in the presence of the Kd8N biosynthetic genes. This could be due to unregulated overconsumption of NAD^+ or glutamate by KdnB or KdnA in the production of the amino sugar (see Figure 3.2). The sugar was instead generated synthetically by the Vahlteich Medicinal Chemistry Core for further studies.

KdsB from *E. coli* is well studied both kinetically and structurally[11, 26, 27], while the KdsB from *S. oneidensis* is poorly characterized. While these proteins show a high amount of conservation (45% identity), it is known that SoKdsB is able to activate Kd8N *in vivo* while EcKdsB is not.[12] In order to verify this selectivity *in vitro* as well as to study the substrate preferences of SoKdsB, the enzymes were probed by the Enzchek pyrophosphate assay. In these assays, it was verified that indeed EcKdsB has a strong preference for Kdo over Kd8N as was suggested by the *in vivo* studies. Additionally,

SoKdsB also seemed to prefer its native substrate of Kd8N over Kdo in our reaction conditions. Both enzymes had virtually no detectable activity for Kd8Z under these conditions. From these enzyme assays it seems that SoKdsB could have a large effect in the overall selectivity of *S. oneidensis* for Kd8N over Kdo in the core oligosaccharide.

Previous attempts to fluorescently label *S. oneidensis* through click-mediated chemistry were unsuccessful.[17] The authors of the study stated that the inability to incorporate Kd8Z into the LPS was due to Kd8N being the native substrate rather than Kdo which prevents the incorporation of Kd8Z. To determine if SoWaaA is capable of incorporating Kd8Z into the LPS of an *E. coli* cell line, KPM56, lacking both *waaA* and *waaC*, was used. Using this cell line, we see that SoWaaA seems to be capable of incorporating Kd8Z such that the cells can be fluorescently labeled in the presence of the azido sugar as visualized by fluorescence microscopy. This shows that SoWaaA is not specific to Kd8N and can incorporate Kdo analogs into LPS. Carrying on from this, we tested the ability of SoWaaA to transfer Kdo into the core oligosaccharide in KPM56 in the absence of both Kd8N or Kd8Z. The mass spectrum of the cell line with the empty vector pET16b shows that the cell line alone is incapable of addition of any sugar into the core oligosaccharide while expression of the pET16b-SoWaaA allows the incorporation of a single Kdo moiety into the LPS. This shows for the first time that SoWaaA is able to act as a monofunctional Kdo transferase *in vivo*.

It seems from these experiments that SoWaaA is capable of accepting a range of Kdo analogs for incorporation into LPS and likely has little to do with Kd8N selectivity.

The genes *kdnA* and *kdnB* are now known to be responsible for the generation of Kd8N in *S. oneidensis*. In their study, Gattis *et al.* removed these genes showing that they are not only dispensable for survival of the cell, but also had a marginal effect on susceptibility to polymyxin B or bile salts, 3-fold and 2-fold increase respectively.[12] One would suspect that if *S. oneidensis* had a strict selectivity for Kd8N over Kdo lack of *kdnA* and *kdnB* would prevent incorporation of Kd8N into the core oligosaccharide leading to complete truncation of the LPS. While, *E. coli* cells lacking the core oligosaccharide show a 125-fold increase in sensitivity to bile salts compared to wild-type;[28] the lack of comparatively similar increase in sensitivity led us to believe that the LPS of the *S. oneidensis* lacking Kd8N maintained the ability to generate a core oligosaccharide. We obtained the mutant strain and grew, purified, and analyzed the LPS along with that of wild-type *S. oneidensis*. As expected both samples showed mass peaks characteristic of rough type LPS samples. Indeed, the samples of the mutant LPS varied only by a $\Delta m/z$ of 0.98 u corresponding to a difference of Kd8N vs Kdo. This shows that not only is Kd8N dispensable for the cells, but also that Kdo is able to substitute completely for Kd8N in the LPS and able to generate fully elaborated core oligosaccharide. This shows that although SoKdsB has

selectivity *in vitro,* it seems to be able to compensate *in vivo* to allow for full LPS elaboration. In the end the true selectivity of Kd8N selection over Kdo selection may be determined solely by the availability of Kd8N over Kdo. *S. putrefaciens* CN-32 is the only known species of *Shewanella* to date which does not incorporate Kd8N into its LPS.[8] A BLAST search of KdnB or KdnA finds no closely related homologs of either in this subspecies of *S. putrefaciens* demonstrating that a native *Shewanella* species can naturally survive without these genes and generate full LPS.

Although most studied strains of *Shewanella* contain Kd8N in their core LPS, this unique sugar is not essential for growth or full LPS elaboration in *S. oneidensis* MR-1. The KdsB of this species *in vitro* seems to have a preference for Kd8N over Kdo, but in the end the overall preference of the cell seems to hinge on the presence or absence of KdnA or KdnB in the cell. Further studies are needed to fully elucidate the reason for which *Shewanella* typically display Kd8N in their outer membrane rather than Kdo, typical of most other Gram-negative species.

Chapter contributions/acknowledgements

We thank Brigitte Kunz and Uwe Mamat of Forschungzentrum Borstel for her assistance in performing mass spectrometry experiments. We thank David

Cech, Joseph Madak, Pan-Fen Wang, and Melissa Holt for assistance with cloning of the Kd8N generating plasmid. We thank Uwe Mamat for assistance with the fluorescence microscopy experiments. We thank Rod Sorenson and Kim Hutchings for synthesis of Kd8Z and Kd8N. All other work and writing was done by ACP.

References

[1] Hau, H. H., and Gralnick, J. A. (2007) Ecology and Biotechnoogy of the Genus Shewanella, *Annual Review of Microbiology 61*, 237-2358.

[2] Hong, Y.-G., and Gu, J.-D. (2010) Physiology and biochemistry of reduction of azo compounds by Shewanella strains relevant to electron transport chain, *Appl Microbiol Biot 88*, 637-643.

[3] Janda, J. M. (2014) Shewanella: a Marine Pathogen as an Emerging Cause of Human Disease, *Clinical Microbiology Newsletter 36*, 25-29.

[4] Gram, L., Ravn, L., Rasch, M., Bruhn, J. B., Christensen, A. B., and Givskov, M. (2002) Food spoilage-interactions between food spoilage bacteria, *International Journal of Food Microbiology 78*, 79-97.

[5] Fredrickson, J. K., Romine, M. F., Beliaev, A. S., Auchtung, J. M., Driscoll, M. E., Gardner, T. S., Nealson, K. H., Osterman, A. L., Pinchuk, G., Reed, J. L., Rodionov, D. A., Rodrigues, J. L. M., Saffarini, D. A., Serres, M. H., Spormann, A. M., Zhulin, I. B., and Tiedje, J. M. (2008) Towards environmental systems biology of Shewanella, *Nature Reviews Microbiology 6*, 592-603.

[6] Raetz, C. R. H., and Whitfield, C. (2002) Lipopolysaccharide Endotoxins, *Annual Review of Biochemistry 71*, 635-700.

[7] Vinogradov, E., Korenevsky, A., and Beveridge, T. J. (2003) The structure of the rough-type lipopolysaccharide from Shewanella oneidensis MR-1, containing 8-amino-8-deoxy-Kdo and an open-chain form of 2-acetamido-2-deoxy-d-galactose, *Carbohydrate Research 338*, 1991-1997.

[8] Vinogradov, E., Korenevsky, A., and Beveridge, T. J. (2002) The structure of the carbohydrate backbone of the LPS from Shewanella putrefaciens CN32, *Carbohydrate Research 337*, 1285-1289.

[9] Claesson, A., Luthman, K., Gustafsson, K., and Bondesson, G. (1987) A 2-deoxy analogue of KDO as the first inhibitor of the enzyme CMP-KDO synthetase, *Biochemical and biophysical research communications 143*, 1063-1068.

[10] Ohrui, H., Morita, M., and Meguro, H. (1992) A New Synthesis of 2-Deoxy-Beta-Kdo, a Potent Cmp-Kdo Synthetase Inhibitor, *Carbohydrate Research 224*, 319-325.

[11] Heyes, D. J., Levy, C., Lafite, P., Roberts, I. S., Goldrick, M., Stachulski, A. V., Rossington, S. B., Stanford, D., Rigby, S. E. J., Scrutton, N. S., and Leys, D. (2009) Structure-based mechanism of CMP-2-keto-3-deoxymanno-octulonic acid synthetase: convergent evolution of a sugar-activating enzyme with DNA/RNA polymerases, *The Journal of biological chemistry 284*, 35514-35523.

[12] Gattis, S., Chung, H., Trent, M., and Raetz, C. (2013) The origin of 8-amino-3,8-dideoxy-D-manno-octulosonic acid (Kdo8N) in the lipopolysaccharide of Shewanella oneidensis, *The Journal of biological chemistry 288*, 9216-9225.

[13] Zachman-Brockmeyer, T. R., Thoden, J. B., and Holden, H. M. (2016) Structures of KdnB and KdnA from Shewanella oneidensis: Key Enzymes in the Formation of 8-Amino-3,8-Dideoxy-D-Manno-Octulosonic Acid, *Biochemistry 55*, 4485-4494.

[14] Helander, I. M., Kato, Y., Kilpeläinen, I., Kostiainen, R., Lindner, B., Nummila, K., Sugiyama, T., and Yokochi, T. (1996) Characterization of Lipopolysaccharides of Polymyxin-Resistant and Polymyxin-Sensitive Klebsiella pneumoniae O3, *European Journal of Biochemistry 237*, 272-278.

[15] Nummila, K., Kilpeläinen, I., Zähringer, U., Vaara, M., and Helander, I. M. (1995) Lipopolysaccharides of polymyxin B-resistant mutants of Escherichia coii are extensively substituted by 2-aminoethyl

pyrophosphate and contain aminoarabinose in lipid A, *Molecular Microbiology 16*, 271-278.

[16] Olaitan, A. O., Morand, S., and Rolain, J.-M. (2014) Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria, *Frontiers in Microbiology 5*, 643.

[17] Dumont, A., Malleron, A., Awwad, M., Dukan, S., and Vauzeilles, B. (2012) Click-Mediated Labeling of Bacterial Membranes through Metabolic Modification of the Lipopolysaccharide Inner Core, *Angewandte Chemie International Edition 51*, 3143-3146.

[18] Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of Molecular Biology 189*, 113-130.

[19] Mamat, U., Schmidt, H., Munoz, E., Lindner, B., Fukase, K., Hanuszkiewicz, A., Wu, J., Meredith, T., Woodard, R., Hilgenfeld, R., Mesters, J., and Holst, O. (2009) WaaA of the hyperthermophilic bacterium Aquifex aeolicus is a monofunctional 3-deoxy-D-manno-oct-2-ulosonic acid transferase involved in lipopolysaccharide biosynthesis, *The Journal of biological chemistry 284*, 22248-22262.

[20] Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd Edition*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[21] Camci-Unal, G., Mizanur, R. M., Chai, Y., and Pohl, N. L. B. (2012) Synthesis of a 3-deoxy-d-manno-octulosonic acid (KDO) building block from d-glucose via fermentation, *Organic & biomolecular chemistry 10*.

[22] Wang, H., Zhou, N., Ding, F., Li, Z., Chen, R., Han, A., and Liu, R. (2011) An efficient approach for site-directed mutagenesis using central overlapping primers, *Anal. Biochem. 418*, 304-306.

[23] Galanos, C., Luderitz, O., and Westphal, O. (1969) A new method for the extraction of R lipopolysaccharides, *Eur J Biochem 9*, 245-249.

[24] Mamat, U., Meredith, T. C., Aggarwal, P., Kühl, A., Kirchhoff, P., Lindner, B., Hanuszkiewicz, A., Sun, J., Holst, O., and Woodard, R. W. (2008) Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-d-manno-oct-2-ulosonic acid-depleted Escherichia coli, *Molecular Microbiology 67*, 633-648.

[25] Mamat, U., Meredith, T. C., Aggarwal, P., Kuhl, A., Kirchhoff, P., Lindner, B., Hanuszkiewicz, A., Sun, J., Holst, O., and Woodard, R. W. (2008) Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-D-manno-oct-2-ulosonic acid-depleted Escherichia coli, *Molecular Microbiology 67*, 633-648.

[26] Goldman, R. C., and Kohlbrenner, W. E. (1985) Molecular cloning of the structural gene coding for CTP:CMP-3-deoxy-manno-octulosonate cytidylyltransferase from Escherichia coli K-12, *J. Bacteriol. 163*, 256-261.

[27] Yi, L., Velasquez, M. S., Holler, T. P., and Woodard, R. W. (2011) A simple assay for 3-deoxy-d-manno-octulosonate cytidylyltransferase and its use as a pathway screen, *Anal. Biochem. 416*, 152-158.

[28] Meredith, T. C., Aggarwal, P., Mamat, U., Lindner, B., and Woodard, R. W. (2006) Redefining the Requisite Lipopolysaccharide Structure in Escherichia coli, *ACS chemical biology 1*, 33-42.

Chapter IV

Summary, conclusions, and future directions

The focus of this dissertation was on the study of variability in the early core oligosaccharide in Gram-negative bacteria. A cell line was constructed which not only lacked the ability to biosynthesize a lipopolysaccharide (LPS) core natively but was shown to also be able to generate unnatural early core structures by expressing exogenous genes. Additionally, we explored the selectivity of the species *Shewanella oneidensis* in the incorporation of the unique sugar 8-amino-3,8-dideoxy-D-*manno*-octulosonic acid (Kd8N).

LPS is an interesting, yet highly complex glycolipid which is found in nearly all Gram-negative bacteria. The presence of LPS is essential to the bacteria for their survival and well-being but plays a large role in bacterial pathogenesis. Since the discovery of the endotoxic properties of LPS vast amounts of research and study have gone into how to alleviate its detrimental properties while simultaneously trying to develop it for the use as an antigen for vaccine therapies. Unfortunately, throughout all of these

previous studies, the ability to generate analogs of LPS has been limited to single knockouts and complex multi-step synthetic production.

With the creation of the WOD02 cell line, it is now possible to obtain core oligosaccharide LPS modifications in a straightforward and simple fashion. This cell line represents a tool which can be carried forward to not only further generate complete LPS core, but also study biosynthetic properties of glycosyltransferases and other LPS modifying enzymes that would have been extremely difficult otherwise. To date, many cell lines have been created to generate and study truncated LPS glycolipids, but all of these cell lines have had drawbacks either in very slow growth rates, lack of ability to generate necessary carbohydrates, requiring supplementation for growth, containing enzymes which complicate proper glycoform production, or are unable to generate non-native glycoforms. The generation of the WOD02 cell line is the first cell line to fully circumvent all of these drawbacks giving a viable cell line with a blank canvas for complete generation and study of LPS and LPS biosynthesis. The availability of such a tool could allow for faster development of non-toxic LPS based vaccines.[1] This cell line represents a step forward in our ability to quickly and easily generate glycoforms essential for research and medicinal usage.

Many questions remain in the realm of LPS biosynthesis. While the biosynthesis of the Lipid A region has largely been elucidated,[2] the biosynthesis of the oligosaccharide core remains mostly unknown. Utilizing

the WOD02 system we have generated, it is now possible as we have shown to create novel glycolipids. This cell line has the potential to be used in order to elucidate the LPS structures of many Gram-negative bacteria which are not directly amenable to growth in laboratories. By inserting the core oligosaccharide biosynthetic genes we could generate a fully elaborated glycolipid in a clean *E. coli* background system giving use useful information on the properties of the outer membrane of a vast number of Gram-negative bacteria. Additionally, the use of this cell line allows us to generate various substrates in order to study enzyme function and selectivity. Studying the selectivity of glycosyltransferases involved in LPS core biosynthesis has been to date a complex undertaking. In order to study these enzymes *in vitro*, one would need to biosynthesize a complex amphipathic glycolipid through a multistep process. Truncated native glycolipids could be generated for *in vivo* studies, but this would require removing genomic copies of genes in the species of choice which could be quite complicated. On the other hand, with the creation of the WOD02 cell line, it is now possible to biosynthesize your substrate of choice simply by expressing the desired genes to generate the glycolipid which could be used for *in vivo* studies or further purified for *in vitro* studies. The ability to simply generate these substrates greatly opens the door for understanding the biosynthesis and selectivity of these late glycosyltransferases which to date have been nearly impossible to properly characterize. The exact function of even some of the core modifying genes

within *E. coli* is not yet fully understood, and only recently with gene expression studies are we beginning to see a connected effect between LPS glycoform and gene regulation.[3] Studies with the WOD02 cell line are allowing us to understand through gene expression studies the effects of LPS modifications on the bacteria themselves which often lead to increased stress responses from LPS truncations. Having a cell line lacking all core sugars opens many doors to studies which could not have been accomplished otherwise.

However, other more unique modifications of core LPS exist as well though which are less directly amenable to the WOD02. In these cases, such as for Kd8N, the organisms in question must be probed. Kd8N exists as one of two rare, alternate Kdo analogs seen in Gram-negative bacteria. The reason that such a sugar is used in these particular Gram-negative bacteria is still unknown, but we have advanced our understanding of the pathway by which Kd8N is biosynthesized[4] and gained insight into how this sugar is selected over Kdo. Further studies are needed to elucidate the effects Kd8N might have on endotoxicity of the LPS upon incorporation, and what function this unique sugar has for these bacteria. By now understanding how the Kd8N is selected in *S. oneidensis*, it could now to possible to combine this selectivity together with the blank canvas generated from WOD02 to create even more unique LPS core glycoforms. This would further push our understanding of outer membrane stability in Gram-

negative bacteria and give us an additional route to create new LPS analogs for vaccine and adjuvant studies.

It is my hope that this technology will be carried on and help us to overcome many of the hurdles we currently have in generating vaccine therapies of LPS against a number of deadly infectious bacteria. With this technology the rapid generation of analogs is now a distinct possibility and should allow us to bring these ever important lifesaving therapies to fruition. This technology opens us up to many new unexplored possibilities that I believe will be a great stepping stone to our full understanding of LPS core function, biosynthesis, and use as a medicine.

References

[1] Cross, A. S. (2013) Anti-endotoxin vaccines: back to the future, *Virulence 5*, 219-225.

[2] Trent, M. S. (2004) Biosynthesis, transport, and modification of lipid A, *Biochem Cell Biol 82*, 71-86.

[3] Klein, G., Stupak, A., Biernacka, D., Wojtkiewicz, P., Lindner, B., and Raina, S. (2016) Multiple Transcriptional Factors Regulate Transcription of the rpoE Gene in Escherichia coli under Different Growth Conditions and When the Lipopolysaccharide Biosynthesis Is Defective, *The Journal of biological chemistry 291*, 22999-23019.

[4] Gattis, S., Chung, H., Trent, M., and Raetz, C. (2013) The origin of 8-amino-3,8-dideoxy-D-manno-octulosonic acid (Kdo8N) in the lipopolysaccharide of Shewanella oneidensis, *The Journal of biological chemistry 288*, 9216-9225.

APPENDICES

Appendix A

Structural and computational dissection of the catalytic mechanism of the inorganic pyrophosphatase from *Mycobacterium tuberculosis*

Published: Pratt, Andrew C., Dewage, Sajeewa W., Pang, Allan H., Biswas, Tapan, Barnard-Britson, Sandra, Cisneros, G. Andres, Tsodikov, Oleg V., (2015), *J. Struct. Biol.*, 192, 1, 76-87

Summary

Family I inorganic pyrophosphatases (PPiases) are ubiquitous enzymes that are critical for phosphate metabolism in all domains of life. The detailed catalytic mechanism of these enzymes, including the identity of the general base, is not fully understood. We determined a series of crystal structures of the PPiase from Mycobacterium tuberculosis (Mtb PPiase) bound to catalytic metals, inorganic pyrophosphate (PPi; the reaction substrate) and to one or two inorganic phosphate ions (Pi; the reaction product), ranging in resolution from 1.85 to 3.30Å. These structures represent a set of major kinetic intermediates in the catalytic turnover pathway for this enzyme and suggest an order of association and dissociation of the divalent metals, the substrate and the two products during the catalytic turnover. The active site of Mtb PPiase exhibits significant structural differences from the well characterized

Escherichia coli PPiase in the vicinity of the bound PPi substrate. Prompted by these differences, quantum mechanics/molecular mechanics (QM/MM) analysis yielded an atomic description of the hydrolysis step for Mtb PPiase and, unexpectedly, indicated that Asp89, rather than Asp54 that was proposed for *E. coli* PPiase, can abstract a proton from a water molecule to activate it for a nucleophilic attack on the PPi substrate. Mutagenesis studies of the key Asp residues of Mtb PPiase supported this mechanism. This combination of structural and computational analyses clarifies our understanding of the mechanism of family I PPiases and has potential utility for rational development of drugs targeting this enzyme.

Introduction

Inorganic pyrophosphate (PP_i) is produced during the metabolism of various biologically important molecules such as proteins, nucleic acids, and fatty acids.[1] The hydrolysis of PP_i into two molecules of inorganic phosphate (P_i) in the cytoplasm is catalyzed by inorganic pyrophosphatase (PPiase). This enzymatic degradation of PP_i alleviates inhibitory effects of PP_i on PP_i-generating reactions and drives them thermodynamically.[2, 3] Inorganic pyrophosphatase displays significant catalysis in the reverse direction, to form PP_i, with a rate constant for the PP_i synthesis step only 5-fold smaller

than that for the hydrolysis step for *Escherichia coli* inorganic pyrophosphatase.[4] Inorganic pyrophosphatases are divided into four families: families I, II and III of cytoplasmic pyrophosphatases and a family of integral membrane H^+/Na^+-dependent inorganic pyrophosphatases. Family I PPiases are ubiquitous and present in most eukaryotes, archaea and bacteria, family II enzymes are found only in some bacteria, archaea and primitive eukaryotes and family III PPiases are relatively unexplored enzymes found in a few bacterial species.[5-9] A small number of bacteria including *Vibrio cholerae* contain both family I and family II PPiases.[10] Even though PPiases from families I and II are not homologous to each other in sequence or overall structure, they employ similar divalent metal-dependent catalysis; family III appear to be modified haloalkane dehalogenases, which function by a distinct mechanism. The catalytic mechanism and the spatial arrangement of the catalytic residues in PPiase families I and II are thought to have emerged through convergent evolution.[8] For family I PPiases, a mechanism was proposed in which the PP_i substrate is activated for hydrolysis by three separate divalent metal ions in the enzyme catalytic site (Figure A.1).[11] Two of these ions (M_1 and M_2) are thought to be involved not only in positioning and activating the substrate, but also in increasing the third divalent ion (M_3) coordinates the PP_i and active site residues, thereby positioning the PP_i for catalysis. In addition to the coordination by the

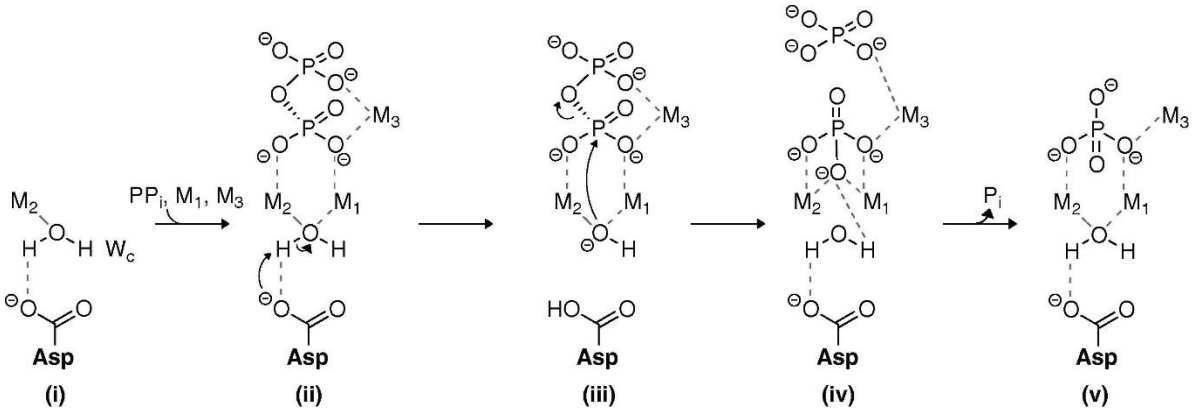


Figure A.1. Previously proposed PPiase mechanism. A catalytic mechanism proposed previously for family I PPiases. The active site already containing the catalytic water (W_c) and divalent metal ion M_2 (i) captures the PP_i substrate along with divalent metal ions M_1 and M_3 (ii). The nucleophilic hydroxide formed via proton abstraction by an Asp residue (Asp67 in *E. coli* PPiase, Asp54 in *Mtb* PPiase) attacks the activated electrophilic phosphate group of the PP_i (iii). Upon hydrolysis (iv), a new catalytic water rebinds and the other phosphate group dissociates, yielding complex (v), which then releases the attacked P_i.

nucleophilicity of a putative water molecule, W_c for the attack on the PP_i. The divalent metals, the PP_i substrate is stabilized by intermolecular salt bridges and hydrogen bonds with protein residues. The catalytic water is thought to be activated both by the bridging metal ions and, through proton abstraction, by one of the nearby aspartate residues (Asp67 in *E. coli* PPiase[11] which corresponds to Asp54 in *Mtb* PPiase), resulting in conversion of the water into a nucleophilic hydroxide, which then hydrolyzes the phosphodiester bond linking the two phosphate groups of the PP_i. The central role of the catalytic water in the mechanism was postulated based on site-directed mutagenesis studies with *E. coli* PPiase,[12] and the location of

this water was inferred to coincide with that of an inhibitory fluoride ion observed in crystal structures of *E. coli* and *Saccharomyces cerevisiae* PPiases in complex with PP_i.[11, 13] Asp67 of *E. coli* PPiase as the catalytic base was proposed based on the direct interaction of this residue with the fluoride ion observed in the crystal structure of *E. coli* PPiase in complex with PP_i, fluoride and Mg^{2+} (Protein Data Bank accession code: 2AUU).

