University of the Basque Country
Sciences and Technology Faculty

# STRUCTURAL BASIS FOR SUBSTRATE RECOGNITION AND CATALYSIS OF THE MYCOBACTERIAL ACYLTRANSFERASE PATA

by

Montse Tersa Peñacoba

Thesis supervisors:
Marcelo E. Guerin (Director)
David Albesa Jové (Co-director)

Structural Glycobiology Group
CIC bioGUNE
Center for Cooperative Research in Biosciences

Montse Tersa Peñacoba, Bizkaia (Spain) 2017

Distributor: Department of Biochemistry and Molecular Biology – University of the Basque Country

# Acknowledgements

A mi director de Tesis Doctoral, Marcelo E. Guerin, gracias por darme la oportunidad de formar parte de tu laboratorio, donde a lo largo de estos años he crecido tanto a nivel profesional como personal. A mi codirector de tesis, David Albesa, que tanto me ha enseñado con su paciencia infinita.

A la Unidad de Biofísica, el departamento de Bioquímica y Biología Molecular de la Universidad del País Vasco (UPV/EHU) y el CIC bioGUNE, instituciones donde he llevado a cabo este trabajo. A la Fundación Biofísica Bizkaia (FBB) y el CIC bioGUNE por su financiación.

To Prof. Lena Mäler and Jobst Liebau from Department of Biochemistry and Biophysics of the Arrhenius Laboratory at Stockholm University for their kind support during my stay and the valuable scientific discussions. I would like also to thank to the EMBO Organization for having awarded me with an EMBO Short Term Fellowship.

A todos los compañeros del *Structural Glycobiology Group* (SGP): Javier O. Cifuente, Mª Ángela Sainz, Natalia Comino y Alberto Marina gracias por todo vuestro apoyo, ayuda y consejos, ya firmo por encontrarme con gente como vosotros en el camino. A Beatriz Trastoy, muchas gracias por todo lo que me has enseñado, dejo a PatA en buenas manos. Gracias también a Enea, por sus consejos y alegría. Con especial cariño a Ane Rodrigo, con la que he compartido muchos momentos y risas dentro y fuera del laboratorio. Muchas gracias a todas aquellas personas que en algún momento formaron parte del laboratorio o estuvieron involucradas de alguna manera en esta tesis; os deseo lo mejor.

A las chicas de la Unidad de Biofísica: Ane M., Ariana y Ornella, por siempre estar ahí. Ane M., suerte en tu nueva andadura vikinga; Ariana y Ornella mucho ánimo en este último tramo que os queda.

A la cocinitas de Paula por sus roscones de reyes y los vinitos mañaneros, dale duro a esa tesis y a ver si coincidimos en el futuro. A los compañeros del GarbiGUNE: Alberto, Mª Ángela, Geor, Bego, Amaia, Miguel, Mikel, Carlos, Iker, Nekane, Bea, Ander y Lucia. Gracias por esos desayunos, comidas y fiestones mortales que hacen que sea una pena acabar el doctorado. Os echaré de menos.

A mis amigas de siempre, Eva, Sara, Claudia y Marina. Aunque nos vemos poco cuando lo hacemos parece que no ha pasado el tiempo. Gracias por estar siempre ahí. Os quiero.

Els meus últims agraïments a la meva familia. Als meus pares, per la seva ajuda i comprensió. Sense vosaltres no hauría arribat fins aquí. Gràcies a les meves germanes, Cris i Meri, i al meus cunyats Joan i Pablo. I també als petits de la casa per estimar-me tant: Aucán, Aimé, Jordà, Guillem i Joana.

Por último al amor de mi vida, por ayudarme, apoyarme y hacer de mí una mejor persona. Te quiero mucho.

***…Nobody said it was easy….***

*The scientist, Coldplay*

# Summary

Glycolipids are prominent constituents of biological membranes, primary localized on the surface of archaeal, prokaryotic and eukaryotic cells[1,2]. Glycolipids play a major role in molecular recognition events including cell-cell interactions and host-pathogen interactions, as well as in the modulation of membrane protein function.

Acyltransferases are enzymes that catalyze the transfer of fatty-acid chains from activated fatty-acid donors to a variety of acceptor molecules of different chemical structure and complexity, playing a key role in the biosynthesis of glycolipids[3]. It is worth noting that fatty-acid chains are activated by esterification of their carboxyl groups with either the thiol group of coenzyme A (CoA) or that of the acyl carrier protein (ACP), yielding acyl-thioesters. Acyltransferases are very often required to access a lipophilic acceptor substrate and to catalyze its reaction with a water-soluble donor in the form of acyl-CoA or acyl-ACP derivatives, promoting their association to the lipid bilayer[4]. Therefore, an understanding of the molecular mechanism by which acyltransferases govern substrate recognition and catalysis remains a major challenge.

*Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB), is the second most deadly infectious agent in the world after HIV[5,6]. The extremely robust and impermeable mycobacterial cell envelope is the dominant feature of the biology of *M. tuberculosis*[7–9]. In this context, the cell envelope of *M. tuberculosis* contains a variety of unique glycolipids/lipoglycans that populate both the inner and outer membranes, including phosphatidyl-*myo*-inositol mannosides (PIMs), lipomannan (LM) and lipoarabinomannan (LAM), which are important modulators of the host immune response during infection[8,10]. PIMs are based on a phosphatidyl-*myo*-inositol (PI) lipid anchor carrying one to six Man*p* residues and up to four acyl chains[11,12]. PIMs are considered not only essential structural components of the cell envelope but also the precursors of LM and LAM[13,14]. The early steps of the PIMs biosynthesis involve the consecutive action of three enzymes: the mannosyltransferases PimA and PimB, and the acyltransferase PatA, to form $Ac_1PIM_2$[11,12]. The crystal structures of the mannosyltransferases PimA and PimB show the typical organization and catalytic machinery of GT-B glycosyltransferases[15,16]. However, to date, no structural information is available for PatA, an enzyme that is a member of a large family of acyltransferases for which the molecular mechanism of substrate recognition and catalysis remains not well understood[2,3,17,18]. Specifically, PatA transfers a palmitoyl moiety at position 6 of the Man*p* residue, transferred by PimA, from Palmitoyl-CoA to $PIM_1$ or $PIM_2$, to form $Ac_1PIM_1$ and $Ac_1PIM_2$[15].

This study describes a detailed investigation of the three-dimensional structure, glycolipid acceptor and acyl donor substrate binding, as well as the catalytic mechanism of PatA from *M. smegmatis*. Using a combination of X-ray crystallography, biochemical and biophysical techniques,

quantum-mechanics/molecular-mechanics (QM/MM) metadynamics and chemical synthesis we propose a model for substrate recognition, catalysis and enzyme inhibition. The implications of this model in the comprehension of the early steps of PIM biosynthesis and the mode of action of other members of the bacterial acyltransferase family of enzymes are discussed.

# Resumen

## Introducción

Los glicolípidos son constituyentes importantes de las membranas biológicas, localizados principalmente en la superficie de las células de archeas, procariotas y eucariotas[1,19]. Éstos componentes desempeñan un papel principal en eventos de reconocimiento molecular, incluidas las interacciones célula-célula y las interacciones huésped-patógeno, así como en la modulación de la función de las proteínas en las membranas.

Las aciltransferasas son enzimas que catalizan la transferencia de cadenas de ácidos grasos de dadores de ácidos grasos activados a una variedad de moléculas aceptoras de diferente estructura química y complejidad, jugando un papel clave en la biosíntesis de los glicolípidos[20]. Vale la pena señalar que las cadenas de ácidos grasos se activan por esterificación de sus grupos carboxilo con el grupo tiol del coenzima A (CoA) o el de la proteína transportadora de acilo (ACP), produciendo acil-tioésteres. Las aciltransferasas a menudo se requieren para acceder a un sustrato aceptor lipófilo y para catalizar su reacción con un donador soluble en agua en forma de derivados de acil-CoA o acil-ACP, promoviendo su asociación a la bicapa lipídica[4]. Por lo tanto, la comprensión del mecanismo molecular por el cual aciltransferasas rigen el reconocimiento del sustrato y la catálisis sigue siendo un desafío importante.

*Mycobacterium tuberculosis*, el agente causal de la tuberculosis (TB), es el segundo agente infeccioso más mortal del mundo después del HIV[5,6]. Una membrana celular micobacteriana extremadamente robusta e impermeable es la característica principal de la biología de *M. tuberculosis*[7–9]. En este contexto, la membrana celular de *M. tuberculosis* contiene una gran variedad de glicolípidos/lipoglicanos únicos que ocupan las membranas interna y externa, entre ellos se incluyen los fosfatidil-mio-inositol manósidos (PIM), lipomanano (LM) y lipoarabinomanano (LAM), que son moduladores importantes de la respuesta inmune del huésped durante la infección[8,10]. Los PIM se basan en un anclaje lipídico de fosfatidil-mio-inositol (PI) que puede llevar asociados de uno a seis residuos de Man*p* y hasta cuatro cadenas acílicas[11,12]. Los PIM se consideran no solamente componentes estructurales esenciales de la membrana celular sino también precursores de LM y LAM[13,14]. Los primeros pasos en la biosíntesis de los PIMs implican la acción consecutiva de tres enzimas: las manosiltransferasas PimA y PimB, y la aciltransferasa PatA, para formar $Ac_1PIM_2$[11,12]. Las estructuras cristalinas de las manosiltransferasas PimA y PimB muestran la organización típica y la maquinaria catalítica de glucosiltransferasas de tipo GT-B[15,16]. Sin embargo, hasta la fecha, no hay información estructural disponible para PatA, una enzima que es miembro de una gran familia de aciltransferasas para las cuales el mecanismo molecular de reconocimiento del sustrato y la catálisis aún no esta muy detallado[2,3,17,18]. En concreto, PatA transfiere un palmitato en la posición 6 del

residuo de Man*p*, transferido por PimA, desde palmitoil-CoA a $PIM_1$ o $PIM_2$, para formar $Ac_1PIM_1$ y $Ac_1PIM_2$[15].

Este estudio describe una investigación detallada de la estructura tridimensional, el sitio de unión al glicolípido aceptor y el sitio de unión al sustrato acilo dador, así como también el mecanismo catalítico de PatA de *M. smegmatis*. Usando cristalografía de rayos X, técnicas bioquímicas y biofísicas, metadinámica de cuántica-mecánica/molecular-mecánica (QM/MM) y síntesis química, se propone un modelo para el reconocimiento de sustrato, la catálisis y la inhibición de la enzima. También se discuten las implicaciones de este modelo en la comprensión de los primeros pasos de la biosíntesis de los PIMs y el modo de acción de otros miembros de la familia de enzimas aciltransferasas bacterianas.

## La estructura de PatA

El núcleo central de PatA consiste en seis hojas β rodeadas de hélices-α. Un surco largo y abierto que corre paralelo a la superficie de la proteína contiene el sitio activo. Este surco también muestra un túnel angosto y profundo, en su mayoría hidrofóbico, que se extiende perpendicularmente desde su base hasta el núcleo central de la proteína. La entrada del túnel hidrofóbico muestra varios residuos cargados compatibles para interactuar con el grupo carboxilato de un ácido graso, sugiriendo que el bolsillo podría jugar un papel en la unión del sustrato dador[21].

## El sitio de unión del sustrato dador

Se identificó una molécula de ácido palmítico en los mapas de diferencia de densidad electrónica de las tres estructuras cristalinas resueltas (PatA-C16). La cadena de acilo está profundamente enterrada en el túnel hidrofóbico y orientada con el grupo carboxilato orientado hacia la ranura principal de la proteína. La comparación entre las tres estructuras cristalinas resueltas reveló flexibilidad conformacional en el grupo carboxilato del palmitato, ubicado muy cerca de la cadena lateral de His126[21].

Cómo reconoce PatA al CoA?

Se resolvió la estructura cristalina de PatA en complejo con S-hexadecil Coenzima A (PatA-S-C16CoA), un análogo no hidrolizable de palmitoil-CoA. La cadena de acilo de S-C16CoA se localiza en el túnel hidrofóbico y se sobrepone muy bien con la cadena de palmitato observada en las tres estructuras anteriores[21]. Vale la pena señalar que la posición del grupo carboxilato del palmitato es diferente a la observada para el tio-éster del complejo PatA-S-C16CoA. El 4-fosfopanteteinato de S-C16CoA está claramente definido en el mapa de densidad electrónica, y se encuentra en la entrada del surco principal. La adenosina 3', 5'-difosfato (3', 5'-ADP) del ligando sobresale del núcleo globular y está expuesto al solvente, como se observa en otros enzimas modificadoras de acil-CoA[21].

Para validar aún más el modelo, se diseñan mutantes puntuales para perjudicar la interacción palmitoil-CoA con PatA, dificultando así la formación del complejo, y se mide la actividad enzimática. Los resultados de los ensayos enzimáticos muestran como las mutantes de PatA no pudieron (i) transferir una cadena de palmítico a $PIM_2$, o (ii) hidrolizar palmitoil-CoA. Estos resultados validan el modelo propuesto[21].

## El sitio de unión del sustrato aceptor

Para describir la arquitectura del sitio de unión del aceptor de PatA, primero se trató de cristalizar PatA en presencia de sus sustratos aceptores naturales $Ac_1PIM_1/Ac_2PIM_1$, $Ac_1PIM_2/Ac_2PIM_2$ o sus análogos deacilados. Sin embargo, estos intentos no tuvieron éxito. Por lo tanto, para describir la arquitectura del sitio de unión del aceptor, PatA se co-cristalizó con D-manopiranosa (Man*p*)[22].

La estructura cristalina de PatA co-cristalizada con Man*p* se resolvió con 4 moléculas por unidad asimétrica. La inspección de los mapas de densidad electrónica reveló la presencia del producto 6-O-palmitoil-α-D-manopiranósido (C16-Man) en tres de las cuatro moléculas (cadenas B, C y D) de la unidad asimétrica (PatA-C16-Man). La cuarta molécula (cadena A) mostró (i) una molécula de palmitato libre localizado en el túnel hidrofóbico y (ii) una molécula de β-D-manopiranosa libre, localizada en la cavidad de unión del aceptor. En las tres moléculas en las que se observa el producto PatA-C16-Man, el grupo de palmitato está profundamente enterrado en el túnel hidrofóbico y con el grupo carboxilato orientado hacia el surco principal, como se observa en los complejos descritos anteriormente PatA-C16 y PatA-S-C16CoA, y el anillo Man*p* está ubicado dentro del surco principal. La comparación estructural de las cuatro moléculas de la unidad asimétrica reveló cambios conformacionales sutiles pero importantes en el sitio activo en la cadena A, en la que no se produce reacción. Específicamente, el grupo carboxilato se aleja del centro catalítico como se observa en el complejo PatA-S-C16CoA. El O1 de la Man*p* en la conformación β se estabiliza ahora mediante nuevas interacciones. El O4 de la Man*p* crea un enlace de hidrógeno con una molécula de agua, que a su vez interactúa fuertemente con el grupo carboxilato de Glu149. De manera similar, el O6 de la Man*p* forma un fuerte enlace de hidrógeno con la cadena lateral de His126[21,22].

Para validar el modelo del sitio de unión del aceptor se realizaron estudios de dinámica molecular (MD) y ensayos enzimáticos.

El residuo de Man*p* observado en el complejo PatA-C16-Man se solapó bien con el residuo de Man*p* correspondiente de la forma deacilada de $PIM_2$ obtenida mediante técnicas de dinámica molecular. La superposición estructural de los complejos PatA-C16-Man y S-C16CoA proporciona una buena visión de los sitios de unión de sustratos donadores y aceptores de acilo[22].

Por otro lado, se diseñan tres mutantes puntuales que se predice que impactarán en la interacción $PIM_2$ con PatA y se mide su actividad enzimática. El resultado del ensayo muestra como la

actividad de transferencia del acilo al sustrato aceptor de todas las mutantes de PatA se ve gravemente comprometida, sin embargo, todas conservan la capacidad de hidrolizar el sustrato dador palmitoil-CoA.

## El mecanismo catalítico de PatA

El análisis del surco que corre sobre la superficie de PatA revela un sitio catalítico que recuerda al observado en la familia de enzimas serina proteasas. En todas las estructuras cristalinas de los complejos PatA-C16 y PatA-S-C16CoA, el átomo de oxígeno carboxilato OE2 de Glu200 se encuentra a 2,8 Å del átomo de nitrógeno ND1 del anillo aromático de la His126. Nuestros complejos binarios corresponden muy probablemente al producto de reacción de hidrólisis de palmitoil-CoA, por lo que uno de los átomos de oxígeno que se encuentra en el palmitato que interactúa con His126 muy probablemente provenga de una molécula de agua activada por His126[22].

Las coordenadas S-C16CoA en la estructura cristalina PatA-S-C16CoA se usan para generar coordenadas atómicas de palmitoil-CoA por sustitución de un átomo C16 con un grupo carbonilo seguido de minimización energética. Los cálculos de acoplamiento colocan el resto de Man*p* unido a la posición 2 del inositol en $PIM_2$ con su átomo de O6 posicionado favorablemente para recibir el grupo de palmitato de palmitoil-CoA.

Se propone un modelo en el que la His126 actúa como la base general para abstraer un protón del grupo hidroxilo en la posición 6 del anillo Man*p* unido a la posición 2 del inositol en $PIM_1$ o $PIM_2$, para facilitar el ataque nucleofílico en el tio-éster de palmitoil-CoA. El Glu200 se involucra en un sistema de relé de carga con His126 y el átomo HO6 del resto Man*p*, contribuyendo a la disposición estructural apropiada del anillo de imidazol del residuo de histidina y modulando su pKa para actuar como una base en el primer paso y como un ácido en el segundo paso, proporcionando asistencia protónica al grupo saliente CoA. Vale la pena señalar que el enlace de hidrógeno His126-Glu200 se encontró en una orientación *syn* relativa al carboxilato. El resultado del ataque nucleofílico es un enlace covalente entre el anillo de manosa de $PIM_1$ o $PIM_2$ y el palmitato.

Para validar el modelo de mecanismo catalítico propuesto se realizaron estudios de dinámica molecular/mecánica quántica (MD/QM) y ensayos enzimáticos con mutantes de PatA.

Con los estudios de dinámica molecular/mecánica quántica se demuestra que PatA puede acilar $PIM_2$ en un mecanismo de reacción de dos etapas en el que His126 actúa como una base general para desprotonar el grupo hidroxilo de un anillo de manosa en $PIM_2$, facilitando el ataque nucleofílico sobre el tio-éster de C16-CoA. La reacción es asistida por Glu200, que modula el pKa de la histidina central como una base general o ácido durante el ciclo catalítico, de forma similar al sistema carga-liberación de las serina proteasas. Estos resultados también proporcionan evidencias de la formación de un estado de transición de tipo tetraedro, como ya han sido encontradas para el

plegamiento de tipo *hotdog* de las acil-CoA tioesterasas. Estos resultados remarcan la importancia del Glu200 en el mecanismo, y ofrecen detalles valiosos sobre el estado de transición de la reacción que serviría para el diseño de nuevos fármacos quimioterapéuticos[21,22].

Para validar aún más el modelo propuesto, se diseñan mutantes puntuales de los posibles residuos catalíticos Asp131Ala, Glu200Ala e His126Ala, y se mide su actividad. Los resultados muestran la importancia del papel de la His126 y Glu200, ya que su sustitución por alanina inactiva completamente la enzima[21,22].

## Mecanismo de acción de PatA

PatA cataliza un paso crítico en la biosíntesis de los PIM en micobacterias [11,12]. La enzima transfiere un grupo palmítico de un sustrato dador soluble, palmitoil-CoA, a la posición 6 del anillo de Man*p* unido a la posición 2 del *myo*-inositol de los glicolípidos de membrana $PIM_1$ o $PIM_2$[21,23]. Por lo tanto, la interacción de la enzima con la cara citosólica de la membrana plasmática micobacteriana podría ser un requisito para la modificación de $PIM_1$ o $PIM_2$ por PatA. De acuerdo con esta suposición, se observa que PatA co-localiza con la fracción de membrana micobacteriana[23,24]. El análisis de la secuencia de aminoácidos de la enzima revela la falta de un péptido señal o segmentos hidrofóbicos transmembrana, sugiriendo que PatA se asocia a un solo lado de la bicapa lipídica, una característica típica de las proteínas de membrana periféricas y monotópicas[25,26]. El potencial de superficie electrostática de PatA revela varios residuos hidrofóbicos/aromáticos intercalados con residuos cargados positivamente en las proximidades del surco principal[21]. Es importante destacar que el lado opuesto de PatA muestra una superficie cargada negativamente, lo que sin duda contribuye a determinar la orientación correcta de la enzima en la membrana.

A pesar de que la transferencia de acilo se cataliza entre el grupo palmitato de palmitoil-CoA y la manosa de $PIM_1$ o $PIM_2$, los datos experimentales indican que la enzima muestra una necesidad absoluta de ambas cadenas de ácido graso de PI para que suceda la reacción, apuntando a un mecanismo que requiera una interfaz lípido-agua. Las cadenas de ácido graso del sustrato aceptor parecen jugar un papel estructural importante en la generación de un sitio activo competente, ayudando/guiando al grupo glicosílico al entorno apropiado del centro catalítico.

## Similitudes estructurales de PatA con otras aciltransferasas

En el momento en que se publicó la primera estructura cristalina de PatA, la búsqueda de homólogos estructurales usando el servidor DALI reveló solamente una proteína con similitud estructural significativa, la glicerol-3-fosfato aciltransferasa de *Cucurbita moschata* (*Cm*GPAT)[27,28]. Sin embargo, en la actualidad, la búsqueda de homólogos estructurales usando el servidor DALI revela tres proteínas con una similitud estructural significativa, *Cm*GPAT, la aciltransferasa de

*Acinetobacter baumannii* LpxM (*Ab*LpxM)[29], y la 1-acil-glicerol-3-fosfato aciltransferasa de *Thermotoga maritima* (*Tm*PlsC)[30].

Los alineamientos de las secuencias de aminoácidos revelan que las enzimas PatA y *Ab*LpxM están relacionadas lejanamente con las familias CmGPAT y TmPlsC de aciltransferasas bacterianas.

La superposición del dominio α/β de PatA, *Ab*LpxM, *Cm*GPAT y *Tm*PlsC confirma un núcleo estructural común, sugiriendo que éstas podrían formar parte de una nueva familia estructural de aciltransferasas.

En la literatura, alineamientos de las secuencias primarias de varias las familias de aciltranferasas muestran residuos altamente conservados distribuidos en cuatro regiones y denominados como Bloques I-IV[18,31]. El alineamiento de las secuencias primarias sugiere que las familias de aciltransferasas PatA, *Cm*GPAT, *Ab*LpxM y *Tm*PlsC muestran un residuo conservado de aspartato/glutamato que participa en el sistema de carga-liberación con la histidina conservada del motivo $HX_4D/E$ (Bloque I)[17,32], como es el caso de las dos aciltransferasas implicadas en la biosíntesis de ácido fosfatídico *Cm*GPAT y *Tm*PlsC. En cambio, estructuras cristalinas muestran que *Ab*LpxM y PatA parecen usar un residuo de glutamato/aspartato localizado en el Bloque III en lugar del perteneciente al Bloque I, lo que sugiere una divergencia entre esta familia de aciltransferasas.

# Publications

This thesis is based on the following two articles, which will be referred in the **4. Results and Discussion** section. A full copy of those articles have been incorporated as **Annex II** and **Annex III**.

**I.** David Albesa-Jové*, Zuzana Svetlíova*, **Montse Tersa***, Enea Sancho-Vaello, Ana Carreras-González, Pascal Bonnet, Pedro Arrasate, Ander Eguskiza, Shiva K. Angala, Javier O. Cifuente, Jana Korduláková, Mary Jackson, Katarına Mikusová & Marcelo E. Guerin. **Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA.** *Nat Commun.* 11:7:10906. doi: 10.1038. (2016)

**II.** Montse Tersa*, Lluís Raich*, David Albesa-Jové*, Beatriz Trastoy, Jacques Prandi, Martine Gilleron, Carme Rovira, Marcelo E. Guerin. The Molecular Mechanism Of Substrate Recognition And Catalysis Of The Membrane Acyltransferase PatA. *ACS Chemical Biology* **Just Accepted Manuscript**. DOI: 10.1021/acschembio.7b00578 (2017)

# Other publications:

**III.** Albesa-Jové D, Chiarelli LR, Makarov V, Pasca MR, Urresti S, Mori G, Salina E, Vocat A, Comino N, Mohorko E, Ryabova S, Pfieiffer B, Lopes Ribeiro AL, Rodrigo-Unzueta A, **Tersa M**, Zanoni G, Buroni S, Altmann KH, Hartkoorn RC, Glockshuber R, Cole ST, Riccardi G, Guerin ME. **Rv2466c mediates the activation of TP053 to kill replicating and non-replicating *Mycobacterium tuberculosis*.** *ACS Chem. Biol.* 9(7):1567-75. (2014)

**IV.** Albesa-Jové D*, Comino N*, **Tersa M*,** Mohorko E, Urresti S, Dainese E, Chiarelli LR, Pasca MR, Manganelli R, Makarov V, Riccardi G, Svergun DI, Glockshuber R, Guerin ME. **The Redox State Regulates the Conformation of Rv2466c to Activate the Antitubercular Prodrug TP053.** *J Biol Chem*. 290(52):31077-89 (2015)

# Common Abbreviations and Symbols

Å ..... Ångström

λ ..... Wavelenght

s ..... Second

ACP ..... Acyl Carrier Protein

$AcPIM_1$ ..... Acylated Phosphoinositolmonomannoside

$AcPIM_2$ ..... Acylated Phosphoinositoldimannoside

$AcPIM_3$ ..... Acylated Phosphoinositoltrimannoside

$Ac_2PIM_1$ ..... Diacylated Phosphoinositolmonomannoside

$Ac_2PIM_2$ ..... Diacylated Phosphoinositoldimannoside

$Ac_2PIM_3$ ..... Diacylated Phosphoinositoltrimannoside

AFM ..... Atomic force microscopy

AG ..... Arabinogalactan

AGP ..... Peptidoglycan-arabinogalactan complex

AGPAT ..... Acyl-Glycerol-3-phosphate Acyltransferase

AM ..... D-arabino-D-mannan

*ca.* ..... *circa*

CCD ..... Charged-coupled device detector

CMP ..... Cytidine monophosphate

CL ..... Cardiolipin

CoA ..... Coenzyme A

C-terminal ..... Carboxyl terminal domain

DAG ..... Diacylglycerol

DAT ..... Diacyltrehalose

DDM ..... n-dodecyl-β-D-maltoside

DHAPAT ..... Dihydroacetone phosphate acyltransferase

DMSO ..... Dimethyl sulfoxide

DOPC ..... 1,2-Dioleoyl-*sn*-glycero-3-phosphocholine

DTNB ..... 5,5'-dithiobis-(2-nitrobenzoic acid)

DTT ..... dithiothreitol

EDTA ..... Ethylenediaminetetraacetic acid

EM ..... Electron microscopy

FA-CoA ..... Fatty acyl-CoA

FAex ..... Exogenous fatty acids

FAK ..... Fatty acid kinase

FakA ..... Fatty acid kinase domain protein

FakB ..... Fatty acid kinase binding protein

FAS-I ..... Fatty acid synthase I

FAS-II ..... Fatty acid synthase II

G3P ..... Glycerol-3-phosphate

Gal ..... Galactose

Galf ..... β-D-galactofuranose

GalNAc ..... *N*-acetyl galactosamine

GlcNAc ..... *N*-acetylglucosamine

Glc ..... D-Glucose

GPAT ..... Glycerol-3-phosphate Acyltransferase

GT ..... Glycosyltransferase

GT-B ..... Glycosyltransferase B fold

HCl ..... Hydrogen chloride

Hep ..... L-glycero-D-mannose-heptose

HEPES ..... 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HIV ..... Human immunodeficiency virus

$HX_4D$ ..... His-$X_4$-Asp motif

KDO ..... 3-deoxy-D-manno-octulosonic acid

KdsB ..... CMP-kdo synthase

HPLC ..... High performance liquid chromatography

IM ..... Inner membrane

IP ..... Image plate detector

K ..... Kelvin

LAM ..... Lipoarabinomannan

| | |
|---|---|
| LB | Luria-Bertani |
| LCP | Lipidic cubic phase |
| LD | Lipid droplet |
| LDAO | Lauryldimethylamine N-oxide |
| LM | Lipomannan |
| LPA | Lysophosphatidic Acid |
| LPAAT | Lysophosphatidic Acid Acyltransferase |
| LPEAT | 2-acylglycero-phosphatidylethanolamine acyltransferase |
| LPS | Lipopolysaccaride |
| MAD | Multi-wavelength anomalous |
| MAG | Monoacylglycerol |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization-Time-Of-Flight |
| Man | Mannose |
| ManLAM | Mannosylated lipoarabinomannan |
| Man*p* | D-mannopyranose |
| *M. bobis* BCG | *Mycobacterium bobis* bacillus Calmette-Guerin |
| MD | Molecular Dynamics |
| MDO | Membrane-derived oligosaccharides |
| Meso-DAP | meso-diaminopimelate |
| MIR | Multiple isomorphous replacement |
| MM | Molecular Mechanics |
| Ms | *Mycobacterium smegmatis* |
| MR | Molecular replacement |
| *Mtb* | *Mycobacterium tuberculosis* |
| MPD | 2-methyl-2,4-pentanediol |
| Mur | Muramic acid |
| MurG | Pyrophosphoryl-undecaprenol N-acetylglucosamine transferase |
| NaCl | Sodium chloride |

NMR ..... Nuclear magnetic resonance

N-terminal ..... Amino terminal domain

OD ..... Optical density

OM ..... Outer membrane

OMP ..... Outer membrane protein

ON ..... Over night

PA ..... Phosphatidic acid

PAD ..... Pixel array detector

PAT ..... Polyacyltrehalose

PatA ..... Phosphatidyl-myo-inositol acyltransferase A

PDB ..... Protein data bank

PE ..... Phosphatidylethanolamine

PEG ..... Polyethylene glycol

PG ..... Peptidoglycan

PI ..... Phosphatidyilinositol

PIM ..... Phosphoinositolmannoside

$PIM_1$ ..... Phosphoinositol monomannoside

$PIM_2$ ..... Phosphoinositol dimannoside

$PIM_6$ ..... Phosphoinositol hexamannoside

PimA ..... Phosphatidyl-*myo*-inositol mannosyltransferase A

PimB' ..... Phosphatidylinositol mannosyltransferase B'

PMf ..... Plasma membrane free

$PO_4$ ..... Phosphate

PPEtn ..... Ethanolamine pyrophosphate

PS ..... Phosphatidylserine

QM/MM ..... Quantum Mechanics/Molecular Mechanics

Rha ..... α-L-rhamnose

r.m.s.d ..... Root Mean Square Deviation

SAD ..... Single-wavelength anomalous

SD ..... Standard Deviation

SGL ..... Sulfoglycolipid

S-HexaCoA ..... S-hexadecyl coenzyme A

SIR ..... Single isomorphous replacement

SLD ..... Small lipid droplet

$S_N1$ ..... Unimolecular nuclephilic substitution

$S_N2$ ..... Bimolecular nuclephilic substitution

$S_ni$ ..... Intramolecular nuclephilic substitution

SUV ..... Small Unilamellar Vesicles

T ..... Temperature

TAG ..... Triacylglycerol

TB ..... Tuberculosis

TMD ..... Trehalose dimycolates

TMH ..... Trans-membrane helix

TMM ..... Trehalose monomycolates

Tris ..... Tris(hydroxymethyl)aminomethane

Trp ..... Tryptophan

UDP ..... Uridine diphosphate

UV ..... UltraViolet

V/V ..... Volume/volume

WT ..... Wild Type

W/V ..... Weight/volume

XFEL ..... X-ray free electron laser

# INDEX

Chapter 1

# INTRODUCTION

# 1. INTRODUCTION

Long-chain fatty acids play a central role in a multitude of important biological processes in all living organisms. They are prominent constituents of biological membranes, mainly in the form of phospholipids, allowing cells to be functionally constituted and differentiated from the environment[20]. Long-chain fatty acids are used as energy storage and metabolic intermediates as well as being modulators of signal transduction pathways. Moreover, the attachment of fatty acids to proteins and glycans generates a significant amount of structural diversity in biological systems[33]. This structural information is particularly apparent in molecular recognition events including cell-cell interactions during critical steps of development and host-pathogen interactions.[20]

Fatty acids are usually activated for subsequent reactions by esterification of their carboxyl groups with the thiol group of coenzyme A (CoA) or that of the acyl carrier protein (ACP), yielding acyl-thioesters. In general, there are two basic routes to provide long-chain fatty acids: via *de novo* fatty acid synthesis from the central metabolite acetyl-CoA, yielding acyl-ACPs, or via uptake of exogenous fatty acids or other compounds that are converted to fatty acids such as alkanes and their conversion to acyl-CoAs by acyl-CoA synthetases[34,35]. Acyltransferases or transacylases use these activated acyl chains and transfer them to a variety of different substrates or polymerize them to lipophilic storage compounds. Lipases form another important enzyme class dealing with fatty acyl chains; however, they cannot be regarded as acyltransferases in the strict sense since they employ a completely different enzyme mechanism to cleave and transfer fatty acids, which employs (i) a catalytic triad and (ii) the formation of covalent intermediate[3].

The aim of this introduction is to provide an overview of those bacterial acyltransferases that (like the mycobacterial acyltransferase of study in this thesis PatA) (i) employ a conserved motif with an invariant histidine residue followed by an acidic aminoacid residue, and (ii) their catalytic mechanism is characterized by a non-covalent transition state. In particular, this group of enzymes includes:

- Glycerol-Phosphate biosynthesis acyltransferases,
- Lipid A biosynthesis acyltransferases,
- and a Mycobacterial Phosphatidyl-*myo*-Innositol Mannosides biosynthesis acyltransferase, PatA.

Interestingly, those enzymes apart from sharing a conserved catalytic mechanism, the amino-acid sequence alignment revealed that PatA has strong resemblance with the Lipid A biosynthesis acyltransferases LaxIy and LpxM[36,37], and that they are distantly related to the glycerol-3-phosphate (GPAT) and lysophosphatidic acid (AGPAT or LPAAT) families of bacterial acyltransferases[18,38]. In concrete, all those enzymes display highly conserved residues distributed in four regions, and named

as Block I to IV. According to multiple sequence alignments weighted by structural alignment of PatA, a common core can be defined.

## 1.1. Glycerol-Phosphate Acyltransferases

Phosphatidic acid (PA) is the universal precursor required for the production of glycerolipids. Glycerolipids include both non-polar lipids and polar phospholipids. Non-polar lipids are generally divided into three different classes depending on the number of fatty acids attached to glycerol: monoacylglycerol (MAG), diacylglycerol (DAG) and triacylglycerol (TAG). PA can be diverted to triacylglycerol (TAG) synthesis, or activated to form cytidine diphosphate-diacylglycerol (CDP-DAG), which is the precursor for the synthesis of phospholipids. Therefore, PA is an important branch point for the synthesis of TAG and phospholipids[39]. An alternative pathway for PA synthesis is phosphorylation of DAG by DAG kinase **(Figure1)**[31].

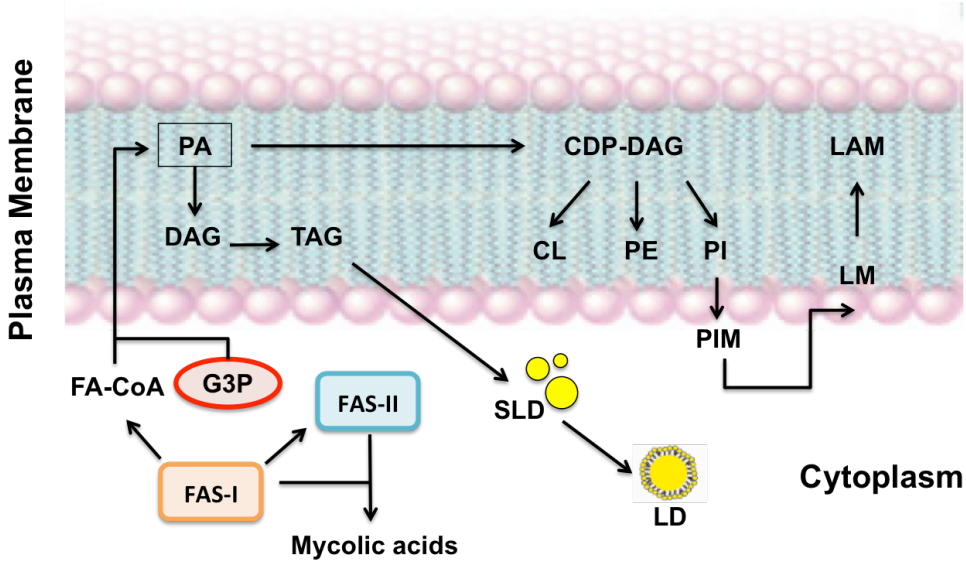



**Figure 1: Mycobacterial plasma membrane with flow of key metabolic pathways.** PA is converted in phospholipids or in TAG. SLD (small lipid droplet); LD (lipid droplet); FA-CoA (fatty acyl-CoA); PIM (phosphatidyl-*myo*-innositol mannoside); LM (Lippomannan); LAM (Lipoarabinomannan); FAS-I (fatty acid synthase I); FAS-II (fatty acid synthase II); CL (cardiolipin); PE (phosphatidylethanolamine); PI (phosphatidylinositol).

PA is the biosynthetic product of the esterification of two fatty acids onto the two hydroxyl groups of *sn*-glycerol-3-phosphate (G3P)[40–43]. Glycerol-3-phosphate is first acylated by glycerol-3-phosphate acyltransferase (GPAT; EC 2.3.1.15) to produce lysophosphatidic acid (LPA), to be further acylated by 1-acylglycerol-3-phosphate acyltransferases or lysophosphatide acyltransferase (AGPAT or LPAAT, respectively; EC 2.3.1.51) to produce PA[44,45].

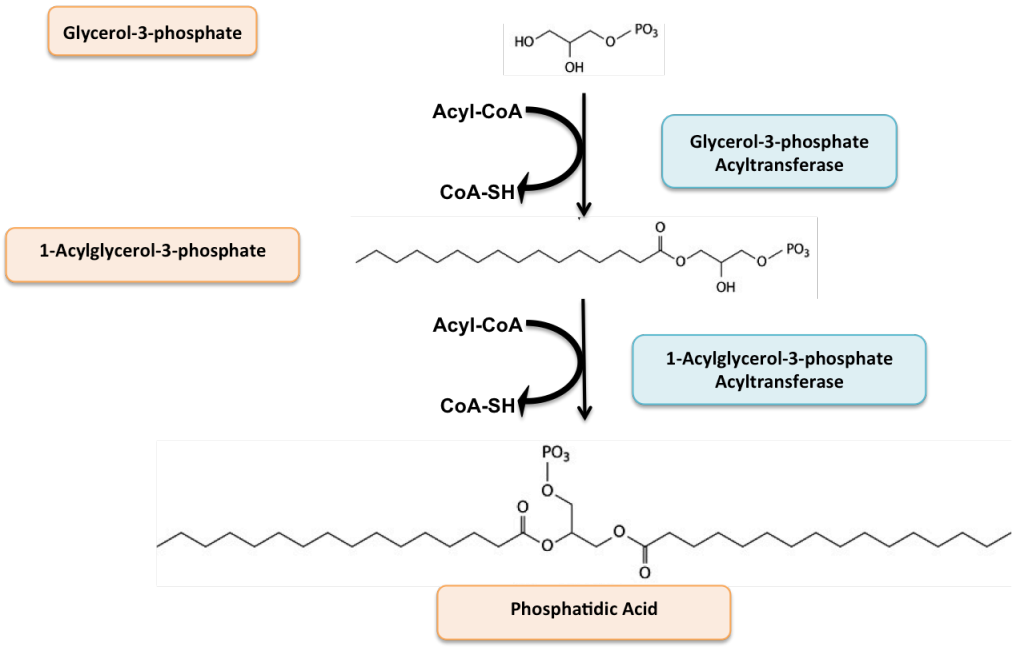



**Figure 2: First two acylation steps in membrane lipid and triacylglycerol synthesis from glicerol-3-phosphate to the central intermediate phosphatidic acid.**

## 1.1.1. Conserved Motifs of Glycerolipid Acyltransferases

Glycerolipid acyltransferases (including GPAT and LPAAT enzymes) from bacteria, plants, yeasts, and animals all share the highly conserved active-site motif HxxxxD/E or $HX_4D/E$, consisting on an invariant histidine and an acidic residue (aspartic/glutamic) separated by four less-conserved amino acids. The highly conserved region containing the $HX_4D$ sequence is referred to as Block I. Furthermore, there are three short regions (Blocks II to IV) exhibiting high conservation in glycerolipid acyltransferases from bacteria, yeasts, and animals (enzymes from plants lack Blocks III and IV)[18].

Four regions of high homology are shared between glycerolipid acyltransferases from different species (**Table 1**). In order to investigate the importance of the conserved amino acids of Blocks I to IV, the *E. coli* PlsB was firstly analyzed as a model for GPAT acyltransferases, and when crystal structure of a distantly related soluble GPAT variant from plant plastids (squash) become available, some more mutagenesis studies on these conserved residues have defined the roles of these blocks in substrate binding and catalysis[17,27,28,45].

The catalytic site constituted by $HX_4D$ motif is located in homology Block I[17,18]. Histidine and aspartate in $HX_4D$ form a catalytic dyad in which the negative residue aspartate forms a charge relay system with histidine that abstract a proton from the hydroxyl group of the acceptor substrate. Subsequently, the nucleophilic oxyanion attacks the carbon atom of the acyl-thioester substrate. The structure confirms that the $HX_4D$ motif forms a charge relay for catalysis similar to serine hydrolases[46]. Blocks II and III participate in G3P binding, with mutations of the arginine residues in the blocks

causing an increase in G3P Km. Block IV is a less conserved hydrophobic patch, with proline as the only invariant residue.

Glycerolipid acyltransferases are responsible for the selection of fatty acids incorporated into membrane phospholipids and are key regulatory points in the pathway[47,48].

| | Sequence | | | |
|---|---|---|---|---|
| **Enzyme** | **Block I** | **Block II** | **Block III** | **Block IV** |
| **E.coli GPAT (PlsB)** | VPCHRSHMDYL | GAFFIRR | YFVEGGRSRTGR | ITLIPIYI |
| **GPAT concensus** | φxx***H***R**S**xφ***D***φφ | **G**xφ**E**I**R**R | φ**F**φ**E****GTR****S**R**x****G**K | φφφφ**P**φφφ |
| **E.coli LPAAT (PlsC)** | IANHQNNYDMV | GNLLLDR | MFPEGTRSR-GR | VPIIPVCV |
| **LPAAT consensus** | φxN***H***Q**S**xφ***D***φφ | **G**xφ**FI****D****R** | φ**FP****EGTR**xxx**G**x | φPφφ**P**φφφ |

**Table 2: Blocks with high degree of conservation in GPAT acyltransferases.** φ, hydrophobic amino acids (Val, Ile, Leu, Phe, Trp, Tyr, or Met). Highly conserved residues are in bold, the catalytically essential histidine and aspartate in Block I in italic, and residues for glycerol-3-phosphate binding are underlined.

## 1.1.2. Two Bacterial Pathways for Phosphatidic Acid (PA) Biosynthesis

There are two distinct pathways to carry out the acylation of the glycerol backbone of PA in bacteria.

- PlsB/PlsC Pathway was initially discovered in *E. coli* and is composed by the PlsB/PlsC acyltransferases. This system is largely limited to Gram-negative bacteria (mostly gamma-proteobacteria), but homologs are also present in mammals[49]. This protein family primarily uses acyl-acyl carrier protein and products of fatty acid synthesis as acyl donors, but may also use acyl-CoA derived from exogenous fatty acids. The PlsB enzyme is responsible for selecting the fatty acids incorporated into the position 1 of phospholipids and is a key point of regulation. The second acyltransferase, PlsC, completes the synthesis of PA, by transfering an acyl group from acyl-ACP or acyl-CoA to position 2 of 1-acyl-glycerol-3-phosphate (**Figure 3**)[49].
- PlsX/PlsY/PlsC Pathway consists of three steps for the formation of PA. It was recently discovered in *Firmicutes*, which include most clinically important Gram-positive bacteria such as *Staphylococcus aureus* and *Streptococcus pneumonia*[49–52]. The first step in this pathway is catalyzed by PlsX, an acyl-ACP:phosphate transacylase, that converts acyl-ACP in acyl-phosphate (acyl-$PO_4$). The second step is catalyzed by the integral membrane protein PlsY, an acyl-$PO_4$-dependent acyltransferase, which transfers the acyl group of acyl-$PO_4$ to position 1 of glycerol-3-phosphate. The last step is similar to the one in Gram-negatives, with the difference that *E. coli*'s PlsC uses acyl-ACP or acyl-CoA as acyl donors, whereas *Streptococcus pneumoniae* and *Bacillus subtilis* PlsC only uses acyl-ACP (**Figure 3**)[50,53].

The glycerol-3-phosphate and 1-acylglycerol-3-phosphate acyltransferases sit at the interface between the soluble type II fatty acid biosynthetic pathway and the creation of a phospholipid molecule that drives membrane expansion. This pivotal position makes glycerolphosphate acyltransferases key regulators of both fatty acid and phospholipid synthesis.

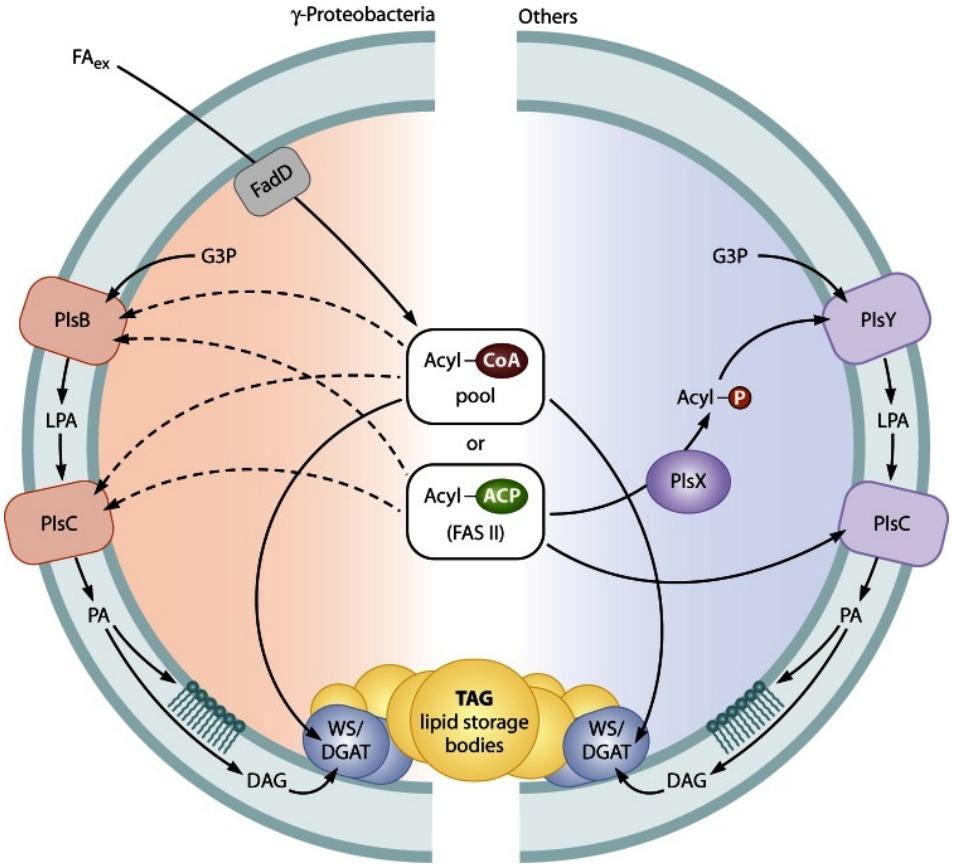



**Figure 3: Overview of pathways for membrane phospholipid and triacylglycerol biosynthesis in bacteria.** G3P, glycerol-3-phosphate; FAex, exogenous fatty acids; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; TAG, triacylglycerol. Modified from[20].

### 1.1.2.1. PlsB/PlsC Pathway of Gram-negative Bacteria PA Biosynthesis

In the first and most detailed analysed model system of *E. coli* and other gamma-proteobacteria, G3P is acylated to form LPA by the acyltransferase PlsB, and further acylated to form PA by the acyltransferase PlsC. PlsB-homologous GPATs have also been identified in mammals and plants[54,55].

#### 1.1.2.1.1. The Glycerol-3-Phosphate (G3P) Acyltransferase PlsB

The first step in PA formation in gamma-proteobacteria is the acylation of the 1-position of G3P by the GPAT named PlsB, the first glycerol-3-phosphate acyltransferase characterized in bacteria[56]. The structure of the soluble GPAT from *Cucurbita moschata* have been solved (*Cm*GPAT; PDB Codes: 1IUQ and 1K30)[27,28]. *Cm*GPAT catalyses the transfer of an acyl chain either from acyl-acyl-carrier protein or acyl-CoA, to the sn-1 position of glycerol-3-phosphate, to form 1-acylglycerol-3-phosphate[57]. Importantly, *Cm*GPAT is able to use palmitoyl-CoA as a donor substrate, using a bi-bi-ordered mechanism[58].

The structure of *Cm*GPAT crystallized as a monomer and it is composed of two domains: a helical domain comprising the first 78 residues of the protein displays a four-helix bundle architecture of unknown function, and an α/β domain, consisting of a nine-stranded continuous β-sheet surrounded by 11 α-helices (**Figure 4**). A combination of site-directed mutagenesis and activity measurements provided experimental support on the location of the acceptor-binding site in *Cm*GPAT[59]. The replacement of Arg235, Arg237 and Lys193 by serine resulted in inactive enzymes. However, the *Cm*GPAT variants retained the ability to bind stoichiometric quantities of acyl-ACPs, consistent with the location of these residues in the positively charged acceptor-binding pocket, and in close proximity to the catalytic $HX_4D$ motif[27,28,59].

The location of the acyl-CoA and acyl-ACP-binding site in *Cm*GPAT has been a matter of strong debate[27,28]. On the basis of sequence conservation analysis, a structural model in which the acyl chain runs over the entrance of the main groove of the protein was first proposed[27]. Molecular surface calculations revealed the existence of three tunnels with sufficient space to accommodate the acyl chain of palmitoyl-CoA[60]. Furthermore, enzymatic analysis of chimeric *Cm*GPAT and *Spinacea oleracea* GPAT revealed that a region comprising residues 128-187 is important for acyl-CoA selectivity. This region completely covered the narrowest and most hydrophobic of the three tunnels which was proposed to be involved in fatty-acid recognition[28]. The catalytic site is located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases[18]. His139 and the adjacent Asp144 were proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the acyl-CoA, similar mechanism that the one observed in serine protease family[17,27,28,61].

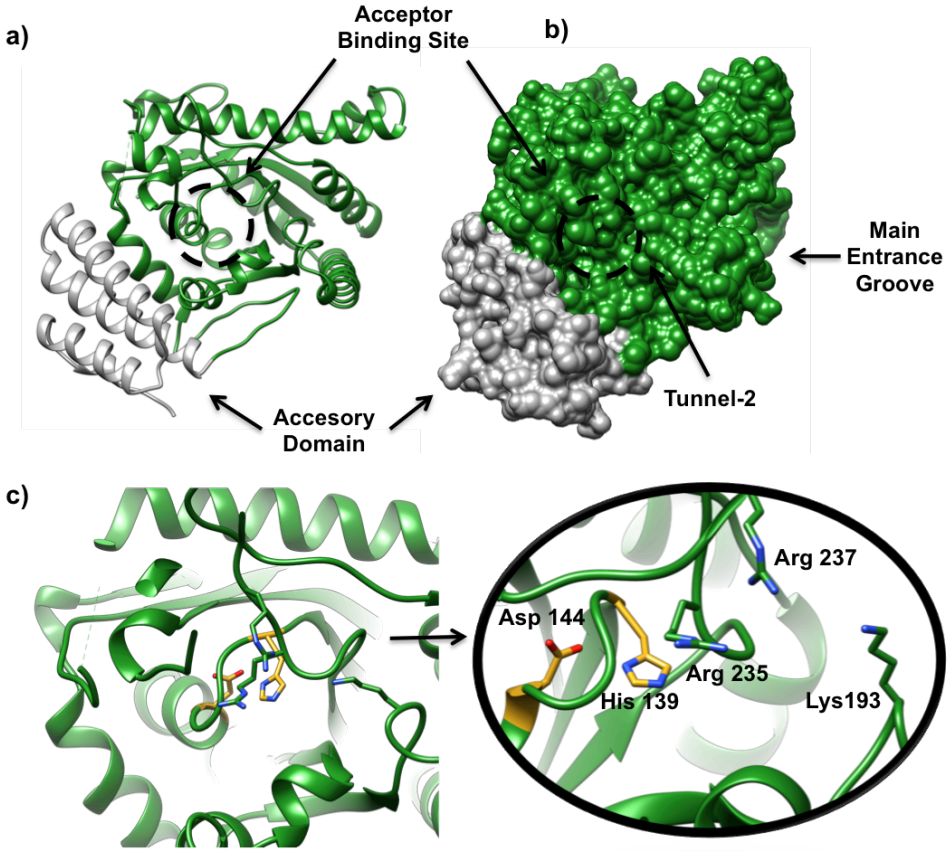



**Figure 4: Tridimensional structure of *Cm*GPAT. a-b)** The active site of *Cm*GPAT is located into a main groove, with a hydrophobic tunnel running perpendicular and deeply buried into the core of the enzyme. The accessory domain is in grey[21]. **c)** Proposed catalytic residues His139 and Asp144 are coloured in yellow, and proposed acceptor-binding site important residues Arg235, Arg237 and Lys193 are shown in green.

### 1.1.2.1.2. The Acyl-G3P Acyltransferase PlsC

The 1-acylglycerol-3-phosphate acyltransferases belong to the evolutionarily conserved lysophospholipid acyltransferase family of intrinsic membrane proteins. This large family is responsible not only for phospholipid synthesis but also for phospholipid remodelling[62–64]. In humans, mutations in AGPAT family members are associated to many human diseases[65]. Very recently, the structure of the PlsC from *Thermotoga maritima* has been solved (*Tm*PlsC; PDB Code 5KYM)[30]. *Tm*PlsC crystalyzed as a monomer and is composed by two domains: the N-terminal domain (residues 1-61) comprises an antiparallel two-helix bundle ($\alpha_1$ and $\alpha_2$), and the C-terminal domain forms an αβ-domain that consists of a seven-stranded β-sheet core surrounded by five α-helices and four $3_{10}$ helices (residues 62-247; **Figure 5 a**)[30].

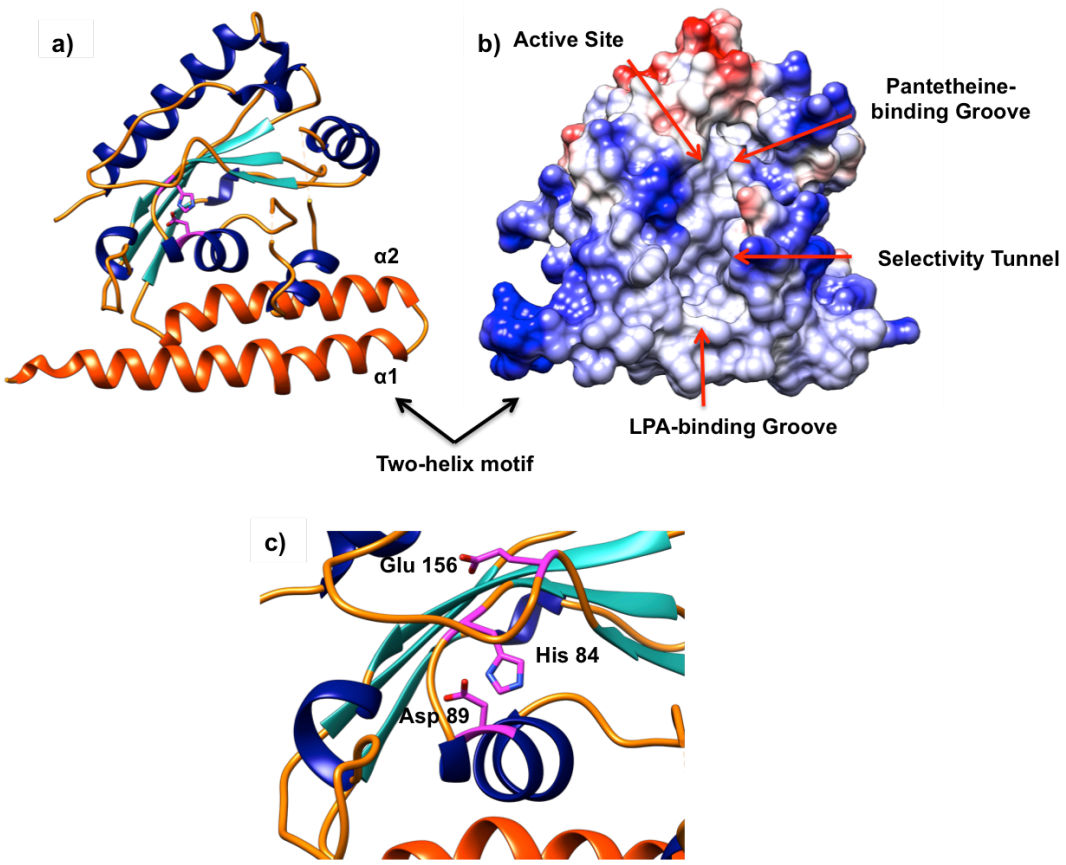



**Figure 5: The tertiary structure of *Tm*PlsC. a)** The overall structure of *Tm*PlsC. The N-terminal two-helix motif is colored orange, and the core catalytic αβ-domain is coloured blue (α-helices), light green (β-strands) and light orange (loops). Proposed catalytic residues are coloured in magenta. **b)** A charged surface representation of *Tm*PlsC showing the substrate (LPA)-binding groove, the entrance to the acyl-chain-selectivity tunnel, the two-helix motif and the active site. **c)** The active site with the catalytic residues His 84 and Asp 89, Glu 156 is proposed to stabilize His 84. Modified from[30].

The two-helix bundle has a very distinctive array of amino acids. The first long α-helix $\alpha_1$ is characterized by an unbroken string of 25 hydrophobic and aromatic residues, a central kink mediated by a pair of glycine residues (Gly25 and Gly26), and basic termini. It is suggested that the kink in helix $\alpha_1$ allows it to enter and exit on the same side of the membrane, while the central hydrophobic

residues and the terminal basic residues interact with the apolar interior and negatively charged surface of the membrane, respectively (**Figure 5 a**)[66,67]. Helix $\alpha_2$ is sandwiched between helix $\alpha_1$ and the αβ-domain, and its exposed surfaces are also populated with basic residues (**Figure 5 a**)[30]. The truncation of this motif reduces the association with the membrane, and a complete removal prevents membrane association.

The α/β-domain includes the location of the conserved catalytic $HX_4D$ motif, and the four conserved blocks within the GPAT and AGPAT family[18,68]. Block I contains the $HX_4D$ motif, and Asn83 and Gln85 flank the catalytic His84 and fix its position via hydrogen bonds with neighboring backbone amide nitrogen atoms. Block III contains Glu156, which forms two stabilizing hydrogen bonds to His84 (**Figure 5 c**). Block IV is within strand $\beta_6$ and interacts with the adjacent strand ($\beta_2$) to cap the active site and further stabilize Block I. Block II is in a separate region and it is suggested that contributes to binding of the phosphopantetheine arm of the acyl donor.

The catalytic site is located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases (**Figure 5 c**)[18]. His84 and the adjacent Asp89 were proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the substrate, similar mechanism that the one observed in serine protease family (**Figure 5 c**) [17,30,61].

### 1.1.2.2. PlsX/PlsY/PlsC Pathway of Gram-positive Bacteria PA Biosynthesis

In Gram-positive pathogens, such as *Streptococcus pneumonia* and *Staphylococcus aureus,* PA is generated by the PlsX/PlsY/PlsC pathway. A comprehensive analysis of bacterial genomes revealed that this alternative PlsX/PlsY pathway is most widely distributed, whereas the PlsB/PlsC pathway seems to be restricted primarily to gamma-proteobacteria. While many of the latter organisms such as *Pseudomonas aeruginosa* possess both alternatives, the *Xanthomonadales* have solely PlsB[56,69].

First, PlsX is a soluble acyl-ACP:$PO_4$ transacylase that converts the acyl-ACP end-products of the *de novo* fatty acid synthesis (FAS-II) to their acyl-$PO_4$ derivatives[70]. These activated fatty acids are then used by the integral membrane acyltransferase PlsY to acylate glycerol-3-phosphate[69,70]. The last acyltransferase of the pathway, PlsC[71], acylates the 2-position and is specific for acyl-ACP thioesters as acyl donors[70] to produce PA. Exogenous fatty acids can access these acyltransferase systems after their uptake by the cell and activation by fatty acid kinase (Fak)[72]. Fatty acid kinase is composed of two subunits: a kinase domain protein (FakA) and a fatty acid binding protein (FakB) that work together to produce acyl-$PO_4$. The acyl-$PO_4$ arising from extracellular fatty acids is either incorporated into the 1-position of G3P by PlsY or converted to acyl-ACP by PlsX. When it is converted to acyl-ACP by PlsX, can be either elongated by FAS-II or used by PlsC[73]. Thus, PlsX is a key enzyme that interconverts the two

acyltransferase acyl donors. Moreover, blocking the PlsX reaction results in the cessation of fatty acid synthesis by FAS-II. Therefore, PlsX seems to be a key regulatory point that synchronizes FAS-II with phospholipid synthesis (**Figure 6**)[53].

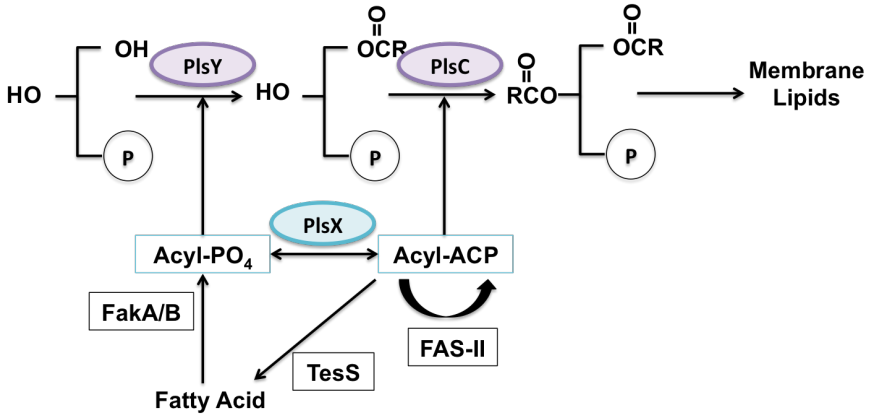



**Figure 6: The role of PlsX in *S. pneumoniae*.** PlsX plays a key role in fatty acid metabolism by interconverting acyl-ACP and acyl-$PO_4$. The liberated free fatty acids are converted to acyl-$PO_4$ by fatty acid kinase (FakA/B) and incorporated into the 1-position of phospholipids by the sn-glycerol-3-phosphate acyltransferase PlsY. The second acyltransferase (PlsC) uses acyl-ACP from FAS-II as the acyl donor.

### 1.1.2.2.1. The Acyltransferase PlsX

PlsX is a acyl-ACP:phosphate transacylase that catalyzes the formation of acyl-$PO_4$ from the acyl-ACP, and plays a key role in fatty acid metabolism by interconverting acyl-ACP and acyl-$PO_4$, so either both molecules can be used as a substrate[70]. Therefore, PlsX links FAS-II pathway to the phospholipid synthesis pathway, and it is crucial for the movement of unsaturated fatty acids into the membrane (**Figure 6**). Nowadays there are two crystal structures available of the acyltransferase PlsX: the structure of the soluble PlsX from *Bacillus subtillis* (*Bs*PlsX; pdb code 1VI1)[74], and PlsX from *Enterococcus faecalis* (*Ef*PlsX; pdb code 1U7N)[75].

PlsX is a α/β/α sandwich that resembles a Rossmann fold and crystallyzed as a dimer. The twisted β-sheet has 11 mostly parallel strands that are flanked by 7 helices on one side and 6 helices on the other side of the sheet (**Figure 7 a**)[74]. The monomer has a well defined hydrophobic core and the structure is stabilized by various intra-subunit salt bridges on the protein surface. The PlsX protein forms an S-shaped dimer using the novel dimerization motif (**Figure 7 b**). Two α-helices $\alpha_9$ and $\alpha_{10}$ form a long hairpin in each monomer that protrudes away from the main domain (**Figure 7 c**). The hairpins associate and form a four-helix bundle ($\alpha_9$, $\alpha_{10}$ and $\alpha_{9'}$, $\alpha_{10'}$) with a well defined hydrophobic core. The loop between helices $\alpha_9$ and $\alpha_{10}$ is disordered in the crystal (**Figure 7 c**)[74]. It has been reported that PlsX associates with the membrane in *Bacillus subtillis*, but the basis for this association has not been explored yet[51].

In the *E. faecalis* PlsX protein, the base of this four-helix bundle includes 12 methionine residues that interact through Van der Waals contacts and form a rather unusual methionine belt (**Figure 7 c-d**). The dimer is additionally stabilized by the interaction of loops and α-helices of the main domain[74].

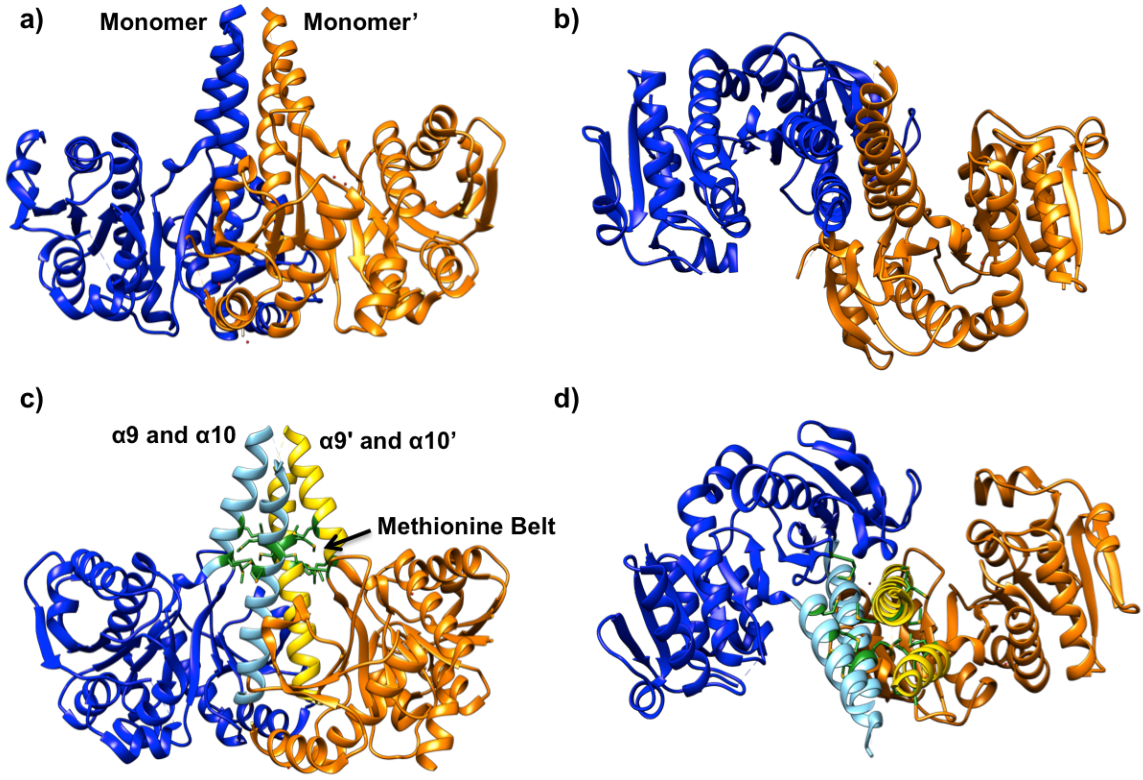



**Figure 7: Structure of *Enterococcus faecalis* PlsX protein dimer. a-b)** Two views of *Ef*PlsX, one monomer is coloured in blue and the other is coloured in orange. **c-d)** α9 and α10 of one monomer are coloured in light blue and α9 and α10 of the other monomer are coloured in yellow. The residues that forms the methionine belt are shown and coloured in green.

In *E. faecalis* the *plsX* gene is co-localized with a bacterial fab gene cluster which encodes several key fatty acid biosynthetic enzymes. The *B. subtilis plsX* is an essential gene and it was annotated based on its similarity to the *E. coli* counterpart[76]. The protein is a member of a large, conserved protein family (Pfam 02504) found exclusively in bacteria. The PlsX sequence homologues include both phosphate acetyltransferases and phosphate butaryltransferases that catalyze the transfer of an acetyl or butaryl group to orthophosphate.

### 1.1.2.2.2. The Gram-positive G3P Acyltransferase PlsY

PlsY is the most widely distributed G3P acyltransferase and it is the sole G3P acyltransferase in Gram-positive bacteria, including medically important bacteria such as *S. pneumoniae* and *S. aureus*[50]. PlsY represents a unique class of GPAT, since not only has a completely different sequence and size - most Gram-negative GPAT PlsB are approximately 90-KDa, whereas Gram-positive GPAT PlsY is approximately 23-KDa - but also uses acyl-phosphate as a novel and unique acyl donor. PlsY protein is an acyl-$PO_4$ dependent G3P acyltransferase. This precursor is provided by the soluble acyltransferase PlsX, previously described, which catalyzes the formation of acyl-$PO_4$ from acyl-ACP (**Figure 6**)[50].

The integral-membrane glycerol 3-phosphate (G3P) acyltransferase PlsY catalyses the committed and essential step in bacterial phospholipid biosynthesis by acylation of G3P, forming lysophosphatidic acid. It contains no known acyltransferase motifs, lacks eukaryotic homologs, and uses the unusual acyl-phosphate as acyl donor, as opposed to acyl-CoA or acyl-carrier protein for other acyltransferases.

Very recently, the structure of *Aquifex aeolicus* PlsY (*Aa*PlsY) has been solved in the presence of substrates and products[77] (PDB Codes 5XJ5, 5XJ6, 5XJ7, 5XJ8, 5XJ9; **Figure 8**). The structure *Aa*PlsY contains a monomer in the asymmetric unit. The overall shape of *Aa*PlsY resembles a funnel composed of three tilted helices ($\alpha_{1\text{-}3}$) sitting in a cup composed of seven TMHs (Trans-membrane helix; **Figure 8**). The membrane-perpendicular $TMH_4$ is sandwiched by two bundles, $TMH_{1\text{-}3}$ and $TMH_{5\text{-}7}$. The striking shortness of $TMH_4$ is compensated by the membrane-embedded Loop3. The overall structure is very compact; apart from two relatively long cytoplasmic loops containing functionally important residues, the helices are stitched together with short linkers (**Figure 8**)[77].

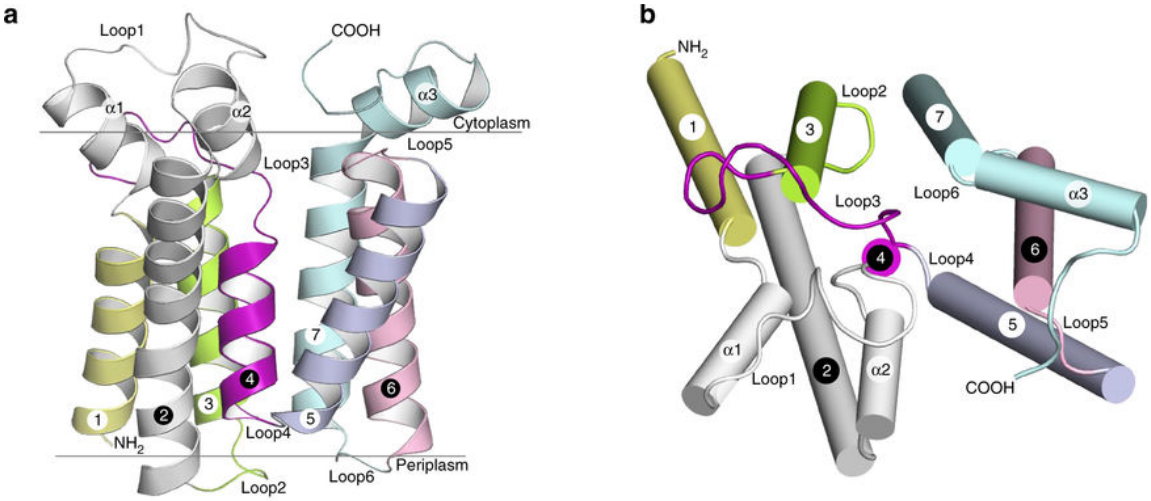



**Figure 8: Crystal structure and membrane topology of *Aa*PlsY. a)** View from the membrane plane. The membrane boundary was defined by the OPM server68. TMHs and loops are appropriately labeled. **b)** View from the cytoplasm. Modified from[77].

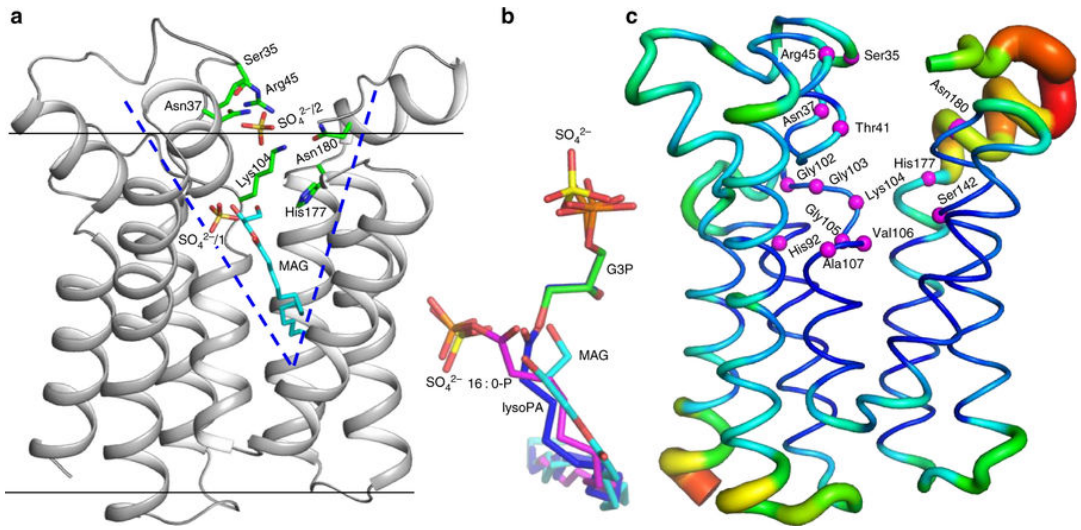



**Figure 9: The active site of *Aa*PlsY. a)** The V-shaped active site hosting a MAG molecule and two $SO_4^{2-}$ ions. Active site residues from a previous report[69], the MAG, and the two sulphates were shown in sticks. **b)** Superimposition of all active site ligands. The MAG, 16:0-P, G3P, and lysoPA are shown in stick representations with the oxygen atoms in red, sulphur atoms in yellow, and carbon atoms in cyan, magenta, green, and blue, respectively. **c)** Putty representation of B-factors of the *Aa*PlsYMAG structure. Blue lines and red tubes indicate low and high B-factors, respectively. The Cα atoms of active site residues are highlighted as magenta spheres. Modified from[77].

Mapping 12 functionally important residues onto *Aa*PlsY reveals a V-shaped cavity that hosts two $SO_4^{2-}$ and a monoacylglycerol (MAG is used for the crystallyzation process by LCP) with well-defined electron density (**Figure 9**). It is noticed that the MAG and $SO_4^{2-}$ /1 are positioned in a way that could mimics the lipid substrate acyl-$PO_4$. As well, the $SO_4^{2-}$ /2 at the membrane interface should mimic the phosphate group of G3P. Therefore, this cavity could be the active site[77](**Figure 9**).

*Aa*PlsY is small (192 residues) and monomeric; yet it crosses membrane seven times and contains pockets that must host two substrates with distinct differences in solubility. This implies an efficient structure build for function. Backbone amides from four consecutive residues (Lys104, Gly105, Val106, and Ala107) in Loop3, the side chain of His92, and two α-helix dipoles, namely the partially positively charged N-termini of $TMH_4$ and $\alpha_2$, form a hydrophilic 'phosphate hole' deep in the membrane, securing the phosphate group of acyl-$PO_4$[77] (**Figure 10 a**). Two backbone amides from Ala40 and Thr41, and the Thr41 hydroxyl, form three H-bonds with the carbonyl. Three water molecules further stabilize the acyl-$PO_4$ head group in this hydrophilic interior. According to the crystal structures solved in complex with different ligands, His92 and Thr41 are the only two residues that contribute to acyl-$PO_4$ binding by side chains. Mutating Thr41 to alanine had severe consequence in catalytic activity, whereas keeping the hydroxyl by serine mutation only reduced half of the activity. Mutation of His92 to alanine also caused a dramatic 80% loss of activity[77] (**Figure 10**).

The *Aa*PlsYG3P G3P structure reveals a G3P pocket composed of a 'phosphate clamp' (Ser35, Arg45, Lys104, and Ans180), which binds the phosphate of G3P, and two side chains (Ser142 and His177), which interact with the 2-OH[77] (**Figure 10 c**). Alanine mutations of the phosphate clamp

residues had a severe consequence on both activity (>92% loss) and G3P Km (5- to 36-fold increase). Mutants Ser35Cys and Ser35Thr show 4% and 17% of the activity compared with the wild type, suggesting that the phosphate clamping is not only hydroxyl selective but also sensitive to the hydroxyl position. Stringent too is the requirement of Arg45, which, when mutated to lysine, caused over 90% loss of activity, and a 14-fold increase of G3P Km. Similarly, mutating Lys104 to arginine reduced the activity to 25%, although with little effect on G3P Km. For Asn180, adding one more carbon to the side chain (Asn180Gln) impaired catalytic activity more than the Asn180Gln mutation, suggesting the G3P-binding pocket is sensitive to subtle changes. Mutating Asn180 to histidine almost abolished activity, probably also for steric reasons. In addition, replacing Asn180 with the spatially similar aspartate caused over 95% loss of activity with a nearly 22-fold increase of Km; presumably, charge-charge repulsion between the acidic residue and the phosphate of G3P impaired the substrate binding. For the binding of the 2-OH group, removing the hydroxyl of Ser142 caused 70% loss of activity. Remarkably, His177 was found to be indispensable with various side chains[77] (**Figure 10 c**).