Despite this proposed role of Asp67 of *E. coli* PPiase, mutations of this residue to a Glu resulted in a modest 5-fold loss in k_{cat} with a compensatory 5-fold gain in affinity.[12] In contrast, mutating other metal ion coordinating active site Asp residues in the same study had a more pronounced effect: a mutation of Asp70 (Asp57 in *Mtb* PPiase) to a Glu decreased k_{cat} more than 2000-fold, a mutation of Asp102 (Asp89 in *Mtb* PPiase) decreased k_{cat} ~7-fold. Another study probed the function of these Asp residues by mutating them to a less sterically perturbing Asn. For Asp67Asn, Asp70Asn and Asp102Asn, the effect was a ~26-fold, a 150-fold and a 90-fold decrease in k_{cat}, respectively.[14] Such modest effects of mutating Asp67 and comparable effects of other mutations suggest that there are back-up water activation mechanisms, which involve other Asp residues. Moreover, due to the involvement of Asp residues in the network of divalent metal coordination, which is critical to both substrate binding and catalysis, mutations of these residues perturb this coordination, complicating interpretation of the mechanistic role of these residues in catalysis. Therefore, the mutagenesis

studies leave the precise mechanism of the water activation unresolved, whereas structural studies alone are indirect.

E. coli,[15] *Mycobacterium tuberculosis* (*Mtb*),[16] as well as *S. cerevisiae*[17] contain family I PPiases, and these enzymes are essential for survival of these organisms. Because of the global clinical significance of *Mtb* as a deadly pathogen, *Mtb* PPiase is considered to be a potential target for novel anti-tuberculosis drugs.[18, 19] In addition to extensive structural and biochemical studies on *E. coli* and *S. cerevisiae* PPiases,[8, 11, 13, 20-22] X-ray crystallographic studies were also carried out with the *Mtb* PPiase,[23, 24] providing structural information on the location of a sole phosphate ion in the active site (PDB ID: 1WCF). The position of this P_i, presumably mimicking the second P_i product to dissociate, agrees with a position of the sulfate ion bound to the *E. coli* PPiase (PDB ID: 1JFD[25]). On the other hand, the opposite order of dissociation of the two P_i products (Figure A.1) was proposed based on another crystal structure of *E. coli* PPiase, obtained by soaking inhibitory fluoride out of the crystals of the complex of this enzyme with PP_i and fluoride (PDB ID: 2AU8[11]). Therefore, the order of product dissociation events in the catalytic turnover of bacterial PPiases remains to be clarified. In addition, the active site of *Mtb* PPiase displays unique features,[23, 24] such as two His residues that are not present in *E. coli* and whose role in catalysis is not clear. These His residues appear to be too far to interact with the substrate directly, but they can help orient the residues

that are directly involved in catalysis. A double mutant of the His residues reduced k_{cat} 5-fold, demonstrating a modest effect.[24]

In this study, we focused on obtaining more complete structural information on *Mtb* PPiase, to elucidate mechanistic details for *Mtb* PPiase and clarify our understanding of bacterial PPiases in general. We determined a set of crystal structures of *Mtb* PPiase bound to divalent metal ions and in complexes with its catalytic substrate, PP_i, and product, P_i. These structures have been employed to investigate details of the hydrolysis step catalyzed by *Mtb* PPiase by QM/MM simulations.

Experimental Procedures

Protein mutagenesis and purification. Mutations Asp54Asn, Asp57Asn and Asp89Asn were introduced by using vector pJexpress411-mtPPiase encoding a hexahistidine-tagged *Mtb* PPiase (a gift of Dr. Luiz Pedro de Carvalho) as a template and performing PCR with overlapping mutagenic primers by using Phusion DNA polymerase (New England Biolabs). The primer pairs were as follows: 5′-GGTGACGATGGCAACCCGCTG GACG-3′ and 5′-CGTCCAGCGGGTTGCCATCGTCACC-3′ for D54N, 5′-GCG ACCCGCTGAACGCGCTGGTG-3′ and 5′-CACCAGCGCGTTCAGCGGGTCGC-3′, for D57N, 5′-CACGGCGGCAACGACAAAGTGCTGTGC-3′ and 5′-GCACAGC

ACTTTGTCGTTGCCGCCGTG-3′ for D89N. The PCR reaction was followed by digestion with DpnI according to the QuikChange kit protocol (Agilent Technologies), and a transformation into chemically competent TOP10 cells (Life Technologies). All DNA sequences were verified by DNA sequencing (Eurofins Genomics).

Hexahistidine-tagged wild-type *Mtb* PPiase and its mutants Asp54Asn, Asp57Asn and Asp89Asn were expressed and purified by Ni^{2+} affinity chromatography with a 5 mL Ni-IMAC HisTrap FF column (GE Healthcare), as described previously.[18] The proteins were then dialyzed against dialysis buffer (40 mM Tris HCl pH 8.0, 100 mM NaCl and 2 mM β-mercaptoethanol). These analogously purified proteins were used for the activity assays. For crystallization, the wild-type protein was further purified on a size exclusion S-200 column (GE Healthcare) equilibrated in gel filtration buffer (40 mM Tris HCl pH 8.0, 100 mM NaCl). Fractions containing *Mtb* PPiase were pooled and concentrated using an Amicon Ultra-15 centrifugal filter device (Millipore) to 750 μM and stored at 4 °C.

Pyrophosphatase activity assay. The measurements of PPi hydrolytic activity of *Mtb* PPiase and its mutants by following a protocol that we recently reported,[6, 18] with minor modifications, as described below in brief. All reactions were carried out in 20 mM Tris HCl pH 7.5 and 1 mM $MgCl_2$. The reactions were stopped after 10 min by adding 150 μL of the

malachite green reagent to a 30 µL reaction. After 2 min, 15 µL of 34% sodium citrate was added for color stability. The color was allowed to develop for 30 min, then the mixture was transferred to a 96-well clear flat-bottom plate and the absorbance at 620 nm was measured with a SpectraMax M5 plate reader. We first established that 2 nM of the wild-type *Mtb* PPiase and *Mtb* PPiase Asp54Asn was an appropriate concentration for a steady-state kinetic measurement of reaction rate (V) with 200 µM PP_i, as this concentration corresponded to the linear range of product accumulation over time. The majority of substrate under these conditions was not hydrolyzed. Similarly, 800 nM of the weakly active mutants Asp57Asn and Asp89Asn yielded steady-state measurements for a 10 min reaction. We then used these enzyme concentrations and the 10 min reaction time to perform measurements of absorbance as a function of PP_i concentration, as specified in the text. All reactions were performed in triplicate.

The values of steady-state kinetic parameters K_m and $V_{max,app}$ were obtained by non-linear regression analysis of absorbances (A) as a function of [PP_i] by using SigmaPlot (SysStat) with the Michaelis–Menten equation $A=A_0+V_{max,app}[PP_i]/(K_m+[PP_i])$, where A_0 is the absorbance in the absence of PP_i. The values of k_{cat} were obtained by dividing $V_{max,app}$ by the enzyme concentration, reaction time and the unit conversion constant from absorbance units to M of PP_i hydrolyzed. The last constant was determined

by calibration measurements of absorbance with increasing concentrations of P_i in the reaction buffer, without enzyme and PP_i.

Crystallization and crystal cryoprotection. All complexes were crystallized by vapor diffusion in hanging drops at 22°C. The drops contained a mixture of 1µL of the reservoir solution and 1µL of concentrated protein (~14mg/mL) equilibrated against the reservoir solution. The crystals of *Mtb* PPiase in complex with PP_i were obtained with the reservoir solution composed of 1.65M $NaKHPO_4$, 100mM HEPES pH 7.75 and 2mM $CaCl_2$, the crystals of *Mtb* PPiase in complex with two P_i ions were obtained with the reservoir solution composed of 1.6M KH_2PO_4, 100mM HEPES pH 7.75 and 2 mM $CaCl_2$, and the crystals of *Mtb* PPiase in complex with one P_i were obtained with the reservoir solution containing 1.57M $NaKHPO_4$, 100mM HEPES pH 7.75 and 2mM $MnCl_2$. In order to obtain crystals of *Mtb* PPiase bound to different divalent metal ions, we first grew crystals by using reservoir solution containing 9% PEG 4000, 100mM HEPES pH 7.0 and 100mM $CaCl_2$ and then soaked these crystals with the reservoir solution, in which Ca^{2+} was substituted with Mg^{2+} or other divalent metal ions. Single crystals were obtained in each condition in 1–5 days. Crystals were gradually transferred over 30 min from their mother liquors to the reservoir solutions containing 20% v/v glycerol and then frozen by quick immersion in liquid nitrogen.

Data collection, and structure determination. X-ray diffraction data for crystals of *Mtb* PPiase–PP_i and *Mtb* PPiase–$2P_i$ complexes were collected at beamline 22-ID-D and all the other data were collected at beamline 21-ID-F of the Advanced Photon Source of the Argonne National Laboratory at 100 K and processed using program HKL2000.[26] The structure of *Mtb* PPiase bound to Mg^{2+} was determined by molecular replacement by using the previously reported structure of apo-*Mtb* PPiase (PDB ID: 1WCF[23]) as a search model, with program Phaser.[27] The other structures were determined by molecular replacement with the structure of the *Mtb* PPiase–Mg^{2+} complex as a search model. The structures were refined and rebuilt iteratively by using programs REFMAC[28] and COOT,[29] respectively. The data collection and refinement statistics are provided in Table A.1. The crystal structure coordinates and structure factors were deposited in the Protein Data Bank with the PDB accession numbers: 4Z70, 4Z71, 4Z72, 4Z73, 4Z74, as specified in Table A.1.

QM/MM simulations. The crystal structure of *Mtb* PPiase bound to PP_i was used for all simulations. For all the QM/MM simulations one hexamer of *Mtb* PPiase was considered. One of the six active sites of the hexamer was treated as a QM subsystem, starting from the experimentally defined locations of the substrate, cations and side chains. The Ca^{2+} ions in the crystal structure were replaced by Mg^{2+} ions. The system was hydrogenated

Data collection	Mtb PPiase–PPi	Mtb PPiase–2Pi	Mtb PPiase–Pi	Mtb PPiase–Ca^{2+}	Mtb PPiase–Mg^{2+}
PDB ID	4Z74	4Z72	4Z73	4Z70	4Z71
Space group	$P2_12_12_1$	$P63_22$	$P2_12_12_1$	$P3_221$	$P3_221$
Monomers (a.u.)	12	1	12	3	3
Unit cell dimensions: a,b,c (Å) α, β, γ (°)	87.0, 105.2, 255.9 90, 90, 90	99.3, 99.3, 96.7 90, 90, 120	85.6, 105.1, 254.7 90, 90, 90	102.1, 102.1, 80.4 90, 90, 120	101.7, 101.7, 81.2 90, 90, 120
Resolution (Å)	50–2.55 (2.59–2.55)[a]	50–2.35 (2.43–2.55)[a]	50–3.3 (3.36–3.30)[a]	50–1.95 (1.98–1.95)[a]	50–1.85 (1.88–1.85)[a]
Rmerge (%)	8.8 (57.9)	12.0 (78.3)	15.7 (54.2)	6.7 (52.6)	7.4 (57.9)
I/σI	28.5 (4.2)	18.1 (2.3)	14.8 (3.9)	44.2 (5.3)	36.8 (4.4)
Completeness (%)	100.0 (100.0)	99.3 (98.9)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Redundancy	7.5 (7.6)	7.1 (7.4)	6.9 (7.0)	12.4 (12.5)	11.1 (11.1)
Refinement					
Resolution (Å)	35–2.55 (2.62–2.55)[a]	40–2.35 (2.41–2.35)[a]	34–3.30 (3.4–3.30)[a]	32–1.95 (2.00–1.95)[a]	39–1.85 (1.90–1.85)[a]
Number of reflections	73,541	11,538	33,599	33,886	39,587
R (%)	20.8 (28.4)	20.8 (29.8)	21.9 (28.9)	20.9 (25.2)	17.4 (20.0)
Rfree (%)	24.9 (34.8)	24.1 (31.1)	25.9 (32.9)	25.4 (27.6)	20.8 (24.0)
Number of atoms/a.u.	16,428	1321	15,551	4627	4500
r.m.s. deviations: Bond length (Å) Bond angle (°)	0.007 1.185	0.007 1.492	0.004 0.980	0.007 1.324	0.004 0.990
Ramachandran plot[b]					
Most favored Additionally allowed: Disallowed:	93.8% 6.2% 0.0%	93.3% 6.7% 0.0%	98.15% 1.85% 0.0%	96.62% 3.38% 0.0%	97.48% 2.52% 0.0%

[a] Values in parentheses refer to the highest resolution shell
[b] PROCHECK output

Table A.1. Crystallographic data of MtbPPiase. Data collection and refinement statistics

using MolProbity[30, 31] and electro-neutralized by adding 51 K^+ ions using the tleap program of AMBER12 package[32]. The subsequent MD simulations with this neutralized system were unstable, as characterized by large fluctuations of density and energies. Therefore, we instead both neutralized the system

and increased its ionic strength to 100 mM by adding 102 K^+ ions and 51 Cl^- ions using tleap function in AMBER12. The system was solvated with TIP3P water[33] with a 15 Å buffer in a final box size of approximately 150 Å×140 Å×147 Å. The total number of atoms in the system used in the MD simulations was 50,396.

The MD simulations were performed using the PMEMD program in the AMBER12 software suite with the ff99SB force field.[34] The SHAKE algorithm was applied for bonds between heavy atoms and hydrogen.[35] All MD simulations were carried out under periodic boundary conditions by the smooth particle mesh Ewald (sPME) method for long-range electrostatic interactions.[36, 37] Systems were minimized using sander function with 50 steps of steepest descent and 450 steps of conjugate gradient with a restraint of 500 kcal mol^{-1} $Å^{-2}$ on the protein atoms. Densities of the systems were equilibrated to 1 g cm^{-3} in the NPT ensemble with pressure being held constant by anisotropic pressure scaling. The systems were gradually heated to 300 K under constant volume conditions with 500 kcal mol^{-1} $Å^{-2}$ restraints on the protein atoms that were gradually removed. Temperature was kept constant by using the Langevin thermostat.[38] Non-restrained production MD was performed for 20 ns for each system with snapshots saved at each 1 ps. Twenty random snapshots were selected from the MD trajectory and the PPiase active sites were observed for waters that could be considered as possible candidates to carry out a nucleophilic attack

on the PP_i. The selection criteria for a candidate water included the distance between the oxygen atom of the candidate water and the closer phosphorus atom of the PP_i as well as the angle of attack formed by these two atoms and the bridging oxygen of the PP_i. Upon selection of such a water, ten snapshots for which the angle of the attack was closest to 180° were selected from the trajectory for QM/MM calculations.

All atoms in the PPiase hexamer, and water molecules that were within 48 Å of the attacked phosphorus atom of PP_i in the selected active site, were considered in the QM/MM calculations. The QM subsystem was chosen to include the side-chains of Asp57, Asp84 and Asp89, PP_i, three Mg^{2+} ions that correspond to M_1, M_2, M_3, and 13 water molecules including the attacking water. All atoms where a covalent bond crosses the QM/MM boundary were treated using the pseudobond approach,[39, 40] yielding a smooth connection at the QM/MM interface. All the QM/MM calculations were performed by means of an in-house program that links modified versions of Gaussian09[41] and TINKER[42] to perform additive QM/MM with electrostatic embedding. An iterative method was used in the QM/MM optimizations and were performed at the B3LYP/631G(d,p) level of theory.[43]

The reaction path was calculated by using the quadratic string method (QSM).[44] The path was created using 18 images, including the optimized reactant and the product structures at each end. The MM subsystem of the

optimized reactant was used as the starting MM subsystem for each of the intermediate structures between the optimized reactant and the product. These starting MM subsystems of the 16 intermediates were then subjected to constrained optimization.[45] The constraints were gradually reduced after each QM/MM optimization cycle. As before, the individual QM/MM optimizations for each of the 16 structures were performed using the same in-house program that links modified versions of Gaussian09 and TINKER. The structure corresponding to the highest energy point in the optimized path calculated using QSM was used as the starting structure for the transition state optimization. Transition state optimization was performed using the QST3 method. No negative frequencies were present in any minima and the TS exhibited only one imaginary frequency corresponding to the vibration connecting the reactant and product valleys.

Results

Structures of *Mtb* PPiase in complex with PP_i and with two P_i ions. A structure of *Mtb* PPiase–PP_i complex was determined at the resolution of 2.55 Å from crystals of *Mtb* PPiase grown at high concentration of sodium/potassium HPO_4^{2-} salt in the presence of Ca^{2+} (Figure A.2). In this

crystal form, there were two *Mtb* PPiase hexamers per asymmetric unit. Surprisingly, a PP_i molecule, instead of the expected P_i, was bound to all 12

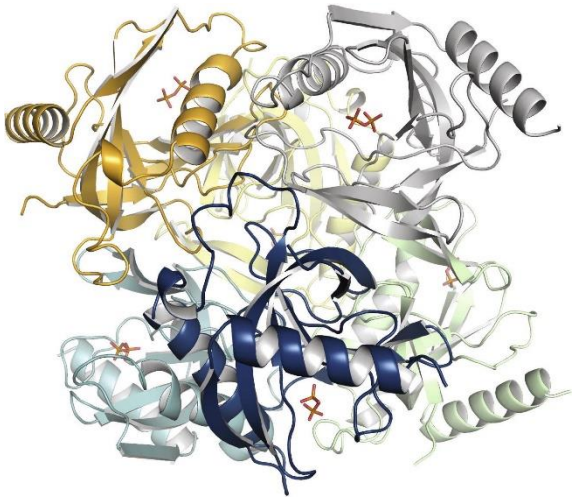


Figure A.2. Mtb PPiase-PPi structure. A cartoon view of an *Mtb* PPiase hexamer from the crystal structure of *Mtb* PPiase–PPi complex. The six monomers are colored differently, and the bound PP_i molecules are shown by orange sticks. The representative monomer colored gray is shown in subsequent figures in a similar orientation.

active sites in the two *Mtb* PPiase hexamers (Figure A.3A). The PP_i was generated from two molecules of P_i by the reaction catalyzed by *Mtb* PPiase in the direction of synthesis, driven by a high P_i concentration. The location and the conformation of the PP_i in all 12 active sites were nearly identical, whereas the location and the occupancy of a Ca^{2+} at position M_1 (most commonly coordinating both Asp57 and Asp89) as well as the conformation of the side chain of Asp89 and, to a lesser extent of Asp57, were somewhat variable among the different active sites. Several stably bound water molecules were well resolved in all active sites in the vicinity of Asp57,

Asp89 and the PP_i. In some active sites, water molecules were found to be directly coordinated to M_1 or M_2 but not to both ions at the same time. Such

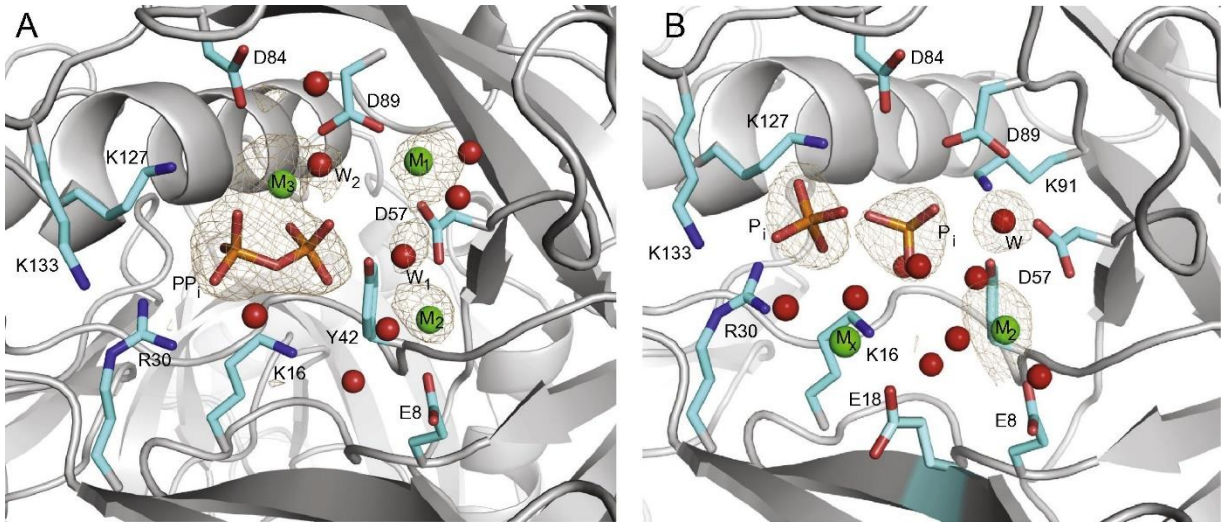


Figure A.3. Mtb PPiase active sight with substrate or product bound. The active site of *Mtb* PPiase in the structures of its complexes with PP_i, $2P_i$ and Ca^{2+}. (A) The complex of *Mtb* PPiase with PP_i. The mesh represents an omit $F_o–F_c$ electron density map contoured at 3.0σ that was generated by using a model devoid of the PP_i, metal ions M_1, M_2 and M_3 and the water molecules (discussed in the text). (B) The complex of *Mtb* PPiase with $2P_i$. The mesh represents an omit $F_o–F_c$ electron density map contoured at 4.0σ, generated without the two P_i, metal ion M_2 and the water molecule. The second Ca^{2+} ion (labeled M_x) coordinated to Glu18 and two water molecules was observed in this structure at a site unrelated to M_1, M_2 or M_3. In both panels, the residues that interact with the substrate, products (orange sticks) or metal ions (green spheres) are shown as cyan sticks. Water molecules are shown as red spheres.

bidentate coordination or even hydrogen bonding by a water molecule would not be possible due to geometrical constraints ($M_1–M_2$ distance of 6.4 Å, with the intervening side chain of Asp57). An example of an active site (which belongs to monomer A in the coordinate file; PDB ID: 4Z74) is shown in

Figure A.3A, where a water molecule (W_1) forms hydrogen bonds with Asp57, an adjacent phosphate group of the PP_i and with other water molecules coordinated to M_1 and M_2 ions. This water molecule superposes nearly

perfectly with the location predicted to be occupied by the catalytic hydroxide ion in structure-based mechanistic studies of *E. coli* PPiase, i.e., matching the site of a fluoride and a water in the respective structures of *E. coli* PPiase structures with bound PP_i (PDB IDs: 2AUU and 1I6T)[11] (Figures A.1 and A.4C). Another water molecule (designated as W_2 in Figure A.3A) was observed coordinated to metal ion M_3, at a hydrogen bonding distance from the PP_i and in the vicinity of Asp89.