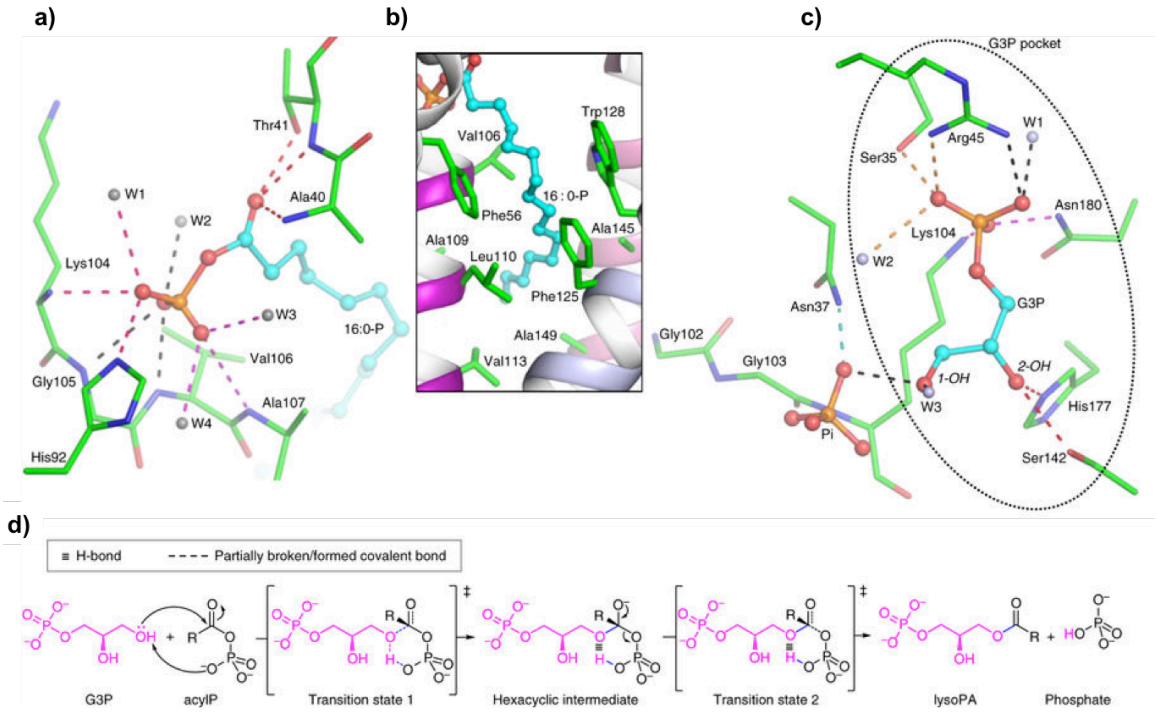



**Figure 10: Interactions between *Aa*PlsY and its substrates acyl-$PO_4$ and G3P**. **a)** Interactions between the acyl-$PO_4$ head group and the enzyme. **b)** The hydrophobic groove that hosts the acyl chain of acyl-$PO_4$. Residues within 4.0 Å of the palmitoyl chain are shown as green sticks. **c)** The G3P binding pocket. Dashed lines in a) and c) indicate distances within 3.2 Å. When appropriate, the dashed lines are colored differently for better visualization. **d)** The proposed 'substrate-assisted catalysis' mechanism for *Aa*PlsY. Modified from[77].

It has been postulated that PlsY acylates G3P in a similar mechanism to the 'Asp-His' dyad thioester GPATs, based on the fact that PlsY contains a critical histidine residue[69] (His177 in *Aa*PlsY). Unlike the thioesters, acyl-$PO_4$ can serve as a proton acceptor itself. The last catalytic model proposed for *Aa*PlsY is a 'substrate-assisted mechanism' for PlsY. G3P 1-OH is clamped in close proximity to the acyl-$PO_4$ phosphate, facilitating its nucleophilic attack on the carbonyl carbon. A chemically favorable hexacyclic transition state then forms, followed by a nucleophilic acyl substitution process through formation and collapse of the carbonyl tetrahedron. The oxyanion intermediate is probably stabilized by the phosphate hole and Asn37. Consistent with this role, Asn37 interacts with neither substrates; yet its alanine mutation caused 80% loss of activity, with no significant G3P Km defects. Moreover, substitution of Asn37 with the spatially similar aspartate almost abolished activity; presumably, aspartate interferes with the intermediates by charge-charge repulsion[77].

#### 1.1.2.2.3. The Acyl-G3P Acyltransferase PlsC

The last step of the pathway is similar to the one in Gram-negative bacteria, with the difference that whereas PlsC from *E. coli* uses acyl-ACP or acyl-CoA as acyl donors, in *S. pneumoniae* and *B. subtilis* PlsC only use acyl-ACP[50,51].

# 1.2. Acyltransferases In Cell Membrane Biosynthesis

## 1.2.1. The Cell Envelope in Gram-Negative Bacteria

Gram-negative bacteria have a complex envelope architecture. Their envelope contains two membranes, different in structure and composition, which delimit an aqueous compartment, the periplasm, containing a thin layer of peptidoglycan (**Figure 9**).

The cytoplasmic or inner membrane (IM) is a symmetric bilayer composed of glycerophospholipids whereas the outer membrane (OM) is highly asymmetrical: it contains phospholipids in the inner leaflet while the outer monolayer mainly consists of lipopolysaccharide (LPS), a glucosamine-based saccharolipid (**Figure 11**)[78]. IM and OM also differ with respect to their integral membrane proteins composition: in the IM integral proteins span the membrane as α-helices almost entirely composed of hydrophobic residues, while the vast majority of integral proteins in the OM, outer membrane proteins (OMPs), consist of amphipathic β-strands which adopt a β-barrel structure[79,80]. A large number of OMPs is constituted by porins, a family of proteins, which form specific and non-specific channels that orchestrate the flux of hydrophilic molecules across the OM[81]. Besides integral membrane proteins, both IM and OM contain lipoproteins that are anchored to the respective membranes via an N-terminal N-acyl-dyacylglycerylcysteine residue[82]. Such lipoproteins are located exclusively at the periplasmic side of the IM, whereas at the OM they can either extend into the periplasm or be exposed at the cell surface (**Figure 11**)[83].

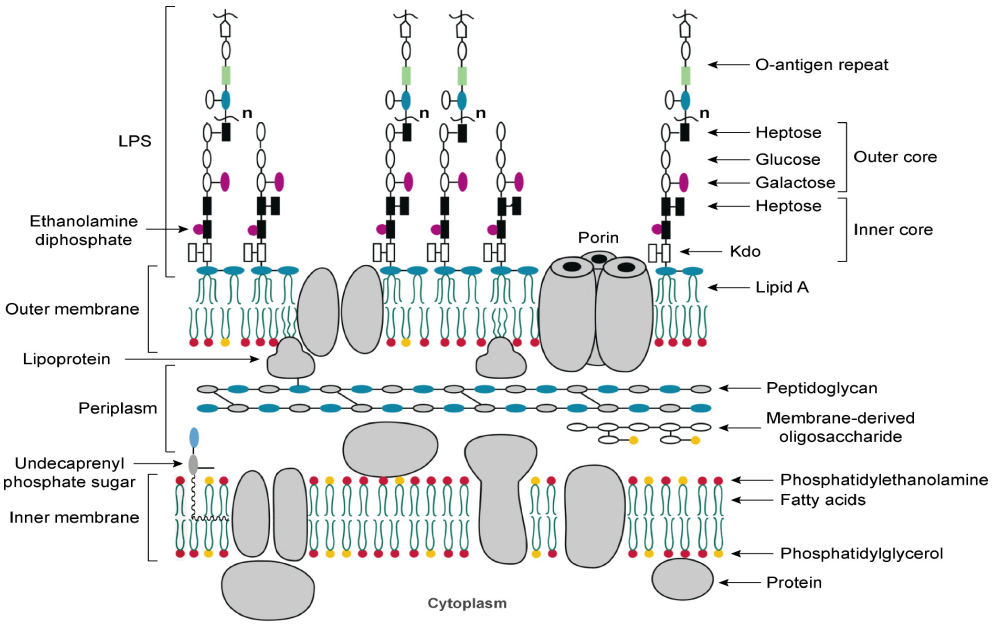



**Figure 11: Schematic structure of the *E. coli* K-12 cell envelope** The structure and biosynthesis of LPS, peptidoglycan, membrane-derived oligosaccharides, lipoproteins and phospholipids. Strains of *E. coli* K-12 normally do not make O-antigen, unless a mutation in the O-antigen operon is corrected. The $KDO_2$-Lipid A region of LPS usually represent the minimal substructure required for growth of Gram-negative bacteria. The red phospholipids represent

phosphatidylethanolamine and the yellow phosphatidylglycerol. Lipopolysaccharide (LPS); membrane-derived oligosaccharides (MDO); Ethanolamine pyrophosphate (PPEtn). Modified from[19].

Lipopolysaccharide (LPS) is a complex glycolipid that forms the major component of the outer leaflet of the OM of most Gram-negative bacteria (**Figure 11**)[84]. It has been estimated that one bacterial cell contains millions of LPS molecules that occupy three quarters of the bacterial surface with the remaining area being filled by proteins[85]. LPS not only provides protection to the cell as a physical barrier but also act as endotoxins in case of Gram-negative pathogens[86,87]. LPS helps stabilize these membranes, protects them from chemical attack, and promotes cell adhesion to various surfaces. It elicits a strong immune response in humans and other animals, and is a main contributor to Gram-negative septic shock, getting detected at picomolar levels by the innate immune system's TLR4 receptor protein[88]. These attributes have made the study of LPS important to the fields of immunology, bacteriology, and drug discovery[19,84,88–90].

LPS molecules are composed of three domains: the Lipid A domain, the core oligosaccharide and the O-antigen polysaccharide (**Figure 12**)[88,89].

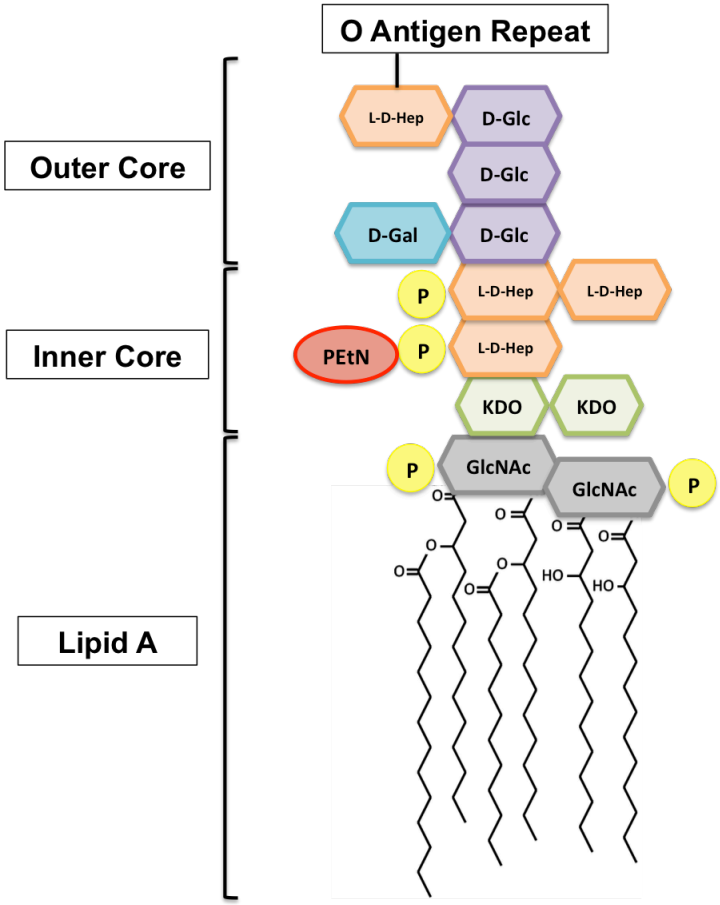



**Figure 12: Structure of *E. coli* LPS.** Smooth-type LPS consists of three distinct domains: Lipid A, inner and outer core oligosaccharides, and the highly variable O-antigen repeat. N-acetylglucosamine (GlcNAc); 3-deoxy-D-manno-octulosonic acid (KDO); L-glycero-D-manno-heptose (Hep); D-glucose (Glc); galactose (Gal). Substitutions of various lipid A and core sugars with phosphate groups (P) and ethanolamine (PEtN) have been indicated with yellow and red ovals, respectively.

In most Gram-negative bacteria, including *E. coli,* Lipid A consists on a bis-phosphorylated glucosamine disaccharide to which six acyl chains are linked by ester and/or amide bonds. The glucosamine and the phosphate head group are covalently connected to a pair of KDO (3-deoxy-D-manno-oct-2-ulosonic acid) sugar residues (**Figure 12** and **Figure 13**)[89]. These KDO residues are connected to several additional sugar residues, and sometimes also to a phosphate (pyrophosphorylethanolamine or phosphorylcholine residues), which together form the core oligosaccharide (**Figure 12**)[88]. The core can be further divided into inner (Lipid A proximal) and outer core. The chemical structure of the outer core is more variable whereas the inner core, that contains at least a residue of KDO and of L-glycero-D-manno-heptose (heptose), is rather conserved across isolates of the same species[91]. The O antigen is the distal portion of the LPS molecule and is constituted of a variable number of repeating units made of one to six sugar residues (**Figure 10**). Notably, the O antigen is absent in many Gram-negative species including pathogens such as *Neisseria meningitidis* and *Bordetella pertussis*[92,93]. Among these three components, the Lipid A moiety is of particular interest because it is the most conserved portion of LPS as well as the only component that is essential for cell viability[86,88,94]. Furthermore, Lipid A is the bioactive component of LPS, which is specifically recognized by the innate immune system. Lipid A is responsible for the endotoxin property of LPS, causing inflammatory response-induced endotoxicity and leading to host tissue damage or septic shock. Thus, knowledge about its biosynthesis and its possible repression represents an important basis for the development of antimicrobial drugs to combat important Gram-negative pathogens[94,95].

Although the overall structure of Lipid A is conserved among Gram-negative bacteria, different species modify this complex molecule upon environmental stimuli by removing or modifying the phosphate groups or acyl chains, respectively. As a consequence, the toxicity caused by Lipid A molecules from different species reveals widely varying extents. These also make its biosynthetic pathway an attractive target for new antibiotics[88,96–100].

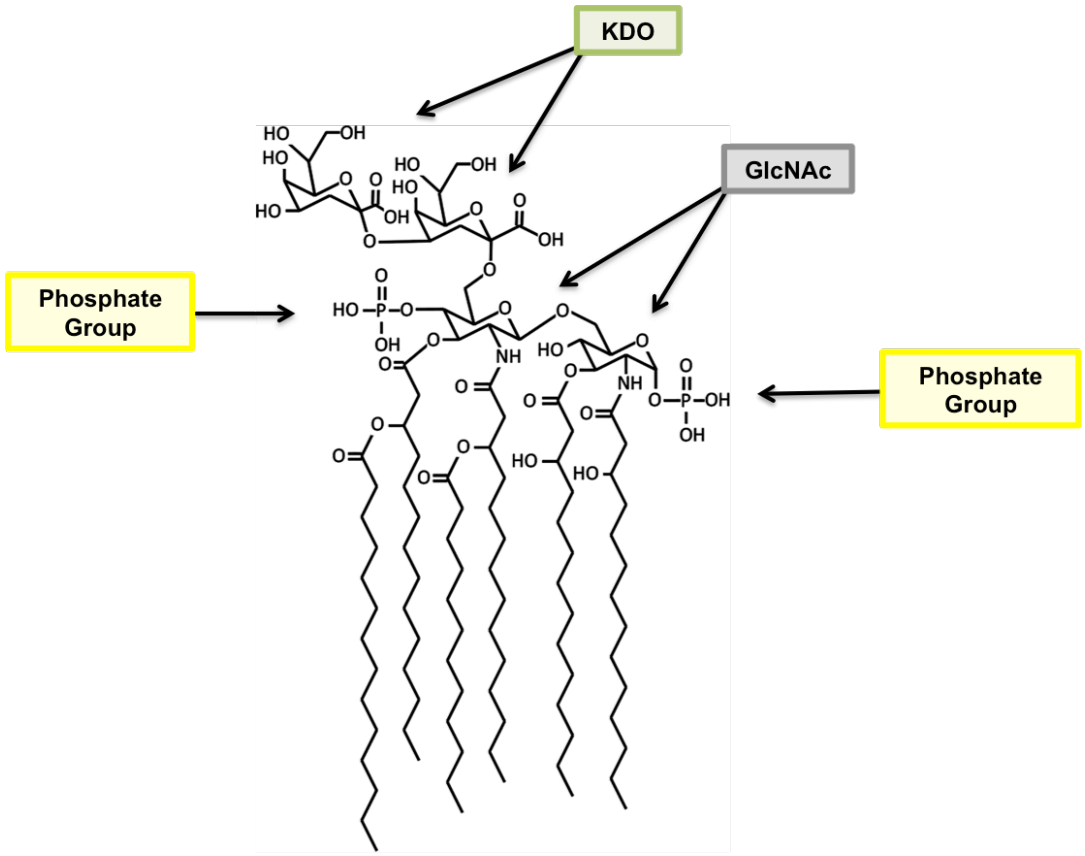



**Figure 13: Structure of $KDO_2$-Lipid A.** The top two sugars are KDO groups, which are part of the core oligosaccharide, while the remainder of the structure represents Lipid A. Modified from[100].

### 1.2.1.1. Lpx Acyltransferases For Synthesis And Modification Of Lipid A

In *E. coli*, nine enzymes are required for the biosynthesis of $KDO_2$-Lipid A, which is sufficient to maintain the viability and fitness of virtually all Gram-negative bacteria. This biosynthetic pathway is known as Raetz Pathway (**Figure 14**)[88]. Of the nine Lipid A enzymes, the first six enzymes (LpxA, LpxC, LpxD, LpxH/I/G, LpxB, and LpxK) are essential, and their deletion cannot be rescued by compensatory mutations. The three remaining enzymes (KdtA, LpxL, and LpxM) are not essential, but they exert a profound effect on the fitness of the bacteria and may serve as anti-virulence targets.

LpxA, C and D are soluble proteins[101–103], LpxB and LpxH are peripheral membrane proteins[104–106], and LpxK, KdtA, LpxL and LpxM are integral inner membrane proteins[29,36,37,107,108]. Their active sites are presumed to face the cytoplasmic surface of the inner membrane, given that their water-soluble co-substrates are cytoplasmic molecules (**Figure 14**)[19].

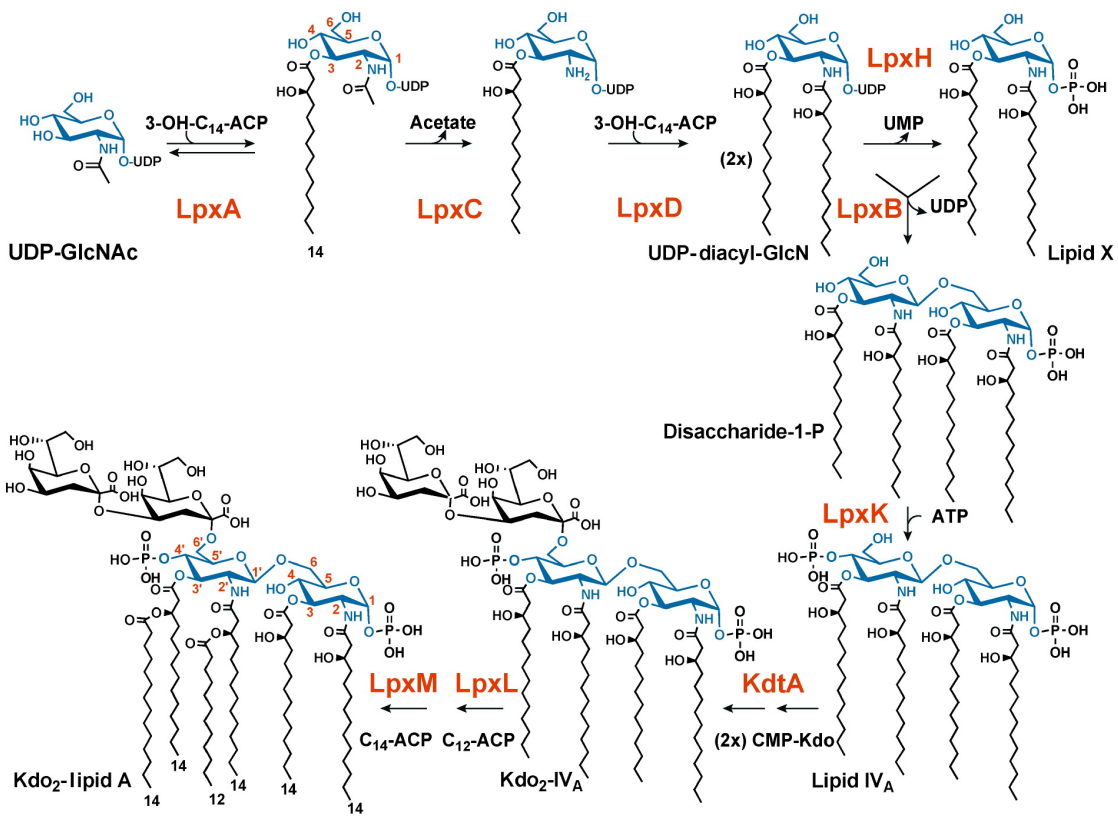



**Figure 14: The constitutive pathway for $KDO_2$-Lipid A biosynthesis in *E. coli*.** Each enzyme of the constitutive Lipid A pathway is encoded by a single structural gene. The glucosamine disaccharide backbone of Lipid A is blue. The KDO disaccharide is black. The characterized enzymes are in red. The red numbers specify the glucosamine ring positions of Lipid A and its precursors. The black numbers indicate the predominant fatty acid chain lengths found in *E. coli* Lipid A. The single molecular species shown at the bottom left represents about 90% of the total Lipid A in *E. coli*, with most of the rest bearing a $C_{12}$ secondary acyl chain at position 3'. Modified from[19].

To initiate the biosynthetic pathway, cytosolic proteins LpxA, LpxC and LpxD catalyze a series of reactions converting uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) into UDP-2,3-diacyl-glucosamine. First, the acyltransferase LpxA transfers the acyl chain from (R)-β-hydroxymyristoyl-ACP to the 3-OH group of UDP-GlcNAc. This acyl transfer reaction is thermodynamically unfavorable. To drive the reaction forward, LpxC, a $Zn^{2+}$-dependent metalloenzyme, hydrolyzes the 2-acetamido functionality, and thus catalyzes the first committed step in the pathway. The acyltransferase LpxD then installs another (R)-β-hydro- xymyristoyl moiety onto the 2-amino group, yielding UDP-2,3-diacyl-glucosamine (**Figure 14**).

The UDP-2,3-diacyl-glucosamine product of LpxA-C-D is transformed by three membrane associated enzymes, LpxH, LpxB and LpxK, into tetra-acylated Lipid IVA. LpxH hydrolyzes UDP-2,3-diacylglucosamine to yield 2,3-diacylglucosamine 1-phosphate. LpxB catalyzes formation of a β-1',6-glycosidic bond to generate the disaccharide backbone. In the final step leading up to Lipid IVA, LpxK phosphorylates the 4'-position of this disaccharide (**Figure 14**).

Prior to late-stage acylation, *E. coli* requires the addition of two KDO sugars to Lipid IVA whereas *Acinetobacter baumanii* does not, at least *in vitro*[29]. A bifunctional enzyme KdtA catalyzes two successive glycosyltransfer reactions to form $KDO_2$-Lipid IVA. This tetra-acyl tetra-saccharide is further modified by late acyltransferases LpxL and LpxM through the addition of lauroyl and myristoyl secondary acyl chains, respectively, to yield hexa-acylated Lipid A (**Figure 14**)[88,109].

After $KDO_2$-Lipid A is synthesized in the cytosol, it is attached with core sugars and flipped out from the cytoplasmic surface of the inner membrane to the periplasmic surface, where it is further modified in some species of bacteria and transported to the outer membrane[19,84].

There is Lipid A diversity among Gram-negative bacteria. This is dependent on the type and length of fatty acids, on the presence of chemical moieties or even on the lack of groups such as phosphates or fatty acids from Lipid A[88,110]. Well-characterized modifications comprise the addition of phosphoethanolamine[111], 4-amino-4-deoxy-l-arabinose[112], palmitoleate (LpxP)[113], and hydroxylation by the $Fe^{2+}$/α-ketoglutarate-dependent dioxygenase enzyme (LpxO)[114]. These modifications provide resistance to stresses such as harsh pH and antimicrobial peptides (APs)[110]. APs are directed to LPS on the bacterial surface by electrostatic interactions[115]. Additionally, these Lipid A changes may help pathogens to modulate the activation of host defense mechanisms by altering the activation of the TLR4/MD-2 receptor complex[116].

#### 1.2.1.1.1. The First Two Acyltransferases of the Raetz Pathway: LpxA and LpxD

LpxA and LpxD UDP-N-acetylglucosamine-O-acyltransferases (LpxA and LpxD) catalyze the first two acylation steps in the Lipid A biosynthetic pathway. They are soluble cytosolic enzymes and usually require 3-hydroxyacyl chains attached to the essential acyl donor ACP (**Figure 14**)[94,117].

**LpxA Acyltransferase**

The structure of the soluble LpxA from differents organisms is available in its apo form, and in complex with different substrates and peptides, thus this acyltransferase has been extensively studied. The first step of the overall Lipid A biosynthesis route, is the reversible O-acylation of the 3-hydroxyl group of the sugar nucleotide UDP-N-acetylglucosamine (GlcNAc) catalyzed by LpxA (**Figure 14**)[118,119].

LpxA is a 28-kDa protein that forms a biological homotrimer in solution. Its molecular architecture resembles a mushroom, with the central 'stem' formed by three copies of a prism-shaped left-handed parallel β-helix domain at the N-terminus, and the 'cap' formed by a four-helix-bundle domain at the C-terminus. The left-handed parallel β-helix domain of LpxA consists of approximately 30 characteristic β-strands, with every three β-strands forming one complete coil of a β-helix in the shape of an equilateral triangle. In the middle of the β-helix domain exist two insertion loops that

connect the subunit interfaces and form the pedestal of the active site for interaction with the UDP-GlcNAc substrate (**Figure 15 A**).

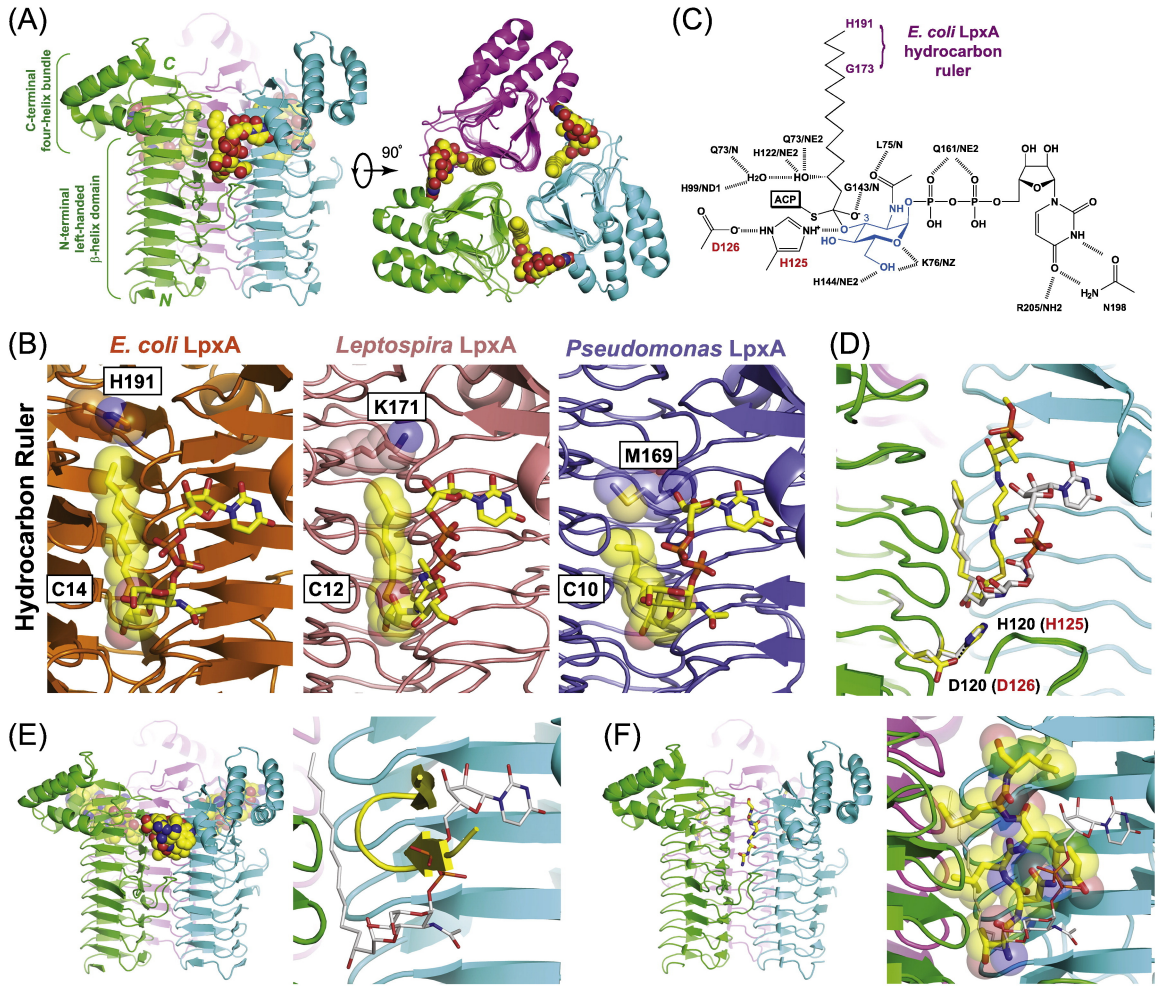



**Figure 15: Structure and inhibition of LpxA. A)** Structure of the *E. coli* LpxA/UDP-3-O-(R-3- hydroxymyristoyl)-GlcNAc complex (PDB: 2qia). LpxA consists of a homotrimer, with each monomer labeled in distinct colors. The domains and termini of the green monomer are labeled. LpxA is shown in the ribbon diagram, whereas the UDP-3-O-(R-3-hydroxymyristoyl)-GlcNAc molecules are shown in the space-filling model. **B)** The hydrocarbon rulers of LpxA enzymes from *E. coli* (PDB: 2qia), *L. interrogans* (PDB: 3i3x), and *P. aeruginosa* (PDB: 5dg3). **C)** LpxA acyl chain selectivity and catalysis. **D)** Overlay of UDP-3-N-(R-3-hydroxylauroyl)-GlcNAc3N (grey) and R-3-hydroxylauroylmethyl-phosphopantetheine (yellow) in the *L. interrogans* LpxA complexes (PDB: 3i3x and 3i3a). UDP-3-N-(R-3-hydroxylauroyl)-GlcNAc3N and R-3-hydroxylauroylmethyl-phosphopantetheine are shown in the stick model. The catalytic dyads are labeled, with *Lepstopira* residue numbers shown in black and *E. coli* residue numbers shown in red. **E)** Peptide 920 bound to *E. coli* LpxA (PDB: 2aq9). **F)** Peptide RJPXD33 bound to *E. coli* LpxA (PDB: 4j09). Carbon atoms of peptides in panels **E)** and **F)** are colored in yellow, whereas carbon atoms of UDP-3-O-acyl- GlcNAc are colored in grey. Modified from[120].

The structure of *E. coli* LpxA in complex with the product UDP-3-O-(R-3- hydroxymyristoyl)-GlcNAc, reveals a product at the interface of two adjacent LpxA subunits. The glucosamine ring is supported by the molecular pedestal formed by two insertion loops mentioned above. The 3-OH group of the glucosamine ring is buried, with its attached fatty acyl chain extending into the active site cleft along the subunit interface, whereas the UDP moiety extends toward the adjacent subunit to the right.

The first four carbon atoms are almost perpendicular to the long axis of the β-helix. The rest of the hydrocarbon chain runs parallel to the β-helix (**Figure 15 A**).

LpxA acyltransferases are highly selective for the acyl chain length of their ACP substrate[117]. This specificity is species dependent and is consistent with the composition of Lipid A isolated from the respective bacterium. In *E. coli* and *L. interrogans*, LpxA prefers to incorporate 14-carbon and 12-carbon R-3-hydroxyacyl chains respectively[117,121], whereas *Pseudomonas aeruginosa* LpxA selects for an acyl chain of 10-carbon atoms (an R-3-hydroxydecanoyl chain) (**Figure 15 B**)[117].

Structural analyses of the corresponding LpxA-product complexes have revealed the presence of an hydrocarbon ruler in all LpxA enzymes: in *E. coli* LpxA, along the hydrophobic cleft formed at the β-helix domain, His191 from the 10$^{th}$ β-coil blocks the further extension of the acyl chain and caps it to 14 carbons[122]; in *L. interrogans* LpxA, Lys171 from the 9$^{th}$ β-coil obstructs the hydrophobic cleft, restricting the length of the acyl chain to 12 carbons[121]; and in *P. aeruginosa* LpxA, Met169 from a neighboring subunit jams the cleft even closer to the glucosamine ring, accommodating the acyl chain of only 10 carbon atoms[123]. Consistent with these structural observations, the elimination of the steric blockage of Met169 by a Met169Gly mutation allows *P. aeruginosa* LpxA to preferentially incorporate a $C_{14}$ acyl chain instead of a $C_{10}$ acyl chain, whereas introducing a steric block (Gly173Met) in *E. coli* LpxA alters its preference from a $C_{14}$ acyl chain toward a $C_{10}$ acyl chain(**Figure 15 C-D**)[117].

Biochemical studies performed in *E. coli* LpxA suggested a mechanism in which His125 in the middle of the central β-helix serves as the catalytic base to active the 3-OH group of the substrate UDP-GlcNAc, which in turn attacks the thioester bond of the second substrate, acyl-ACP to replace ACP(**Figure 15 C**)[117]. Such a mechanism is supported by the observation of a hydrogen bond between His125 and the 3-O atom of the glucosamine ring of UDP-3-O-acyl-GlcNAc in complex with *E. coli* LpxA[122]. The structures of *L. interrogans* LpxA in complex with UDP-3-N-(R-3-hydroxylauroyl)-GlcNAc3N and with R-3-hydroxylauroylmethyl-phosphopantetheine, a mimetic of acyl-ACP, have also been determined[121], and strongly suggest an ordered bi-substrate mechanism for LpxA: the binding of acyl-ACP occurs first, which is followed by the binding of UDP-GlcNAc/UDP-GlcNAc3N; subsequently the acyl chain transfer occurs via a substrate-assisted mechanism facilitated by the catalytic histidine residue.

Potent peptide inhibitors of LpxA have been reported (**Figure 15 E-F**)[124,125]. The Peptide 920 competes with acyl-ACP for LpxA binding. It adopts a β-hairpin conformation binding at the deep cleft of the LpxA trimer interface near the active site[124,126]. Interestingly, peptide 920 not only occupies the space of the pantetheine group of the acyl-ACP analog, acyl-phosphopantetheine[121], but also overlaps in space with the substrate UDP-GlcNAc and the UDP-GlcNAc moiety of the product UDP-3-O-acyl-GlcNAc. Hence, the observation that peptide 920 is a competitive inhibitor against acyl-ACP, but not UDP-GlcNAc, is consistent with an ordered sequential bi-substrate binding process, in which acyl-ACP

binds LpxA first, and UDP-GlcNAc binds second (**Figure 15 E**). RJPXD33 is a dual-specific peptide inhibitor of LpxA and of the second acyltransferase of the pathway, LpxD[125], however it is a weaker inhibitor of LpxA compared to peptide 920. The structure of RJPXD33 reveals an extended conformation that is distinct from the hairpin structure of peptide 920[127], the peptide backbone of RJPXD33 runs parallel with the acyl chain of the UDP-3-O-acyl-GlcNAc molecule, with its amino acid side chains reaching out to the space occupied by the acyl chain and the UDP-GlcNAc moiety (**Figure 15 F**).

LpxA homologs not only are very widespread among nearly all Gram-negative bacteria but also occur in the majority of higher plants, although their function remains unclear[128].

**LpxD Acyltransferase**

The LpxA product is deacetylated by LpxC to UDP-3-acylglucosamine, and a second acyl chain is attached via an amide linkage at position 2 of the glucosamine ring by the N-acyltransferase LpxD, yielding UDP-2,3-diacylglucosamine (**Figure 14**).

The gene encoding LpxD was initially discovered as *firA*, for its role in reversal of certain rifampin- resistant mutations[129]. By 1993, *firA* was biochemically characterized and renamed *lpxD*, which is the gene that encodes LpxD[102]. LpxD and LpxA are encoded in the same operon and share significant sequence homology (45% similarity) to each other. Therefore, it is supposed that both enzymes result from gene duplications[102,118,130].

The first structure of LpxD (36-kDa) from *C. trachomatis* (PDB Code: 2IUA) was reported in 2007[131], though the topological arrangement of its C-terminal domain is sharply different from all other LpxD ortholog structures from *E. coli* (PDB Code*:* 3EH0)[132], *P. aeruginosa* (PDB Code: 3PMO)[133] and *A. baumannii* (PDB Code: 4E75)[134]. The *E. coli* LpxD structure, shares a common domain arrangement with LpxD orthologs from *P. aeruginosa* and *A. baumannii*, and is described as an homotrimer structure with a characteristic central axis of the left-handed β-turn helix similar to that seen in LpxA (**Figure 15 A** and **Figure 16 A**). Each monomer also contains an N-terminal uridine-binding domain and a C-terminal helix domain that engages ACP (**Figure 16 D**).

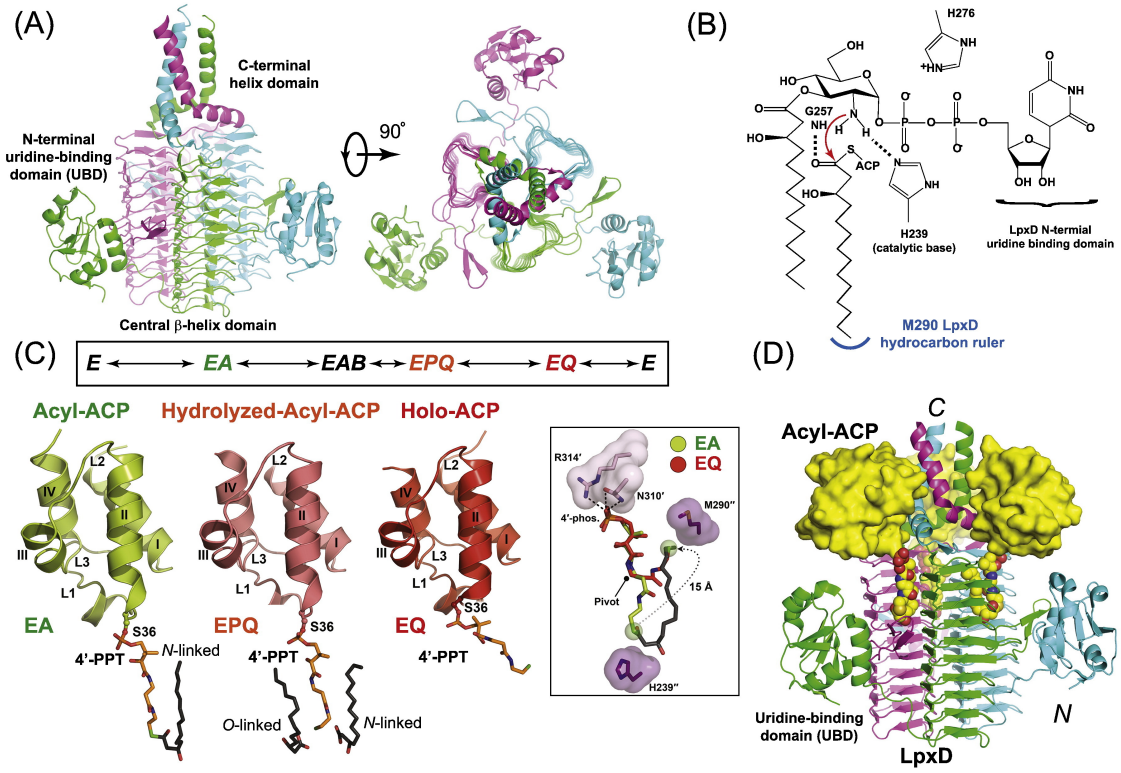



**Figure 16: LpxD structure and catalysis. A)** Structure of *E. coli* LpxD (PDB: 3eh0), revealing an intertwined trimer structure. **B)** Proposed catalytic mechanism of LpxD. **C)** A compulsory ordered bi-bi substrate mechanism. The LpxD-bound conformations of acyl-ACP (PDB: 4ihf), hydrolyzed acyl-ACP with a fatty acid (PDB: 4ihg), and the holo-ACP (PDB: 4ihh) are shown below the reaction. The movement of the 4'-PPT group from the acyl-ACP complex to the holo-ACP complex is shown in the inset. **D)** Structure of the LpxD/acyl-ACP complex (PDB: 4ihf). Modified from[120].

The hydrocarbon ruler of LpxD was tested due to its highly selectivity for the acyl chain length. *E. coli* LpxD prefers to transfer an acyl chain of 14 carbon atoms over a chain of 16 carbon, such selectivity is caused by the steric blockage of the acyl chain binding grove by a bulky residue, Met290. Removal of the steric blockage by the Met290Ala mutation alters the selectivity of *E. coli* LpxD towards a 16-carbon unit acyl chain (**Figure 16 B**)[132].

Biochemical analysis suggests that LpxD utilizes a substrate-assisted catalytic mechanism (**Figure 16 C**)[135]. His239 located in the middle of the left-handed β-turn helix of LpxD serves as the catalytic base to deprotonate the 2-amino group of the glucosamine ring of UDP-3-O-acyl-GlcN, which then attacks the carbonyl carbon of the thiol-ester bond of β-hydroxyacyl-ACP[135], with the oxyanion being stabilized by the backbone of Gly257.

The LpxD reaction undergoes a imperative ordered bi-substrate mechanism (**Figure 16 C**)[135]. In such a mechanism, R-3-OH-$C_{14}$-ACP (acyl-ACP) binds LpxD first to form the EA complex; this is followed by the binding of UDP-3-O-(R-3-OH-$C_{14}$)-GlcN to form the bi-substrate complex EAB; the acyl chain transfer yields the bi-product complex EPQ; one of the products, UDP-DAGn, dissociates first, leaving behind the LpxD-holoACP complex (EQ); finally, the holoACP departs to regenerate LpxD. The tight affinities between various forms of ACP and LpxD have enabled structural snapshots of the

intermediates of the LpxD/acyl-ACP complex (EA), the LpxD/holo-ACP (EQ) complex, as well as a hydrolyzed acyl-ACP LpxD complex that resembles the post-catalysis EPQ complex during the catalytic cycle[136]. In all of these structural states, the ACP docks on top of the C-terminal ACP-recognition domain formed by the trimeric C-terminal helical domain and the last β-coil of the left-handed β-helix domain (**Figure 16 C-D**).

### 1.2.1.1.2. The Last Two Acyltransferases of the Raetz Pathway: LpxL and LpxM

LpxH hydrolyzes the LpxD product, UDP-2,3-diacylglucosamine, to yield 2,3-diacylglucosamine 1-phosphate. LpxB catalyzes formation of a β-1',6-glycosidic bond to generate the disaccharide backbone. In the final step leading up to Lipid IVA, LpxK phosphorylates the 4'-position of this disaccharide (**Figure 14**).

After Lipid IVA is formed, *E. coli* requires the addition of two KDO sugars to Lipid IVA to be subsequently acylated by LpxL and LpxM, however it has been recently discovered that *Acinetobacter baumanii* is able to acylate Lipid IVA without the addition of KDO sugars *in vitro*[29]. In *E. coli*, a bifunctional enzyme KdtA catalyzes two successive glycosyltransfer reactions to form $KDO_2$-Lipid IVA. This tetra-acyl tetra-saccharide is further modified by late acyltransferases LpxL and LpxM through the addition of lauroyl and myristoyl secondary acyl chains, respectively, to yield hexa-acylated Lipid A (**Figure 12**)[88,109].

#### LpxL Acyltransferase

LpxL-like acyltransferases comprises membrane-bound enzymes which do not exhibit any similarity to the soluble LpxA and LpxD. While in *E. coli* and related bacteria the two successive acylations, first laurate at position 2' then myristate at position 3', are carried out by homologous LpxL and LpxM, other Gram-negative bacteria possess solely one or more *lpxL*-orthologous sequences but no *lpxM* gene[36,37,130]. *P. aeruginosa*, for example, possesses two LpxL versions, each of which is presumed to transfer one of the secondary lauroyl groups to Lipid IVA[86]. The *lpxM* gene is a product of *lpxL* gene duplication, and thus LpxM shares approximately 47% similarity to LpxL and represents a more recently emerged, specialized version characterized as myristoyl transferase[37,130]. It is worth to mention that *lpxL* and *lpxM* were initially discovered as *htrB* and *msbB*, respectively, but when their function was known they were renamed LpxL and LpxM[36,37,107].

In addition to *lpxM*, *lpxP* represents a third *lpxL* ortholog in *E. coli* and closely related bacteria[130]. It encodes a palmitoleoyl transferase with a high similarity to LpxL (74% similar amino acids). As a response to low growth temperatures (12°C), the function of LpxL can be replaced by LpxP. LpxP from *E. coli* attaches an unsaturated palmitoleoyl ($C_{16:1}$) residue instead of laurate at position 2' to Lipid A, which probably helps to adapt the outer membrane fluidity to low temperatures[108,137].

Although LpxL crystal structure has not been reported, LpxL from *E. coli* has been purified and studied in detail[138]. Purified LpxL not only was able to perform the first lauroylation step but could also replace LpxM by performing a slow second lauroylation *in vitro*. Furthermore, although lauroyl-ACP is obviously the naturally preferred substrate of LpxL, it can alternatively utilize several acyl-CoAs instead of acyl-ACPs at a lower rate *in vitro*. However, the utilization of acyl-CoAs strongly depends on their concentration, as they can also act as inhibitors[138].

**LpxM Acyltransferase**

*Ab*LpxM performs the secondary acylation of the R-3-hydroxyacyl chain at position 3' of KDO-Lipid IVA (**Figure 14**). The structure of the acyltransferase LpxM from *Acinetobacter baumannii* (*Ab*LpxM; PDB Code 5KNK)[29] has been solved. *Ab*LpxM adopts a globular structure bisected by a large seven-stranded β-sheet. The predicted transmembrane domain is composed of a single α-helix protruding from the globular domain and forms a substantial crystal contact with *Ab*LpxM in a neighboring asymmetric unit (**Figure 17 A**).

Electrostatic surface visualization of the *Ab*LpxM structure revealed a very large hydrophobic pocket, which bound a co-purifying n-dodecyl-β-D-maltoside (DDM) molecule. Due to the hydrophobicity and size of the lipid substrate of *Ab*LpxM, and given the lack of any other large hydrophobic surface, this pocket likely represents the binding site of the acyl chain acceptor (**Figure 15 B**). This large pocket also contains several deep hydrophobic channels, which may serve as hydrocarbon rulers, and which may determine specificity for acyl chain length on either acyl-ACP or on the Lipid A precursors (**Figure 17 C**)[29].

LpxM also contains a conserved $HX_4D/E$ motif, using glutamate in the *A. baumannii* ortholog, which is present within the deep hydrophobic cleft, suggesting that the cleft could be the the location of the active site. The conserved residue Arg159 also resides within the cleft and is positioned very close to the putative catalytic His122 and Glu127. The interaction of this residue with the $C_6$-hydroxyl of the maltose from DDM suggest that Arg159 could interact with the $C_1$ phosphate on the diglucosamine head group of Lipid IVA, thus it may play a role in substrate binding. The transferase activity of both mutants Arg159Ala and Glu127Ala greatly decrease, however whereas the Glu127Ala variant strongly respond to increasing Lipid IVA substrate, the Arg159Ala variant displayed a much more attenuated response. This is in accordance with the hypothesis that the conserved Arg159 residue contributes to either substrate recognition and binding or to the proper positioning of substrates before the acyl chain transfer[29].

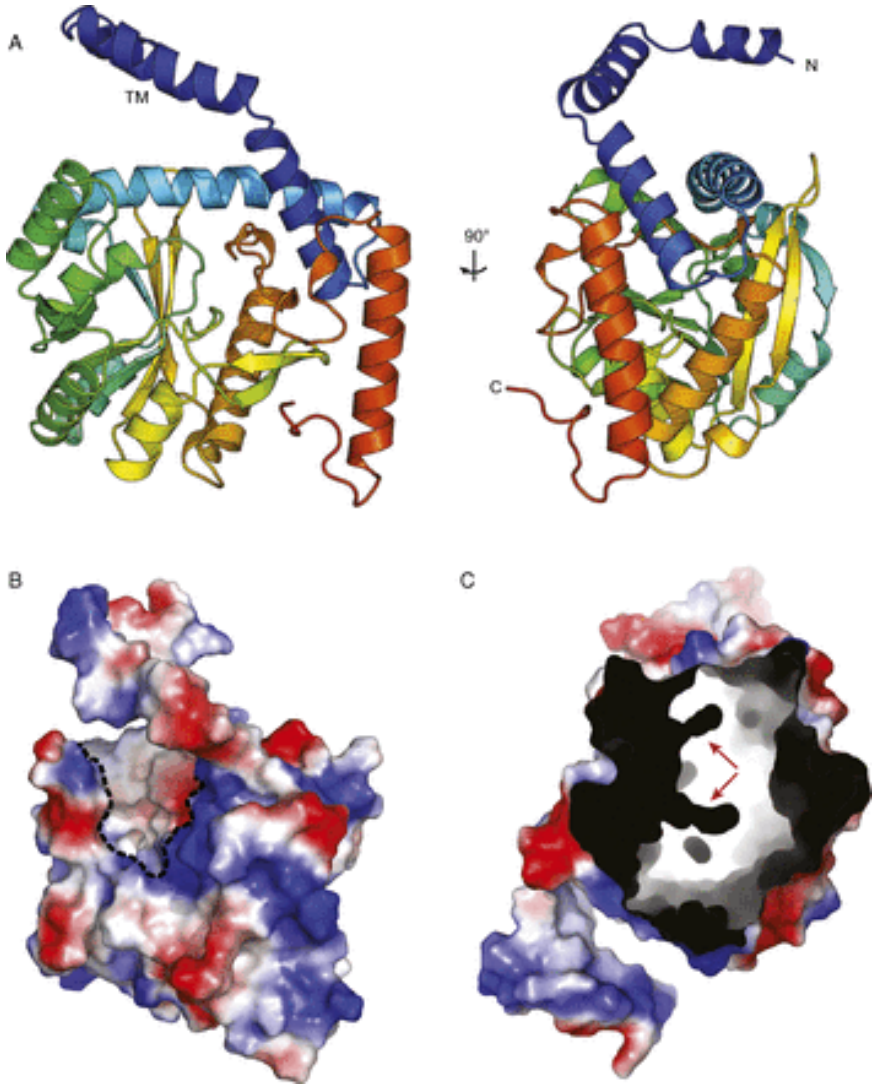



**Figure 17: The structure of *Ab*LpxM. A)** Cartoon schematic of *Ab*LpxM. From the N-terminus to C-terminus, the color shifts from blue to red, and transmembrane helixes (TM) in dark blue. Two angles are shown. **B)** Predicted surface electrostatics map generated by PyMol. The putative binding cleft is indicated with a dashed line. **C)** Cutaway diagram showing deep pockets within the putative binding cleft (pockets indicated with red arrows; cutaway in black to show the depth of the binding cleft).

The acyltransferase activity for *Ab*LpxM, as reported for *E. coli* LpxL[138], is capable of using Lipid IVA as a substrate, even lacking the KDO moieties. Indeed, *Ab*LpxM produces both penta- and hexa-acylated Lipid A species when incubate with lauryl-ACP and Lipid IVA. It is also found that *Ab*LpxM could produce holo-ACP in the absence of any acyl chain acceptor, suggesting an acyl-protein thioesterase activity. Moreover, *Ab*LpxM produce free lauric acid when incubate with lauryl-ACP, however, this activity would be observed only by direct measurement of holo-ACP production, and not by tracking radiolabeled lipid A precursors as has previously been done[29,138,139]. *Ab*LpxM display a strong preference for lauryl-ACP over the other acyl chain lengths. As the acyl chain length deviate from 12 carbons in either direction, the Km increase and the specific activity decrease.

Whereas LpxM is typically depicted as accepting $KDO_2$-lauryl-lipid IVA as its substrate, making it the terminal enzyme in the Raetz Pathway, many LpxM orthologs exhibit substantial substrate promiscuity. For example, *Neisseria meningitides* produces fully acylated Lipid A even when the KDO transferase, *kdtA*, is knocked out, suggesting LpxL and LpxM can use Lipid IVA in this species[140]. Similarly, *P. aeruginosa* can produce fully acylated Lipid A even when CMP-Kdo synthase (KdsB) is inhibited or when acting upon purified Lipid IVA[141,142], suggesting that the presence of the KDO moieties on the lipid acceptor is dispensable for some LpxL/LpxM orthologs. In *A. baumannii*, LpxM adds lauryl groups to the R-3-hydroxyacyl chains at both the 3'-and 2-positions of Lipid A precursors, producing hepta-acylated Lipid A[143]. This secondary activity may necessitate flexibility in the active site to accommodate multiple conformations of Lipid A, so as to position the substrate properly[29].

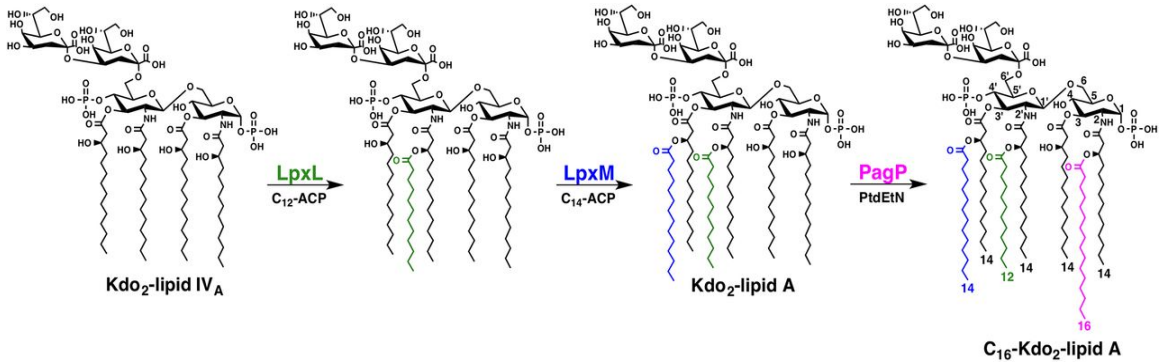



**Figure 18: Synthesis of hexa- and hepta-acylated Lipid A in *E. coli* K-12.** In *E. coli*, LpxL catalyzes transfer of a laurate ($C_{12}$) group from an acyl carrier protein (ACP) onto the R-2=-hydroxymyristate acyl chain of $KDO_2$-Lipid IVA. Subsequently, LpxM-dependent addition of myristate ($C_{14}$) onto the R-3=-hydroxymysristate residue completes synthesis of hexa-acylated $KDO_2$-Lipid A. Upon sensing specific environmental cues, PagP activation results in palmitoylation ($C_{16}$) of the R-2-hydroxymyristate primary acyl chain, altering membrane fluidity and increasing resistance to antimicrobial peptides. Phosphatidylethanolamine (PtdEtN)[143].

### 1.2.1.1.3. Modification of Lipid A by the Acyltransferase PagP

As an adaptation to changes of the environmental conditions, the Lipid A structure can be further modified by the cell after its transport across the membrane. A mechanism involving additional acylation is exerted by the outer membrane acyltransferase PagP.

PagP is an outermembrane palmitoyltransferase that catalyzes the transfer of a palmitate chain (16:0) from the sn-1 position of a glycerophospholipid to the free hydroxyl group of the (R)-3-hydroxymyristate chain at position 2 of Lipid A (**Figure 18**)[2,144,145]. The resulting hepta-acylated Lipid A can provide a more tightly packed LPS layer and higher integrity of the outer membrane[145,146].

The structure of *E. coli* PagP has been determined both by solution nuclear magnetic resonance spectroscopy (PDB Code 1MM4 and 1MM5) and by X-ray crystallography (PDB Code 1THQ and 3GP6), revealing an eight-strand antiparallel β-barrel preceded by an N-terminal amphipathic α-helix [147–149].

Like other β-barrel membrane proteins, the center of the PagP β-barrel is relatively rigid with more flexible external loops where the active site residues are localized (**Figure 19**).

Even though the catalytic mechanism remains to be elucidated, PagP alternates between two dynamically distinct states likely representing latent and active conformations in the outer membrane environment[150]. PagP activity is triggered by outer membrane lipid asymmetry perturbations that enable the donor phospholipid and acceptor Lipid A substrates to have access to the active site from the external leaflet[151–153]. Phospholipid and Lipid A binding occurs by lateral diffusion through two gateways in the β-barrel wall known as the crenel and embrasure, respectively[154]. A solved crystal structure of PagP inhibited by the detergent lauryldimethylamine N-oxide (LDAO) revealed a buried LDAO molecule within the rigid β-barrel core, a site corresponding with the palmitoyl group binding pocket known as the hydrocarbon ruler (**Figure 19**)[149]. A X-ray structure of PagP crystallized from a mixture of sodium dodecyl sulfate (SDS) and 2-methyl-2,4-pentanediol (MPD) has revealed a crenel gating mechanism for the transit of the phospholipid between the hydrocarbon ruler and the outer membrane external leaflet[147]. The interior of the PagP region, which is exposed to the inner leaflet of the membrane, is mostly polar, whereas the interior of the other, LPS-exposed, region is strictly hydrophobic. Punctuations of the β-strands by proline residues enable the lateral access of lipid substrates from the inner leaflet of the membrane[148,149,155].

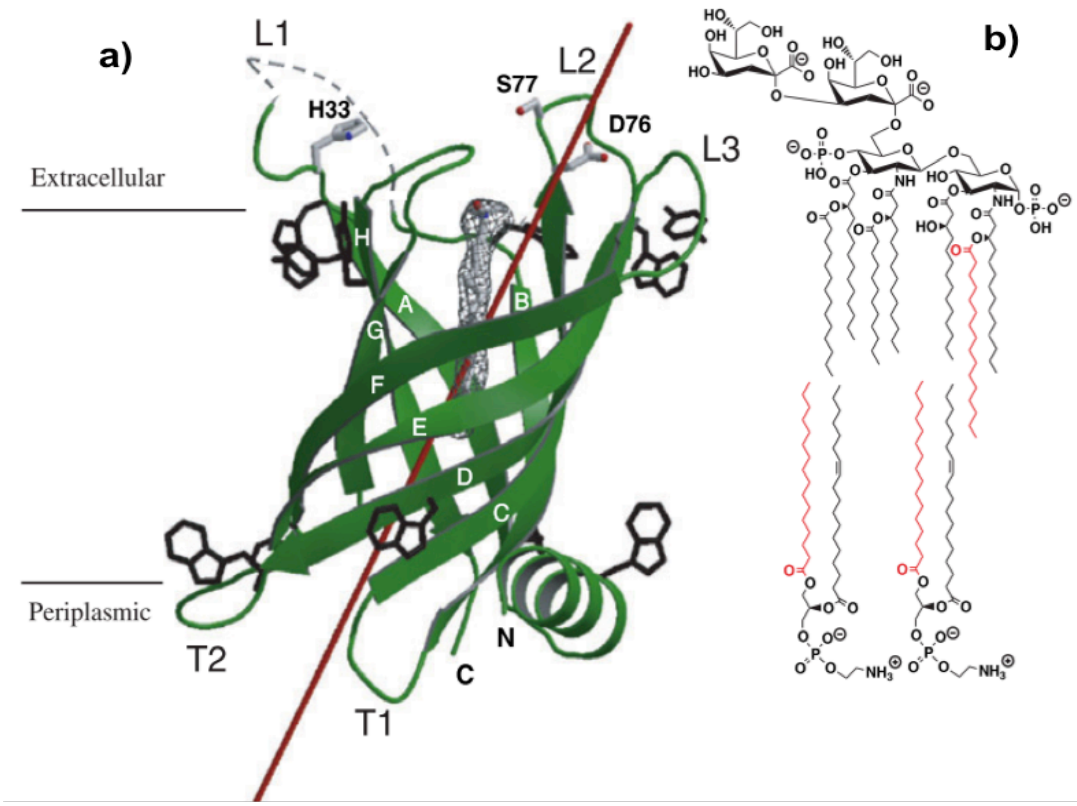



**Figure 19: PagP crystal structure. a)** Ribbon diagram of PagP. Electron density for the internal LDAO ligand is shown in gray mesh, and the aromatic residues located at the presumed membrane/water interfaces are shown in black. Residues 38-47 of extracellular loop L1 are not visible in the electron density maps and are indicated by the dashed

segment. The barrel axis is tilted approximately 25° from the membrane normal and is shown as a red line. The horizontal lines represent the presumed position of the outer membrane bilayer. **b)** The asymmetric outer membrane of Gram-negative bacteria consists of an inner phospholipid leaflet (indicated by the two molecules of phosphatidylethanolamine in the lower part of the panel) and an outward-facing leaflet of lipid A bearing minimally two 3-deoxy-d-manno-2-octulosonic acid (KDO) sugars (indicated by a $KDO_2$-Lipid A in the upper part of the panel, shown as the palmitoylated product of the PagP reaction). PagP catalyzes the transfer of a palmitate chain (shown in red) from the sn-1 position of a phospholipid to the Lipid A acceptor. This transfer requires the migration of the phospholipid donor to the outer leaflet of the outer membrane, presumably by mechanisms that do not directly involve PagP. Modified from[149].

PagP is remarkably selective for a 16-carbon palmitate chain because its hydrocarbon ruler excludes lipid acyl chains differing in length by a solitary methylene unit[149,156]. Mutation of Gly88 lining the hydrocarbon ruler floor can modulate lipid acyl chain selection[156]. Appropriate amino acid substitutions shorten the selected acyl chain by a degree predictable from the expected rise in the hydrocarbon ruler floor.

However, although the structure and dynamics of PagP have been studied in detail, its exact catalytic mechanism remains uncertain[144]. The putative catalytically active residues His33, Asp76, and Ser77 were initially thought to exert a mechanism similar to that of classical serine esterases as a catalytic triad. However, detailed structural analysis revealed that they are not organized correctly to form such a triad (**Figure 19**)[149]. Moreover, it was observed that typical inhibitors acting against serine esterases were not effective against PagP[144]. It is assumed that PagP can exist in different conformational states and performs an induced-fit mechanism[19,144,157].

In *E. coli* and related pathogenic Gram-negative bacteria, PagP evades host immune defenses by resisting antimicrobial peptides[113,158–160] and attenuating the inflammatory response to infection triggered by lipopolysaccharide through the TLR4/MD2 pathway[161–163]. Transcription of the pagP gene is governed by the PhoP/PhoQ regulon, which controls expression of bacterial Lipid A modification enzymes induced during infection by host antimicrobial peptides[164,165]. PagP is thus a target for the development of anti-infective agents and a tool for the synthesis of novel vaccine adjuvants and endotoxin antagonists[19,144,157].

## 1.2.2. The Gram-positive Mycobacteria Cell Envelope

*Mycobacterium tuberculosis*, the causative agent of tuberculosis, is the second most deadly infectious agent in the world after HIV[5,6]. The extremely robust and impermeable mycobacterial cell envelope, based on sugars and lipids of unique chemical structure, is the dominant feature of the biology of *M. tuberculosis*[7–9]. With this in mind, considerable effort has been placed on investigating the cell envelope structure and its biosynthesis to identify attractive drug targets[166].

According to the last model, mycobacteria cell envelope comprises four main components: the plasma membrane or inner membrane, the peptidoglycan-arabinogalactan complex (AGP), the outer membrane which is covalently bound to AGP through the mycolic acids, and the external capsular-like structure (**Figure 21** and **Figure 22**)[167–169].

The IM is composed of the standard phospholipids, such as cardiolipin (CL), phosphatidyl-*myo*-inositol (PI) and phosphatidylethanolamine (PE), constituting ~1.16%, ~0.37 and ~0.52% of the dry cell mass, respectively[169]. However, phosphatidyl-*myo*-innositol mannosides (PIMs) are the most abundant of all lipids extracted from the IM. PIMs are unique glycolipids found in abundant quantities in the inner and outer membranes of the cell envelope of all *Mycobacterium* species. They are based on a PI lipid anchor carrying one to six mannose (Man*p*) residues and up to four acyl chains. In more detail, just the tetra-acylated form of PI dimannosides ($Ac_2PIM_2$) account for up to 42% of all lipids extracted from de IM (**Figure 20**). PIMs are considered not only essential structural components of the cell envelope but also the structural basis of the lipoglycans, lipomannan and lipoarabinomannan[167,170,171].

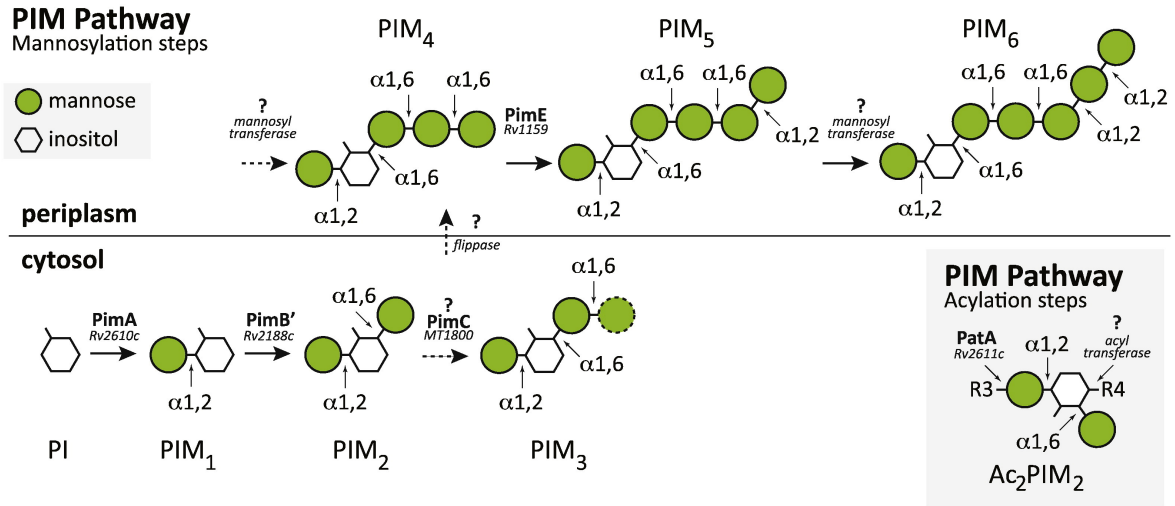



**Figure 20: The PIM Biosynthetic Pathway.** Scheme showing the mannosyltransferases involved in the biosynthesis of the two major products of the pathway, $PIM_2$ and $PIM_6$. The currently accepted model propose that PimA and PimB catalyze the two first consecutive mannosylation steps in the cytoplasmic face of the plasma membrane. The identity and location of the enzyme/s responsible for the third and fourth mannosylation steps is under debate, as well the existence of a flippase. The fifth mannosylation step is catalyzed by PimE in the periplasmic face of the plasma membrane. The identity of the enzyme responsible for the sixth mannosylation step is unknown. PIMs can be also acylated by the action of two acyltransferases: PatA catalyzes the transfer of an acyl chain to the mannose ring

transferred by PimA at the cytosolic face of the plasma membrane, whereas an unknown acyltransferase catalyze the transfer of an acyl chain to the inositol ring of the glycolipid. In the inset the acylation steps on $Ac_2PIM_2$ are shown as an example.

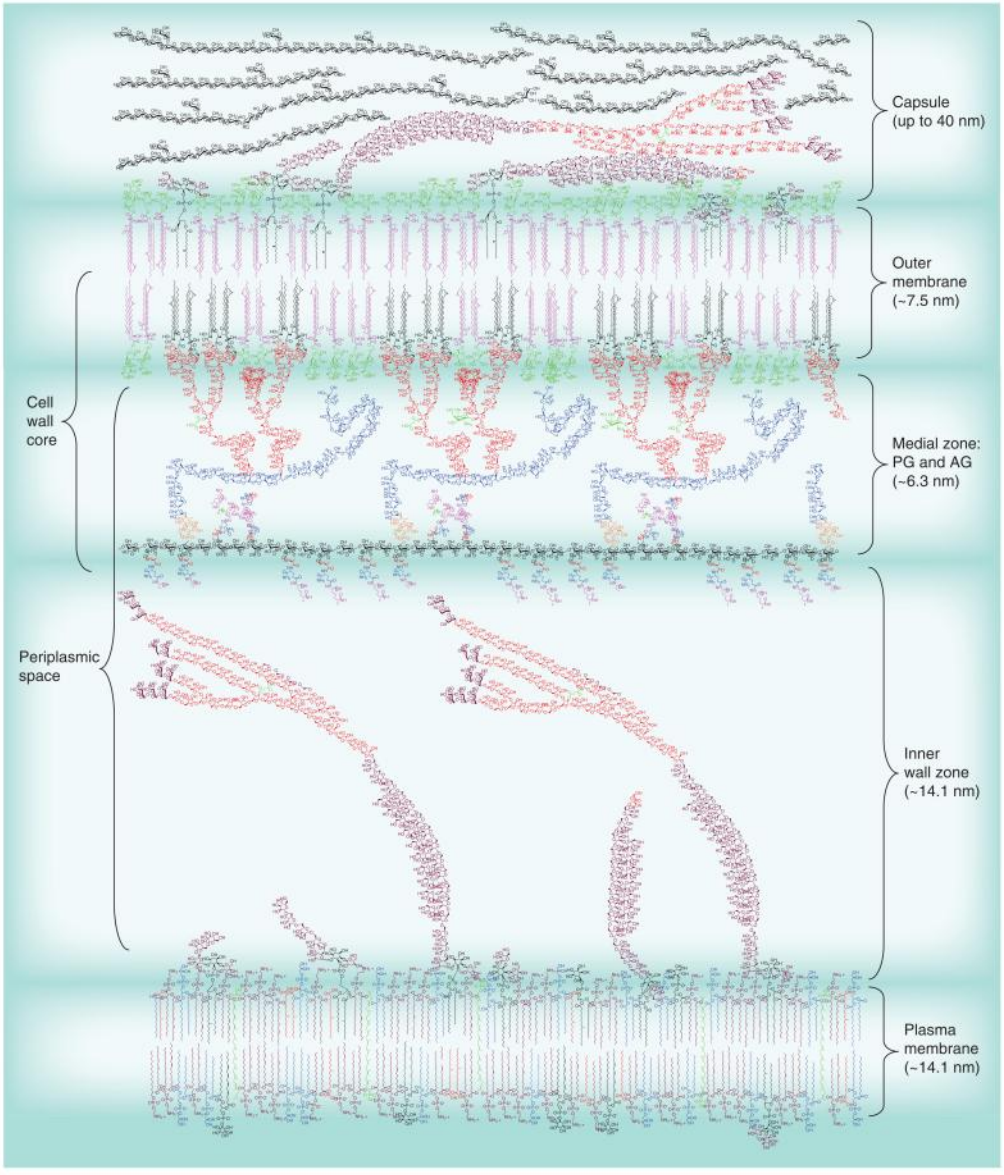



**Figure 21: A depiction of the Mycobacterium tuberculosis cell envelope.** The cell wall core consists of the PG (black with colored peptides extending from it) attached to the mycolate layer (inner leaflet of the outer membrane) via the AG. The galactofuranosyl residues are shown in blue and the arabinofuranosyl residues in red. The mycolic acids attached to the arabinan are shown in black and are in a folded 'W' configuration; the pink mycolic acids are attached to trehalose (trehalose monomycolate with a single mycolic acid and trehalose dimycolate with two mycolic acids) and are expected to be found in both leaflets of the outer membrane. Lipoarabinomannan (arabinofuranosyl residues in red as in AG and mannopyranosyl residues in purple) is shown both in the outer membrane and plasma membrane as is lipomannan. Phospholipids are only shown in the plasma membrane, but are also found in the outer membrane. The capsule is shown containing glucan (black), mannan (purple) and arabinomannan (purple and red). The dimensions of the various layers are from published electron micrographic studies. Molecular modeling suggests that the length of the medial zone could be as large as 20 nm; the arabinan and galactan are quite flexible and a very dense contracted configuration is required for the much smaller 6.5 nm medial zone suggested by the electron microscopic studies; this perhaps could change under different growth conditions. Modified from[172].

In the outer part of the plasma membrane, the AGP complex comprises the peptidoglycan (PG) in phosphodiester linkage to a linear galactan, to which several strands of highly branched arabinan are attached (**Figure 21** and **Figure 22**)[173,174]. In more depth, PG consists of a glycan backbone of alternating units of N-acetylglucosamine (GlcNAc) linked via β-(1→4) to modified muramic acid (Mur)[167,175]. These Mur residues can be either N-acetylated (MurNAc) or N-glycosylated (MurNGlyc), this MurNGlyc modification is supposed to increase the strength of mycobacterial PG[176]. Tetrapeptide side chains attach to the lactyl moiety of Mur components of the backbone, typically consisting of L-alanyl-D-glutamine-meso-diaminopimelyl-D-alanyl-D-alanine, that may be cross-linked[167]. Cross-links can comprise either (3→3) linkages between the meso-diaminopimelate (meso-DAP) residues of adjacent peptides[177] or (4→3) linkages between the D-Ala at position 4 of one peptide and the meso-DAP at position 3 of an adjacent peptide[178]. A small percentatge, about 10-12%, of the Mur residues of PG are covalently attached to arabinogalactan (AG) by a specific linker made of a GlcNAc-1-phosphate residue attached to a α-L-rhamnose (Rha)[179]. The galactan core is composed by a linear chain of approximately 30 alternating 5- and 6-linked β-D-galactofuranose (Galf) residues[180]. The highly branched arabinan chains are connected to the galactan core at $C_5$ of some of the β-(1→6) Galf residues[167,168,181].

The inner leaflet of the OM is formed by arabinogalactan esterified at its non-reducing ends to α-alkyl, β-hydroxy long-chain ($C_{60}$-$C_{90}$) mycolic acids[8,182–184]. To form the outer leaflet of the OM, mycolic acids can also be unbound and esterified with a variety of non-covalently attached glycolipids: trehalose mono- and di-mycolates (TMM and TDM, respectively), glycerol monomycolate, glucose monomycolate, PIMs, lipomannan (LM) and lipoarabinomannan (LAM), and lipoproteins/proteins some of which are glycosylated (**Figure 22**)[8,167,185,186]. The organization and composition of this asymmetrical outer membrane known as ‘‘mycomembrane’’[187,188] confer mycobacteria a high intrinsic resistance to many therapeutic agents and host defense mechanisms (**Figure 21** and **Figure 22**)[172,189,190].

Finally, the capsular structure outside the outer membrane mostly consists of polysaccharides and proteins with only minor amounts of lipids[191–193]. The three major polysaccharides found in the capsular material of *M. tuberculosis* consist of a high molecular weight α-glucan with a structure similar to that of glycogen, a D-arabino-D-mannan (AM), and a D-mannan (**Figure 21**)[192–194]. The α-glucans were recently suggested to be required for virulence of *M. tuberculosis*[195–197].

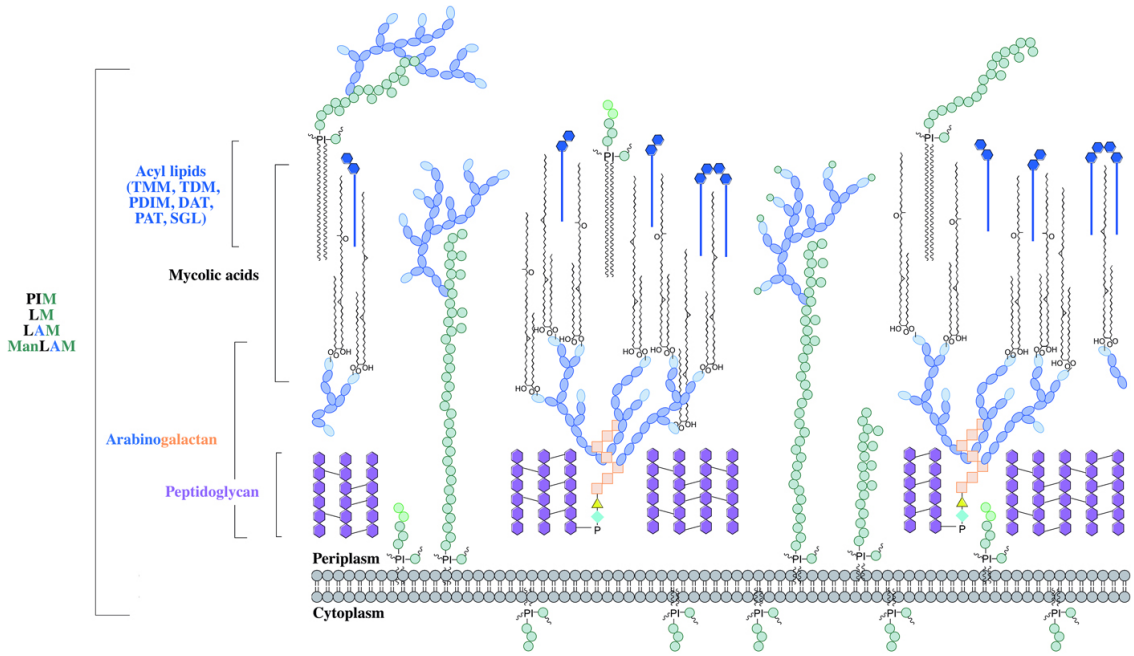



**Figure 22: The mycobacterial cell wall.** A schematic representation of the mycobacterial cell wall, including the glycolipids (Phosphatidyl-*myo*-inositol mannosides (PIMs), Lipomannan (LM), Lipoarabinomannan (LAM), Mannosylated lipoarabinomannan (ManLAM)), peptidoglycan, arabinogalactan and mycolic acids. Intercalated into the mycolate layer are the acyl lipids (including trehalose monomycolate (TMM), trehalose dimycolate (TDM); diacyltrehalose (DAT); polyacyltrehalose (PAT), phthiocerol dimycocerosate (PDIM), sulfoglycolipid (SGL)). The capsular material is not illustrated. Modified from[198].

### 1.2.2.1. Phosphatidyl-*myo*-innositol Mannosides (PIMs)

PIMs are based on PI lipid anchor carrying one to six Man*p* residues and up to four acyl chains (**Figure 20**)[167,170]. Based on a conserved mannosyl-PI anchor, PIMs are thought to be the precursors of the two major mycobacterial lipoglycans, LM and LAM (**Figure 22**)[13,14,199]. PIMs, LM, and LAM are considered not only essential structural components of the mycobacterial cell envelope, but also important molecules implicated in host-pathogen interactions in the course of tuberculosis and leprosy, which have been the object of several reviews[9,167,200–202].

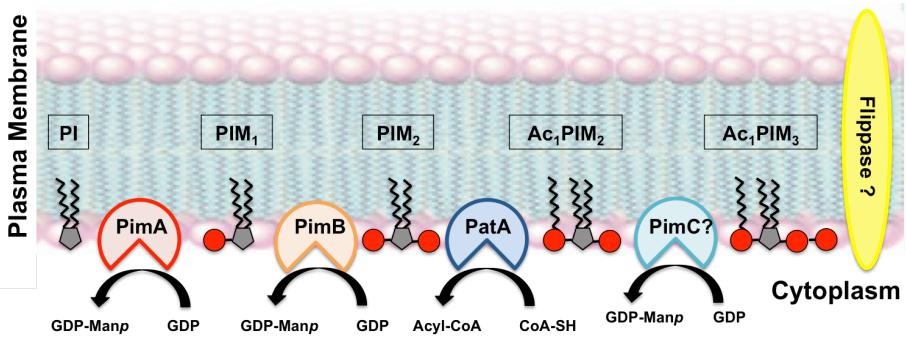



**Figure 23: Overview of the early stages of PIM biosynthesis in *Mycobacterium tuberculosis*.** On the cytosolic side of the plasma membrane, PI is glycosylated by PimA and PimB, and acylated by PatA to form $Ac_1PIM_2$. The identity and location of the enzyme/s responsible for the third and fourth mannosylation steps is under debate, as well the existence of a flippase.

The PIM family of glycolipids comprises PI mono-, di-, tri-, tetra-, penta-, and hexamannosides with different degrees of acylation(**Figure 24**)[170,203]. The acylated forms of $PIM_2$ and $PIM_6$ are the two most abundant classes found in *Mycobacterium bovis* bacillus Calmette-Guerin (BCG), *M. tuberculosis* H37Rv, *M. smegmatis* 607 and *M. smegmatis* $mc^2$ [204,205]. The complete chemical structures of the acylated forms of $PIM_2$ and $PIM_6$ were unequivocally established in *M. bovis* BCG[199,202]. $PIM_2$ is composed of two mannose (Man) residues attached to positions 2 and 6 of the *myo*-inositol ring of PI (**Figure 24**), whereas $PIM_6$ is composed of a pentamannosyl group, t-α-Man(1→2)-α-Man(1→2)-α-Man(1→6)-α- Man(1→6)-α-Man(1→, attached to position 6 of the *myo*-inositol ring, in addition to the Man*p* residue present at position 2)[170]. The triacylated forms of $PIM_2$ and $PIM_6$ ($Ac_1PIM_2$ and $Ac_1PIM_6$) show major acyl forms containing two palmitic acid residues ($C_{16}$) and one tuberculostearic acid residue ($C_{19}$), where one fatty acyl chain is linked to the Man*p* residue attached to position 2 of *myo*-inositol, and two fatty acyl chains are located on the glycerol moiety (**Figure 24**). The tetra-acylated forms, $Ac_2PIM_2$ and $Ac_2PIM_6$, are present predominantly as two populations bearing either three $C_{16}$/one $C_{19}$ or two $C_{16}$/two $C_{19}$[205,206]. Two fatty acyl chains are located on the glycerol moiety, another fatty acyl chain is linked to the Man*p* residue attached to position 2 of *myo*-inositol, and finally, the last fatty acyl chain is attached to position 3 of the myo-inositol unit. $Ac_1PIM_4$/$Ac_2PIM_4$ seems to be a branch point intermediate in $Ac_1PIM_6$/$Ac_2PIM_6$ and LM/LAM biosynthesis. The addition of two α-(1→2)-Man residues to $Ac_1PIM_4$/$Ac_2PIM_4$, it is worth to mention that this combination is not found in the mannan backbone of LM and LAM, leads to the formation of the higher order mannosides $Ac_1PIM_6$ and $Ac_2PIM_6$[14]. Although LM and LAM have variations in their mannan chains, in both the overall size and additional acylation, they share a core based on an α-(1→6)-linked backbone of 20-25 Man residues frequently branched with single α-(1→2)-Man residues[9,207,208].

### 1.2.2.1.1. PIMs Biosynthesis

According to the currently accepted model, PIMs biosynthesis takes place on both sides of the mycobacteria plasma membrane[170,203,209] but is initiated on the cytoplasmic side by the mannosylation of PI, specifically with the transfer of a Man*p* residue from GDP-Man to the 2-position of the myo-inositol ring of PI, yielding PI monomannoside ($PIM_1$)[210]. According to genetic, enzymatic and structural evidence, the enzyme responsible for this catalytic step is PimA[211–213] (*Rv2610c* and *MSMEG_2935* in *M. tuberculosis* H37Rv and *M. smegmatis* $mc^2$155 strains, respectively). This enzyme proved to be essential in *M. tuberculosis* and *M. smegmatis* growth *in vitro* and *in vivo* (**Figure 24 b**)[211,214,215]. The second enzyme of the biosynthetic pathway, PimB (*Rv2188c* and *MSMEG_4253* in *M. tuberculosis* H37Rv and *M. smegmatis* $mc^2$155 strains, respectively), transfers a Man*p* residue from GDP-Man to the 6-position of the myo-inositol ring of $PIM_1$, generating $PIM_2$[213,216,217]. PimB was found to be essential in *M. smegmatis* growth[213]. Both $PIM_1$ and $PIM_2$ can be further acylated by the Phosphatidylinositol Mannosides Acyltransferase A or PatA (**Figure 24 b**)[170].

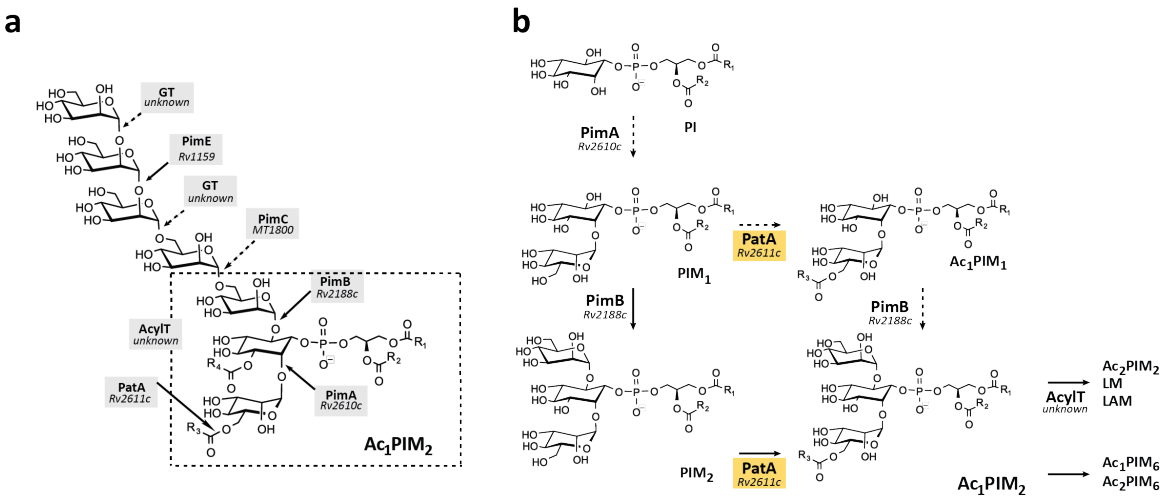



**Figure 24: PIM biosynthetic pathway in mycobacteria**. **a)** Chemical structure of $PIM_{2/6}$ and their acylated forms $Ac_{1/2}PIM_2$ and $Ac_{1/2}PIM_6$. **b)** Proposed pathway for the early steps of PIM biosynthesis in mycobacteria. The two pathways originally proposed for the biosynthesis of $Ac_1PIM_2$ in mycobacteria are shown: PI is mannosylated to form $PIM_1$. $PIM_1$ is then mannosylated to $PIM_2$, which is acylated to form $Ac_1PIM_2$; $PIM_1$ is first acylated to $Ac_1PIM_1$ and then mannosylated to $Ac_1PIM_2$. Our experimental evidence suggests that although both pathways might co-exist in mycobacteria, the PI→$PIM_1$→$PIM_2$→$Ac_1PIM_2$ pathway is favored. As an important part of the literature concerning PIMs studies refers to the nomenclature based on the *Mtb* H37Rv sequences, we also include the Rv numbers to identify the proteins.

Two pathways were originally proposed for the biosynthesis of $Ac_1PIM_2$ in mycobacteria: In the first one, PI is mannosylated by PimA to form $PIM_1$, $PIM_1$ is then mannosylated by PimB to form $PIM_2$, which is further acylated to form $Ac_1PIM_2$; and in the second one, $PIM_1$ is first acylated to $Ac_1PIM_1$ and then mannosylated to form $Ac_1PIM_2$. The experimental evidence indicates that although both pathways might co-exist in mycobacteria, the sequence of events PI→$PIM_1$→$PIM_2$→$Ac_1PIM_2$ is favoured, by the consecutive action of PimA, followed by PimB, and finally the acyltransferase PatA[170,213]. Finally, $Ac_1PIM_2$ can be further acylated at position 3 of the myo-inositol ring by the action of a second acyltransferase not described yet yielding $Ac_2PIM_2$ (**Figure 24 b**).

The degree of acylation of PIMs seems to impact on the permeability and fluidity of the cell envelope[218], as well as during the glycolipid antigen presentation to the immune system[219]. $Ac_1PIM_2$ and $Ac_2PIM_2$ can be further elongated with additional Man*p* residues to form higher PIM species (such as $Ac_1PIM_3$-$Ac_1PIM_6$/$Ac_2PIM_3$-$Ac_2PIM_6$), LM, and LAM. In *M. tuberculosis* CDC1551 strain, PimC (MT1800, GT-B, GT4 family) has been reported to transfer the third Manp residue from GDP-Man to the 6 position of the Manp residue transferred by PimB, elongating the glycolipid until $Ac_1PIM_3$/$Ac_2PIM_3$ (**Figure 20**)[220]. The fact that this enzyme has no ortholog in *M. tuberculosis* H37Rv strain nor in *M. smegmatis*, suggests that an alternative pathway with an unknown glycosyltransferase should exist in these species[170,221]. Likewise, the fourth mannosyltransferase remains to be identified.

PimE (*Rv1159* and *MSMEG_5149* in *M. tuberculosis* H37Rv and *M. smegmatis* mc$^2$155 strains, respectively), a polytopic membrane protein, was identified as the mannosyltransferase involved in the synthesis of higher forms of PIMs. Specifically, PimE transfers a Manp residue from polyprenol-phosphate-mannose to the 2 position of $Ac_1PIM_4$ to form $Ac_1PIM_5$ at the periplasmic face of the plasma membrane (**Figure 20**)[209,222,223]. The mannosyltransferase responsible for the formation of $PIM_6$ from $PIM_5$ is not yet known.

The fundamentals of how PIMs are assembled and exported are yet not fully understood. The first two or three mannosylation steps of the PIM pathway seems to occur on the cytoplasmic face of the plasma membrane. Supporting this notion, enzymes involved in the synthesis of early PIM intermediates were localized to a membrane subdomain termed Plasma Membrane free (PMf) which was clearly resolved from the cell wall[224,225]. The inner leaflet of the inner membrane is predominantly composed of $Ac_1PIM_2$/$Ac_2PIM_2$ and all these steps involve the action of GDP-Man-dependent enzymes[170,171]. Further steps in the biosynthesis of higher forms of PIMs ($PIM_5$ and $PIM_6$) appear to rely upon integral membrane GT-C-type glycosiltransferases on the periplasmic side of the plasma membrane[171,209]. Since the unassisted translocation of apolar PIMs across the plasma membrane is energetically unfavorable[226,227] such a compartmentalization implies that an unknown flippase translocates the glycolipid intermediate/s from the cytoplasmic to the periplasmic side of the plasma membrane (**Figure 20**)[167].

### 1.2.2.1.2. The Phosphatidylinositol Mannosides Acyltransferase A (PatA)

The acyltransferase PatA (*Rv2611c* and *MSMEG_2934* in *M. tuberculosis* H37Rv and *M. smegmatis* mc$^2$155 strains, respectively)[21] transfers a palmitoyl moiety from palmitoyl-CoA to the 6-position of the mannose ring linked to 2-position of inositol in $PIM_1$ or $PIM_2$ resulting in $Ac_1PIM_1$ or $Ac_1PIM_2$, repectively (**Figure 24 b**)[170].

The *M.tuberculosis* PatA gene *Rv2611c* is the third gene of a cluster of five ORFs potentially organized as a single transcriptional unit[23,228–230]. The second and fourth ORFs encode the phosphatidylinositol phosphate synthase PgsA (*Rv2612c*)[229–232] and the mannosyltransferase PimA, respectively (*Rv2610c*)[228,233]. In addition, the mannosyltransferase PimB is encoded by the *Rv2188c* gene[216,217]. Interestingly, the four enzymes involved in the biosynthetis of the metabolic end product $Ac_1PIM_2$ (PgsA, PimA, PimB and PatA) were found to be essential for the growth of *M. smegmatis* and/or *M. tuberculosis*[23,215,228,230,233]. Specifically, a *patA* mutant of *M. smegmatis* mc$^2$155 exhibited severe growth defects[23], whereas *patA* was found to be an essential gene in *M. tuberculosis* H37Rv, by a combination of transposon high-density mutagenesis and deep-sequencing approaches[234,235].

This thesis is focused on describing the three-dimensional structure, glycolipid acceptor and acyl donor substrate binding, as well as the catalytic mechanism of PatA from *M. smegmatis*. We propose a

model for substrate recognition, catalysis and enzyme inhibition using a combination of X-ray crystallography, biochemical and biophysical techniques, quantum-mechanics/molecular-mechanics (QM/MM) metadynamics and chemical synthesis. The implications of this model in the comprehension of the early steps of PIM biosynthesis and the mode of action of other members of the bacterial acyltransferase family of enzymes are discussed.

Chapter 2

# OBJECTIVES AND HYPOTHESIS

# 2. OBJECTIVES AND HYPOTHESIS

During recent times major efforts have been made to understand the early steps of the PIM biosynthetic pathway at the molecular level. In that sense, the crystal structures of the mannosyltransferases PimA and PimB were reported, showing the typical organization and catalytic machinery of GT-B glycosyltransferases. However, to date, no structural information is available for PatA, an enzyme that is a member of a large family of acyltransferases for which the molecular mechanism of substrate recognition and catalysis remains not well understood.

This thesis is focused on the structural characterization of the essential membrane associated acyltransferase involved in the biosynthesis of mycobacterial phosphatidyl-myo-inositol mannosides (PIMs) of *M. smegmatis*, the acyltransferase PatA. These studies would like to establish the mechanistic basis of substrate/membrane recognition and catalysis for an important family of acyltransferases, providing exciting possibilities for inhibitor design.

The objectives of these thesis are the following:

- ✓ to determine the crystal structure of PatA from *Mycobacterium smegmatis* in complex with the donor and acceptor substrates
- ✓ to validate the structural information by direct mutagenesis and biophysical/biochemical approaches
- ✓ to propose a model of action of the acyltransferase PatA

Chapter 3

# EXPERIMENTAL TECHNIQUES

# 3. EXPERIMENTAL TECHNIQUES

In this section, an overview of Macromolecular X-ray Crystallography is given. A careful description of the materials and other techniques used during these studies can be found in **Chapter 4. Structural Insights of the Acyltransferase PatA**.

## 3.1. Protein Crystallography

In the late 1950s, the first crystal structures of proteins were solved. Sir John Cowdery Kendrew[236] solved the first protein structure, sperm whale myoglobin[236], for which he shared the Nobel Prize in Chemistry with Max Perutz in 1962. Since that success, about 120,791 X-ray crystal structures of proteins, nucleic acids and other biological molecules have been determined, and 14 Nobel prizes in chemistry or medicine have been awarded to protein crystallographers. These structures are important for the understanding of the processes that take place in living organisms and also for practical applications such as drug design.

Over the years, X-ray crystallography has become one of the most powerful tecnique for the structural resolution of macromolecules. It has great advantages in comparison with other techniques focused on the same purpose, since it provides information on an atomic scale, obtaining a 'three-dimensional photograph' of the crystallized material. Regarding to proteins, their folding and three-dimensional structure are obtained, as well as important physical and chemical information such as distances and bonding angles, atom packing and vibrational thermal state, all of which are crucial for the comprehension of their biological function.

An alternative technique used in structural biology for determining macromolecular structures is Nuclear Magnetic Resonance (NMR). The main advantatge of this technique is that macromolecules are in solution thus, dymanics can be determined. However, the structural information attainable by NMR is limited when working with large macromolecules (above *ca.* 200 kDa). Additionally, other techniques such as electron microscopy (EM) or atomic force microscopy (AFM) offer alternatives to the structural study of macromolecules. Finally, bioinformatics techniques can also provide structural information about a protein if structural data of this protein or similar is already available. It is worth mentioning a new technique in the field of X-ray crystallography, X-ray Free Electron Laser (XFEL), that works at the femtosecond time scale. XFEL consists of a free-electron laser that generates high-intensity electromagnetic radiation by accelerating electrons to relativistic speeds and directing them through special magnetic structures. The extremely bright X-ray pulses produced are approximately 1,000 million times higher than the one currently produced in synchrotrons. The technique is still in

early stages and its implementation is expected to be a qualitative jump in the field of structural biology[237].

From 1958 to the present, the number of macromolecular structures has grown exponentially and 90% of them has been solved by X-ray crystallography. However, this collection of structures represents only a small portion of the hundreds of thousands of different proteins that play some role in the processes of life. Therefore, structural determination of proteins is an essential part of the research developed in hundreds of laboratories around the world.

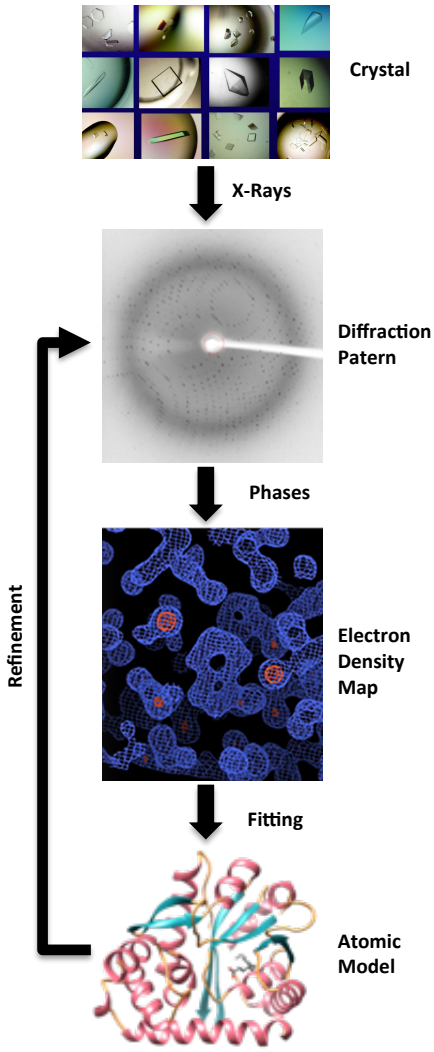



**Figure 25: Workflow for solving a molecular structure using X-ray crystallography.**

The obtention of crystals is the first step for the application of X-ray crystallography technique. These crystals need to be sufficiently large, unique and homogeneous for the diffraction experiments. Obtaining single crystals of sufficient quality for diffraction is currently the bottleneck of molecular crystallography (**Figure 25**).

A crystal is defined as a solid body constituted by an ordered group of atoms, ions or molecules that are repeated and regularly distributed in space. This minimum unit of repetition is called unit or elementary cell. The unit cell is defined by three axes (a,b,c) and three angles (α,β,γ), which determine the reference angle of the atomic coordinates. Furthermore, there are other repetition operations, in addition to the repetitions by translation, that cause repetitions within the elementary cell: those are the elements of symmetry. An asymmetric unit is defined as the minimum set of atoms that, applying the symmetry operations of a given space group, generates the total content of atoms in the unit cell (**Figure 26**).

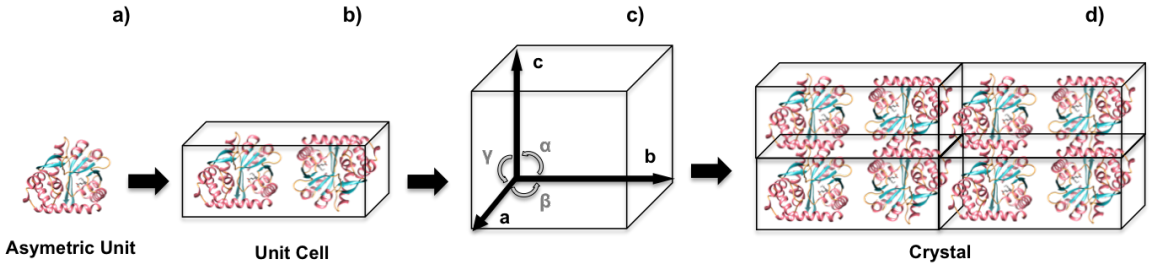



**Figure 26: The crystal concept. a)** The asymmetric unit, **b)** repeats itself through the elements of symmetry, generating the unit or elementary cell. **c)** The parameters that characterize its shape and size are three axes (a,b,c) and three angles (α,β,γ). **d)** The repetition of the elementary cell, in the three directions of space, generates the crystal.

The set of elements of symmetry in a crystal and the different cell types give rise to the space groups, which represent 230 different ways in which a motif can be repeated in a three-dimensional space. These 230 groups are classified in the International Tables for X-ray Crystallography. Among those 230 groups, only 65 groups that do not contain symmetry planes or inversion centers are compatible with biological macromolecules because of the intrinsic chirality of these molecules.

It is important to highlight that protein crystals present certain peculiarities in comparison with the inorganic crystals and small molecules. One remarkable difference is that, in general, protein crystals are smaller and more brittle than non-protein crystals. Moreover, they have weaker optical properties and tend to dehydrate easily due to the high content of solvent (between 25-90%, 50% average) of crystals. What *a priori* seems to be a problem, this high percentage of solvent (mostly disordered) that usually results in weak diffraction pattern, could also be an advantage because the presence of solvent maintains the protein in an environment similar to that found in solution[238].

## 3.1.1. Principles of Crystallisation

In general, crystallisation of a macromolecules is not very different from the crystallisation of small molecules. Crystals of any substance are formed from a supersaturated solution, which triggers the spontaneous formation of the initial crystallisation nucleus. When the solubility limit (supersaturation) is exceeded, the molecules that are in excess are separated from the aqueous phase

and move to the solid phase, establishing an equilibrium between both phases. Therefore, the supersaturation conditions are necessary for the crystal formation, since with only saturation this change would not occur, and thus, the crystal would not be formed (**Figure 27**).

So, the strategy is to obtain a supersaturated solution that generates a few crystallisation nuclei on which a few large and ordered crystals grow. In other words, the protein should be maintained near the solubility curve, but in the supersaturation zone (**Figure 27**).

Crystallisation consists on two well differentiated phases: nucleation and crystalline growth. On one hand, thermodynamically stable crystalline aggregates appear in nucleation. On the other hand, protein molecules are incorporated into the surface of the crystal during the growth, displacing part of the water molecules. This second phase ends when a thermodynamic equilibrium is achieved or when crystal defects accumulate. Nucleation can occur spontaneously or can be induced artificially. Therefore, the study of nucleation is a fundamental part of crystalline growth. Several theories have been proposed to explain this phenomenon, among them, the Gibbs-Thompson equation is elementary to explain nucleation events (**Figure 27**).

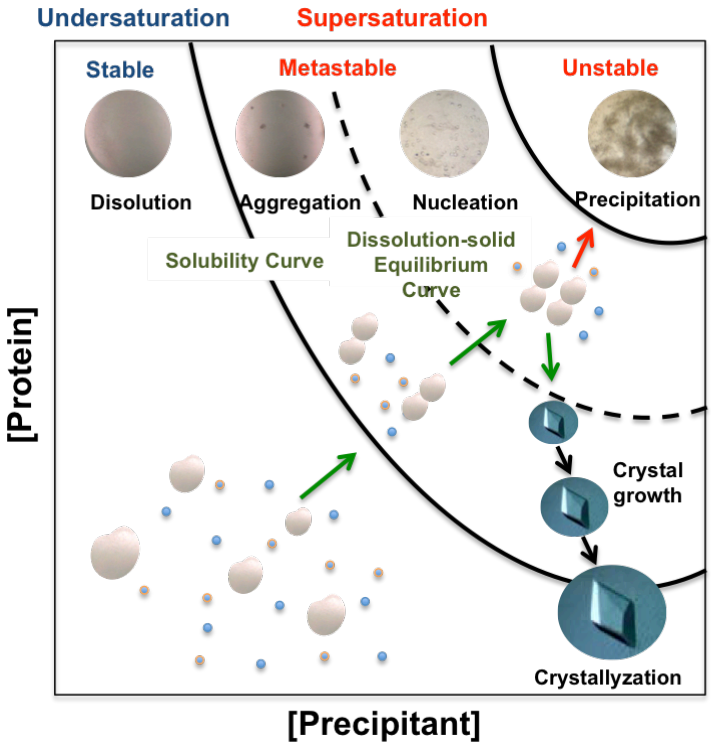



**Figure 27: Diagram of phases of crystallisation process.** The aqueous and saturated phases are represented in addition to the zones of aggregation, nucleation, crystalline growth and precipitation. The green arrows show the usual evolution of a vapor diffusion experiment to the formation of crystals. If it formed precipitate, it would follow the path of the red arrow. The blue circles represent the solvent and the oranges the precipitate.

## 3.1.2. Requirements Previous to Crystallisation

The first thing to consider for crystallisation is whether the sample fullfil some important requirements. There are five main aspects to have into account: concentration, purity, homogeneity, stability and quantity[239].

- ✓ **Concentration**: As supersaturation is needed, so a high concentration of protein is required. Concentrations of 5-20 mg/ml are typically considered depending on protein solubility.
- ✓ **Purity**: Maximum purity is required, although 95% is acceptable. This is usually determined by electrophoresis in the polyacrylamide gel. The impurities can stop the growth of the crystal and produce irregularities in it, which translates into a worse quality of the diffraction pattern.
- ✓ **Homogeneity**: The sample must be homogeneous and any factor that leads to heterogeneity must be evaluated and considered before the crystallisation experiment.
- ✓ **Stability**: A special care must be taken not to alter the stability of the protein during the purification, freezing/thawing and crystallisation processes.
- ✓ **Quantity**: It is important to have the maximum amount of protein as possible in order to do the maximum crystallisation trials. It is not recommended to use protein sample from different batches of purification, since this variable can be crucial in the crystallisation process.

Once the protein meets all the requirements, the most popular strategy used to achieve supersaturation is based on: increasing the protein concentration by solvent evaporation, and incorporating additives to decrease protein solubility. Usually, the factors that affect the obtention of saturated protein solutions are the following: protein concentration, pH and solution buffer, temperature, ionic strength, precipitating agent type and concentration, additives and ligands.

## 3.1.3. Crystallisation Methodology

There are several techniques for protein crystallisation: vapour diffusion, dialysis, microbatch under oil and counterdiffusion. The most common technique is vapour diffusion which is the one used in this thesis.

**Vapour Diffusion**

In this system, a drop of protein solution together with a drop of the precipitating agent are placed in a closed container, and are equilibrated against the precipitating solution contained in a reservoir. In this way, as the precipitant concentration in the drop is less than the one in the reservoir, evaporation occurs. It results in the decrease of the volume of the drop and an increase of the protein concentration until it reaches the equilibrium. Thus, first the saturation conditions will be reached and

then supersaturation, getting into the nucleation zone. The final objective is to reach the nucleation and crystalline growth zone, and this could be fulfilled by knowing the system and adjusting the appropriate parameters in different experiments (**Figure 28**).

There are two experimental modalities for the assembly of this system (**Figure 28**):

- **Sitting Drop**: The drop of protein and precipitant is placed on a flat or concave elevation next to the reservoir of precipitating solution. The system is finally sealed from the outside with a flexible adhesive sheet that covers the entire well.
- **Hanging Drop**: The drop of protein and precipitant is placed on a coverslip, which is previously siliconized to prevent dispersion. The precipitant solution is deposited in a well which will be closed with the coverslip containing the attached droplet in an inverted position. Vacuum grease is used to seal it tightly.

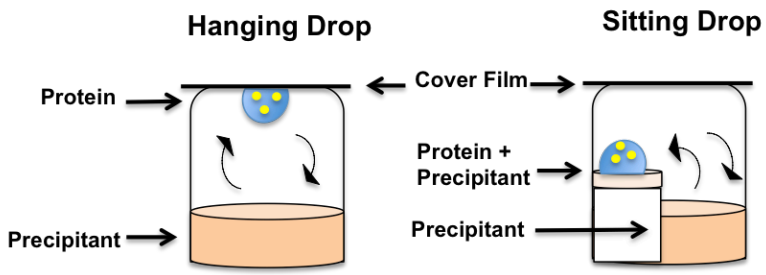



**Figure 28: Diagram of vapour diffusion crystallisation technique in sitting drop and hanging drop.**

## 3.1.4. Optimal Crystallisation Conditions

Crystallisation involves two main experimental stages: search for crystallisation and optimization conditions.

Protein crystallisation experiments require some preliminary trials that allow testing as much conditions as possible. There are multiple variables that affect crystallisation that must be taken under consideration, such as chemical (pH, ionic strength, concentration, impurities, etc.), physical (temperature, pressure, vibration, etc.) and biochemical factors (purity, measurements, ligands, etc.) must be taken under consideration.

Nowadays, high throughput techniques allow the completion of multiple experiments in a short time and also to spend the minimum amount of sample. Moreover, there are a large number of commercial kits that allow analysing, in a single experiment, almost 100 different crystallisation conditions. The crystallisation conditions often contain a buffer that maintains the pH and one or more precipitating agents. In the last years, the appearance of robots to automate and monitor the search

for conditions and to store crystallisation plates have increased the number of crystallized biomolecules.

**Crystal Optimization**

In general, the crystals obtained in this first stage are small and/or have imperfections, so it is necessary to improve their quality. Thus, the most common practice is scaling. This technique consists in increasing the work volumes of the protein and precipitant. The aim is to improve crystals size and their degree of order, to finally improve the quality of the diffraction pattern.

At the same time as trying to reproduce the crystals, small changes are usually introduced into the crystallisation solution (pH, temperature, precipitant concentration, etc.), the proportion of the drop components (protein/precipitant; 1:2, 2:1, 1:3, etc.), and small amounts of additives or ligands in order to improve the quality of the crystals.

When the use of the above strategies is not effective, a technique called seeding can be useful[239]. This technique consists in reducing the supersaturation of the system with the aim of placing it in the metastable zone inside the solubility curve. Then, previously formed nuclei or micro crystals are added, in such a way that the protein present in the solution can be incorporated into them and generate bigger crystals. This seeding technique include three variants: micro-seeding, macro-seeding and streak-seeding.

- **Micro-seeding** consists in crushing a defective crystal previously obtained, and using its nuclei as a basis for the growth of new crystals[240].
- **Macro-seeding** is based on manipulating a crystal of acceptable size but with imperfections. The technique consists in performing surface washings of the crystal and using it as a crystallisation nucleus.
- **Streak-seeding** consists in introducing a natural fibre into a solution with crystalline nuclei or small crystals, and dragging the nuclei attached to the fibre to the solution of interest.

## 3.1.5. Protein-Ligand Complex Crystallisation

To get a better understanding of the mechanism of action of a biomolecule, a key point is to have the description of the three-dimensional structure of the protein interacting with ligands, such as substrates or products of the reaction. For the protein-ligand complexes crystallisation there are two main strategies: soaking and co-crystallisation[241].

- **Soaking**: The crystals of protein are introduced into a solution containing the ligand at high concentration. The ligand can diffuse through the channels that already exist in the crystal to get to the active or binding site of the crystallized protein. The main disadvantage of this

strategy is that many times, crystals are damaged by the entry of the substrate or, the substrate cannot reach the active/binding site due to the crystal packaging.

- **Co-crystallisation**: In this strategy, the ligand is added to the protein drop at the beginning of the crystallisation experiment. In general, the same crystallisation conditions of the apo protein can be used, but many times, for example due to conformational changes, new crystallisation conditions must be found.

## 3.1.6. Cryo-crystallography

To determine the crystal structure of a macromolecule, a complete X-ray diffraction data set is required. Typically, this involves exposing crystals to X-rays radiation over a few minutes, leading to radiation damage in the crystal and reduction of diffraction quality. This radiation effect can be mitigated at cryogenic temperatures. Nowadays, X-ray diffraction experiments are carried out at temperatures of about 100 K, thus the crystalline sample is able to withstand intense beams of light such as those produced by synchrotron radiation.

One limitation of this technique is the risk of the formation of ice crystals, which affects the diffraction pattern, and can even destroy the crystal. To avoid this, the sample must be treated with cryopreservation conditions. Cryo-protectants prevent the formation of ice crystals[242]. In general, the selection of an appropriate cryo-protectant is based on the composition of the crystallisation condition of the protein. The most common cryoprotectants are usually effective at a concentration between 15-25% (v/v) and are included in the mother solution of crystallisation. The most common are glycerol[243], ethylene glycol, mineral oil and PEG 400, but some salts are also effective, such as malonate[244].

# 3.2. X-Ray Diffraction

## 3.2.1. The Diffraction Phenomenon

X-ray diffraction is the result of the interaction of X-rays with the electrons of the atoms in the crystal. This interaction produces an electron-vibration coupling with variation of the electric field. Thus, part of the X-rays pass through the crystal and part are scattered by the electrons. These scattered photons can be annulled (if their waves interfere) or, due to their periodic nature, they can emerge in phase and reinforce themselves, generating the X-ray diffraction spectrum. This phenomenon can be explained geometrically by the Bragg's law (William Lawrence and William Henry Bragg, Nobel Prize in Physics in 1915):

$$2d_{hkl}sen\theta = n\lambda$$

**Equation 1. Bragg's law.**

In this expression, different parallel atomic planes (*hkl*) separated by a distance (*d*) are considered, with an angle of incidence of the X-ray beam and the wavelength of the beam ($\lambda$) (**Figure 29**).

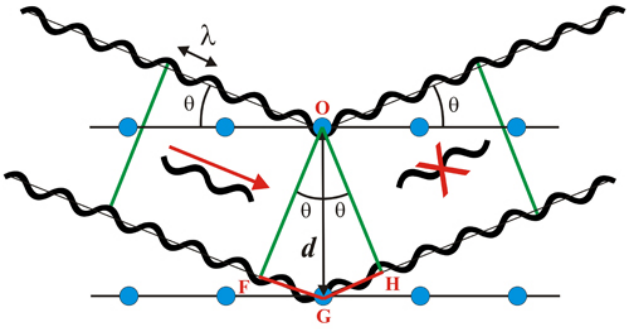



**Figure 29: Bragg's law scheme.** The wave that falls on the lower plane delays a whole number of wavelength and leaves in phase with the upper wave. In this way, it reinforces the signal.

To better explain the diffraction phenomenon, the reciprocal network concept and the Ewald model are used (**Figure 30**). In this model, each family of Bragg's planes is replaced by a geometric point, so that all the families of planes contribute to a reciprocal network, closely related to the crystal lattice. In this way, the diffraction will occur when, during turning the crystal in data collection, each one of the points of the reciprocal network interact with the surface of the Ewald sphere (whose radius is equal to the inverse of the wavelength). Finally, the diffracted beam can be collected by placing the detector in the proper position (**Figure 30**).

To perform the diffraction experiment, an X-rays source, a goniometer, a liquid nitrogen based low temperature system and a detector are required.

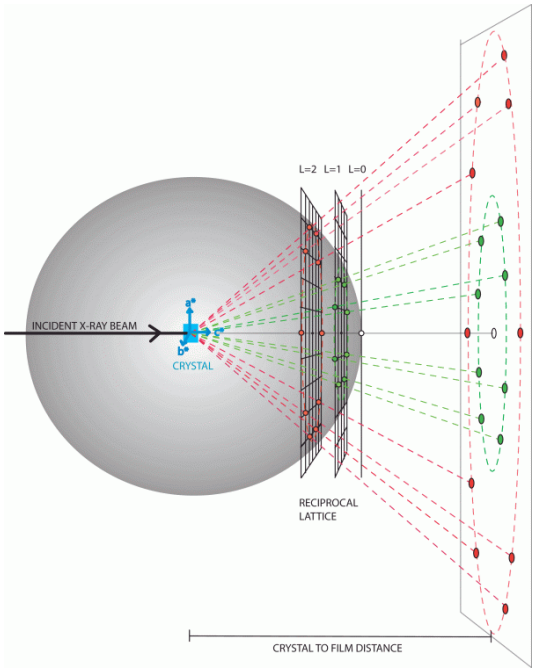



**Figure 30: Representation of the Ewald's sphere.** When a reciprocal point meets Ewald's sphere Bragg's law is fulfilled.

## 3.2.2. X-Ray Radiation Sources

The history of X-rays began with William Crookes in the 19$^{th}$ century. This physical-chemist was studying the generation of high voltage currents in a vacuum tube when he saw that if it was close to photographic films, fuzzy images were generated in them. From there, Wilhem Röntgen continued with Crookes's studies and discovered these unknown (until then) X-rays. With this discovery he was awarded with the Nobel Prize in Physics in 1901.

X-rays are extraordinarily penetrating electromagnetic waves, that pass through certain bodies, produced by the emission of the internal electrons in the atom. In this way, they are generated by the rapid deceleration of very energetic electrons when hitting a metallic target. Normally, the most used wavelength in protein crystallography is of the order of 1 Å, suitable for atomic distances. In general, two main X-ray sources are used: anode generators and synchrotron radiation stations.

- **Rotatory Anode Generators**: The conventional diffractometers typically found in small molecule crystallography laboratories contain high voltage tubes with a metal plate (usually copper) situated at the end, which is bombarded with accelerated electrons generated from an incandescent filament. When these excited electrons return to the fundamental state, X-rays are emitted with a frequency proportional to the energy jump that triggered their excitation. This process generates a lot of heat and therefore the rotatory anode generators were introduced, in

which the metallic anode is a cylinder that rotates to achieve a better cooling, allowing to obtain much more intense X-rays than from a conventional tube.

- **Synchrotron Radiation**: A synchrotron is a particle (usually electrons) accelerator in which these move in a circular orbit by the action of a magnetic field and are accelerated at speeds close to light by an electric field synchronized with orbital motion. At the moment in which the rectilinear trajectory is broken the electrons emit a very energetic radiation, synchrotron radiation, which encompasses different wavelengths, from microwaves to X-rays (**Figure 31**).

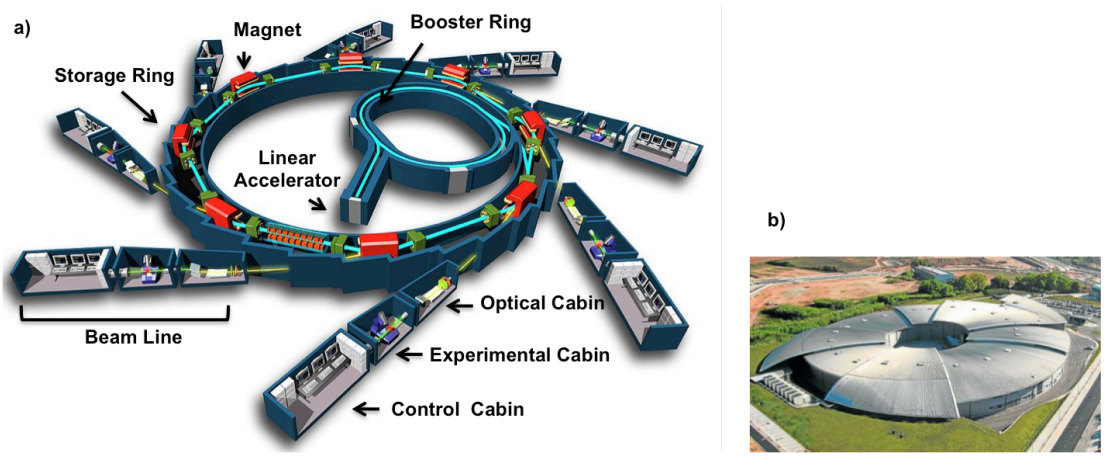



**Figure 31: X-ray radiation synchrotron source. a)** General scheme of a synchrotron. **b)** Exterior image of ALBA Synchrotron (Barcelona)**.**

### Cryo-Protection

The protein and solvent atoms can suffer irreparable damage due to radiation, because heat and free radicals that degrade the crystal are generated[245]. To solve this, a low temperature system is used. This system cools the crystal with a stream of liquid nitrogen reaching temperatures of 100 K. This method can increase the average life of the crystal and its resolution, since flexible zones are fixed in the protein due to the decrease of its thermal vibration. To work at these temperatures the crystal must be previously cryo-protected. After setting the correct cryoprotection of the crystal, it is captured in a nylon or similar material loop, and placed in a device called the goniometer. The goniometer has several degrees of freedom, therefore it allows to orient the crystal in small intervals of 0.2-1$^{o}$ while the X-rays strike it, so that the crystal can be rotated to obtain a complete set of diffraction data (**Figure 32**).

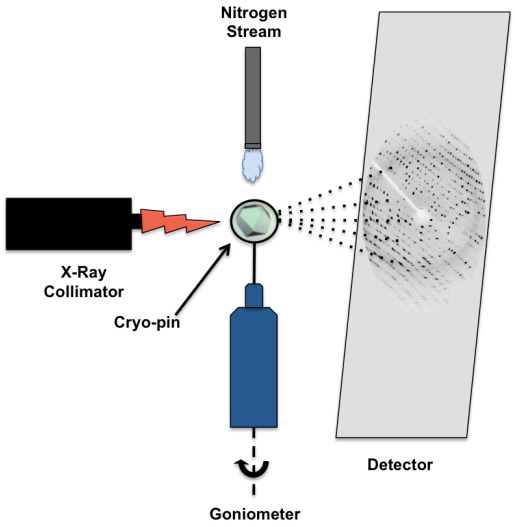



**Figure 32: Diagram of the assembly and data collection of a crystal.**

### Detector

Currently, almost all detectors used in macromolecular crystallography are area detectors. The three best known are:

- **Image Plate Detectors (IP):** There is a plate that contains a lanthanide salt. Atoms are excited when they are irradiated by a photon from the diffracted X-ray beam. Next, a laser sweeps the plate which causes a light emission proportional to the number of photons absorbed. This light is registered by a photomultiplier and becomes an image of the diffraction pattern. The process lasts about 4-9 minutes and requires to erase the plate in each measurement, so its use is limited to rotating anode generators.
- **Charged-Coupled Device Detectors (CCD):** They are based on the photoelectric effect and convert the received light in the form of photons into electric current. Therefore, they consist of a sensor with small photoelectric cells that register the image. These devices are more sensitive and have shorter exposure times than IP detectors. They are usually used in synchrotron installations.
- **Pixel Array Detectors (PAD):** This device contains a very sensitive silicon layer that absorbs X-rays and converts them into an electrical signal. The reading system is extremely efficient, so they are present in synchrotron installations. Another great advantage is a very little background noise.

## 3.2.3. Data Collection

In this stage, a series of consecutive diffraction images are collected for different crystal positions. The goal is to get a complete data set. It is essential to use a good data collection strategy

since the subsequent process of structural resolution will depend on the quality of data. To do this, a set of parameters must be controlled such as exposure time, angle of rotation per image (Δφ), angle of total rotation, distance to the detector, etc.[246].

# 3.3. Structure Determination

## 3.3.1. Data Processing

Once the diffraction experiment is finished, all the images obtained are processed to extract the net intensities of the diffracted beams (**Figure 33**).

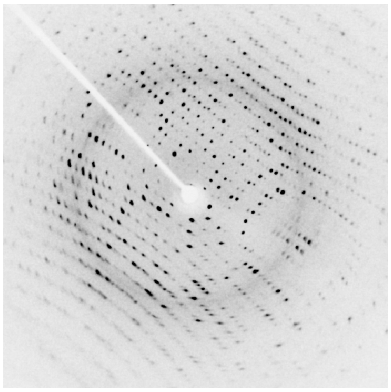


**Figure 33: Diffraction Pattern.**

This process is constituted by the following stages:

- **Indexing**: The geometric arrangement (position) of the reflections in the diffraction pattern allows us to extract information from the unit cell of the crystal lattice and the symmetry of the crystal. Thus, the Miller indices (*hkl*) are assigned to each of the reflections, and the unit cell, space group and the mosaicity determined[247,248] (the mosaicity estimates the grade of disorder in a crystal).

- **Integration**: In this stage the intensities of each reflection and relative error are estimated. Therefore, we obtain the set of all the reflections (*hkl*) together with their intensities ($I_{hkl}$) and their estimated errors ($\sigma_{hkl}$).

- **Scaling**: In this stage the intensities are scaled and normalized, reducing the data to a set of unique reflections – reflections not related to each other by a symmetry element. This normalization is necessary due to experimental factors (temperature, intensity of the source and decrease of the intensities by destruction of the crystalline order). Moreover, due to the symmetry contained in the crystal there is a high redundancy of data, which allows obtaining a set of statistical parameters that give an idea of the quality of the data in terms of resolution. They are the following:

- **Multiplicity**: It is the number of times that a certain reflection has been recorded. The higher is this value, the better the estimate of errors associated with the reflection measure.
- **Completeness**: Indicates the percentage of reciprocal space that has been analysed.
- **<II/σ(I)>**: Represents the relationship between intensity and background noise. Values greater than *ca.* 2 for the last resolution layer are typically accepted. The expression is (being N the number of reflections):

$$< \frac{|I|}{\sigma(I)} >= \frac{1}{N} \sum_{hkl}^{N} \frac{|I_{hkl}|}{\sigma(I_{hkl})}$$

**Equation 2. Relationship between intensity and background noise.**

- **$R_{merge}$ and $R_{pim}$ Factors**: $R_{merge}$ evaluates the difference between the intensity of a reflection measured N times and its average (**Equation 3**). There is a new indicator called $R_{pim}$, which also takes into account the total number of reflections (**Equation 4**). Maximum values of *ca.* 0.3-0.6 for the last resolution shell (high-resolution data) are typically accepted.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I_{(hkl)i} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl}}$$

**Equation 3. $R_{merge}$.**

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_{i=1}^{N} |I_{(hkl)i} - \bar{I}_{(hkl)}|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl}}$$

**Equation 4. $R_{pim}$.**

- $CC_{1/2}$ and CC: Those are two parameters based on the calculation of the Pearson correlation coefficient (CC)[249]. These new parameters consider that if the data set is divided in two groups, the errors of each one would be random.

$$CC^{*} = \sqrt{\frac{2CC_{1/2}}{1 + CC_{1/2}}}$$

**Equation 5. Pearson correlation coefficient.**

- **Structure Factor Calculation ($F_{hkl}$)**: It is defined as the result of the dispersion of the X-rays by all the atoms of the cell in a determined direction. The resultant is a wave so it has an associated vector that can be described by its module, $|F_{hkl}|$, and its phase $\phi_{hkl}$. The modules represent the amplitudes of the diffracted beams and are obtained directly from the collected intensities. However, the $\phi_{hkl}$ phases cannot be measured directly and different methods are required for their calculation.

The analytical expression of the structure factors is shown in the **Equation 6**. It shows the magnitude *f*, which is the atomic scattering factor and represents the amount of X-rays atomic electrons are able to scatter:

$$F_{hkl} = \sum_j f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

**Equation 6. Structure Factor ($F_{hkl}$).**

## 3.3.2. Solving the Phase Problem

The transformation of the reciprocal space (diffraction spectrum) to the crystalline space (electron density function) involves the resolution of a function called the Fourier transform. This equation relates the electron density, $\rho_{xyz}$ for each point of the crystalline unit cell, with the reciprocal space, which is defined by the structure factors $F_{hkl}$. It is expressed as:

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F_{hkl}|\, e^{-2\pi i(h_x + k_y + l_z - \emptyset_{hkl})}$$

**Equation 7. Fourier transform.** Where V is the volume of the unit cell**.**

Therefore, to calculate the electronic density maps on which the atomic model is going to be built, it is necessary to know both the amplitudes associated with the diffracted beam $|F_{hkl}|$ extracted from the diffraction pattern, and its phase $\phi_{hkl}$ which is unknown. This constitutes the so-called "phase problem" in protein crystallography, which can be solved by three methods:

i) **Molecular Replacement (MR)**: A homologous protein of known structure is used as the initial structural model. This method is based on the assumption that homologous proteins in peptide sequence have similar folding and, therefore, can be used as structural models. This strategy is also used to solve the structure of protein-ligand complexes in the case that the crystals belong to a different space group than the crystals of the protein. The technique consists in placing the known macromolecule in the crystal of the unknown protein, for which it is necessary to know its orientation and specific position. The procedure is based on the use of the Patterson function, which is essentially the Fourier transform of the intensities rather than the structure factors, and the calculation of rotation and translation functions (**Figure 34**).

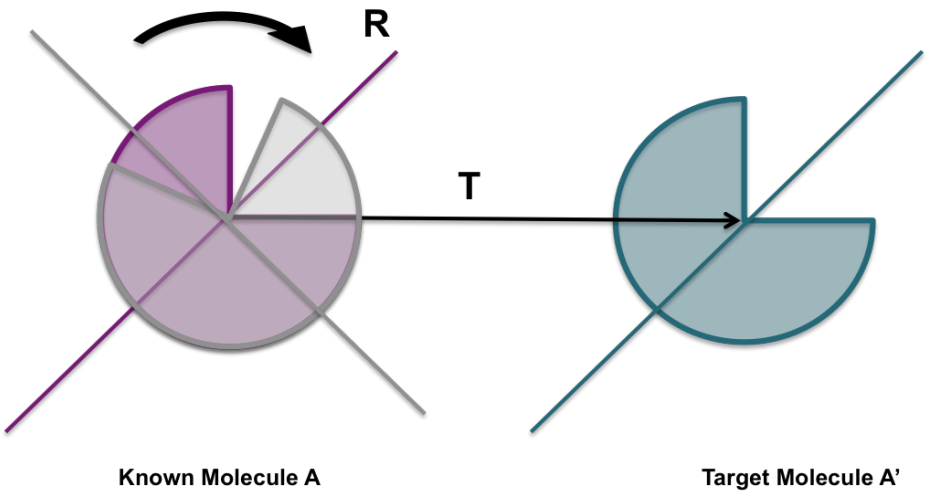



**Figure 34: Outline of molecular replacement method.** The molecule of known structure and model protein (A) is rotated by operation (R) and is transferred by operation (T), to bring it to the position of the unknown molecule (A'). Next step is evaluating if this is the correct position or if it should keep looking for the right position.

ii) **Single/Multiple Isomorphous Replacement (SIR/MIR)**: Heavy atoms with high atomic scattering factors are introduced in the crystal. Using the Patterson function and due to the difference of diffracted intensities between the heavy atom-containing crystal (derivative) and without heavy atoms (native), the position of the heavy atoms inside the unit cell can be determined (known as the substructure). From the positions of these heavy atoms, phases can be estimated to all reflections in the dataset. SIR: a single crystal; MIR: several derivative crystals.

iii) **Single/Multi-wavelength Anomalous Diffraction (SAD/MAD)**: The effect of X-ray absortion by heavy atoms can be exploited to obtain phase information experimentally. To perform a SAD/MAD experiment, incident radiation is matched to that of the absorption maximum of the heavy atom introduced in a derivative crystal. As a result, the atomic scattering factor is affected by i) a reduction in scattering power, since some of the incident radiation is absorved and therefore not able to scatter, and ii) a phase-delay of 90 degrees in the scattered radiation. These changes in the atomic scattering factors can be used to determine the substructure of a derivative crystal by measuring X-ray diffraction data at one (SAD) or several (MAD) wavelengths, requiring only one derivative crystal.

In addition, the isomorphic replacement and anomalous dispersion methods (SIRAS/MIRAS) can be combined. Those are very powerful techniques to solve *ab initio* the phase problem for novel proteins[250].

### 3.3.3. Phases Improvement

Once the problem of the phase has been solved, the electronic density function can be calculated. The graphic representation of this function originates the electronic density map, which provides the necessary information for building of a protein model. Normally, the initial phases are just an approximation, and the improvement of the electronic density map is needed for interpretation.

The two most used techniques for the improvement of phases are:

- **Solvent Flattening**: This technique is based on the creation of a mask that separates protein and solvent, setting a constant value for the density of the solvent.
- **Non-Crystallographic Symmetry Averaging**: This technique averages density when there are several copies of the protein inside the asymmetric unit, reducing the noise and increasing the constraints on the phases.

### 3.3.4. Model Building and Refinement

Once the best possible density map ($2F_o$-$F_c$) has been calculated, the polypeptide chain is modelled on the electron density maxima, which is usually done manually. However, in case of having high resolution data the automatic procedures work well. In cases where MR has been used, the initial model contains an important bias towards the homologous protein and it is necessary to alternate a series of atomic coordinate refinement cycles, with a manual building. Manual reconstruction is the only way to eliminate serious errors in the model that refinement cannot correct.

Refinement involves the adjustment of several parameters such as the coordinate of each atom (x,y,z), the thermal B-factor (thermal vibration) and the occupancy (fraction of molecules that occupy a position), in order to adjust as much as possible the model to the information obtained from the diffraction data. During refinement, stereochemical constraints are imposed to maintain distances, bond angles and appropriate torsions. The refinement is usually initiated as a rigid body model in which the protein is considered as a whole, and is followed by statistical models of maximum likelihood. In the last stages, it continues with the addition of ligands, ions, other molecules present in the crystal and solvent molecules. A complementary technique is to apply simulated cooling models using molecular dynamics, especially in the initial stages since they have a better degree of convergence and can be effective in eliminating important model bias.

To evaluate the fit of the refined model into the experimental data, the reliability or residual factor (**Equation 8**) is calculated. The *R*-factor measures the difference between the observed amplitudes $F_o$ and calculated $F_c$ from a model.

$$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$$

**Equation 8. *R*-factor.**

To avoid bias derived from over-parametrization during the refinement (overfitting), a statistical method of cross validation is applied. To do this, the diffraction data is divided into two groups and two *R*-factors are introduced. The first group of reflections contains 90-95% of the data (working set, $R_{work}$) and are used in the model refinement. The second group contains the remaining 5-10% (test set, $R_{free}$) and is used to verify that the refinement course actually leads to a model that better fits the experimental data[251].

Finally, differential electron density maps ($F_o$-$F_c$) are used in the manual construction of the model. In those the peaks of positive density reveal that the model is incomplete, while the negatives peaks show incorrect positions in the model.

The model can be validated when there are no residues of unmodeled electron density, the geometry is correct, and the $R_{work}$ and $R_{free}$ values are appropiate ($R_{work}$ <20-25% and $R_{free}$ <30%).

## 3.3.5. Model Validation

The validation constitutes the last stage of the structural resolution. In this phase the geometry of all the molecules must be checked as well as their correspondence with the electron density. Thus, the possible existence of clashes between atoms in the crystal, distances and bonding angles of the protein, conformation of the acceptable rotamers for the different amino acids, and torsion of the peptide bond (angles *phi,ϕ*: N-Cα and *psi, ψ*: Cα-C) that must meet the requirements of Ramachandran can be determined (**Figure 35**)[252].

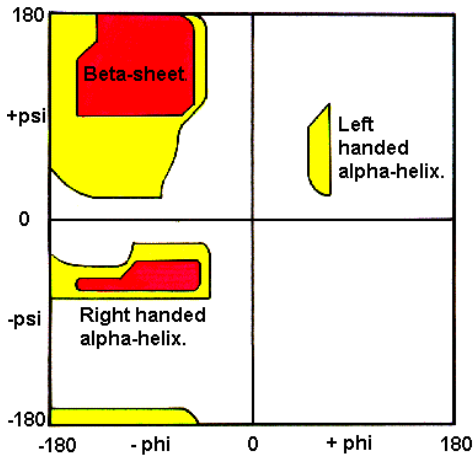



**Figure 35: Ramachandran Plot.** The most favourable conformations in terms of *phi,ϕ* and *psi, ψ* of the peptide bond are highlighted.

The final model is deposited in the Protein Data Bank (PDB) database, where it is subjected to new tests. This file has a .pdb extension and contains the crystallographic information and coordinates of the macromolecule, which can be visualized with numerous graphic softwares. Once the file is reviewed and approved, an identification code is assigned and becomes of public access.

Chapter 4

# STRUCTURAL INSIGHTS OF THE ACYLTRANSFERASE PatA

# 4. STRUCTURAL INSIGHTS OF THE ACYLTRANSFERASE PatA

## 4.1. Materials and Methods

### 4.1.1. Expression and Purification of PatA from *M. smegmatis*

#### 4.1.1.1. Truncated Version of PatA from *M. smegmatis*

A truncated version of *M. smegmatis* PatA lacking the first 12 residues (PatA; MSMEG_2934) was ligated into pJAM2 vector. *M. smegmatis* $mc^2155$ cells transformed with the corresponding plasmid pJAM2-PatA were grown in LB medium supplemented with 0.05% (v/v) Tween 80 10% and 50 µg/ml of kanamycin at 37°C for 4 days. Then, LB medium supplemented with 0.05% (v/v) tween 80 10% and 50 µg/ml of kanamycin was inoculated for an $OD_{600}$=0.2 at 30°C and 180 rpm for 4-6 hours until it reached a final $OD_{600}$=0.5-6. Finally, expression of PatA was induced by the addition of 0.2% acetamide 40% at 30 °C and 120 rpm overnight (ON). Cells were collected at 5,000 xg for 15 minutes and resuspended in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM imidazole (solution A) containing protease inhibitors (Complete EDTA-free; Roche). Then, cells were disrupted by sonication in 15 cycles of 60 seconds pulses with 90 seconds cooling intervals between the pulses. PatA was centrifuged at 20,000 xg for 20 minutes. The supernatant was filtered first with a 0.4 µm and finally with a 0.22 µm filter and then the truncated version of PatA was applied to a HisTrap Chelating column (1 ml, GE HealthCare) previously equilibrated with solution A. The column was then washed with solution A until no absorbance at 280 nm was detected. Elution was performed with a linear 20 ml gradient of 40–250 mM imidazole in solution A at 1 ml/min. The resulting PatA preparation displayed a single protein band when run on a 12% NuPAGE Bis-Tris precast gel stained with SimplyBlue SafeStain (Invitrogen). The purified recombinant PatA protein was concentrated and diluted several times for a buffer exchange to eliminate the imidazole by using a Amicon-Ultra spin concentrator (Merck Millipore) with a 30-kDa-molecular mass cutoff and the final buffer 50 mM Tris-HCl pH 7.5, 500 mM NaCl.

#### 4.1.1.2. PatA Full-Length from *M. smegmatis*

The full length version of the acyltransferase PatA (MSMEG_2934)[21,24] was purified. *M. smegmatis* $mc^2155$ cells transformed with the corresponding plasmid pJAM3-PatA were grown in MM63 medium (15 mM $(NH4)_2SO_4$, 10 mM $KH_2PO_4$, 1 mM $MgSO_4.7H_2O$ and 17 mM succinic acid pH 7.0) supplemented with 0.05% tween 80 10% and 50 µg/ml of kanamycin at 37°C for 4 days. Then, MM63 medium (15 mM $(NH4)_2SO_4$, 10 mM $KH_2PO_4$, 1 mM $MgSO_4.7H_2O$ and 17 mM succinic acid pH 7.0) supplemented with 0.05% tween 80 10% and 50 µg/ml of kanamycin was inoculated for an $OD_{600}$=0.2 at 30°C and 180 rpm for 4-6 hours until it reached a final $OD_{600}$=0.5-6. Finally, expression of

PatA was induced by the addition of 0.2% Acetamide 40% at 30$^{o}$C and 120 rpm overnight (ON). Cells were collected at 5,000 xg for 15 minutes and resuspended in 50 mM Tris-HCl pH 7.5 and 500 mM NaCl (solution A) containing protease inhibitors (Complete EDTA-free, Roche). The cells were disrupted by sonication in 15 cycles of 60 seconds pulses with 90 seconds cooling intervals between the pulses. PatA was solubilized from the mycobacterial membrane by the addition of 2 mM CHAPS and 10% glycerol (solution B). The suspension was gently stirred during 1 hour at 4$^{o}$C and centrifuged at 20,000 xg for 30 minutes. The supernatant was applied to 5 ml TALON Superflow Metal Affinity Resin (Clontech) previously equilibrated with solution B. The suspension was gently stirred during 1 hour at 4$^{o}$C. Finally, the suspension is packed by gravity in a plastic column and washed with 30 ml of 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10% glycerol (solution C) and 40 mM imidazole. Elution was performed by gravity gradient with steps of 20 ml of solution C with 100 mM imidazole, 200 mM imidazole and 500 mM imidazole and fractions of 1 ml are manually taken. The resulting fractions are analysed by 12% NuPAGE Bis-Tris precast gel stained with SimplyBlue SafeStain (Invitrogen), and fractions that correspond to PatA molecular weight are concentrated and diluted several times for a buffer exchange to eliminate the imidazole by using a Amicon-Ultra spin concentrator (Merck Millipore) with a 30-kDa-molecular mass cutoff and the final buffer 50 mM Tris-HCl pH 7.5 and 500 mM NaCl.

## 4.1.2. Crystallisation and Data Collection

### 4.1.2.1. PatA-$C_{16}$ Complex

Three crystal forms were obtained, referred thereafter as PatA-$C_{16}$-1, PatA-$C_{16}$-2 and PatA-$C_{16}$-3. The first and second crystal forms were obtained by mixing 0.25 ml of the truncated version of PatA at 5 mg/ml in 20 mM Tris-HCl pH 7.5 with 0.25 ml of mother liquor containing 100 mM Tris-HCl pH 7.0, 230 mM $MgCl_2$ and 12-16% (w/v) PEG 8,000 using the sitting drop vapour diffusion method. Crystals grew in 7-15 days and were transferred to a cryo-protectant solution containing 25% ethylene glycol and frozen under liquid nitrogen. Complete X-ray diffraction data sets were collected at beamline I03 (Diamond Light Source, Oxfordshire, UK) and processed with XDS program[253]. The second crystal form of PatA diffracted to a maximum resolution of 2.9 Å and crystallized with four molecules in the asymmetric unit and space group $P2_1$ (PatA-$C_{16}$-2). The first crystal form of PatA diffracted to a maximum resolution of 2.06 Å and crystallized with two molecules in the asymmetric unit and space group C2 (PatA-$C_{16}$-1). The third crystal form was obtained by mixing 0.25 ml of the truncated version of PatA at 5 mg/ml in 25 mM Tris-HCl pH 7.5, 150 mM NaCl with 0.25 ml of mother liquor containing 100 mM Tris-HCl pH 8.5 and 20% ethanol. Crystals grew in 3 days and were transferred to a cryo-protectant solution containing 25% sucrose and frozen under liquid nitrogen. A complete set of X-ray diffraction data were collected at beamline X06SA Swiss Light Source (Villigen, Switzerland) and processed with XDS program[253]. PatA-$C_{16}$-3 crystal diffracted to a maximum resolution of 2.43 Å and

crystallized with one molecule in the asymmetric unit and space group $P4_22_12$ (**Table 3**).

| | PatA-$C_{16}$-1 | PatA-$C_{16}$-2 | PatA-$C_{16}$-3 |
|---|---|---|---|
| **Beamline** | I03 (DLS) | I03 (DLS) | PXI-X06SA (SLS) |
| **Wavelength (Å)** | 0.97625 | 0.97625 | 0.99996 |
| **Resolution range (Å)** | 56.60 – 2.06 | 92.37 – 2.90 | 40.19 – 2.43 |
| **Space group** | *C* 2 | *P* $2_1$ | *P* $4_2$ $2_1$ 2 |
| **Unit cell** | 93.60 71.09 76.2 90 91.95 90 | 70.83 84.48 98.80 90 110.78 90 | 80.30 80.30 113.77 90 90 90 |
| **Total reflections** | 96616 (4273) | 73942 (5870) | 153438 (11538) |
| **Unique reflections** | 30533 (1973) | 23420 (1759) | 14652 (1057) |
| **Multiplicity** | 3.2 (2.2) | 3.2 (3.3) | 10.5 (10.9) |
| **Completeness (%)** | 98.5 (86.7) | 96.2 (99.2) | 100.0 (100.0) |
| **Mean I/sigma(I)** | 10.6 (2.0) | 6.6 (1.7) | 4.6 (1.1) |
| **Wilson B-factor** | 21.81 | 54.57 | 41.10 |
| **R-sym** | 0.068 (0.410) | 0.115 (0.508) | 0.122 (0.669) |
| **R-factor** | 0.1714 (0.2609) | 0.2389 (0.3409) | 0.2236 (0.2854) |
| **R-free** | 0.2320 (0.3323) | 0.2737 (0.4251) | 0.2793 (0.3599) |
| **Number of atoms** | 4441 | 7720 | 2098 |
| **Macromolecules** | 4055 | 7548 | 1989 |
| **Ligands** | 37 | 72 | 22 |
| **Water** | 349 | 100 | 87 |
| **Protein residues** | 509 | 1003 | 256 |
| **RMS(bonds)** | 0.004 | 0.005 | 0.003 |
| **RMS(angles)** | 0.90 | 0.99 | 0.76 |
| **Ramach. favored (%)** | 99 | 99 | 97 |
| **Ramach. outliers (%)** | 0 | 0 | 0.39 |
| **Clashscore** | 1.49 | 2.89 | 1.01 |
| **Average B-factor** | 20.10 | 36.60 | 53.70 |
| **Macromolecules** | 19.40 | 36.90 | 53.90 |
| **Ligands** | 26.00 | 23.30 | 55.90 |
| **Solvent** | 26.40 | 20.60 | 46.60 |

**Table 3: PatA-$C_{16}$ complex data collection and refinement statistics.** Statistics for the highest-resolution shell are shown in parentheses.

### 4.1.2.2. PatA-S-$C_{16}$CoA Complex

One crystal form was obtained by mixing 0.25 μl of the truncated version of PatA at 5.1 mg/ml in 1 mM S-hexadecyl Coenzyme A (S-$C_{16}$CoA; stock solution at 10 mM in 20 mM Tris-HCl pH 7.5) and 20 mM Tris-HCl pH 7.5 with 0.25 ml of mother liquor containing 100 mM HEPES pH 7.5, 500 mM ammonium sulfate and 30% (v/v) 2-methyl-2,4-pentanediol using the sitting drop vapour diffusion method. Crystals grew in 7-15 days and were directly frozen under liquid nitrogen. A complete X-ray diffraction data set was collected at beamline I03 (Diamond Light Source) and processed with XDS program[253]. PatA-S-$C_{16}$CoA diffracted to a maximum resolution of 3.28 Å and crystallized with four molecules in the asymmetric unit and space group $P2_1$ (**Table 4**).

| PatA-S-$C_{16}$CoA | |
|---|---|
| **Beamline** | I03 (DLS) |
| **Wavelength (Å)** | 0.978000 |
| **Resolution range (Å)** | 47.12 - 3.28 |
| **Space group** | P $2_1$ |
| **Unit cell** | 81.08 80.27 97.38 90 110.92 90 |
| **Total reflections** | 60657 (5816) |
| **Unique reflections** | 17931 (1708) |
| **Multiplicity** | 3.4 (3.4) |
| **Completeness (%)** | 99.13 (94.36) |
| **Mean I/sigma(I)** | 11.18 (1.47) |
| **Wilson B-factor** | 103.69 |
| **R-sym** | 0.091 (0.898) |
| **R-factor** | 0.2640 (0.4051) |
| **R-free** | 0.3060 (0.4495) |
| **Number of atoms** | 7597 |
| **Macromolecules** | 7400 |
| **Ligands** | 174 |
| **Water** | 23 |
| **Protein residues** | 1011 |
| **RMS(bonds)** | 0.004 |
| **RMS(angles)** | 1.08 |
| **Ramach. favored (%)** | 99 |
| **Ramach. outliers (%)** | 0 |
| **Clashscore** | 2.46 |
| **Average B-factor** | 111.90 |
| **Macromolecules** | 111.30 |
| **Ligands** | 139.40 |
| **Solvent** | 86.30 |

**Table 4: PatA-S-$C_{16}$-CoA data collection and refinement statistics.** Statistics for the highest-resolution shell are shown in parentheses.

### 4.1.2.3. PatA-$C_{16}$-Man Complex

One crystal form was obtained by mixing 0.25 μl of the truncated version of PatA at 5 mg/ml in 5 mM D-(+)-mannose and 20 mM Tris-HCl buffer pH 7.5 with 0.25 μl of a mother liquor containing, 100 mM Tris-HCl pH 7.0, 225 mM $MgCl_2$ and 16% (w/v) PEG 8,000 using the sitting drop vapour diffusion method. Crystals grew in 7-15 days and was transferred to a cryo-protectant solution containing 25% ethylene glycol and frozen under liquid nitrogen. on a PILATUS3 6M pixel detector (100 Hz max. frame rate) at the microfocus I24 beamline (λ=0.9686 Å -Diamond Light Source, UK) and integrated with XDS[253] following standard procedures. PatA-$C_{16}$-Man diffracted to a maximum resolution of 2.42 Å and crystallized with four molecules in the asymmetric unit and space group $P2_1$ (**Table 5**).

| PatA-$C_{16}$-Man | |
|---|---|
| **Beamline** | I24 (DLS) |
| **Wavelength (Å)** | 0.9686 |
| **Resolution range (Å)** | 28.77-2.41 (2.50-2.41) |
| **Space group** | P 1 21 1 |
| **Unit cell** | 70.69 83.86 98.43 90 110.98 90 |
| **Total reflections** | 133880 (12180) |
| **Unique reflections** | 40216 (3754) |
| **Multiplicity** | 3.3 (3.3) |
| **Completeness (%)** | 97.17 (90.55) |
| **Mean I/sigma(I)** | 12.76 (2.44) |
| **Wilson B-factor** | 42.4 |
| **R-merge** | 0.1125 (0.4939) |
| **$CC_{1/2}$** | 0.989 (0.773) |
| **CC*** | 0.997 (0.934) |
| **R-work** | 0.1817 (0.2748) |
| **R-free** | 0.2127 (0.3055) |
| **Number of non-H atoms** | 8086 |
| **Macromolecules** | 7824 |
| **Ligands** | 117 |
| **Protein residues** | 1009 |
| **RMS (bonds)** | 0.003 |
| **RMS (angles)** | 0.57 |
| **Ramachandran favored (%)** | 98 |
| **Ramachandran allowed (%)** | 2.3 |
| **Ramachandran outliers (%)** | 0 |
| **Rotamers outliers (%)** | 0.91 |
| **Clashscore** | 1.68 |
| **Average B-factor** | 44.52 |
| **Macromolecules** | 44.42 |
| **Ligands** | 54.39 |
| **Solvent** | 42.10 |

**Table 5: PatA-$C_{16}$-Man data collection and refinement statistics.** Statistics for the highest-resolution shell are shown in parentheses.

## 4.1.3. Structure Determination and Refinement

### 4.1.3.1. PatA-$C_{16}$ Complex by Single-wavelength Anomalous Diffraction (SAD)

PatA crystals of form C2 (PatA-$C_{16}$-1) were soaked with 10 different platinum salts at 1 mM concentration for a time period of 130-145 min (HR2-442, Hampton Research) followed by flash freezing in liquid-$N_2$. Anomalous data were collected at the theoretical L–I absorption edge of Pt (13,879.9 eV-0.8933 Å). Oscillation images were collected at I04 beamline (Diamond Light Source) with an oscillation angle of 0.2 for a total of 1,800 images using a Pilatus 6M-F pixel detector. Data were collected with an attenuated X-ray beam (5% transmission) and a $0.04^{-s}$ exposure time per image. Data were integrated and scaled in XDS and experimental phases determined using the SHELXC/D/E package[254]. Data of a PatA crystal soaked with 1 mM $K_2PtCl_4$ for 130 min were used for experimental phasing with a 2.5-Å data cutoff applied, giving a mean value $|\Delta F|/\sigma(\Delta F)$ in the highest resolution shell of 0.9. The substructure determination located two Pt atoms in the asymmetric unit (CC=33.64, CC(weak)=20.71 and CFOM=54.35). Experimental phases were determined and subsequently used for initial cycles of model building and density modifications by SHELXE. Buccaneer[255] and the CCP4 suite[256] were used for further model extension (**Table 6**).

The structure determination of PatA-$C_{16}$-1, PatA-$C_{16}$-2 and PatA-$C_{16}$-3 were carried out by molecular replacement using Phaser[257] and the PHENIX suite[258] and the PatA-Pt structure as model. Followed by cycles of manual rebuilding and refinement using Coot[259] and phenix.refine[260], respectively. The structures were validated by MolProbity[261].

| PatA·Pt | |
|---|---|
| **Beamline** | I04 (DLS) |
| **Wavelength (Å)** | 0.8933 |
| **Photon energy (keV)** | 13.8799 |
| **Exposure (s)** | 0.040 |
| **Transmission (%)** | 5.01 |
| **Beamsize (µm)** | 90 x 45 |
| **Total images** | 1800 |
| **Ω Osc (°)** | 0.20 |
| **Resolution range (Å)** | 78.09 - 1.88 |
| **SigAno # > 0.88 (Å)** | 78.09 - 2.33 |
| **Space group** | *C*2 |
| **Unit cell** | 95.99 71.04 78.200 90 93.01 90 |
| **Total reflections** | 254696 (8905) |
| **Unique reflections** | 41213 (2343) |
| **Multiplicity** | 6.2 (3.8) |
| **Completeness (%)** | 96.4 (74.6) |
| **Mean I/sigma(I)** | 15.0 (2.1) |
| **Wilson B-factor** | 20.44 |
| **R-sym** | 0.076 (0.462) |

**Table 6: Data collection for experimental phasing.** # Mean anomalous difference in units of its estimated standard deviation (|F(+)-F(-)|/Sigma).

#### 4.1.3.2. PatA-S-$C_{16}$ Complex by Molecular Replacement

The structure determination of PatA-S-$C_{16}$CoA was carried out by molecular replacement using Phaser[257] and the PHENIX suite[258] and the PatA-Pt structure as model. Followed by cycles of manual rebuilding and refinement using Coot[259] and phenix.refine[260], respectively. The structures were validated by MolProbity[261].

#### 4.1.3.3. PatA-$C_{16}$-Man Complex by Molecular Replacement

Structure determination of PatA was resolved using as a template the previously reported PatA structure in complex with palmitate (PDB 5F2T)[21] and molecular replacement methods implemented in Phaser[257] and the PHENIX suite[258]. Model rebuilding was carried out with Buccaneer[255] and the CCP4 suite[256]. The final manual building was performed with Coot[259] and refinement with phenix.refine[260]. The structure was validated by MolProbity[261]. Atomic coordinates and structure factors have been deposited with the Protein Data Bank, accession code 5OCE. Molecular graphics and structural analyses were performed with the UCSF Chimera package[262].

### 4.1.4. N-Terminal Sequence Analyses

Samples were applied to a NuPAGE 4–12% gel. The gel was then washed with NuPAGE transfer buffer (Invitrogen) for 15 min at room temperature. Proteolytic fragments were electrotransferred to a PVDF membrane using the iBlot dry blotting system (Invitrogen) for 7 min. The PVDF membrane was then washed for 10 min with Milli-Q purified water. Bands were stained with a solution containing 0.1% Coomassie brilliant blue R-250, 40% methanol and 1% acetic acid and subjected to N-terminal sequence analysis using Applied Biosystems 494 procise cLC protein sequencer in the Functional Genomics Center Zurich at the ETH Zurich, Switzerland. Two sequences were found, MVTLSG and VTLSGR, indicating that the purified protein corresponds to full-length PatA, with the N-terminal Met cleaved off by about 50%.

### 4.1.5. PatA Acyltransferase Activity Assay

The truncated version of PatA transferase activity was measured in the assay with mycobacterial membranes[24]. Briefly, *M. smegmatis* $mc^2155$ cells were broken by sonication and the membrane (100,000 xg pellet) fraction was obtained by differential centrifugation. The reaction mixtures contained 250 mg of membrane proteins, 1.2 mg of purified $PimA_{SM}$, 10 mg of purified PatA or the mutated versions, 0.1 mCi GDP-[$^{14}$C]mannose (specific activity of 55 mCi/mmol, ARC Inc.), 0.12 mM palmitoyl-CoA (Sigma-Aldrich) in DMSO with final concentration in the reaction mixture 2% (v/v), 62 mM ATP, 10 mM $MgCl_2$, and 25 mM Tris-HCl pH 7.5 in the final volume of 50 ml. Reactions were incubated 100 min at 37$^{o}$C and stopped with 300 ml of $CHCl_3$/$CH_3OH$ (2:1, by volume). The samples

were left rocking 30 min at room temperature, and centrifuged at 1,500 xg for 10 min. The organic phase (bottom) was analysed by thin layer chromatography on aluminium-coated silica 60 $F_{254}$ plates (Merck) developed in $CHCl_3/CH_3OH$/conc. $NH_4OH/H_2O$ (65:25:0.5:4), and quantified by scintillation spectrometry[24]. All enzymatic activity measurements were determined in duplicates. Following the same procedure, palmitate was assayed as a possible inhibitor or substrate at different concentrations.