When KH_2PO_4 was used as a precipitant in the presence of Ca^{2+}, we obtained a different crystal form of *Mtb* PPiase, with only one PPiase monomer per asymmetric unit. The *Mtb* PPiase in this crystal form contained two P_i ions bound to the active site, mimicking the products of pyrophosphorolysis (Figure A.3B). The P–P distance between the two P_i ions in this complex is 4.8 Å, whereas the P–P distance in a PP_i molecule is 2.86 Å, clearly indicating the absence of covalent connectivity between the two P atoms. This structure thus appears to correspond to a catalytic state immediately after the PP_i cleavage, but before dissociation of either P_i. The P_i that is adjacent to Asp57/Asp89 is at approximately the same location as the

respective phosphate group of the PP_i in the previous structure. The other P_i is shifted by about 2 Å from the position of the respective phosphate group of the PP_i towards Lys133, and it forms a hydrogen bond with this residue. This P_i interacts both electrostatically and via hydrogen bonds with Arg30, Lys127 and Lys133, while maintaining an electrostatic interaction with Lys16. This disposition of the two bound P_i molecules is similar to that observed in the active site of *S. cerevisiae* PPiase,[46] supporting the notion that this structure is a mimic of the first product complex after PP_i hydrolysis.

As expected, metal ion M_3, which was chelated by the two phosphate groups of the PP_i substrate in the previous structure, was not observed in this *Mtb* PPiase–$2P_i$ complex (Figure A.3B). M_2 ion remains bound in this structure, but is shifted by 2.5 Å from the location in the previous structure and is no longer directly coordinated directly to Glu8 and Asp57, but still maintains salt bridges with these residues. In this complex, M_1 ion is also no longer present, but a water molecule is found in the vicinity, with strong hydrogen bonds to Asp57, Asp89 and to the closest O atom of the adjacent P_i (the O–O distances of 2.6–2.8 Å). Both the backbone and the side chain of Asp89 are somewhat shifted towards Asp57 and the P_i from the conformation of Asp89 in the *Mtb* PPiase–PP_i complex.

PPiases from *Mtb* and *E. coli* are very highly homologous in sequence (46% sequence identity and 62% sequence homology) and in structure

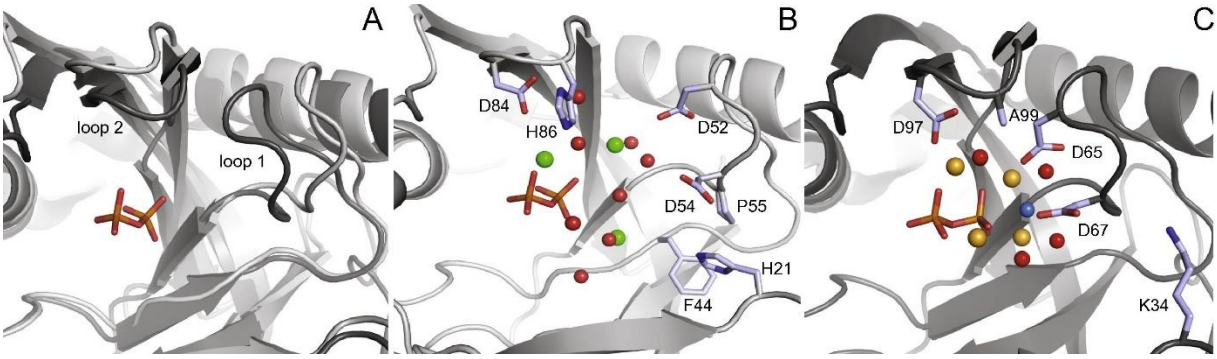


Figure A.4. Mtb PPiase and E. coli PPiase comparison. Comparison of the active sites of *Mtb* PPiase (PDB ID: 4Z74; this study) and *E. coli* PPiase (PDB ID: 2AUU) when bound to PP_i. The ligands are highlighted as in Figure A.1. (A) Conformational differences between the backbones *Mtb* PPiase (light gray) and *E. coli* PPiase (dark gray). Loops 1 (residues 52–59 of *Mtb* PPiase) and 2 (residues 83–90) are labeled. (B) The active site of *Mtb* PPiase. (C) The active site of *E. coli* PPiase. A bound fluoride is shown as a blue sphere. The residues whose identity or disposition with respect to the PP_i differ from those of *E. coli* are shown explicitly in panels (B) and (C).

(RMSD of Cα–Cα distances is 1.0 Å), and all the active site residues that contact PP_i or divalent metal ions are the same in the two enzymes. While the active site of *Mtb* PPiase and the bound PP_i display a high degree of similarity to that observed in the crystal structure of *E. coli* PPiase in complex with PP_i, fluoride and Mg^{2+} (PDB ID: 2AUU), there are significant conformational differences in two loops (loop 1: residues 52–59 and loop 2: residues 83–90), which contain some of the residues that were previously

proposed to play key catalytic roles for the *E. coli* enzyme (Figure A.4A). These differences are not due to crystal packing interactions, as these regions of the *Mtb* PPiase are not involved in crystal packing, nor are they due to a different divalent metal ion, as another crystal structure of *Mtb* PPiase with bound PP_i and the catalytic Mn^{2+} (described in the next section) is nearly perfectly superimposable onto this structure. Moreover, the conformations of these loops in the crystal structure of *E. coli* PPiase in complex with PP_i, fluoride and Ca^{2+} (PDB ID: 1I6T) are nearly identical to those in the structure with Mg^{2+}. Major differences are between the backbone conformations of these loops; the side chain rotamers of the conserved residues of the two PPiases are nearly identical to each other. These dissimilarities are also reflected in the geometry of coordination of the metal ions and hydrogen bonds with waters by the key metal Asp residues residing in these loops (Figure A.4B and C). Most prominently, Asp54 and Asp52 are shifted away from the PP_i by 3.1 Å and 4.7 Å from its *E. coli* counterparts, Asp67 and Asp65, respectively. As a consequence, Asp54 does not form a hydrogen bond with the positionally conserved water W_1, which was proposed to be activated by Asp67 of *E. coli* by proton abstraction. Other differences include the lack of M_1 and M_3 coordination by Asp52 and Asp84, respectively. This geometry makes the mechanism of water activation by Asp54 unlikely. Remaining candidate residues for proton

abstraction from a catalytic water, based on proximity of such residues to the PP_i, are Asp57 and Asp84.

These conformational differences may result from evolutionary divergent sequences of *Mtb* and *E. coli* PPiases. For example, His21, one of the two His unique to *Mtb*,[24] is wedged between Asp54 and Phe44, forming a salt bridge with Asp54, a hydrogen bond with the carbonyl oxygen of Val19 stacking perpendicularly with Phe44 and sterically interacting with Pro55. Thus, His21 appears to keep loop 1 away from the PP_i. In contrast, Lys34 of *E. coli* points out of the protein interior into the solvent. In loop 2, the other unique His, His86, corresponds to the much smaller Ala99 of *E. coli.* The Cα of His86 is shifted by 2.7 Å from the respective Cα of Ala99 away from the PP_i, likely to preserve metal ion coordination by Asp residues, and, as a result, loop 2 is also shifted. These significant conformational differences imply potential mechanistic differences between the *Mtb* and the *E. coli* PPiases.

Structures of *Mtb* PPiase in complex with one P_i and apo-*Mtb* PPiase. When crystals of *Mtb* PPiase were grown at high HPO_4^{2-} concentration in the presence of Mn^{2+}, the crystals again contained two hexamers of the enzyme per asymmetric unit. The structure, determined at a resolution of 3.3 Å revealed that active sites of 9 of the 12 monomers contained one P_i ion and the other 3 sites contained a PP_i molecule. Because

of the lower resolution of this structure, we were able to distinguish only the PP_i/P_i ligands and Mn^{2+} ions. The locations of the Mn^{2+} ions were also confirmed by the anomalous difference Fourier map. Bound water was not resolvable. The PP_i was bound similarly to its position in the structure described in the previous section, indicating that the differences between the

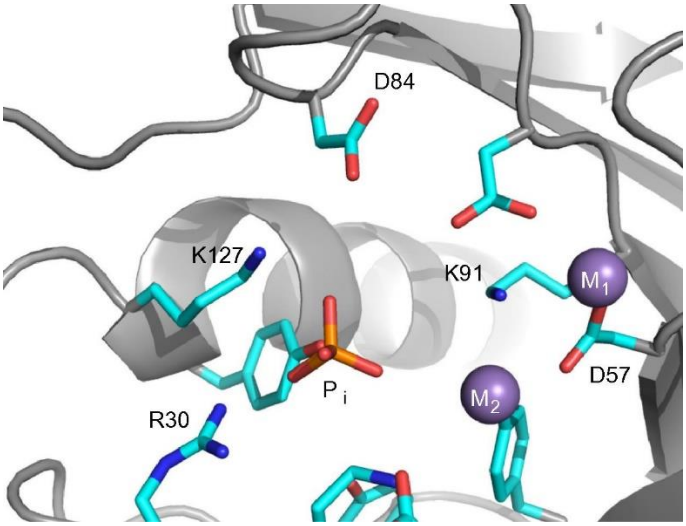


Figure A.5. Mtb PPiase bound with Pi and Mn2+. The active site of *Mtb* PPiase in the structure of its complex with P_i and Mn^{2+}. The active site contains a P_i molecule bound at a location in a positively charged cleft that overlaps with the phosphate group of the PP_i that is not directly attacked. Metal ion M_1 occupies a similar position in both the P_i and the PP_i bound structures; however, metal ion M_2 has shifted closer to the bound P_i. No other metal ion is observed in this structure.

structures of *Mtb* and *E. coli* PPiase are not due to different divalent metal ions. The location of the bound P_i was almost exactly in the middle of the two P_i ions seen in the complex of *Mtb* PPiase with two P_i ions. This P_i position closely corresponds to the location of the P_i group of the bound PP_i

that is not directly attacked. In this structure, the P_i still interacts with Arg30, Lys127, and Lys16, but not with Lys133 (Figure A.5). This structure is similar to that reported previously in the absence of divalent metal ions (PDB ID: 1WCF).[23] Notably, the location of this remaining P_i is also consistent with that seen in structures of *S. cerevisiae* in an elegant crystallographic study by Oksanen.[46] Therefore, this structure is likely a

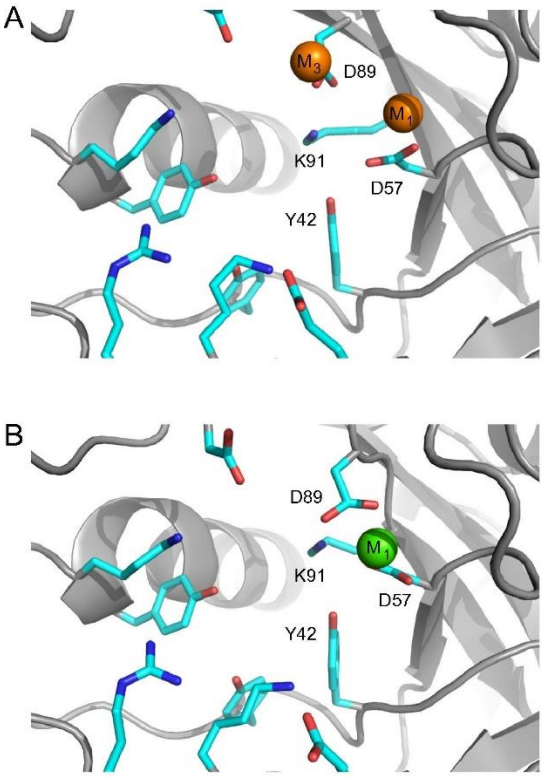


Figure A.6. Metal binding positions of Mtb PPiase. (A) The active site of *Mtb* PPiase in the structure of its complex with Mg^{2+}. The active site contains metal ions in the M_1 and M_3 positions. (B) The active site of *Mtb* PPiase in the structure of its complex with Ca^{2+}. The sole bound metal ion is located at the M_1 position.

mimic of a subsequent mechanistic intermediate, formed after one of the two P_i products has dissociated. Taken together, the above structures of the *Mtb* PPiase–PP_i/$2P_i$/P_i complexes strongly suggest that the directly attacked phosphate group of the PP_i is first to dissociate as a P_i, since the remaining P_i maintains most of its electrostatic interactions in this mechanism, as described above. Site M_1 is occupied (by Mn^{2+}), and the other bound metal ion is shifted from the previously observed site M_2 by 3.2 Å towards the P_i.

Crystal structures of *Mtb* PPiase were also determined in the presence of only divalent metal ions, without the substrate or the product (PP_i or P_i), to capture the state(s) of the enzyme prior to binding of the PP_i substrate or after the dissociation of the reaction products. Specifically, structures of apo-*Mtb* PPiase bound to Mg^{2+} ion (Figure A.6A) and to Ca^{2+} ion (Figure A.6B) were determined at resolutions of 1.85 Å and 1.95 Å, respectively. In the complex with Mg^{2+}, the active site contains two bound Mg^{2+} ions. One cation is bound at the M_1 site, with the same adjacent water molecule as in the structures of complexes with PP_i and P_i. The other Mg^{2+} is bound to the protein exclusively through water mediated interactions, at a site shifted from the original M_3 site by 3.8 Å towards Asp89. The structure of the *Mtb* PPiase in complex with Ca^{2+} contains only one metal ion in the active site near the M_1 site, where the Ca^{2+} is coordinated to Asp57 and Asp52 instead of Asp89 due to a minor positional shift. Minor differences were observed in the conformations of the side chains coordinating the divalent metal ions,

which may be due to different sizes or coordination propensities of Mg^{2+} and Ca^{2+}. The remaining metal ion, M_2, is observed only together with bound PP_i, $2P_i$ or P_i and, therefore, this metal ion and PP_i likely bind the enzyme in a cooperative fashion.

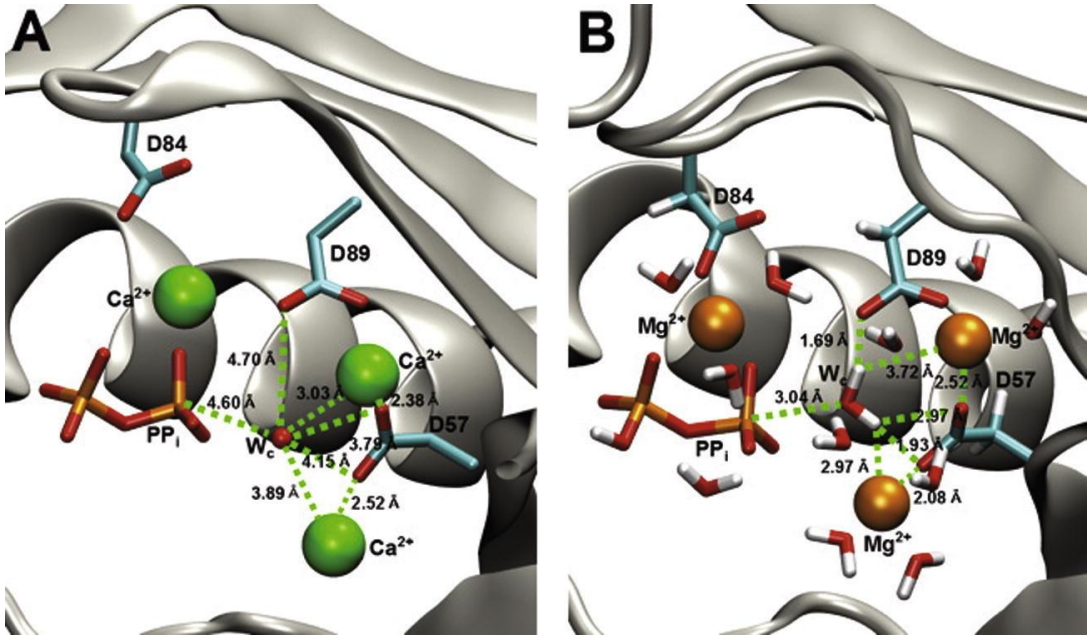


Figure A.7. Optimized reactant structure comparison. Comparison of the immediate active site between the crystal structure (A) and the optimized reactant structure (B). The catalytic water is labeled as W_c.

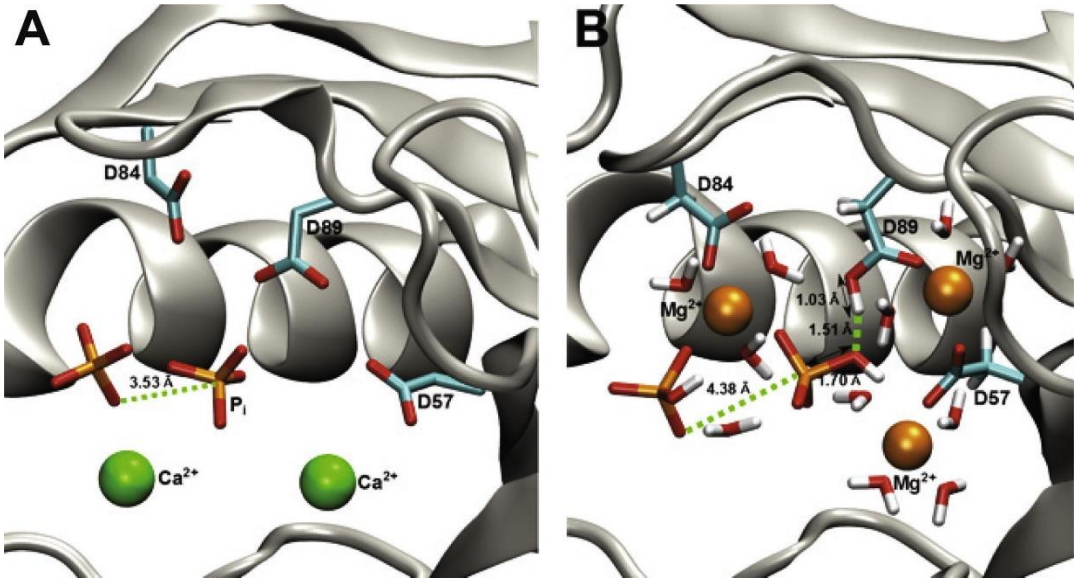


Figure A.8. Comparison of optimized product structure. Comparison of the immediate active site between the crystal structure (A) and the optimized product (B) structure. The phosphoanhydride bond of the PP_i has been broken by W_c.

QM/MM calculations of the central catalytic step. As the mechanism of water activation by *Mtb* PPiase appears to be distinct from the *E. coli* homolog, to characterize this process for *Mtb* PPiase, we carried out QM/MM simulations. At the initial stage, the QM/MM simulations involved optimization of the reactant and product structures. The reactant structure was optimized starting from the structure of *Mtb* PPiase–PP_i complex (Figure A.7A). The final optimized reactant structure (Figure A.7B) had an overall backbone RMSD deviation of 0.9 Å and an RMSD for all atoms in the active site of 0.6 Å with respect to the crystal structure of the *Mtb* PPiase–PP_i complex. As an internal consistency check, the product structure was

modeled *in silico* from the optimized reactant structure and subjected to QM/MM optimization.

The initial calculations were performed with the PP_i completely de-protonated (−3 charge), which did not yield an optimized product structure starting from the optimized reactant. Subsequently, a proton was added to one of the oxygen atoms on the phosphate group that was not directly attacked (Figure A.7B), which enabled the optimization of both the reactant and product structures. The optimized product structure (Figure A.7B) had an overall backbone RMSD of 0.9 Å and an all-atom RMSD for the active site of 0.8 Å with respect to the crystal structure of the *Mtb* PPiase–$2P_i$ complex (Figure A.8A). Thus, neither optimized structures were significantly perturbed relatively to the reference crystal structures. Moreover, the low RMS deviation of the optimized product structure with respect to the *Mtb* PPiase–$2P_i$ crystal structure underscores the applicability of our method. The overall calculated reaction energy for this process was −11.24 kcal/mol.

Upon replacing the three Ca^{2+} atoms in the active site with Mg^{2+} and subjecting the structure to MD and QM/MM optimization, a shift in the active site is observed. In the optimized reactant structure (Figure A.7B), the change in cations results in a tighter active site, as an apparent result of the shorter coordination bonds between the carboxylate oxygens of the aspartates and the metal ions. For example, the distance between Asp57

and M_1 changes from 2.52 Å to 2.08 Å. This tightening promotes a rearrangement of the active site, which results in the loss of coordination between Asp89 and M_3, while improving coordination of this residue with M_1. The catalytic water molecule is observed to travel starting from a site occupied by water W_1 in the *Mtb* PPiase–PP_i complex (Figure A.3A) in the reactant species and get activated (in the TS) at a site approximately halfway between W_1 and W_2 in that complex. In addition, the attacking water is observed to form hydrogen bonds to both Asp89 and Asp57. The hydrogen bond between the attacking water, W_c, and Asp89 is slightly stronger than to Asp57 as judged by the respective O–O distances (Figure A.7B). The optimized reactant and product structures were used as starting points for the Quadratic String Method (QSM) calculation.[44] Once the path

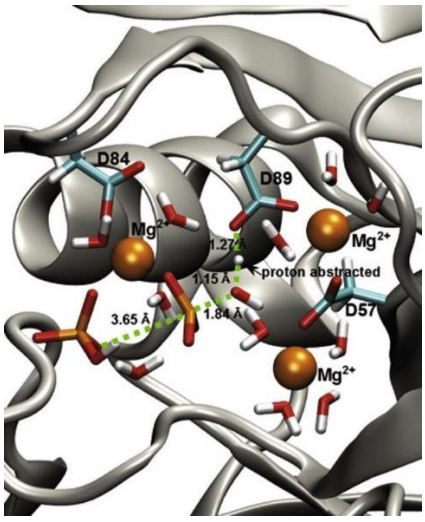


Figure A.9. Optimized transition state of Mtb PPiase, Optimized transition state corresponding to a concerted mechanism where the nucleophilic water is activated as it attacks the adjacent phosphate group of the PP_i.

was optimized, the highest energy point was selected and optimized to the transition state (TS) (Figure A.9) by a QST3 optimization, resulting in a final structure with one single imaginary frequency of 240.8*i* cm^{-1}. The final calculated barrier for the hydrolysis of PP_i by *Mtb* PPiase is 6.6 kcal/mol. There have been a number of theoretical studies on other phosphatases and phosphate hydrolysis.[47-52] However, to our knowledge, only one study has been reported on an inorganic pyrophosphatase, with *E. coli* PPiase.[53] The authors of this study reported a calculated barrier (for a cluster model with continuum solvation) of ≈ 9 kcal/mol. This barrier is similar to our calculated barrier of 6.6 kcal/mol, especially taking into account that the present results explicitly include the entire enzyme and solvent into the calculation.