## 4.1.6. PatA Palmitoyl-CoA Hydrolytic Activity Assay

### 4.1.6.1. Hydrolytic Activity of the Truncated PatA in the Absence $PIM_2$

The hydrolytic activity of the truncated version of PatA and PatA mutants against palmitoyl-CoA was measured as following a methodology described for other acyltransferases[263]. A typical reaction contained 20 mM Tris-HCl pH 8.3, 0.2 mM disulfide [5,5'-dithiobis-(2-nitrobenzoic acid)] (DTNB), 0.2 mM palmitoyl-CoA and 4 µM PatA or its variants. All reactions were carried out at 37°C in a CARY 300 Bio UV Visible/Spectrophotometer. The spectrum of the product $TNB^{-2}$ formed after the interaction of the DTNB and the CoA liberated from the hydrolysis of the substrate palmitoyl-CoA by the enzyme, was measured continuously at 412 nm during 20 min. All enzymatic activity measurements were determined in duplicates.

### 4.1.6.2. Hydrolytic Activity of the Full-length PatA in the Presence of $PIM_2$

Enzymatic activity of full-length PatA against palmitoyl-CoA by using diacylated and deacylated $PIM_2$ as acceptors of the reaction was measured as following a methodology previously described for other acyltransferases[263]. Reactions were performed in: i) 5% (v/v) DMSO in order to solubize $PIM_2$ or ii) 1 mM DOPC-SUVs with incorporation of $PIM_2$ by vortexing the mixture for 4 min. Both reactions contained 20 mM Tris-HCl pH 8.3, 0.6 mM disulfide [5,5'-dithiobis(2-nitrobenzoic acid)] (DTNB), 0.2 mM palmitoyl-CoA, 0.2 mM $PIM_2$ or decetylated $PIM_2$ and 3.5 µM PatA. All reactions were carried out at room temperature in a UV-Vis NanoDrop™ One/OneC. The spectrum of the product $TNB^{-2}$ formed after the interaction of the DTNB and the CoA liberated from the hydrolysis of the substrate palmitoyl-CoA by the enzyme, was measured continuously at 412 nm during 20 min. All enzymatic activity measurements were determined in triplicates.

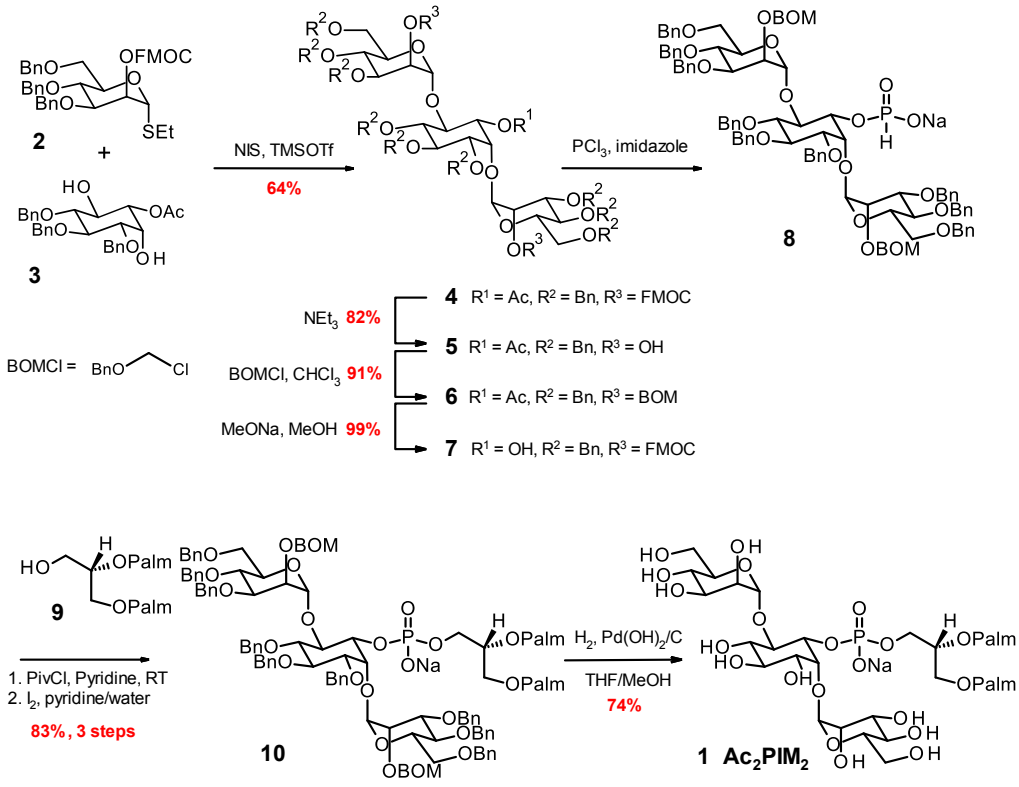
## 4.1.7. Preparation of Small Unilamellar Vesicles (SUVs)

1,2-dioleoyl-sn-glycero-3-phosphoglycerol (DOPC) lipid vesicles were prepared in 20 mM Tris-HCl pH 8.3. A dry lipid film was prepared from a chloroform:methanol (2:1) DOPC solution by removing the solvent under a stream of $N_2$ and high vacuum. DOPC SUVs were prepared by dispersion of the lipids in buffer by vigorous vortexing and following sonication with a probe tip sonicator (MSE Soniprep 150 (MSE, UK)) for 10 min (5 seconds on and off cycles) at 100 µm amplitude.

## 4.1.8. Chemical Synthesis of $PIM_2$

$PIM_2$ (**1**) was obtained using a route adapted from published procedures[264,265], starting from mannose donor **2**, known inositol derivative **3** and (*S*)-1,2-di-*O*-palmitoyl glycerol **9**[266]. **2** was easily obtained in two steps from ethyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside.[267] Removal of the 2-*O*-acetate group in basic methanol (MeONa in MeOH) and protection of the liberated alcohol with FMOC chloroformate and pyridine in dichloromethane gave **2** in 75% overall yield. Chiral inositol **3** was prepared from methyl α-D-glucopyranoside following a biomimetic approach described in the literature[268]. Finally, known dipalmitoylglycerol **9** was prepared from commercially available (*S*)-(+) 1,2-isopropylidene glycerol by a four-steps sequence. Benzylation of the primary position of (*S*)-(+) 1,2-isopropylidene glycerol (BnBr, NaH, DMF), acidic hydrolysis of the ketal (trifluoroacetic acid in a THF/ $H_2O$ mixture, 88% yield for the two steps), bis-palmitoylation with palmitic acid and DCC/DMAP in refluxing toluene (90%) and removal of the benzyl protecting group by hydrogenolysis ($H_2$, catalytic $Pd(OH)_2/C$, 70%) gave **9**.

Treatment of *myo*-inositol derivative **3** with 3.0 equiv. of mannose donor **2** under NIS/TMSOTf activation in dichloromethane gave the bis-α-glycosylated inositol **4** in good yield. The FMOC protecting groups were removed with $NEt_3$ in THF (82%) and the liberated alcohol functions of **5** were protected with freshly distilled benzyloxymethyl chloride (BOM-Cl) in a concentrated chloroform solution to give **6** in high yield. These experimental conditions are crucial to get a high yield of the bis-BOM ether. Elaboration of the phosphodiester bond of **1** was carried out without purification of the intermediates. Removal of the acetate group on **6** was done by treatment of **6** with sodium methanolate in methanol and **7** was obtained in quantitative yield. Alcohol **7** was then transformed into H-phosphonate **8** with $PCl_3$, triethylamine and imidazole in toluene. After work-up of the reaction, crude H-phosphonate **8** was reacted with **9** under pivaloyl chloride activation in pyridine, followed, without work-up, by oxidation of the intermediate H-phosphonate to the phosphate **10** with iodine in a pyridine/water mixture. **10** was obtained in a rewarding 83% isolated yield for the three steps from **7**. Final deprotection of all protecting groups was carried out in one step by hydrogenolysis with 20% palladium hydroxide on carbon as catalyst in a 2/3 THF/methanol mixture and gave the targeted $Ac_2PIM_2$ **1** in 74% isolated yield.




For further details please see **Annex I.**

## 4.1.9. Preparation of Deacylated $PIM_2$

Deacylated $PIM_2$ was prepared by mild alkali hydrolysis of acylated $PIM_2$ purified from *M. tuberculosis* (BEI Resources, NIAID, NR-14846). To this end, 1 mg of dried sample was resuspended in 100 µl of a chloroform:methanol (2:1). After the addition of 100 µl of 0.2 M NaOH in methanol, the sample was incubated during 30 min at 37°C. Then, the sample was carefully neutralized with glacial acetic acid to pH 7.0. The deacylated $PIM_2$ was extracted by sequentially adding 200 µl of chloroform:methanol (2:1), 200 µl of chloroform and 100 µl of water. The sample was mixture by using a Vortex device and then centrifuged at 5,000 xg for 5 min. The upper aqueous phase contained the deacylated form of $PIM_2$. $PIM_2$ extraction was repeated by adding 100 µl of water to the bottom organic phase. The two water phases were pooled and the identity of deacylated $PIM_2$ was confirmed by MALDI-TOF.

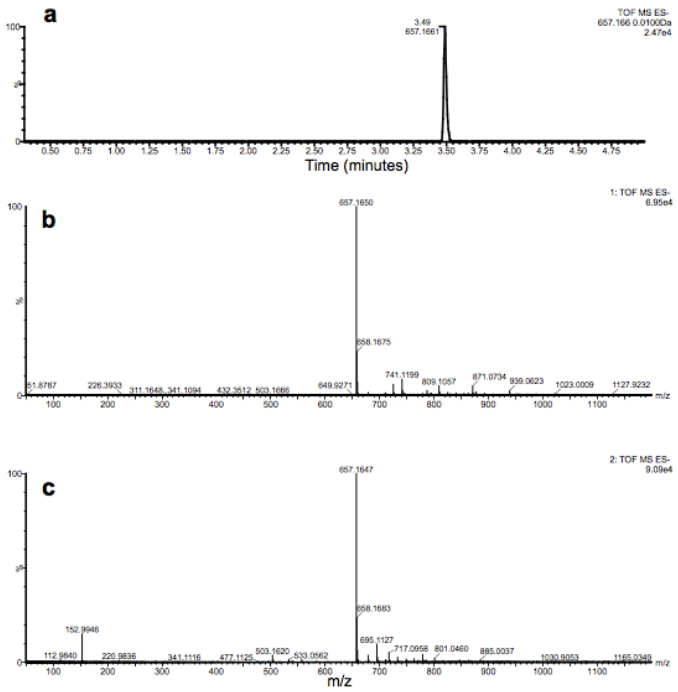



**Figure 36: Mass spectrometry of deacylated $PIM_2$. a)** Extracted ion chromatogram at m/z 657.166, representing $PIM_2$ deacylated. **b)** Low collision energy spectrum of the XIC peak at 3.49. **c)** High collision energy spectrum of the XIC peak at 3.49.

## 4.1.10. Molecular Docking Calculations

The crystal structure of mycobacterial PatA in complex with S-$C_{16}$CoA (PatA-S-$C_{16}$CoA) was investigated using the structure preparation function in MOE2013.08[269]. First, S-$C_{16}$CoA coordinates in PatA-S-$C_{16}$CoA crystal structure (chain A) were used to generate palmitoyl-CoA atomic coordinates by substitution of $C_{16}$ atom with a carbonyl group. Then the model was prepared using the Amber12EHT force field, an all-atom force field, combining two-dimensional Extended Hueckel Theory (EHT) and Amber12 force field, with Born solvation, and hydrogen atoms were added using Protonate3D function[270,271]. The docking site of $PIM_1$ and $PIM_2$ was defined by 14 residues of the PatA-S-$C_{16}$CoA crystal structure. $PIM_2$ structure was retrieved from the PDB (ligand code XPX) and after substitution of acyl chains with acetyl groups, the molecule was energy minimized using MOE with a 0.1 kcal/mol·Å r.m.s. gradient threshold. $PIM_1$ structure was constructed by removing one mannose ring of $PIM_2$. The $PIM_1$ and $PIM_2$ structures were submitted to conformational search using LowModeMD with default parameters in MOE2013.08[272]. The same procedure was carried out with $PIM_1$, in which a mannose residue GOLD (Genetic Optimization for Ligand Docking; Cambridge Crystallographic Data Center (CCDC), version 5.2.2) was used with default genetic algorithm parameter settings[273,274] for all calculations, with the exception that the search efficiency parameter was set to 200% to improve predictive accuracy by calculating the optimal number of genetic algorithm operations for the ligand due to their large flexibility. The ASP scoring function implemented in GOLD was used to rank the

docked poses; this fitness function has been optimized for the prediction of ligand-binding positions[275,276]. $PIM_1$ and $PIM_2$ were docked into the PatA-palmitoyl-CoA complex. Only docking poses having a C6-hydroxyl group of the mannose closed to the thioester of palmitoyl-CoA were kept for analysis. The best solutions were assessed by their respective docking score and by visual inspection.

### 4.1.11. Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations

The initial structure of the PatA-$PIM_2$-$C_{16}$-CoA enzymatic complex was obtained by molecular docking in our previous work.

Protonation states of all ionisable residues were selected from close inspection of intermolecular interactions. The total model was composed by 49,756 atoms, including 11 Na+ counterions and 15,192 water molecules surrounding the enzyme.

Classical molecular dynamics (MD) simulations were performed to equilibrate the ternary complex using the Amber11 software. The FF99SB[277], GLYCAM06[278], and TIP3P[279] force-fields were selected to describe the protein, carbohydrates and water molecules, respectively. Gaff parameters[280] were used to describe $C_{16}$-CoA. Weak spatial constraints were applied to the Man1 sugar in the first 25 ns and released afterwards. Analysis of the trajectories was carried out using standard tools of AMBER[281] and VMD[282].

QM/MM MD simulations were done using the Car-Parrinello code[283] which is based on Density Functional Theory (DFT). The QM zone was composed by 84 atoms in total, including the reactive mannose (Man1) and half of the inositol ring that is attached to it, a segment of the Palmitoyl-CoA substrate, the catalytic histidine (His126) and also the charge relay glutamate Glu200. Capping hydrogens were used to define the QM/MM interface. The QM region was enclosed in an isolated supercell of size 22.0×20.0×20.7 Å[253]. Kohn–Sham orbitals were expanded in a planewave basis set with a kinetic energy cut-off of 70 Ry. Norm-conserving Troullier-Martins[284] *ab initio* pseudopotentials were used for all elements. The calculations were performed using the Perdew, Burke, and Ernzerhoff (PBE) generalized gradient-corrected approximation. A fictitious electronic mass of 700 au and a time step of 5 au were used to ensure an adiabaticity of $7.4 \cdot 10^{-5}$ au·$ps^{-1}$ ·$atom^{-1}$ for the fictitious kinetic energy. The reaction free energy landscape (FEL) was explored using metadynamics[284]. Three collective variables, as described in the main text, were used to monitor the reaction. A hill height of 1.0 kcal/mol and the deposition time of 30 fs (250 MD steps) was considered. Gaussian widths were tuned according to the oscillations of each collective variable (0.45 Å, 0.45 Å and 0.65 Å for CV1, CV2 and CV3 respectively). An upper wall at 3.2 Å was set up for the first collective variable, and lower walls at 2.6 Å and -3.2 Å for the second and the third collective variables. The first step of the reaction was observed

after 511 deposited Gaussians, reaching the reaction intermediate. The simulation was stopped after the deposition of 367 more Gaussians in the intermediate well, when the S-CoA leaving group, negatively charged, flipped up and performed strong hydrogen bonds with Arg201. The reaction transition state was further verified by isocommittor analysis using 20 trajectories. To reduce the computational cost of the 3D exploration, we took a snapshot of the reaction intermediate and we performed a second metadynamics simulation in which we onlyt activated CV3. In this case, since we expected a lower free energy barrier, we decreased the hill height to 0.6 kcal/mol, keeping the same deposition time as in the previous metadynamics simulation. The Gaussian width was set to 0.20 Å and an upper and lower wall were defined at 2.0 Å and -2.0 Å. First crossing criterion was used to stop the simulation[285].

## 4.1.12. Sequence Alignment and Structural Analysis

The sequence of PatA from *M. smegmatis* (A0QWG5) was subjected to basic local alignment search tool (BLAST) and several orthologues in *Mycobacteria* sp. were found. Afterwards, they were aligned using the ClustalW server (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

The structure weighted sequence alignment was performed using PROMALS3D[286]. For labelling the conserved and similar residues, BoxShade server was used (http://embnet.vital-it.ch/software/BOX_form.html). Structural analysis and graphics for publications were performed with PyMOL (version 0.99)[287] and Chimera[262].

## 4.2. Results and Discussion

The comprehension of the early steps of PIM biosynthetic pathway has been a major goal during the recent years due to its importance into the search of drug targets against *M. tuberculosis*. The crystal structures of the glycosyltransferases PimA and PimB have been reported[15,16,25,212]. However, to date, no structural information is available for PatA, an enzyme that is a member of a large family of acyltransferases for which the molecular mechanism of substrate recognition and catalysis remains not well understood[2,3,17,18]. Specifically, PatA transfers a palmitate at position 6 of the Man*p* residue, transferred by PimA, from Palmitoyl-CoA to $PIM_1$ or $PIM_2$, to form $Ac_1PIM_1$ and $Ac_1PIM_2$[15].

This chapter provides a detailed investigation of the three-dimensional structure, glycolipid acceptor and acyl donor substrate binding, as well as the catalytic mechanism of PatA from *M. smegmatis*, as well as a model for substrate recognition, catalysis and enzyme inhibition. The implications of this model in the comprehension of the early steps of PIM biosynthesis and the mode of action of other members of the bacterial acyltransferase family of enzymes are discussed.

The references of this chapter are based on the publications, mentioned in "Publications" section, *Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA* (Albesa-Jové *et al*. 2016) and *The Molecular Mechanism Of Substrate Recognition And Catalysis Of The Membrane Acyltransferase PatA* (Tersa *et al.* 2017).

## 4.2.1. Overall Structure of PatA

Despite much effort, we were unable to produce recombinant PatA from *M. tuberculosis*. In contrast, recombinant PatA from *Mycobacterium smegmatis* was purified to homogeneity following the protocol described in section **4.1. Materials and Methods**. The amino-acid sequences of *M. tuberculosis* and *M. smegmatis* versions of PatA displayed 74% sequence identity and 84% sequence similarity. All residues that participate in both of the catalytic and substrate recognition mechanisms are strictly conserved between both proteins.

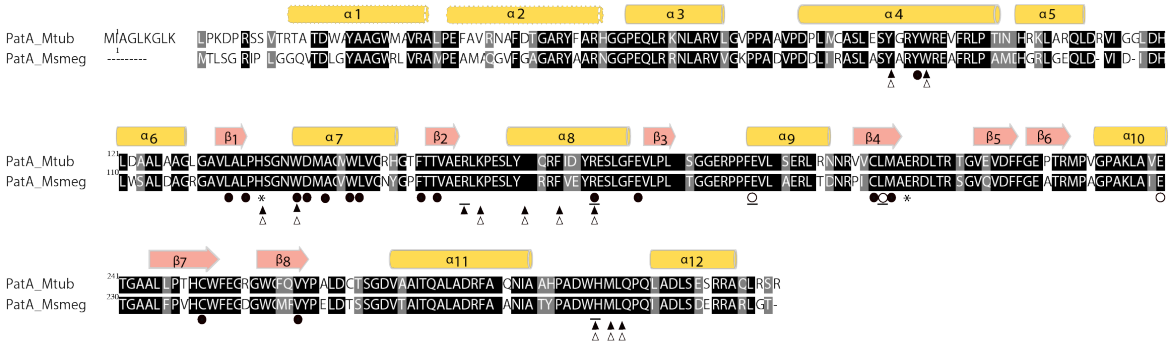



**Figure 37: Structural similarity of *M. tuberculosis* H37Rv and *M. smegmatis* mc$^2$155 PatA.** Catalytic amino acids are indicated as asterisks. Amino acids involved in palmitate and pantotheinate binding are indicated as black and white circles, respectively. Residues proposed to interact with $PIM_1$ and $PIM_2$ are shown as white and black triangles, respectively. Amino acids mutated are underlined.

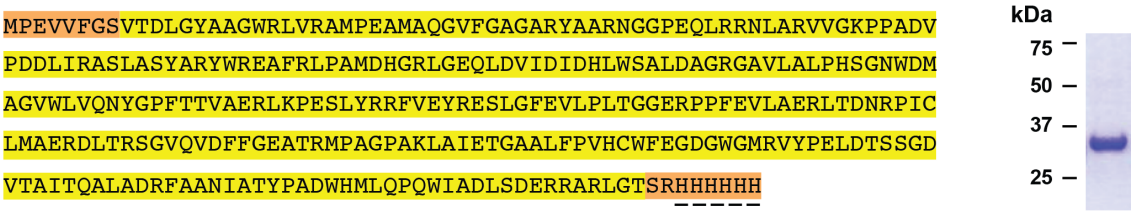



**Figure 38: Recombinant production of truncated version of PatA.** On the left side of the figure, amino acid sequence of PatA construct used for production in *M. smegmatis*. The recombinant PatA construct lacks the first 12 residues of the protein. On the right side of the figure, SDS-PAGE showing purified monomeric PatA of a molecular mass of 34 kDa.

The crystal structure of PatA from *M. smegmatis* was solved using single-wavelength anomalous dispersion (SAD) with a $K_2PtCl_4$ derivative at 2.06 Å resolution in C2 space group (PatA-$C_{16}$-1; PDB: 5F2T; **Table 6**). Two other crystal forms were obtained in $P2_1$ (PatA-$C_{16}$-2; PDB: 5F2Z) and $P4_22_12$ (PatA-$C_{16}$-3; PDB: 5F31) space groups, and the corresponding crystal structures solved at 2.90 and 2.43 Å resolution, respectively, by using molecular replacement methods. The high quality of the electron density maps allowed the trace of residues 41 to 302 (PatA-$C_{16}$-1), 48 to 295 (PatA-$C_{16}$-2) and 48 to 303 (PatA-$C_{16}$-3). PatA-$C_{16}$-1 crystal form will be used for the structure description since it displays the highest resolution (**Table 3**).

A close inspection of the three crystal structures revealed that the protein crystallized as a monomer, displaying a high degree of structural flexibility in the N- and C-terminal regions (coloured in green in **Figure 39** (root-mean-square deviation (r.m.s.d.) value of 1.6 and 8.4 Å for Cα atoms in 41–47 and 288–302 residue ranges for the N- and C-terminus, respectively).

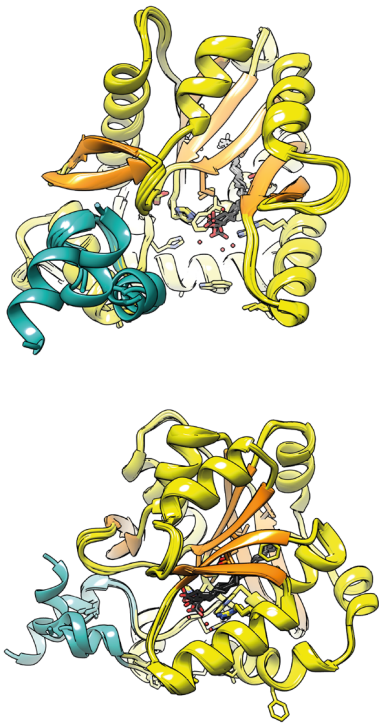


**Figure 39: Structural flexibility at the N- and C-terminus of PatA-C16-1, -2 and -3 complexes.** Cartoon representation showing the superimposed monomers obtained from crystal structures of PatA-$C_{16}$-1, PatA-$C_{16}$-2, and PatA-$C_{16}$-3. The N- and C-terminus are coloured blue.

The central core of PatA consists on a six-stranded β-sheet with topology $\beta_1$-$\beta_2$-$\beta_3$-$\beta_4$-$\beta_7$-$\beta_8$ ($\beta_8$ is antiparallel) surrounded by α-helices, with an overall size of 45X40X40 Å (**Figure 40 a-b**). A long and open groove that runs parallel to the protein surface contains the active site. This groove is flanked by $\alpha_4$, $\beta_2$, $\alpha_8$, $\beta_4$, $\alpha_9$, $\alpha_{10}$ and the connecting loops $\beta_1$-$\alpha_7$ (residues 124-130), $\beta_2$-$\alpha_8$ (residues 148-154), $\beta_3$-$\alpha_9$ (residues 173-181), $\beta_4$-$\beta_5$ (residues 198-208), $\beta_6$-$\alpha_{10}$ (residues 219-221) and $\alpha_{11}$-$\alpha_{12}$ (residues 282-291; **Figure 40 c-d**). Strikingly, the groove also displays a narrow and deep, mostly hydrophobic tunnel that runs perpendicular from its floor to the central core of the protein (**Figure 40 e-f**). The walls of the tunnel comprise the entire central β-sheet, two α-helices $\alpha_7$ and $\alpha_8$, and the connecting loops $\beta_1$-$\alpha_7$ (residues 124-130) and $\alpha_8$-$\beta_6$ (residues 167-170). The cavity extends entirely through the core and is closed at the bottom by $\beta_8$, the connecting loop $\alpha_5$-$\alpha_6$ and part of the $\alpha_6$ (**Figure 40 e-f**). Interestingly, the entrance of the hydrophobic pocket displays several charged residues compatible to interact with a buried carboxylate group of a fatty acid moiety, strongly suggesting the pocket might play a role in the donor substrate binding.

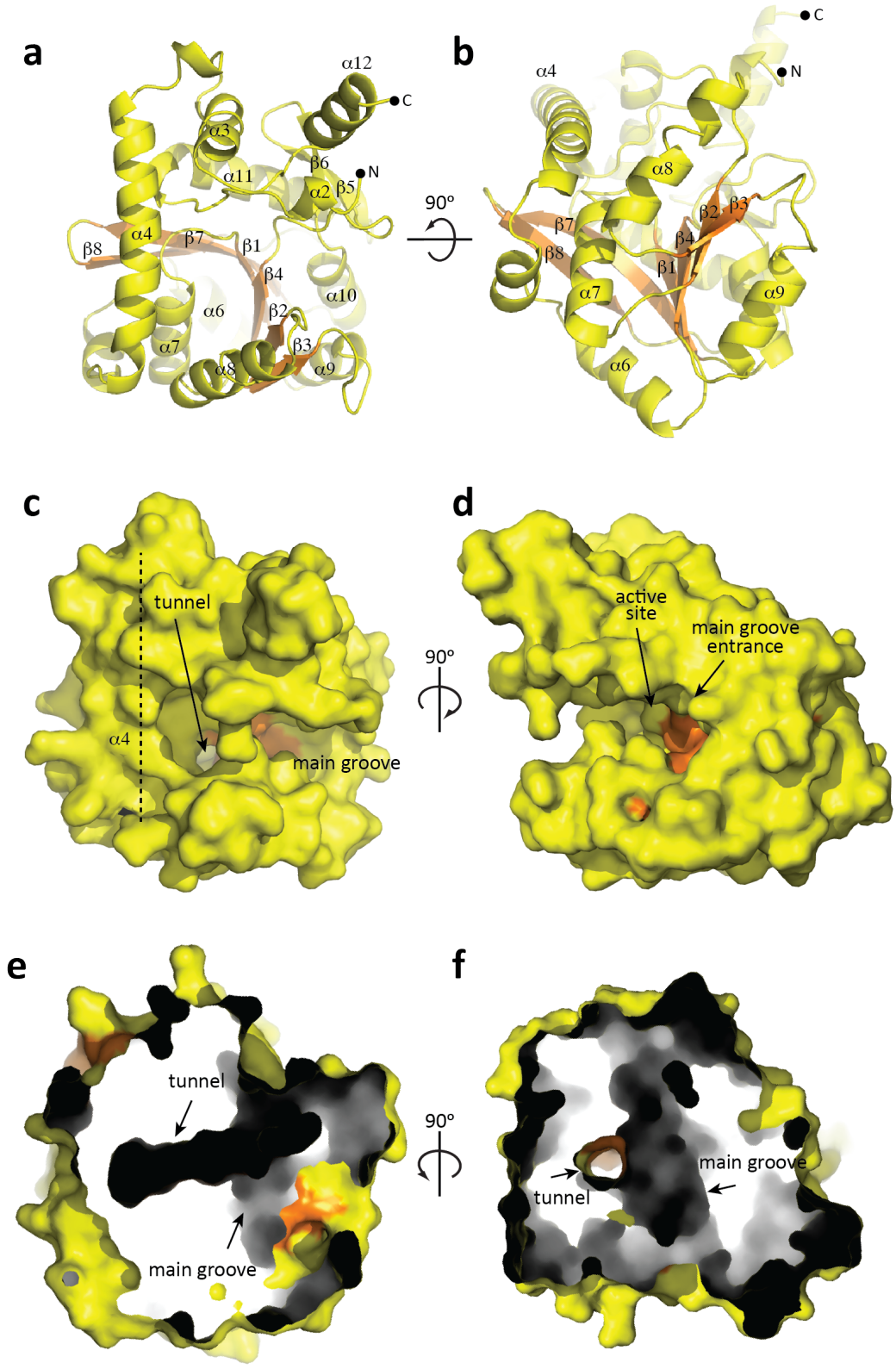



**Figure 40: Overall structure of PatA. a-b)** Cartoon representation showing the general fold and secondary structure organization of PatA. Secondary structure elements are labelled. The central core β-sheet is shown in orange. **c-d)** Surface representation of PatA showing the location of the main groove and the active site. The groove entrance is flanked by two important α-helices, $\alpha_{11}$-$\alpha_{12}$. The groove ends up into a cavity mainly flanked by $\alpha_4$. **e-f)** The main groove runs perpendicular to a hydrophobic tunnel, which is deeply buried into the core of PatA.

## 4.2.2. The Acyl-CoA Binding Site of PatA

Strikingly, one molecule of palmitic acid was unambiguously identified in the difference electron density maps of PatA-$C_{16}$-1, PatA-$C_{16}$-2 and PatA-$C_{16}$-3 crystal structures (**Figure 41 a-d**). It is thought that acyl molecule is associated to the enzyme due to the hydrolysis of palmitoyl-CoA during the isolation and purification of PatA from *M. smegmatis* $mc^2155$. Supporting this notion, the chemical structure of $Ac_1PIM_2$ was clearly established by using a combination of analytical techniques including mass spectrometry and two-dimensional NMR[205]. The major acyl form observed, corresponded to $PIM_2$ with the glycerol moiety being di-acylated by $C_{16}/C_{19}$ and the mannose residue transferred by PimA bearing a $C_{16}$. This structural profile was also identified in its hexamannosylated derivative $Ac_1PIM_2$[206].

The acyl chain is deeply buried into the hydrophobic pocket, and oriented with the carboxylate group facing the groove and the acyl tail extending into the globular core of the monomer (**Figure 41 a-d**).

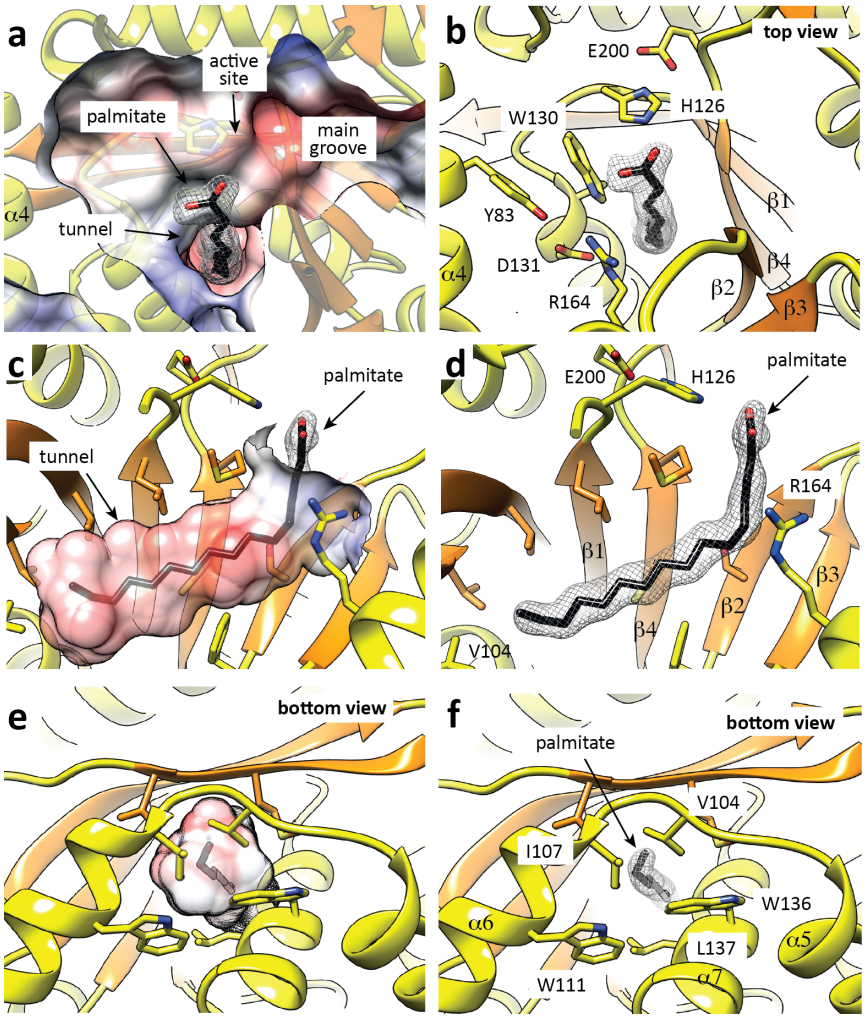



**Figure 41: The palmitate binding site of PatA. a-d)** Four views of PatA-$C_{16}$-1 crystal structure showing the palmitate chain deeply buried into the hydrophobic tunnel. **e-f)** Two views of PatA showing the cap which closes the bottom of the hydrophobic tunnel. The $2mF_o$-$DF_c$ electron density map countered at 1 σ for the palmitate ligand is shown.

Interestingly, the comparison of the three crystal structures revealed conformational flexibility in the carboxylate moiety of the palmitate (**Figure 42**), located in close proximity with the lateral chain of His126. The Arg164 guanidinium group engages the side chain of Tyr83 and Asp131. This tryptophan residue, together with Met198, makes an important Van der Waals interaction with the acyl chain. The acyl chain undergoes a kink at position C6 and terminates in a pocket mainly formed by hydrophobic residues, including Leu122 and Leu124 ($\beta_1$), Ala133, Trp136 and Leu137 ($\alpha_7$), Phe144 ($\alpha_7$-$\beta_2$ loop), Thr146 ($\beta_2$), Phe169 ($\alpha_8$-$\beta_3$ loop), Phe235 and Val237 ($\beta_7$), Met248 and Val250 ($\beta_8$), and two cysteine residues Cys196 ($\beta_4$) and Cys239 ($\beta_7$).

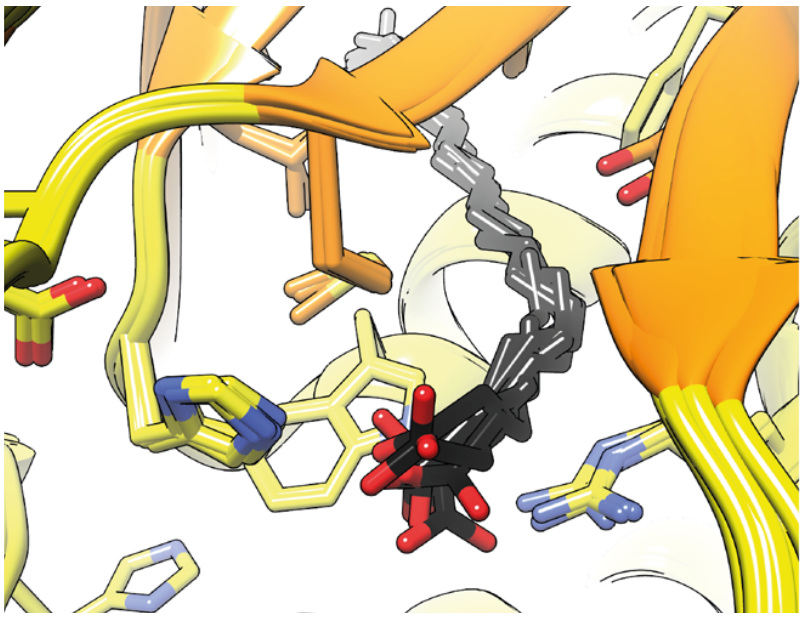


**Figure 42: Structural superposition of the palmitate molecules as visualized in the PatA-$C_{16}$-1, PatA-$C_{16}$-2, and PatA-$C_{16}$-3 structures.**

The connecting loop $\alpha_5$-$\alpha_6$ (residues 101-106) and part of $\alpha_6$ form a flexible and mostly hydrophobic cap that closes the bottom of the cavity, suggesting that PatA might be able to discriminate the length of the acyl chain groups (**Figure 41 e-f**). It is worth noting that $Ac_1PIM_2$ was the main product formed in the reaction when endogenous or crude mycobacterial phospholipids from *Mycobacterium phlei* were used as the lipid acceptors and a series of acyl-CoA derivatives of fatty acids were used as $C^{14}$-labelled donor substrates[204,288]. Palmitoyl-CoA ($C_{16}$) gave higher incorporation than myristyl-CoA ($C_{14}$). Interestingly, the oleyl-CoA ($C_{18:1}$) was a much better substrate than the saturated counterpart stearyl-CoA ($C_{18}$). Finally, the tuberculostearic acid ($C_{19:1}$) had a low specific activity and the small incorporation of label that was observed may be not significant[288].

### 4.2.2.1. How does PatA recognize CoA?

To this end, the crystal structure of PatA in complex with S-hexadecyl Coenzyme A (S-$C_{16}$CoA), a nonhydrolyzable analogue of palmitoyl-CoA, was solved at 3.28 Å resolution in $P2_1$ space group (PatA-S-$C_{16}$CoA; PDB: 5F34; **Table 4**). The acyl chain of S-$C_{16}$CoA is localized into the hydrophobic tunnel, and superimposes very well with the acyl chain moiety of palmitoyl observed in the PatA-$C_{16}$-1, PatA-$C_{16}$-2 and PatA-$C_{16}$-3 complexes (**Figure 43 a-d**). It is worth noting that the position of the carboxylate group of palmitate is different to that observed for the thioether in the PatA-S-$C_{16}$CoA complex. The 4-phosphopantetheinate moiety of S-$C_{16}$CoA is clearly defined in the electron density map, and located at the entrance of the main groove, in close contact with a highly conserved region flanked by the $\beta_2$-$\alpha_8$ (residues 149-153), $\beta_3$-$\alpha_9$ (residues 174-180) and $\beta_4$-$\beta_5$ (residues 199-207) loops, and two alpha helices, $\alpha_9$ (residues 181-190) and $\alpha_{10}$ (residues 221-230).

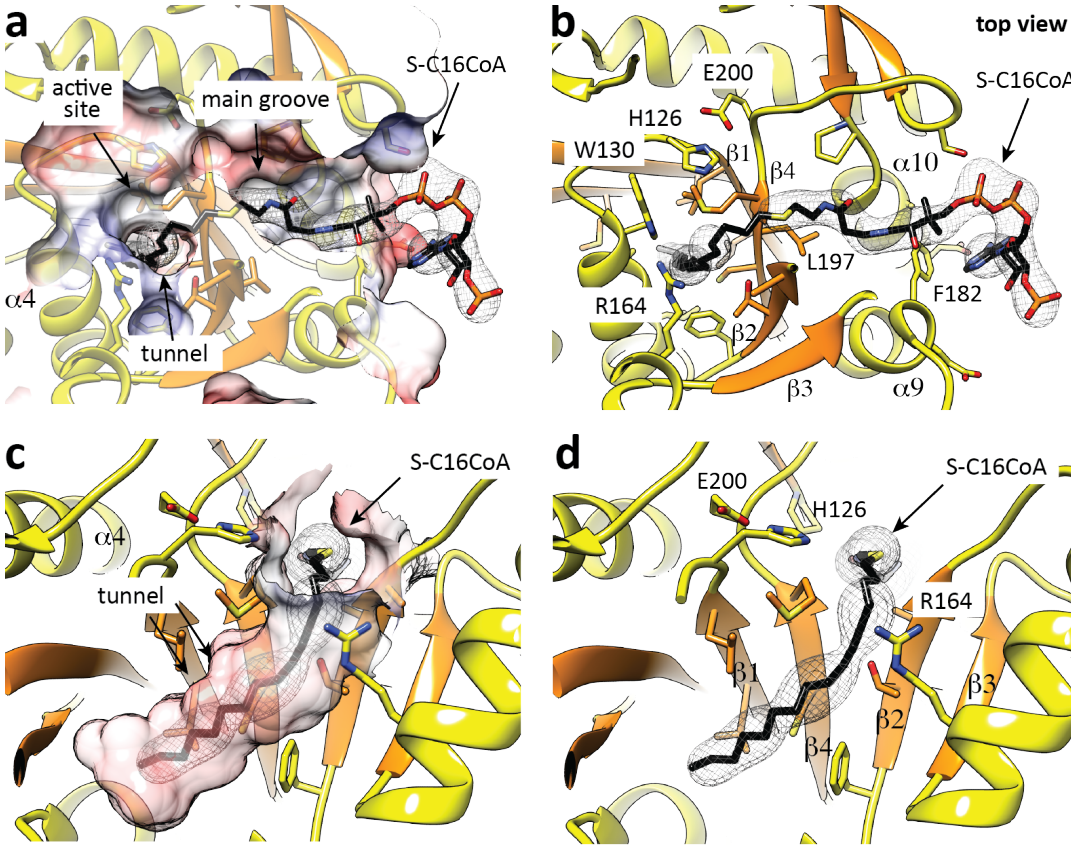



**Figure 43: The palmitoyl-CoA binding site of PatA. a-d)** Four views of PatA-S-$C_{16}$CoA crystal structure showing the binding mechanism of S-$C_{16}$CoA. The palmitoyl moiety is deeply buried into the hydrophobic tunnel, whereas the coenzyme A moiety extends outwards through the side of the main groove. The $2mF_o$-$DF_c$ electron density map countered at 1 σ for the palmitate ligand is shown.

The adenosine 3',5'-diphosphate (3',5'-ADP) moiety of the ligand (disordered in other monomers of the asymmetric unit) sticks out from the globular core and is exposed to the bulk solvent, as observed in other acyl-CoA modifying enzymes[289] (**Figure 44 a-c**). The approximate volume of the palmitoyl-CoA-binding pocket was *ca.* 2,801 Å[290].

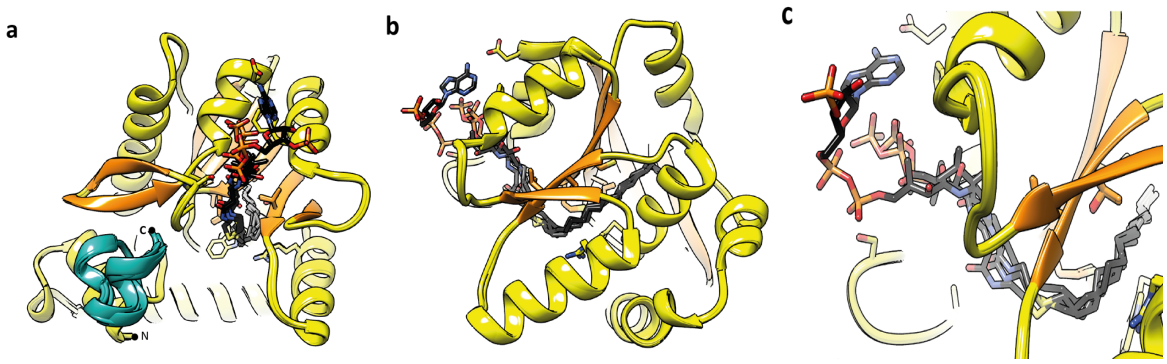



**Figure 44: Structural flexibility of S-$C_{16}$CoA as visualized in the PatA-S-$C_{16}$CoA complex**. **a-b)** Cartoon representation showing the four superimposed monomers obtained from the crystal structure PatA-S-$C_{16}$CoA. N- and C-terminus are indicated. **c)** Close view of the S-$C_{16}$CoA binding site in PatA. Both palmitoyl and pantetheine moieties binding to PatA are conserved in all four monomers in the asymmetric unit. The phosphorylated ADP of S-$C_{16}$CoA was visible in one molecule of the asymmetric unit, exposed to the bulk solvent.

To further validate the model, we designed site-directed mutations predicted to impair the palmitoyl-CoA interaction with PatA. Thus, the double substitution Phe182Trp/Leu197Trp would block the groove region of PatA, hindering the formation of the complex (**Figure 43 b**). As depicted in **Figure 45**, the PatA Phe182Trp/Leu197Trp variant could not (i) transfer a palmitoyl moiety to $PIM_2$, or (ii) hydrolyse palmitoyl-CoA, supporting the proposed model.

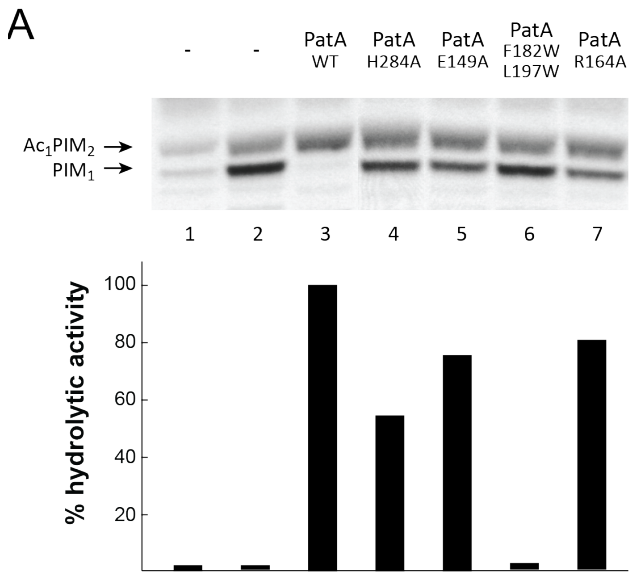



**Figure 45: Enzymatic activity of selected PatA variants.** In the upper panel, the acyltransferase activity of purified PatA and variants involved in substrate binding are shown. Reaction mixtures contained crude membranes from *M. smegmatis* $mc^2155$ and GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with PimA (lanes 2-7) and palmitoyl-CoA (lanes 3-7) and purified PatA (lane 3), PatA-His284Ala (lane 4), PatA-Asp149Ala (lane 5), double mutant PatA-Phe182Trp/Leu197Trp (lane 6) and PatA-Arg164Ala (lane 7). The lipids were extracted from reaction mixtures and analysed by TLC and autoradiography. In the lower panel, the hydrolytic activity against palmitoyl-CoA is shown as determined by spectrophotometric analysis. All enzymatic activities measurements were determined in duplicates.

The residues that contact the bound palmitic acid are highly conserved in the PatA mycobacterial homologues (**Figure 46**).

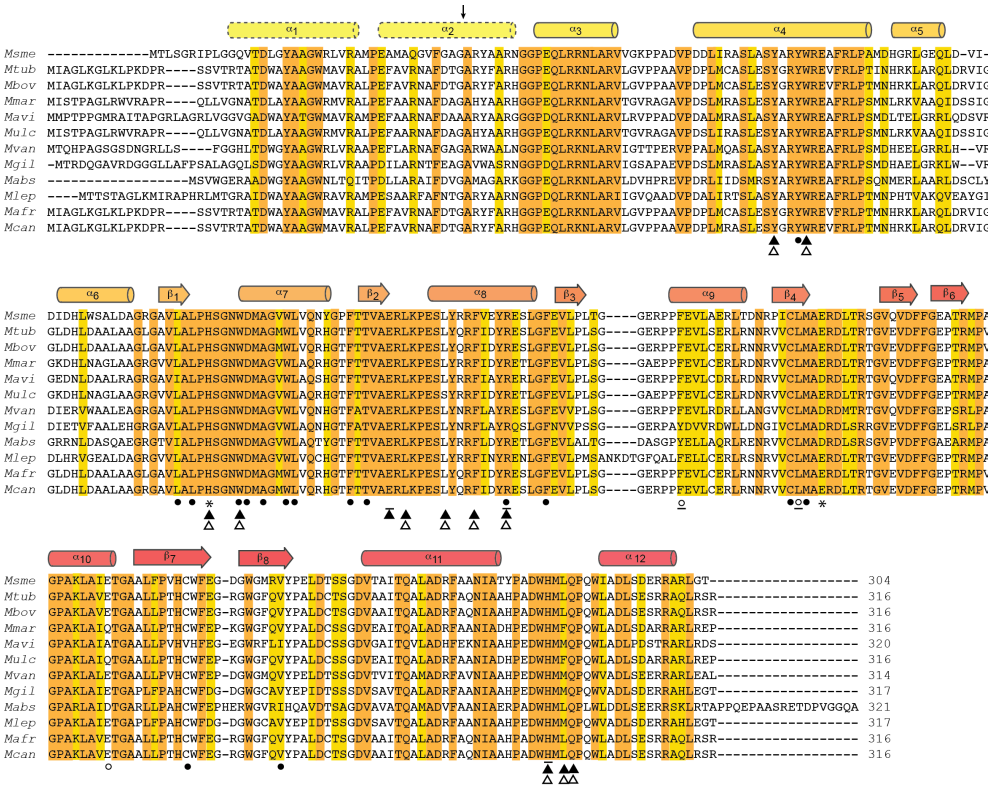


**Figure 46: Multiple sequence alignment of PatA with mycobacterial orthologs.** Orthologs shown are from *Mycobacterium smegmatis* (*Msme*, A0QWG5), *M. tuberculosis* (*Mtub*, O06203), *M. bovis* (*Mbov*, Q7TY87), *M. marinum* (*Mmar*, B2HN43), *M. avium* (*Mavi*, AIV25898.1), *M. ulcerans* (*Mulc*, ABL05460.1), *M. vanbaalenii* (*Mvan*, ABM13371.1), *M. gilvum* (*Mgil*, ADT99794.1), *M. abscessus* (*Mabs*, AIV11722.1), *M. leprae* (*Mlep*, Q7AQJ0), *M. africanum* (*Mafr*, F8M2W4), and *M. canettii* (*Mcan*, L0PY90). Strictly conserved positions are shown in orange background. The secondary structural elements corresponding to the three distinct 3D structures of PatA are shown above the alignment. Catalytic amino acids are indicates as asterisks. Amino acids involved in palmitate and pantotheinate binding are indicated as black and white circles, respectively. Residues proposed to interact with $PIM_1$ and $PIM_2$ based on the dockings are shown as white and black triangles, respectively. Amino acids mutated are underlined.

## 4.2.3. The Phosphatidylinnositol Mannosides Binding Site

To describe the architecture of the PatA acceptor binding site, it was first tried to crystallize PatA in the presence of $Ac_1PIM_1/Ac_2PIM_1$, $Ac_1PIM_2/Ac_2PIM_2$ or their deacylated analogues. However, these attempts were unsuccessful. Therefore, to describe the architecture of the acceptor binding site, PatA was co-crystallized with D-mannopyranose (Man*p*).

The crystal structure of PatA from *M. smegmatis* co-crystallized with Man*p* was solved by using molecular replacement methods (**Figure 47**). PatA crystallized in the $P2_1$ space group with four molecules (ca 260 residues each) in the asymmetric unit, and diffracted to a maximum resolution of 2.42 Å (**Figure 49** and **Figure 50**). Careful inspection of the electron density maps revealed the presence of the product 6-*O*-palmitoyl-α-D-mannopyranoside ($C_{16}$-Man) in three of the four molecules (chains B, C and D) of the asymmetric unit (PatA-$C_{16}$-Man). The fourth molecule (chain A) showed (i) one molecule of free palmitate located in the hydrophobic pocket and (ii) one molecule of free β-D-mannopyranose, located in the acceptor binding cavity (**Figure 49 a-b** and **Figure 50**). Because palmitoyl-CoA was not present during the crystal screening experiments, we believe that the palmitate molecule is associated with the enzyme during the isolation and purification of PatA from *M. smegmatis* $mc^2155$, as previously described[21]. This suggests that PatA is mediating the formation of $C_{16}$-Man in three of the four molecules of PatA in the asymmetric unit.

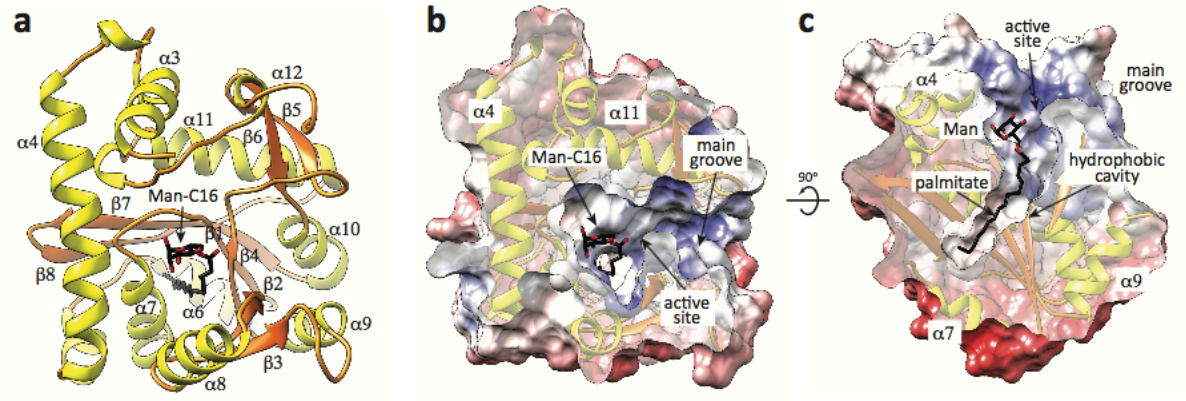



**Figure 47: Overall structure of the PatA-$C_{16}$-Man complex. a)** Cartoon representation showing the general fold and secondary structure organization of PatA. β-strands and α-helices are shown in orange and yellow respectively. The $C_{16}$-Man product is shown in black and red. **b-c)** Two views of the electrostatic surface representation of PatA showing the location of the $C_{16}$-Man product. The Man moiety of $C_{16}$-Man localizes into the main groove, in close proximity to $\alpha_4$. In contrast, the palmitate moiety $C_{16}$-Man localizes in a hydrophobic cavity, deeply buried into the core of PatA.

PatA in complex with Man*p* crystallized as a monomer, displaying a conserved α/β architecture[21]. The central core comprises a six-stranded β-sheet ($\beta_1$-$\beta_2$-$\beta_3$-$\beta_4$-$\beta_7$-$\beta_8$ of with $\beta_8$ is antiparallel) bound by several α-helices (**Figure 47 a**). A surface-exposed, open groove with an overall size of 25x10x10 Å contains the active site. This long groove is flanked by several secondary structure elements including $\alpha_4$, $\beta_2$, $\alpha_8$, $\beta_4$, $\alpha_9$, $\alpha_{10}$ and the connecting loops $\beta_1$-$\alpha_7$ (residues 124-130), $\beta_2$-$\alpha_8$ (residues 148-154), $\beta_3$-$\alpha_9$ (residues 173-181), $\beta_4$-$\beta_5$ (residues 198-208), $\beta_6$-$\alpha_{10}$ (residues 219-221) and $\alpha_{11}$-$\alpha_{12}$ (residues 282-291)[21]. A deep hydrophobic cavity extends through the core of PatA, with the entrance located in the floor of the main groove and closed at the bottom by $\beta_8$, the connecting loop $\alpha_5$-$\alpha_6$ and part of the $\alpha_6$ (**Figure 47** and **Figure 48**).

Strikingly, the careful inspection of the electron density maps revealed the presence of the product $C_{16}$-Man in three of the four molecules of the asymmetric unit, and in the fourth molecule of the asymmetric unit the reaction between palmitoyl and D-mannopyranose did not take place, showing one molecule of free palmitate located in the hydrophobic pocket and one molecule of free β-D-mannopyranose, located in the acceptor binding cavity (**Figure 49 a-b** and **Figure 50**).

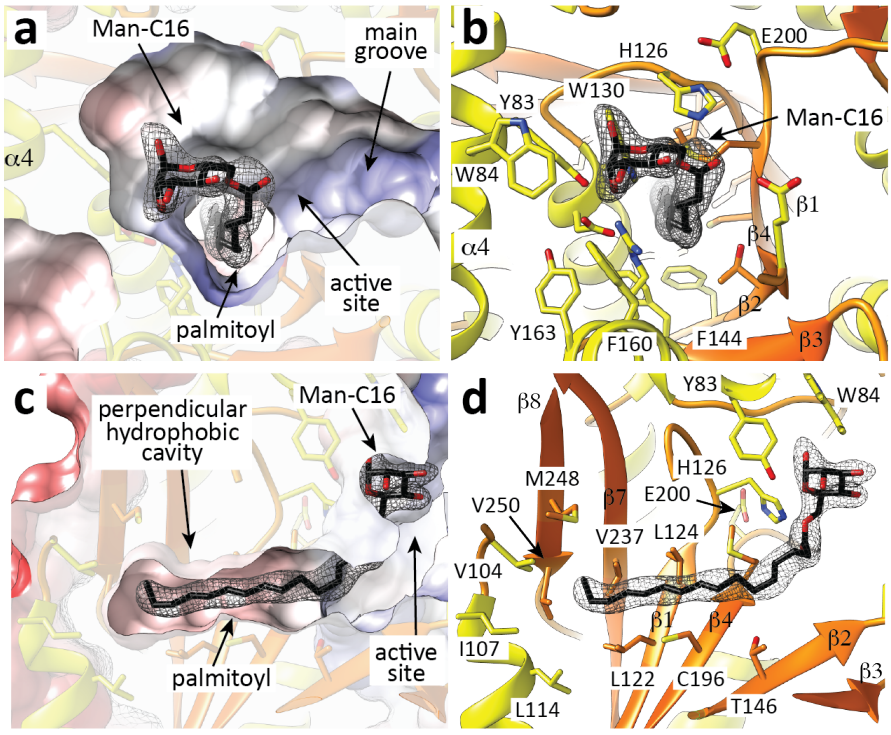



**Figure 48: The acceptor binding site of PatA. a-d)** Four views of the PatA-$C_{16}$-Man crystal structure showing the palmitoyl chain deeply buried into the hydrophobic cavity and the mannose unit located in the acceptor binding site, at the end of the main groove. The final electron density maps ($2mF_o$-$DF_c$ contoured at 1σ) of the 6-*O*-palmitoyl-α-D-mannopyranoside is coloured in grey.

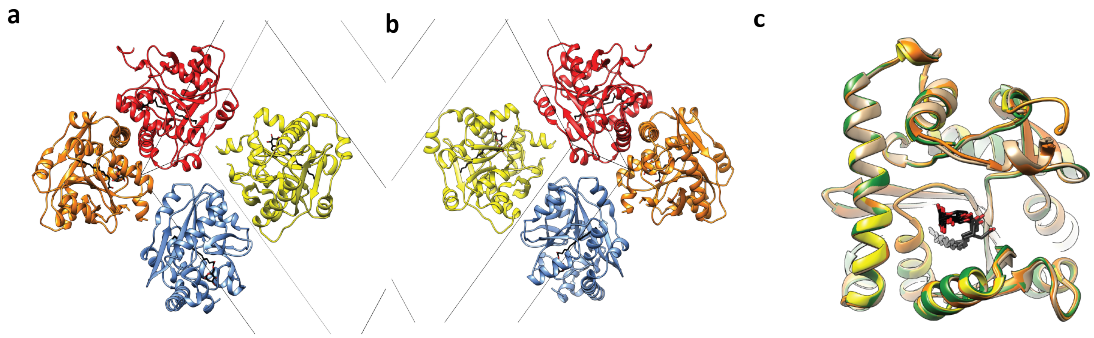



**Figure 49: Four crystallographically independent PatA-$C_{16}$-Man complex protomers reveal two different states. a-b)** The asymmetric unit of the $P2_1$ crystal form contains four PatA molecules (shown here in orange, red, yellow and blue). Three independent molecules coloured in red, yellow and blue were observed in complex with the reaction product $C_{16}$-Man. The fourth molecule in the asymmetric unit (coloured in orange) was observed in complex with β-D-Man and palmitoyl. **c)** Structural superposition of the four PatA protomers observed in the asymmetric unit.

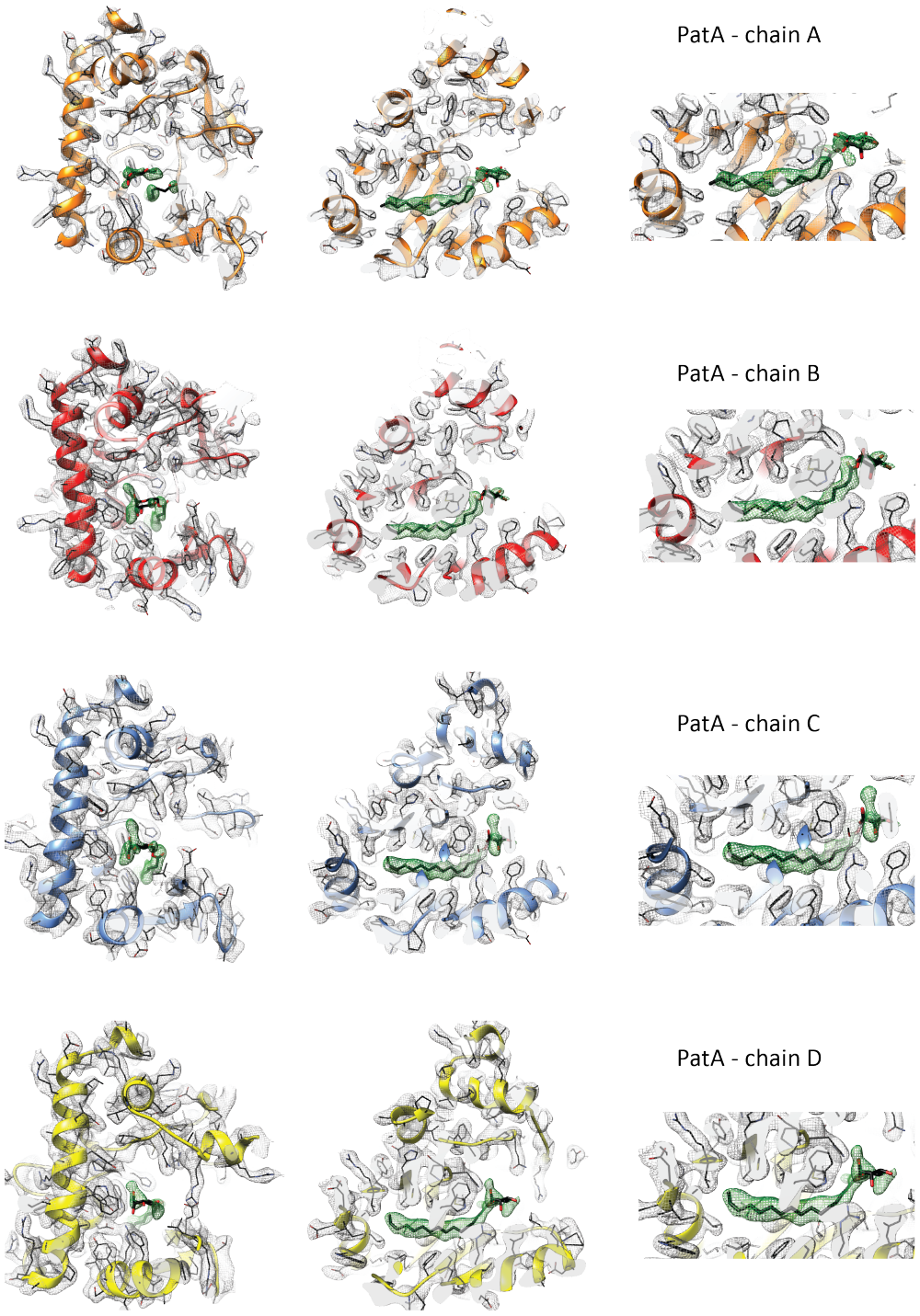



**Figure 50: Electron density maps of the refined PatA-$C_{16}$-Man complex.** Two views of the final electron density maps ($2mF_o$-$DF_c$ contoured at 1σ) corresponding to the PatA-$C_{16}$-Man structure (reference chain).

It is believed that the palmitate molecule is associated to the enzyme due to the hydrolysis of palmitoyl-CoA during the isolation and purification of PatA from *M. smegmatis* $mc^2155$, as previously described[21]. Supporting this notion, the major acyl forms observed in $Ac_1PIM_2$ corresponded to $PIM_2$ with the glycerol moiety being di-acylated by $C_{16}/C_{19}$ and the mannose residue transferred by PimA bearing a $C_{16}$[205]. The palmitate group is deeply buried into the hydrophobic cavity, and oriented with the carboxylate group facing the groove and the acyl tail extending into the globular core of the monomer, as observed in the PatA-$C_{16}$ and PatA-S-hexadecyl Coenzyme A (S-$C_{16}$CoA) complexes (**Figure 47** and **Figure 48**)[21]. Man*p* ring is located within a cavity located at the end of the main groove and comprising helices $\alpha_4$ and $\alpha_8$, and the connecting loops $\beta_1$-$\alpha_7$ (residues 83-90), $\beta_2$-$\alpha_8$ (residues 148-154) and $\alpha_{11}$-$\alpha_{12}$ (residues 282-291). Man*p* ring is stabilized by Van der Waals interactions with the side chains of four aromatic residues Tyr83, Trp84, Trp130, and Phe160. The O6 atom of the Man*p* ring makes an important hydrogen bond with NE2 atom of His126, whereas O2 atom makes a weak hydrogen bond with $NH_2$ atom of Arg164 (**Figure 48**).

It is worth noting that the Man*p* residue observed in the PatA-$C_{16}$-Man complex superimposed well with the corresponding Man*p* residue of the deacylated form of $PIM_2$ obtained by docking calculations[21]. The structural superposition of the PatA-$C_{16}$-Man and S-$C_{16}$CoA complexes gives a very nice view of the acyl donor and acceptor substrates binding sites (**Figure 51**). The 4-phosphopantetheinate moiety of S-$C_{16}$CoA localizes in the opposite side, at the entrance of the main groove, flanked by the $\beta_2$-$\alpha_8$ (residues 149-153), $\beta_3$-$\alpha_9$ (residues 174-180) and $\beta_4$-$\beta_5$ (residues 199-207) loops, and two alpha helices, $\alpha_9$ (residues 181-190) and $\alpha_{10}$ (residues 221-230). The adenosine 3',5'-diphosphate (3',5'-ADP) moiety of S-$C_{16}$CoA (disordered in three of the four protomers of the asymmetric unit) protrudes from the groove and is exposed to the bulk solvent, as observed in other acyl-CoA modifying enzymes (**Figure 51**)[18,289].

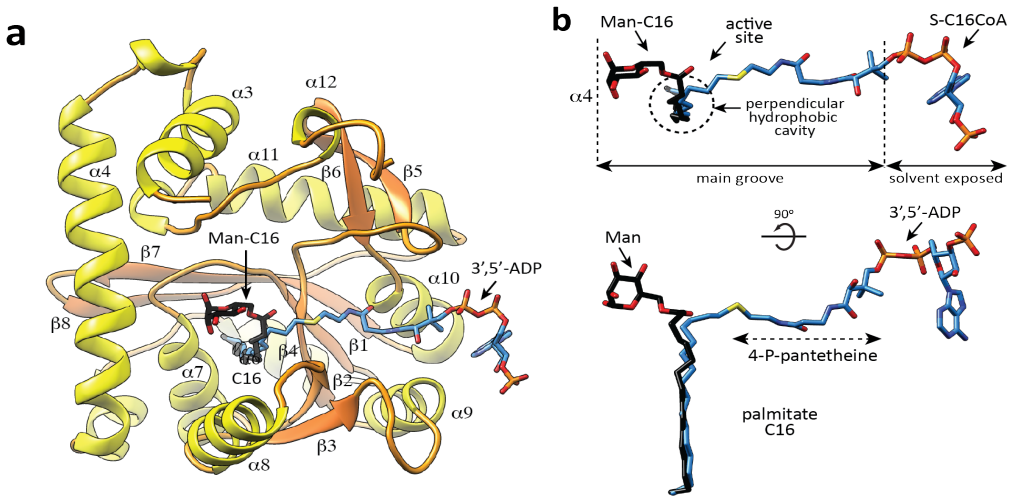



**Figure 51: The donor and acceptor binding mechanism. a)** Cartoon representation showing the location of $C_{16}$-Man and S-$C_{16}$CoA into PatA. β-strands and α-helices are shown in orange and yellow respectively. **b-c)** Two views showing the structural superposition of $C_{16}$-Man and S-$C_{16}$CoA.

### 4.2.3.1. Inhibition Mecanism of PatA

The crystal structure of PatA-$C_{16}$-Man certainly provide one possible mechanism of enzyme inhibition. The PatA-$C_{16}$-Man complex revealed the presence of the product $C_{16}$-Man in three of the four molecules (chains B, C and D) of the asymmetric unit, as a consequence of the chemical reaction between the palmitate and α-D-mannopyranose. Interestingly, careful inspection of the electron density maps revealed that in the fourth molecule (chain A) of the assymetric unit the reaction between palmitate and D-mannopyranose did not take place, showing (i) one molecule of free palmitate located in the hydrophobic pocket and (ii) one molecule of free β-D-mannopyranose, located in the acceptor binding cavity (**Figure 52 a**).

The structural comparison of the four molecules of the asymmetric unit revealed subtle but important conformational changes in the active site (**Figure 52 b**). Specifically, the carboxylate moiety moves away from the catalytic centre as observed in the PatA-S-$C_{16}$CoA complex, in closed contact with the main chain of residues Val147 and Glu149[21]. Importantly, the Man O1 in β configuration is now stabilized by new interactions with the side chains of Tyr83, Trp130 and Arg164 residues (**Figure 52 b**). The Man O4 makes hydrogen bond with a water molecule, which in turn strongly interact with the carboxylate moiety of Glu149. Similarly, the Man O6 makes a strong hydrogen bond with the side chain of His126 (**Figure 52 b**).

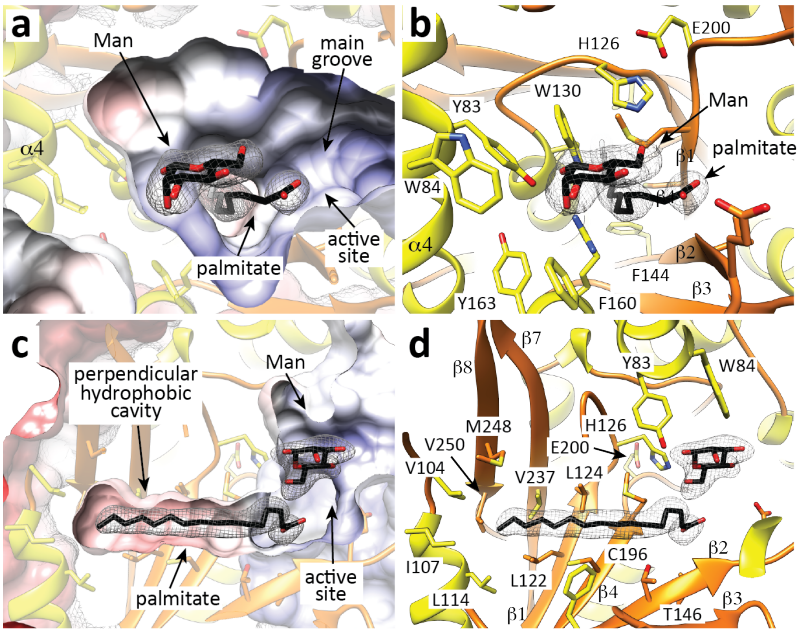



**Figure 52: The crystal structure of PatA-$C_{16}$-Man suggest a possible mechanism of enzyme inhibition. a-d)** Four views of PatA-$C_{16}$-Man crystal structure showing the palmitoyl chain deeply buried into the hydrophobic cavity and the β-Man residue located in the acceptor binding site, at the end of the main groove.

$Ac_1PIM_2$ appears to be a metabolic end product that accumulates at high steady-state levels in the cells as well as a precursor for more polar forms of PIMs, lipomannan and LAM[12]. Interestingly, the four enzymes involved in the biosynthesis of $Ac_1PIM_2$, the phosphatidyl-*myo*-inositol synthase PgsA1,

PimA, PimB and PatA, were found to be essential for the growth of *M. smegmatis* and/or *M. tuberculosis*[12,215,228,230]. Thus, the structural data presented here offers exciting possibilities for inhibitor design and the discovery of chemotherapeutic agents against this major human pathogen.

### 4.2.3.2. Molecular Docking

Docking calculations placed the polar head of $PIM_2$, the better substrate of PatA making important interactions within a region located at the end of the main groove and comprising helices $\alpha_4$ and $\alpha_{8'}$ and the connecting loops $\beta_1$-$\alpha_7$ (residues 83-90), $\beta_2$-$\alpha_8$ (residues 148-154) and $\alpha_{11}$-$\alpha_{12}$[24,233] (residues 282-291; **Figure 53**). As a consequence, the O6 atom of the Manp ring linked to position 2 of *myo*-inositol in $PIM_2$ is predicted to be positioned favourably for activation by His126 and to receive the palmitate group from palmitoyl-CoA (**Figure 53**; distance His126 NE2 atom to Man*p* O6 atom is 2.8 Å).

The model also predicts an important role of residues Glu149, Arg164 and His284 to bind $PIM_2$ in the active site. In the model, Glu149 OE1 atom is found at 2.9 Å of the Man*p* residue O4 atom and Arg164 $NH_2$ atom is placed at 3.1 Å of the O2 atom of the Man*p* ring, whereas His284 ND1 atom is at only 2.5 Å of *myo*-inositol O3 atom (**Figure 54**). The *myo*-inositol moiety and the Man*p* ring linked to position 6 of *myo*-inositol in $PIM_2$ also interact with Tyr80 and Gln287. The approximate volume of the $PIM_2$-binding pocket was ca. 1,741 Å (**Figure 54**)[290].

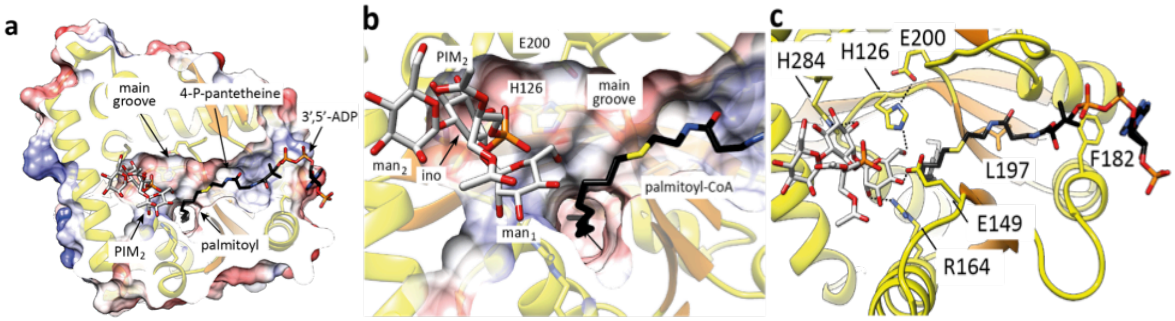



**Figure 53: Docking calculations of the phosphatidylinositol mannosides binding site. a-b)** Two views showing the docking calculations in which $PIM_2$ attaches to the end side of the main groove, close to the hydrophobic tunnel. **c)** Close view of the binding site, showing the location of predicted residues involved in $PIM_2$ interaction that have been mutated (His126, Glu149, Arg164, His284) along with mutated residues involved in CoA binding (Leu197 and Phe182).

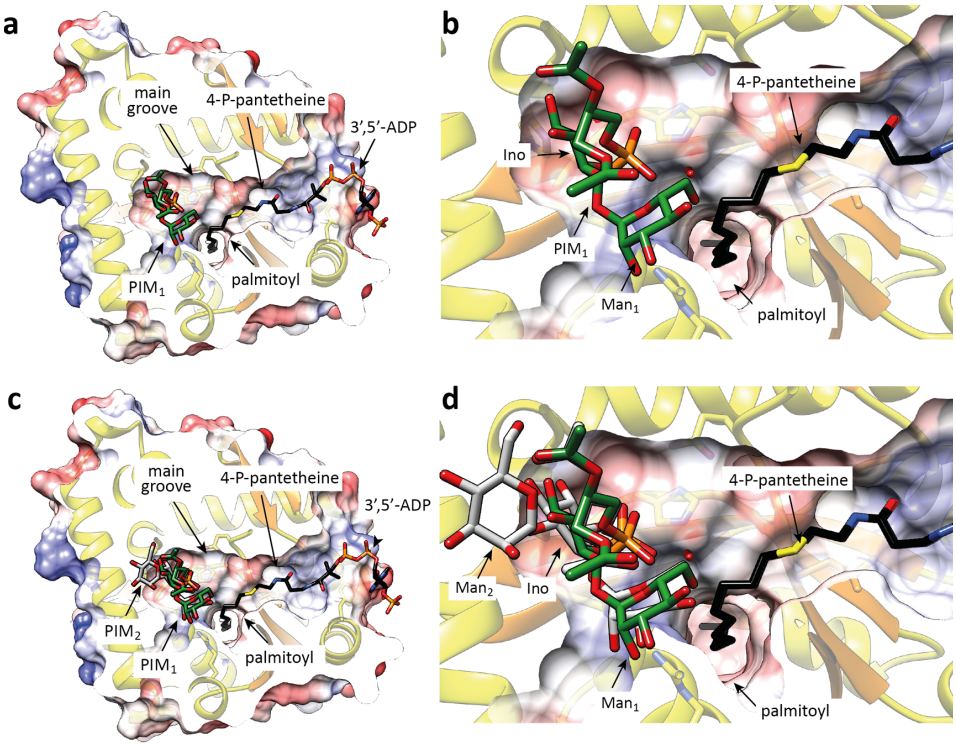



**Figure 54: Structural comparison of the $PIM_1$ and $PIM_2$ binding sites. a-b)** Two views showing the docking calculations in which $PIM_1$ (green) attaches to the end side of the main groove, and close to the hydrophobic tunnel. **c-d)** Two views showing the comparison of the docking calculations for $PIM_1$ and $PIM_2$ (grey).

### 4.2.3.3. Enzymatic Assays

To further validate the proposed model, we designed three single-point substitution, Glu149Ala, Arg164Ala and His284Ala, predicted to impact the $PIM_2$ interaction with PatA (**Figure 53 c**). As depicted in **Figure 55**, the transferase activity of all PatA variants was severely compromised, nevertheless preserving the capability to hydrolyse palmitoyl-CoA.