These calculations indicate that the carboxylate oxygen of Asp89 forming the hydrogen bond with W_c abstracts a proton from this water, as this residue becomes both nucleophilic due to the loss of one of the two coordinating metal ions and due to its proper positioning with respect to W_c. In contrast, both oxygens of Asp57 remain coordinated to two divalent metal ions, apparently remaining insufficiently nucleophilic. Based on the minimum energy path, the nucleophilic water transfers a proton to Asp89 to yield a hydroxyl anion and almost concomitantly carries out the nucleophilic attack on the adjacent P atom of the PP_i.

To test directly the possibility for proton abstraction by Asp57, an initial structure was generated where the phosphodiester bond was broken and the proton from the nucleophilic water was transferred to the closes carboxyl oxygen of Asp57. This initial structure was subjected to the same optimization procedure as that used above. Interestingly, the final optimized structure involved the return of the proton from Asp57 to the water O and the transfer of the other proton to Asp89. This suggests, that the carboxyl oxygen on Asp89 that is not coordinated to Mg^{2+} is a better proton acceptor.

Effects of mutations of Asp54, Asp57 and Asp89 on PPi hydrolysis by *Mtb* PPiase. To test the mechanism suggested by the

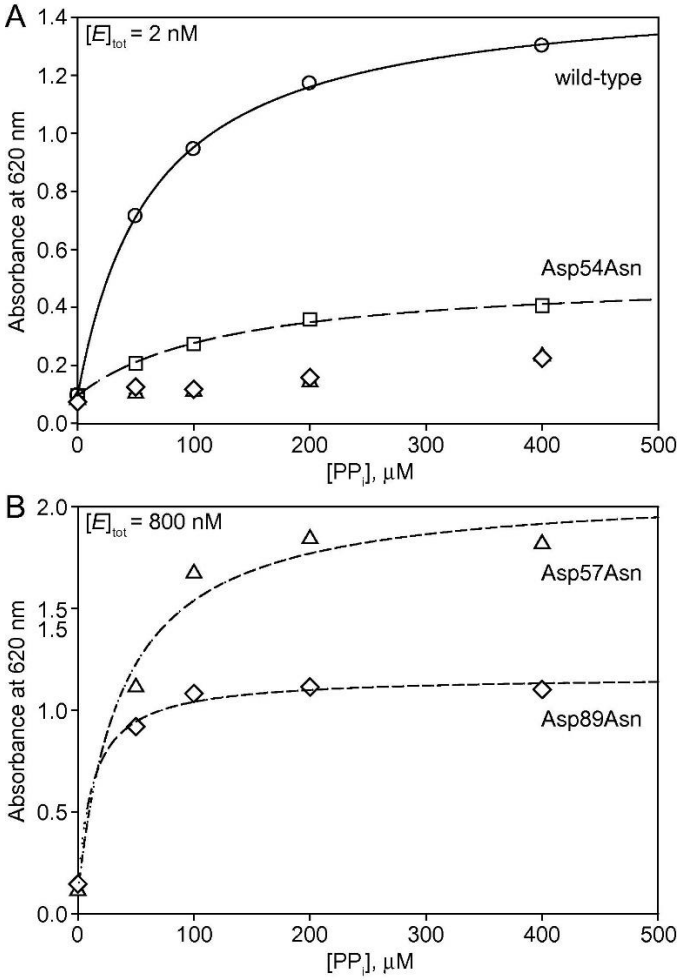


Figure A.10. Mtb PPiase and mutant kinetic curves. The analysis of the steady-state hydrolysis by wild-type *Mtb* PPiase (circles) and its three mutants: Asp54Asn (squares), Asp57Asn (triangles) and Asp89Asn (diamonds). (A) Activity measurements for reactions with 2 nM of each enzyme. The curves show the best fit to the Michaelis–Menten equation with K_m = (64 ± 3) μM and k_{cat} = (48 ± 1) s^{-1} for wild-type *Mtb* PPiase and K_m = (131 ± 19) μM and k_{cat} = (14 ± 1) s^{-1} for Asp54Asn mutant. The amount of product formed by the other two mutant enzymes, Asp57Asn and Asp89Asn was insignificant at these enzyme concentrations. (B) Activity measurements for reactions with 800 nM of Asp57Asn and Asp89Asn mutants. The best-fit curves correspond to K_m = (38 ± 17) μM and k_{cat} = (0.17 ± 0.02) s^{-1} for Asp57Asn and K_m = (14 ± 5) μM and k_{cat} = (0.087 ± 0.005) s^{-1} for Asp89Asn mutant. The standard deviations were invariably of the size of the data symbols or smaller and, therefore, the error bars were not shown.

structural and computational studies, we performed steady-state measurements of the pyrophosphatase activity by wild-type *Mtb* PPiase and its point mutants of three relevant Asp residues to an Asn: Asp54Asn, Asp57Asn and Asp89Asn. An Asn was chosen, since it is approximately isosteric with an Asp and is not a proton abstractor. The activity of the Asp54Asn is only modestly perturbed, as establish by steady-state kinetic measurements with the concentration of the enzymes of 2 nM (Figure A.10A): the value of k_{cat} for this mutant is 3-fold smaller and the value of K_m is 2-fold larger than those for the wild-type enzyme. The activity of the other two mutant enzymes, Asp57Asn and Asp89Asn, is insignificant at this concentration (Figure A.10A). These results, in agreement with our proposal, indicate that Asp54 does not play a major catalytic role, and another residue (*e.g.*, Asp89) acts as a base instead. The kinetics of the two weakly active mutants were measured quantitatively in assay where they were present at 800 nM each (Figure A.10B). While their K_m values are within 2- to 3-fold range from that of the wild-type enzyme, the k_{cat} values for Asp57Asn and Asp89Asn mutants are 300-fold and 600-fold smaller, respectively. Even though these observations are also consistent with our proposed mechanism, due to a possible perturbation of Mg^{2+} coordination in these mutants, their interpretation is not unambiguous.

Discussion

It was previously reported that family I PPiases could catalyze the reverse reaction to combine two molecules of P_i into a PP_i with a remarkable catalytic efficiency.[4] The use of P_i, a sole reactant of this reverse reaction at a high concentration (in the molar range) in the crystallization solution allowed us to capture *Mtb* PPiase with a PP_i bound in the active site in the crystalline form. We also obtained crystals of *Mtb* PPiase bound to reactants and products together with divalent metal ions and coordinated water molecules. While previous structural studies of *E. coli* and *S. cerevisiae* PPiases postulated the location of a nucleophilic water based on the observation of an inhibitory fluoride ion, considered to be a mimic of the hydroxide nucleophile,[11, 13] the present study addressed structural and mechanistic details of water activation through both crystal structure analysis and computational QM/MM simulations. Significant structural differences in the active sites of *Mtb* and *E. coli* PPiases strongly suggest a distinct water activation mechanism for the *Mtb* homolog. We observed at least two water molecules in the structure of *Mtb* PPiase–PP_i complex that could in principle serve the catalytic role, based on their proximity to the bound PP_i (Figure A.3A). Because Asp54 (a counterpart of Asp67 proposed as the general base for *E. coli* PPiase) is too far from the PP_i and the bound water molecule to serve as a proton abstractor, Asp57 and Asp89 are

potential candidates for this role, as they are located in close enough proximity to PP_i. Indeed, our biochemical studies demonstrated that the Asp54Asn mutant retains significant catalytic activity (the mutation decreases k_{cat} only 3-fold), implicating either Asp54 or Asp89 in proton abstraction. The QM/MM simulations resolved this ambiguity and established that Asp89 could function to deprotonate the catalytic water, with very minor conformational changes in the active site. Even though for *E. coli* PPiase the effect of the analogous mutation (Asp67Asn) is greater (a 26-fold decrease in k_{cat}),[14] even in that system it is too small to account for this Asp as the sole catalytic base. Furthermore, analogs of Asp57Asn and Asp89Asn mutants in *E. coli* PPiase (Asp70Asn and Asp102Asn) resulted in lesser defects in the hydrolytic activity (150-fold and a 90-fold, respectively) than in *Mtb* PPiase (300- and 600-fold). The potential of Asp89 to act as an alternative general base is a novel finding, which not only suggests a likely activation mechanism for the *Mtb* PPiase, but provides an alternative, back-up mechanism for *E. coli* PPiase, explaining the modest effects on catalysis of mutating Asp67 in that homolog. This back-up catalytic strategy is likely common among enzymes of this family ensuring their robustness, as there is usually no redundant cytoplasmic phosphatase that is able to hydrolyze inorganic pyrophosphate. A similar catalytic redundancy was previously

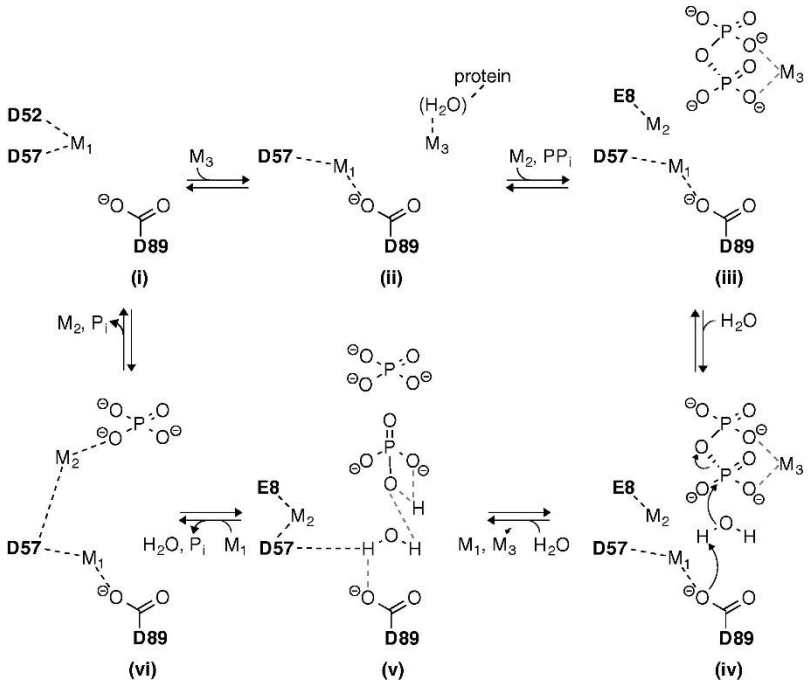


Figure A.11. Proposed Mtb PPiase mechanism. A proposed mechanism of the catalytic turnover of PPiase. Complex i: PPiase bound to metal ion M_1. Complex ii: PPiase bound to metal ions M_1 and M_3. Complex iii: PPiase bound to all three metal ions and the substrate, PP_i. Complex iv: Pre-catalytic complex formed as a result of binding of the catalytic water. Complex v: Post-catalytic complex PPiase–$2P_i$ formed after hydrolysis and the release of metal ions M_1 and M_3. Complex vi: Post-catalytic complex formed after release of one of the P_i products. Metal ion M_2 is shifted closer to the remaining P_i. Release of this P_i and the metal ion M_2 returns the enzyme to its initial state.

proposed for a catalytic Asp of an unrelated enzyme, triacylglycerol lipase from *Pseudomonas glumae*.[54]

A collection of crystal structures of *Mtb* PPiase – alone, bound to a PP_i, to two P_i ions and to one P_i ion provided us with a complete set of major mechanistic intermediates in the enzymatic turnover pathway for this enzyme (Figure A.11). These structures allowed us to refine the previous

mechanistic proposal (species i and v in Figure A.1), and they strongly suggest that the catalytic water is not bound in the active site prior to the PP_i binding, or if it is bound, it is positioned appropriately only upon PP_i binding. The structures also suggest that M_3 ion binds prior to PP_i, but that M_3 binding is likely stabilized in the presence of PP_i. M_3 dissociates after the PP_i hydrolysis step, likely because the proper bidentate coordination to the two P_i products cannot be maintained. In contrast, M_2 binding depends strictly on binding of PP_i or the presence of at least one of the P_i products and is also likely cooperative with PP_i/P_i binding, as M_2 ion is observed only in the presence of phosphate ligands and not in apo-enzyme structures.

After the hydrolysis, but before dissociation of the first P_i, M_1 ion is apparently destabilized and dissociates, likely because of disruptions in the hydrogen bonding and coordination with water as a result of hydrolysis. The P_i that is adjacent to Asp57 and Asp89 likely dissociates first of the two P_i products. The dissociation of the first P_i is likely driven by the repulsion between the two P_i products and, likely, by destabilizing structural perturbations at the catalytic center (near Asp57 and Asp89), whereas the remaining P_i ion maintains its favorable electrostatic interactions with the protein. This order of dissociation of the two Pi products is the opposite of that previously depicted in the *E. coli* PPiase mechanism (Figure A.1, species iv) proposed based on studies of crystals generated in the presence of inhibitory fluoride and then gradual soaking this fluoride out of the

crystals.[11] On the other hand, our proposed order of P_i dissociation is consistent with other structural studies of both *E. coli* and *S. cerevisiae* PPiases.[25, 46] After the first P_i dissociates, M_1 rebinds, as its coordination can likely be again satisfied due to increased flexibility of the Asp side chains and water molecules in the active site. On the other hand, metal ion M_2 rather than dissociating, progressively shifts its position towards the remaining P_i, as seen in the structures of *Mtb* PPiase–2P_i (Figure A.2B) and *Mtb* PPiase–P_i complexes (Figure A.5). This shift is sterically allowed by the dissociation of the first P_i.

The structural and computational information gained in this and earlier studies allow us to propose a more elaborate mechanistic scheme (Figure A.11). In this scheme, divalent metal ion M_3 binds to PPiase that is already bound to divalent metal ion M_1 (Figure A.11, species i and ii), likely in an equilibrium process. Binding of M_1 and M_3 enables binding of PP_i and divalent metal ion M_2 (Figure A.11, species iii). These binding events are followed by binding and/or repositioning of the already bound catalytic water molecule (Figure A.11, species iv), which can then become deprotonated by Asp89 and attack the adjacent phosphate group of PP_i. After catalysis, ion M_3, which was chelated by both phosphate groups of the PP_i, leaves the active site (Figure A.11, species v). Upon destabilization of the electrostatic interactions and the hydrogen bonding network due to the hydrolysis, the P_i closest to M_1 and M_2 dissociates from the active site leaving the other P_i

along with original metal ion M_1 as well as metal ion M_2, which is shifted from its original location towards the remaining P_i (Figure A.11, species vi). Finally, the P_i and M_2 dissociate from the active site thereby restoring the initial state of PPiase for a new catalytic cycle.

Two histidine residues (His21 and His86) in the vicinity of the catalytic center have been of interest, as they are major features distinguishing the *Mtb* PPiase from other characterized PPiases.[23, 24] Kinetic studies of mutant *Mtb* PPiase have shown that His21 and His86 likely function in divalent metal ion selectivity, as wild-type *Mtb* PPiase is inhibited by Zn^{2+} but not Mg^{2+}, whereas the single and the double alanine mutants of these residues are active with both divalent metals.[24] Our structural data support the idea that these His residues are not directly involved in normal catalysis. In addition, our structures suggest that these His residues are involved in the distinct conformation of the active site loops bearing Asp residues that are engaged in metal ion coordination or water activation, ultimately dictating distinct mechanistic features. In summary, this study combining structural and computational analyses, clarifies the description of the water-mediated hydrolysis in family I PPiases and suggests a defined sequence of cofactor and substrate binding and dissociation events during the catalytic turnover.

Bacterial PPiases are essential enzymes and, therefore, are potential targets for inhibition by small molecules. Discovery of such inhibitors should

be enabled by development of assays amenable to high-throughput screening (HTS).[18] Low-μM diphosphonate inhibitors selective for mammalian inorganic pyrophosphatases have been reported.[55] A recent HTS campaign yielded the first non-ionic inhibitory compounds with IC_{50}~30 μM.[19] Through a deeper understanding of the catalytic mechanism, the unique features of *Mtb* PPiase, such as the disposition of the active site Asp residues and the two His residues discussed above, may be used as specificity determinants in future development of selective inhibitory compounds against this enzyme. The structures described here could guide rational chemical optimization of such compounds.

Chapter contributions/acknowledgements

We thank Dr. Luiz Pedro de Carvalho for the *Mtb* PPiase plasmid and an initial protein preparation, Nathan Scharf for assistance with the refinement of the structure of product-bound *Mtb* PPiase, Dr. Sylvie Garneau-Tsodikova for help with figure preparation, the staff of sectors SER-CAT and LS-CAT at the Advanced Photon Source of the Argonne National Laboratories for assistance with X-ray diffraction data collection and Wayne State C&IT for the computational time.

References

[1] Kornberg, A. (1948) The participation of inorganic pyrophosphate in the reversible enzymatic synthesis of diphosphopyridine nucleotide, *J. Biol. Chem. 176*, 1475-1476.

[2] Kornberg, A. (1962) Horizons in Biochemistry, *Academic Press, Inc*.

[3] de Meis, L. (1984) Pyrophosphate of high and low energy. Contributions of pH, Ca2+, Mg2+, and water to free energy of hydrolysis, *The Journal of biological chemistry 259*, 6090-6097.

[4] Baykov, A. A., Shestakov, A. S., Kasho, V. N., Vener, A. V., and Ivanov, A. H. (1990) Kinetics and thermodynamics of catalysis by the inorganic pyrophosphatase of Escherichia coli in both directions, *European Journal of Biochemistry 194*, 879-887.

[5] Fabrichniy, I. P., Lehtiö, L., Salminen, A., Zyryanov, A. B., Baykov, A. A., Lahti, R., and Goldman, A. (2004) Structural studies of metal ions in family II pyrophosphatases: the requirement for a Janus ion, *Biochemistry 43*, 14403-14411.

[6] Gajadeera, C. S., Zhang, X., Wei, Y., and Tsodikov, O. V. (2015) Structure of inorganic pyrophosphatase from Staphylococcus aureus reveals conformational flexibility of the active site, *Journal of structural biology 189*, 81-86.

[7] Kajander, T., Kellosalo, J., and Goldman, A. (2013) Inorganic pyrophosphatases: one substrate, three mechanisms, *FEBS letters 587*, 1863-1869.

[8] Merckel, M. C., Fabrichniy, I. P., Salminen, A., Kalkkinen, N., Baykov, A. A., Lahti, R., and Goldman, A. (2001) Crystal Structure of Streptococcus mutans Pyrophosphatase: A New Fold for an Old Mechanism, *Structure 9*, 289-297.

[9] Rantanen, M. K., Lehtiö, L., Rajagopal, L., Rubens, C. E., and Goldman, A. (2007) Structure of the Streptococcus agalactiae family II inorganic pyrophosphatase at 2.80 Å resolution, *Acta Crystallographica Section D: Biological Crystallography 63*, 738-743.

[10] Salminen, A., Ilias, M., Belogurov, G. A., Baykov, A. A., Lahti, R., and Young, T. (2006) Two soluble pyrophosphatases in Vibrio cholerae: Transient redundancy or enduring cooperation?, *Biochemistry (Moscow) 71*, 978-982.

[11] Samygina, V. R., Moiseev, V. M., Rodina, E. V., Vorobyeva, N. N., Popov, A. N., Kurilova, S. A., Nazarova, T. I., Avaeva, S. M., and Bartunik, H. D. (2007) Reversible Inhibition of Escherichia coli Inorganic Pyrophosphatase by Fluoride: Trapped Catalytic Intermediates in Cryo-crystallographic Studies, *Journal of Molecular Biology 366*, 1305-1317.

[12] Salminen, T., Käpylä, J., Heikinheimo, P., Kankare, J., Goldman, A., Heinonen, J., Baykov, A. A., Cooperman, B. S., and Lahti, R. (1995) Structure and function analysis of Escherichia coli inorganic pyrophosphatase: is a hydroxide ion the key to catalysis?, *Biochemistry 34*, 782-791.

[13] Heikinheimo, P., Tuominen, V., Ahonen, A.-K., Teplyakov, A., Cooperman, B. S., Baykov, A. A., Lahti, R., and Goldman, A. (2001) Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase, *Proceedings of the National Academy of Sciences of the United States of America 98*, 3121-3126.

[14] Avaeva, S., Ignatov, P., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Oganessyan, V., and Harutyunyan, E. (1996) Escherichia coli inorganic pyrophosphatase: site directed mutagenesis of the metal binding sites, *FEBS letters 399*, 99-102.

[15] Chen, J., Brevet, A., Fromant, M., Leveque, F., Schmitter, J. M., Blanquet, S., and Plateau, P. (1990) Pyrophosphatase is essential for growth of Escherichia coli, *J. Bacteriol. 172*, 5686-5689.

[16] Griffin, J. E., Gawronski, J. D., DeJesus, M. A., Ioerger, T. R., Akerley, B. J., and Sassetti, C. M. (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism, *PLoS Pathogens 7*, 1-9.

[17] Lundin, M., Baltscheffsky, H., and Ronne, H. (1991) Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function, *The Journal of biological chemistry 266*, 12168-12172.

[18] Biswas, T., Resto-Roldán, E., Sawyer, S. K., Artsimovitch, I., and Tsodikov, O. V. (2012) A novel non-radioactive primase–pyrophosphatase activity assay and its application to the discovery of inhibitors of Mycobacterium tuberculosis primase DnaG, *Nucleic acids Research 41*, e56.

[19] Lv, W., Banerjee, B., Molland, K. L., Seleem, M. N., Ghafoor, A., Hamed, M. I., Wan, B., Franzblau, S. G., Mesecar, A. D., and Cushman, M. (2014) Synthesis of 3-(3-aryl-pyrrolidin-1-yl)-5-aryl-1,2,4-triazines that have antibacterial activity and also inhibit inorganic pyrophosphatase, *Bioorganic & Medicinal Chemistry 22*, 406-418.

[20] Harutyunyan, E. H., Kuranova, I. P., Vainshtein, B. K., Höhne, W. E., Lamzin, V. S., Dauter, Z., Teplyakov, A. V., and Wilson, K. S. (1996) X Ray Structure of Yeast Inorganic Pyrophosphatase Complexed with Manganese and Phosphate, *European journal of biochemistry 239*, 220-228.

[21] Harutyunyan, E. H., Oganessyan, V. Y., Oganessyan, N. N., Avaeva, S. M., Nazarova, T. I., Vorobyeva, N. N., Kurilova, S. A., Huber, R., and

Mather, T. (1997) Crystal structure of holo inorganic pyrophosphatase from Escherichia coli at 1.9 Å resolution. Mechanism of hydrolysis, *Biochemistry 36*, 7754-7760.

[22] Pohjanjoki, P., Fabrichniy, I. P., Kasho, V. N., Cooperman, B. S., Goldman, A., Baykov, A. A., and Lahti, R. (2001) Probing Essential Water in Yeast Pyrophosphatase by Directed Mutagenesis and Fluoride Inhibition Measurements, *J. Biol. Chem. 276*, 434-441.

[23] Benini, S., and Wilson, K. (2011) Structure of the Mycobacterium tuberculosis soluble inorganic pyrophosphatase Rv3628 at pH 7.0, *Acta Crystallographica Section F: Structural Biology and Crystallization Communications F67*, 866-870.

[24] Tammenkoski, M., Benini, S., Magretova, N. N., Baykov, A. A., and Lahti, R. (2005) An Unusual, His-dependent Family I Pyrophosphatase from Mycobacterium tuberculosis, *J. Biol. Chem. 280*, 41819-41826.

[25] Avaeva, S., Kurilova, S., Nazarova, T., and Rodina, E. (1997) Crystal structure of Escherichia coli inorganic pyrophosphatase complexed with SO4 2−, *FEBS …*.

[26] Otwinowski, Z., and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode, *Methods in Enzymology 276*, 307-326.

[27] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.

[28] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica Section D: Biological Crystallography 53*, 240-255.

[29] Emsley, P., Cowtan, K., Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallographica Section D: Biological Crystallography D60*, 2126-2132.

[30] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2012) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallographica Section D Biological Crystallography D66*, 694-701.

[31] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Research 35*, W375-W383.

[32] Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, R., Duke, R., Luo, M., Crowley, R., Walker, W., Zhang, K. M., Merz, B., Wang, S., Hayik, A., Roitberg, G., Seabra, I., Kolossvary, K. F., Wong, F., Paesani, J., Vanicek, X., Wu, S., Brozell, T., Steinbrecher, H., Gohlke, L., Yang, C., Tan, J., Mongan, V., Hornak, G., Cui, D. H., Mathews, M. G., Seetin, C., Sagui, V., Babin, P., and Kollman, P. A. (2010) Amber 11, *University of California, San Francisco*.

[33] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics 79*, 926-935.

[34] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, T., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Comput Chem 26*, 1668-1688.

[35] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.

[36] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method, *J. Chem. Phys. 103*, 8577-8593.

[37] York, D. M., Darden, T. A., and Pedersen, L. G. (1993) The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods, *The Journal of Chemical Physics 99*, 8345.

[38] Loncharich, R. J., Brooks, B. R., and Pastor, R. W. (1992) Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide, *Biopolymers 32*, 523-535.

[39] Parks, J. M., Hu, H., Cohen, A. J., and Yang, W. (2008) A pseudobond parametrization for improved electrostatics in quantum mechanical/molecular mechanical simulations of enzymes, *The Journal of Chemical Physics 129*, 154106.

[40] Zhang, Y., Lee, T.-S., and Yang, W. (1999) A pseudobond approach to combining quantum mechanical and molecular mechanical methods, *The Journal of Chemical Physics 110*, 46.

[41] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Jr., J. A. M., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L.,

Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, *Gaussian Inc, Wallingford, CT, USA*.

[42] Ren, P., and Ponder, J. W. (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations, *Journal of Computational Chemistry 23*, 1497-1506.

[43] Zhang, Y., Liu, H., and Yang, W. (2000) Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface, *The Journal of Chemical Physics 112*, 3483.

[44] Burger, S. K., and Yang, W. (2006) Quadratic string method for determining the minimum-energy path based on multiobjective optimization, *The Journal of chemical physics 124*, 054109.

[45] Cisneros, G. A., and Yang, W. (2009) Comparison of reaction barriers in energy and free energy for enzyme catalysis, *Multi-scale Quantum Models for Biocatalysis*, 57-78.

[46] Oksanen, E., Ahonen, A.-K., Tuominen, H., Tuominen, V., Lahti, R., Goldman, A., and Heikinheimo, P. (2007) A complete structural description of the catalytic cycle of yeast pyrophosphatase, *Biochemistry 46*, 1228-1239.

[47] Hou, G., and Cui, Q. (2011) QM/MM analysis suggests that alkaline phosphatase (AP) and nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution: Implication for catalytic promiscuity in the AP superfamily, *Journal of the American Chemical Society 134*, 229-246.

[48] Hou, G., and Cui, Q. (2013) Stabilization of different types of transition states in a single enzyme active site: QM/MM analysis of enzymes in

the alkaline phosphatase superfamily, *Journal of the American Chemical Society 135*, 10457-10469.

[49] Koivisto, P., Duncan, T., Lindahl, T., and Sedgwick, B. (2003) Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase, *J. Biol. Chem. 278*, 44348-44354.

[50] López-Canut, V., Roca, M., Bertrán, J., Moliner, V., and Tuñón, I. (2010) Theoretical study of phosphodiester hydrolysis in nucleotide pyrophosphatase/phosphodiesterase. Environmental effects on the reaction mechanism, *Journal of the American Chemical Society 132*, 6955-6963.

[51] Prasad, R. B., Plotnikov, N. V., and Warshel, A. (2012) Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping, *The journal of physical chemistry. B 117*, 153-163.

[52] Wong, K. Y., and Gao, J. (2011) Insight into the phosphodiesterase mechanism from combined QM/MM free energy simulations, *FEBS Journal 278*, 2579-2595.

[53] Yang, L., Liao, R.-Z., Yu, J.-G., and Liu, R.-Z. (2009) DFT study on the mechanism of Escherichia coli inorganic pyrophosphatase, *The journal of physical chemistry. B 113*, 6505-6510.

[54] Noble, M. E. M., Cleasby, A., Johnson, L. N., Egmond, M. R., and Frenken, L. G. J. (1993) The crystal structure of triacylglycerol lipase from Pseudomonas glumae reveals a partially redundant catalytic aspartate, *FEBS Letters 331*, 123-128.

[55] Smirnova, I. N., Kudryavtseva, N. A., Komissarenko, S. V., Tarusova, N. B., and Baykov, A. A. (1988) Diphosphonates are potent inhibitors of mammalian inorganic pyrophosphatase, *Archives of Biochemistry and Biophysics 267*, 280-284.

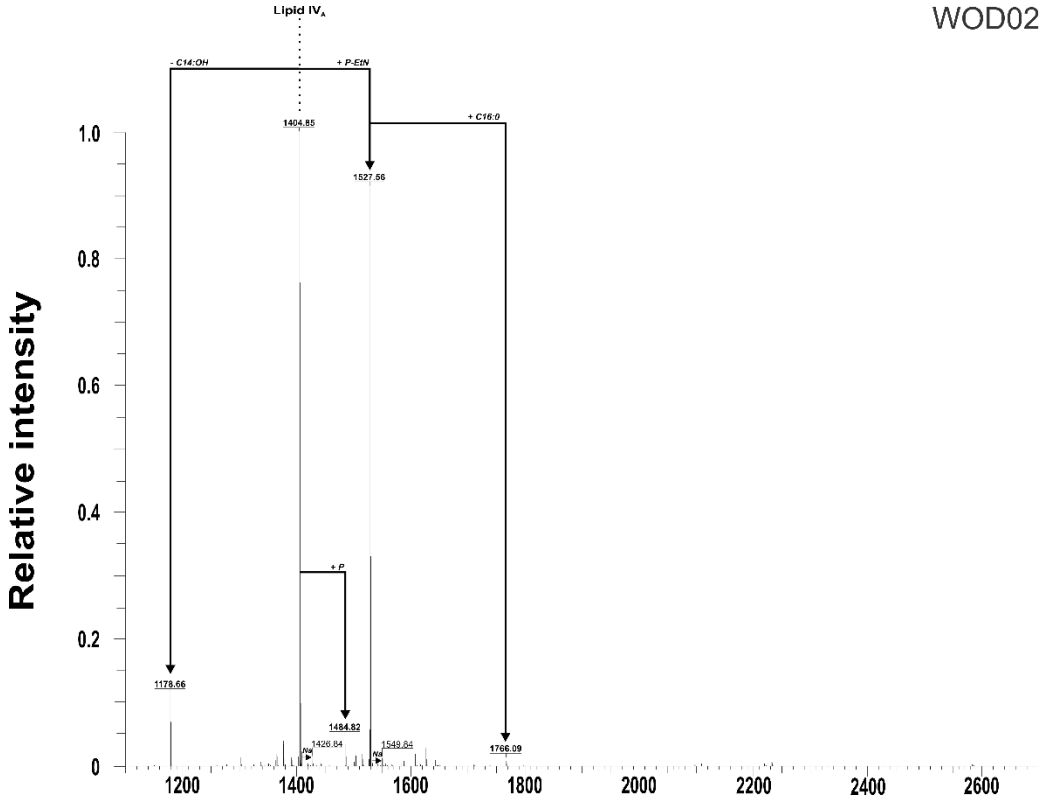
Appendix B

ESI FT-ICR MS SPECTRA OF LPS SAMPLES GENERATED FROM WOD02

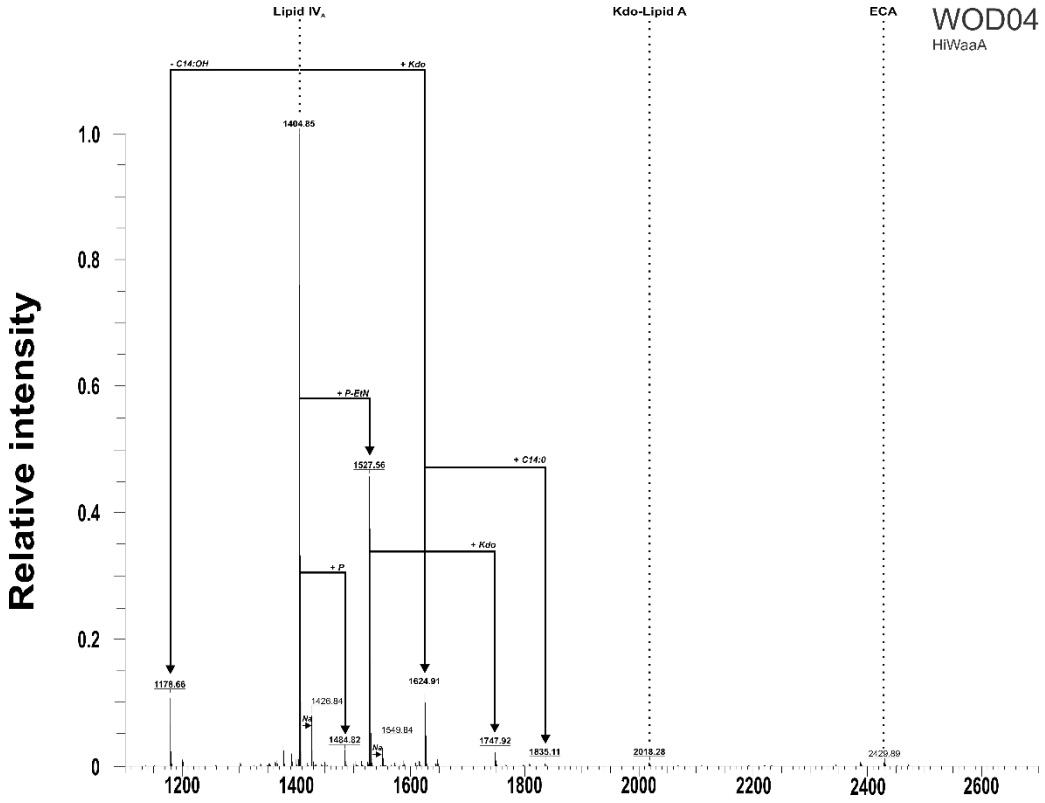
Obs Mass	Calc Mass	Chemical Composition
1178.66	1178.661	2*GlcN, 2*P, 3*(OH)-14:0
1404.85	1404.854	2*GlcN, 2*P, 4*(OH)-14:0
1484.82	1484.820	2*GlcN, 3*P, 4*(OH)-14:0
1527.56	1527.863	2*GlcN, 2*P, 4*(OH)-14:0, 1*P-EtN
1624.91	1624.912	2*GlcN, 2*P, 4*(OH)-14:0, 1*Kdo
1747.92	1747.920	2*GlcN, 2*P, 4*(OH)-14:0, 1*Kdo, 1*P-EtN
1766.09	1766.092	2*GlcN, 2*P, 4*(OH)-14:0, 1*16:0
1800.94	1800.944	2*GlcN, 2*P, 3*(OH)-14:0, 1*12:0, 2*Kdo
1816.98	1816.976	2*GlcN, 2*P, 4*(OH)-14:0, 1*Kdo, 1*Hep
1835.11	1835.111	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*Kdo
1844.97	1844.971	2*GlcN, 2*P, 4*(OH)-14:0, 2*Kdo
1881.91	1880.911	2*GlcN, 3*P, 3*(OH)-14:0, 1*12:0, 2*Kdo
1994.01	1993.008	2*GlcN, 2*P, 3*(OH)-14:0, 1*12:0, 2*Kdo, 1*Hep
2000.15	1999.142	2*GlcN, 2*P, 4*(OH)-14:0, 1*12:0, 1*Kdo, 1*Hep
2018.28	2017.278	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 1*Kdo
2028.14	2027.137	2*GlcN, 2*P, 4*(OH)-14:0, 1*12:0, 2*Kdo
2073.98	2072.974	2*GlcN, 3*P, 3*(OH)-14:0, 1*12:0, 2*Kdo, 1*Hep
2108.11	2107.104	2*GlcN, 3*P, 4*(OH)-14:0, 1*12:0, 2*Kdo
2141.29	2140.286	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 1*Kdo, 1*Hep
2220.21	2219.201	2*GlcN, 2*P, 4*(OH)-14:0, 1*12:0, 1*Kdo, 1*Hep
2238.34	2237.336	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 2*Kdo
2248.20	2247.196	2*GlcN, 2*P, 4*(OH)-14:0, 1*12:0, 3*Kdo
2300.17	2299.167	2*GlcN, 3*P, 4*(OH)-14:0, 1*12:0, 1*Kdo, 1*Hep
2318.31	2317.302	2*GlcN, 3*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 2*Kdo
2328.17	2327.162	2*GlcN, 3*P, 4*(OH)-14:0, 1*12:0, 3*Kdo
2429.89	2429.89	ECA
2430.40	2429.399	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 2*Kdo, 1*Hep
2458.40	2457.394	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 3*Kdo
2510.37	2509.366	2*GlcN, 3*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 2*Kdo, 1*Hep
2538.36	2537.360	2*GlcN, 3*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 3*Kdo
2650.46	2649.458	2*GlcN, 2*P, 4*(OH)-14:0, 1*14:0, 1*12:0, 3*Kdo, 1*Hep

Table B.1. Mass spectrometric data table. List of mass peaks of LPS analogs identified. Due to isotopic distributions above 1844.97 major peaks are 1 u higher.

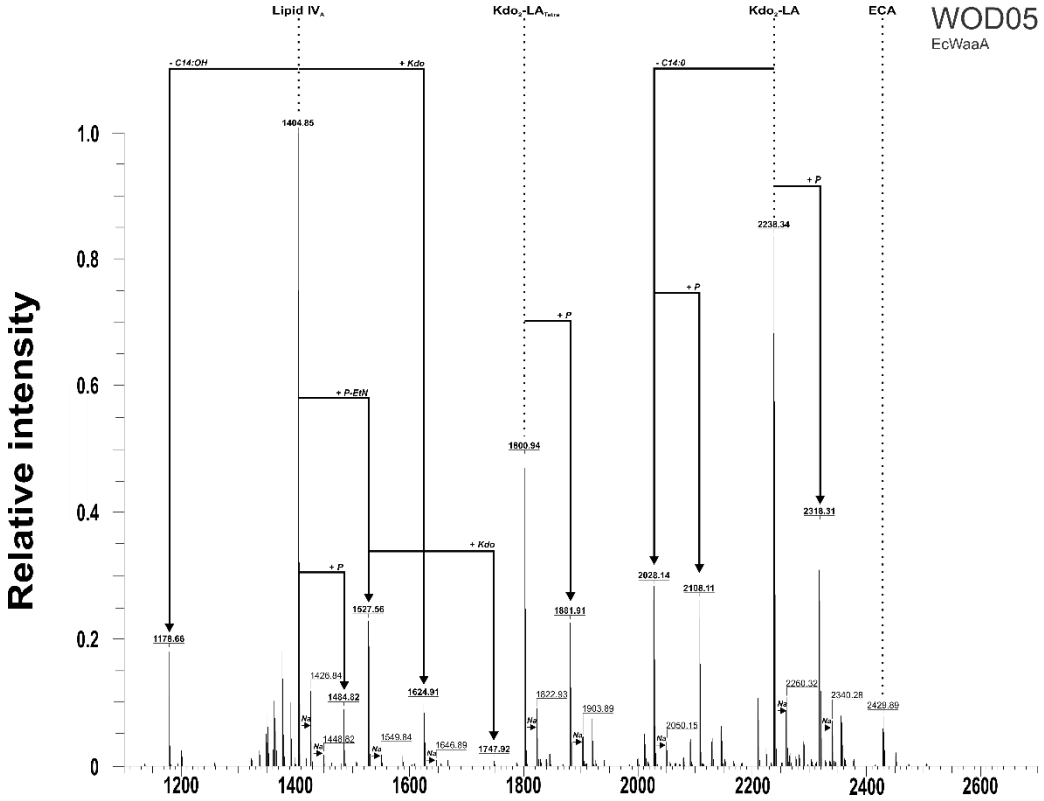
Figure B.1. Mass spectra from WOD02 and derivatives. (A-U) Mass spectra of purified LPS samples from WOD02 derivatives from m/z 1100-2700 u.



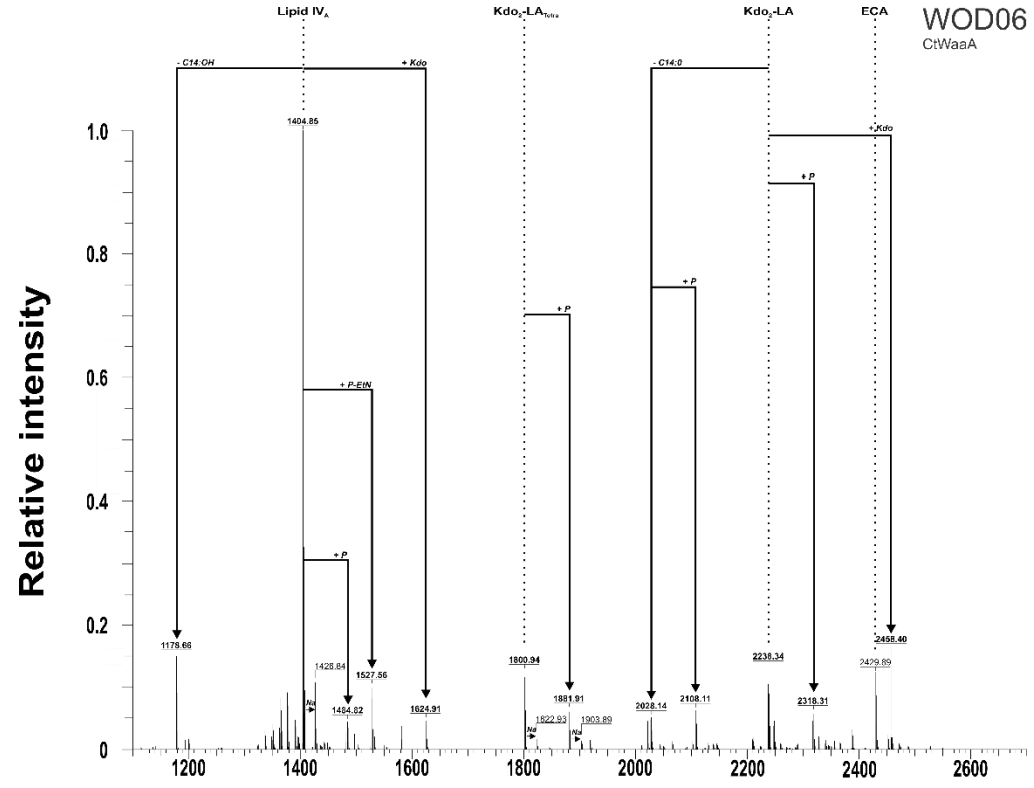
A. Mass spectrum of WOD02 with mass units of m/z.



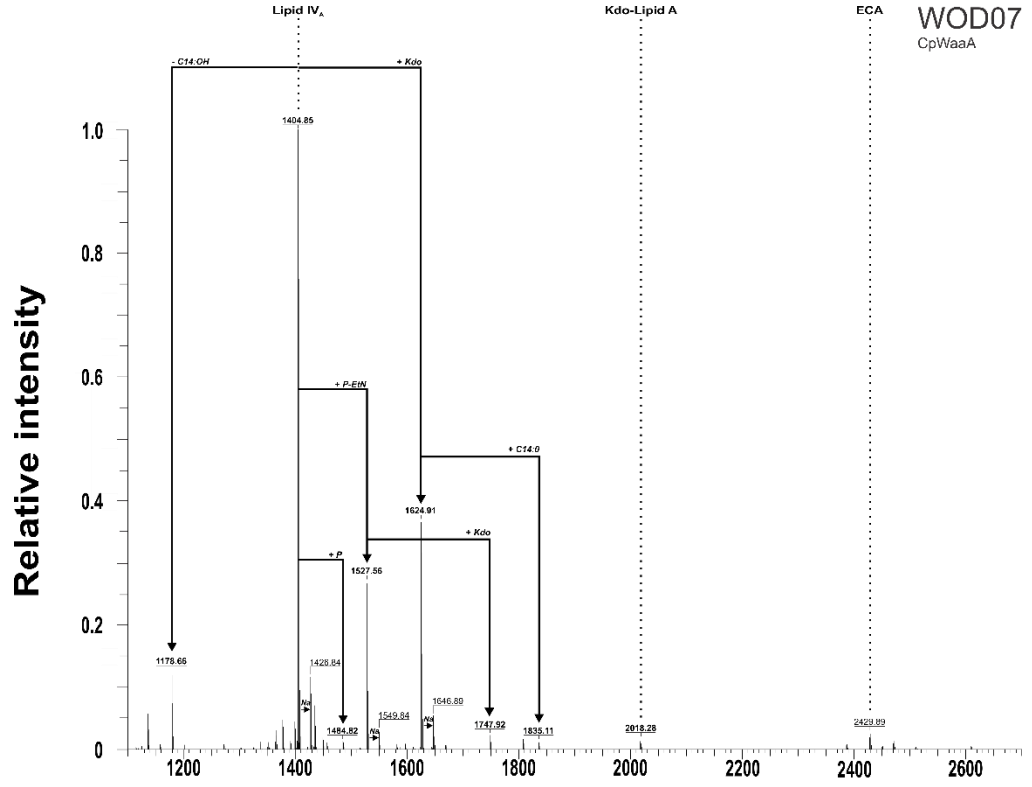
B. Mass spectrum of WOD04 with mass units of m/z.