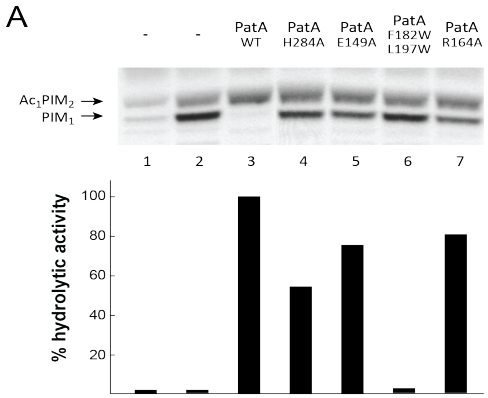



**Figure 55: Duplicate enzymatic activity measurements of selected PatA variants.** In the upper panel, the acyltransferase activity of purified PatA and variants involved in catalysis are shown. Reaction mixtures contained crude membranes from *M. smegmatis* $mc^2155$ and GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with PimA (lane 2 to 6) and palmitoyl-CoA (lanes 3 to 6) and purified PatA (lane 3), PatA-Asp131Ala (lane 4), PatA-Glu200Ala (lane 5) and PatA-His126Ala (lane 6). In the lower panel, the hydrolytic activity against palmitoyl-CoA is shown.

## 4.2.4. The Catalytic Mechanism of PatA

Analysis of the cleft running over the surface of PatA revealed a catalytic site reminiscent to that observed in the serine protease family of enzymes. In serine proteases, the cleavage of the peptide bond is mediated by nucleophilic attack of the serine hydroxyl group on the scissile carbonyl bond. The active site comprises a catalytic triad consisting of the Oγ atom of the serine, the imidazole ring of a histidine, and the carboxylate group of an aspartic/glutamic acid, involved in a charge relay system that increases the nucleophilicity of the serine hydroxyl and modulates the pKa of the central histidine as a general base or acid during the catalytic cycle[60].

In all crystal structures of palmitoyl-PatA and PatA-S-$C_{16}$CoA complexes, the carboxylate OE2 oxygen atom of Glu200 was found at 2.8 Å of the ND1 nitrogen atom of the aromatic imidazole ring of the invariant His126 (**Figure 56**). It is worth noting that when the individual palmitate molecules observed in the three crystal structures of PatA-$C_{16}$ were superimposed, they showed important structural flexibility at the carboxylate region (**Figure 42**), suggesting that the carboxylate group is not in a catalytically competent position in the crystal structures. Moreover, the palmitate is not a substrate neither an inhibitor of the reaction catalysed by PatA (**Figure 57**). Our binary complexes correspond most likely to the hydrolysis reaction product of palmitoyl-CoA, thus one of the oxygen atoms found in the palmitate that interacts with His126 most likely come from a water molecule activated by His126.

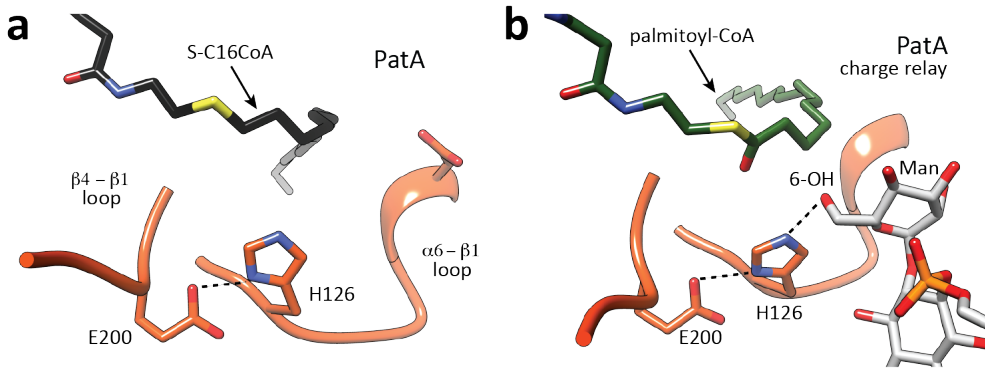



**Figure 56: Catalytic mechanism of PatA. a)** Active site configuration of PatA, as visualized in the crystal structure in complex with S-$C_{16}$CoA (PatA-S-$C_{16}$CoA). **b)** Active site configuration of PatA as visualized in the docking of palmitoyl-CoA and $PIM_2$ (the atomic coordinates of palmitoyl-CoA were derived from S-$C_{16}$CoA following in silico addition of an oxygen atom at position $C_{16}$ of S-$C_{16}$CoA). The charged relay system formed between Glu200, His126 and the 6-OH of the mannose linked to the 2-position of inositol (ino) is highlighted.

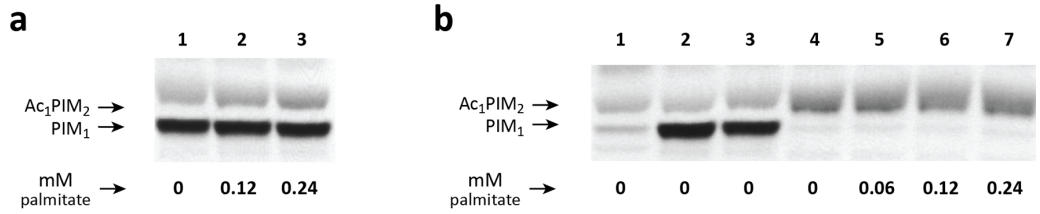



**Figure 57: Palmitate is neither a substrate or inhibitor of PatA. a)** Reaction mixtures contained 250 µg of crude membranes from *M. smegmatis* mc$^2$155 and 0.1 µCi GDP-[$^{14}$C]-mannose as a tracer supplemented with purified PimA and PatA (lane 1 to 3), and 120 µM (lane 2) and 240 µM (lane 3) palmitate. The lipids were extracted from reaction mixtures and analysed by TLC and autoradiography as described in Methods section. **b)** Reaction mixtures contained 250 µg crude membranes from *M. smegmatis* mc$^2$155 and 0.1 µCi GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with purified PimA (lanes 2 to 7), purified PatA (lanes 3 to 7), 120 µM palmitoyl-CoA (lanes 4 to 7), and palmitate at 60 µM, 120 µM and 240 µM, respectively.

The S-$C_{16}$CoA coordinates in PatA-S-$C_{16}$CoA crystal structure were used to generate palmitoyl-CoA atomic coordinates by substitution of $C_{16}$ atom with a carbonyl group followed by energy minimization. Docking calculations placed the Man*p* moiety attached to the 2-position of *myo*-inositol in $PIM_2$ with its O6 atom favourably positioned to receive the palmitate group from palmitoyl-CoA (**Figure 56 b**).

In the proposed model His126 acts as the general base to abstract a proton from the hydroxyl group at position 6 of the Man*p* ring linked to the 2-position of inositol in $PIM_1$ or $PIM_2$, to facilitate the nucleophilic attack on the thioester of palmitoyl-CoA (**Figure 58**). The Glu200 gets involved in a charge relay system with His126 and the HO6 atom from the Man*p* moiety, contributing to the appropriate structural arrangement of the imidazole ring of the histidine residue and modulating its pKa to act as a base in the first step and as an acid in the second step, providing protonic assistance to the departing CoA leaving group. It is worth noting that the His126-Glu200 hydrogen bond was found in a *syn* orientation relative to the carboxylate[60]. The result of the nucleophilic attack is a covalent bond between the mannose ring of $PIM_1$ or $PIM_2$ and palmitate.

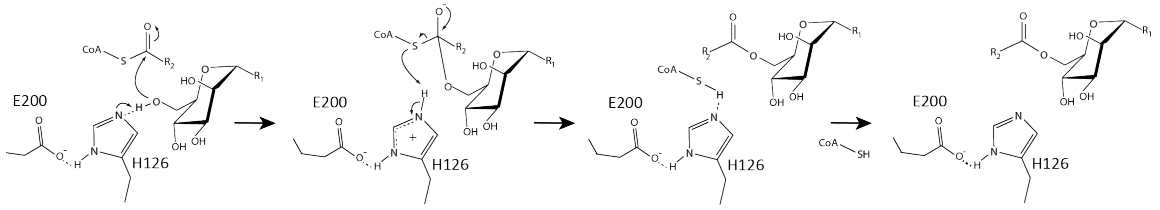



**Figure 58: Proposed catalytic mechanism for PatA.**

### 4.2.4.1. Molecular Dynamics

To gain insight into the chemical reaction mechanism catalysed by PatA, a PatA-$PIM_2$-$C_{16}$CoA ternary complex was modelled, taking into account the experimental structural data available (**Figure 59**)[21].

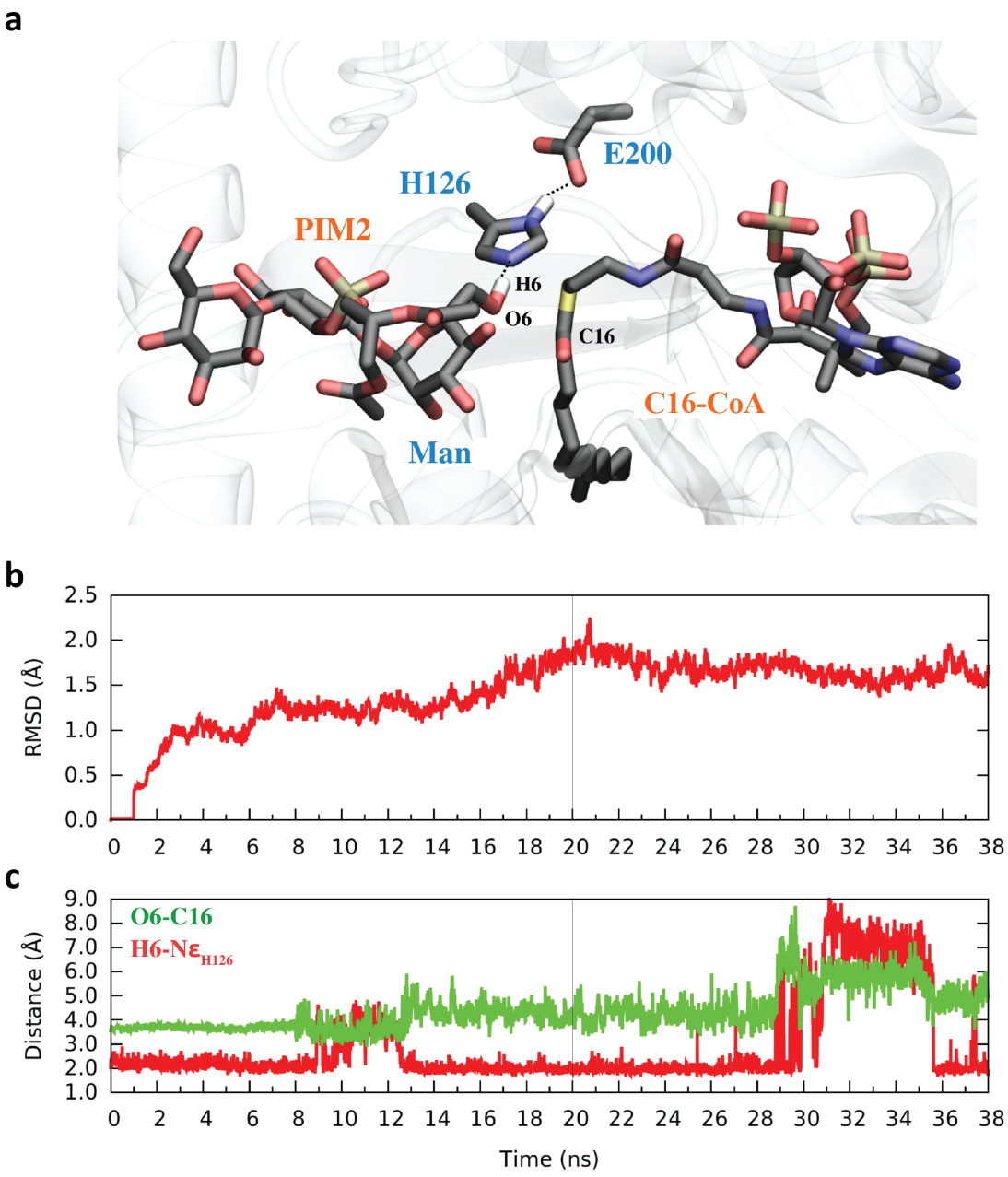



**Figure 59: PatA-$PIM_2$-$C_{16}$CoA ternary complex modelled with molecular dynamics. a)** Ternary complex of PatA with $PIM_2$ and $C_{16}$-CoA. **b)** Time evolution of the RMSD during the classical molecular dynamics simulation of the PatA complex with $PIM_2$ and $C_{16}$-CoA. The system was considered equilibrated after 20 ns. **c)** Evolution of the distance corresponding to the nucleophilic attack of Man1 of $PIM_2$ on the thioester carbonyl of $C_{16}$-CoA (green line) and hydrogen bond distance between O6H and His126 (red line).

Molecular dynamics (MD) simulations (38 ns) were performed at room temperature to relax the system and sample all the accessible enzyme-substrate conformations (**Figure 59**). A snapshot from the equilibrated complex was used as a starting point for the modelling of the enzymatic reaction using QM/MM metadynamics[291]. Three collective variables (CVs; **Figure 60**) were selected to account for the main changes taking place during the reaction. The first collective variable (CV1) was defined as the difference of distances between O6-$C_{16}$ and $C_{16}$-S. This variable describes the nucleophilic attack of the O atom of the $CH_2OH$ group of one of the mannosyl residues of $PIM_2$ (Man1 in **Figure 59 a**) on the carbonyl of the thioester bond of $C_{16}$-CoA. We named this variable as "nucleophile attack". The second collective variable (CV2) was selected as the difference between the O6-H6 and $N_\varepsilon$-H distances. This variable accounts for the proton transfer between the $CH_2OH$ group of Man1 and His126, thus we named it as "$CH_2OH$ proton transfer". The third collective variable (CV3), named as "His proton transfer", was taken as the distance difference between $N_\varepsilon$-H and H-S, thus it accounts for the transfer of the His126 proton to the S atom of $C_{16}$CoA. Similar ideas were previously used with success to model chemical reactions in enzymes[292]. It is important to note that the CVs used do not self-select any specific reaction pathway: it does not impose whether proton transfer takes place in concert with nucleophilic attack or is a later event along the reaction coordinate. Likewise, the CVs used do not impose whether the reaction is fully concerted or an intermediate species forms along the reaction coordinate.

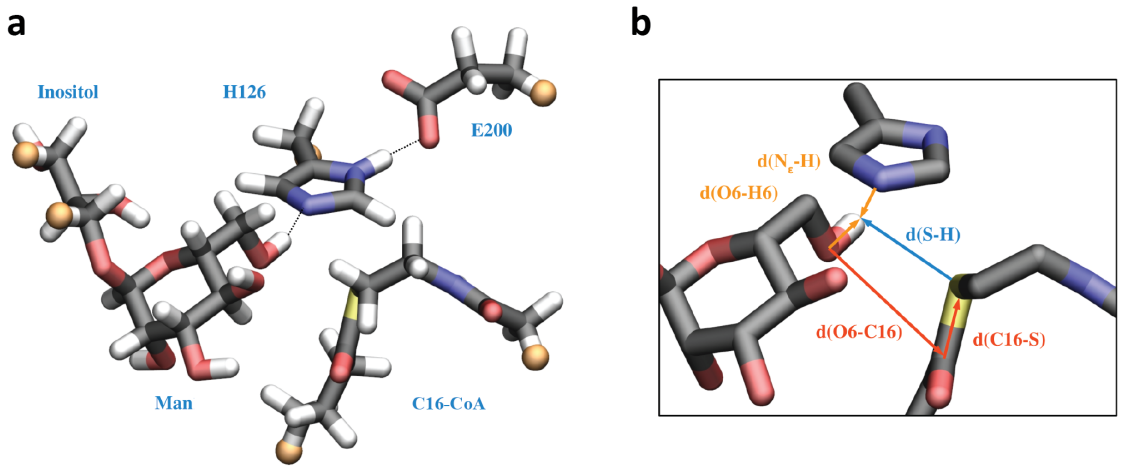



**Figure 60: Collective variables used in the metadynamics simulation.**

The free energy landscape reconstructed from the QM/MM metadynamics simulation (**Figure 61**) shows two minima connected by a unique transition state, suggesting a concerted reaction. However, analysis of the structures along the minimum energy pathway reveals that the minimum on the left-hand side of the FEL does not correspond to the products of the reaction. Instead, the thioester bond is cleaved but the leaving group bears a negative charge (i.e. a thiolate leaving group). Additional analysis by free QM/MM MD confirmed that this species is a stable reaction intermediate, thus the reaction is clearly stepwise. Apparently, reorientation of the thiolate after cleavage of the C-S bond,

forming a strong interaction with Arg201 (**Figure 62**), makes proton transfer from His126 difficult. Therefore, we decided to model the second step of the reaction (protonation of the thiolate leaving group) separately, using as unique collective variable the distance difference between $N_{\varepsilon}$-H and H-S. Analysis of the corresponding free energy profile (**Figure 61**) reveals a concerted proton transfer step, leading to the reaction products $Ac_1PIM_2$ and CoA.

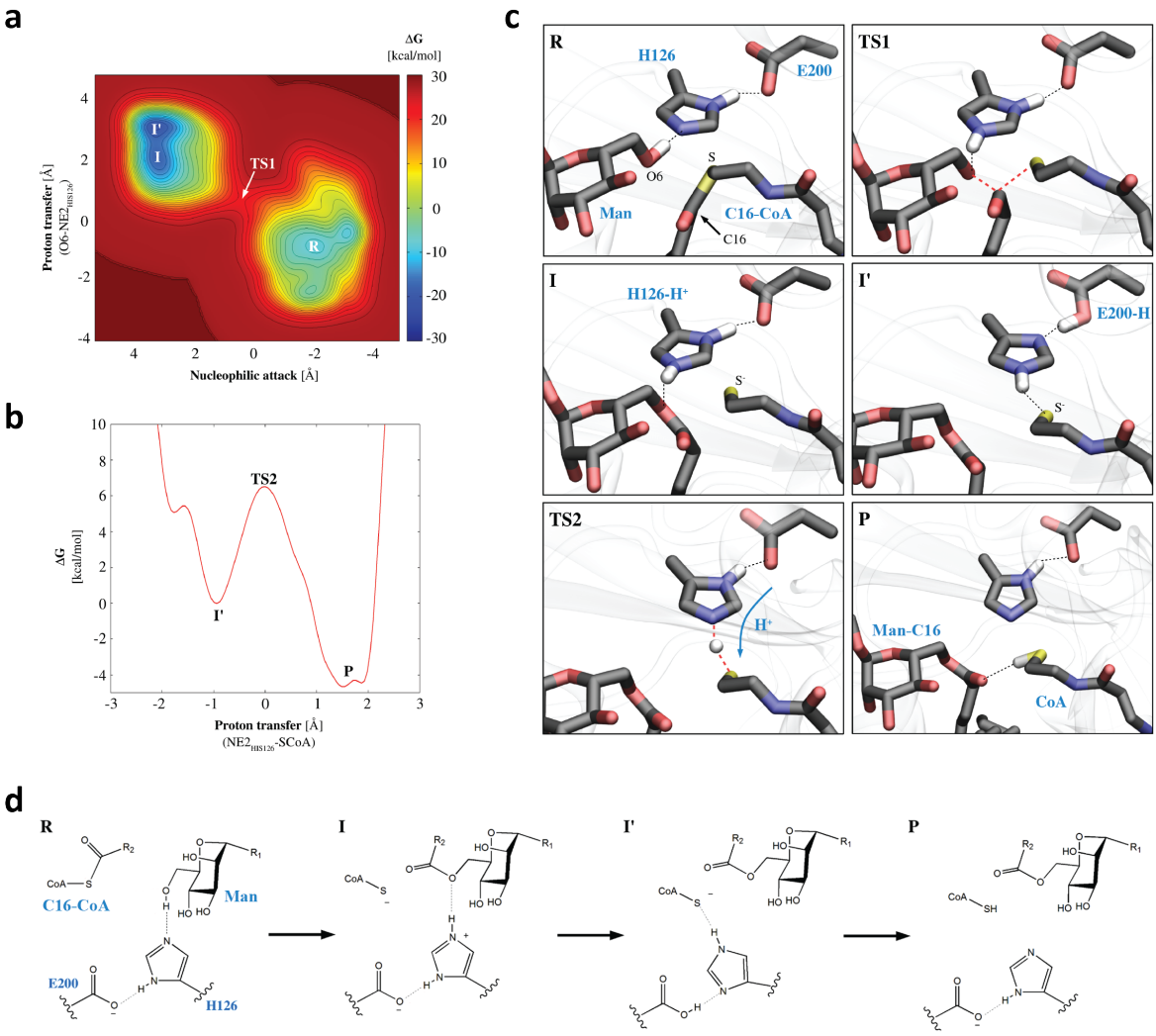



**Figure 61: QM/MM studies on PatA mediated catalysis. a)** Free energy landscape reconstructed from the metadynamics simulation of the first step of the enzymatic reaction (projection on CV1 and CV2). **b)** Free energy profile corresponding to the second reaction step. **c)** Representative structures along the reaction coordinate, obtained from the minimum free energy reaction pathway. Red broken lines represent bonds that are being broken or formed, whereas dotted black lined represent relevant hydrogen bond interactions. **d)** Proposed catalytic mechanism for PatA.

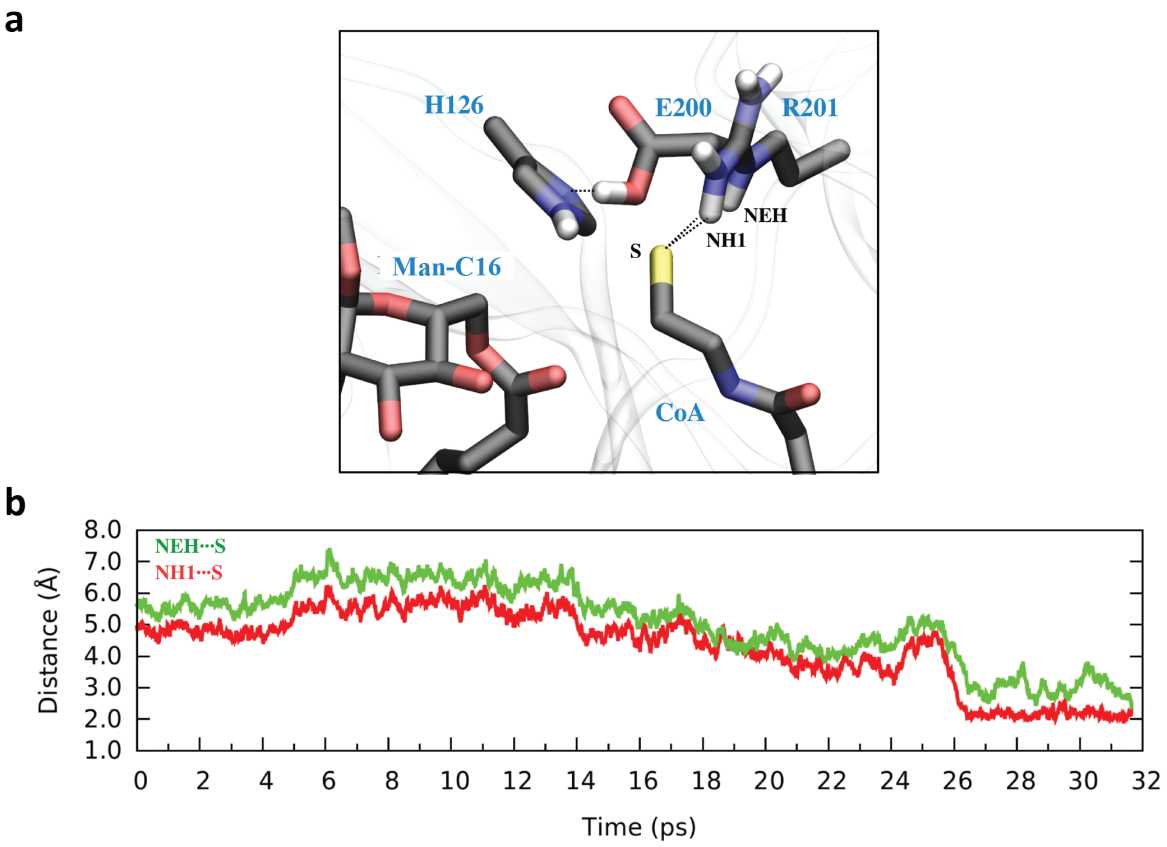



**Figure 62: Interaction of the thiolate leaving group with Arg201.**

To get a detailed picture of the atomic reorganization during the enzymatic reaction, we extracted configurations along the reaction coordinate of both reaction steps (**Figure 61**). The reaction starts with the approach of the mannosyl $CH_2OH$ to the carbonyl of the thioester of Palm-CoA: the O6···C16 distance evolves from 4 Å to 2.8 Å (**Figure 59**). At this point, the C-S thioester bond starts to stretch as the catalytic histidine (His126) abstracts a proton from the mannosyl $CH_2OH$, leading to a transition state (TS; 32 kcal/mol above the reactants) in which the thioester carbon atom exhibits a 4-fold tetrahedral coordination (**Figure 61**). The transition state decays towards a reaction intermediate in which the C-S bond is broken and the mannosyl residue is acylated. The $N_\delta$-H of the catalytic histidine is involved in a low barrier hydrogen bond with the carboxylate side chain of Glu200 (**Figure 61** and **Figure 63**). In the second reaction step, a concerted double proton transfer between Glu200, His126 and the thiolate leaving group takes place (GluH···HisH ···$S^-$ → $Glu^-$···HHis···HS), restoring the original tautomeric state and protonation of His126 and Glu200, respectively (**Figure 61**). The free energy barrier of this reaction step is 6.5 kcal/mol and the products state is 4.6 kcal/mol below the reaction intermediate.

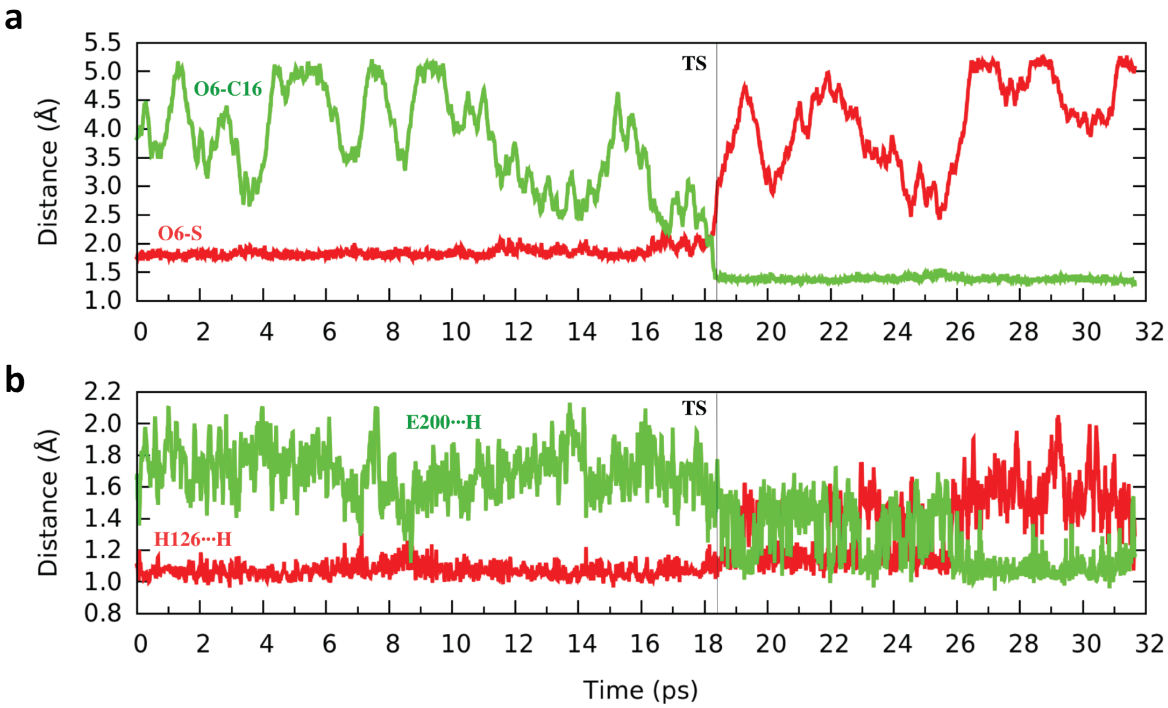



**Figure 63: QM/MM distances along the reaction. a)** Distances involved in the nucleophilic attack. **b)** Distances involves in the charge relay system.

Therefore, it has been shown that PatA is able to acylate $PIM_2$ in a two-step reaction mechanism in which His126 acts as a general base to deprotonate the hydroxyl group of a mannose ring in $PIM_2$, facilitating the nucleophilic attack on the thioester of $C_{16}$-CoA. The reaction is assisted by Glu200, which modulates the pKa of the central histidine as a general base or acid during the catalytic cycle, similarly to the charge-relay system of serine proteases[293]. These results also provide evidence of the formation of a tetrahedral-like transition state, as previously found for the closely related hotdog-fold acyl-CoA thioesterase[294]. These results emphasize the relevance of the charge-relay glutamate, Glu200, and offer valuable detail on the reaction transition state that would serve for the design of new chemotherapeutic drugs.

### 4.2.4.2. Enzymatic Assays

To further validate the proposed model, we designed three single-point substitution, Asp131Ala, Glu200Ala and His126Ala, predicted to impact the transferase and the hydrolytic activity of PatA (**Figure 53 c**). As depicted in **Figure 64**, the functional role of His126 and Glu200 was clearly confirmed, since their substitution by alanine completely inactivated the enzyme.

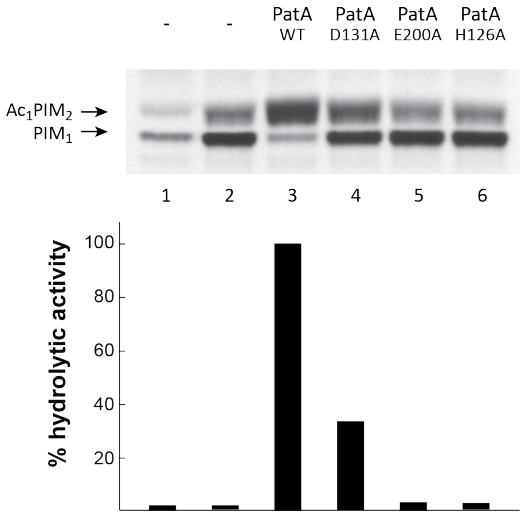



**Figure 64: Enzymatic activity of selected PatA variants.** In the upper panel, the acyltransferase activity of purified PatA and variants involved in catalysis are shown. Reaction mixtures contained crude membranes from *M. smegmatis* $mc^2155$ and GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with PimA (lane 2-6) and palmitoyl-CoA (lanes 3-6) and purified PatA (lane 3), PatA-Asp131Ala (lane 4), PatA-Glu200Ala (lane 5) and PatA-His126Ala (lane 6). In the lower panel, the hydrolytic activity against Palmitoyl-CoA is shown as determined by spectrophotometric analysis. All enzymatic activities measurements were determined in duplicates.

## 4.2.5. A Model of Action for PatA

Biological membranes are primarily used as physico-chemical barriers allowing cells to be functionally constituted and differentiated from the environment. Biological membranes also play critical roles in all living organisms, mediating the selective transport of molecules across the cell, the modulation of the cellular response, as well as the occurrence of a diverse and important set of biochemical reactions[295,296]. Many of the enzymes that work in biological membranes are very often required to interact/recognize both hydrophobic and hydrophilic substrates, located in the membrane and the aqueous environment of the cytosol/organelles, respectively[4,297].

To perform their biochemical functions, these enzymes interact with membranes by two different mechanisms. Whereas integral membrane enzymes are permanently attached (e.g. through hydrophobic α-helices and transmembrane β-sheets), peripheral proteins temporarily associate to membranes by (i) a stretch of hydrophobic residues exposed to the solvent, (ii) electropositive surface patches that interact with acidic phospholipids (e.g. amphipathic α-helices), (iii) protein-protein interactions and/or post-translational modifications.

PatA catalyses a critical step in the biosynthesis of PIMs in mycobacteria[11,12]. The enzyme transfers a palmitoyl moiety from a water-soluble substrate donor palmitoyl-CoA, to the 6-position of the Man*p* ring linked to 2-position of *myo*-inositol of the membrane glycolipids $PIM_1$ or $PIM_2$[21,23]. Therefore, a close interaction of the enzyme with the cytosolic face of the mycobacterial plasma membrane might be a strict requirement for $PIM_1$ or $PIM_2$ modification by PatA. Supporting this notion, PatA was found to co-localize with the mycobacterial membrane fraction[23,24]. Analysis of the amino-acid sequence of the enzyme revealed the lack of a signal peptide or hydrophobic transmembrane segments, suggesting that PatA associates to only one side of the lipid bilayer, a typical feature of peripheral and monotopic membrane proteins[25,26].

To perform their biochemical functions, these proteins very often display a high content of solvent exposed positively charged residues in the form of amphipathic helices, promoting membrane surface interaction with anionic phospholipids. Protein-membrane interaction appears to be mediated following different thermodynamic steps: first, peptide binding is initiated by the electrostatic attraction of the positively charged residues to the anionic membrane, second, most likely followed by the transition of the peptide into the plane of binding, and third, a change of the conformation of the bound peptide[298]. The penetration of the protein depends on the chemical nature of the lipids and peptides involved and also on the mechanistic nature of the processes involved, in which both location and timing of membrane association can be tightly controlled.

Interestingly, the electrostatic surface potential of PatA revealed several hydrophobic/aromatic residues interspersed with positively charged residues in the vicinity of the major groove (**Figure 65**)[21]. This region comprises the amphipathic helices $\alpha_3$, $\alpha_4$, $\alpha_8$, and $\alpha_{12}$ the connecting loop $\beta_2$-$\alpha_8$ (residues 148-154). Importantly, the opposite side of PatA displays a negatively charged surface, certainly contributing to determine the correct orientation of the enzyme into the membrane (**Figure 65**).

Moreover, the structural comparison between PatA and a lipid A biosynthesis acyltransferase *Ab*LpxM crystal structures suggest that $\alpha_1$ and $\alpha_2$ (N-terminal helices) are deeply inserted into the lipid bilayer. Altogether, the structural data support a model in which PatA is permanently associated to the membrane (**Figure 65**).

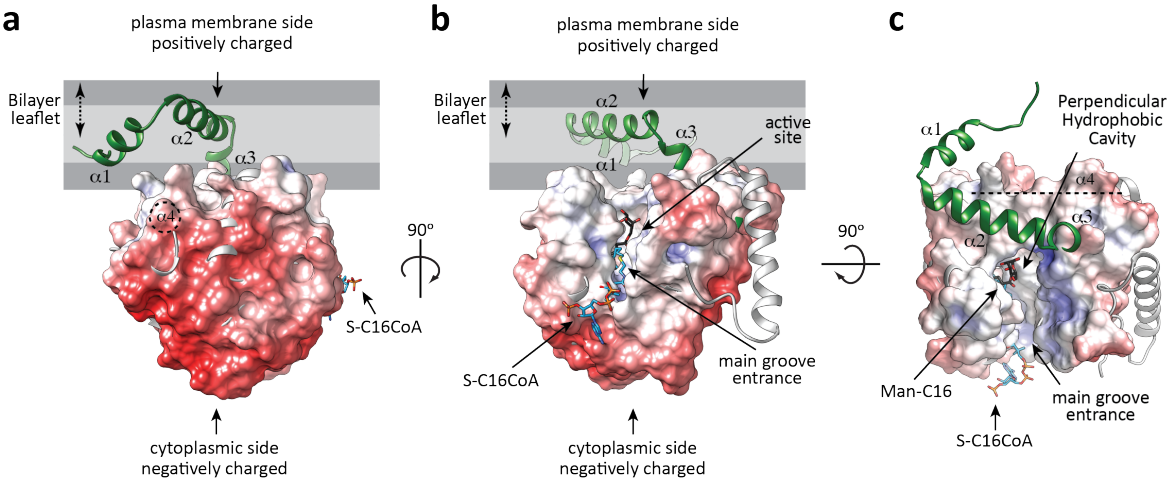



**Figure 65: A model of action for PatA. a)** Electrostatic surface representation of PatA. PatA is a polar protein, with hydrophobic and positively charged residues oriented towards the plasma membrane. Negatively charged residues are facing the cytosol. The first N-terminal three α-helices of *Ab*LpxM are shown in green, whereas the remaining secondary structure elements are shown in grey. **b-c)** Two views of the active site of PatA and the location of the acyl donor and acceptor binding sites.

Despite the fact that acyl transfer is catalysed between the palmitate moiety of palmitoyl-CoA and the mannosyl group of $PIM_1$ or $PIM_2$, our experimental data indicate the enzyme displays an absolute necessity of both fatty acid chains of PI in order for the transfer reaction to take place, pointing to a mechanism requiring a lipid-water interface (**Figure 66**). The fatty acid chains of the acceptor substrate seem to play an important structural role in the generation of a competent active site, assisting/guiding the glycosylated moiety to the appropriate environment in the catalytic centre.

### 4.2.5.1. $PIM_2$ Acyl Chains are Essential for PatA Enzymatic Activity

To further advance in the understanding of how does PatA recognizes its lipid acceptor substrate, the contribution of the fatty acid moieties of $PIM_2$ in the enzymatic activity was explored. It is worth noting that previous reports indicate that $PIM_2$ is the preferred substrate of PatA *in vitro*[24,233]. Consequently, $PIM_2$ was chemically synthesized containing two palmitate groups as fatty acid chains and compared its acceptor capacity with that of the deacylated version of $PIM_2$. The full-length PatA specific hydrolysis activity against palmitoyl-CoA by using diacylated and deacylated $PIM_2$ as acceptors of the reaction and in the presence and absence of DOPC SUVs was studied. Only residual specific hydrolysis activity (up to 5%) of PatA against deacylated $PIM_2$ was observed, similar to the one observed in abscense of any acceptor. Maximum hydrolysis activity in the presence of DOPC SUVs and diacylated $PIM_2$ (0.09 ± 0.004 µmol CoA/min*mg PatA), about 10-fold higher than without DOPC SUVs was observed. Thus, the enzymatic activity assays indicate that (i) the presence of the $PIM_2$ acyl groups and (ii) an artificial model lipid membrane (DOPC SUVs) are essential for the enzymatic activity of PatA, strongly supporting the notion that the acyl chains of the acceptor glycolipid play a prominent structural role in the arrangement of its sugar moiety/moieties into the catalytic centre and the generation of a competent active site (**Figure 66**).

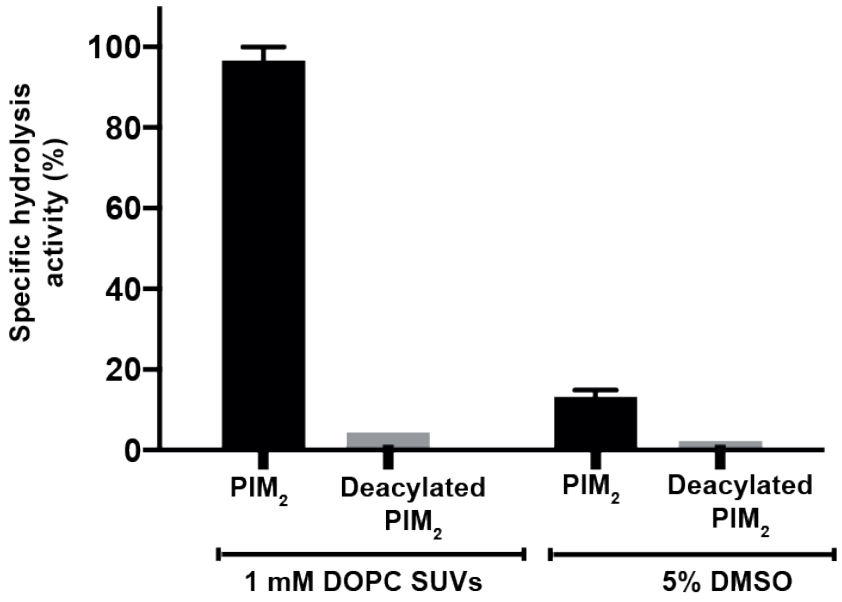



**Figure 66: Enzymatic activity of full-length PatA against palmitoyl-CoA by using diacylated and deacylated $PIM_2$ as acceptors of the reaction and in the presence and absence of DOPC SUVs.**

Thus, a model of interfacial catalysis was proposed, in which PatA recognizes the fully acylated $PIM_1$ or $PIM_2$ substrates with their polar heads within the acceptor binding pocket located into the main groove and the fatty acid moieties likely in close contact with $\alpha_1$, $\alpha_2$ and $\alpha_4$. The conformation of the first two α-helices $\alpha_1$ and $\alpha_2$ in a membrane environment remains still intriguing. The palmitoyl

CoA binding site involves two grooves (**Figure 47** and **Figure 51**). The palmitate chain is deeply buried into the hydrophobic pocket that runs perpendicular to the main groove, in the core of the enzyme. The 4-phosphopantetheinate moiety is located at the entrance of the main groove, in close contact with a highly conserved region flanked by the $\beta_2$-$\alpha_8$ (residues 149-153), $\beta_3$-$\alpha_9$ (residues 174-180) and $\beta_4$-$\beta_5$ (residues 199-207) loops, and two alpha helices, $\alpha_9$ (residues 181-190) and $\alpha_{10}$ (residues 221-230). The adenosine 3',5'-diphosphate (3',5'-ADP) moiety of the ligand (disordered in three of the four protomers of the asymmetric unit) sticks out from the globular core and is exposed to the bulk solvent. This structural configuration of the acyl-CoA binding site seems to be preserved in the homolog enzymes *Tm*PlsC, *Ab*LpxM and *Cm*GPAT, supporting a common binding mode for acyl-CoAs[28–30].

Interestingly, *Tm*PlsC, *Ab*LpxM and *Cm*GPAT prefer acyl-ACP as donor substrates, nevertheless they can also function with acyl-CoA derivatives[49,58,122], suggesting the enzymes preserve common structural features to recognize both acyl donor substrates, which might certainly involve important conformational changes.

## 4.2.6. Structural Similarities of PatA with Other Acyltransferases

### 4.2.6.1. Structural Homologs of PatA: *Cm*GPAT, *Ab*LpxM and *Tm*PlsC

At the time that PatA crystal structure was published, a structure homologue search using DALI server revealed only one protein with significant structural similarity, the glycerol-3-phosphate acyltransferase from *Cucurbita moschata* (**Figure 4**)[27,28]. However, at the present time, structure homologue search using DALI server reveal three proteins with significant structural similarity:

i) Glycerol-3-phosphate acyltransferase from *Cucurbita moschata*[27,28] (*Cm*GPAT; **Figure 4**):

   *Cm*GPAT catalyses the transfer of an acyl chain either from acyl-ACP or acyl-CoA, to the *sn*-1 position of glycerol-3-phosphate, to form 1-acylglycerol-3-phosphate[57,58]. *Cm*GPAT is able to use palmitoyl-CoA as a donor substrate, as PatA does, following a bi-bi ordered mechanism[58].

ii) Acyltransferase LpxM from *Acinetobacter baumannii*[29] (*Ab*LpxM*;* **Figure 18**):

   *Ab*LpxM catalyses the transfer of two lauroyl ($C_{12}$) groups at the R-3'- and R-2-hydroxymyrsitate positions of lipid A[19,84,143,299,300]. LpxM from *Escherichia coli* (*Ec*LpxM) prefers acyl-acyl-carrier protein (acyl-ACP) donors but can also function with acyl-CoA substrates[19].

iii) 1-acyl-glycerol-3-phosphate acyltransferase from *Thermotoga maritime*[30] (*Tm*PlsC; **Figure 5**):

   *Tm*PlsC catalyzes the transfer of a palmitate chain ($C_{16}$) either from acyl-ACP or acyl-CoA, to the *sn*-2 position of 1-acylglycerol-3-phosphate to form PA. *Tm*PlsC prefers acyl-acyl-carrier protein (acyl-ACP) donors but can also function with acyl-CoA substrates[30].

PatA and *Ab*LpxM enzymes are distantly related to *Cm*GPAT and *Tm*PlsC families of bacterial acyltransferases, and to the dihydroacetone phosphate (DHAPAT) and 2-acylglycero-phosphatidylethanolamine (LPEAT) families of eukaryotic acyltransferases[18,21,38] (**Figure 67**).

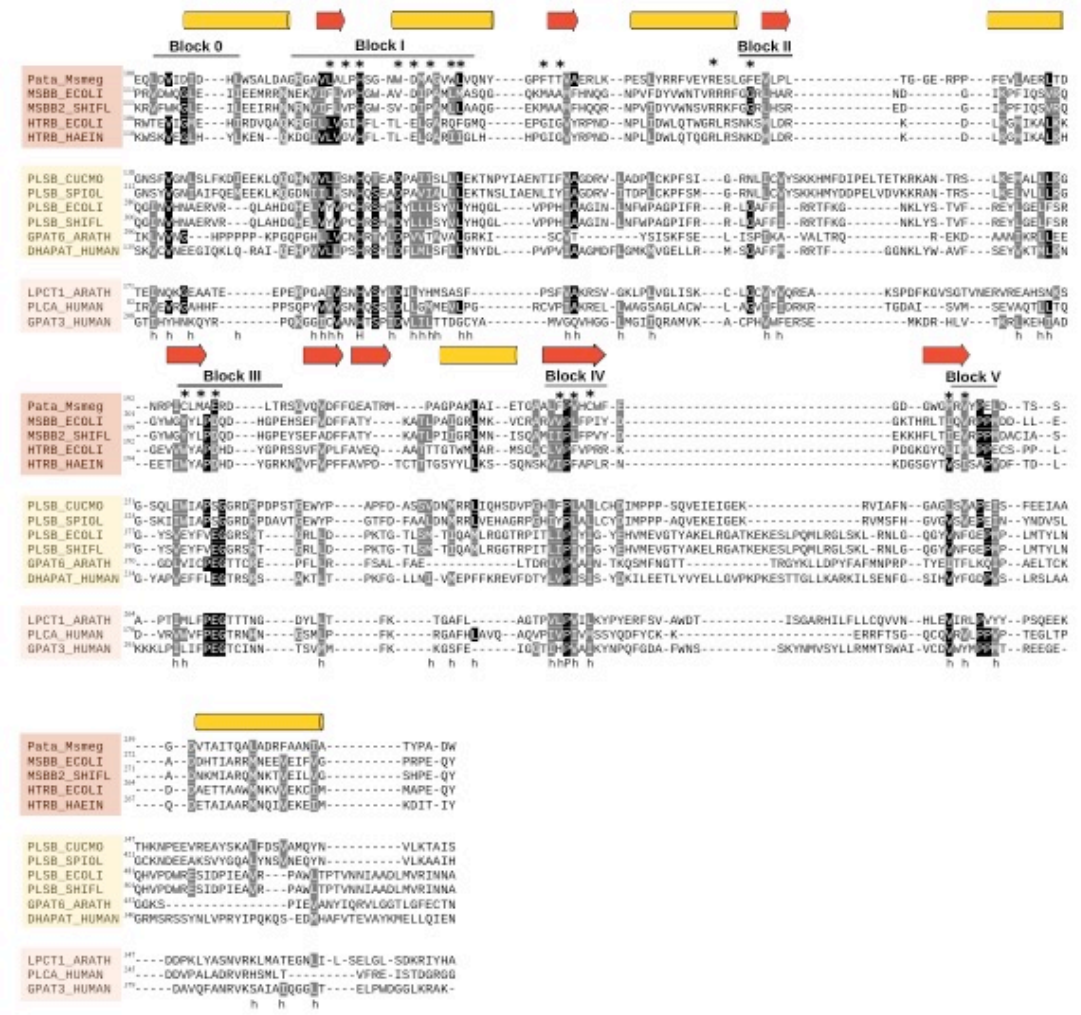



**Figure 67: Structure weighted sequence alignment of PatA with prokaryotic/eukaryotic acyltransferases.** The protein sequences were extracted from UniProt accession numbers A0QWG5 from *M. smegmatis* (PatA_Msmeg); P24205 from *Escherichia coli* (MSBB or LpxM_ECOLI); O06659 from *Shigella flexneri* (MSBB or LpxM_SHIFL); P0ACV0 from *E. coli* (HTRB or LpxL_ECOLI); P45239 from *Haemophilus influenzae* (HTRB or LpxL_HAEIN); P10349 from C. moscata (GPAT_CUCMO); Q43869 from *Spinacia oleracea* (GPAT_SPIOL); P0A7A7 PlsB from *E. coli* (PLSB_ECOLI); Q7UBC6 PlsB from *S. flexneri* (PLSB_SHIFL); O80437 from *Arabidopsis thaliana* (GPAT6_ARATH); O15228 from *Homo sapiens* (DHAPAT_HUMAN); Q8L7R3 from *Arabidopsis thaliana* (LPCT1_ARATH); Q99943 from *H. sapiens* (PLCA_HUMAN); and Q53EU6 from *H. sapiens* (GPAT3_HUMAN). Conserved positions are shown in black and grey background. Conserved hydrophobic residues are indicated with an h. The secondary structural elements corresponding to the 3D structure of PatA are shown above the alignment. Catalytic amino acids and those involved in palmitate binding are indicated as asterisks and black circles, respectively.

GPAT, LPAAT, DHAPAT, and LPEAT display highly conserved residues distributed in four regions, and named as Blocks I-IV[18,31]. According to multiple sequence alignments among the LpxM, GPAT, LPAAT, DHAPAT and LPEAT families, weighted by structural alignment of PatA and *Cm*GPAT, a common core can be defined (**Figure 67**). It is worth noting that the classical names of Blocks I-IV have been respected as reported in the literature, introducing two new regions as Blocks 0 and V.

According to the literature and structural analysis, Block I is localized in the long groove and it contains the catalytic motif HxxxD/E. Block II is placed in a region near the donor substrate, it could help to

stack the acyl chain. Block III is localized adjacent to the catalytic site and contains the catalytic glutamate/aspartate that is porposed to protonate the catalytic histidine in PatA and *Ab*LpxM. Block IV is adjacent to Block V and could it has a structural function. Block O and V are two adjacent regions localized in an area that close the hydrophobic tunnel that accommodates the acyl chain (**Figure 67**).

As depict in **Figure 68**, the structural alignment of the PatA homologs detected by DALI server reveal several regions with similar physic-chemical properties (highlighted in yellow). The conserved blocks previously described are conserved, however, new regions between these homologs seems to be outlined, as α-helices 9 and 10.

Critical residues and their interactions in the reaction centre and in the donor binding site are essentially preserved in all these acyltransferase families, strongly supporting a common structural core and catalytic mechanism. Therefore, the high variability of acceptor substrates of these acyltransferases families is reflected in the lack of conserved regions regarding the acceptor binding site.

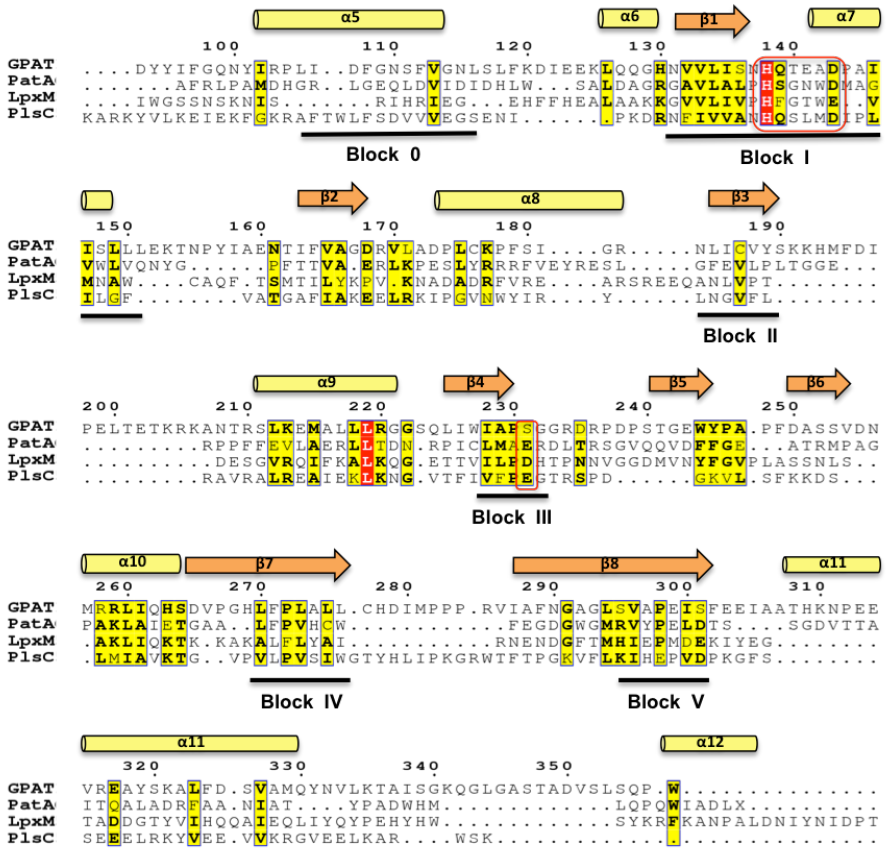



**Figure 68: Structure-based Alignment of PatA, *Ab*LpxM, *Cm*GPAT and *Tm*PlsC.** The uniprot accession numbers A0QWG5 from *M. smegmatis* (*Ms*PatA), S3TFW2 from *A. baumannii* (*Ab*LpxM), P10349 from *C. moschata* (*Cm*GPAT) and Q9X219 from *T. maritime* (*Tm*PlsC); and the X-ray crystal structures with pdb codes 5f2t (*Ms*PatA), 5kn7 (*Ab*LpxM), 1iuq (*Cm*GPAT) and 5kym (*Tm*PlsC) were used to generated the structure-based alignment. Conserved positions are shown in highlighted in red, equivalent residues from physic-chemical point of view are highlighted in yellow and catalytic residues are remarked in a red square. The secondary structural elements corresponding to the 3D structure of PatA are shown above the alignment.

### PatA and *Cm*GPAT

*Cm*GPAT belongs to a large family of glycerol-3-phosphate acyltransferases (GPAT), which are critical in the biosynthesis and regulation of phospholipids composition in prokaryotic and eukaryotic cells[31,56,301]. The structure of *Cm*GPAT is composed of two domains: (i) a helical domain comprising the first 78 residues of the protein displays a four-helix bundle architecture of unknown function, and (ii) an α/β domain, consisting of a nine-stranded continuous β-sheet surrounded by 11 α-helices (**Figure 70 a**). As depicted in **Figure 70 a**, the central β-sheet of *Cm*GPAT superimposes well with that observed in PatA, with the exception of its outermost strands $\beta_1$, $\beta_2$ and $\beta_9$, which are missing in PatA (**Figure 70**).

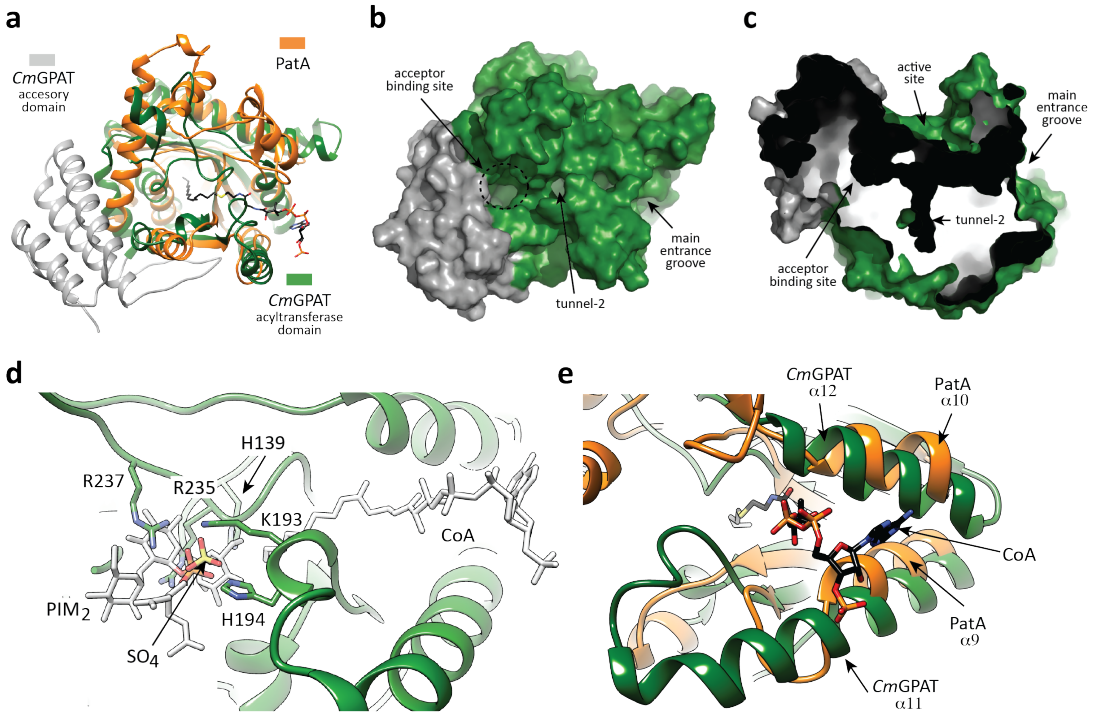



**Figure 69: Structural similarities between PatA and *Cm*GPAT. a)** Structural superposition of PatA and *Cm*GPAT. Secondary structure elements are labelled. The localization of the fatty acid and acceptor binding sites in *Cm*GPAT are indicated. **b-c)** The active site of *Cm*GPAT is located into a main groove, with a hydrophobic tunnel running perpendicular and deeply buried into the core of the enzyme. **d)** Structural superposition of the proposed acceptor binding site in PatA and *Cm*GPAT, showing donor and acceptor PatA substrates for spatial reference in white colour. **e)** Structural superposition of the CoA binding site in PatA and *Cm*GPAT.

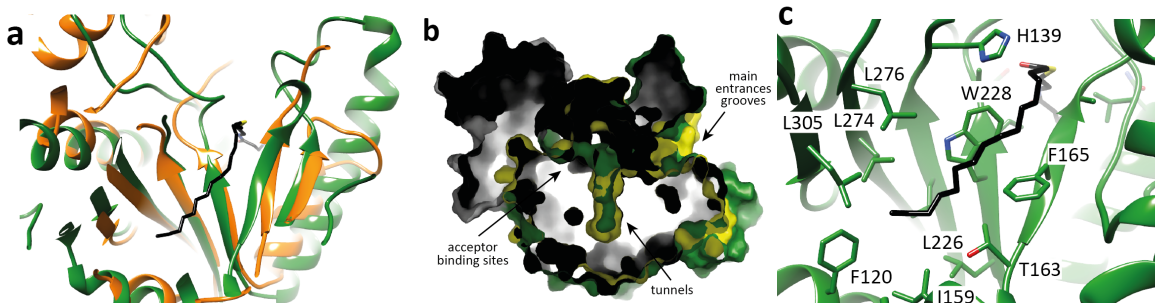



**Figure 70: Structural comparison between PatA and *Cm*GPAT. a)** Structural superposition of the β-sheet core of PatA (orange) and that of the *Cm*GPAT (green). **b)** Structural superposition of the acyl binding tunnels in PatA (yellow) and *Cm*GPAT (green). **c)** Localization of the fatty acid binding site in *Cm*GPAT is revealed by docking calculations.

Interestingly, the suggested $PIM_1$/$PIM_2$-binding site in PatA is located in the same region of the glycerol-3-phosphate-binding site in *Cm*GPAT (**Figure 70 d**). A sulphate ion observed in the crystal structure of *Cm*GPAT superimposed very well with the phosphate moiety of the inositol ring of $PIM_1$ or $PIM_2$ (**Figure 70 d**). It is worth noting that, as acceptors exhibit a marked diversity of chemical structures compared with acyl-CoA or acyl-ACP donors, the acceptor-binding sites reflect this variability by showing different rearrangements of secondary structural elements (**Figure 70 d**).

A region comprising residues 128-187 is important for acyl-CoA selectivity. This region completely covered the narrowest and most hydrophobic of the three tunnels, tunnel-2 (**Figure 70 b-c**), which was proposed to be involved in fatty-acid recognition[28]. Strikingly, the acyl-binding pocket identified in the crystal structures of the palmitoyl-PatA and PatA-S-$C_{16}$CoA complexes superimposed very well with tunnel-2 of *Cm*GPAT (**Figure 71**). On the basis of the experimental location of the S-$C_{16}$CoA in PatA, a palmitoyl-CoA molecule was fitted into tunnel-2 and subjected to energy minimization (**Figure 70** and **Figure 71**). First, the palmitate moiety accommodates into the hydrophobic tunnel. The walls of the tunnel are covered with mainly conserved, hydrophobic and aromatic residues including Leu135 ($\beta_3$), Pro145 and Ile148 ($\alpha_7$), Ile159 ($\alpha_8$), Thr163 ($\alpha_8$-$\beta_4$ loop), Phe165 ($\beta_4$), Phe180 ($\alpha_{10}$), Leu186 ($\beta_5$), Leu226 and Trp228 ($\beta_6$), Leu274 and Leu276 ($\beta_7$) and Leu305, with Phe120 and Ile123 ($\alpha_6$) making up the cap.

Interestingly, the CoA-binding site observed in PatA is also conserved in *Cm*GPAT. Specifically, the acyl donor is located in the region corresponding to the entrance of the main groove (**Figure 70 c-e**). The two α-helices $\alpha_9$ and $\alpha_{10}$, flanking the entrance of acyl-CoA to the main groove of PatA, are structurally equivalent to $\alpha_{11}$ (residues 201–222) and $\alpha_{12}$ (residues 251–265) in *Cm*GPAT (**Figure 70 e**). The substitution of Leu261 ($\alpha_{12}$) by an aromatic residue, which is located at the interface of $\alpha_{11}$-$\alpha_{12}$ helices, caused major changes in the selectivity of *Cm*GPAT for acyl-ACP derivatives[59]. Overall, major secondary structural elements are structurally preserved across both families of acyltransferases supporting a common binding mode for palmitoyl-CoA.

In *Cm*GPAT, the catalytic site is also located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases[18]. His139 and the adjacent Asp144 were proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the acyl-CoA[17,27,28,61]. In PatA, His126 occupies the same location to that observed for His139 in *Cm*GPAT. However, the aspartic acid Asp131 of the $HX_4D$ motif displays a completely different arrangement when compared with Asp144 (**Figure 71**). Specifically, the carboxylate group of Asp131 makes a strong hydrogen bond with the side chain OH of Tyr83, and additional electrostatic interactions with the lateral chains of Trp130, Tyr163 and Arg164, suggesting a structural role for this residue. In that sense, the replacement of Asp131 by alanine preserved 37% of the enzymatic activity for $PIM_2$ (**Figure 64**).

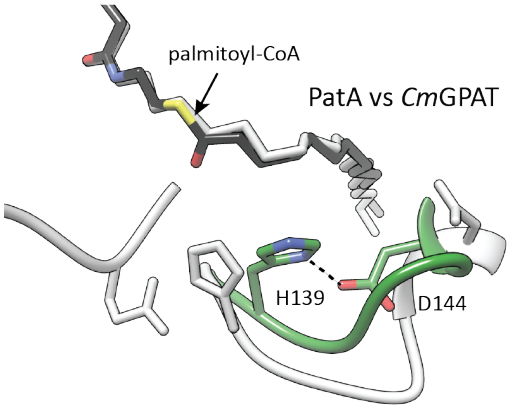



**Figure 71: Structural superposition of the catalytic site of PatA (grey) and *Cm*GPAT (green) with palmitoyl-CoA docked into *Cm*GPAT.**

### PatA and *Ab*LpxM

A detailed comparison of the PatA-Man-$C_{16}$ complex with that of *Ab*LpxM in complex with a n-dodecyl-β-D-maltoside molecule, provide insights into the molecular mechanism of acceptor recognition (**Figure 72 a-d**). As depicted in **Figure 72**, the central β-sheet of *Ab*LpxM superimposes very well with that observed in PatA. Strikingly, the main groove and the acyl-binding pocket identified in the crystal structures of PatA-Man-$C_{16}$, PatA-S-$C_{16}$CoA and PatA-$C_{16}$ complexes superimposed very well with the corresponding main groove and a hydrophobic cavity observed in *Ab*LpxM (**Figure 72** and **Figure 73**).

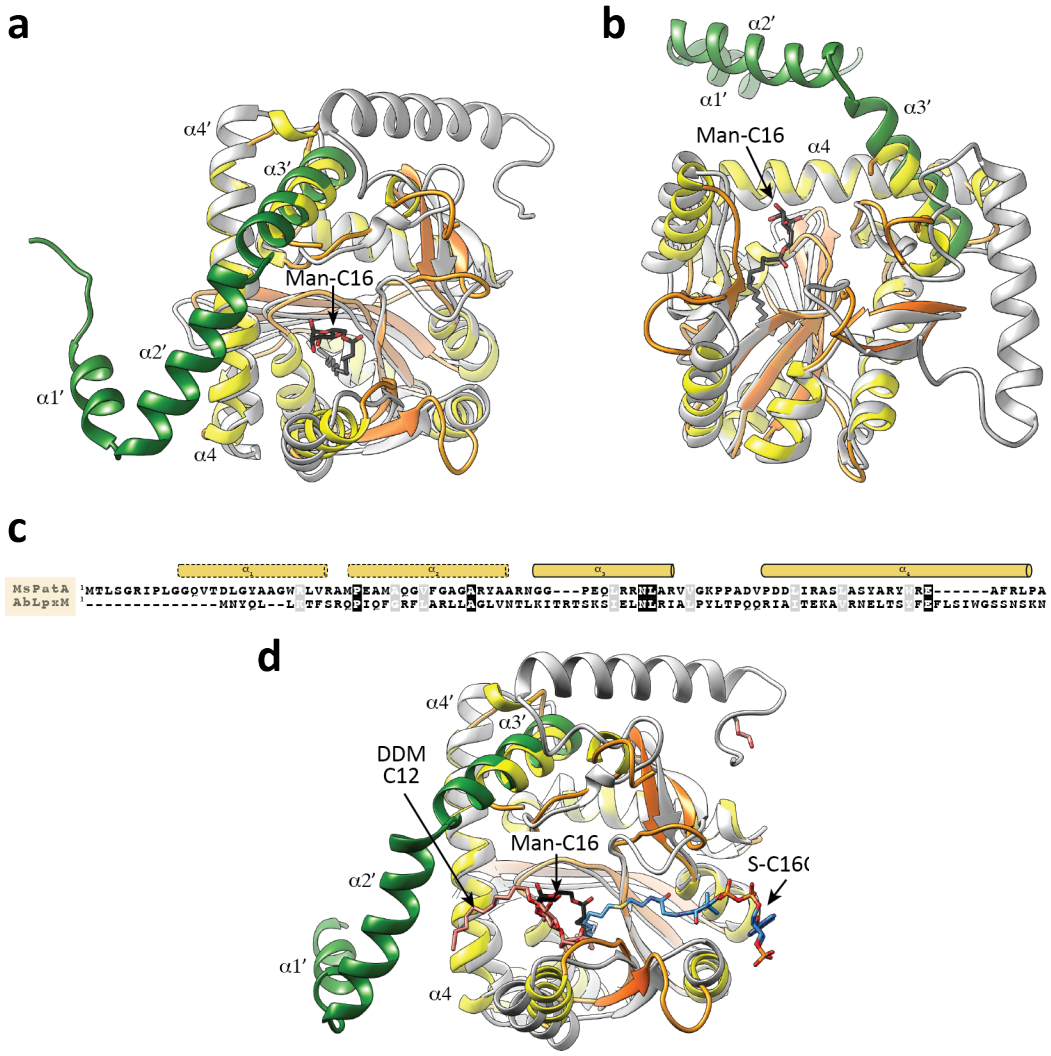



**Figure 72: Structural similarities between PatA and *Ab*LpxM. a-b)** Cartoon representation showing two views of the structure weighted sequence alignment of PatA with *Ab*LpxM. PatA β-strands and α-helices are shown in orange and yellow, respectively. The Man-$C_{16}$ product is shown in black and red. The first N-terminal three α-helices of *Ab*LpxM are shown in green, whereas the remaining secondary structure elements are shown in grey. **c)** Structure weighted sequence alignment of a selected region of PatA and *Ab*LpxM, comprising residues that code for the four N-terminal α-helices. **d)** Cartoon representation showing the structural superposition of PatA and *Ab*LpxM. Both structures are coloured as in panel a). The Man-$C_{16}$ product is shown in black and red. The S-$C_{16}$CoA acyl donor substrate derivative is shown in light blue and red. The DDM molecule is shown in orange.

On the basis of the experimental location of the S-$C_{16}$CoA in PatA, and taking into account that *Ab*LpxM displayed a strong preference for lauryl chains ($C_{12}$) over other acyl chain lengths tested ($C_{10}$, $C_{14}$ and $C_{16}$)[29], a lauryl-CoA molecule was fitted into equivalent cavity and subjected to energy minimization (**Figure 73**). The lauryl moiety accommodates into the hydrophobic cavity, which is shorter than that observed in the PatA-S-$C_{16}$CoA complex, pointing to the occurrence of a hydrocarbon ruler mechanism[29,149,302]. The walls of the tunnel are covered with hydrophobic residues including Leu118 ($\beta_2$), Val120 ($\beta_2$-$\alpha_7$ loop), Trp126 and Met129 ($\alpha_7$), Met139 and Ile141($\beta_3$), Val188 ($\beta_5$), and Leu190 ($\beta_5$-$\beta_6$ loop); with Ile97 and Ile100 ($\beta_1$), Phe106 ($\alpha_6$), Trp132 ($\alpha_7$), Phe136 ($\alpha_7$-$\beta_3$ loop), Leu229 ($\beta_8$), and Ile243 and Met241 ($\beta_9$), building the bottom cap[29].

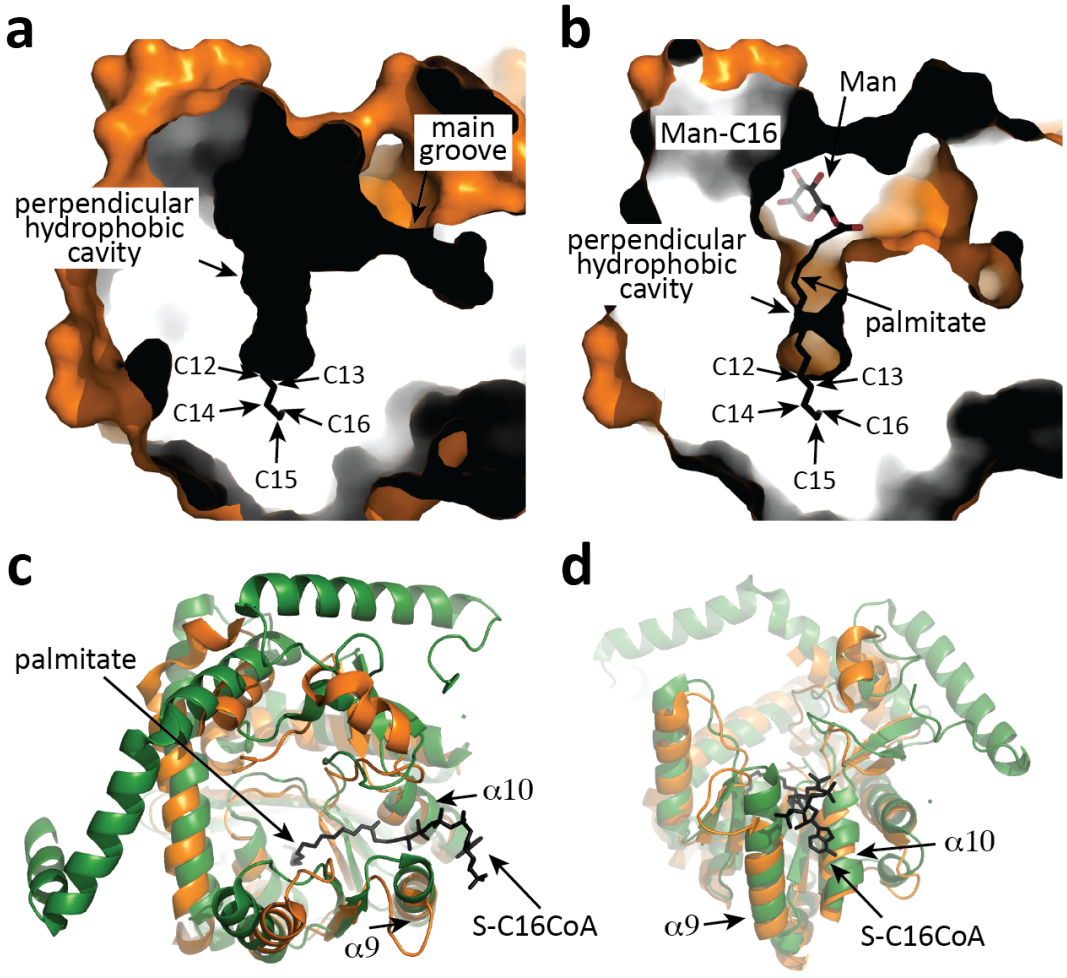



**Figure 73: PatA and *Ab*LpxM share a common acyl donor binding site. a-b)** Structural superposition of the acyl binding tunnel in *Ab*LpxM (orange) and the palmitate residue of $C_{16}$-CoA (black) observed in PatA (yellow) **c-d)** Structural superposition of PatA (orange) and *Ab*LpxM (green) and the localization of the $C_{16}$-CoA binding site.

The CoA-binding site observed in PatA is also conserved in *Ab*LpxM. Specifically, the 4-phosphopantetheinate moiety of S-$C_{16}$CoA is located at the entrance of the main groove (**Figure 73 c-d**). The two α-helices $\alpha_9$ and $\alpha_{10}$, flanking the entrance of acyl-CoA to the main groove of PatA, are structurally equivalent to $\alpha_9$ (residues 170-183) and $\alpha_{10}$ (residues 213-223) in *Ab*LpxM, as also

observed in *Cm*GPAT, pointing to a common binding mode (**Figure 73 c-d**)[21,29]. *Ab*LpxM crystallized with one molecule of n-dodecyl-β-D-maltoside, with the glucosyl residues deeply buried into a hydrophobic pocket located in one side of the main groove. Strikingly, the first glucose moiety of DDM superimposed well with the Man residue of Man-$C_{16}$ located in the acceptor binding site of PatA. The second glucose moiety of DDM superimposed well with the inositol residue of $PIM_2$[11,21]. Interestingly, the $C_{12}$ acyl chain of DDM localizes in a hydrophobic cavity formed by $\alpha_1$, $\alpha_2$ and $\alpha_4$ in *Ab*LpxM, conserved in PatA, suggesting a binding mode for the acyl chains of the lipid acceptor substrates[29].

In *Ab*LpxM, the catalytic site is also located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases[18]. His122 from Block I, and the adjacent Asp192 from Block III, are proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the acyl-CoA[17,27,28,61]. In PatA, His126 occupies the same location to that observed for His122 in *Ab*LpxM and PatA Glu200 is in the same position of Asp192, showing a very similar catalytic mechanism (**Figure 74**).

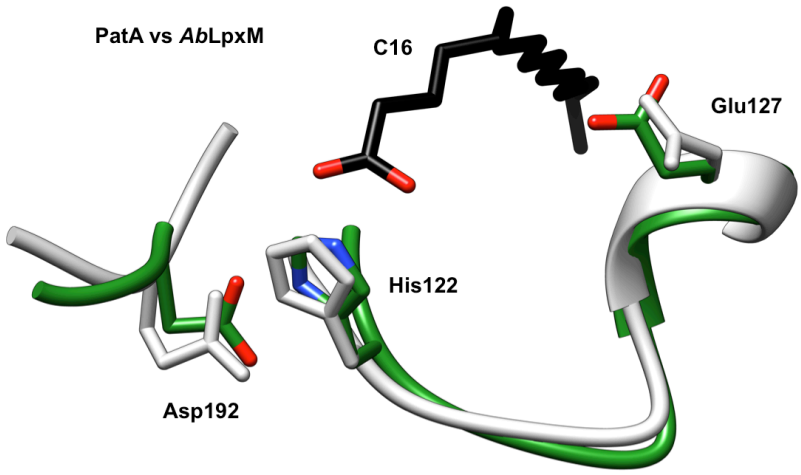



**Figure 74: Structural superposition of the catalytic site of PatA (grey) and *Ab*LpxM (green) with palmitoyl docked into *Ab*LpxM.**

### PatA and *Tm*PlsC

The structure of *Tm*PlsC is composed of two domains: (i) a helical domain comprising the first 61 residues of the protein displays a two-helix bundle architecture and it has been proved to be inserted into the lipid bilayer, and (ii) an α/β domain that comprise the following 62-247 residues, consisting of a seven-stranded continuous β-sheet surrounded by nine α-helices (**Figure 5 a**). As depicted in **Figure 75**, the central β-sheet of *Tm*PlsC superimposes well with that observed in PatA, with the exception of its outermost strands $\beta_6$, $\beta_5$ and $\alpha_4$ which are missing in *Tm*PlsC (**Figure 75**).

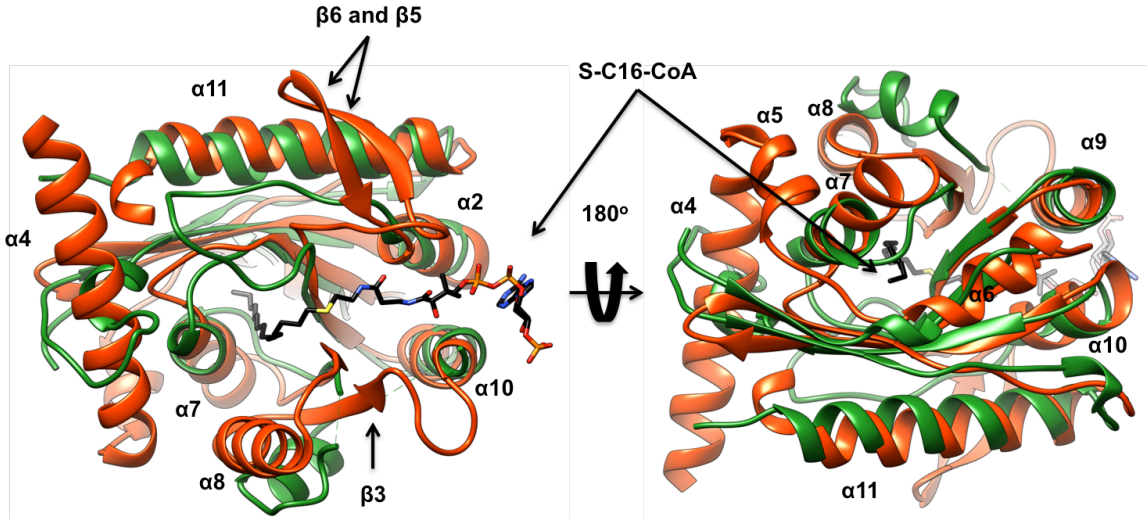



**Figure 75: Superposition of the α/β domain of PatA and *Tm*PlsC, a common main groove and a common structural core is share in both enzymes.** Structural superposition of the core of PatA (orange), the α/β domain of *Tm*PlsC (green) and the S-$C_{16}$-CoA (black).