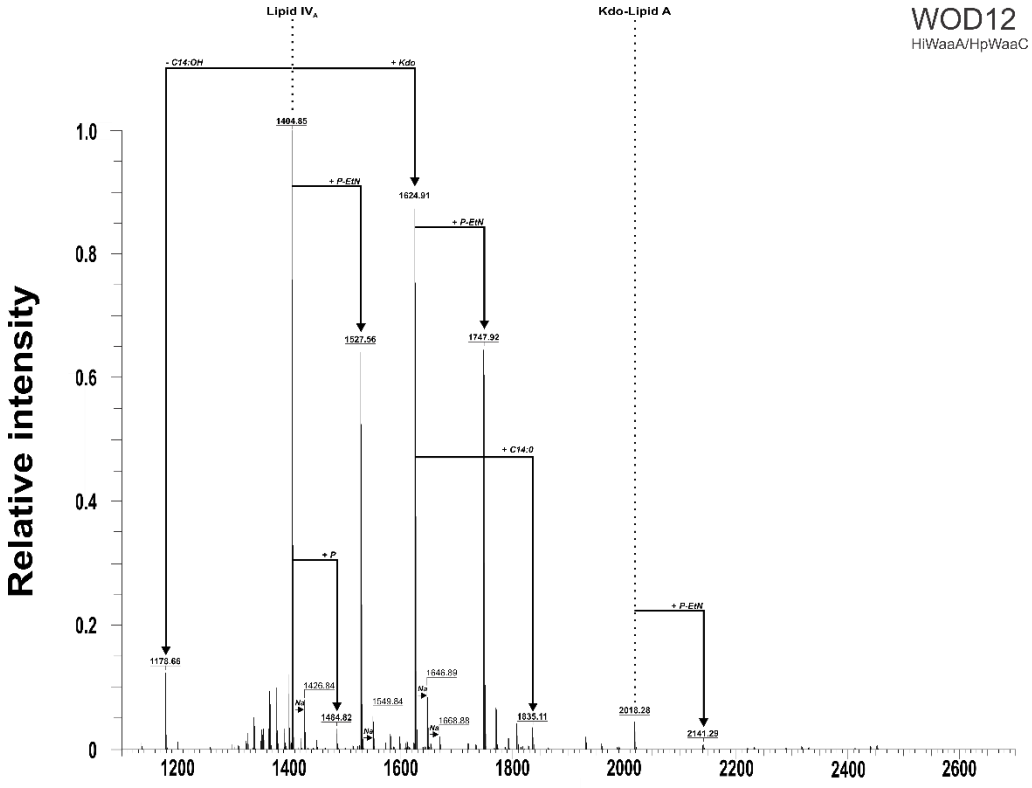
C. Mass spectrum of WOD05 with mass units of m/z.



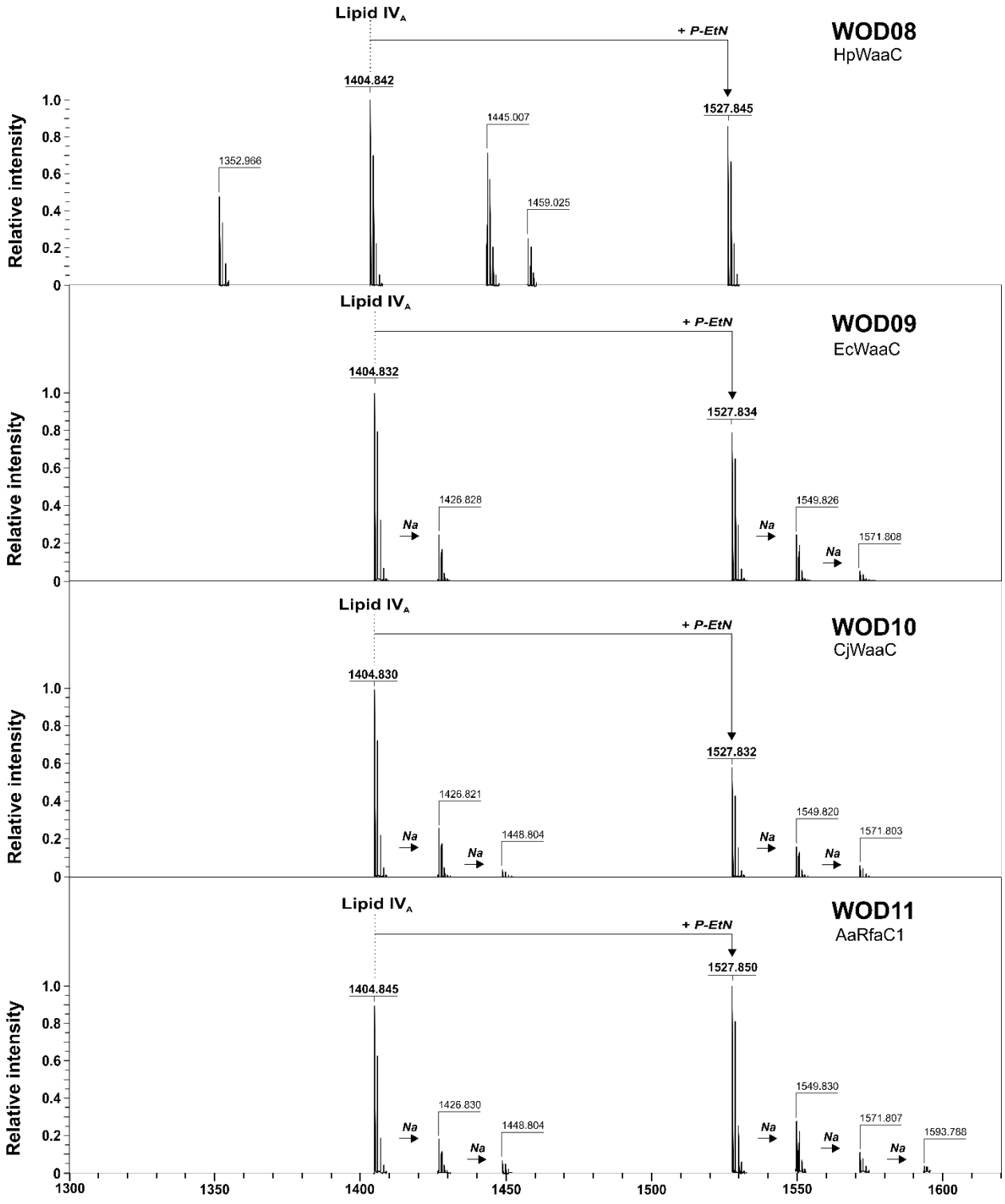
D. Mass spectrum of WOD06 with mass units of m/z.



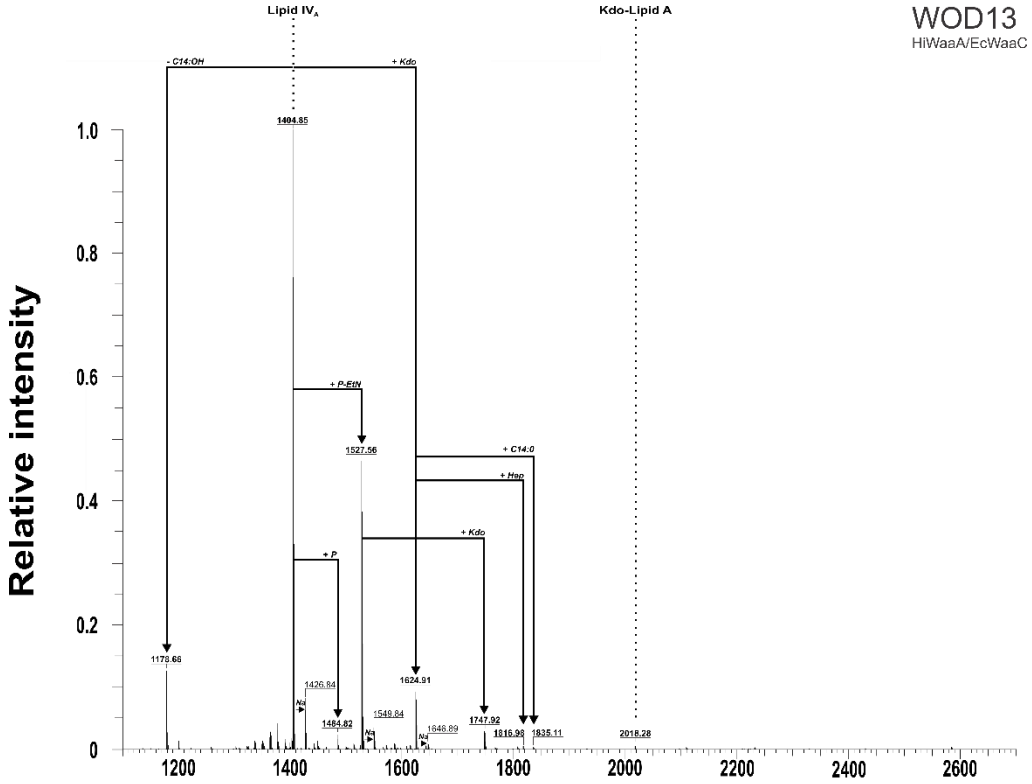
E. Mass spectrum of WOD07 with mass units of m/z.



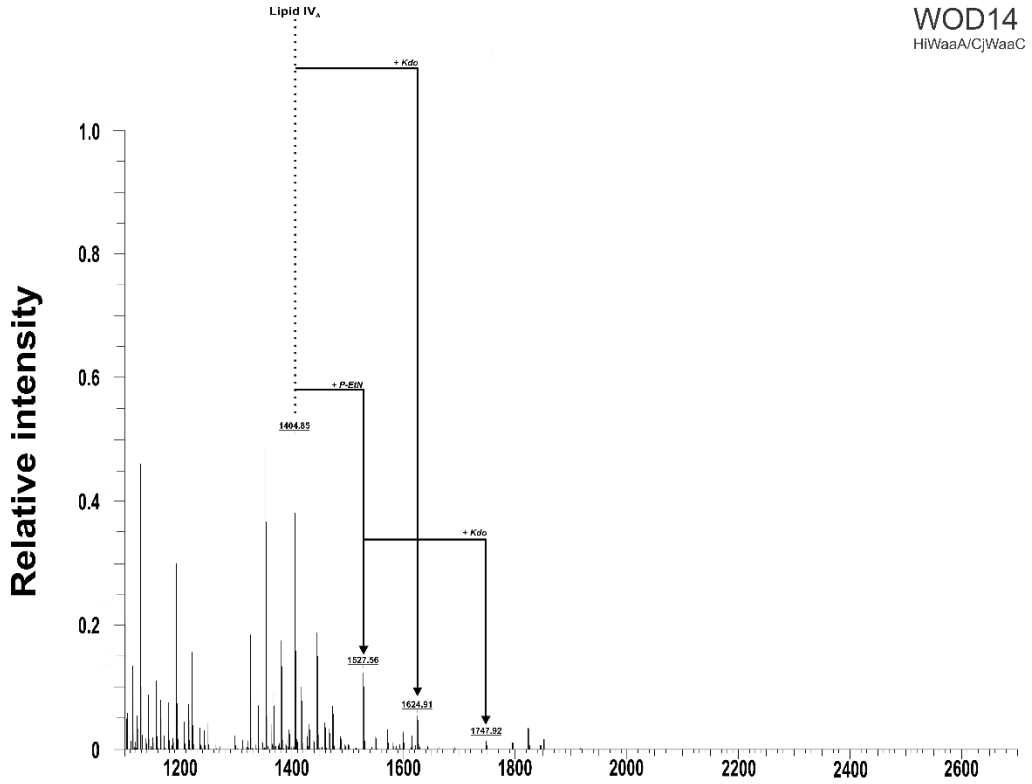
F. Mass spectrum of WOD08 with mass units of m/z.



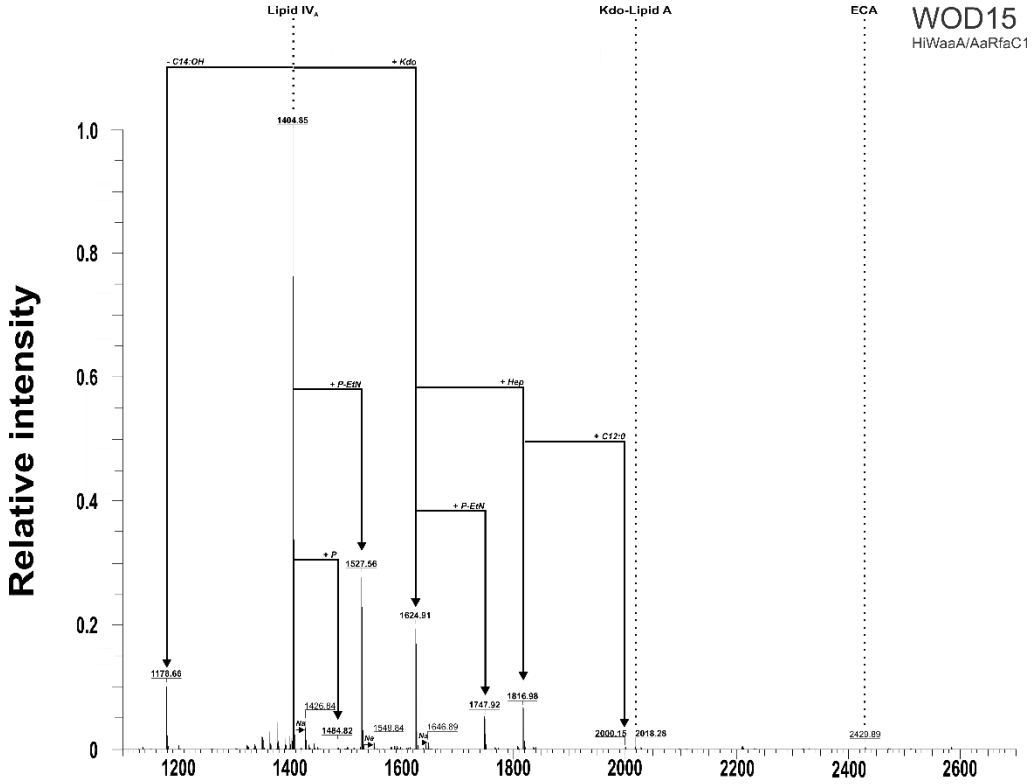
G. Mass spectra of WOD08-11 which lack Kdo transferases with mass units of m/z.



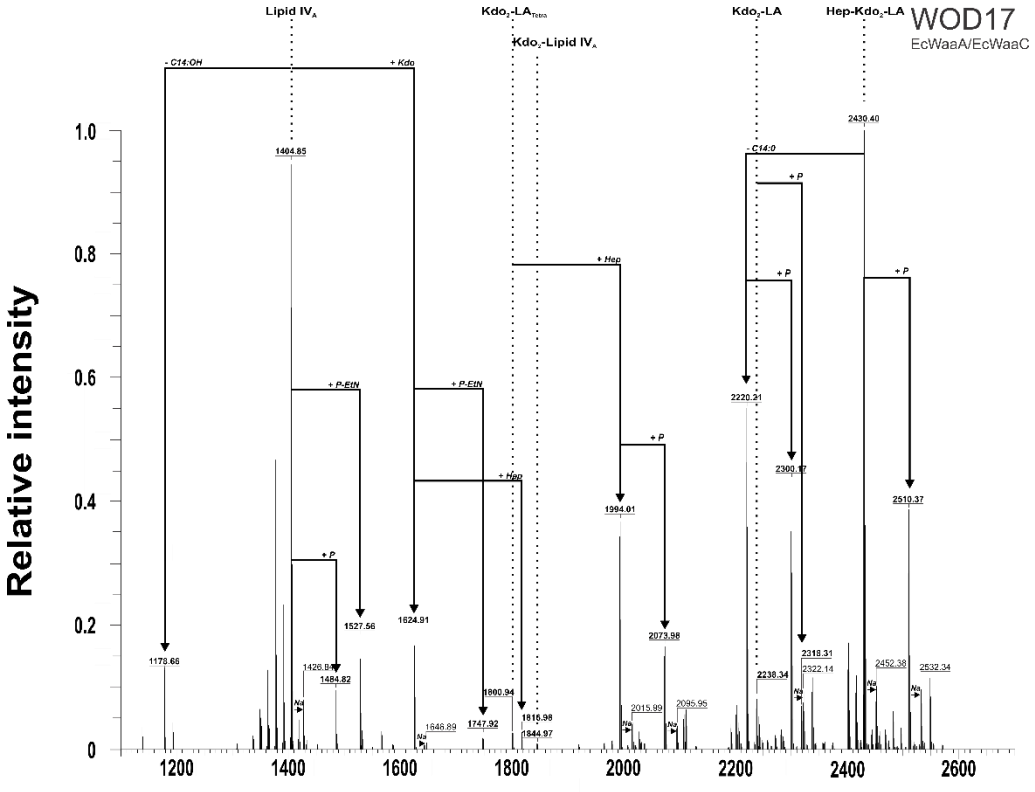
H. Mass spectra of WOD13 with mass units of m/z.



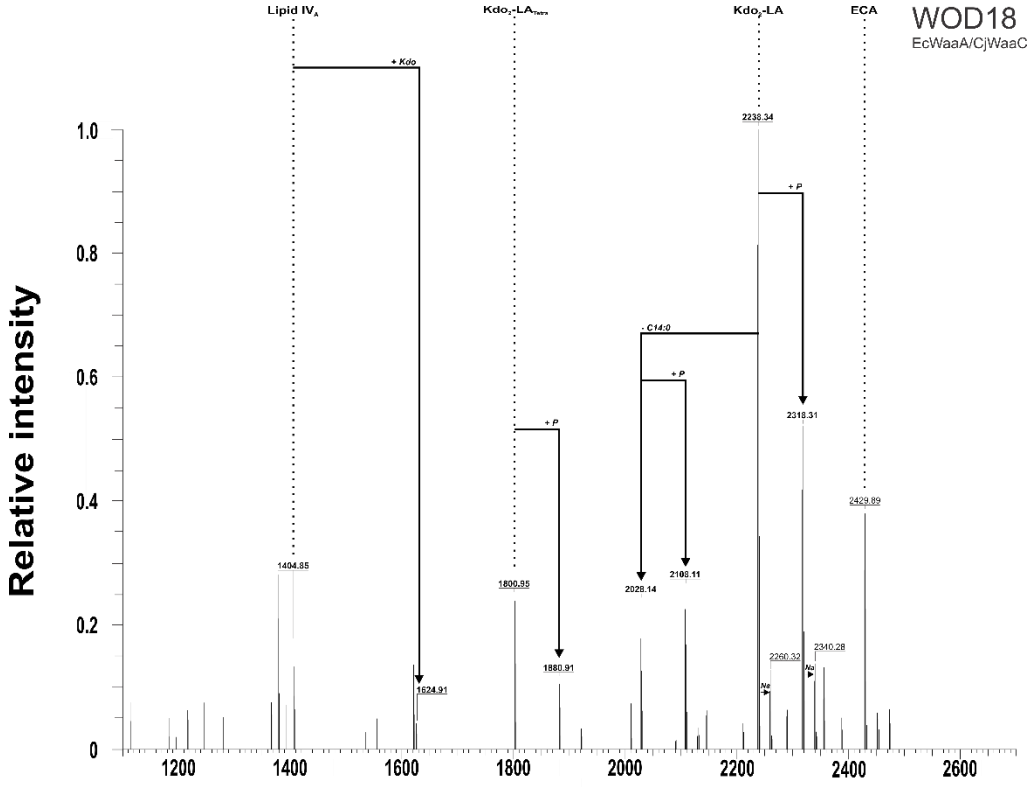
I. Mass spectra of WOD14 with mass units of m/z.



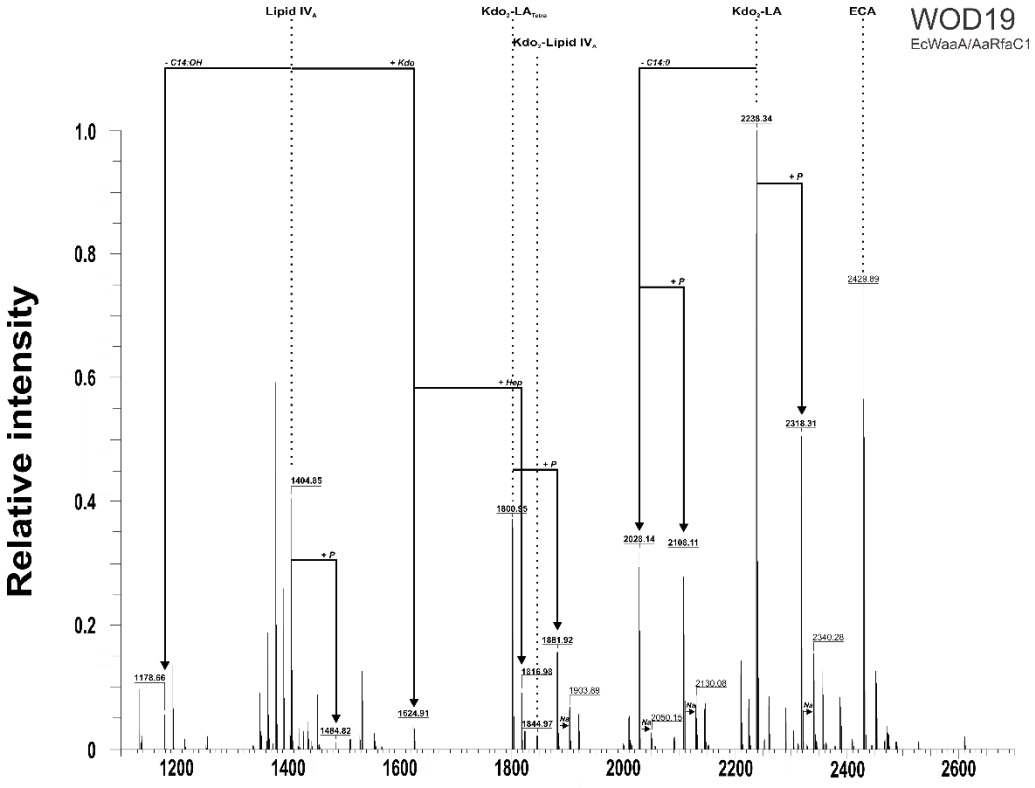
J. Mass spectra of WOD15 with mass units of m/z.



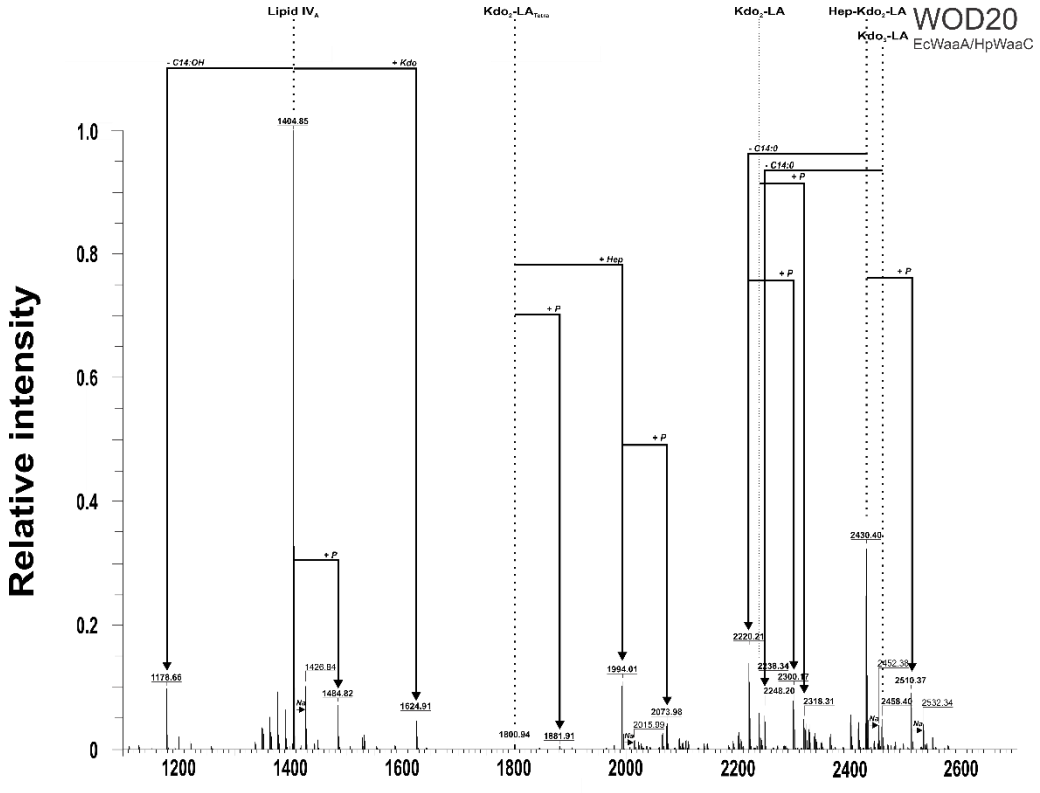
K. Mass spectra of WOD17 with mass units of m/z.



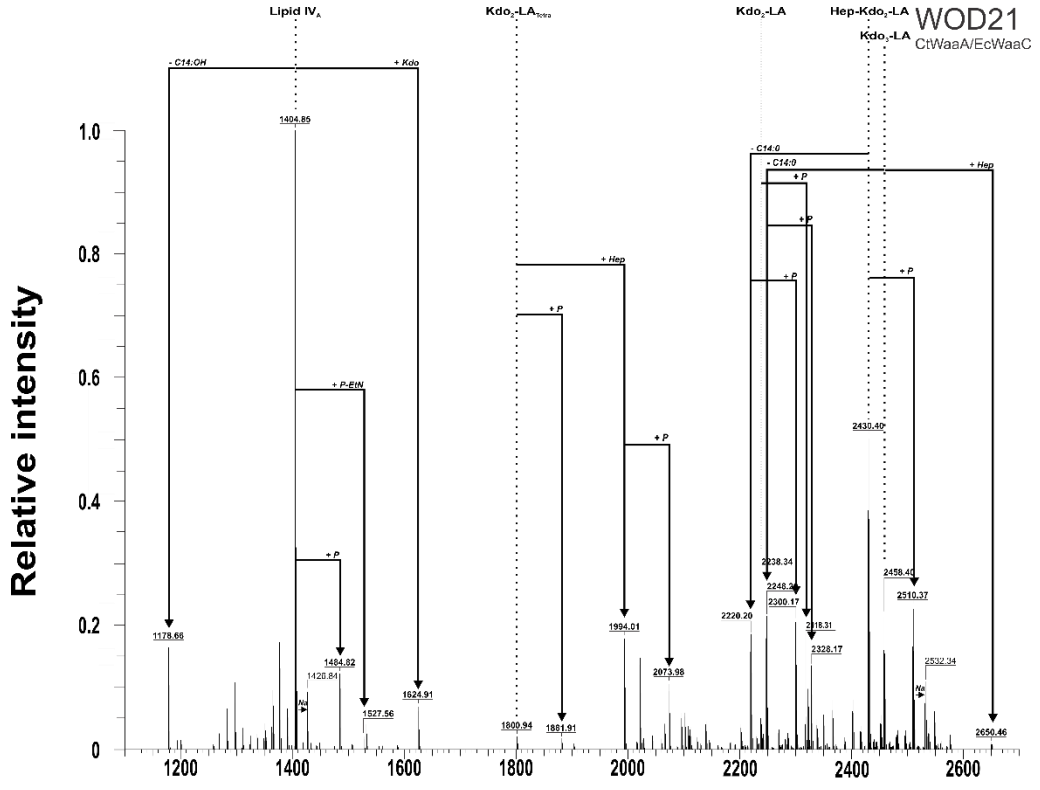
L. Mass spectra of WOD18 with mass units of m/z.



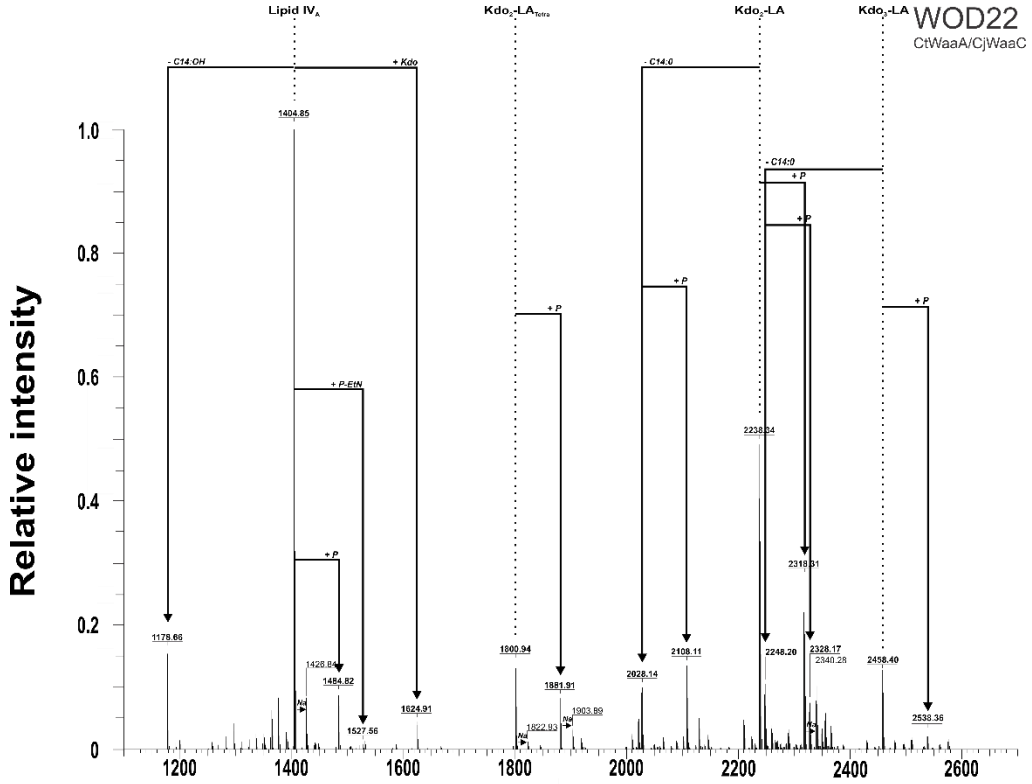
M. Mass spectra of WOD19 with mass units of m/z.



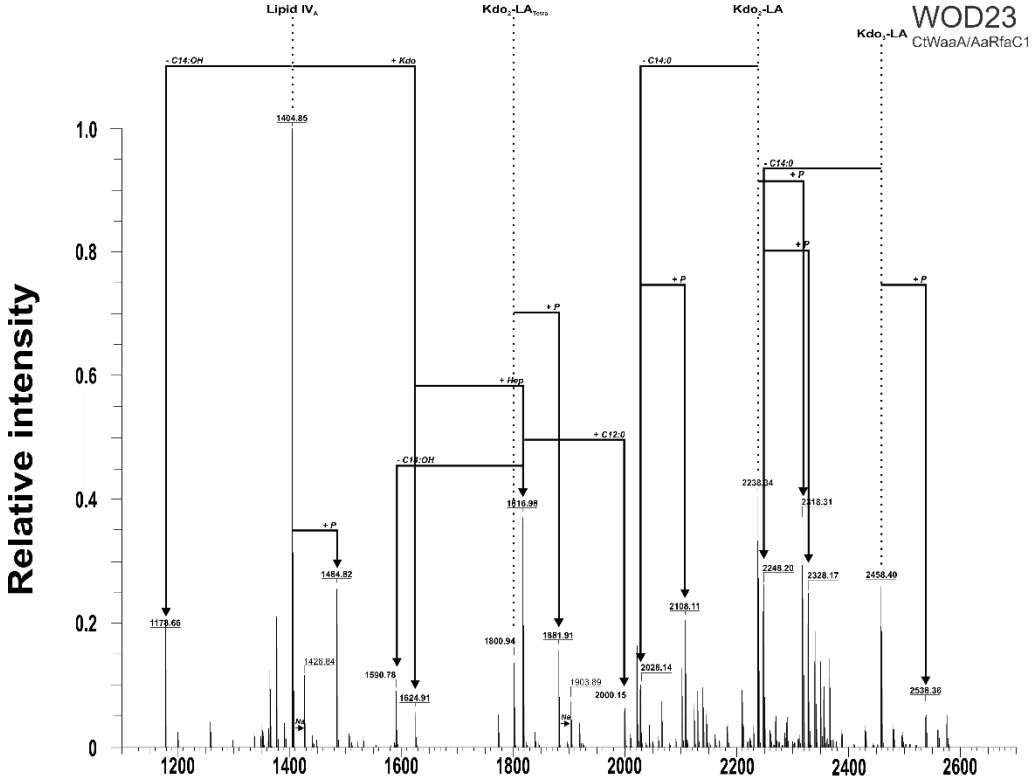
N. Mass spectra of WOD20 with mass units of m/z.



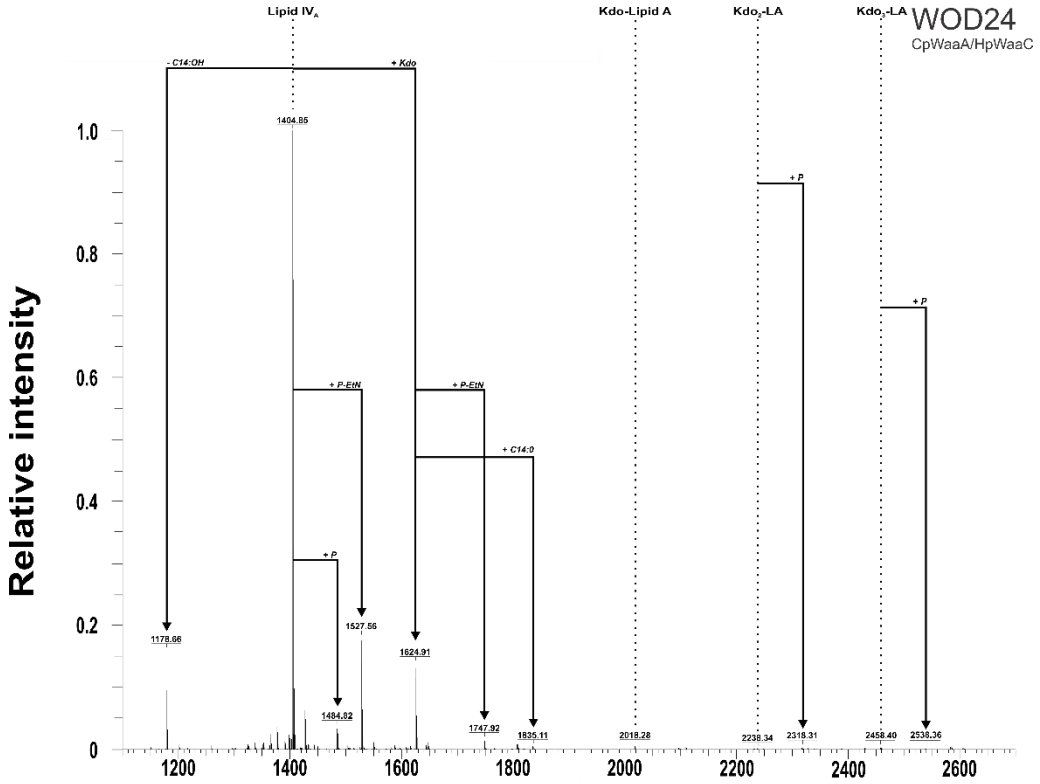
O. Mass spectra of WOD21 with mass units of m/z.



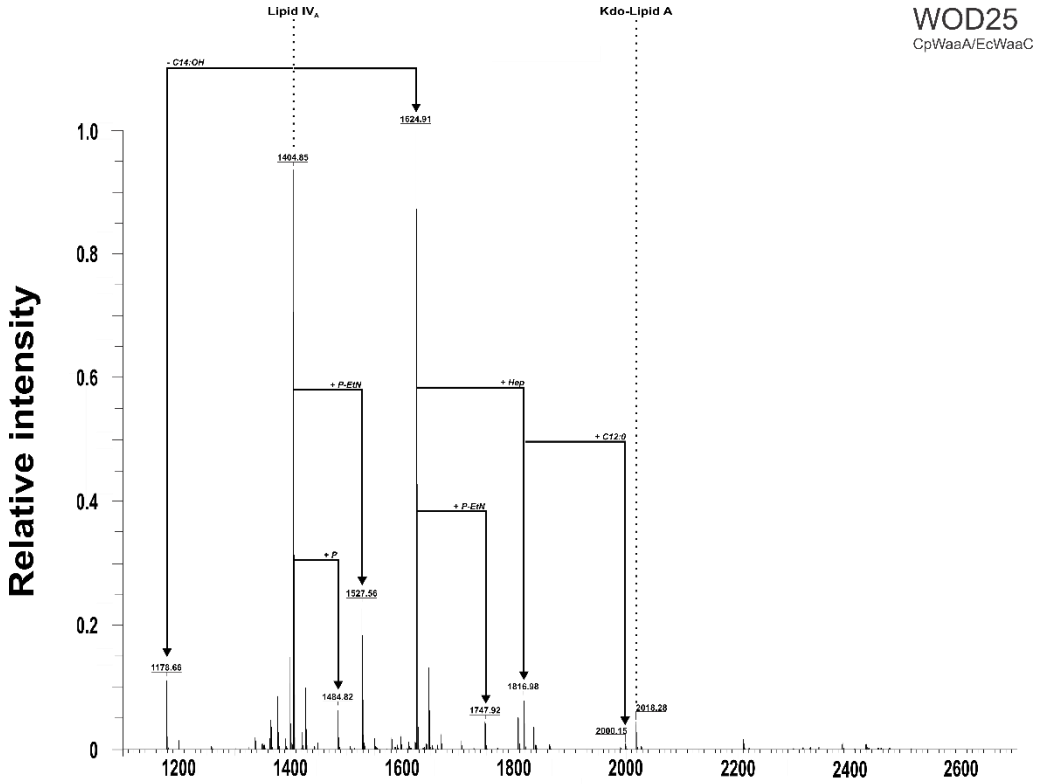
P. Mass spectra of WOD22 with mass units of m/z.



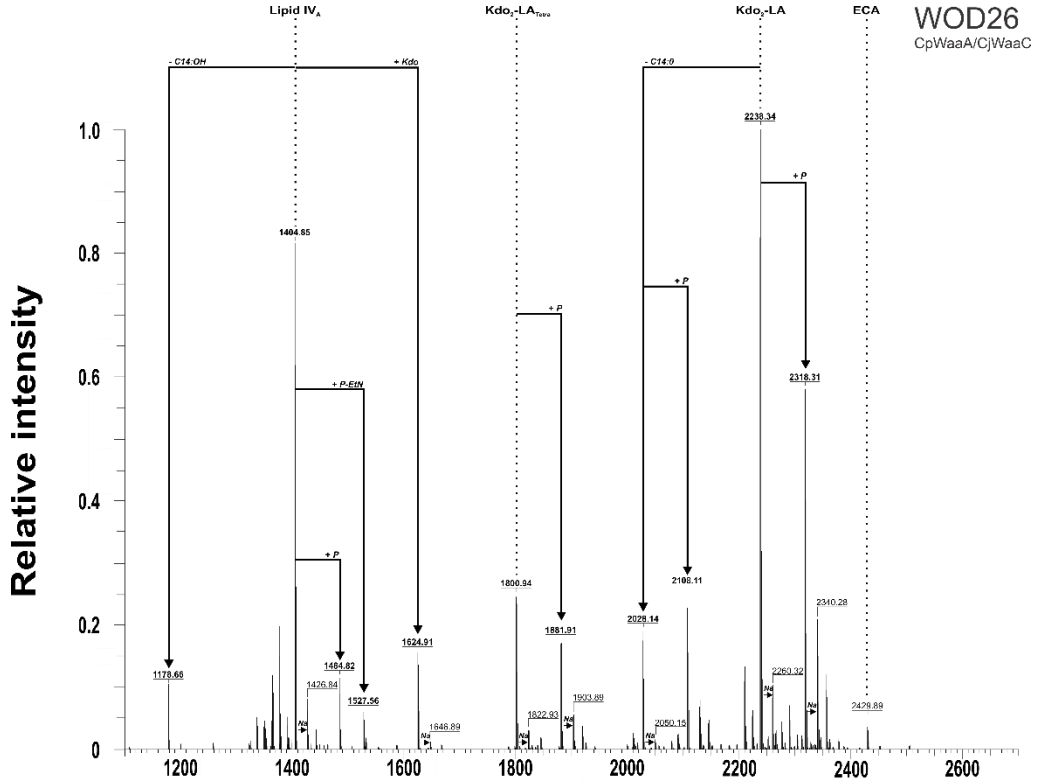
Q. Mass spectra of WOD23 with mass units of m/z.



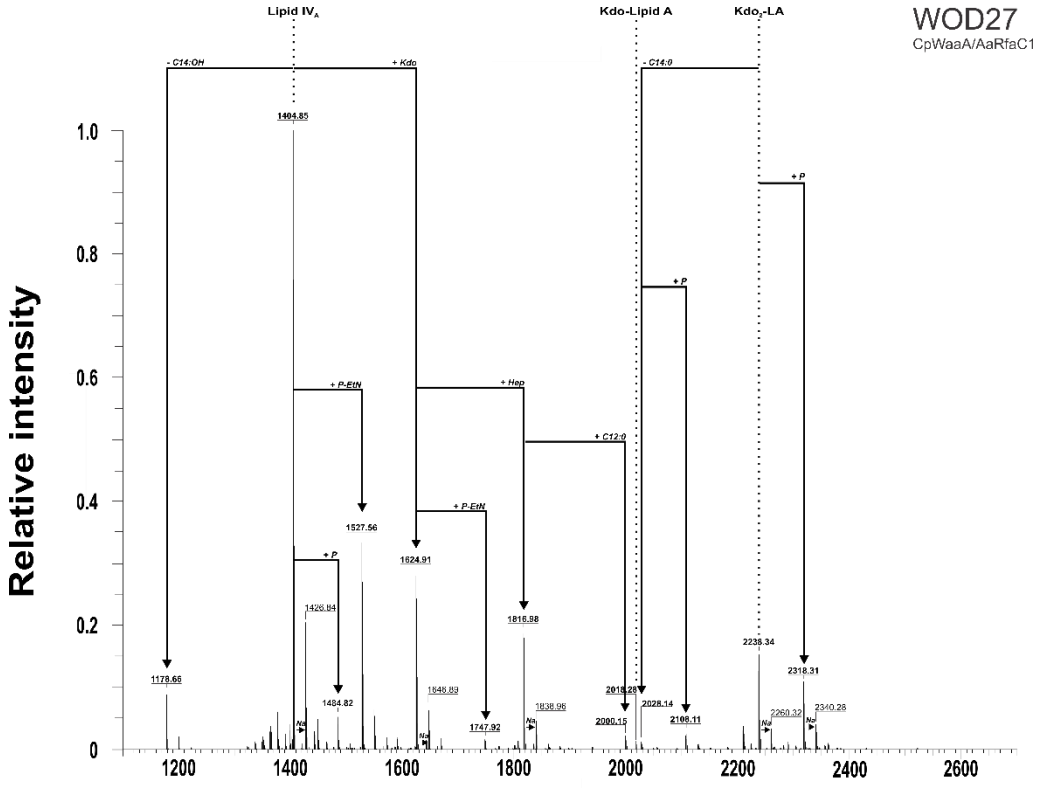
R. Mass spectra of WOD24 with mass units of m/z.



S. Mass spectra of WOD25 with mass units of m/z.



T. Mass spectra of WOD26 with mass units of m/z.



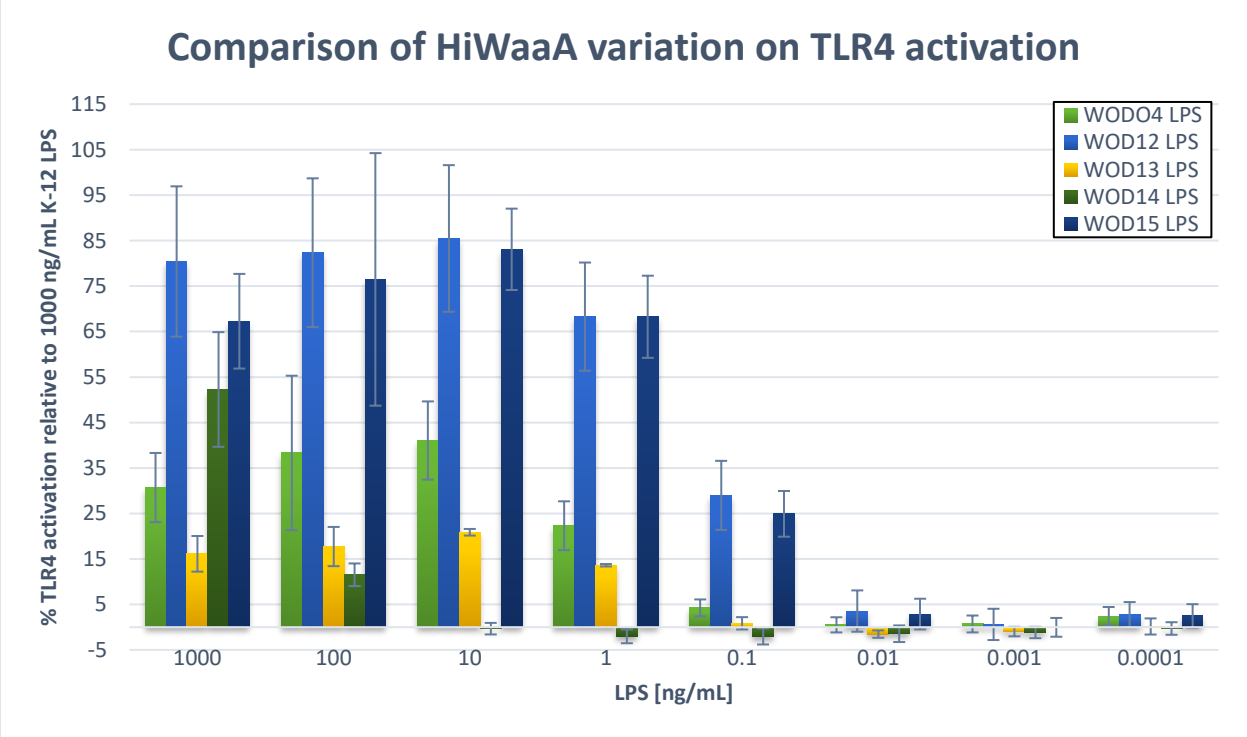
U. Mass spectra of WOD27 with mass units of m/z.