A detailed comparison of the PatA-S-$C_{16}$-CoA complex with α/β domain of *Tm*PlsC, provide insights into the molecular mechanism of acceptor recognition (**Figure 75**). As depicted in **Figure 75**, four out of the six PatA β-strands superimposes very well the central β-sheets of *Tm*PlsC ($\beta_8$, $\beta_7$, $\beta_4$ and $\beta_1$). However, and $\beta_2$ and $\beta_3$ are not observed in *Tm*PlsC central β-sheet core.

Interestingly, Robertson et al. proposed a model in which *Tm*PlsC acyl-binding poket is different than the acyl-binding poket of PatA (**Figure 76 a**)[30]. In their model the acyl chain runs parallel to main groove, not perpendicular as is the case of PatA (**Figure 76 b**). In this model the hydrophobic tunnel proposed to accommodate the substrate in *Tm*PlsC includes residues Tyr20, Ile21, Gly25, Ile49 and Phe52 (coloured in yellow in **Figure 76 b**). The model is based in an MD simulation and enzymatic activity measurements, Gly25 is proposed to close the hydrocarbon ruler hydrophobic tunnel of the $C_{16}$ acyl chain, which is the preferred chain length for *Tm*PlsC[29,30,149,302]. The mutation of this Gly25 by a bulkier residue (Gly25Met) leads to an increase of the $C_{14}$-CoA activity and a decrease of $C_{16}$-CoA activity[30].

It is important having into account that *Tm*PlsC substrate perform the second acylation of PA, thus it seems that two acyl chains need to be accommodated into the enzyme, this could involve the existence of two acyl-binding sites. Even though the surface analysis of *Tm*PlsC reveals the collapse of the PatA acyl-binding poket in *Tm*PlsC (**Figure 76 c**), when the S-$C_{16}$-CoA molecule from the PatA-S-$C_{16}$-CoA complex is fitted into *Tm*PlsC it accommodates well. Interestingly, the residues that are in contact with the palmitate are Val80, Ile90, Ile93, Val97, Ala98, Phe102, Ile152, Phe154, Thr158, Val186 and catalytic His84 (coloured in red in **Figure 76 b**). To further clarify the acyl-binding site of *Tm*PlsC, enzymatic assays with a *Tm*PlsC mutant of residue Ala98, which is located at the end of the PatA acyl-binding poket in *Tm*PlsC, for a bulkier residue could reveal the use of this conserved tunnel in *Tm*PlsC in solution.

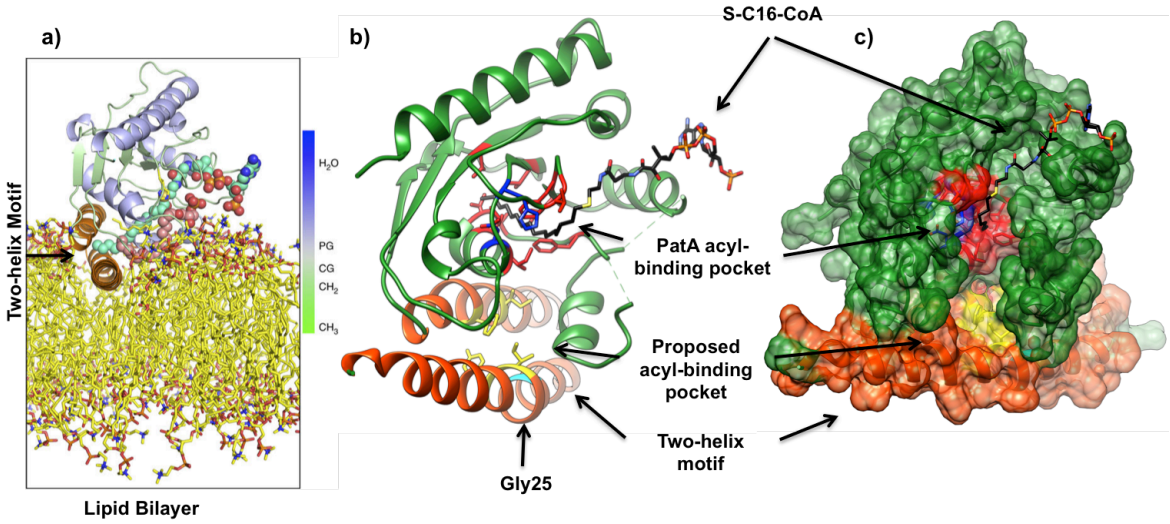



**Figure 76: *Tm*PlsC and PatA acyl-binding poket models. a)** An MD-derived model of *Tm*PlsC bound in the lipid bilayer. The color-coding is as follows: N-terminal two-helix motif, orange; core catalytic αβ-domain, blue and green; active site histidine, yellow; LPA substrate, salmon carbons; acyl donor, teal carbons; membrane phospholipids, yellow acyl chains, red/orange phosphate groups and blue head groups. The blue-to-green scale reflects the water-to-lipid occupancy of one half of the lipid bilayer relative to the proposed location of the *Tm*PlsC-substrate complex. The scale is based on biophysical studies31 in which $H_2O$, phospholipid head group (PG), carbonyl glycerol (CG) and hydrocarbon CH2 and CH3 zones identify the highest probability of finding these moieties within one leaflet of the bilayer. Modified from[30]. **b)** *Tm*PlsC residues proposed to be interacting with the donor substrate acyl-chain (in yellow) and residues that would interact with the acyl chain in the case that *Tm*PlsC would have the same acyl-binding pocket than PatA (in red). Catalytic residues His84 and Asp89 are coloured in dark blue and hydrocarbon ruler proposed residue Gly25 is coloured in cyan. **c)** *Tm*PlsC surface analysis of the acyl-binding pocket model. In red, residues that would interact with the acyl chain in the case that *Tm*PlsC would have the same acyl-binding poket than PatA. In yellow, proposed residues of the acyl-binding pocket of *Tm*PlsC. In orange, the two-helix motif proposed to interact with the lipid bilayer.

In *Tm*PlsC, the catalytic site is also located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases[18]. His184 from Block I, and the adjacent Asp89 from Block I, are proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the acyl-CoA[17,27,28,61]. In PatA, His126 occupies the same location to that observed for His84 in *Tm*PlsC. However, the PatA Asp131 of the $HX_4D$ motif displays a completely different arrangement when compared with Asp89 (**Figure 77**). The catalytic mechanism proposed for *Tm*PlsC is very similar to the one proposed for *Cm*GPAT, while PatA and *Ab*LpxM share the same model.

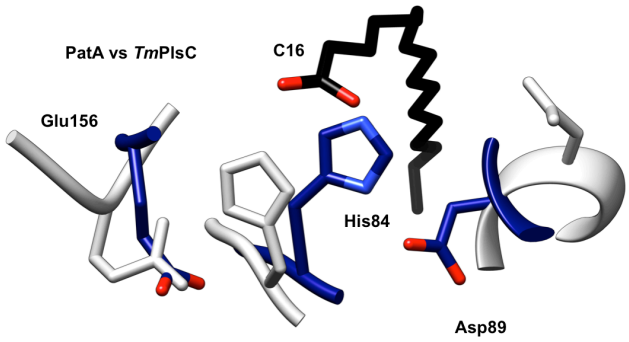



**Figure 77: Structural superposition of the catalytic site of PatA (grey) and *Tm*PlsC (green) with palmitate docked into *Tm*PlsC.**

### 4.2.6.2. New Family of Acyltransferases with a Common Structural Core

In **Figure 76** is depict the superposition of the α/β domain of PatA, *Ab*LpxM, *Cm*GPAT and *Tm*PlsC. The four acyltransferases seem to share a common structural core of six α-helices ($\alpha_{11}$, $\alpha_{10}$, $\alpha_9$, $\alpha_8$, $\alpha_7$ and $\alpha_6$) and four β-sheets ($\beta_1$, $\beta_4$, $\beta_7$ and $\beta_8$). PatA and *Ab*LpxM share four more β-strands ($\beta_2$, $\beta_3$, $\beta_5$ and $\beta_6$), being $\beta_2$ and $\beta_3$ also shared wih *Cm*GPAT (**Figure 78**).

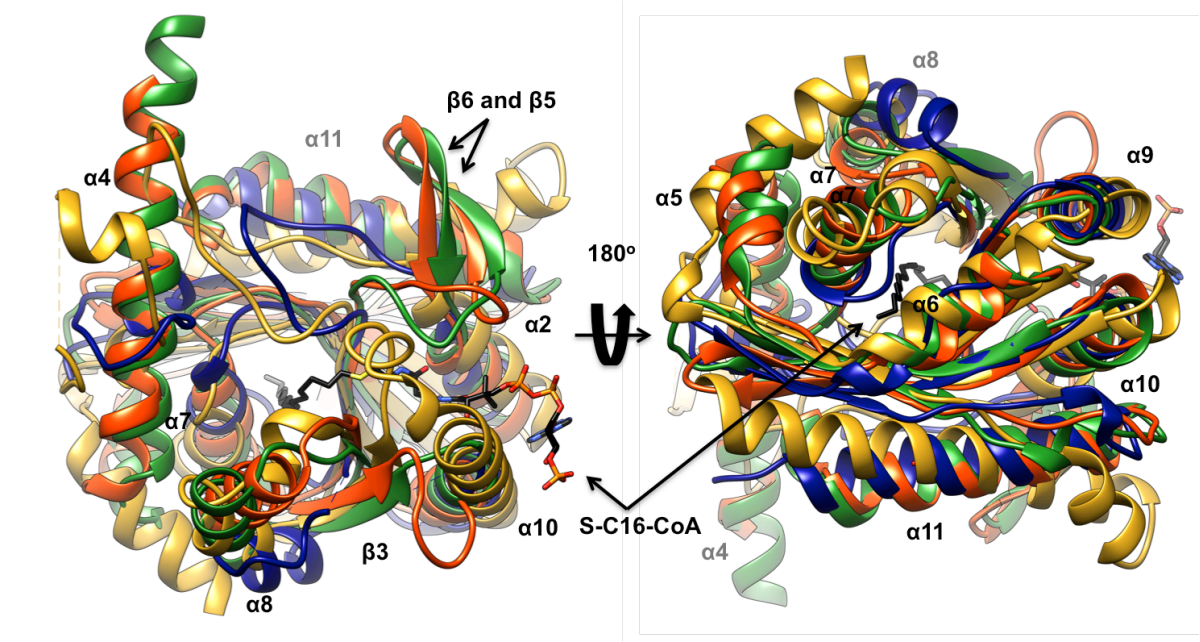



**Figure 78: PatA, *Ab*LpxM, *Tm*PlsC *and* *Cm*GPAT have a common structural core.** Structural superposition of the structural core of PatA (orange), *Ab*LpxM (green), *Tm*PlsC (blue), *Cm*GPAT (yellow) and S-C16-CoA (black).

The superposition of the α/β domain of PatA, *Ab*LpxM, *Cm*GPAT and *Tm*PlsC confirms a common structural catalytic core, however the α-helix motif of the N-terminal domain that those enzymes display is still intriguing. *Ab*LpxM has a long α-helix at the Nt domain[29], *Tm*PlsC has a two-helix motif that is predicted to interact with the lipid bilayer[30], *Cm*GPAT has a four-helix bundle reported with unknown function[28], and the first 50 N-terminal residues of PatA cannot be observed in a crystal structure yet. In that sense, membrane association studies could be performed to shed some light on how these enzymes associate with the lipid bilayer, and how the different number of α-helices affect to the membrane association.

Interestingly, the primary sequence alignment suggests that PatA, *Cm*GPAT, *Ab*LpxM and *Tm*PlsC families of acyltransferases display a conserved aspartate/glutamate residue that participates in the charge relay system with the conserved histidine of the $HX_4D/E$ motif (Block I)[17,32]. This is the case of the two acyltransferases involved in PA biosynthesis *Cm*GPAT and *Tm*PlsC (**Figure 79 a**). In contrast, according to the crystal structures *Ab*LpxM and PatA seem to use a glutamate/aspartate residue located in Block III (**Figure 79 b**), suggesting a divergence among this family of acyltransferases. The structural divergence of the carboxylate group acting as a pKa modulator of the catalytic histidine

residue might be due to the requirement of the acyltransferases to accommodate acceptor molecules of different nature, as observed in PatA and *Ab*LpxM.

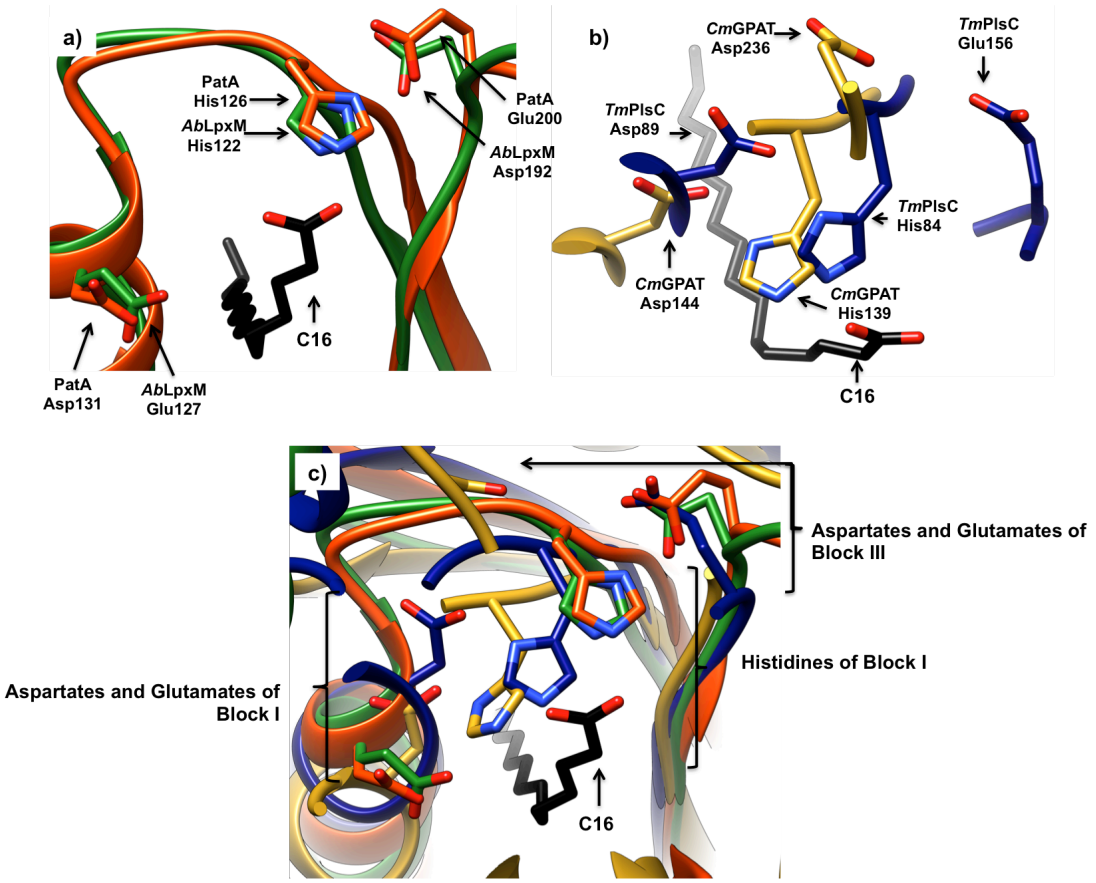



**Figure 79: Divergence in the catalytic mechanism of acyltranferases, PatA is in orange, *Ab*LpxM is in green, *Tm*PlsC is in dark blue and *Cm*GPAT is in yellow. a)** Catalytic mechanism of PatA and *Ab*LpxM, histidine of Block I is activated by PatA glutamate/*Ab*LpxM aspartate of Block III. **b)** Catalytic mechanism of *Tm*PlsC and *Cm*GPAT, histidine of Block I is activated by aspartates of Block I from the motif $HX_4D$. **c)** PatA, *Ab*LpxM, *Tm*PlsC and *Cm*GPAT catalytic mechanisms.

To further confirm the existence of this new structural family of acyltransferases with a common structural core but a divergence in the catalytic mechanism new crystal structures needed. However, this new family of acyltransferases seems to be emerging from the current available data.

Chapter 5

# CONCLUSIONS

# 5. CONCLUSIONS

During recent years major efforts have been made to understand the early steps of the PIM biosynthetic pathway at the molecular level. In this work, we focused in the study of the acyltransferase PatA. PatA is an essential membrane associated acyltransferase involved in the biosynthesis of mycobacterial phosphatidyl-*myo*-inositol mannosides. The enzyme transfers a palmitoyl moiety from palmitoyl-CoA to the 6-position of the mannose ring linked to 2-position of inositol in $PIM_1$/$PIM_2$.

The conclusions of this thesis are the following:

I. After years of research, we have been able to determine the crystal structure of PatA in complex with palmitoyl, a nonhydrolisable analogue of palmitoyl-CoA (S-hexadecyl Coenzime A), the product 6-*O*-palmitoyl-α-D-mannopyranoside, and, one molecule of palmitate and one molecule of free β-D-mannopyransoide.

II. PatA crystallized as a monomer, displaying a high degree of structural flexibility in the N- and C-terminal regions. The central core of PatA consists on a six-stranded β-sheet surrounded by α-helices. A long and open groove that runs parallel to the protein surface contains the active site. The groove also displays a narrow and deep, mostly hydrophobic tunnel that runs perpendicular from its floor to the central core of the protein.

III. The donor binding site is defined by a molecule of palmitic acid carried over from the purification and bound into a hydrophobic tunnel of all PatA. The acyl chain is deeply buried into the hydrophobic pocket, and oriented with the carboxylate group facing the groove and the acyl tail extending into the globular core of the monomer.

IV. The acceptor binding site is defined by a Man*p* residue located within a cavity located at the end of the main groove as well as the product 6-*O*-palmitoyl-α-D-mannopyranoside.

V. Enzyme catalysis is mediated by an unprecedented charge relay system, which markedly diverges from the canonical $HX_4D$ motif. We propose a model in which His126 acts as the general base to abstract a proton from the hydroxyl group at position 6 of the Man*p* ring linked to the 2-position of inositol in $PIM_1$ or $PIM_2$, to facilitate the nucleophilic attack on the thioester of palmitoyl-CoA. The Glu200 gets involved in a charge relay system with His126 and the HO6 atom from the Man*p* moiety, contributing to the appropriate structural arrangement of the imidazole ring of the histidine residue and modulating its pKa to act as a base in the first step and as an acid in the second step, providing protonic assistance to the departing CoA leaving group. The result of the nucleophilic attack is a covalent bond between the mannose ring of $PIM_1$ or $PIM_2$ and the palmitoyl residue.

VI. Despite the fact that acyl transfer is catalysed between the palmitoyl moiety of palmitoyl-CoA and the mannosyl group of $PIM_1$ or $PIM_2$, our experimental data indicate the enzyme displays an absolute necessity of both fatty acid chains of PI in order for the transfer reaction to take place, pointing to a mechanism requiring a lipid-water interface. The fatty acid chains of the acceptor substrate seems to play an important structural role in the generation of a competent active site, assisting/guiding the glycosylated moiety to the appropriate environment in the catalytic centre.

VII. The crystal structure of PatA in the presence of β-D-mannopyranose and palmitate suggest an inhibitory mechanism for the enzyme. These studies establish the mechanistic basis of substrate/membrane recognition and catalysis for an important family of acyltransferases, providing exciting possibilities for inhibitor design and the discovery of chemotherapeutic agents against this major human pathogen.

VIII. The search of structural homologs reveals that *Ab*LpxM, *Cm*GPAT and *Tm*PlsC are homologs of PatA. By the structural comparison of the α/β domain of these enzymes with PatA the existence of a common core can be defined.

IX. The primary sequence alignment suggests that PatA, *Cm*GPAT, *Ab*LpxM and *Tm*PlsC families of acyltransferases display a conserved aspartate/glutamate residue that participates in the charge relay system with the conserved histidine of the $HX_4D/E$ motif (Block I). However, crystal structures of these enzymes reveal that *Cm*GPAT and *Tm*PlsC mechanisms are in accordance with the primary secuence alignments. In contrast, *Ab*LpxM and PatA seem to use a glutamate/aspartate residue located in Block III, suggesting a divergence among this family of acyltransferases.

Chapter 6

# BIBLIOGRAPHY

# 6. BIBLIOGRAPHY


1. Varki, A. *et al. Essentials of Glycobiology*. *Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Print* (1999). doi:10.1016/S0962-8924(00)01855-9
2. Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A Modification Systems in Gram-Negative Bacteria. *Annu. Rev. Biochem.* **76,** 295–329 (2007).
3. Rottig, A. & Steinbuchel, A. Acyltransferases in Bacteria. *Microbiol. Mol. Biol. Rev.* **77,** 277–321 (2013).
4. Forneris, F. Enzymes Without Borders: Mobilizing Substrates, Delivering Products. *Science (80-. ).* **321,** 213–216 (2011).
5. *World Health Organization*. (2016).
6. Zumla, A., Raviglione, M., Hafner, R. & von Reyn, C. F. Tuberculosis. *N. Engl. J. Med.* **368,** 745–755 (2013).
7. Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H. & Jackson, M. The cell envelope glycoconjugates of Mycobacterium tuberculosis. *Crit Rev Biochem Mol Biol* **9238,** 1–39 (2014).
8. Jackson, M. The Mycobacterial Cell Envelope - Lipids. *CSH Perspect. Med.* **83,** 317–317 (2014).
9. Jankute, M., Cox, J. A. G., Harrison, J. & Besra, G. S. Assembly of the Mycobacterial Cell Wall. *Annu. Rev. Microbiol.* **69,** 405–23 (2016).
10. Ishikawa, E., Mori, D. & Yamasaki, S. Recognition of Mycobacterial Lipids by Immune Receptors. *Trends Immunol.* **38,** 66–76 (2017).
11. Sancho-Vaello, E., Albesa-Jové, D., Rodrigo-Unzueta, A. & Guerin, M. E. Structural basis of phosphatidyl-myo-inositol mannosides biosynthesis in mycobacteria. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **16,** S1388-1981 (2016).
12. Guerin, M. E., Korduláková, J., Alzari, P. M., Brennan, P. J. & Jackson, M. Molecular basis of phosphatidyl-myo-inositol mannoside biosynthesis and regulation in mycobacteria. *J. Biol. Chem.* **285,** 33577–83 (2010).
13. Chatterjee, D., Hunter, S. W., McNeil, M. & Brennan, P. J. Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. *J. Biol. Chem.* **267,** 6228–6233 (1992).
14. Khoo, K. H., Dell, A., Morris, H. R., Breman, P. J. & Chatterjee, D. Structural definition of acylated phosphatidylinositol mannosides from Mycobacterium tuberculosis: Definition of a common anchor for lipomannan and lipoarabinomannan. *Glycobiology* **5,** 117–127 (1995).
15. Giganti, D. *et al.* Secondary structure reshuffling modulates glycosyltransferase function at the membrane. *Nat. Chem. Biol.* **11,** 16–18 (2015).
16. Batt, S. M. *et al.* Acceptor substrate discrimination in phosphatidyl-myo-inositol mannoside synthesis: structural and mutational analysis of mannosyltransferase Corynebacterium glutamicum PimB'. *J. Biol. Chem.* **285,** 37741–37752 (2010).
17. Heath, R. J. & Rock, C. O. A conserved histidine is essential for glycerolipid acyltransferase catalysis. *J. Bacteriol.* **180,** 1425–1430 (1998).
18. Lewin, T. M., Wang, P. & Coleman, R. A. Analysis of amino acid motifs diagnostic for the sn-glycerol-3-phosphate acyltransferase reaction. *Biochemistry* **38,** 5764–5771 (1999).
19. Raetz, C. R. H., Reynolds, M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76,** 295–329 (2007).
20. Röttig, A. & Steinbüchel, A. Acyltransferases in Bacteria. doi:10.1128/MMBR.00010-13
21. Albesa-Jové, D. *et al.* Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nat. Commun.* **7,** 10906 (2016).

22. Tersa, M. *et al.* The Molecular Mechanism of Substrate Recognition and Catalysis of the Membrane Acyltransferase PatA from Mycobacteria. *ACS Chem. Biol.* acschembio.7b00578 (2017). doi:10.1021/acschembio.7b00578

23. Korduláková, J. *et al.* Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of Mycobacterium species. *J. Biol. Chem.* **278,** 36285–36295 (2003).

24. Svetlíková, Z., Baráth, P., Jackson, M., Korduláková, J. & Mikušová, K. Purification and characterization of the acyltransferase involved in biosynthesis of the major mycobacterial cell envelope glycolipid - Monoacylated phosphatidylinositol dimannoside. *Protein Expr. Purif.* **100,** 33–9 (2014).

25. Albesa-Jové, D., Giganti, D., Jackson, M., Alzari, P. M. & Guerin, M. E. Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology* **24,** 108–24 (2014).

26. Blobel, G. Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1496–500 (1980).

27. Turnbull, A. P. *et al.* Analysis of the structure, substrate specificity, and mechanism of squash glycerol-3-phosphate (1)-acyltransferase. *Structure* **9,** 347–353 (2001).

28. Tamada, T. *et al.* Substrate recognition and selectivity of plant glycerol-3-phosphate acyltransferases (GPATs) from Cucurbita moscata and Spinacea oleracea. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 13–21 (2004).

29. Dovala, D. *et al.* Structure-guided enzymology of the lipid A acyltransferase LpxM reveals a dual activity mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **Published,** E6064–E6071 (2016).

30. Robertson, R. M. *et al.* A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. *Nat. Struct. Mol. Biol.* **24,** 666–671 (2017).

31. Yao, J. & Rock, C. O. Phosphatidic Acid Synthesis in Bacteria. *Biochim. Biophys. Acta* **1831,** 495–502 (2013).

32. Neuwald, A. F. Barth syndrome may be due to an acyltransferase deficiency. *Curr. Biol.* **7,** 465–466 (1997).

33. Cronan, J. E. Bacterial Membrane Lipids: Where Do We Stand? *Annu. Rev. Microbiol.* **57,** 203–224 (2003).

34. White, S. W., Zheng, J., Zhang, Y.-M. & Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **74,** 791–831 (2005).

35. Black, P. N. & DiRusso, C. C. Transmembrane movement of exogenous long-chain fatty acids: proteins, enzymes, and vectorial esterification. *Microbiol. Mol. Biol. Rev.* **67,** 454–72, table of contents (2003).

36. Clementz, T., Bednarski, J. J. & Raetz, C. R. H. Function of the htrB high temperature requirement gene of Escherichia coli in the acylation of lipid A: HtrB catalyzed incorporation of laurate. *J. Biol. Chem.* **271,** 12095–12102 (1996).

37. Clementz, T., Zhou, Z. & Raetz, C. R. H. Function of the Escherichia coli msbB gene, a multicopy suppressor of htrB knockouts, in the acylation of lipid A. Acylation by MsbB follows laurate incorporation by HtrB. *J. Biol. Chem.* **272,** 10353–10360 (1997).

38. Rottig, A. & Steinbuchel, A. Acyltransferases in bacteria. *Microbiol Mol Biol Rev* **77,** 277–321 (2013).

39. Athenstaedt, K. & Daum, È. Phosphatidic acid , a key intermediate in lipid metabolism. *Eur. J. Biochem.* **266,** 1–16 (1999).

40. Bell, R. M. Mutants of Escherichia coli defective in membrane phospholipid synthesis: macromolecular synthesis in an sn-glycerol 3-phosphate acyltransferase Km mutant. *J. Bacteriol.* **117,** 1065–1076 (1974).

41. Nilsson, R. P., Beijer, L. & Rutberg, B. The glpP and glpF genes of the glycerol regulon in Bacillus subtilis. *J. Gen. Microbiol.* **139,** 349–359 (1993).

42. Kito, M. & Pizer, L. I. Dehydrogenase from Escherichia coli Purification and Regulatory Properties S-Phosphate Dehydrogenase of the Biosynthetic from Escherichia coli. *J. Biol. Chem.* **244,** 3316–3323 (1968).

43. Ray, T. K. & Cronan, J. E. Acylation of glycerol 3-phosphate is the sole pathway of de novo phospholipid synthesis in Escherichia coli. *J. Bacteriol.* **169,** 2896–2898 (1987).

44. Wilkison, W. O. & Bell, R. M. sn-Glycerol-3-phosphate acyltransferase from Escherichia coli. *Biochim. Biophys. Acta - Lipids Lipid Metab.* **1348,** 3–9 (1997).

45. Coleman, J. Characterization of Escherichia coli cells deficient in 1-acyl-sn-glycerol-3-phosphate acyltransferase activity. *J. Biol. Chem.* **265,** 17215–21 (1990).

46. Polgár, L. The catalytic triad of serine peptidases. *Cell. Mol. Life Sci.* **62,** 2161–2172 (2005).

47. Rock, C., Goelzg, S. E. & Grant, A. Phospholipid synthesis in Escherichia coli. Characteristics of fatty acid transfer from acyl-acyl carrier protein to sn-glycerol-3-phosphate. *J. Biol. Chem.* **258,** 462–464 (1981).

48. Lightner, V. A. *et al.* Membrane phospholipid synthesis in Escherichia coli: Cloning of a structural gene (plsB) of the sn-glycerol-3-phosphate acyltransferase. *J. Biol. Chem.* **255,** 9413–9420 (1980).

49. Parsons, J. B. & Rock, C. O. Bacterial lipids: Metabolism and membrane homeostasis. *Prog. Lipid Res.* **52,** 249–276 (2013).

50. Lu, Y. J. *et al.* Acyl-Phosphates Initiate Membrane Phospholipid Synthesis in Gram-Positive Pathogens. *Mol. Cell* **23,** 765–772 (2006).

51. Paoletti, L., Lu, Y. J., Schujman, G. E., De Mendoza, D. & Rock, C. O. Coupling of fatty acid and phospholipid synthesis in Bacillus subtilis. *J. Bacteriol.* **189,** 5816–5824 (2007).

52. Schujman, G. E. & de Mendoza, D. Solving an old puzzle in phospholipid biosynthesis. *Nat. Chem. Biol.* **2,** 573–574 (2006).

53. Paoletti, L., Lu, Y.-J., Schujman, G. E., de Mendoza, D. & Rock, C. O. Coupling of fatty acid and phospholipid synthesis in Bacillus subtilis. *J. Bacteriol.* **189,** 5816–24 (2007).

54. Ganesh Bhat, B. *et al.* Rat sn-glycerol-3-phosphate acyltransferase: Molecular cloning and characterization of the cDNA and expressed protein. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1439,** 415–423 (1999).

55. Hanke, C., Wolter, F. P., Coleman, J., Peterek, G. & Frentzen, M. A Plant Acyltransferase Involved in Triacylglycerol Biosynthesis Complements an Escherichia Coli sn-1-acylglycerol-3-phosphate Acyltransferase Mutant. *Eur. J. Biochem.* **232,** 806–810 (1995).

56. Zhang, Y.-M. & Rock, C. O. Acyltransferases in bacterial glycerophospholipid synthesis. *J. Lipid Res.* **49,** 1867–1874 (2008).

57. Murata, N. & Tasaka, Y. Glycerol-3-phosphate acyltransferase in plants. *Biochim. Biophys. Acta - Lipids Lipid Metab.* **1348,** 10–16 (1997).

58. Hayman, M. W., Fawcett, T. & Slabas, A. R. Kinetic mechanism and order of substrate binding for sn-glycerol-3-phosphate acyltransferase from squash (Cucurbita moschata). *FEBS Lett.* **514,** 281–284 (2002).

59. Slabas, A. R. *et al.* Squash glycerol-3-phosphate (1)-acyltransferase: Alteration of substrate selectivity and identification of arginine and lysine residues important in catalytic activity. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M206429200

60. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102,** 4501–4523 (2002).

61. Bellizzi, J. J. *et al.* The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 4573–4578 (2000).

62. Zhang, Y.-M. & Rock, C. O. Thematic review series: Glycerolipids. Acyltransferases in bacterial glycerophospholipid synthesis. *J. Lipid Res.* **49,** 1867–74 (2008).

63. Shindou, H., Hishikawa, D., Harayama, T., Eto, M. & Shimizu, T. Generation of membrane diversity

by lysophospholipid acyltransferases. *J. Biochem.* **154,** 21–28 (2013).

64. Yamashita, A. *et al.* Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms. *Prog. Lipid Res.* **53,** 18–81 (2014).

65. Agarwal, A. K. *et al.* AGPAT2 is mutated in congenital generalized lipodystrophy linked to chromosome 9q34. *Nat. Genet.* **31,** 21–23 (2002).

66. Emptage, R. P., Daughtry, K. D., Pemble, C. W. & Raetz, C. R. H. Crystal structure of LpxK, the 4'-kinase of lipid A biosynthesis and atypical P-loop kinase functioning at the membrane interface. *Proc. Natl. Acad. Sci.* **109,** 12956–12961 (2012).

67. Viklund, H., Granseth, E. & Elofsson, A. Structural Classification and Prediction of Reentrant Regions in alpha-Helical Transmembrane Proteins: Application to Complete Genomes. *J. Mol. Biol.* **361,** 591–603 (2006).

68. Yamashita, A. *et al.* Topology of acyltransferase motifs and substrate specificity and accessibility in 1-acyl-sn-glycero-3-phosphate acyltransferase 1. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1771,** 1202–1215 (2007).

69. Lu, Y. J., Zhang, F., Grimes, K. D., Lee, R. E. & Rock, C. O. Topology and active site of PlsY: The bacterial acylphosphate:glycerol-3- phosphate acyltransferase. *J. Biol. Chem.* **282,** 11339–11346 (2007).

70. Lu, Y. J. *et al.* Acyl-Phosphates Initiate Membrane Phospholipid Synthesis in Gram-Positive Pathogens. *Mol. Cell* **23,** 765–772 (2006).

71. Coleman, J. Characterization of the Escherichia coli gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (plsC). *Mol. Gen. Genet. MGG* **232,** 295–303 (1992).

72. Parsons, J. B. *et al.* Identification of a two-component fatty acid kinase responsible for host fatty acid incorporation by Staphylococcus aureus. *Proc. Natl. Acad. Sci.* **111,** 10532–10537 (2014).

73. Parsons, J. & Frank, M. Metabolic basis for the differential susceptibility of Gram-positive pathogens to fatty acid synthesis inhibitors. *Proc.* **108,** 15378–15383 (2011).

74. Szabo, F. K. & Hoffman, G. E. Crystal structure of fatty acid/phospholipid synthesis protein PlsX from Enterococcus faecalis. *J Struct Funct Genomics.* **37,** 62–70 (2012).

75. Badger, J. *et al.* Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins Struct. Funct. Genet.* **60,** 787–796 (2005).

76. Kobayashi, K. *et al.* Essential Bacillus subtilis genes. *Proc. Natl. Acad. Sci.* **100,** 4678–4683 (2003).

77. Li, Z. *et al.* Structural insights into the committed step of bacterial phospholipid biosynthesis. *Nat. Commun.* **8,** 1691 (2017).

78. Silhavy, T., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2,** 1–16 (2010).

79. Fairman, J. W., Noinaj, N. & Buchanan, S. K. The Structural Biology of β-barrel membrane proteins: a summary of recent reports. *Curr. Opin. Struct. Biol.* **21,** 523–531 (2012).

80. Schulz, G. E. The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta* **1565,** 308–317 (2002).

81. Zeth, K. & Thein, M. Porins in prokaryotes and eukaryotes: common themes and variations. *Biochem. J.* **431,** 13–22 (2010).

82. Sankaran, K. & Wu, H. C. Lipid modification of bacterial prolipoprotein. *J. Biol. Chem.* **269,** 19701–6 (1994).

83. Konovalova, A. & Silhavy, T. J. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** 20150030 (2015).

84. Whitfield, C. & Trent, M. S. Biosynthesis and Export of Bacterial Lipopolysaccharides. *Annu. Rev. Biochem.* **83,** 99–128 (2014).

85. Nikaio, H. & Vaara, M. Molecular basis of bacterial outer membrane permeability. *Microbiol. Rev.* **49,** 1–32 (1985).

86. King, J. D., Kocíncová, D., Westman, E. L. & Lam, J. S. Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa. *Innate Immun.* **15,** 261–312 (2009).

87. Raetz, C. R. H. Molecular genetics of membrane phospholipid synthesis. *Annu. Rev. Genet.* **20,** 253–295 (1986).

88. Raetz, C. & C, W. Lipopolysaccharide Endotoxins. *Annu Rev Biochem.* **71,** 635–700 (2002).

89. Raetz, C. R. H. *et al.* Discovery of new biosynthetic pathways: the lipid A story. *J. Lipid Res.* **50 Suppl,** S103–S108 (2009).

90. Wang, X. & Quinn, P. J. Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog. Lipid Res.* **49,** 97–107 (2010).

91. Holst, J. J. The Physiology of Glucagon-like Peptide 1. *Physiol. Rev.* 1409–1439 (2007). doi:10.1152/physrev.00034.2006.

92. Griffiss, J. M. *et al.* The immunochemistry of neisserial LOS. *Antonie Van Leeuwenhoek* **53,** 501–507 (1987).

93. Peppler, M. S. Two physically and serologically distinct lipopolysaccharide profiles in strains of Bordetella pertussis and their phenotype variants. *Infect. Immun.* **43,** 224–232 (1984).

94. Trent, M. S. Biosynthesis, transport, and modification of lipid A. *Biochem. Cell Biol.* **82,** 71–86 (2004).

95. GALANOS, C. *et al.* Synthetic and natural Escherichia coli free lipid A express identical endotoxic activities. *Eur. J. Biochem.* **148,** 1–5 (1985).

96. Barb, A. W., Mcclerren, A. L., Snehelatha, K., Reynolds, C. M. & Raetz, C. R. H. Inhibition of Lipid A Biosynthesis as the Primary Mechanism of CHIR-090 Antibiotic Activity in Escherichia coli. *Biochemistry* **46,** 3793–3802 (2007).

97. Jackman, J. E. *et al.* Antibacterial Agents That Target Lipid A Biosynthesis in Gram-negative Bacteria. *J. Biol. Chem.* **275,** 11002–11009 (2000).

98. Onishi, H. R. *et al.* Antibacterial Agents That Inhibit Lipid A Biosynthesis. *Science (80-. ).* **274,** 980–982 (1996).

99. Sperandeo, P., Martorana, A. M. & Polissi, A. Lipopolysaccharide biogenesis and transport at the outer membrane of Gram-negative bacteria. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* (2016). doi:10.1016/j.bbalip.2016.10.006

100. Emiola, A., George, J. & Andrews, S. S. A complete pathway model for lipid a biosynthesis in Escherichia coli. *PLoS One* **10,** 1–28 (2015).

101. Anderson, M. S. *et al.* UDP-N-acetylglucosamine acyltransferase of Escherichia coli: The first step of endotoxin biosynthesis is thermodynamically unfavorable. *J. Biol. Chem.* **268,** 19858–19865 (1993).

102. Kelly, T. M., Stachula, S. A., Raetz, C. R. H. & Anderson, M. S. The firA gene of Escherichia coli encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase: The third step of endotoxin biosynthesis. *J. Biol. Chem.* **268,** 19866–19874 (1993).

103. Young, K. *et al.* The envA Permeability / Cell Division Gene of Escherichia coli Encodes the Second Enzyme of Lipid A Biosynthesis. *J. Biol. Chem.* **270,** 30384–30391 (2000).

104. Radika, K. & Raetz, C. R. Purification and properties of lipid A disaccharide synthase of Escherichia coli. *J. Biol. Chem.* **263,** 14859–67 (1988).

105. Babinski, K. J., Ribeiro, A. a & Raetz, C. R. H. The Escherichia coli gene encoding the UDP-2,3-diacylglucosamine pyrophosphatase of lipid A biosynthesis. *J. Biol. Chem.* **277,** 25937–25946 (2002).

106. Babinski, K. J., Kanjilal, S. J. & Raetz, C. R. H. Accumulation of the lipid A precursor UDP-2,3-diacylglucosamine in an Escherichia coli mutant lacking the lpxH gene. *J. Biol. Chem.* **277,** 25947–25956 (2002).

107. Garrett, T. A., Kadrmas, J. L. & Raetz, C. R. H. Identification of the Gene Encoding the Escherichia coli Lipid A 4'-Kinase. *J. Biol. Chem.* **272,** 21855–21864 (1997).

108. Carty, S. M., Sreekumar, K. R. & Raetz, C. R. H. Effect of Cold Shock on Lipid A Biosynthesis in Escherichia coli. **274,** 9677–9685 (1999).

109. Xiao, X., Sankaranarayanan, K. & Khosla, C. Biosynthesis and structure–activity relationships of the lipid a family of glycolipids. *Curr. Opin. Chem. Biol.* **40,** 127–137 (2017).

110. Needham, B. D. & Trent, M. S. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat. Rev. Microbiol.* **11,** 467–481 (2013).

111. Lee, H., Hsu, F., Turk, J. & Groisman, E. A. The PmrA-Regulated pmrC Gene Mediates Phosphoethanolamine Modification of Lipid A and Polymyxin Resistance in Salmonella enterica. *J. Bacteriol.* **186,** 4124–4133 (2004).

112. Bauman, D. E. & Davis, C. L. Regulation of Lipid A Modifications by Salmonella typhimurium Virulence Genes phoP-phoQ. *Science (80-. ).* **276,** 496–509 (1997).

113. Guo, L. *et al.* Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* **95,** 189–198 (1998).

114. Gibbons, H. S., Lin, S., Cotter, R. J. & Raetz, C. R. H. Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in Salmonella typhimurium lipid A: Function of LpxO, a new Fe2+/??-ketoglutarate-dependent dioxygenase homologue. *J. Biol. Chem.* **275,** 32940–32949 (2000).

115. Nizet, V. Antimicrobial peptide resistance mechanisms of human bacterial pathogens. *Curr. Issues Mol. Biol.* **8,** 11–26 (2006).

116. Maeshima, N. & Fernandez, R. C. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. *Front. Cell. Infect. Microbiol.* **3,** 1–13 (2013).

117. Wyckoff, T. J. O., Lin, S., Cotter, R. J., Dotson, G. D. & Raetz, C. R. H. Hydrocarbon Rulers in UDP- N -acetylglucosamine Acyltransferases *. *J. Biol. Chem.* **273,** 32369–32372 (1998).

118. Coleman, J. & Raetz, C. R. H. 1st Committed Step of Lipid-a Biosynthesis in Escherichia-Coli - Sequence of the Lpxa Gene. *J Bacteriol* **170,** 1268–1274 (1988).

119. Galloway, S. M. & Raetz, C. R. H. A Mutant of Escherichia coli Defective in the First Step of Endotoxin Biosynthesis. *J. Biol. Chem.* **265,** 6394–6402 (1990).

120. Zhou, P. & Zhao, J. Structure, inhibition, and regulation of essential lipid A enzymes. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **16,** 1388–1981 (2016).

121. Robins, L. I., Williams, A. H. & Raetz, C. R. H. Structural Basis for the Sugar Nucleotide and Acyl Chain Selectivity of Leptospira interrogans LpxA. *Biochemistry* **48,** 6191–6201 (2009).

122. Williams, A. H. & Raetz, C. R. H. Structural basis for the acyl chain selectivity and mechanism of UDP-N-acetylglucosamine acyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 13543–50 (2007).

123. Smith, E. W. *et al.* Structures of Pseudomonas aeruginosa LpxA Reveal the Basis for Its Substrate Selectivity. *Biochemistry* **54,** 5937–5948 (2015).

124. Williams, A. H., Immormino, R. M., Gewirth, D. T. & Raetz, C. R. H. Structure of UDP-N-acetylglucosamine acyltransferase with a bound antibacterial pentadecapeptide. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10877–10882 (2006).

125. Dotson, G. D. & Jenkins, R. J. Dual Targeting Antibacterial Peptide Inhibitor of Early Lipid A Biosynthesis. *ACS Chem Biol* (2012). doi:10.1021/cb300094a

126. Benson, R. E., Gottlin, E. B., Christensen, D. J. & Hamilton, P. T. Intracellular expression of peptide fusions for demonstration of protein essentiality in bacteria. *Antimicrob. Agents Chemother.* **47,** 2875–2881 (2003).

127. Jenkins, R. J., Heslip, K. A., Meagher, J. L., Stuckey, J. A. & Dotson, G. D. Structural basis for the recognition of peptide RJPXD33 by acyltransferases in lipid a biosynthesis. *J. Biol. Chem.* **289,** 15527–15535 (2014).

128. Li, C., Guan, Z., Liu, D. & Raetz, C. R. H. Pathway for lipid A biosynthesis in Arabidopsis thaliana resembling that of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11387–11392 (2011).

129. Dicker, I. B. & Seetharam, S. Cloning and nucleotide sequence of the firA gene and the firA200(Ts) allele from Escherichia coli. *J. Bacteriol.* **173,** 334–344 (1991).

130. Opiyo, S. O., Pardy, R. L., Moriyama, H. & Moriyama, E. N. Evolution of the Kdo2-lipid A biosynthesis in bacteria. *BMC Evol. Biol.* **10,** 362 (2010).

131. Buetow, L., Smith, T. K., Dawson, A., Fyffe, S. & Hunter, W. N. Structure and reactivity of LpxD, the N-acyltransferase of lipid A biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4321–4326 (2007).

132. Bartling, C. M. & Raetz, C. R. H. Crystal Structure and Acyl Chain Selectivity of Escherichia coli LpxD, the N-Acyltransferase of Lipid A Biosynthesis. *Biochemistry* **48,** 8672–8683 (2009).

133. Badger, J. *et al.* The structure of LpxD from Pseudomonas aeruginosa at 1.3 A resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 749–752 (2011).

134. Badger, J. *et al.* Structure determination of LpxD from the lipopolysaccharide-synthesis pathway of Acinetobacter baumannii. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69,** 6–9 (2013).

135. Bartling, C. M. & Raetz, C. R. H. Steady-State Kinetics and Mechanism of LpxD, the N-Acyltransferase of Lipid A Biosynthesis. *Biochemistry* **47,** 5290–5302 (2008).

136. Masoudi, A., Raetz, C. R. H., Zhou, P. & Pemble Iv, C. W. Chasing Acyl-Carrier-Protein Through a Catalytic Cycle of Lipid A Production. *Nature* **505,** 422–426 (2014).

137. Vorachek-Warren, M. K., Ramirez, S., Cotter, R. J. & Raetz, C. R. H. A triple mutant of Escherichia coli lacking secondary acyl chains on lipid A. *J. Biol. Chem.* **277,** 14194–14205 (2002).

138. Six, D. A., Carty, S. M., Guan, Z. & Raetz, C. R. H. Purification and Mutagenesis of LpxL, the Lauroyltransferase of Escherichia coli Lipid A Biosynthesis. *Biochemistry* **47,** 8623–8637 (2008).

139. Tran, A. X. *et al.* Resistance to the antimicrobial peptide polymyxin requires myristoylation of Escherichia coli and Salmonella typhimurium lipid A. *J. Biol. Chem.* **280,** 28186–28194 (2005).

140. Tzeng, Y. L., Datta, A., Kumar Kolli, V., Carlson, R. W. & Stephens, D. S. Endotoxin of Neisseria meningitidis composed only of intact lipid A: Inactivation of the meningococcal 3-deoxy-D-manno-octulosonic acid transferase. *J. Bacteriol.* **184,** 2379–2388 (2002).

141. Goldman, R. C., Doran, C. C., Kadam, S. K. & Capobianco, J. O. Lipid A precursor from Pseudomonas aeruginosa is completely acylated prior to addition of 3-deoxy-D-manno-octulosonate. *J. Biol. Chem.* **263,** 5217–5223 (1988).

142. Mohan, S. & Raetz, C. R. H. Endotoxin biosynthesis in Pseudomonas aeruginosa: Enzymatic incorporation of laurate before 3-deoxy-D-manno-octulosonate. *J. Bacteriol.* **176,** 6944–6951 (1994).

143. Boll, J. M. *et al.* Reinforcing lipid a acylation on the cell surface of acinetobacter baumannii promotes cationic antimicrobial peptide resistance and desiccation survival. *MBio* **6,** 1–11 (2015).

144. Bishop, R. E. The lipid A palmitoyltransferase PagP: Molecular mechanisms and role in bacterial pathogenesis. *Mol. Microbiol.* **57,** 900–912 (2005).

145. Bishop, R. E. *et al.* Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. *EMBO J.* **19,** 5071–5080 (2000).

146. Brozek, K. A., Bulawa, C. E. & Raetz, C. R. Biosynthesis of lipid A precursors in Escherichia coli. A membrane-bound enzyme that transfers a palmitoyl residue from a glycerophospholipid to lipid X. *J. Biol. Chem.* **262,** 5170–5179 (1987).

147. Cuesta-seijo, J. A. *et al.* PagP Crystallized from SDS/Cosolvent Reveals the Route for Phospholipid Access to the Hydrocarbon Ruler. *Structure* **18,** 1210–1219 (2010).

148. Hwang, P. M. *et al.* Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 13560–5 (2002).

149. Ahn, V. E. *et al.* A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J.* **23,** 2931–2941 (2004).

150. Hwang, P. M., Bishop, R. E. & Kay, L. E. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9618–23 (2004).

151. Bishop, R. E. Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim. Biophys. Acta* **1778,** 1881–1896 (2008).

152. Jia, W. *et al.* Lipid trafficking controls endotoxin acylation in outer membranes of escherichia coli. *J. Biol. Chem.* **279,** 44966–44975 (2004).

153. Smith, A. E. *et al.* PagP Activation in the Outer Membrane Triggers R3 Core Oligosaccharide Truncation in the Cytoplasm of Escherichia coli O157:H7. *J. Biol. Chem.* **283,** 4332–4343 (2008).

154. M. Adil, K. & Bishop, R. E. Molecular Mechanism for Lateral Lipid Diffusion between the Outer Membrane External Leaflet and a β -Barrel Hydrocarbon Ruler. *Biochemistry* **48,** 9745–9756 (2009).

155. Evanics, F., Hwang, P. M., Cheng, Y., Kay, L. E. & Prosser, R. S. Topology of an outer-membrane enzyme: Measuring oxygen and water contacts in solution NMR studies of PagP. *J. Am. Chem. Soc.* **128,** 8256–8264 (2006).

156. Khan, M. A. *et al.* Gauging a Hydrocarbon Ruler by an Intrinsic Exciton Probe. *Biochemistry* **46,** 4565–4579 (2007).

157. Zelmann, R. *et al.* Inscribing the Perimeter of the PagP Hydrocarbon Ruler by Site- Specific Chemical Alkylation. *Biochemistry* **123,** 106–116 (2010).

158. Preston, A. *et al.* Bordetella bronchiseptica PagP is a Bvg-regulated lipid A palmitoyl transferase that is required for persistent colonization of the mouse respiratory tract. *Mol. Microbiol.* **48,** 725–736 (2003).

159. Pilione, M. R., Pishko, E. J., Preston, A., Maskell, D. J. & Harvill, E. T. pagP Is Required for Resistance to Antibody-Mediated Complement Lysis during Bordetella bronchiseptica Respiratory Infection. *Infect. Immun.* **72,** 2837–2842 (2004).

160. Robey, M., Connell, W. O. & Cianciotto, N. P. Identification of *Legionella pneumophila rcp*, a pagP-like gene that confers resistance to cationic antimicrobial peptides and promotes intracellular infection. *Infect. Immun.* **69,** 4276–4286 (2001).

161. Muroi, M., Ohnishi, T. & Tanamoto, K. I. MD-2, a novel accessory molecule, is involved in species-specific actions of Salmonella lipid A. *Infect. Immun.* **70,** 3546–3550 (2002).

162. Tanamoto, K. & Azumi, S. Salmonella-type heptaacylated lipid A is inactive and acts as an antagonist of lipopolysaccharide action on human line cells. *J. Immunol.* **164,** 3149–56 (2000).

163. Kawasaki, K., Ernst, R. K. & Miller, S. I. 3-O-Deacylation of Lipid A by PagL, a PhoP/PhoQ-regulated Deacylase of Salmonella typhimurium, Modulates Signaling through Toll-like Receptor 4. *J. Biol. Chem.* **279,** 20044–20048 (2004).

164. Murata, T., Tseng, W., Guina, T., Miller, S. I. & Nikaido, H. PhoPQ-mediated regulation produces a more robust permeability barrier in the outer membrane of Salmonella enterica serovar typhimurium. *J. Bacteriol.* **189,** 7213–7222 (2007).

165. Bader, M. W. *et al.* Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell* **122,** 461–472 (2005).

166. Mary Jackson, M. R. M. & P. J. B. Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. *Futur. Microbiol.* **8,** 855–875 (2013).

167. Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H. & Jackson, M. The cell envelope glycoconjugates of Mycobacterium tuberculosis. *Crit Rev Biochem Mol Biol* **49,** 361–399 (2014).

168. Jankute, M., Cox, J. A. G., Harrison, J. & Besra, G. S. Assembly of the Mycobacterial Cell Wall. *Annu. Rev. Microbiol.* **69,** 405–423 (2015).

169. Bansal-Mutalik, R. & Nikaido, H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc. Nat. Acad. Sci. U. S. A.* **111,** 4958–4963 (2014).

170. Guerin, M. E., Korduláková, J., Alzari, P. M., Brennan, P. J. & Jackson, M. Molecular Basis of Phosphatidyl-myo-inositol Mannoside Biosynthesis and Regulation in Mycobacteria. *Journal of Biological Chemistry* **285,** 33577–33583 (2010).

171. Bansal-Mutalik, R. & Nikaido, H. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4958–4963 (2014).

172. Jackson, M., Mcneil, M. R. & Brennan, P. J. Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. *Future Microbiol.* **8,** 1–32 (2013).

173. Crick, D. C., Mahapatra, S. & Brennan, P. J. Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*. *Glycobiology* **11,** 107R–118R (2001).

174. Alderwick, L. J., Harrison, J., Lloyd, G. S. & Birch, H. L. The Mycobacterial Cell Wall--Peptidoglycan and Arabinogalactan. *Cold Spring Harb. Perspect. Med.* **5,** a021113 (2015).

175. Heijenoort, J. v. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* **11,** 25R–36R (2001).

176. Raymond, J. B., Mahapatra, S., Crick, D. C. & Pavelka, M. S. Identification of the namH gene, encoding the hydroxylase responsible for the N-glycolylation of the mycobacterial peptidoglycan. *J. Biol. Chem.* **280,** 326–333 (2005).

177. Kumar, P. *et al.* Meropenem Inhibits D,D-Carboxipeptidase Activity In Mycobacterium Tuberculosis. *Mol. Microbiol.* **86,** 367–381 (2012).

178. Crick, D. C., Pavelka Jr., M. S. & Mahapatra, S. Genetics of Peptidoglycan Biosynthesis. *Microbiol. Spectr.* **2,** 1–20 (2014).

179. McNeil, M., Daffe, M. & Brennan, P. J. Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. *J. Biol. Chem.* **265,** 18200–18206 (1990).

180. Daffé, M., Brennan, P. J. & McNeil, M. Predominat Structural features of The Cell wall arabinogalactan of Mycobacterium fragments by gas chromatography / mass Predominant Structural Features of the Cell Wall Arabinogalactan Mycobacterium tuberculosis as Revealed through Characterization Oligog. *J. Antimicrob. Chemother.* **265,** 6734–6743 (1990).

181. Besra, G. S. *et al.* A New Interpretation of the Structure of the Mycolyl—Arabinogalactan Complex of Mycobacterium Tuberculosis as Revealed Through Characterization of Oligoglycosylalditol Fragments by Fast-Atom Bombardment Mass Spectrometry and 1H Nuclear Magnetic Resonance . *Biochemistry* **34,** 4257–4266 (1995).

182. Marrakchi, H., Lanéelle, M. A. & Daffé, M. Mycolic acids: Structures, biosynthesis, and beyond. *Chemistry and Biology* (2014). doi:10.1016/j.chembiol.2013.11.011

183. Nataraj, V. *et al.* Mycolic acids: Deciphering and targeting the Achilles' heel of the tubercle bacillus. *Mol. Microbiol.* **98,** 7–16 (2015).

184. Quémard, A. New Insights into the Mycolate-Containing Compound Biosynthesis and Transport in Mycobacteria. *Trends Microbiol.* **24,** 725–738 (2016).

185. Ortalo-Magné, A. *et al.* Identification of the surface-exposed lipids on the cell envelopes of Mycobacterium tuberculosis and other mycobacterial species. *J. Bacteriol.* **178,** 456–461 (1996).

186. Pitarquea, S. *et al.* The immunomodulatory lipoglycans, lipoarabinomannan and lipomannan, are exposed at the mycobacterial cell surface. *Tuberculosis* **88,** 560–565 (2008).

187. Zuber, B. *et al.* Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J. Bacteriol.* **190,** 5672–5680 (2008).

188. Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. M. & Engelhardt, H. Disclosure of the mycobacterial outer membrane: Cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3963–3967 (2008).

189. Jarlier, V. Mycobacterial cell wall: Structure and role in natural resistance to antibiotics. *FEMS*

*Microbiol. Lett.* **123,** 11–18 (1994).

190. Barry, C. E. 3rd, Crick, D. C. & McNeil, M. R. Targeting the formation of the cell wall core of M. tuberculosis. *Infect. Disord. Drug Targets* **7,** 182–202 (2007).

191. Sani, M. *et al.* Direct visualization by Cryo-EM of the mycobacterial capsular layer: A labile structure containing ESX-1-secreted proteins. *PLoS Pathog.* **6,** (2010).

192. Ortalo-Magne, A. *et al.* Molecular composition of the outermost capsular material of the tubercle bacillus. *Microbiology* **141,** 1609–1620 (1995).

193. Lemassu, A. & Daffé, M. Structural features of the exocellular polysaccharides of Mycobacterium tuberculosis. *Biochem. J.* **297 ( Pt 2,** 351–357 (1994).

194. Dinadayala, P. *et al.* Revisiting the structure of the anti-neoplastic glucans of Mycobacterium bovis Bacille Calmette-Guérin: Structural analysis of the extracellular and boiling water extract-derived glucans of the vaccine substrains. *J. Biol. Chem.* **279,** 12369–12378 (2004).

195. Kalscheuer, R. *et al.* Self-poisoning of Mycobacterium tuberculosis by targeting GlgE in an α-glucan pathway. *Nat. Chem. Biol.* **6,** 376–384 (2010).

196. Koliwer-Brandl, H. *et al.* Metabolic Network for the Biosynthesis of Intra- and Extracellular α-Glucans Required for Virulence of Mycobacterium tuberculosis. *PLoS Pathog.* **12,** 1–26 (2016).

197. Bornemann, S. α-Glucan biosynthesis and the GlgE pathway in Mycobacterium tuberculosis. *Biochem. Soc. Trans.* **44,** 68–73 (2016).

198. Abrahams, K. A. & Besra, G. S. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. *Parasitology* 1–18 (2016). doi:10.1017/S0031182016002377

199. Gilleron, M., Nigou, J., Nicolle, D., Quesniaux, V. & Puzo, G. The acylation state of mycobacterial lipomannans modulates innate immunity response through toll-like receptor 2. *Chem. Biol.* **13,** 39–47 (2006).

200. Torrelles, J. B. & Schlesinger, L. S. Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. *Tuberculosis* **90,** 84–93 (2010).

201. Neyrolles, O. & Guilhot, C. Recent advances in deciphering the contribution of Mycobacterium tuberculosis lipids to pathogenesis. *Tuberculosis* **91,** 187–195 (2011).

202. Briken, V., Porcelli, S. A., Besra, G. S. & Kremer, L. Mycobacterial lipoarabinomannan and related lipoglycans: From biogenesis to modulation of the immune response. *Mol. Microbiol.* **53,** 391–403 (2004).

203. Morita, Y. S. *et al.* Inositol lipid metabolism in mycobacteria: Biosynthesis and regulatory mechanisms. *Biochimica et Biophysica Acta - General Subjects* **1810,** 630–641 (2011).

204. Brennan, P. & Ballou, C. E. Biosynthesis of Mannophosphoinositides by Mycobacterium phlei. *J. Biol. Chem.* **243,** 2975–2984 (1967).

205. Gilleron, M. *et al.* Acylation State of the Phosphatidylinositol Mannosides from Mycobacterium bovis Bacillus Calmette Guérin and Ability to Induce Granuloma and Recruit Natural Killer T Cells. *J. Biol. Chem.* **276,** 34896–34904 (2001).

206. Gilleron, M., Quesniaux, V. F. J. & Puzo, G. Acylation state of the phosphatidylinositol hexamannosides from Mycobacterium bovis bacillus Calmette Guérin and Mycobacterium tuberculosis H37Rv and its implication in toll-like receptor response. *J. Biol. Chem.* **278,** 29880–29889 (2003).

207. Sena, C. B. C. *et al.* Controlled expression of branch-forming mannosyltransferase is critical for mycobacterial lipoarabinomannan biosynthesis. *J. Biol. Chem.* **285,** 13326–13336 (2010).

208. Kaur, D. *et al.* A single arabinan chain is attached to the phosphatidylinositol mannosyl core of the major immunomodulatory mycobacterial cell envelope glycoconjugate, lipoarabinomannan. *J. Biol. Chem.* **289,** 30249–30256 (2014).

209. Berg, S., Kaur, D., Jackson, M. & Brennan, P. J. The glycosyltransferases of Mycobacterium

tuberculosis - roles in the synthesis of arabinogalaetan, lipoarahinolrnlannan, and other glyeoeonjugates. *Glycobiology* **17,** 35R–56R (2007).

210. Hill, D. L. & Ballou, C. E. Biosynthesis of mannophospholipids by Mycobacterium phlei. *J. Biol. Chem.* **241,** 895–902 (1966).

211. Kordulakova, J. *et al.* Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species. *J. Biol. Chem.* **278,** 36285–36295 (2003).

212. Guerin, M. E. *et al.* Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from mycobacteria. *J. Biol. Chem.* **282,** 20705–20714 (2007).

213. Guerin, M. E. *et al.* Substrate-induced conformational changes in the essential peripheral membrane-associated mannosyltransferase PimA from mycobacteria. Implications for catalysis. *J. Biol. Chem.* **284,** 21613–21625 (2009).

214. Sassetti, C. M., Boyd, D. H. & Rubin, E. J. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* **48,** 77–84 (2003).

215. Boldrin, F. *et al.* The phosphatidyl-myo-inositol mannosyltransferase PimA is essential for Mycobacterium tuberculosis growth in vitro and in vivo. *J. Bacteriol.* **196,** 3441–3451 (2014).

216. Lea-Smith, D. J. *et al.* Analysis of a new mannosyltransferase required for the synthesis of phosphatidylinositol mannosides and lipoarbinomannan reveals two lipomannan pools in corynebacterineae. *J. Biol. Chem.* **283,** 6773–6782 (2008).

217. Mishra, A. K., Batt, S., Krumbach, K., Eggeling, L. & Besra, G. S. Characterization of the Corynebacterium glutamicum ΔpimB' ΔmgtA double deletion mutant and the role of Mycobacterium tuberculosis orthologues Rv2188c and Rv0557 in glycolipid biosynthesis. *J. Bacteriol.* **191,** 4465–4472 (2009).

218. Larrouy-Maumus, G. *et al.* Cell-envelope remodeling as a determinant of phenotypic antibacterial tolerance in mycobacterium tuberculosis. *ACS Infect. Dis.* **2,** 352–360 (2016).

219. Gilleron, M. *et al.* Lysosomal Lipases PLRP2 and LPLA2 Process Mycobacterial Multi-acylated Lipids and Generate T Cell Stimulatory Antigens. *Cell Chem. Biol.* **23,** 1147–1156 (2016).

220. Kremer, L. *et al.* Characterization of a putative alpha-mannosyltransferase involved in phosphatidylinositol trimannoside biosynthesis in Mycobacterium tuberculosis. *Biochem. J.* **363,** 437–447 (2002).

221. Morita, Y. S., Patterson, J. H., Billman-Jacobe, H. & Mcconville, M. J. Biosynthesis of mycobacterial phosphatidylinositol mannosides. *Biochem. J* **378,** 589–597 (2004).

222. Morita, Y. S. *et al.* PimE is a polyprenol-phosphate-mannose-dependent mannosyltransferase that transfers the fifth mannose of phosphatidylinositol mannoside in mycobacteria. *J. Biol. Chem.* **281,** 25143–25155 (2006).

223. Crellin, P. K. *et al.* Mutations in pimE restore lipoarabinomannan synthesis and growth in a Mycobacterium smegmatis lpqW mutant. *J. Bacteriol.* **190,** 3690–3699 (2008).

224. Morita, Y. S. *et al.* Compartmentalization of lipid biosynthesis in mycobacteria. *J. Biol. Chem.* **280,** 21645–21652 (2005).

225. Hayashi, J. M. *et al.* Spatially distinct and metabolically active membrane domain in mycobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 5400–5 (2016).

226. Daleke, D. L. Phospholipid flippases. *J. Biol. Chem.* **282,** 821–825 (2007).

227. Sanyal, S., Frank, C. G. & Menon, A. K. Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum. *Biochemistry* **47,** 7937–7946 (2008).

228. Korduláková, J. *et al.* Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis: PimA is essential for growth of mycobacteria. *J. Biol. Chem.* **277,** 31335–31344 (2002).

229. Cole, S. T. *et al.* Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393,** 537–544 (1998).

230. Jackson, M., Crick, D. C. & Brennan, P. J. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J. Biol. Chem.* **275,** 30092–30099 (2000).

231. Morii, H. *et al.* Studies of inositol 1-phosphate analogues as inhibitors of the phosphatidylinositol phosphate synthase in mycobacteria. *J. Biochem.* **153,** 257–266 (2013).

232. Clarke, O. B. *et al.* Structural basis for phosphatidylinositol-phosphate biosynthesis. *Nat. Commun.* 8505 (2015). doi:10.1038/ncomms9505

233. Guerin, M. E. *et al.* Substrate-induced conformational changes in the essential peripheral membrane-associated mannosyltransferase PimA from mycobacteria: implications for catalysis. *J. Biol. Chem.* **284,** 21613–25 (2009).

234. Sassetti, C. M. & Rubin, E. J. Genetic requirements for mycobacterial survival during infection. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12989–12994 (2003).

235. Griffin, J. E. *et al.* High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **7,** 1–9 (2011).

236. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181,** 662–666 (1958).

237. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470,** 73–77 (2011).

238. McPherson, A. Introduction to protein crystallization. *Methods* **34,** 254–265 (2004).

239. Bergfors, T. M. Protein crystallization. (2009). doi:10.1016/0022-0248(95)00921-3

240. Luft, J. R. & DeTitta, G. T. A method to produce microseed stock for use in the crystallization of biological macromolecules. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 988–993 (1999).

241. Hassell, A. M. *et al.* Crystallization of protein-ligand complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63,** 72–79 (2006).

242. Rubinson, K. A., Ladner, J. E., Tordova, M. & Gilliland, G. L. Cryosalts: Suppression of ice formation in macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 996–1001 (2000).

243. Kempkes, R., Stofko, E., Lam, K. & Snell, E. H. Glycerol concentrations required for the successful vitrification of cocktail conditions in a high-throughput crystallization screen. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64,** 287–301 (2008).

244. Holyoak, T. *et al.* Malonate: A versatile cryoprotectant and stabilizing solution for salt-grown macromolecular crystals. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59,** 2356–2358 (2003).

245. Owen, R. L., Rudino-Pinera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl. Acad. Sci.* **103,** 4912–4917 (2006).

246. Dauter, Z. Data-collection strategies. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 1703–1717 (1999).

247. Bellamy, H. D., Snell, E. H., Lovelace, J., Pokross, M. & Borgstahl, G. E. O. The high-mosaicity illusion: Revealing the true physical characteristics of macromolecular crystals. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56,** 986–995 (2000).

248. Harp, J. M., Timm, D. E. & Bunick, G. J. Macromolecular crystal annealing: Overcoming increased mosaicity associated with cryocrystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 622–628 (1998).

249. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science (80-. ).* **336,** 1030–1033 (2012).

250. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 325–338 (2010).

251. Brünger, A. T. Free R-Value - A Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature* **355,** 472–475 (1992).

252. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7,** 95–99 (1963).

253. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

254. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 479–485 (2010).

255. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 1002–1011 (2006).

256. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

257. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).

258. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

259. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

260. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68,** 352–367 (2012).

261. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

262. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

263. Rodgers, J. B. Assay of acyl-CoA : monoglyceride acyltransferase from rat small intestine using continuous recording spectrophotometry. **10,** 427–432 (1969).

264. Stadelmaier, A. & Schmidt, R. R. Synthesis of phosphatidylinositol mannosides (PIMs). *J. Chinese Chem. Soc.* **338,** 2557/ 2569 (2003).

265. Dyer, B. S. *et al.* Synthesis and structure of phosphatidylinositol dimannoside. *J. Org. Chem.* **72,** 3282–3288 (2007).

266. Martin, S. F., Josey, J. A., Wong, Y. & Dean, D. W. General Method for the Synthesis of Phospholipid Derivatives of 1,2-O-Diacyl-sn-glycerols. *J. Org. Chem.* **59,** 4805–4820 (1994).

267. Peters, T. Synthesis and Conformational Analysis of Methyl 2-O-(a-~-Mannopyranosyl)-a- D-mannop yranoside. 135–141 (1991).

268. Bender, S. L. & Budhu, R. J. Biomimetic synthesis of enantiomerically pure D-myo-inositol derivatives. *J. Am. Chem. Soc.* **113,** 9883–9885 (1991).

269. Computing, C. & Inc., G. Molecular Operating Environment (MOE). 2013.08. (2015).

270. Gerber, P. R. & Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided. Mol. Des.* **9,** 251–268 (1995).

271. Case, D., Darden, T. & III, T. C. Amber 12 Reference Manual. *Univ. California, San …* (2012).

272. Labute, P. LowModeMD - Implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *J. Chem. Inf. Model.* **50,** 792–800 (2010).

273. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved Protein – Ligand Docking Using GOLD. *Proteins Struct. Funct. Bioinforma.* **623,** 609–623 (2003).

274. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. *J. Mol. Biol.* **267,** 727–748 (1997).

275. Cheng, T., Li, X., Li, Y., Liu, Z. & Wang, R. Comparative assessment of Sscoring Functions on a Diverse Test set. *J. Chem. Inf. Model.* **49,** 1079–93 (2009).

276. Mooij, W. T. M. & Verdonk, M. L. General and targeted statistical potentials for protein-ligand

interactions. *Proteins Struct. Funct. Genet.* **61,** 272–287 (2005).

277. Pearlman, D. A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **91,** 1–41 (1995).

278. Kirschner, K. N. *et al.* GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **29,** 622–655 (2008).

279. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

280. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. a & Case, D. a. Development and testing of a general amber force field. *J. Comput. Chem.* **25,** 1157–1174 (2004).

281. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C., Wang, J., Duke, R., Luo, R., Crowley, M. F., Walker, R., Zhang, W., et al. Amber 11. *Univ. California, San Fr.* (2010).

282. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14,** 33–38 (1996).

283. Car, R. & Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **55,** 2471–2474 (1985).

284. Troullier, N. & Martins, J. L. Efficient pseudopotentials for plane-wave calculations. II. Operators for fast iterative diagonalization. *Phys. Rev. B* **43,** 8861–8869 (1991).

285. Ensing, B., Laio, A., Parrinello, M. & Klein, M. L. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B* **109,** 6676–6687 (2005).

286. Pei, J., Grishin, N. V & Road, F. P. Multiple Sequence Alignment Methods. **1079,** 263–271 (2014).

287. DeLano, W. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **700,** (2002).

288. Brennan, P. & Ballou, E. Biosynthesis of Mannophosphoinositides by Mycobacterium phlei. *J. Biol. Chem.* **243,** 2975–2984 (1968).

289. Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. & Holden, H. M. The Structure of 4-Hydroxybenzoyl-CoA Thioesterase from Arthrobacter sp. strain SU. *J. Biol. Chem.* **278,** 43709–43716 (2003).

290. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10,** 168 (2009).

291. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1,** 826–843 (2011).

292. Ardèvol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **137,** 7528–7547 (2015).

293. Botos, I. & Wlodawer, A. The expanding diversity of serine hydrolases. *Curr. Opin. Struct. Biol.* **17,** 683–690 (2007).

294. Cantu, D. C., Ardevol, A., Rovira, C. & Reilly, P. J. Molecular mechanism of a hotdog-fold Acyl-CoA thioesterase. *Chem. - A Eur. J.* **20,** 9045–9051 (2014).

295. McMahon, H. T. & Gallop, J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438,** 590–596 (2005).

296. Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **438,** 578–580 (2005).

297. Dufrisne, M. B., Petrou, V. I., Clarke, O. B. & Mancia, F. Structural basis for catalysis at the membrane-water interface. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1862,** 1368–1385 (2017).

298. White, S. & Wimley, W. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 319–365 (1999).

299. Whitfield, C. Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli. *Annu. Rev. Biochem.* **75,** 39–68 (2006).

300. Pelletier, M. R. *et al.* Unique structural modifications are present in the lipopolysaccharide from colistin-resistant strains of Acinetobacter baumannii. *Antimicrob. Agents Chemother.* **57,** 4831–4840 (2013).

301. Wendel, A. A., Lewin, T. M. & Coleman, R. A. Glycerol-3-phosphate acyltransferases: Rate limiting enzymes of triacylglycerol biosynthesis. *Biochim. Biophys. Acta* **179,** 501–506 (2009).

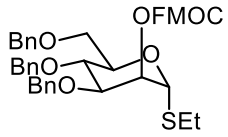
302. Cuesta-Seijo, J. A. *et al.* PagP crystallized from SDS/Cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure* **18,** 1210–1219 (2010).

# ANNEX I

# ANNEX I

Chemical synthesis of $PIM_2$[22].

**Scheme 1**




**$C_{44}H_{44}O_7S$**

M = 716.89 g.mol$^{-1}$

Ethyl 3,4,6-tri-*O*-benzyl-2-*O*-(9-fluorenylmethyloxycarbonyl)-1-thio-

α-D-mannopyranoside

(**2**)

A freshly prepared sodium methoxide solution (30 ml, 0.5 M in methanol) was added to a solution of ethyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-1-thio-α-D-mannopyranoside (13.90 g, 0.026 mol) in methanol (100 ml) at 0°C. The reaction mixture was stirred at room temperature and the progress of the reaction was monitored by TLC. The methanol was removed *in vacuo* and the residue was extracted by dichloromethane. The organic phase was washed with water, dried over $MgSO_4$, filtered and concentrated *in vacuo* to give a yellow oil which was used without further purification. To a solution of the oil in anhydrous dichloromethane (60 ml) was added pyridine (4 ml) followed by 9-fluorenylmethyl chloroformate (FMOC-Cl) (10.06 g, 0.039 mol). The reaction mixture was stirred at room temperature overnight. The reaction mixture was quenched with water and extracted with dichloromethane. The organic phase was washed with water, dried over $MgSO_4$, filtered and concentrated *in vacuo*. The residue was chromatographed on silica gel (Petroleum ether/dichloromethane = 1:2) to afford **2** (13.87 g, 75% yield over two steps).

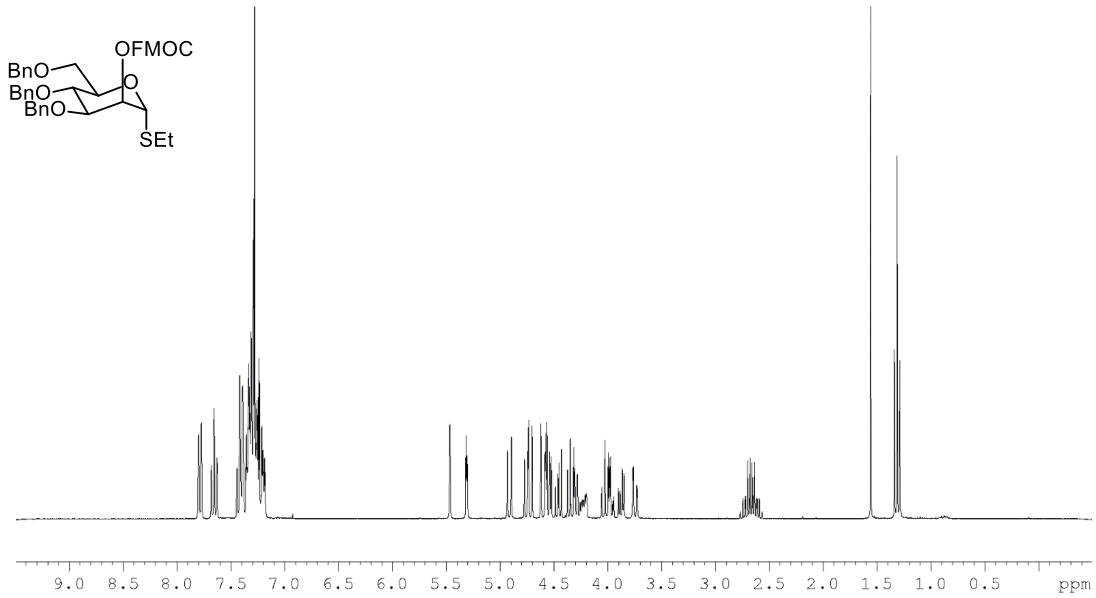
$[\alpha]_D^{25}$ = + 37.7° (*c* = 1.0, $CH_2Cl_2$)

$^1$H NMR (400MHz, $CDCl_3$): δ 1.33 (t, 3H, $^3J$ = 7.5 Hz, $SCH_2CH_3$), 2.59-2.77 (m, 2H, $SCH_2CH_3$), 3.76 (dd, 1H, $^2J$ = 11.0 Hz, $^3J$ = 1.5 Hz, H-6), 3.89 (dd, 1H, $^2J$ = 11.0 Hz, $^3J$ = 4.5 Hz, H-6), 3.97 (dd, 1H, $^3J$ = 9.5, 2.5 Hz, H-3), 4.04 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4), 4.23 (ddd, 1H, $^3J$ = 9.5, 4.5, 1.5 Hz, H-5), 4.30 (dd, 1H, $^3J$ = 7.5, 7.5 Hz, C*H*-Fmoc), 4.36 (dd, 1H, $^2J$ = 10.0 Hz, $^3J$ = 7.0 Hz, $CH_2$-Fmoc), 4.47 (dd, 1H, $^2J$ = 10.0 Hz, $^3J$ = 7.0 Hz, $CH_2$-Fmoc), 5.32 (dd, 1H, $^3J$ = 2.5, 1.5 Hz, H-2), 5.48 (d, 1H, $^3J$ = 1.5 Hz, H-1).