Appendix C

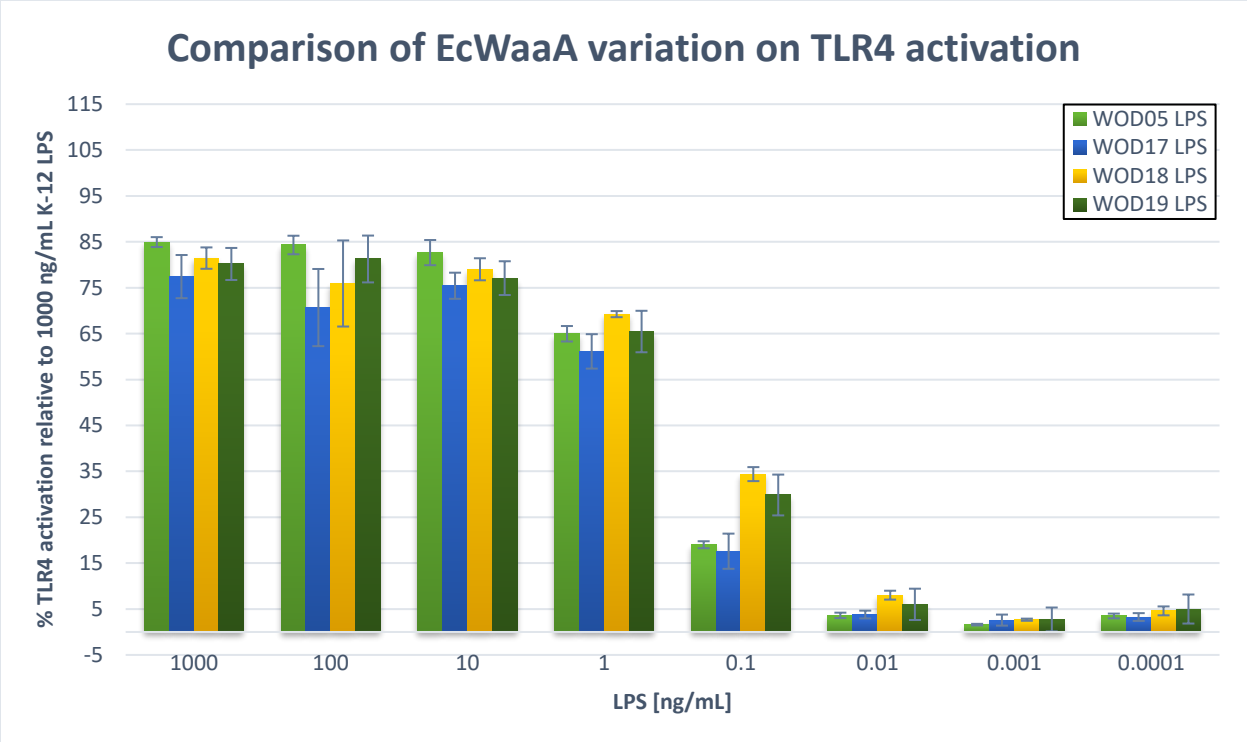
hTLR4 ACTIVATION BY VARIABLE LPS OF WOD02 DERIVATIVES

Figure C.1. Activation of hTLR4 by LPS from WOD02 derivatives. Measure of hTLR4 activation by SEAP secretion in triplicate. Total activities are compared to commercial full-length *E. coli* K-12 LPS (100% activation). Samples include Rasor02 with (A) HiWaaA, (B) EcWaaA, (C) CtWaaA, (D) CpWaaA, each with HpWaaC, EcWaaC, CjWaaC, or AaRfaC1.

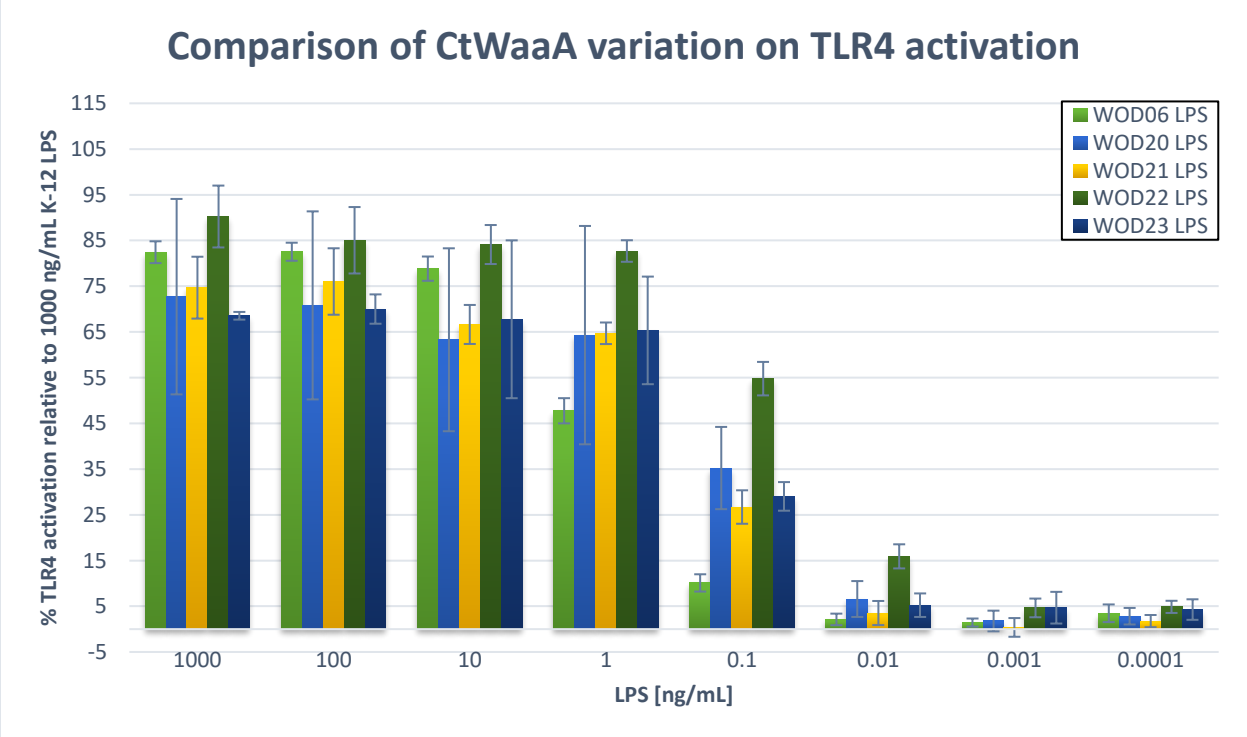
A



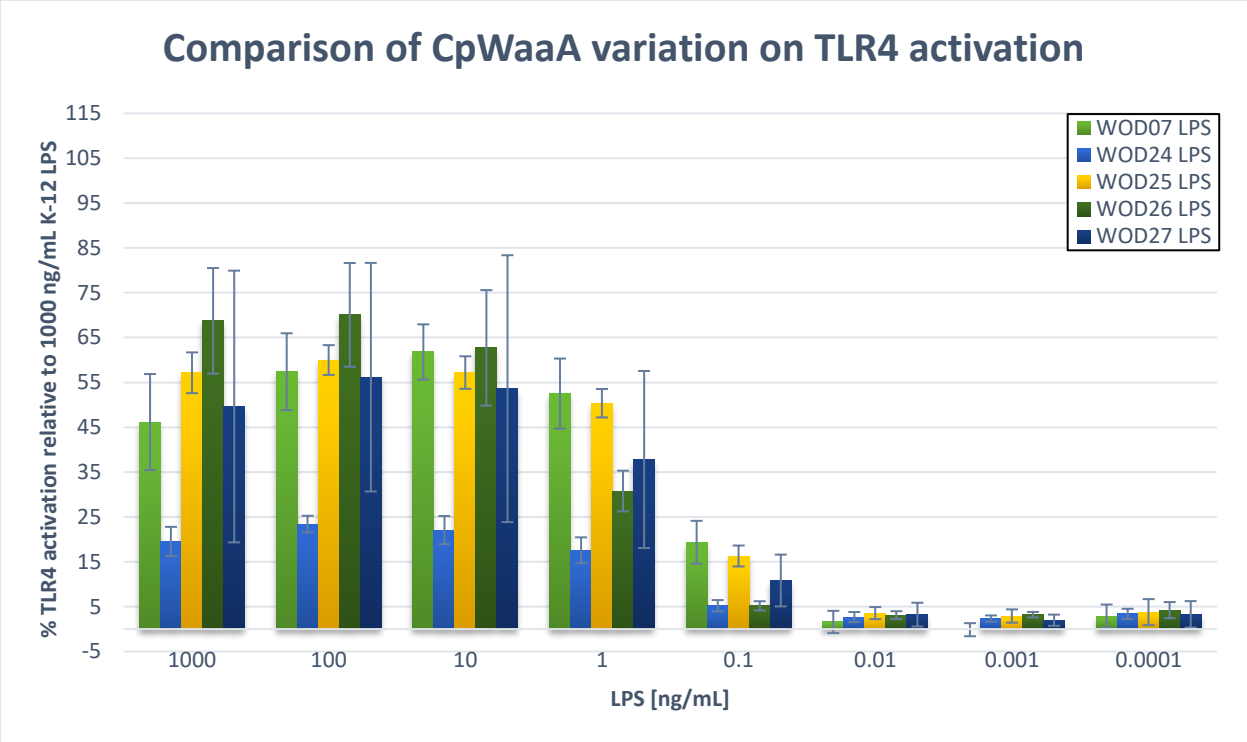
B


Comparison of EcWaaA variation on TLR4 activation
% TLR4 activation relative to 1000 ng/mL K-12 LPS
115
105
95
85
75
65
55
45
35
25
15
5
-5
1000
100
10
1
0.1
0.01
0.001
0.0001
LPS [ng/mL]
WOD05 LPS
WOD17 LPS
WOD18 LPS
WOD19 LPS

C


Comparison of CtWaaA variation on TLR4 activation
% TLR4 activation relative to 1000 ng/mL K-12 LPS
115
105
95
85
75
65
55
45
35
25
15
5
-5
1000
100
10
1
0.1
0.01
0.001
0.0001
LPS [ng/mL]
WOD06 LPS
WOD20 LPS
WOD21 LPS
WOD22 LPS
WOD23 LPS

D


Comparison of CpWaaA variation on TLR4 activation
% TLR4 activation relative to 1000 ng/mL K-12 LPS
115
105
95
85
75
65
55
45
35
25
15
5
-5
1000
100
10
1
0.1
0.01
0.001
0.0001
LPS [ng/mL]
WOD07 LPS
WOD24 LPS
WOD25 LPS
WOD26 LPS
WOD27 LPS

Appendix D

DNA SEQUENCE OF KD8N PRODUCING SYNTHETIC OPERON

Figure D.1. Sequence of synthetic genes for Kd8N generation. Synthetic codon optimized sequence of the engineered Kd8N biosynthetic genes from *S. oneidensis* MR-1

```
   1 TAATACGACT CACTATAGGG GAATTGTGAG CGGATAACAA TTCCCCTCTA GAAATAATTT
  61 TGTTTAACTT TAAGAAGGAG ATATACATAT GGTGGATCAA TCACAACTGC GTCAATGGGG
 121 CTGTAAAGTC ATCGACATCG AAAAATCGGC ACTGGACAAC CTGTATCAAT ATGTTGATTC
 181 TGCCGAATTT GCACAGGCTT GCGAACTGAT TCTGAACTGT TCAGGCAAAG TGATCGTTAT
 241 GGGCATGGGT AAATCGGGCC ATATTGGTAA TAAAATCAGC GCTACCCTGG CCTCTACCGG
 301 TACGCCGGCC TTTTTCGTTC ATCCGGGTGA AGCAAGTCAC GGCGATCTGG GTGTCCTGTC
 361 CGATAACGAC ATTATCCTGG CCATTTCAAA TTCGGGCGAA AGCTCTGAAA TTCTGGCACT
 421 GATGCCGGTG ATCCAGCGTA AAGCTATTCC GGTTATCGCG ATGACCGGTA AACCGGAAAG
 481 CACGATGGCC CGCCTGGCAA AAATTCATCT GTGCATCGAA GTTCCGGAAG AAGCCTGTCC
 541 GCTGGGTCTG GCACCGACCA GTTCCACCAC GGCCACGCTG GTGATGGGCG ATGCGATGGC
 601 CATTGCACTG CTGCAAGCAA AAGGTTTTAC CCGTGATGAC TTCGCTATGA GCCACCCGGG
 661 CGGTGCGCTG GGCCGTAAAC TGCTGCTGAA AGTCTCTAAC GTGATGCATT GCGGTGATGA
 721 CCTGCCGCTG GTTAAACACG ATATTTGTAT CACCGACGCT CTGTATGAAA TTAGCAAGAA
 781 AGGCCTGGGC ATGACGGCGA TTATCGATGA ACAGAACAAA CTGGTGGGCA TTTTCACCGA
 841 TGGTGACCTG CGTCGCGTTA TCGATGCTCA AGTCAATCTG CGTACCACGC CGATTGCGGA
 901 CGTGATGACC CGCAACTGCG TTACCATCAC GGAAAATGTC CTGGCGGCCC AGGCGCTGCA
 961 AGTCATGGAT AGCCGTAACA TTAATGGCCT GATTGTGATC GACAAAGAAA ATCATCCGGT
1021 TGGTGCCCTG AATATGCTGG ACATGGTCAA AGCGGGTGTT ATCTGAGAAT TCTAATACGA
1081 CTCACTATAG GGGAATTGTG AGCGGATAAC AATTCCCCAC TAGAAATAAT TTTGTTTAAC
1141 TTTAAGAAGG AGTATACACA TGAGCAACAA AATTATTAAA CTGGGTTCTA TTGAAATCGC
1201 AAATGACAAA CCGTTCGTCC TGTTCGGCGG CATGAATGTC CTGGAAAGTC GTGATCTGGC
1261 GATGTCCATT GCCGAAACCT ATGCAGAAGT GACGCAGAAA CTGGGCATCC CGTACGTTTT
1321 TAAAGCCTCA TTCGACAAAG CAAACCGTAG CTCTGTGAAT TCGTATCGCG GCCCGGGTAT
1381 GGAAGAAGGT CTGAAAATCT TTGAAGAAAT TAAGAAAACC TTCAACCTGC CGCTGATTAC
1441 CGATGTCCAT GAAACCTATC AGTGTGCCCC GGTGGCCGAA GTGGTTGATA TTATCCAGCT
1501 GCCGGCATTT CTGGCTCGCC AAACCGACCT GGTCGTGGCG ATGGCCAAAA CGGGCGCGAT
1561 CATCAACGTT AAAAAACCGC AATTCCTGGC CCCGCATGAA ATGCGTCACA TTATCACCAA
1621 ATTCAACGAA GCGGGTAATG ATGAAATTAT CCTGTGCGAA CGCGGCTCCT GTTTTGGTTA
1681 TAACAATCTG GTTGTCGATA TGCTGGGCAT GGACGAAATG AAACAGAGCG GTTACCCGGT
1741 CATTTTCGAT GCGACCCATG CACTGCAACG TCCGGGCGGT CGTGCAGACA GCGCTGGCGG
1801 TCGTCGCGCA CAGGCTACCG AACTGGCCCG TTCTGGTATG GCTCTGGGCC TGGCGGGTCT
1861 GTTCATCGAA GCTCATCCGG ACCCGGACAA CGCAAAATGT GATGGTCCGT GTGCACTGCC
1921 GCTGCACCAA CTGGAAAATT ATCTGAAACA GATGAAAGCG ATTGATGACC TGGTGAAAAG
1981 TTTTGAACCG ATTGATACCT CGAAATAAGG AGAGCTCTCT AATACGACTC ACTATAGGGG
2041 AATTGTGAGC GGATAACAAT TCCCCACTAG AAATAATTTT GTTTAACTTT AAGAAGGAGA
2101 TATACACATG CCGCAACAGG GCTTCTACGG TCCGGTTAGC GATGATGTGT GGCAACGTGC
2161 TCAAAAAATT AAACTGCTGA TCTGCGATGT GGATGGCGTG TTTAGCGATG GTCGTATTTA
2221 TCTGAGCAAC TCTGGCGAAG AACTGAAAGC ATTTCATACC CGTGACGGCT ACGGTGTGCG
```

2281 CAGTCTGCTG ACCAGCGGCT TCCACCTGGC AGTGATTACG GGTCGTCAGT CACGCATCGT
2341 TGAAAATCGC ATGACCGCTC TGGGCGTCAC GCATATCTAC CAAGGTGTGG ATAACAAATT
2401 CGTTCCGTAT GAAGAACTGC TGTCGATCTA CAATGTCACC GCGGAAGAAG TGGCCTACAT
2461 TGGCGATGAC ATCGTTGATC TGCCGGTCAT GAACGTGGTT GGCCTGGCAG TCTGCGTGGC
2521 TGACGGTCAT CCGTATGTTC GTCAGCATGC GCACTTCGTC ACGCAACTGA ATGGCGGTCA
2581 CGGTGCCCTG CGCGAACTGA CGGACCTGCT GCTGCTGTCG CAAAATAAAT TCACCTCGGC
2641 TCACGGCATG TCTATCTAAG GAAGATCTGT CGACTCTAAT ACGACTCACT ATAGGGGAAT
2701 TGTGAGCGGA TAACAATTCC CCCCTAGGAA TAATTTTGTT TAACTTTAAG AAGGAGATAT
2761 TAATATGTCG TTCAAAAACT TCAAAGTGGT GGAAAAAATG ATCTTCGGTC GTGGCTCGTT
2821 CGTTCAACTG GATGATGTCC TGGCGGCTCA GCGTAAAGCG GATGACGATT TTGTGGTTTT
2881 CCTGGTCGAC GATGTGCATC AGGGTAAACC GCTGGAAGCA CGCATTCCGG TGAAAGCTCA
2941 GGATCTGCTG ATCTGGGTTA ACGTCGACGA AGAACCGAGC ACCATTCAGA TCGATGCGCT
3001 GACGGAACAG GTTCAAGCCT TTAATGGTAA ACTGCCGGTC AGCGTCGTGG GCCTGGGCGG
3061 TGGCTCTACC ATGGATGTTG CGAAAGCCGT CAGTCTGATG CTGACGAACC CGGGTGGCTC
3121 CGCGATGTAC CAGGGTTGGG ATCTGATCAA AAAACCGGCC GTTCATCACA TTGGCATCCC
3181 GACCATCAGC GGTACGGGCG CAGAAGCTTC GCGTACCGCA GTGCTGTGCG GTCCGGTTCG
3241 CAAACTGGGC CTGAACAGCG ACTATACGGT CTTTGATCAG ATTATCATGG ACTCTGAACT
3301 GATTGATGGT GTTGAAACCG ACCAATGGTT CTACACGGGC ATGGATTGCT ATATCCATTG
3361 TGTGGAATCA CTGGAAGGCA CCTTTCTGAA TGAATTCTCG AAAGCGTACG CCGAAAAAGC
3421 CATGGATCTG TGTCGTCAGG TTTATCTGGA AGACCATCCG GAAAAAGACG ATAAACTGAT
3481 GATGGCATCA TTTATGGGTG GCATGAGCAT TGCTTACTCT CAAGTCGGTG CATGCCACGC
3541 TGTGAGTTAT GGTCTGTCCT ACATCCTGGG CTATCATCAC GGTATTGGCA ACTGTATCGC
3601 ATTTGATGTG CTGGAAGAAT TCTATCCGGA AGGCGTTGCT GAATTCCGCC TGATGATGAA
3661 AAAACACAAC ATTACCCTGC CGAAAAATAT CTGCAAAGAC CTGCCGGATG AAACCATTGC
3721 GAAAATGGTG GCCGTTACGA AATCCATGGG TCCGCTGTGG GCCAATGTGT ATGGCCCGAC
3781 CTGGGAAGAA AAAGTGACCG ATGAAATGCT GACCGCCCTG TTCCGCCGTA TCTGAGGATC
3841 TAATACGACT CACTATAGGG GAATTGTGAG CGGATAACAA TTCCCCACTA GAAATAATTT
3901 TGTTTAACTT TAAGAAGGAG ATATACATAT GCCGGGCTTT GAACTGTTTG GTCCGGAAGA
3961 AAAACAAGAA GTCGCAGACG TGATGGAACA TGGCTTTACC TTTCGCTACA ACTTTGATCA
4021 CATGCGTAAC GACCGCTGGA AAACCCGTGA TATGGAACAG CTGCTGTGCG AAAAAATGAA
4081 TGTGAAACAT GCACACCTGC TGAGCTCTGG TACCGCGGCC CTGCAAACGG CGATGATGGC
4141 AGCTGGCATT GGTGCCGGCG ATGAAGTGAT CGTTCCGCCG TTTACCTTCG TCGCGAGTGT
4201 GGAAGCCATT TTTATGGCAG GCGCTGTTCC GATTTTCGCC GAAATCGACG AAACGCTGTG
4261 CCTGTCCCCG GAAGGTATTG AAGCAGTCAT CACCCCGCGC ACGAAAGCTA TTAACCTGGT
4321 GCACATGTGT GGCTCAATGG CAAAAATGGA TGAAATCAAA GCTATCTGCA AAAAACATAA
4381 TCTGGTTCTG CTGGAAGACG CATGTCAGGC TATTGGCGGT TCGTATAAAG GCCAAGCGCT
4441 GGGTACCATC GGCGATGTTG GTTGCTACAG TTTTGACTCC GTCAAAACCA TTACGTGTGG
4501 TGAAGGCGGT GCCGTCATTA CCAACAATAC GGAAATCTAT GATAACGCAC ACATGTTTTC
4561 CGATCATGGC CACGACCATA TCGGCAAAGA CCGTGGTGCG GAAAGCCATC CGATTATGGG
4621 CCTGAACTTC CGCATCTCTG AAATGAACGC CGCCCTGGGT CTGGCCCAGC TGCGTAAACT
4681 GGATACCATC ATCGACATCC AACGCAAAAA CAAGAAAGCG ATCAAAGATG CGATGGCCAG
4741 TATCCCGGAA GTGTCCTTTC GTGAAATCCC GGACCCGGAA GGCGACTCAG CAGGTTTTCT
4801 GTCGTTCATG CTGCCGACCG AAGCTCGCAC GCAGGAAATT AGCAAAAAAC TGGCAGCTAA
4861 CGGCGTGGAT GGTTGTTTCT ATTGGTATGT GAACAATTGG CACTATCTGA AAAATTGGAA
4921 ACATATTCAG GAACTGAAAG CCCCGGCCGC ACTGCCGATT ACCCTGATCG CCGATCGTCC
4981 GGACTACACG CAAATTTCAG TTCCGAAATC GGATGCGATC ATGAGCCGCA CCATTTCTAT
5041 GCTGATCAAA CTGAGCTGGA CGGACGCTCA GATTGCAGAA CGCATTGAAA ACATTAAGAA
5101 GGCATTCGCA CAGTGAGGAT CCAGGAATTC GAGCACCGCT GAGCAATAAC TAGCATAACC
5161 CCTTGGGGCC TCTAAACGGG TCTTGAGGGG TTTTTTG