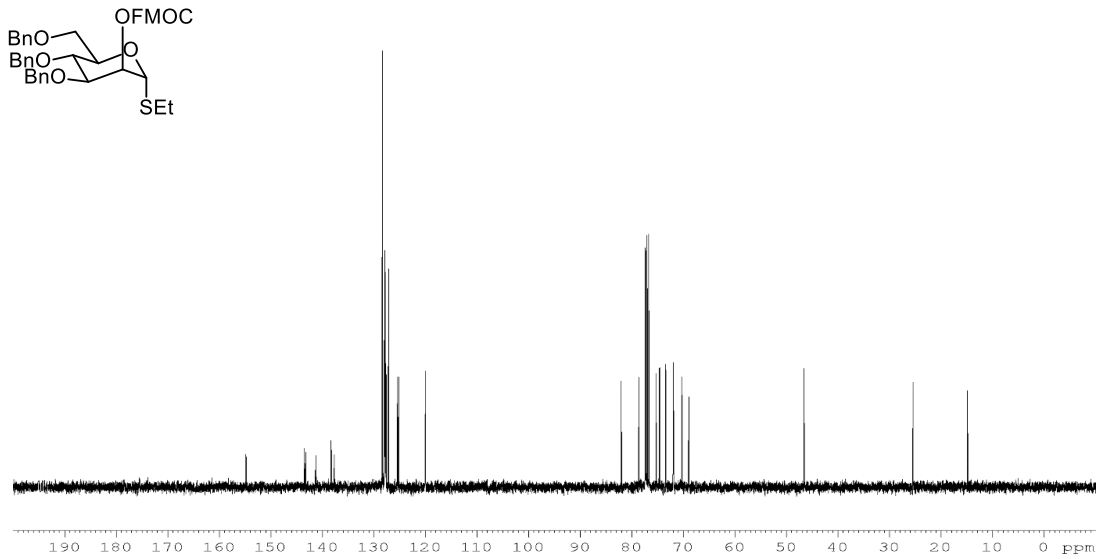
$^{13}$C NMR (100MHz, $CDCl_3$): δ 14.9 ($SCH_2CH_3$), 25.5 ($SCH_2CH_3$), 46.6 (*C*H-Fmoc), 68.9 (C-6), 70.3 ($CH_2$-Fmoc), 72.0, 73.4 and 75.3 ($CH_2Ph$), 72.0 (C-5), 74.6 (C-4), 74.6 (C-2), 78.6 (C-3), 82.1 (C-1), 154.8 (*C*O-Fmoc).

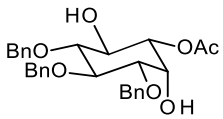
HRMS $C_{44}H_{44}O_7SNa$ : Calcd. $[M + Na]^+$ 739.2706. Found 739.2709.

**$^1$H NMR**




**$^{13}$C NMR**

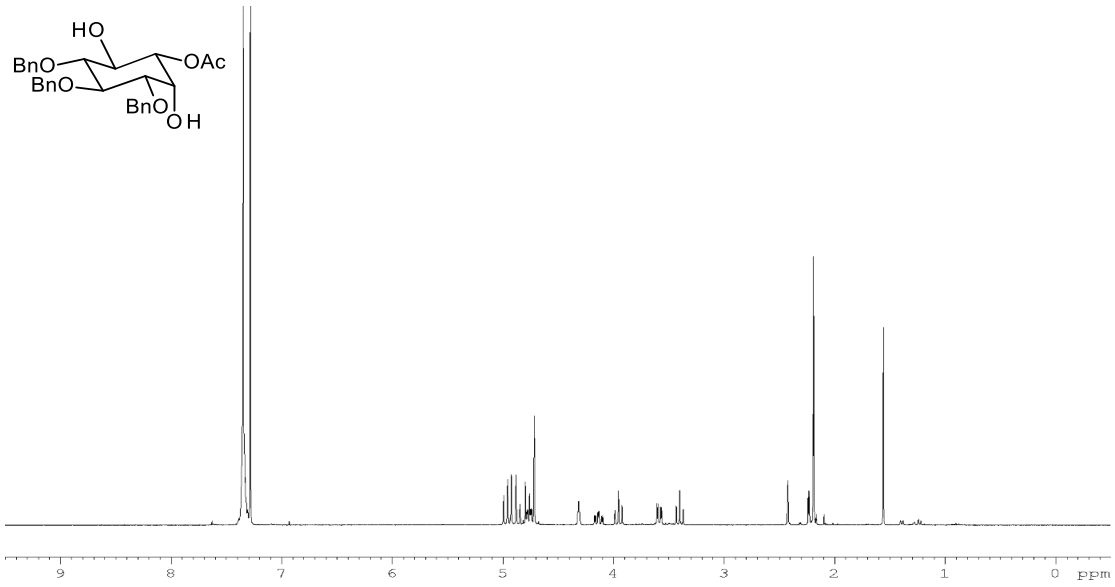


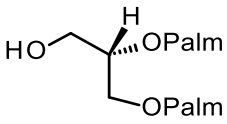




(**3**)

Known compound: prepared according to Bender and Budhu[268].

**$^1$H NMR**



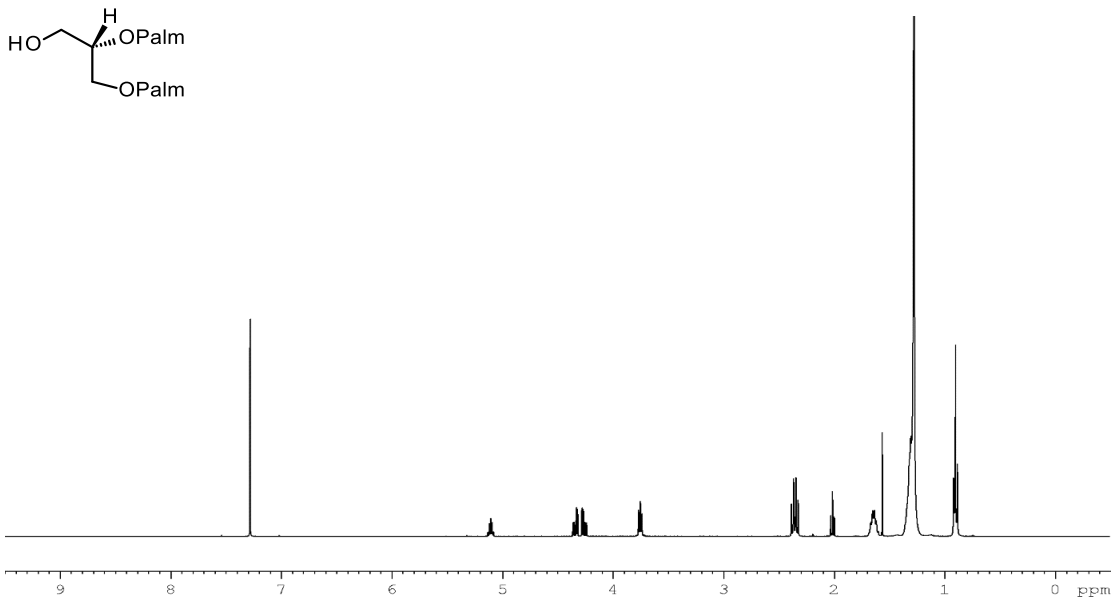



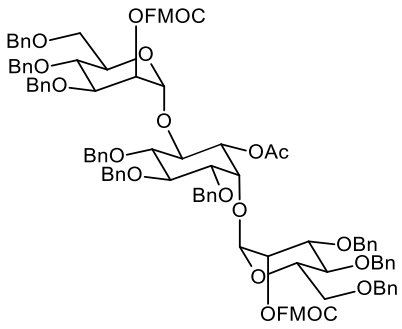

(*S*)-1,2-di-*O*-palmitoyl glycerol

(**9**)

Known compound: prepared according to Martin and colleagues[266].

**$^1$H NMR**






**$C_{113}H_{108}O_{21}$**

M = 1802.09 g.mol$^{-1}$

(**4**)

*Myo*-inositol derivative **3** (1.02 g, 2.1 mmoles) and thiomannoside **2** (4.54 g, 6.3 mmoles) were dried under vaccum for three hours, they were then dissolved in $CH_2Cl_2$ (26 ml) and powdered 4Å molecular sieves was added. The mixture was stirred at room temperature for 15 minutes. The reaction mixture was cooled to -15°C and were sequentially added N-iodosuccinimide (1.86 g, 8.24 mmoles) and TMSOTf (1M solution in $CH_2Cl_2$, 0.65 ml, 0.65 mmoles). The dark red mixture was stirred for three hours and allowed to warm to room temperature. The reaction was quenched by adding a phosphate buffer solution (pH = 7) and the aqueous phase was extracted twice with $CH_2Cl_2$. The combined organic phase was washed with saturated $Na_2S_2O_3$ and brine, dried with $MgSO_4$ and the solvent evaporated. Chromatography (toluene/ethyl acetate 98/2 to 92/8) gave **4** as colourless oil in two fractions (1.10 g, pure by NMR, 32% and 1.32 g, > 85% pure, 32%, total yield 64%).

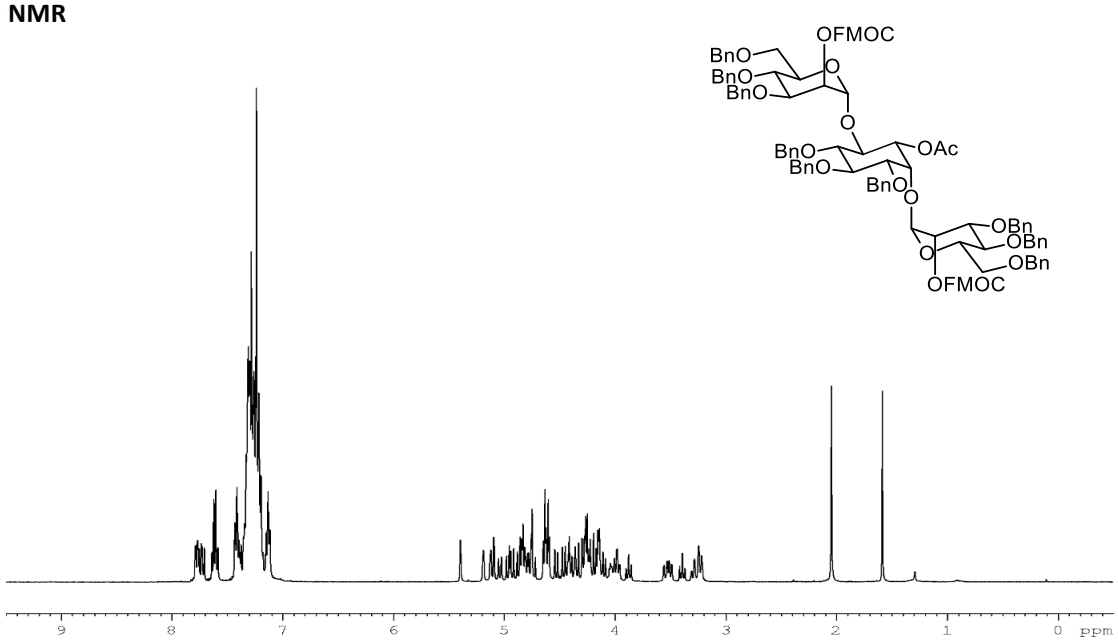
$[\alpha]_D^{25}$ = + 14.5° (*c* 1.17, chloroform)

$^1$H NMR (400MHz, $CDCl_3$: δ 2.5 (s, 3H, Ac); 3.23 (m, 2H, 2 x H-6M); 3.30 (dd, 1H, $^2J$ = 11.5, $^3J$ = 3.0 Hz, H-6M); 3.39 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-5I); 3.50 (dd, 1H, $^3J$ = 10.0, 2.5 Hz, H-3I); 3.55 (dd, 1H, $^2J$ = 11.0, $^3J$ = 2.0 Hz, H-6M); 3.88 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4I); 3.97 (dd, 1H, $^3J$ = 8.5, 2.5 Hz, H-3M); 3.99-4.08 (m, 2H, H-3M, H-5M); 4.11 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4M); 4.13-4.21 (m, 3H, H-4M, H-5M, H-6I); 4.36 (dd, 1H, $^3J$ = 2.5, 2.5 Hz, H-2I); 4.82 (dd, 1H, $^3J$ = 9.5, 2.5 Hz, H-1I); 5.09 (d, 1H, $^3J$ = 1.5 Hz, H-1M); 5.12 (dd, 1H, $^3J$ = 2.5, 1.5 Hz, H-2M); 5.19 (dd, 1H, $^3J$ = 2.5, 1.5 Hz, H-2M); 5.40 (d, 1H, $^3J$ = 1.5 Hz, H-1M).

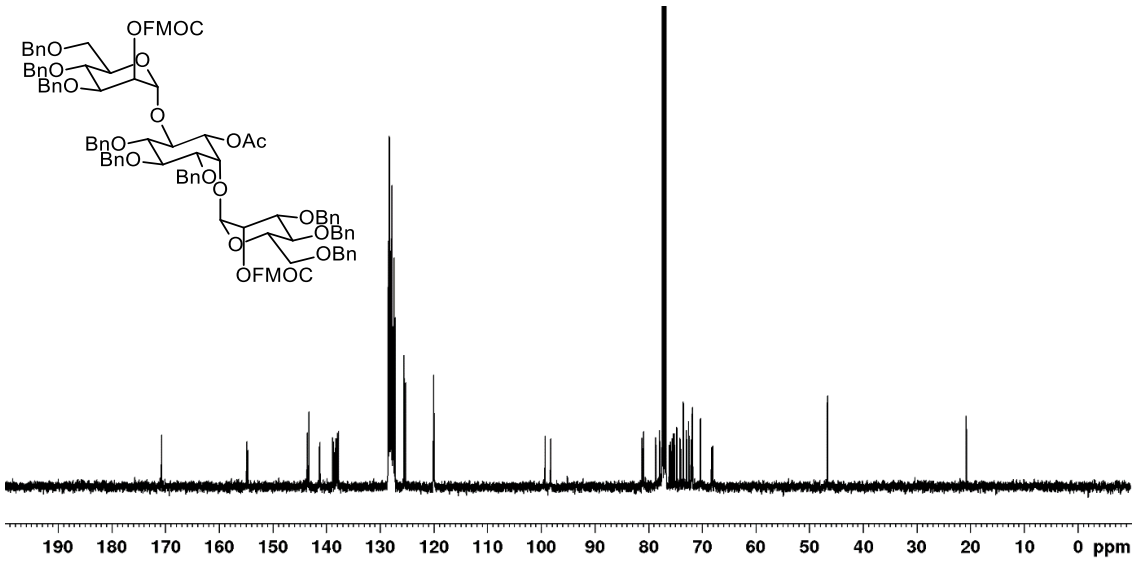
$^{13}$C NMR (100MHz, $CDCl_3$): δ 20.6 ($CH_3$-Ac), 46.5 and 46.6 (*C*H-Fmoc), 67.9 and 68.1 (C-6M), 70.2 and 70.2 (*C*$H_2$-Fmoc), 71.7 and 71.9 (C-5M), 71.7, 71.8, 72.4, 73.3, 73.4, 74.6, 75.1, 75.3 and 75.8 (*C*$H_2$Ph), 72.8 and 72.8 (C-2M), 73.9 and 74.0 (V-4M), 74.6 (C-2I), 75.5 (C-6I), 76.0 (C-1I), 77.2 and 77.8 (C-3M), 78.5 (C-3I), 80.9 (C-5I), 81.1 (C-4I), 98.1 and 99.2 (C-1M), 154.6 and 154.8 (*C*O-Fmoc), 170.8 (*C*O-Ac).

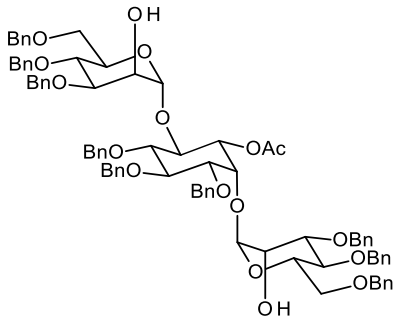
HRMS $C_{113}H_{108}O_{21}Na$ : Calcd. $[M + Na]^+$ 1823.7281. Found 1823.7264.

**$^{1}$H NMR**




**$^{13}$C NMR**






**$C_{83}H_{88}O_{17}$**

M = 1357.60 g.mol$^{-1}$

(**5**)

**4** (1.23 g, 0.68 mmoles) was dissolved in THF (10 ml) and treated with $NEt_3$ (1.2 ml, 8.3 mmoles) at room temperature for 4 hours. The mixture was concentrated *in vacuo*. Chromatography on silica (1% to 5% methanol in $CH_2Cl_2$) gave **5** (765 mg, 82%).

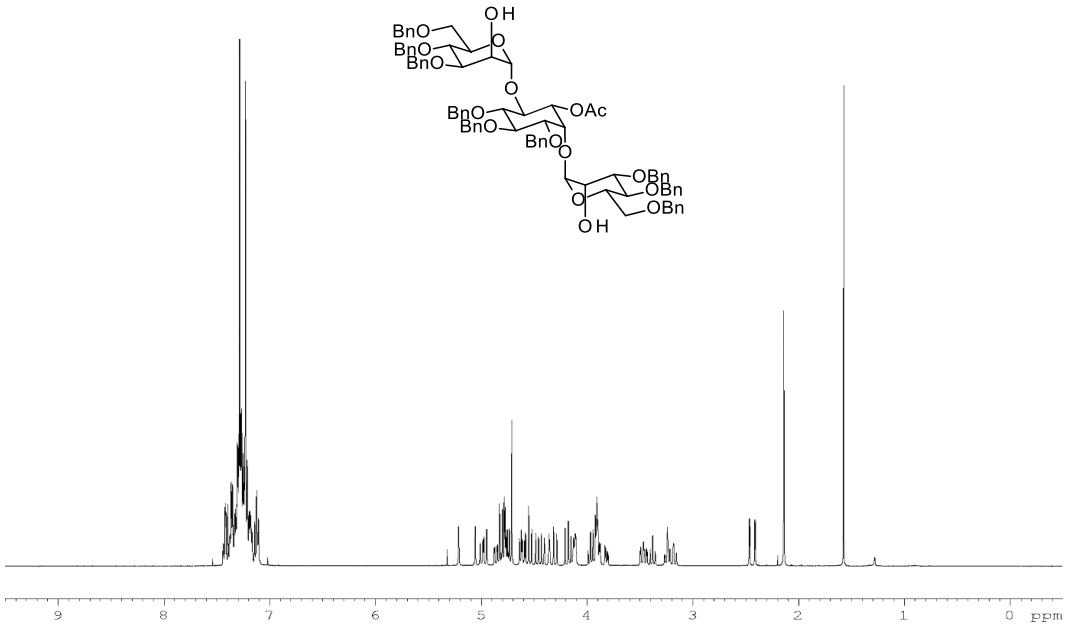
$[\alpha]_D^{25}$ = +55.0° (*c* = 1.45, chloroform)

$^1$H NMR (400MHz, $CDCl_3$: δ 2.15 (s, 3H, Ac); 2.41 and 2.46 (2d, 2 x 1H, $^3J$ = 2.5 Hz, OH-M); 3.17 (dd, 1H, $^2J$ = 10.5, $^3J$ = 1.5 Hz, H-6M); 3.24 (dd, 1H, $^2J$ = 9.5, $^3J$ = 2.0 Hz, H-6M); 3.27 (dd, 1H, $^2J$ = 9.5, $^3J$ = 2.5 Hz, H-6M); 3.39 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-5I); 3.46 (dd, 1H, $^2J$ = 10.5, $^3J$ = 2.5 Hz, H-6M); 3.50 (dd, 1H, $^3J$ = 10.0, 2.5 Hz, H-3I); 3.83 (dd, 1H, $^3J$ = 8.5, 2.5 Hz, H-3M); 3.88-3.96 (m, 5H, H-2M, H-3M, H-4M, H-4I, H-5M); 3.98 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4M); 4.11-4.15 (m, 2H, H-2M, H-5M); 4.16 (dd, 1H, $^3J$ = 10.0, 9.5 Hz, H-6I); 4.37 (dd, 1H, $^3J$ = 2.5, 2.5 Hz, H-2I); 4.87 (dd, 1H, $^3J$ = 10.0, 2.5 Hz, H-1I); 5.07 (d, 1H, $^3J$ = 1.5 Hz, H-1M); 5.22 (d, 1H, $^3J$ = 1.5 Hz, H-1M).

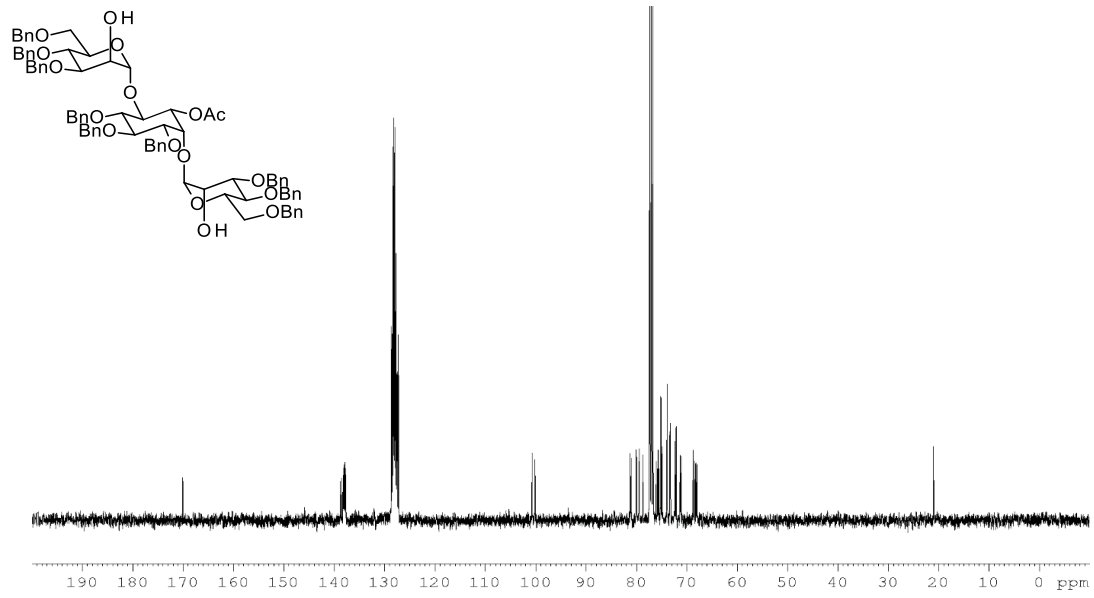
$^{13}$C NMR (100MHz, $CDCl_3$): δ 21.0 ($CH_3$-Ac), 67.9 and 68.2 (C-6M), 68.4 and 68.7 (C-5M), 71.1 and 71.3 (C-2M), 72.0, 72.1, 72.3, 73.3, 73.4, 75.1, 75.7 and 76.0 ($CH_2$-Bn), 73.3 (C-2I), 73.9 and 73.9 (C-4M), 74.9 (C-1I), 75.5 (C-6I), 78.7 (C-3I), 79.4 and 80.0 (C-3M), 81.0 (C-5I), 81.2 (C-4I), 100.1 and 100.7 (C-1M), 170.1 (C=O).

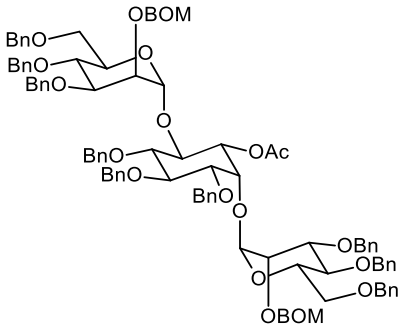
HRMS $C_{83}H_{88}O_{17}Na$. Calcd. $[M + Na]^+$ 1379.5919. Found 1379.5896.

**$^{1}$H NMR**




**$^{13}$C NMR**






**$C_{99}H_{104}O_{19}$**

M = 1597.90 g.mol$^{-1}$

(**6**)

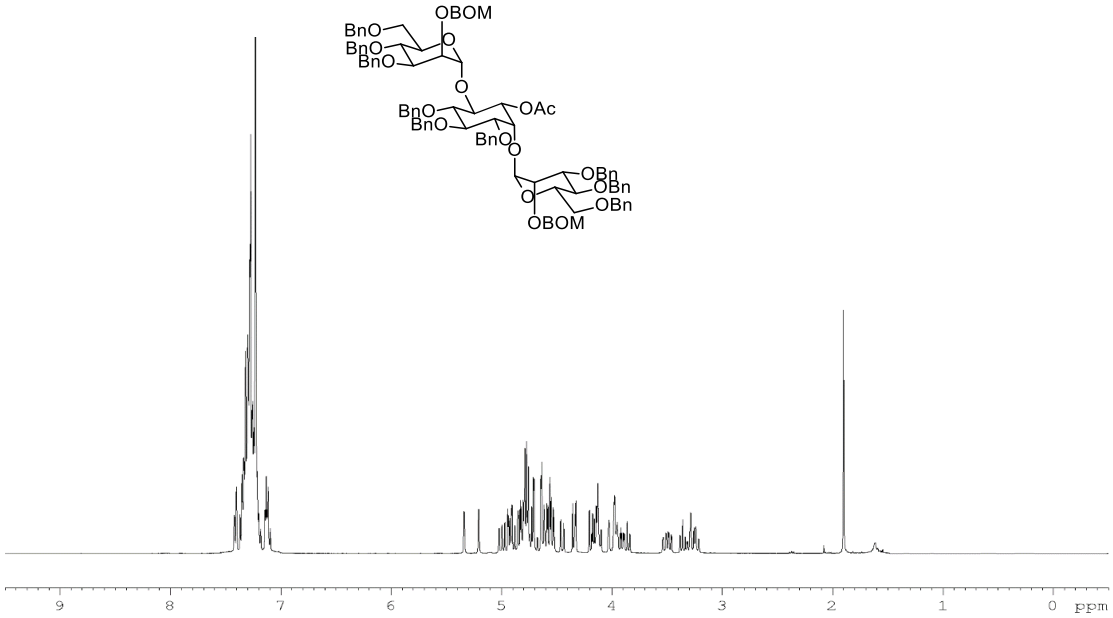
**5** (950.3 mg, 0.78 mmoles) was dissolved in anhydrous chloroform (4 ml) and treated sequentially with tetrabutylammonium iodide (144.1 mg, 0.39 mmoles) and diisopropylethylamine (453 µL, 3.12 mmoles) at room temperature. The mixture was brought to 40°C and freshly distilled benzyloxymethyl chloride (BOM-Cl, 330 µL, 2.34 mmoles) was added. The reaction was left at 40°C overnight. At that time, some monoprotected compounds were still present, BOM-Cl (110 µL, 0.78 mmoles) was added and the mixture left at 40°C until completion of the reaction. The mixture was quenched with water and diluted with $CH_2Cl_2$. The water phase was extracted twice with $CH_2Cl_2$, the organic phase was dried on $MgSO_4$ and the solvent evaporated. Chromatography on silica (petroleum ether/Ethyl acetate 85/ to 72/25) gave the product as a white solid (1.026 g, 91%).

$[\alpha]_D^{25}$ = +63.2° (*c* 1.70, chloroform)

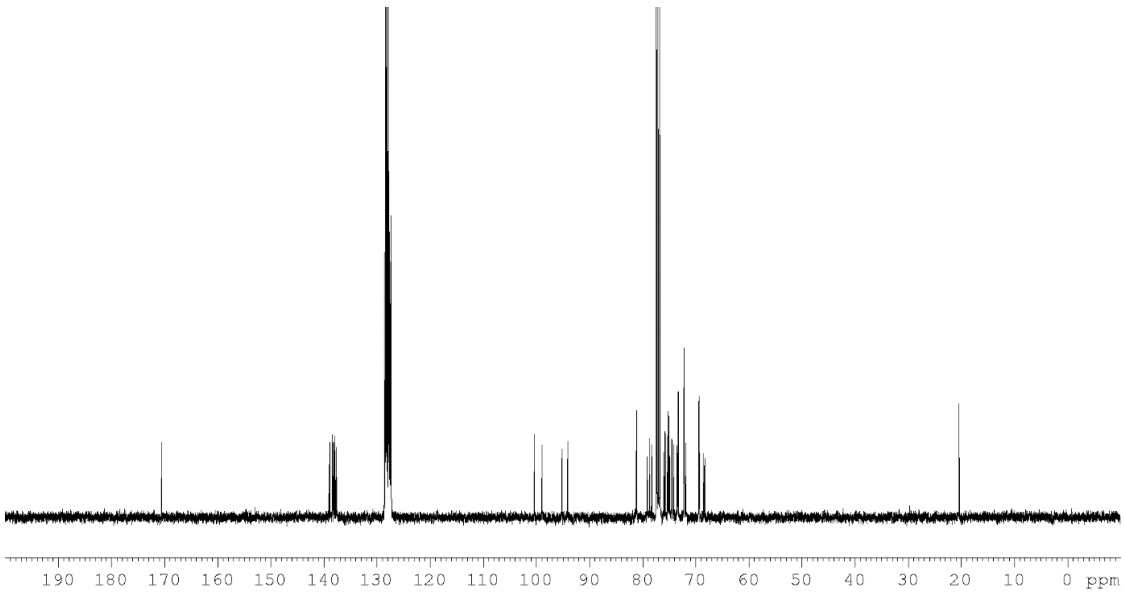
$^1$H NMR (400MHz, $CDCl_3$: δ 1.90 (s, 3H, Ac); 3.23 (dd, 1H, $^2J$ = 11.0, $^3J$ = 1.5 Hz, H-6M); 3.27 (br d, 1H, $^2J$ = 10.5 Hz, H-6M); 3.30 (dd, 1H, $^2J$ = 11.0, $^3J$ = 2.5 Hz, H-6M); 3.36 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-5I); 3.47 (dd, 1H, $^3J$ = 10.0, 2.5 Hz, H-3I); 3.52 (br d, 1H, $^2J$ = 11.0 Hz, H-6M); 3.86 (dd, 1H, $^3J$ = 10.0, 9.5 Hz, H-4I); 3.90 (dd, 1H, $^3J$ = 9.5, 3.0 Hz, H-3M); 3.96-4.00 (m, 3H, H-2M, H-3M, H-5M); 4.03 (dd, 1H, $^3J$ = 3.0, 1.5 Hz, H-2M); 4.09-4.19 (m, 4H, 2 x H-4M, H-5M, H-6I); 4.32 (dd, 1H, $^3J$ = 2.5, 2.5 Hz, H-2I); 4.76, 4.80, 4.84 and 4.92 (4d, 4 x 1H, $^2J$ = 7.5 Hz, C$H_2$-BOM); 5.21 (d, 1H, $^3J$ = 1.5 Hz, H-1M); 5.34 (d, 1H, $^3J$ = 1.5 Hz, H-1M).

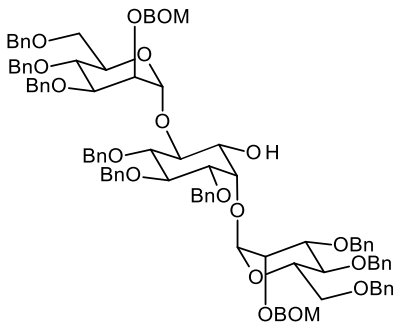
$^{13}$C NMR (100MHz, $CDCl_3$): δ 20.3 ($CH_3$-Ac), 68.3 and 68.5 (C-6M), 69.3, 69.4, 72.1, 72.1, 72.1, 73.2, 73.3, 74.89 75.1, 75.1, 76.0 (*C*$H_2$-Bn), 71.8 and 72.1(C-5M), 73.5 (C-2I), 74.1 and 75.7 (C-2M), 74.4 and 74.5 (C-4M), 75.2 (C-6I), 75.7 (C-1I), 78.2 and 79.1 (C-3M), 78.7 (C-3I), 81.1 (C-4I and C-5I), 94.1 and 95.1 (*C*$H_2$-BOM), 98.9 and 100.3 (C-1M), 170.5 (C=O). HRMS $C_{99}H_{104}O_{19}Na$: Calcd. $[M + Na]^+$ 1619.7045. Found 1619.7070.

**$^{1}$H NMR**




**$^{13}$C NMR**



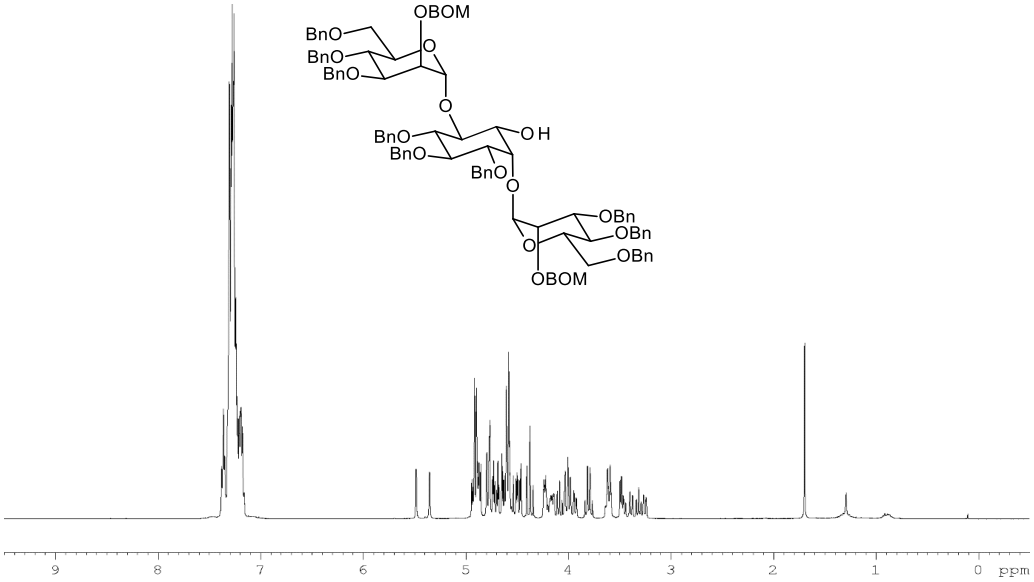
**$C_{97}H_{102}O_{18}$**

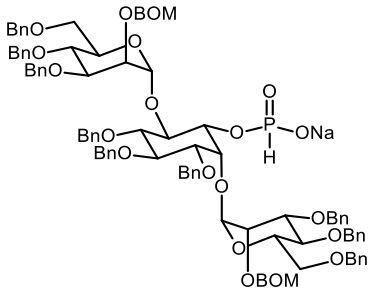
M = 1555.87 g.mol$^{-1}$

(**7**)

**6** (468 mg, 0.29 mmoles) was dissolved in 10 ml of a sodium methanolate solution in methanol and the mixture was stirred at room temperature overnight. The reaction was quenched with a bit of dry ice and the solvent evaporated. The residue was taken up in $CH_2Cl_2$ and washed with water. The water phase was extracted two times with $CH_2Cl_2$ and the organic phase dried on $MgSO_4$. Evaporation of the solvent gave **7** (453 mg, 99%) which was used for the next step without any purification.

**$^1$H NMR**




**$C_{97}H_{102}O_{20}PNa$**

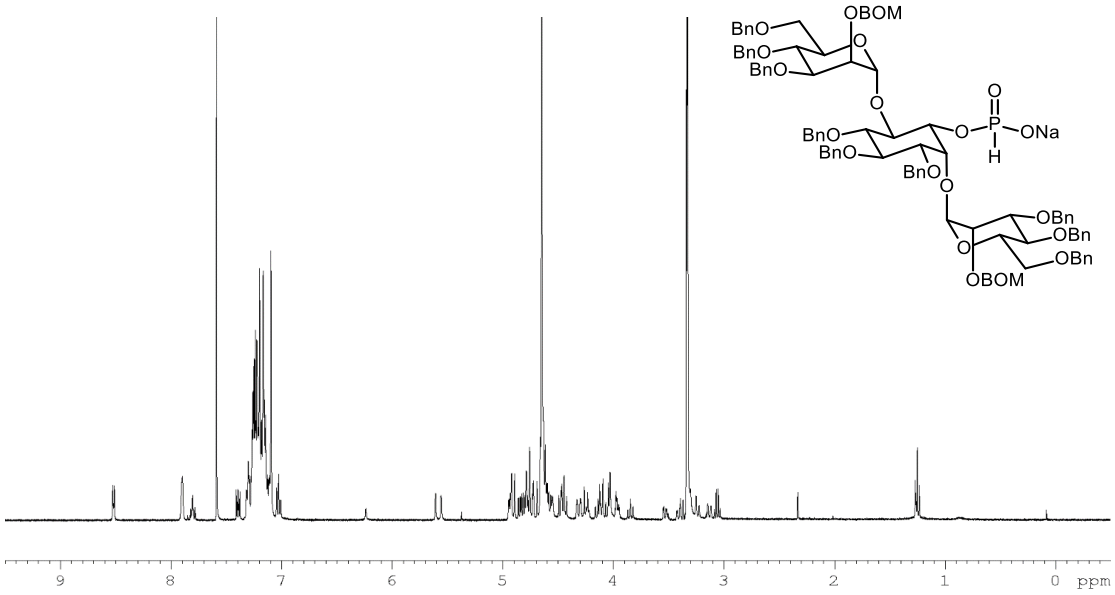
M = 1641.83 g.mol$^{-1}$

(**8**)

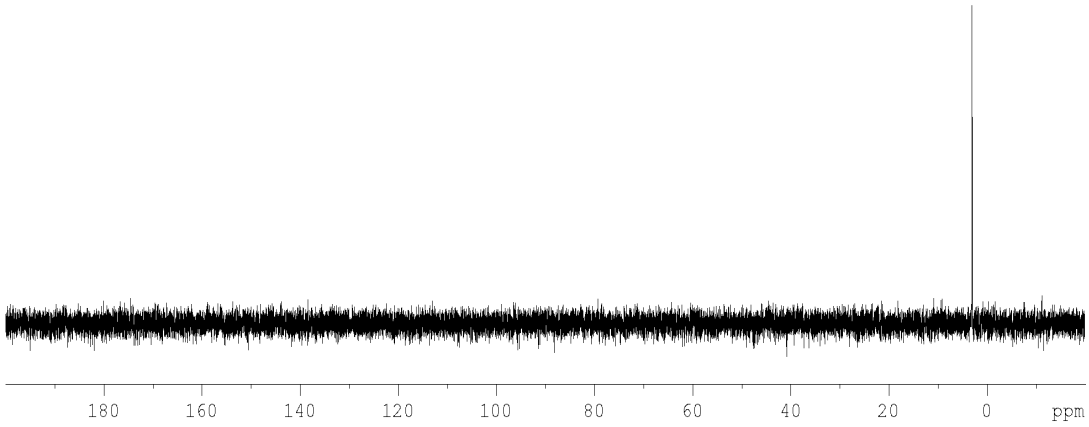
Imidazole (317 mg, 0.46 mmoles) was dissolved in toluene (10 ml) at -10°C and treated with triethylamine (370µL, 2.7 mmoles) and $PCl_3$ (87 µL, 1 mmole). To this solution was added **7** (512.0 mg, 0.31 mmoles) dissolved in toluene (5 ml + 5 ml rinse) and the reaction was left to come at room temperature overnight. The reaction was treated a 4/1 pyridine/water mixture (17 ml) and stirred for 10 min at room temperature. $CH_2Cl_2$/water extraction and solvent evaporation gave crude **8** which was used directly for the next step.

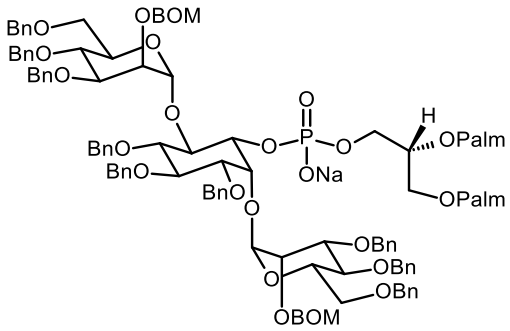
$^{31}$P NMR (162MHz, $CDCl_3/CD_3OD$ 1/1): δ 2.98.

**$^{1}$H NMR**




**$^{31}$P NMR**






**$C_{132}H_{168}O_{25}PNa$**

M = 2208.73 g.mol$^{-1}$

(**10**)

Crude **8** (0.31 mmoles) and **9** (567 mg, 1.0 mmole) were dissolved in anhydrous pyridine (10 ml) and treated with pivaloyl chloride (410 µL, 3.3 mmoles) at room temperature overnight. At that time, TLC show the reaction to be finished. The reaction mixture was treated with a solution of iodine (172 mg, 0.66 mmoles) in 2% aqueous pyridine for 15 minutes. The mixture was diluted with chloroform and washed with 10% aqueous sodium thiosulfate solution. The water phase was reextracted twice with chloroform and the combined organic phase washed with a 1M triethylammonium hydrogen carbonate aqueous solution. This aqueous phase was extracted twice with chloroform and the combined organic phase was dried with $MgSO_4$ and the solvent evaporated. Chromatography on silica (1% to 8% methanol in $CH_2Cl_2$) gave **10** as a sticky oil (574 mg, 83% overall from **7**).

$[\alpha]_D^{25}$ = +20.1° (*c* = 1.25, chloroform/methanol 1/1)

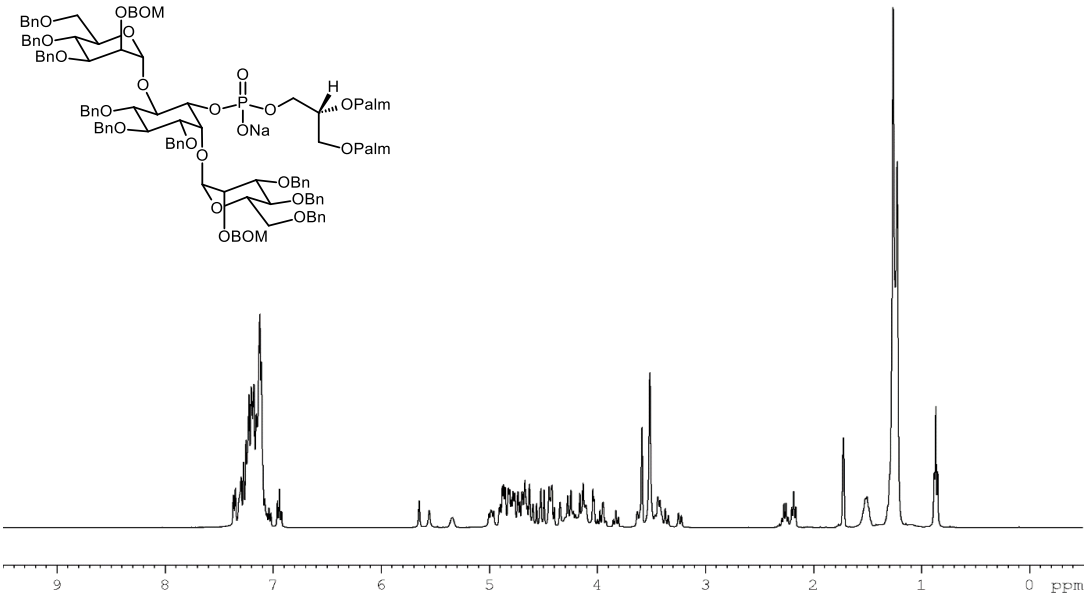
$^1$H NMR (400MHz, THF/$D_2O$ 85/15, 313 K): δ 0.90 (t, 6H, $^3J$ = 6.5 Hz, $CH_3$-palm); 1.22-1.35 (m, 48H, $CH_2$-palm); 1.48-1.57 (m, 4H, $CH_2$-palm); 2.21 (t, 2H, $^3J$ = 7.5 Hz, $CH_2$-C=O palm); 2.26 and 2.31 (2dt, 2 x 1H, $^2J$ = 15.0 Hz, $^3J$ = 7.5 Hz, $CH_2$-C=O palm); 3.26 (br d, 1H, $^2J$ = 10.5 Hz, H-6M); 3.41-3.51 (m, 4H, 3 x H-6M, H-5I); 3.64 (m, 1H, H-3I); 3.85 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4I); 3.96 (dd, 1H, $^3J$ = 9.5, 2.5 Hz, H-3M); 3.99 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4M); 4.03-4.10 (m, 2H, H-3M, H-4M); 4.10-4.21 (m, 3H, 2 x H-5M, CH-benzyl); 4.21-4.34 (m, 4H, 2 x H-1G, H-6I, CH-benzyl); 4.37 (m, 1H, H-2M); 4.42-4.50 (m, 4H, 2 x H-3G, H-2M, CH-benzyl); 4.51-4.94 (m, 23H, H-2I, H-1I, 19 x CH-benzyl, 2 x CH-BOM); 5.00 and 5.02 (2d, 2 x 1H, $^2J$ = 13.5 Hz, 2 x CH-BOM); 5.36 (m, 1H, H-2G), 5.59 and 5.67 (2 br s, 2 x 1H, 2 x H-1M).

$^{13}$C NMR (100MHz, THF/$D_2O$ 85/15, 313 K): δ 14.1 ($CH_3$-palm), 23.2, 25.5, 29.8, 29.9, 30.0, 3.01, 30.2, 30.2, 30.3, 32.5 ($CH_2$-palm), 34.3 and 34.6 (*C*$H_2$-C=O-palm), 63.4 (C-3G), 64.6 (C-1G), 69.5 and 69.8 (C-6M), 69.5, 69.8, 71.7, 71.5 (C-2G), 71.7 or 72.3 (C-2I), 72.3 and 72.5 (C-5M), 73.7 and 73.8 (C-2M), 75.3 and 75.7 (C-4M), 76.6 (C-1I), 78.4 (C-6I), 79.2 and 80.1 (C-3M), 79.9 (C-3I), 82.1 (C-4I, C-5I), 94.2 and 94.6 (*C*$H_2$-O-BOM), 99.3 and 100.2 (C-1M), 173.5 and 173.5 (C=O-palm).

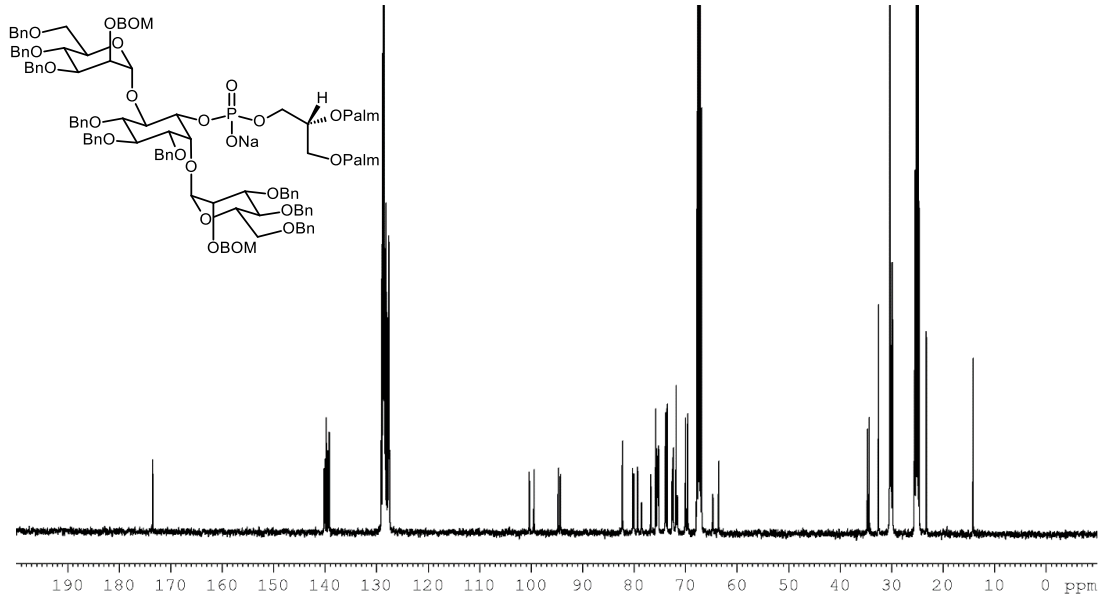
$^{31}$P NMR (162MHz, THF/$D_2O$ 85/15, 313 K): δ -2.31.

HRMS $C_{132}H_{169}O_{25}P$ : Calcd. $[M-1]^-$ 2184.1612. Found 2184.1672.

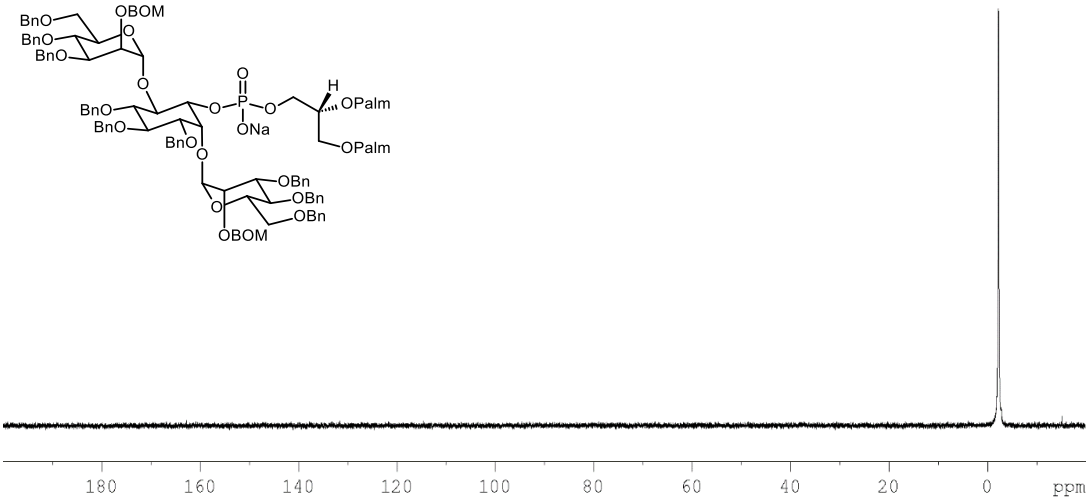
**$^{1}$H NMR**

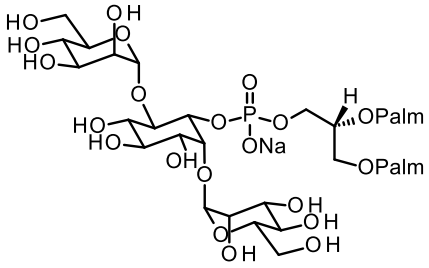



**$^{13}$C NMR**



**$^{31}$P NMR**



**$C_{53}H_{98}O_{23}PNa$**

M = 1157.31 g.mol$^{-1}$

(**1**)

**10** (256 mg, 0.12 mmoles) was dissolved in a 3/2 methanol/THF mixture and treated at room temperature with hydrogen and a catalytic amount of 20% $Pd(OH)_2/C$. The reaction was monitored by TLC and allowed to proceed for 5 hours. The catalyst was filtered on Celite (1/1 chloroform/methanol elution) and the filtrate was evaporated. Chromatography ($CHCl_3$/MeOH/ $H_2O$ 60/35/1 to 60/35/8) gave **1** as a white solid (98 mg, 74%).

$[\alpha]_D^{25}$ = +9.3° (*c* = 0.44, dichlorométhane/methanol 1/1)

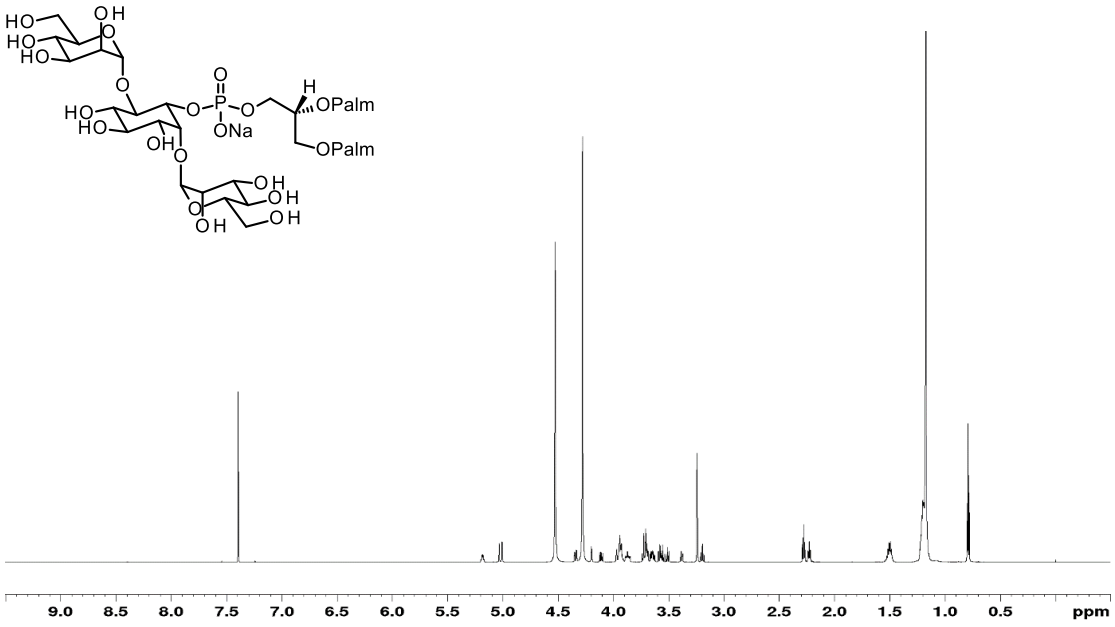
$^1$H NMR (700MHz, $CDCl_3/CD_3OD$): δ 0.79 (t, 6H, $^3J$ = 7.0 Hz, $CH_3$-palm); 1.17 (m, 48H, $CH_2$-palm); 1.50 (m, 4H, $CH_2$-palm); 2.22 (m, 2H, $CH_2$-C=O palm); 2.27 (t, 2H, $^3J$ = 7.5 Hz, $CH_2$-C=O palm); 3.19 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-5I); 3.37 (dd, 1H, $^3J$ = 10.0, 2.5 Hz, H-3I); 3.50 (dd, 1H, $^3J$ = 9.5, 9.5 Hz, H-4I); 3.55 and 3.57 (2dd, 2 x 1H, $^3J$ = 9.5, 9.5 Hz, 2 x H-4M); 3.63 and 3.67 (2dd, 2 x 1H, $^3J$ = 9.5, 5.0 Hz, 2 x H-6M); 3.67-3.74 (m, 5H, 2 x H-3M, 2 x H-6M, H-6I); 3.85 and 3.88 (2ddd, 2 x 1H, $^3J$ = 10.0, 5.0, 2.5 Hz, 2 x H-5M); 3.89-3.96 (m, 4H, 2 x H-1G, H-1I, H-2M); 3.97 (dd, 1H, $^3J$ = 3.0, 1.5 Hz, H-2M); 4.10 (dd, 1H, $^2J$ = 12.0 Hz, $^3J$ = 7.5 Hz, H-3G); 4.19 (br, 1H, H-2I); 4.34 (dd, 1H, $^2J$ = 12.0 Hz, $^3J$ = 2.5 Hz, H-3G); 5.00 and 5.03 (2d, 2 x 1H, $^3J$ = 1.0 Hz, 2 x H-1M); 5.18 (m, 1H, H-2G).

$^{13}$C NMR (176MHz, $CDCl_3/CD_3OD$): δ 13.8, 22.6, 24.8, 24.9, 29.1, 29.1, 29.3, 29.4, 29.5, 29.6, 29.6, 31.8, 34.0, 34.2, 60.7 and 61.0 (2 x C-6M), 62.6 (C-3G), 63.1 (C-1G), 66.5 and 66.6 (2 x C-4M), 69.7 and 69.9 (2 x C-2M), 69.9 (C-3I), 70.2 (C-2G, 2 x C-3M), 72.3 and 72.6 (2 x C-5M), 72.6 (C-4I), 72.9 (C-5I), 76.4 (C-1I), 77.9 (C-6I), 78.2 (C-2I), 101.1 (2 x C-1M), 173.8 and 174.1 (C=O).

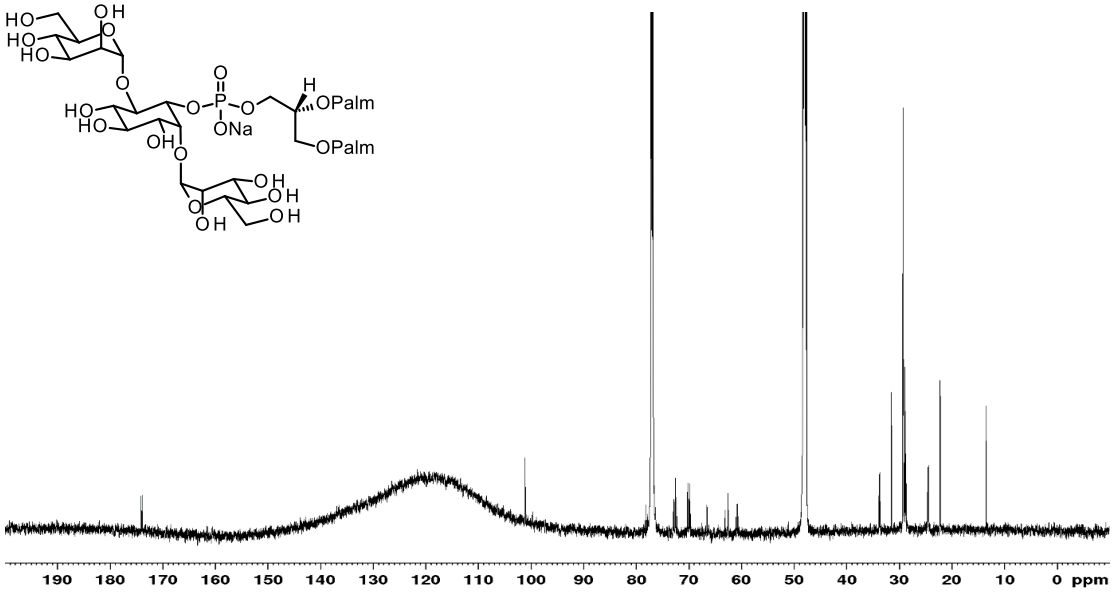
$^{31}$P NMR (162MHz, $CDCl_3/CD_3OD$ 1/1): δ -0.55.

HRMS $C_{53}H_{99}O_{23}P$ : Calcd. $[M-1]^-$ 1133.6237. Found 1133.6218.

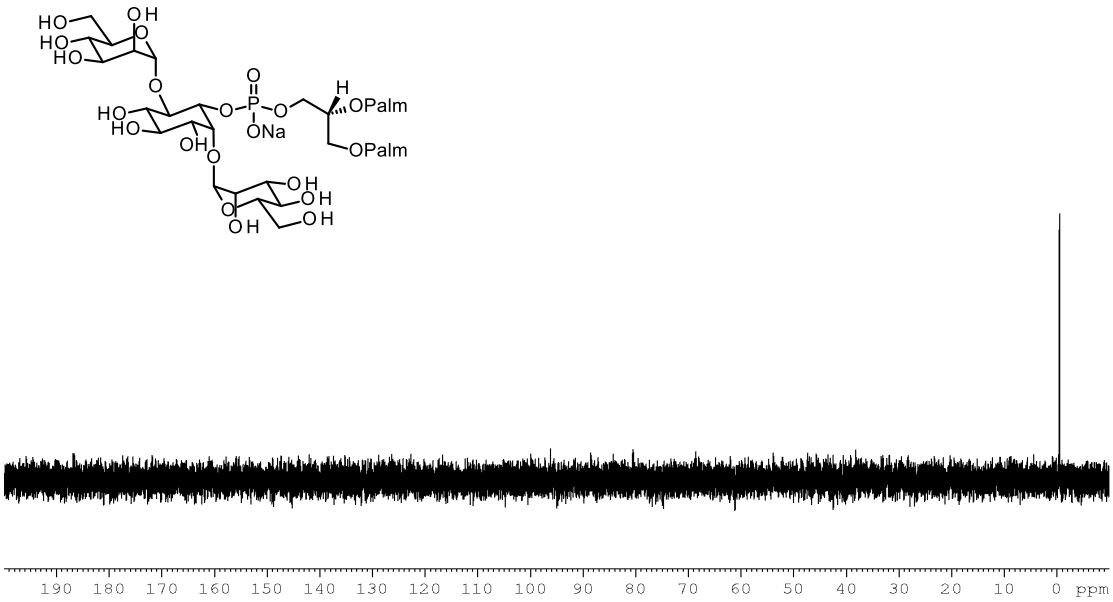
**$^1$H NMR**




**$^{13}$C NMR**



**$^{31}$P NMR**


HO
HO
HO
OH
O
HO
HO
OH
O
O
P
O
ONa
H
OPalm
OPalm
OH
OH
O
OH
OH
OH
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 ppm

# ANNEX II

# ANNEX II

nature COMMUNICATIONS

ARTICLE

Received 5 Nov 2015 | Accepted 30 Jan 2016 | Published 11 Mar 2016

DOI: 10.1038/ncomms10906 OPEN

# Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA

David Albesa-Jové[1,2,3,*], Zuzana Svetlíková[4,*], Montse Tersa[1,2,*], Enea Sancho-Vaello[1,2], Ana Carreras-González[1,2], Pascal Bonnet[5], Pedro Arrasate[1,2], Ander Eguskiza[1,2], Shiva K. Angala[6], Javier O. Cifuente[1,2], Jana Korduláková[4], Mary Jackson[6], Katarína Mikušová[4] & Marcelo E. Guerin[1,2,3,7]

The biosynthesis of phospholipids and glycolipids are critical pathways for virtually all cell membranes. PatA is an essential membrane associated acyltransferase involved in the biosynthesis of mycobacterial phosphatidyl-*myo*-inositol mannosides (PIMs). The enzyme transfers a palmitoyl moiety from palmitoyl–CoA to the 6-position of the mannose ring linked to 2-position of inositol in $PIM_1$/$PIM_2$. We report here the crystal structures of PatA from *Mycobacterium smegmatis* in the presence of its naturally occurring acyl donor palmitate and a nonhydrolyzable palmitoyl–CoA analog. The structures reveal an $\alpha/\beta$ architecture, with the acyl chain deeply buried into a hydrophobic pocket that runs perpendicular to a long groove where the active site is located. Enzyme catalysis is mediated by an unprecedented charge relay system, which markedly diverges from the canonical $HX_4D$ motif. Our studies establish the mechanistic basis of substrate/membrane recognition and catalysis for an important family of acyltransferases, providing exciting possibilities for inhibitor design.

[1] Unidad de Biofísica, Consejo Superior de Investigaciones Científicas—Universidad del País Vasco/Euskal Herriko Unibertsitatea (CSIC,UPV/EHU), Barrio Sarriena s/n, Leioa, 48940 Bizkaia, Spain. [2] Departamento de Bioquímica, Universidad del País Vasco, Leioa, 48940 Bizkaia, Spain. [3] IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain. [4] Department of Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, 84215 Bratislava, Slovakia. [5] Université d'Orléans et CNRS, ICOA, UMR 7311, F-45067 Orléans, France. [6] Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado 80523-1682, USA. [7] Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, 48160 Derio, Spain. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.E.G. (email: mrcguerin@cicbiogune.es).

Long-chain fatty acids play a central role in a variety of important biological processes in all living organisms. They are prominent constituents of biological membranes, mainly in the form of phospholipids, allowing cells to be functionally constituted and differentiated from the environment[1]. Long-chain fatty acids are used as energy storage and metabolic intermediates as well as being modulators of signal transduction pathways[2,3]. Moreover, the attachment of fatty acids to proteins and glycans generates a significant amount of structural diversity in biological systems[4]. This structural information is particularly apparent in molecular recognition events including cell–cell interactions during critical steps of development and host–pathogen interactions. Fatty acids are usually activated for subsequent reactions by esterification of their carboxyl groups with the thiol group of coenzyme A (CoA) or that of the acyl carrier protein (ACP), yielding acyl-thioesters[5]. Acyltransferases are key enzymes that catalyse the transfer of activated acyl chains to acceptor molecules of different chemical structure and complexity[6]. Importantly, acyltransferases are involved in the biosynthesis of triacylglycerols and a diverse group of naturally occurring polyesters composed of 3-hydroxyalkanoic acids, the most important storage lipids found in eukaryotes and prokaryotes, respectively[5]. Moreover, acyltransferases participate in the biosynthesis of the lipid A moiety of lipopolysaccharide, modulating virulence of some Gram-negative human pathogens[7,8]. Therefore, the understanding of the mechanism of action for these enzymes at the molecular level, particularly those firmly associated to the lipid bilayer, represents a major challenge.

The phosphatidyl-*myo*-inositol mannosides (PIMs) are glycolipids of exceptional chemical structure found in abundant quantities in the inner and outer membranes of the cell envelope of all *Mycobacterium* species[9]. PIMs are based on a phosphatidyl-*myo*-inositol (PI) lipid anchor carrying one to six Man*p* residues and up to four acyl chains, with tri- and tetra-acylated phosphatidyl-*myo*-inositol dimannoside ($PIM_2$) and phosphatidyl-*myo*-inositol hexamannoside ($PIM_6$) as the predominant species (Fig. 1a)[10,11]. PIMs are thought to be the precursors of the two major mycobacterial lipoglycans, lipomannan and lipoarabinomannan (LAM). PIMs, lipomannan and LAM are considered not only essential structural components of the mycobacterial cell envelope[10–13] but also important molecules implicated in host–pathogen interactions in the course of tuberculosis and leprosy[14,15]. The biosynthesis of PIMs is initiated by the phosphatidyl-*myo*-inositol mannosyltransferase PimA (Rv2610c in *Mycobacterium tuberculosis* (*M. tuberculosis*) H37Rv), which catalyses the transfer of Man*p* residue from GDP-Man to the 2-position of the *myo*-inositol ring of PI, to form phosphatidyl-*myo*-inositol monomannoside ($PIM_1$) on the cytoplasmic face of the plasma membrane (Fig. 1b)[12,16]. PimA was found to be essential for *Mycobacterium smegmatis* (*M. smegmatis*) $mc^2155$ and *M. tuberculosis* growth *in vitro* and *in vivo*[12,17]. The second mannosylation step involves the action of another essential enzyme PimB (Rv2188c in *M. tuberculosis* H37Rv), which transfers a Man*p* residue to the 6-position of the *myo*-inositol ring of $PIM_1$ to form $PIM_2$ (refs 17,18). Both $PIM_1$ and $PIM_2$ can be acylated with palmitate at position 6 of the Man*p* residue transferred by PimA by the acyltransferase PatA (*P*hosphatidyl-*myo*-inositol mannosides *A*cyl*T*ransferase *A*; Rv2611c in *M. tuberculosis* H37Rv), to form $Ac_1PIM_1$ and $Ac_1PIM_2$, respectively[19–21]. This enzyme has been proved (i) to be important for the optimal growth of *M. smegmatis* $mc^2155$ and (ii) an essential enzyme for the growth of *M. tuberculosis in vitro*[10,21]. Two models were originally proposed for the biosynthesis of $Ac_1PIM_2$ in mycobacteria. In the first model, PI is mannosylated to form $PIM_1$. $PIM_1$ is then mannosylated to form $PIM_2$, which is acylated to form $Ac_1PIM_2$. In the second model, $PIM_1$ is first acylated to $Ac_1PIM_1$ and then mannosylated to $Ac_1PIM_2$. Recent evidence indicates that although both pathways might co-exist in mycobacteria, the sequence of events $PI \rightarrow PIM_1 \rightarrow PIM_2 \rightarrow Ac_1PIM_2$ is favoured (Fig. 1b)[13]. Finally, $Ac_1PIM_2$ can be further acylated on position 3 of the *myo*-inositol ring to form $Ac_2PIM_2$. However, this acyltransferase as well as most of the mannosyltransferases that catalyse the formation of higher PIMs still remains to be identified[10].

During recent years major efforts have been made to understand the early steps of the PIM biosynthetic pathway at the molecular level. In that sense, the crystal structures of the mannosyltransferases PimA and PimB were reported, showing the typical organization and catalytic machinery of GT-B glycosyltransferases[22–25]. However, to date, no structural information is available for PatA, an enzyme that is a member of a large family of acyltransferases for which the molecular mechanism of substrate recognition and catalysis remains not well understood[5,7,26,27]. Here, X-ray crystallography, site-directed mutagenesis and enzymatic activity data are used to define the three-dimensional structure, acceptor/donor substrate binding, membrane binding and catalytic mechanism of PatA from *M. smegmatis*. Our results reveal that the enzyme has an α/β architecture and shares with other important eukaryotic/prokaryotic acyltransferases an acid/base catalytic mechanism involving conserved histidine and aspartic/glutamic acid residues.

## Results

**Overall structure of PatA**. The crystal structure of PatA from *M. smegmatis* was solved using single-wavelength anomalous dispersion with a $K_2PtCl_4$ derivative at 2.06 Å resolution in *C* 2 space group (PatA–C16-1). Two other crystal forms were obtained in *P* $2_1$ (PatA–C16-2) and *P* $4_22_12$ (PatA–C16-3) space groups, and the corresponding crystal structures solved at 2.90 and 2.43 Å resolution, respectively, by using molecular replacement methods (see Methods and Supplementary Information for details; Supplementary Fig. 1; Supplementary Tables 1 and 2). The high quality of the electron density maps allowed the trace of residues 41 to 302 (PatA–C16-1), 48 to 295 (PatA–C16-2) and 48 to 303 (PatA–C16-3; Supplementary Fig. 2). We have decided to use the PatA–C16-1 crystal form for our description since it displays the highest resolution. A close inspection of the three crystal structures revealed that the protein crystallized as a monomer, displaying a high degree of structural flexibility in the N- and C-terminal regions (coloured in green in Supplementary Fig. 3; root-mean-square deviation (r.m.s.d.) value of 1.6 and 8.4 Å for Cα atoms in 41–47 and 288–302 residue ranges for the N and C terminus, respectively). The central core of PatA consists of a six-stranded β-sheet with topology $\beta_1-\beta_2-\beta_3-\beta_4-\beta_7-\beta_8$ ($\beta_8$ is antiparallel) surrounded by α-helices, with an overall size of 45 × 40 × 40 Å (Fig. 2a,b). A long and open groove that runs parallel to the protein surface contains the active site. This groove is flanked by $\alpha_4$, $\beta_2$, $\alpha_8$, $\beta_4$, $\alpha_9$, $\alpha_{10}$ and the connecting loops $\beta_1$–$\alpha_7$ (residues 124–130), $\beta_2$–$\alpha_8$ (residues 148–154), $\beta_3$–$\alpha_9$ (residues 173–181), $\beta_4$–$\beta_5$ (residues 198–208), $\beta_6$–$\alpha_{10}$ (residues 219–221) and $\alpha_{11}$–$\alpha_{12}$ (residues 282–291; Fig. 2c,d). Strikingly, the groove also displays a narrow and deep, mostly hydrophobic tunnel that runs perpendicular from its floor to the central core of the protein (Fig. 2e,f). The walls of the tunnel comprise the entire central β-sheet, two α-helices $\alpha_7$ and $\alpha_8$, and the connecting loops $\beta_1$–$\alpha_7$ (residues 124–130) and $\alpha_8$–$\beta_6$ (residues 167–170). The cavity extends entirely through the core and is closed at the bottom by $\beta_8$, the connecting loop $\alpha_5$–$\alpha_6$ and part of the $\alpha_6$ (Fig. 2b,e). Interestingly, the entrance of the hydrophobic pocket displays several charged residues compatible to interact with a buried carboxylate group of a fatty acid moiety, strongly suggesting the pocket might play a role in the donor substrate binding.

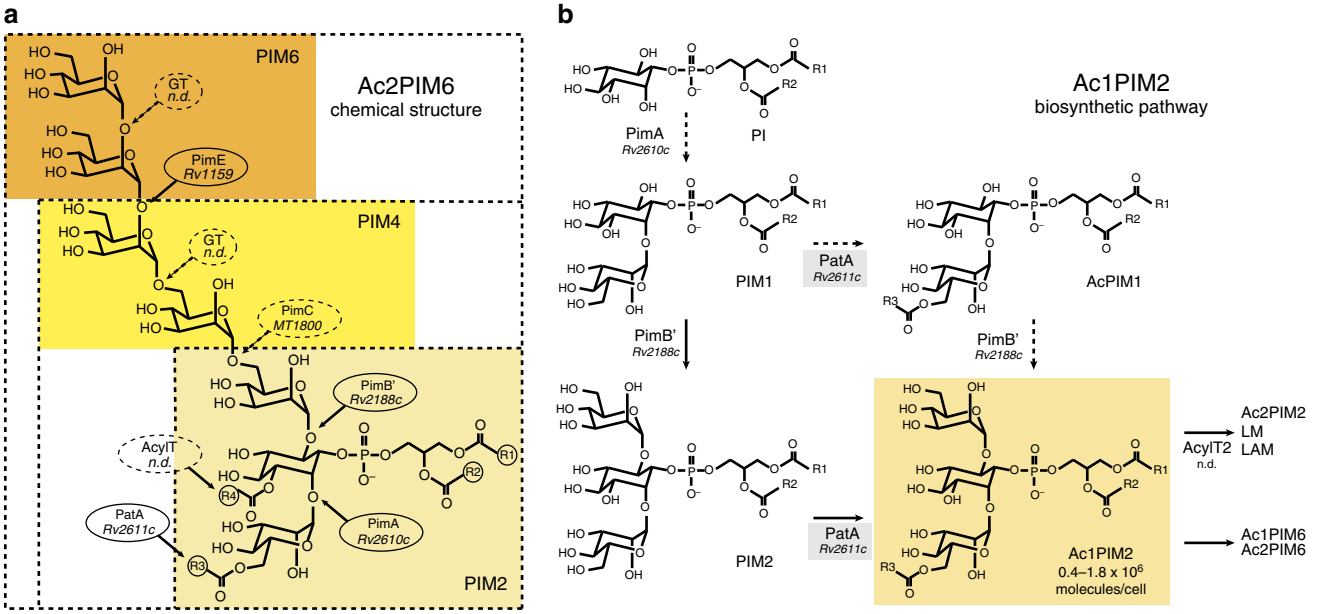



**Figure 1 | Proposed pathway for the early steps of PIM biosynthesis in mycobacteria.** (**a**) Chemical structure of $PIM_{2/6}$ and their acylated forms $Ac_{1/2}PIM_2$ and $Ac_{1/2}PIM_6$. (**b**) The two pathways originally proposed for the biosynthesis of $Ac_1PIM_2$ in mycobacteria are shown: (i) PI is mannosylated to form $PIM_1$. $PIM_1$ is then mannosylated to $PIM_2$, which is acylated to form $Ac_1PIM_2$; (ii) $PIM_1$ is first acylated to $Ac_1PIM_1$ and then mannosylated to $Ac_1PIM_2$. Our experimental evidence indicates that although both pathways might co-exist in mycobacteria, the sequence of events PI → $PIM_1$ → $PIM_2$ → $Ac_1PIM_2$ is favoured. As an important part of the literature concerning PIMs studies refers to the nomenclature based on the *Mtb* H37Rv sequences, we also include the Rv numbers to identify the proteins.

**Membrane association**. PatA catalyses an essential step in the biosynthesis of PIMs in *M. tuberculosis*[10]. The enzyme transfers a palmitoyl moiety from palmitoyl–CoA to the 6-position of the Man*p* ring linked to 2-position of *myo*-inositol in $PIM_1$ or $PIM_2$. A close interaction of the enzyme with the cytosolic face of the mycobacterial plasma membrane might be a strict requirement for $PIM_1$ or $PIM_2$ modification by PatA. Supporting this notion, PatA was found to co-localize with the mycobacterial membrane fraction[19,28]. Analysis of the amino-acid sequence of the enzyme revealed the lack of a signal peptide or hydrophobic transmembrane segments, suggesting that PatA associates to only one side of the lipid bilayer, a typical feature of peripheral and monotopic membrane proteins[24,29]. To perform their biochemical functions, these proteins very often display a high content of solvent exposed positively charged residues in the form of amphiphatic helices, promoting membrane surface interaction with anionic phospholipids[30]. Protein–membrane interaction appears to be mediated following different thermodynamic steps: (i) peptide binding is initiated by the electrostatic attraction of the positively charged residues to the anionic membrane, (ii) most likely followed by the transition of the peptide into the plane of binding and (iii) a change of the conformation of the bound peptide[31]. The penetration of the protein depends on the chemical nature of the lipids and peptides involved and also on the mechanistic nature of the processes involved, in which both location and timing of membrane association can be tightly controlled. The electrostatic surface potential of PatA revealed a clear solvent-exposed area adjacent to the major groove that contain several hydrophobic patches interspersed with clusters of positively charged residues[32]. Specifically, this region comprises the $\alpha_2$, the amphiphatic helices $\alpha_3$, $\alpha_4$ and $\alpha_8$, the connecting loop $\beta_2$–$\alpha_8$ (residues 148–154), and likely the $\alpha_1$ and $\alpha_{12}$ helices located at the N and C terminus of the protein (Supplementary Fig. 4). Interestingly, the opposite side of PatA displays a negatively charged surface, which would generate a significant electrostatic repulsion with the lipid bilayer. Thus, the polar character of PatA certainly contributes to determine the correct orientation of the enzyme into the membrane (Supplementary Fig. 4).

**The acyl–CoA-binding site**. Strikingly, one molecule of palmitic acid was unambiguously identified in the difference electron density maps of PatA–C16-1, PatA–C16-2 and PatA–C16-3 crystal structures (Fig. 3a–d). We believe that acyl molecule is associated to the enzyme due to the hydrolysis of palmitoyl–CoA during the isolation and purification of PatA from *M. smegmatis* $mc^2155$. Supporting this notion, the chemical structure of $Ac_1PIM_2$ was clearly established by using a combination of analytical techniques including mass spectrometry and two-dimensional NMR[20]. The major acyl form observed, corresponded to $PIM_2$ with the glycerol moiety being di-acylated by $C_{16}/C_{19}$ and the mannose residue transferred by PimA bearing a $C_{16}$. This structural profile was also identified in its hexamannosylated derivative $Ac_1PIM_6$ (ref. 21). The acyl chain is deeply buried into the hydrophobic pocket, and oriented with the carboxylate group facing the groove and the acyl tail extending into the globular core of the monomer (Fig. 3). The residues that contact the bound palmitic acid are highly conserved in the PatA mycobacterial homologues (Supplementary Fig. 5). Interestingly, the comparison of the three crystal structures revealed conformational flexibility in the carboxylate moiety of the palmitate (Supplementary Fig. 3), located in close proximity with the lateral chain of H126. The R164 guanidinium group engages the side chain of Y83 and D131. This tryptophan residue, together with M198, makes an important van der Waals interaction with the acyl chain. The acyl chain undergoes a kink at position C6 and terminates in a pocket mainly formed by hydrophobic residues, including L122 and L124 ($\beta_1$), A133, W136 and L137 ($\alpha_7$), F144 ($\alpha_7$–$\beta_2$ loop), T146 ($\beta_2$), F169 ($\alpha_8$–$\beta_3$ loop), F235 and V237 ($\beta_7$), M248 and V250 ($\beta_8$), and two cysteine residues C196 ($\beta_4$) and C239 ($\beta_7$).

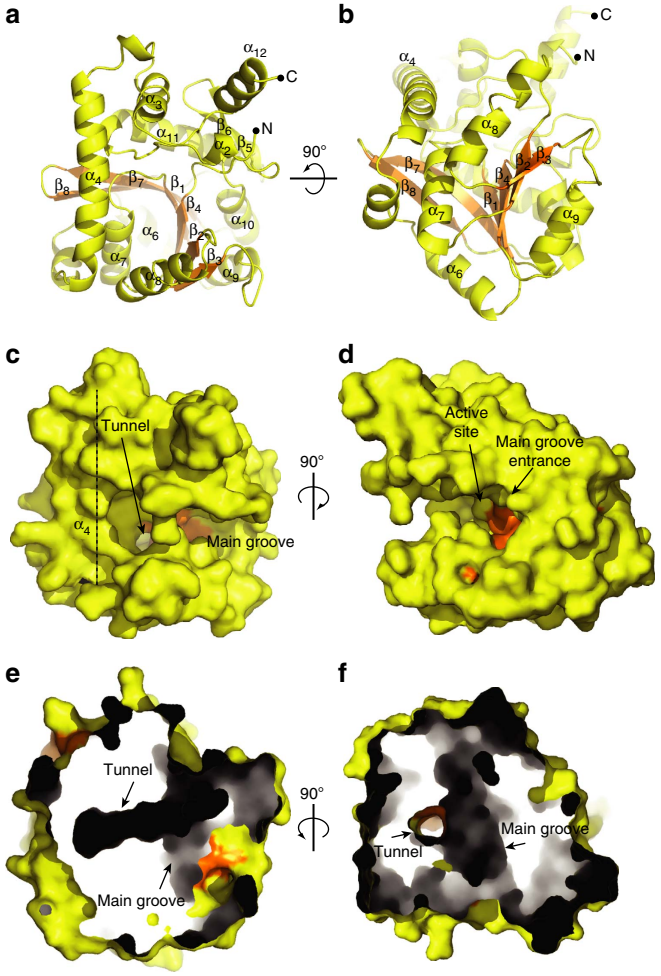



**Figure 2 | Overall structure of PatA.** (**a**,**b**) Cartoon representation showing the general fold and secondary structure organization of PatA. Secondary structure elements are labelled. The central core β-sheet is shown in orange. (**c**,**d**) Surface representation of PatA showing the location of the main groove and the active site. The groove entrance is flanked by two important α-helices, $\alpha_{11}$–$\alpha_{12}$. The groove ends up into a cavity mainly flanked by $\alpha_4$. (**e**,**f**) The main groove runs perpendicular to a hydrophobic tunnel, which is deeply buried into the core of PatA.

Interestingly, the connecting loop $\alpha_5$–$\alpha_6$ (residues 101–106) and part of $\alpha_6$ form a flexible and mostly hydrophobic cap that closes the bottom of the cavity, suggesting that PatA might be able to discriminate the length of the acyl chain groups (Fig. 3e,f). It is worth noting that $Ac_1PIM_2$ was the main product formed in the reaction when endogenous or crude mycobacterial phospholipids from *Mycobacterium phlei* were used as the lipid acceptors and a series of acyl–CoA derivatives of fatty acids were used as $^{14}$C-labelled donor substrates[33,34]. Palmitoyl–CoA (C16:0) gave higher incorporation than myristyl–CoA (C14:0). Interestingly, the oleyl–CoA (C18:1) was a much better substrate than the saturated counterpart stearyl–CoA (C18:0). Finally, the tuberculostearic acid (C19:1) had a low specific activity and the small incorporation of label that was observed may be not significant[34].

How does PatA recognize CoA? To this end, the crystal structure of PatA in complex with *S*-hexadecyl Coenzyme A (S-C16CoA), a nonhydrolyzable analogue of palmitoyl–CoA, was solved at 3.28 Å resolution in $P\,2_1$ space group (PatA–S-C16CoA; Supplementary Table 1; Fig. 4). The acyl chain of S-C16CoA is localized into the hydrophobic tunnel, and superimposes very well with the acyl chain moiety of palmitate observed in the PatA–C16-1, PatA–C16-2 and PatA–C16-3 complexes. It is worth noting that the position of the carboxylate group of palmitate is different to that observed for the thioether in the PatA–S-C16CoA complex. The 4-phosphopantetheinate moiety of S-C16CoA is clearly defined in the electron density map, and located at the entrance of the main groove, in close contact with a highly conserved region flanked by the $\beta_2$–$\alpha_8$ (residues 149–153), $\beta_3$–$\alpha_9$ (residues 174–180) and $\beta_4$–$\beta_5$ (residues 199–207) loops, and two alpha helices, $\alpha_9$ (residues 181–190) and $\alpha_{10}$ (residues 221–230; Supplementary Fig. 6). The adenosine 3′,5′-diphosphate (3′,5′-ADP) moiety of the ligand (disordered in other monomers of the asymmetric unit) sticks out from the globular core and is exposed to the bulk solvent, as observed in other acyl–CoA modifying enzymes (Supplementary Fig. 7)[35]. The approximate volume of the palmitoyl–CoA-binding pocket was *ca.* 2,801 Å$^3$(ref. 36) To further validate the model, we designed site-directed mutations predicted to impair the palmitoyl–CoA interaction with PatA. Thus, the double substitution F182W/L197W would block the groove region of PatA, hindering the formation of the complex (Fig. 4b and Supplementary Fig. 8). As depicted in Fig. 5a, the PatA F182W/L197W variant could not (i) transfer a palmitoyl moiety to $PIM_2$, or (ii) hydrolyze palmitoyl–CoA, supporting the proposed model (Supplementary Fig. 9).

**The phosphatidylinositol mannosides binding site**. Although we were unable to co-crystallize PatA in complex with $Ac_1PIM_1$/$Ac_2PIM_1$, $Ac_1PIM_2$/$Ac_2PIM_2$ or their deacylated analogs, the three-dimensional structure suggests the possible binding mode for the polar head of $PIM_1$ or $PIM_2$ acceptor substrates within the active site. Docking calculations placed the polar head of $PIM_2$, the better substrate of PatA making important interactions within a region located at the end of the main groove and comprising helices $\alpha_4$ and $\alpha_8$, and the connecting loops $\beta_1$–$\alpha_7$ (residues 83–90), $\beta_2$–$\alpha_8$ (residues 148–154) and $\alpha_{11}$–$\alpha_{12}$ (residues 282–291; Fig. 6)[13,28]. As a consequence, the O6 atom of the Man*p* ring linked to position 2 of *myo*-inositol in $PIM_2$ is predicted to be positioned favourably for activation by H126 and to receive the palmitate group from palmitoyl–CoA (Fig. 6; distance H126 NE2 atom to Man*p* O6 atom is 2.8 Å). The model also predicts an important role of residues E149, R164 and H284 to bind $PIM_2$ in the active site. In the model, E149 OE1 atom is found at 2.9 Å of the Man*p* residue O4 atom and R164 NH2 atom is placed at 3.1 Å of the O2 atom of the Man*p* ring, whereas H284 ND1 atom is at only 2.5 Å of *myo*-inositol O3 atom. The *myo*-inositol moiety and the Man*p* ring linked to position 6 of *myo*-inositol in $PIM_2$ also interact with Y80 and Q287. The approximate volume of the $PIM_2$-binding pocket was *ca.* 1741 Å$^3$ (ref. 36). Interestingly, docking calculations put the *myo*-inositol and Man*p* moieties of $PIM_1$ in an equivalent position to that observed for $PIM_2$, leaving free space in the pocket, a fact that might account for the acceptor substrate specificity of PatA (Supplementary Fig. 10). To experimentally validate the proposed model, we designed three single-point substitution, E149A, R164A and H284A, predicted to impact the $PIM_2$ interaction with PatA (Fig. 6c; Supplementary Fig. 8). As depicted in Fig. 5a, the transferase activity of all PatA variants was severely compromised, nevertheless preserving the capability to hydrolyse palmitoyl–CoA (Supplementary Fig. 9).

**Structural similarity with other acyltransferases**. Structure homologue search using DALI server revealed only one protein with significant structural similarity, that of the glycerol-3-phosphate acyltransferase from *Cucurbita moschata* (*Cm*GPAT;

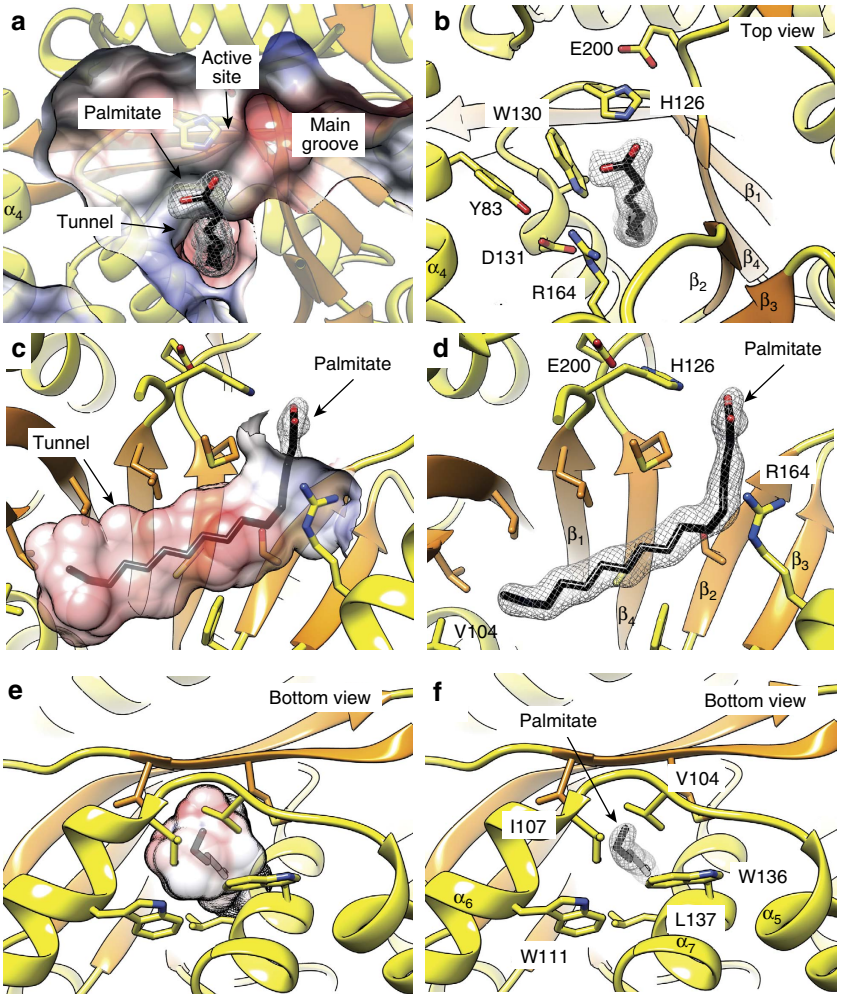



**Figure 3 | The palmitate binding site of PatA.** (**a**–**d**) Four views of PatA–C16-1 crystal structure showing the palmitate chain deeply buried into the hydrophobic tunnel. (**e**,**f**) Two views of PatA showing the cap which closes the bottom of the hydrophobic tunnel. The 2mFo-DFc electron density map countered at 1 σ for the palmitate ligand is shown.

pdb codes 1IUQ and 1K30; *Z*-score of 8.9; r.m.s.d. value of 3.9 Å for 164 aligned residues; Fig. 7a)[37,38]. *Cm*GPAT catalyses the transfer of an acyl chain either from acyl–acyl-carrier protein (acyl–ACP) or acyl–CoA, to the *sn*-1 position of glycerol-3-phosphate, to form 1-acylglycerol-3-phosphate[39]. Importantly, *Cm*GPAT is able to use palmitoyl–CoA as a donor substrate, as PatA does, using a bi–bi-ordered mechanism[40]. *Cm*GPAT belongs to a large family of glycerol-3-phosphate acyltransferases (GPAT), which are critical in the biosynthesis and regulation of phospholipids composition in prokaryotic and eukaryotic cells[41–43]. The structure of *Cm*GPAT is composed of two domains: (i) a helical domain comprising the first 78 residues of the protein displays a four-helix bundle architecture of unknown function, and (ii) an α/β domain, consisting of a nine-stranded continuous β-sheet surrounded by 11 α-helices (Fig. 7a). As depicted in Fig. 7a, the central β-sheet of *Cm*GPAT superimposes well with that observed in PatA, with the exception of its outermost strands $\beta_1$, $\beta_2$ and $\beta_9$, which are missing in PatA (see also Supplementary Fig. 11).

A combination of site-directed mutagenesis and activity measurements provided experimental support on the location of the acceptor-binding site in *Cm*GPAT (Fig. 7b–d)[44]. The replacement of R235, R237 and K193 by serine resulted in inactive enzymes. However, the *Cm*GPAT variants retained the ability to bind stoichiometric quantities of acyl–ACPs, consistent with the location of these residues in the positively charged acceptor-binding pocket, and in close proximity to the catalytic $HX_4D$ motif[37,38,44]. Interestingly, the suggested $PIM_1$/$PIM_2$-binding site in PatA is located in the same region of the glycerol-3-phosphate-binding site in *Cm*GPAT (Fig. 7d). A sulfate ion observed in the crystal structure of *Cm*GPAT superimposed very well with the phosphate moiety of the inositol ring of $PIM_1$ or $PIM_2$ (Fig. 7d). It is worth noting that, as acceptors exhibit a marked diversity of chemical structures compared with acyl–CoA or acyl–ACP donors, the acceptor-binding sites reflect this variability by showing different rearrangements of secondary structural elements (Fig. 7d).

The location of the acyl–CoA and acyl–ACP-binding site in *Cm*GPAT has been a matter of strong debate[37,38]. On the basis of sequence conservation analysis, a structural model in which the acyl chain runs over the entrance of the main groove of the protein was first proposed[37]. Molecular surface calculations revealed the existence of three tunnels with sufficient space to accommodate the acyl chain of palmitoyl–CoA[45]. Furthermore, enzymatic analysis of chimeric *Cm*GPAT and *Spinacea oleracea* GPAT revealed that a region comprising residues 128–187 is important for acyl–CoA selectivity. This region completely covered the narrowest and most hydrophobic of the three tunnels, tunnel-2 (Fig. 7b,c), which was proposed to be involved in fatty-acid recognition[38]. Strikingly, the acyl-binding pocket

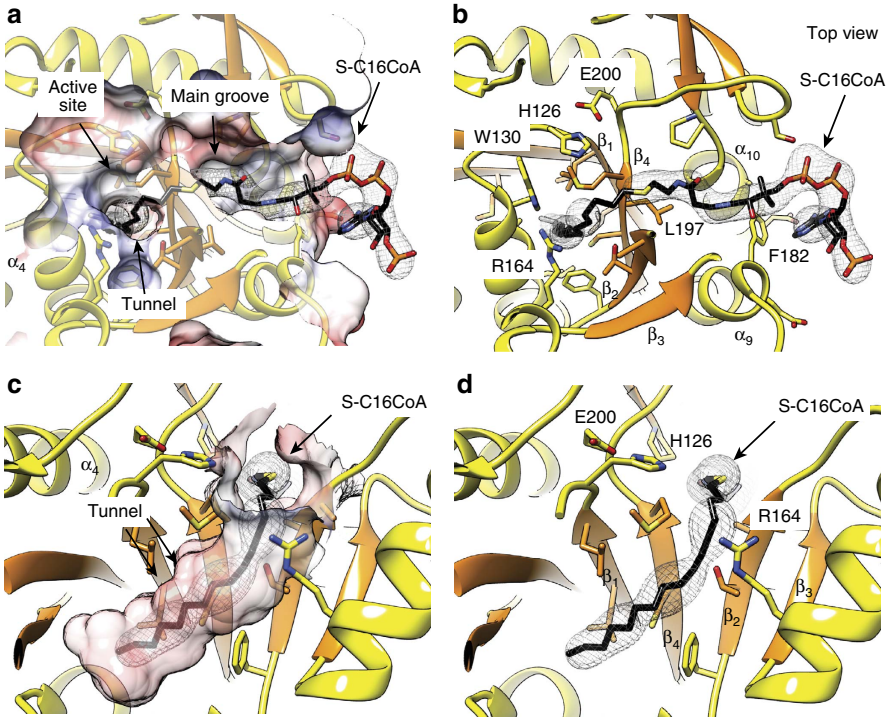



**Figure 4 | The palmitoyl-CoA binding site of PatA.** (**a**–**d**) Four views of PatA-S-C16CoA crystal structure showing the binding mechanism of S-C16CoA. The palmitoyl moiety is deeply buried into the hydrophobic tunnel, whereas the coenzyme A moiety extends outwards through the side of the main groove. The 2mFo-DFc electron density map countered at 1 σ for the palmitate ligand is shown.

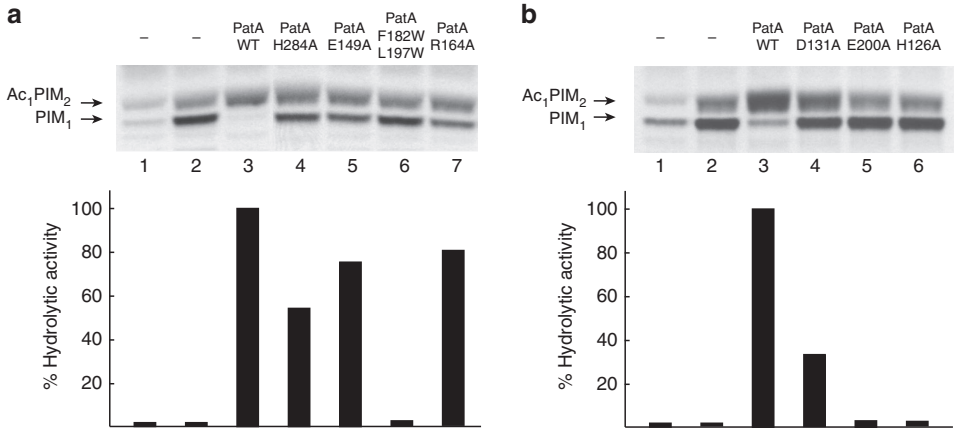



**Figure 5 | Enzymatic activity of selected PatA variants.** (**a**) In the upper panel, the acyltransferase activity of purified PatA and variants involved in substrate binding are shown (see Supplementary Fig. 9 for duplicates). Reaction mixtures contained crude membranes from *M. smegmatis* mc$^2$155 and GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with PimA (lanes 2–7) and palmitoyl-CoA (lanes 3–7) and purified PatA (lane 3), PatA-H284A (lane 4), PatA-E149A (lane 5), double mutant PatA-F182W/L197W (lane 6) and PatA-R164A (lane 7). The lipids were extracted from reaction mixtures and analysed by TLC and autoradiography as described in Methods section. In the lower panel, the hydrolytic activity against palmitoyl-CoA is shown. (**b**) In the upper panel, the acyltransferase activity of purified PatA and variants involved in catalysis are shown (see Supplementary Fig. 9 for duplicates). Reaction mixtures contained crude membranes from *M. smegmatis* mc$^2$155 and GDP-[$^{14}$C]-mannose as a tracer (lane 1), supplemented with PimA (lane 2–6) and palmitoyl-CoA (lanes 3–6) and purified PatA (lane 3), PatA-D131A (lane 4), PatA-E200A (lane 5) and PatA-H126A (lane 6). In the lower panel, the hydrolytic activity against palmitoyl-CoA is shown as determined by spectrophotometric analysis. All enzymatic activities measurements were determined in duplicates (Supplementary Fig. 9)

identified in the crystal structures of the palmitoyl–PatA and PatA-S-C16CoA complexes superimposed very well with tunnel-2 of *Cm*GPAT (Supplementary Fig. 11). On the basis of the experimental location of the S-C16CoA in PatA, a palmitoyl–CoA molecule was fitted into tunnel-2 and subjected to energy minimization (Fig. 7d,e; Supplementary Fig. 11; see Methods for details). First, the palmitate moiety accommodates into the hydrophobic tunnel. The walls of the tunnel are covered with mainly conserved, hydrophobic and aromatic residues including L135 ($\beta_3$), P145 and I148 ($\alpha_7$), I159 ($\alpha_8$), T163 ($\alpha_8$–$\beta_4$ loop), F165 ($\beta_4$), F180 ($\alpha_{10}$), L186 ($\beta_5$), L226 and W228 ($\beta_6$), L274 and L276 ($\beta_7$) and L305, with F120 and I123 ($\alpha_6$) making up the cap. Interestingly, the CoA-binding site observed in PatA is also conserved in *Cm*GPAT. Specifically, the acyl donor is located in the region corresponding to the entrance of the main groove (Fig. 7c,e). The two α-helices $\alpha_9$ and $\alpha_{10}$, flanking the entrance of

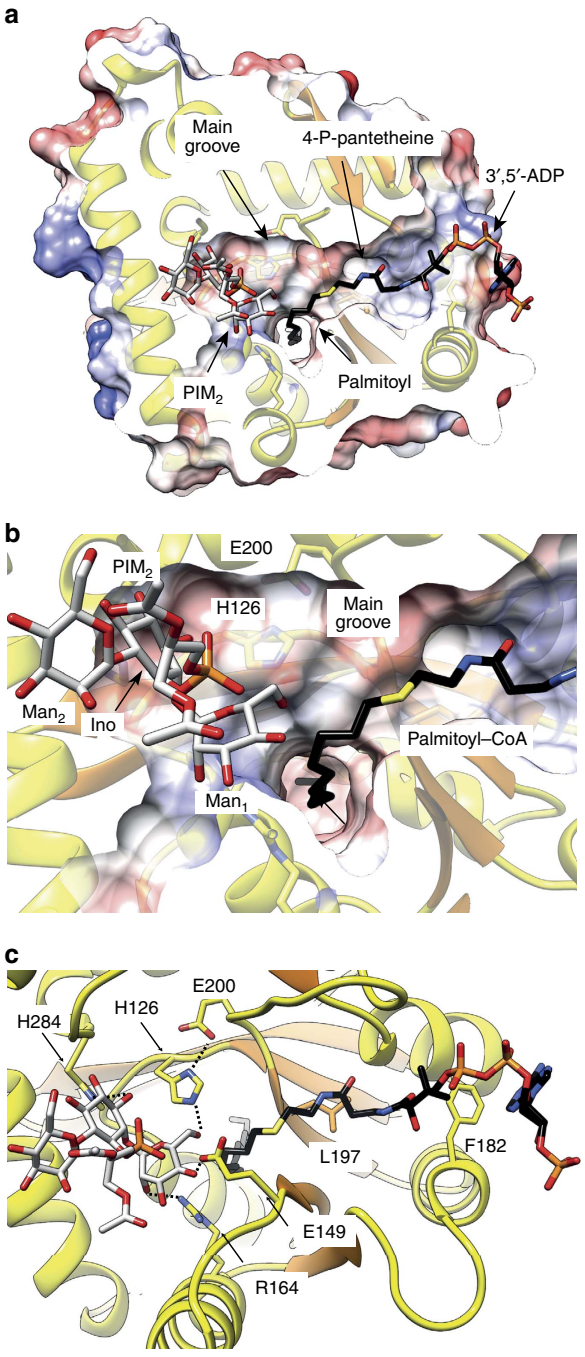



**Figure 6 | The phosphatidylinositol mannosides binding site.** (**a**,**b**) Two views showing the docking calculations in which $PIM_2$ attaches to the end side of the main groove, close to the hydrophobic tunnel. (**c**) Close view of the binding site, showing the location of predicted residues involved in $PIM_2$ interaction that have been mutated (H126, E149, R164, H284) along with mutated residues involved in CoA binding (L197 and F182).

acyl–CoA to the main groove of PatA, are structurally equivalent to $\alpha_{11}$ (residues 201–222) and $\alpha_{12}$ (residues 251–265) in *Cm*GPAT (Fig. 7e). The substitution of L261 ($\alpha_{12}$) by an aromatic residue, which is located at the interface of $\alpha_{11}$–$\alpha_{12}$ helices, caused major changes in the selectivity of *Cm*GPAT for acyl–ACP derivatives[44]. Overall, major secondary structural elements are structurally preserved across both families of acyltransferases supporting a common binding mode for palmitoyl–CoA.

**The catalytic mechanism of PatA**. Analysis of the cleft running over the surface of PatA revealed a catalytic site reminiscent to that observed in the serine protease family of enzymes. In serine proteases, the cleavage of the peptide bond is mediated by nucleophilic attack of the serine hydroxyl group on the scissile carbonyl bond. The active site comprises a catalytic triad consisting of the Oγ atom of the serine, the imidazole ring of a histidine, and the carboxylate group of an aspartic/glutamic acid, involved in a charge relay system that increases the nucleophilicity of the serine hydroxyl and modulates the pKa of the central histidine as a general base or acid during the catalytic cycle[45]. In all crystal structures of palmitoyl–PatA and PatA–S-C16CoA complexes, the carboxylate OE2 oxygen atom of E200 was found at 2.8 Å of the ND1 nitrogen atom of the aromatic imidazole ring of the invariant H126 (Fig. 8a). It is worth noting that when the individual palmitate molecules observed in the three crystal structures of palmitoyl–PatA were superimposed, they showed important structural flexibility at the carboxylate region (Supplementary Fig. 3c), suggesting that the carboxylate group is not in a catalytically competent position in the crystal structures. Moreover, the palmitate is not a substrate neither an inhibitor of the reaction catalysed by PatA (Supplementary Fig. 12). Our binary complexes correspond most likely to the hydrolysis reaction product of palmitoyl–CoA, thus one of the oxygen atoms found in the palmitate that interacts with H126 most likely come from a water molecule activated by H126. The S-C16CoA coordinates in PatA–S-C16CoA crystal structure were used to generate palmitoyl–CoA atomic coordinates by substitution of C16 atom with a carbonyl group followed by energy minimization. Docking calculations placed the Man*p* moiety attached to the 2-position of *myo*-inositol in $PIM_2$ with its O6 atom favourably positioned to receive the palmitate group from palmitoyl–CoA (Fig. 8b; see Online Methods for details). We propose a model in which H126 acts as the general base to abstract a proton from the hydroxyl group at position 6 of the Man*p* ring linked to the 2-position of inositol in $PIM_1$ or $PIM_2$, to facilitate the nucleophilic attack on the thioester of palmitoyl–CoA (Fig. 8d). The E200 gets involved in a charge relay system with H126 and the HO6 atom from the Man*p* moiety, contributing to the appropriate structural arrangement of the imidazole ring of the histidine residue and modulating its pKa to act as a base in the first step and as an acid in the second step, providing protonic assistance to the departing CoA leaving group. It is worth noting that the H126-E200 hydrogen bond was found in a *syn* orientation relative to the carboxylate[45]. The result of the nucleophilic attack is a covalent bond between the mannose ring of $PIM_1$ or $PIM_2$ and palmitate. As depicted in Fig. 5b, the functional role of H126 and E200 was clearly confirmed, since their substitution by alanine completely inactivated the enzyme (Supplementary Fig. 9).

In *Cm*GPAT, the catalytic site is also located at the base of the large groove of the protein, displaying the sequence $HX_4D$, a well-conserved motif among the GPAT family of acyltransferases[27]. H139 and the adjacent D144 were proposed to promote a charge-relay system to facilitate the nucleophilic attack on the thioester of the acyl–CoA[26,37,38,46]. In PatA, H126 occupies the same location to that observed for H139 in *Cm*GPAT. However, the aspartic acid D131 of the $HX_4D$ motif displays a completely different arrangement when compared with D144 (Fig. 8c). Specifically, the carboxylate group of D131 makes a strong hydrogen bond with the side chain OH of Y83, and additional electrostatic interactions with the lateral chains of W130, Y163 and R164, suggesting a structural role for this residue. In that sense, the replacement of D131 by alanine preserved 37% of the enzymatic activity for $PIM_2$ (Fig. 5b;

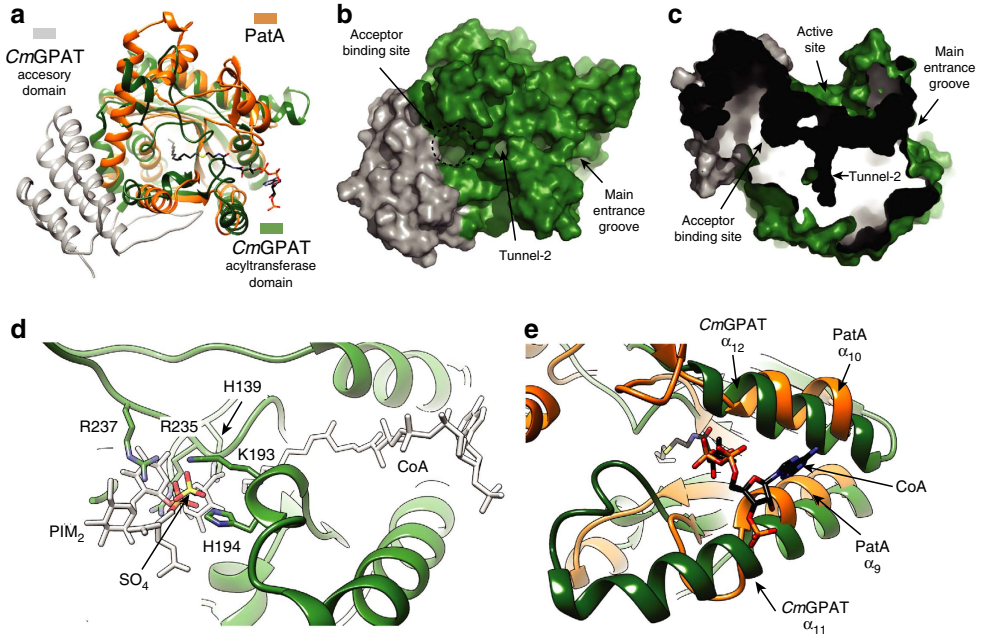



**Figure 7 | Structural similarities between PatA and *Cm*GPAT.** (**a**) Structural superposition of PatA and *Cm*GPAT. Secondary structure elements are labelled. The localization of the fatty acid and acceptor binding sites in *Cm*GPAT are indicated. (**b**,**c**) The active site of *Cm*GPAT is located into a main groove, with a hydrophobic tunnel running perpendicular and deeply buried into the core of the enzyme. (**d**) Structural superposition of the proposed acceptor binding site in PatA and *Cm*GPAT, showing donor and acceptor PatA substrates for spatial reference in white colour. (**e**) Structural superposition of the CoA binding site in PatA and *Cm*GPAT.

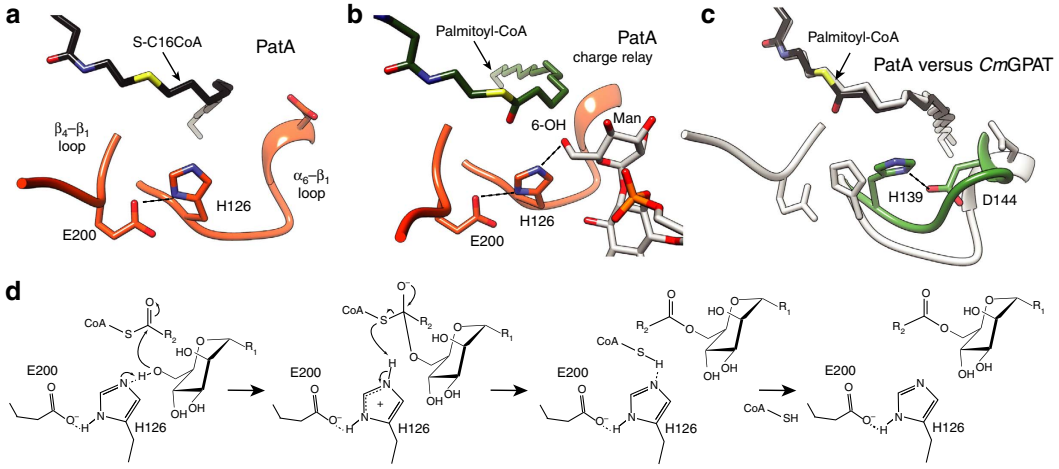


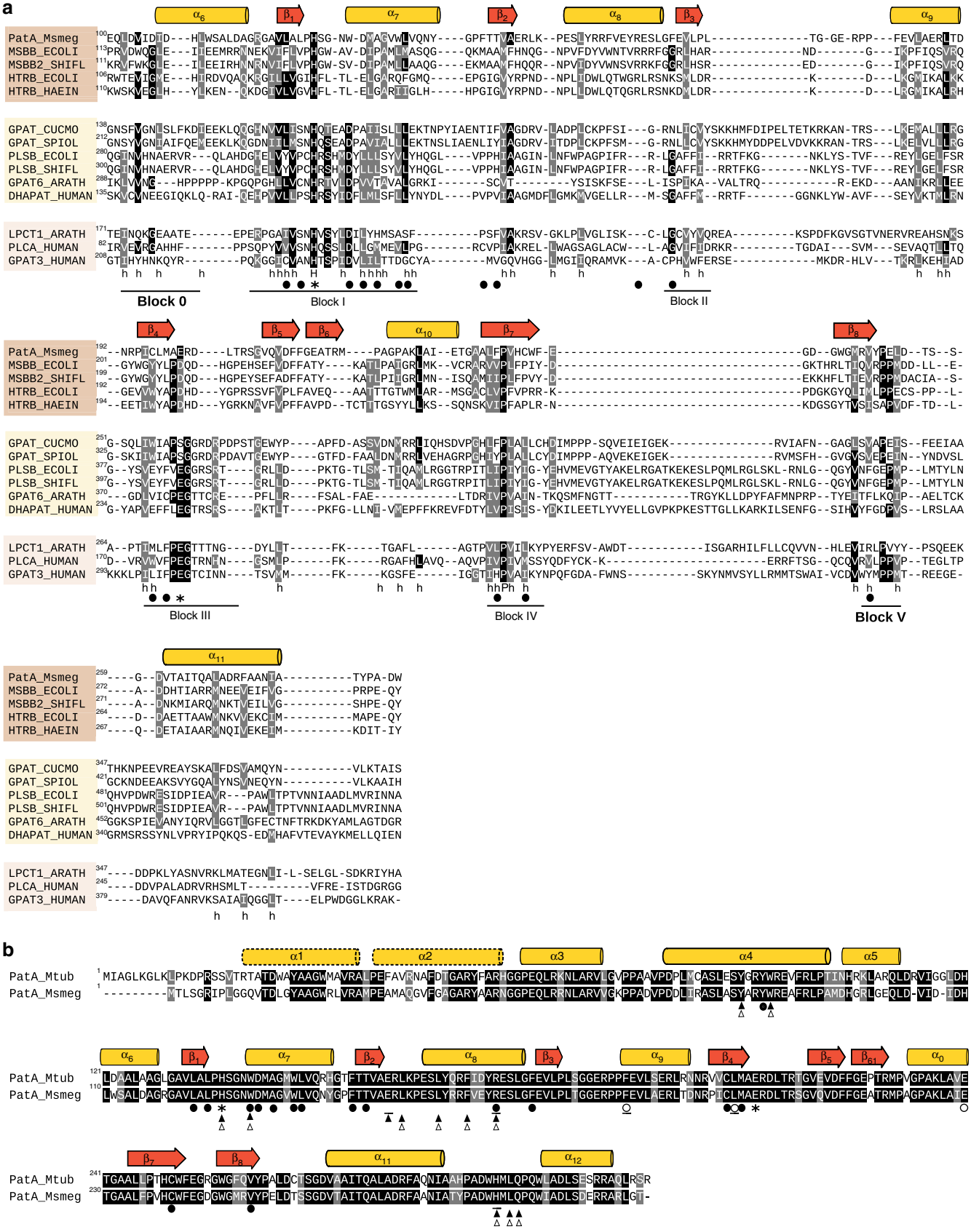

**Figure 8 | Catalytic mechanism of PatA.** (**a**) Active site configuration of PatA, as visualized in the crystal structure in complex with S-C16CoA. (PatA–S-C16CoA) (**b**) Active site configuration of PatA as visualized in the docking of palmitoyl-CoA and $PIM_2$ (the atomic coordinates of palmitoyl-CoA were derived from S-C16CoA following *in silico* addition of an oxygen atom at postion C16 of S-C16CoA). The charged relay system formed between E200, H126 and the 6-OH of the mannose linked to the 2-position of inositol (ino) is highlighted. (**c**) Structural superposition of the catalytic site of PatA (grey) and *Cm*GPAT (green) with palmityol-CoA docked into *Cm*GPAT. (**d**) Proposed catalytic mechanism for PatA.

**Figure 9 | Structure weighted sequence alignment of PatA with eukaryotic/prokaryotic acyltransferases.** (**a**) The protein sequences were extracted from UniProt accession numbers A0QWG5 from *M. smegmatis* (PatA_Msmeg); P24205 from *Escherichia coli* (MSBB_ECOLI); O06659 from *Shigella flexneri* (MSBB2_SHIFL); P0ACV0 from *E. coli* (HTRB_ECOLI); P45239 from *Haemophilus influenzae* (HTRB_HAEIN); P10349 from *C. moscata* (GPAT_CUCMO); Q43869 from *Spinacia oleracea* (GPAT_SPIOL); P0A7A7 PlsB from *E. coli* (PLSB_ECOLI); Q7UBC6 PlsB from *S. flexneri* (PLSB_SHIFL); O80437 from *Arabidopsis thaliana* (GPAT6_ARATH); O15228 from *Homo sapiens* (DHAPAT_HUMAN); Q8L7R3 from *Arabidopsis thaliana* (LPCT1_ARATH); Q99943 from *H. sapiens* (PLCA_HUMAN); and Q53EU6 from *H. sapiens* (GPAT3_HUMAN). Conserved positions are shown in black and grey background. Conserved hydrophobic residues are indicated with an *h*. The secondary structural elements corresponding to the 3D structure of PatA are shown above the alignment. Catalytic amino acids and those involved in palmitate binding are indicated as asterisks and black circles, respectively. (**b**) Structural similarity of *M. tuberculosis* H37Rv and *M. smegmatis* $mc^2155$ PatA. Catalytic amino acids are indicated as asterisks. Amino acids involved in palmitate and pantotheinate binding are indicated as black and white circles, respectively. Residues proposed to interact with $PIM_1$ and $PIM_2$ based on the dockings are shown as white and black triangles, respectively. Amino acids mutated are underlined.

Supplementary Fig. 9). Altogether, the structural information strongly supports a common catalytic mechanism for both families of enzymes.

## Discussion

Amino-acid sequence alignment revealed that PatA has strong resemblance to HtrB and MsbB, two key acyltransferases involved



in the biosynthesis of bacterial lipopolysaccharides (Fig. 9a)[47,48]. HtrB and MsbB catalyse the last steps of Kdo2-lipid A biosynthesis in Gram-negative bacteria, consecutively adding the secondary lauroyl and myristoyl residues to the distal glucosamine unit. Both enzymes prefer acyl–ACP donors but can also function with acyl–CoA substrates[7,8]. The fact that HtrB, MsbB and PatA preferentially use lauroyl (C12), myristoyl (C14) or palmitoyl (C16) derivatives might suggest the occurrence of a hydrocarbon ruler mechanism for acyl moieties recognition. PatA, HtrB and MsbB enzymes are distantly related to the glycerol-3-phosphate (GPAT), lysophosphatidic acid (LPAAT), dihydroacetone phosphate (DHAPAT) and 2-acylglycero-phosphatidylethanolamine (LPEAT) families of acyltransferases (Fig. 9)[5,27]. GPAT, LPAAT, DHAPAT, and LPEAT display highly conserved residues distributed in four regions, and named as blocks 1–4 (refs 27,41). According to multiple sequence alignments among the HtrB, MsbB, GPAT, LPAAT, DHAPAT and LPEAT families, weighted by structural alignment of PatA and *Cm*GPAT, a common core can be defined (Fig. 9). It is worth noting that we respected the classical names of blocks 1–4 as reported in the literature, introducing two new regions as blocks 0 and 5. Critical residues and their interactions in the reaction centre are essentially preserved in all these acyltransferase families, strongly supporting a common catalytic mechanism. Interestingly, the alignment suggests that members of the GPAT, LPAAT, DHAPAT and LPEAT families display a conserved aspartate residue that participates in the charge relay system with the conserved histidine of the $HX_4D$ motif (block 1)[26,49]. In contrast, HtrB, MsbB and PatA seem to use a glutamate/aspartate residue located in block 3, suggesting a divergence among these families of acyltransferases. The structural divergence of the carboxylate group acting as a pKa modulator of the catalytic histidine residue might be due to the requirement of the acyltransferases to accommodate acceptor molecules of different nature, as observed in PatA and *Cm*GPAT. Nevertheless, the hydrophobic nature of residues involved in fatty-acid recognition is well conserved, suggesting a common binding mode.

Finally, $Ac_1PIM_2$ appears to be a metabolic end product that accumulates at high steady-state levels in the cells as well as a precursor for more polar forms of PIMs, lipomannan and LAM[10]. Interestingly, the four enzymes involved in the biosynthetis of $Ac_1PIM_2$, the phosphatidyl-*myo*-inositol synthase PgsA1 (Rv2612c in *M. tuberculosis* H37Rv), PimA (Rv2610c), PimB (Rv2188c in *Mtb* H37Rv) and PatA (Rv2611c), were found to be essential for the growth of *M. smegmatis* and/or *M. tuberculosis*[10,12,17,50]. Importantly, the amino-acid sequences of *M. smegmatis* and *M. tuberculosis* versions of PatA displayed 74% identity and 84% similarity (Fig. 9b). All residues that participate in the catalytic mechanism and palmitoyl binding, as well as those proposed to interact with the CoA and $PIM_2$ substrates are strictly conserved between both proteins. Thus, the structural data presented here offers exciting possibilities for inhibitor design and the discovery of chemotherapeutic agents against this major human pathogen.

## Methods

**Expression and purification of PatA in *M. smegmatis*.** A truncated version of PatA lacking the first 12 residues of the protein (PatA; MSMEG_2934)[28], was purified as previously described with the following modifications. *M. smegmatis* $mc^2155$ cells transformed with the corresponding plasmid pJAM2-*patA* were grown in MM63 medium (15 mM $(NH4)_2SO_4$, 10 mM $KH_2PO_4$, 18 μM $FeSO_4.7H_2O$, pH 7.0) supplemented with 1 mM $MgSO_4$, 0.025% (v/v) tyloxapol and 0.2% (w/v) succinate and 20 μg ml$^{-1}$ of kanamycin. When the culture reached $OD_{600} = 0.6$ the expression of PatA was induced by adding 0.2% acetamide. After 16 h at 37 °C, cells were collected at 4,000*g* for 10 min and resuspended in 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 40 mM imidazole (solution A) containing protease inhibitors (Complete EDTA-free, Roche). The cells were resuspended in solution A (1 g of cells per 5 ml of solution A) and disrupted by sonication in 15 cycles of 60 s pulses with 90 s cooling intervals between the pulses. PatA was solubilized from the mycobacterial membrane by the addition of 2 mM CHAPS. The suspension was gently stirred during 1 h at 4 °C and centrifuged at 11,000*g* for 20 min. The supernatant was applied to a HisTrap Chelating column (1 ml, GE HealthCare) equilibrated with solution A. The column was then washed with solution A until no absorbance at 280 nm was detected. Elution was performed with a linear gradient of 40–500 mM imidazole in solution A at 1 ml min$^{-1}$. The resulting PatA preparation displayed a single protein band when run on a 12% NuPAGE Bis–Tris precast gel stained with SimplyBlue SafeStain (Invitrogen). The purified recombinant PatA protein was stored at 4 °C and then concentrated for crystallization screening by using a Vivaspin 20 spin concentrator (Vivascience) with a 10-kDa-molecular mass cutoff.

**Site-directed mutagenesis.** The PatA-H126A, PatA-D131A, PatA-E149A, PatA-R164A, PatA-E200A and PatA-H284A mutants, and double mutant PatA-F182W/L197W were synthetized by GenScript using the pJAM2-*patA* construct, and further expressed and purified to apparent homogeneity as described for the recombinant PatA enzyme.

**PatA–C16 complex crystallization and data collection.** Three crystal forms were obtained, referred thereafter as PatA–C16-1, PatA–C16-2 and PatA–C16-3. The first and second crystal forms were obtained by mixing 0.25 μl of PatA at 5 mg ml$^{-1}$ in 20 mM Tris-HCl pH 7.5 with 0.25 μl of mother liquor containing 100 mM Tris-HCl pH 7.0, 230 mM $MgCl_2$ and 12–16% (w/v) PEG 8,000. Crystals grew in 7–15 days and were transferred to a cryo-protectant solution containing 25% ethylene glycol and frozen under liquid nitrogen. Complete X-ray diffraction data sets were collected at beamline I03 (Diamond Light Source, Oxfordshire, UK) and processed with XDS program[51]. The second crystal form of PatA diffracted to a maximum resolution of 2.9 Å and crystallized with four molecules in the asymmetric unit and space group $P\,2_1$ (Supplementary Table 1; PatA–C16-2). The first crystal form of PatA diffracted to a maximum resolution of 2.06 Å and crystallized with two molecules in the asymmetric unit and space group $C\,2$ (Supplementary Table 1; PatA–C16-1). The third crystal form was obtained by mixing 0.25 μl of PatA at 5 mg ml$^{-1}$ in 25 mM Tris-HCl pH 7.5, 150 mM NaCl with 0.25 μl of mother liquor containing 100 mM Tris-HCl pH 8.5 and 20% ethanol. Crystals grew in 3 days and were transferred to a cryo-protectant solution containing 25% sucrose and frozen under liquid nitrogen. A complete set of X-ray diffraction data were collected at beamline X06SA Swiss Light Source (Villigen, Switzerland) and processed with XDS program. PatA–C16-3 crystals diffracted to a maximum resolution of 2.43 Å and crystallized with one molecule in the asymmetric unit and space group $P\,4_2\,2_1\,2$ (Supplementary Table 1).

**PatA-S-C16CoA complex crystallization and data collection.** One crystal form was obtained by mixing 0.25 μl of PatA at 5.1 mg ml$^{-1}$ in 1 mM *S*-hexadecyl Coenzyme A (S-C16CoA; stock solution at 10 mM in 20 mM Tris-HCl pH 7.5) and 20 mM Tris-HCl pH 7.5 with 0.25 μl of mother liquor containing 100 mM HEPES pH 7.5, 500 mM ammonium sulfate and 30% (v/v) 2-methyl-2,4-pentanediol. Crystals grew in 7–15 days and were directly frozen under liquid nitrogen. A complete X-ray diffraction data set was collected at beamline I03 (Diamond Light Source) and processed with XDS program[51]. PatA-S-C16CoA diffracted to a maximum resolution of 3.28 Å and crystallized with four molecules in the asymmetric unit and space group $P\,2_1$ (Supplementary Table 1).

**PatA structure determination and refinement.** PatA crystals of form $C\,2$ (PatA–C16-1) were soaked with 10 different platinum salts at 1 mM concentration for a time period of 130–145 min (HR2-442, Hampton Research) followed by flash freezing in liquid-$N_2$. Anomalous data were collected at the theoretical L–I absorption edge of Pt (13,879.9 eV–0.8933 Å). Oscillation images were collected at I04 beamline (Diamond Light Source) with an oscillation angle of 0.2 for a total of 1,800 images using a Pilatus 6M-F pixel detector. Data were collected with an attenuated X-ray beam (5% transmission) and a 0.04-s exposure time per image. Data were integrated and scaled in XDS and experimental phases determined using the SHELXC/D/E package[52]. Data of a PatA crystal soaked with 1 mM $K_2PtCl_4$ for 130 min were used for experimental phasing with a 2.5-Å data cutoff applied, giving a mean value $|\Delta F|/\sigma(\Delta F)$ in the highest resolution shell of 0.9. The substructure determination located two Pt atoms in the asymmetric unit (CC = 33.64, CC(weak) = 20.71 and CFOM = 54.35). Experimental phases were determined and subsequently used for initial cycles of model building and density modifications by SHELXE. Buccaneer[53] and the CCP4 suite[54] were used for further model extention. The structure determination of PatA–C16-1, PatA–C16-2, PatA–C16-3 and PatA–S-C16CoA were carried out by molecular replacement using Phaser[55] and the PHENIX suite[56] and the PatA–Pt structure as model (Supplementary Table 2). Followed by cycles of manual rebuilding and refinement using Coot[57] and phenix.refine[58], respectively. The structures were validated by MolProbity[59].

**Molecular docking calculations.** The crystal structure of mycobacterial PatA in complex with S-C16CoA (PatA–S-C16CoA) was investigated using the structure

preparation function in MOE2013.08 (ref. 60). First, S-C16CoA coordinates in PatA–S-C16CoA crystal structure (chain A) were used to generate palmitoyl–CoA atomic coordinates by substitution of C16 atom with a carbonyl group. Then the model was prepared using the Amber12EHT force field, an all-atom force field, combining two-dimensional Extended Hueckel Theory (EHT) and Amber12 force field, with Born solvation, and hydrogen atoms were added using Protonate3D function[61,62]. The docking site of $PIM_1$ and $PIM_2$ was defined by 14 residues of the PatA–S-C16CoA crystal structure. $PIM_2$ structure was retrieved from the PDB (ligand code XPX) and after substitution of acyl chains with acetyl groups, the molecule was energy minimized using MOE with a 0.1 kcal $mol^{-1}$ $Å^{-1}$ r.m.s. gradient threshold. $PIM_1$ structure was constructed by removing one mannose ring of $PIM_2$. The $PIM_1$ and $PIM_2$ structures were submitted to conformational search using LowModeMD with default parameters in MOE2013.08 (ref. 63). The same procedure was carried out with $PIM_1$, in which a mannose residue GOLD (Genetic Optimization for Ligand Docking; Cambridge Crystallographic Data Center (CCDC), version 5.2.2) was used with default genetic algorithm parameter settings[64,65] for all calculations, with the exception that the search efficiency parameter was set to 200% to improve predictive accuracy by calculating the optimal number of genetic algorithm operations for the ligand due to their large flexibility. The ASP scoring function implemented in GOLD was used to rank the docked poses; this fitness function has been optimized for the prediction of ligand-binding positions[66,67]. $PIM_1$ and $PIM_2$ were docked into the PatA–palmitoyl–CoA complex. Only docking poses having a C6-hydroxyl group of the mannose closed to the thioester of palmitoyl–CoA were kept for analysis. The best solutions were assessed by their respective docking score and by visual inspection.

**PatA acyltransferase activity assay.** PatA transferase activity was measured in the assay with mycobacterial membranes[28]. Briefly, *M. smegmatis* $mc^2155$ cells were broken by sonication and the membrane (100,000*g* pellet) fraction was obtained by differential centrifugation. The reaction mixtures contained 250 µg of membrane proteins, 1.2 µg of purified $PimA_{SM}$, 10 µg of purified PatA or the mutated versions, 0.1 µCi GDP-[$^{14}$C]mannose (specific activity of 55 mCi $mmol^{-1}$, ARC Inc.), 0.12 mM palmitoyl–CoA (Sigma-Aldrich) in DMSO with final concentration in the reaction mixture 2% (v/v), 62 µM ATP, 10 mM $MgCl_2$, and 25 mM Tris-HCl pH 7.5 in the final volume of 50 µl. Reactions were incubated 100 min at 37 °C and stopped with 300 µl of $CHCl_3$/$CH_3OH$ (2:1, by volume). The samples were left rocking 30 min at room temperature, and centrifuged at 1,500*g* for 10 min. The organic phase (bottom) was analysed by thin layer chromatography on aluminium-coated silica 60 $F_{254}$ plates (Merck) developed in $CHCl_3$/$CH_3OH$/conc. $NH_4OH$/$H_2O$ (65:25:0.5:4), and quantified by scintillation spectrometry[28]. All enzymatic activity measurements were determined in duplicates. Following the same procedure, palmitate was assayed as a possible inhibitor or substrate at different concentrations (Supplementary Fig. 12).

**PatA palmitoyl-CoA hydrolytic activity assay.** The hydrolytic activity of PatA and PatA variants against palmitoyl–CoA was measured as following a methodology described for other acyltransferases[68,69]. A typical reaction contained 20 mM Tris-HCl pH 8.3, 0,2 mM disulfide (5,5′-dithiobis-(2-nitrobenzoic acid)) (DTNB), 0,06 mM palmitoyl–CoA and 4 µM PatA or its variants. All reactions were carried out at 37 °C in a CARY 300 Bio UV Visible/Spectrophotometer. The spectrum of the product $TNB^{-2}$ formed after the interaction of the DTNB and the CoA liberated from the hydrolysis of the substrate palmitoyl–CoA by the enzyme, was measured continuously at 412 nm during 20 min. All enzymatic activity measurements were determined in duplicates (Fig. 5).

**Structural analysis and sequence alignment.** The sequence of PatA from *M. smegmatis* (A0QWG5) was subjected to basic local alignment search tool (BLAST) and several orthologs in *Mycobacteria* sp. were found. Afterwards, they were aligned using the ClustalW server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The structure weighted sequence alignment was performed using PROMALS3D server (http://prodata.swmed.edu/promals3d/promals3d.php). For labelling the conserved and similar residues, BoxShade server was used (http://embnet.vital-it.ch/software/BOX_form.html). Structural analysis and graphics for publications were performed with PyMOL (version 0.99) and Chimera[70].

## References


1. White, S. W., Zheng, J., Zhang, Y. M. & Rock, C. O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **74,** 791–831 (2005).
2. de Mendoza, D. Temperature sensing by membranes. *Annu. Rev. Microbiol.* **68,** 101–116 (2014).
3. Grevengoed, T. J., Klett, E. L. & Coleman, R. A. Acyl-CoA metabolism and partitioning. *Annu. Rev. Nutr.* **34,** 1–30 (2014).
4. Cronan, J. E. Bacterial membrane lipids: where do we stand? *Annu. Rev. Microbiol.* **57,** 203–224 (2003).
5. Röttig, A. & Steinbüchel, A. Acyltransferases in bacteria. *Microbiol. Mol. Biol. Rev.* **77,** 277–321 (2013).
6. Chan, D. I. & Vogel, H. J. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem. J.* **430,** 1–19 (2010).
7. Raetz, C. R., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* **76,** 295–329 (2007).
8. Whitfield, C. & Trent, M. S. Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem.* **83,** 99–128 (2014).
9. Kaur, D., Guerin, M. E., Skovierova, H., Brennan, P. J. & Jackson, M. Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*. *Adv. Appl. Microbiol.* **69,** 23–78 (2009).
10. Guerin, M. E., Korduláková, J., Alzari, P. M., Brennan, P. J. & Jackson, M. Molecular basis of phosphatidyl-*myo*-inositol mannoside biosynthesis and regulation in mycobacteria. *J. Biol. Chem.* **285,** 33577–33583 (2010).
11. Morita, Y. S. *et al.* Inositol lipid metabolism in mycobacteria: biosynthesis and regulatory mechanisms. *Biochim. Biophys. Acta* **1810,** 630–641 (2011).
12. Korduláková, J. *et al.* Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. *J. Biol. Chem.* **277,** 31335–31344 (2002).
13. Guerin, M. E. *et al.* New insights into the early steps of phosphatidylinositol mannoside biosynthesis in mycobacteria: PimB' is an essential enzyme of *Mycobacterium smegmatis*. *J. Biol. Chem.* **284,** 25687–25696 (2009).
14. Doz, E. *et al.* Mycobacterial phosphatidylinositol mannosides negatively regulate host Toll-like receptor 4, MyD88-dependent proinflammatory cytokines, and TRIF-dependent co-stimulatory molecule expression. *J. Biol. Chem.* **284,** 23187–23196 (2009).
15. Cala-De Paepe, D. *et al.* Deciphering the role of CD1e protein in mycobacterial phosphatidyl-myo-inositol mannosides (PIM) processing for presentation by CD1b to T lymphocytes. *J. Biol. Chem.* **287,** 31494–31502 (2012).
16. Hill, D. L. & Ballou, C. E. Biosynthesis of mannophospholipids by *Mycobacterium phlei*. *J. Biol. Chem.* **241,** 895–902 (1966).
17. Boldrin, F. *et al.* The phosphatidyl-*myo*-inositol mannosyltransferase PimA is essential for *Mycobacterium tuberculosis* growth in vitro and in vivo. *J. Bacteriol.* **196,** 3441–3451 (2014).
18. Lea-Smith, D. J. *et al.* Analysis of a new mannosyltransferase required for the synthesis of phosphatidylinositol mannosides and lipoarbinomannan reveals two lipomannan pools in corynebacterineae. *J. Biol. Chem.* **283,** 6773–6782 (2008).
19. Korduláková, J. *et al.* Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species. *J. Biol. Chem.* **278,** 36285–36295 (2003).
20. Gilleron, M. *et al.* Acylation state of the phosphatidylinositol mannosides from Mycobacterium bovis bacillus Calmette Guérin and ability to induce granuloma and recruit natural killer T cells. *J. Biol. Chem.* **276,** 34896–34904 (2001).
21. Gilleron, M., Quesniaux, V. F. & Puzo, G. Acylation state of the phosphatidylinositol hexamannosides from Mycobacterium bovis bacillus Calmette Guerin and *Mycobacterium tuberculosis* H37Rv and its implication in Toll-like receptor response. *J. Biol. Chem.* **278,** 29880–29889 (2003).
22. Guerin, M. E. *et al.* Molecular recognition and interfacial catalysis by the essential phosphatidylinositol mannosyltransferase PimA from mycobacteria. *J. Biol. Chem.* **282,** 20705–20714 (2007).
23. Batt, S. M. *et al.* Acceptor substrate discrimination in phosphatidyl-*myo*-inositol mannoside synthesis: structural and mutational analysis of mannosyltransferase *Corynebacterium glutamicum* PimB'. *J. Biol. Chem.* **285,** 37741–37752 (2010).
24. Albesa-Jové, D., Giganti, D., Jackson, M., Alzari, P. M. & Guerin, M. E. Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology* **24,** 108–124 (2014).
25. Giganti, D. *et al.* Secondary structure reshuffling modulates glycosyltransferase function at the membrane. *Nat. Chem. Biol.* **11,** 16–18 (2015).
26. Heath, R. J. & Rock, C. O. A conserved histidine is essential for glycerolipid acyltransferase catalysis. *J. Bacteriol.* **180,** 1425–1430 (1998).
27. Lewin, T. M., Wang, P. & Coleman, R. A. Analysis of amino acid motifs diagnostic for the *sn*-glycerol-3-phosphate acyltransferase reaction. *Biochemistry* **38,** 5764–5771 (1999).
28. Svetlíková, Z., Baráth, P., Jackson, M., Korduláková, J. & Mikušová, K. Purification and characterization of the acyltransferase involved in biosynthesis of the major mycobacterial cell envelope glycolipid-monoacylated phosphatidylinositol dimannoside. *Protein Expr. Purif.* **100,** 33–39 (2014).
29. Blobel, G. Intracellular protein topogenesis. *Proc. Natl Acad. Sci. USA* **77,** 1496–1500 (1980).
30. McLaughlin, S. The electrostatic properties of membranes. *Annu. Rev. Biophys. Biophys. Chem.* **18,** 113–136 (1989).
31. White, S. H. & Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 319–365 (1999).
32. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98,** 10037–10041 (2001).
33. Brennan, P. J. & Ballou, C. E. Biosynthesis of mannophosphoinositides by *Mycobacterium phlei*. The family of dimannophosphoinositides. *J. Biol. Chem.* **242,** 3046–3056 (1967).

34. Brennan, P. J. & Ballou, C. E. Biosynthesis of mannophosphoinositides by *Mycobacterium phlei*. Enzymatic acylation of the dimannophosphoinositides. *J. Biol. Chem.* **243,** 2975–2984 (1968).
35. Thoden, J. B., Zhuang, Z., Dunaway-Mariano, D. & Holden, H. M. The structure of 4-hydroxybenzoyl-CoA thioesterase from arthrobacter sp. strain SU. *J. Biol. Chem.* **278,** 43709–43716 (2003).
36. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10,** 168 (2009).
37. Turnbull, A. P. *et al.* Analysis of the structure, substrate specificity, and mechanism of squash glycerol-3-phosphate (1)-acyltransferase. *Structure* **9,** 347–353 (2001).
38. Tamada, T. *et al.* Substrate recognition and selectivity of plant glycerol-3-phosphate acyltransferases (GPATs) from *Cucurbita moscata* and *Spinacea oleracea*. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 13–21 (2004).
39. Murata, N. & Tasaka, Y. Glycerol-3-phosphate acyltransferase in plants. *Biochim. Biophys. Acta* **1348,** 10–16 (1997).
40. Hayman, M. W., Fawcett, T. & Slabas, A. R. Kinetic mechanism and order of substrate binding for sn-glycerol-3-phosphate acyltransferase from squash (*Cucurbita moschata*). *FEBS Lett.* **514,** 281–284 (2002).
41. Yao, J. & Rock, C. O. Phosphatidic acid synthesis in bacteria. *Biochim. Biophys. Acta* **1831,** 495–502 (2013).
42. Wendel, A. A., Lewin, T. M. & Coleman, R. A. Glycerol-3-phosphate acyltransferases: rate limiting enzymes of triacylglycerol biosynthesis. *Biochim. Biophys. Acta* **1791,** 501–506 (2009).
43. Zhang, Y. M. & Rock, C. O. Thematic review series: glycerolipids. Acyltransferases in bacterial glycerophospholipid synthesis. *J. Lipid Res.* **49,** 1867–1874 (2008).
44. Slabas, A. R. *et al.* Squash glycerol-3-phosphate (1)-acyltransferase. Alteration of substrate selectivity and identification of arginine and lysine residues important in catalytic activity. *J. Biol. Chem.* **277,** 43918–43923 (2002).
45. Hedstrom, L. Serine protease mechanism and specificity. *Chem. Rev.* **102,** 4501–4524 (2002).
46. Bellizzi, III J. J. *et al.* The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc. Natl Acad. Sci. USA* **97,** 4573–4578 (2000).
47. Clementz, T., Bednarski, J. J. & Raetz, C. R. Function of the *htrB* high temperature requirement gene of Escherichia coli in the acylation of lipid A: HtrB catalyzed incorporation of laurate. *J. Biol. Chem.* **271,** 12095–12102 (1996).
48. Clementz, T., Zhou, Z. & Raetz, C. R. Function of the *Escherichia coli msbB* gene, a multicopy suppressor of *htrB* knockouts, in the acylation of lipid A. Acylation by MsbB follows laurate incorporation by HtrB. *J. Biol. Chem.* **272,** 10353–10360 (1997).
49. Neuwald, A. F. Barth syndrome may be due to an acyltransferase deficiency. *Curr. Biol.* **7,** R465–R466 (1997).
50. Jackson, M., Crick, D. C. & Brennan, P. J. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J. Biol. Chem.* **275,** 30092–30099 (2000).
51. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 125–132 (2010).
52. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 479–485 (2010).
53. Cowtan, K. The Buccaneer software for automated model building1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011 (2006).
54. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242 (2011).
55. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr* **40,** 658–674 (2007).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
57. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
58. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 352–367 (2012).
59. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 (2010).
60. Molecular Operating Environment (MOE). 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7 (2015).
61. Gerber, P. R. & Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9,** 251–268 (1995).
62. Case, D. A. *et al. AMBER 12* (University of California, San Francisco, 2012).
63. Labute, P. LowModeMD--implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops. *J. Chem. Inf. Model.* **50,** 792–800 (2010).
64. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267,** 727–748 (1997).
65. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins* **52,** 609–623 (2003).
66. Mooij, W. T. M. & Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins* **61,** 272–287 (2005).
67. Cheng, T., Li, X., Li, Y., Liu, Z. & Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **49,** 1079–1093 (2009).
68. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82,** 70–77 (1959).
69. Rodgers, Jr J. B. Assay of acyl-CoA:monoglyceride acyltransferase from rat small intestine using continuous recording spectrophotometry. *J. Lipid Res.* **10,** 427–432 (1969).
70. Pettersen, E. F. *et al.* UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).


## Acknowledgements

This work was supported by the European Commission Contract HEALTH-F3-2011-260872, the Spanish Ministry of Economy and Competitiveness Contract BIO2013-49022-C2-2-R, and the Basque Government (to M.E.G.); Slovak Research and Development Agency Contract No. DO7RP-0015-11 (to K.M.) and the NIH/NIAID grant AI064798 (to M.J.). D.A.-J. acknowledges the support from Fundación Biofisica Bizkaia. We gratefully acknowledge Sonia López-Fernández (Unit of Biophysics, CSIC,UPV/EHU, Spain), Drs E. Ogando and T. Mercero (Scientific Computing Service UPV/EHU, Spain) for technical assistance. We thank the Swiss Light Source (SLS), and the Diamond Light Source (DLS) for granting access to synchrotron radiation facilities and their staff for the onsite assistance. We specially thank the BioStruct-X project to support access to structural biology facilities. We also acknowledge all members of the Structural Glycobiology Group (Spain) for valuable scientific discussions. The following reagent was obtained through BEI Resources, NIAID, NIH: *Mycobacterium tuberculosis*, Strain H37Rv, Purified Phosphatidylinositol Mannosides 1 and 2 (PIM1,2), NR-14846.

## Author contributions

M.E.G. and K.M. conceived the project. D.A.-J., Z.S., M.T., E.S.-V., A.C.-G., P.B., P.A., A.E., S.K.A. and J.O.C. performed the experiments. D.A.J., E.S., M.J., J.K., K.M. and M.E.G. analysed the results. M.E.G. wrote the manuscript.

## Additional information

**Accession codes:** Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB) with accession codes 5F2T, 5F2Z, 5F31 and 5F34.

**Supplementary Information** accompanies this paper at http://www.nature.com/naturecommunications

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article**: Albesa-Jové, D. *et al.* Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nat. Commun.* 7:10906 doi: 10.1038/ncomms10906 (2016).

# ANNEX III

# ANNEX III

ACS chemical biology

Articles

Cite This: ACS Chem. Biol. XXXX, XXX, XXX−XXX

# The Molecular Mechanism of Substrate Recognition and Catalysis of the Membrane Acyltransferase PatA from Mycobacteria

Montse Tersa,[†,○] Lluís Raich,[‡,○] David Albesa-Jové,[†,§,∥,⊥,○] Beatriz Trastoy,[†] Jacques Prandi,[#] Martine Gilleron,[#] Carme Rovira,[‡,▽] and Marcelo E. Guerin*[,†,§,∥,⊥]

[†]Structural Biology Unit, CIC bioGUNE, Bizkaia Technology Park, 48160 Derio, Spain

[‡]Departament de Química Inorgànica i Orgànica (Secció de Química Orgànica) and IQTCUB, Universitat de Barcelona, 08028 Barcelona, Spain

[§]Unidad de Biofísica, Centro Mixto Consejo Superior de Investigaciones Científicas - Universidad del País Vasco/Euskal Herriko Unibertsitatea (CSIC, UPV/EHU), Barrio Sarriena s/n, Leioa, Bizkaia 48940, Spain

[∥]Departamento de Bioquímica, Universidad del País Vasco, Leioa, Spain

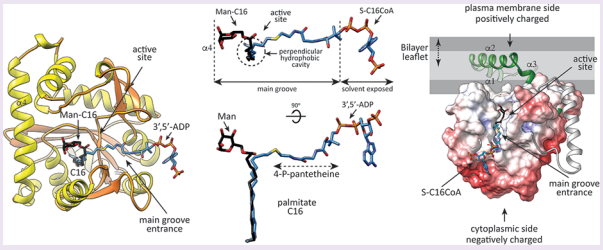
[⊥]IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain

[#]Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, 205 route de Narbonne, F-31077 Toulouse, France

[▽]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08020 Barcelona, Spain

Supporting Information

**ABSTRACT:** Glycolipids play a central role in a variety of important biological processes in all living organisms. PatA is a membrane acyltransferase involved in the biosynthesis of phosphatidyl-*myo*-inositol mannosides (PIMs), key structural elements, and virulence factors of *Mycobacterium tuberculosis*. PatA catalyzes the transfer of a palmitoyl moiety from palmitoyl-CoA to the 6-position of the mannose ring linked to the 2-position of inositol in $PIM_1/PIM_2$. We report here the crystal structure of PatA in the presence of 6-*O*-palmitoyl-$\alpha$-D-mannopyranoside, unraveling the acceptor binding mechanism. The acceptor mannose ring localizes in a cavity at the end of a surface-exposed long groove where the active site is located, whereas the palmitate moiety accommodates into a hydrophobic pocket deeply buried in the $\alpha/\beta$ core of the protein. Both fatty acyl chains of the $PIM_2$ acceptor are essential for the reaction to take place, highlighting their critical role in the generation of a competent active site. By the use of combined structural and quantum-mechanics/molecular-mechanics (QM/MM) metadynamics, we unravel the catalytic mechanism of PatA at the atomic-electronic level. Our study provides a detailed structural rationale for a stepwise reaction, with the generation of a tetrahedral transition state for the rate-determining step. Finally, the crystal structure of PatA in the presence of $\beta$-D-mannopyranose and palmitate suggests an inhibitory mechanism for the enzyme, providing exciting possibilities for inhibitor design and the discovery of chemotherapeutic agents against this major human pathogen.




Glycolipids (GLs) are prominent constituents of biological membranes, primarily localized on the surface of archaeal, prokaryotic, and eukaryotic cells.[1,2] GLs play a major role in molecular recognition events including cell−cell interactions and host−pathogen interactions, as well in the modulation of membrane protein function. Acyltransferases are enzymes that catalyze the transfer of fatty-acid chains from activated fatty-acid donors to a variety of acceptor molecules of different chemical structure and complexity, playing a key role in the biosynthesis of glycolipids.[3] It is worth noting that fatty-acid chains are activated by esterification of their carboxyl groups with either (i) the thiol group of coenzyme A (CoA) or (ii) that of the acyl carrier protein (ACP), yielding acyl-thioesters. Acyltransferases are very often required to access a lipophilic acceptor substrate and to catalyze its reaction with a water-soluble donor in the form of acyl-CoA or acyl-ACP derivatives, promoting their association to the lipid bilayer.[4] Therefore, an understanding of the molecular mechanism by which acyltransferases govern substrate recognition and catalysis remains a major challenge.

Mycobacteria display a characteristic cell envelope based on sugars and lipids of exceptional chemical structure. The cell envelope delineates the biology and pathogenesis of *M. tuberculosis*, the causative agent of tuberculosis (TB), also contributing to its intrinsic resistance to chemotherapeutic agents.[5] Specifically, the cell envelope of mycobacteria comprises four main components: the plasma membrane or

Received: July 9, 2017
Accepted: November 29, 2017
Published: November 29, 2017




DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

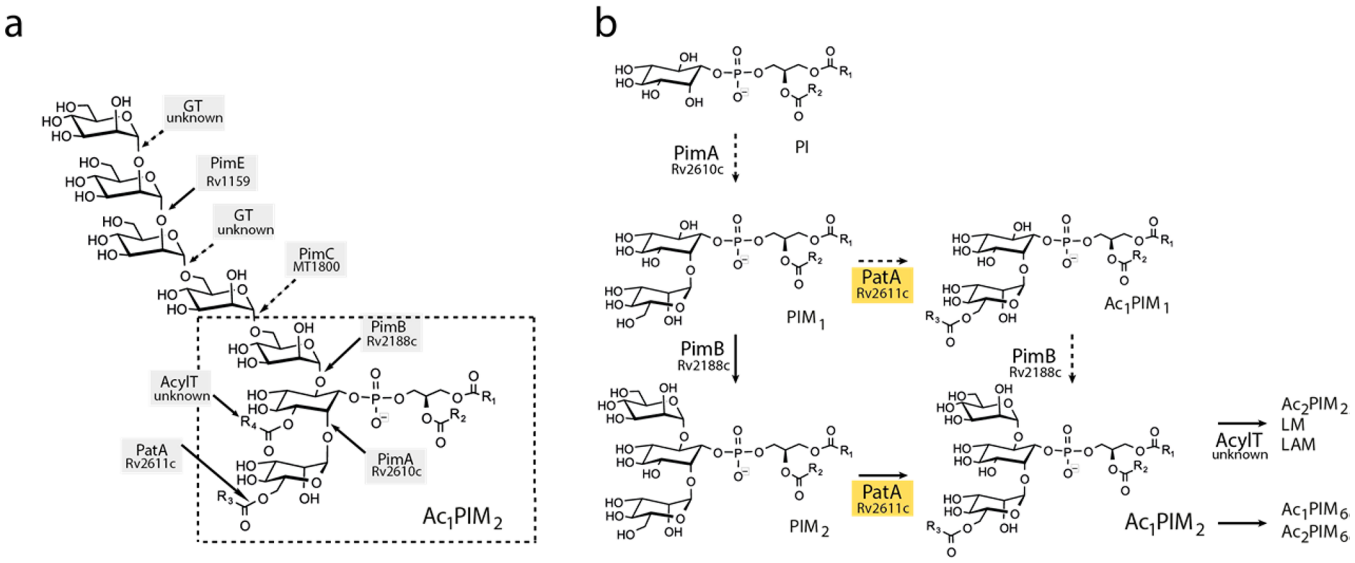



**Figure 1.** PIM biosynthetic pathway in mycobacteria. (a) Chemical structure of $PIM_{2/6}$ and their acylated forms $Ac_{1/2}PIM_2$ and $Ac_{1/2}PIM_6$. (b) The two proposed pathways for the early steps of PIM biosynthesis in mycobacteria are shown: (i) PI is mannosylated to form $PIM_1$. $PIM_1$ is then mannosylated to $PIM_2$, which is acylated to form $Ac_1PIM_2$. (ii) $PIM_1$ is first acylated to $Ac_1PIM_1$ and then mannosylated to $Ac_1PIM_2$. Our experimental evidence suggests that although both pathways might coexist in mycobacteria, the PI → $PIM_1$ → $PIM_2$ → $Ac_1PIM_2$ pathway is favored. H37Rv numbers are shown.

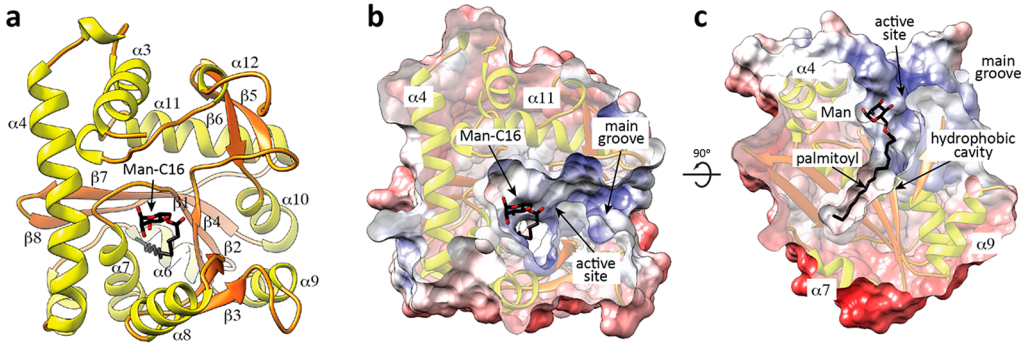



**Figure 2.** Overall structure of the PatA-Man-C16 complex. (a) Cartoon representation showing the general fold of PatA. $\beta$-strands and $\alpha$-helices are shown in orange and yellow, respectively. The Man-C16 product is shown in black and red. (b,c) Two views of the electrostatic surface representation of PatA showing the location of the Man-C16 product. The Man*p* localizes into the main groove, in close proximity to $\alpha_4$. The palmitoyl moiety Man-C16 localizes in a hydrophobic cavity, deeply buried into the core of PatA.

inner membrane (IM), the peptidoglycan−arabinogalactan complex (AGP), the outer membrane (OM) that is covalently linked to the AGP through mycolic acids, and the external capsular-like structure.[6−8] In this context, the cell envelope of *M. tuberculosis* contains a variety of unique glycolipids/lipoglycans that populate both the inner and outer membranes, including phosphatidyl-*myo*-inositol mannosides (PIMs), lipomannan (LM), and lipoarabinomannan (LAM), which are important modulators of the host immune response during infection.[9−11] PIMs are based on a phosphatidyl-*myo*-inositol (PI) lipid anchor carrying one to six Man*p* residues and up to four acyl chains.[8,12] Tri- and tetra-acylated phosphatidyl-*myo*-inositol dimannoside ($Ac_1PIM_2$ and $Ac_2PIM_2$, respectively) and phosphatidyl-*myo*-inositol hexamannoside ($Ac_1PIM_6$ and $Ac_2PIM_6$ respectively) are the predominant species (Figure 1a).[13−15] PIMs are considered not only essential structural components of the cell envelope but also the precursors of LM and LAM.[16,17]

According to the currently accepted model, the PIM biosynthetic pathway takes place on both sides of the mycobacterial inner membrane.[13,18] The stages of the synthesis of the apolar (PI−$PIM_2$) and polar ($PIM_3$−$PIM_6$) forms of PIMs are carried out on the cytoplasmic and periplasmic faces of the inner membrane, respectively. The early steps involve the consecutive action of three enzymes, the mannosyltransferases PimA and PimB and the acyltransferase PatA, to form $Ac_1PIM_2$.[8,12] PimA catalyzes the transfer of a Man*p* residue from GDP-Man to the 2-position of the *myo*-inositol ring of PI, yielding PI monomannoside ($PIM_1$; Figure 1b).[19−21] PimB transfers a Man*p* residue from GDP-Man to the 6-position of the *myo*-inositol ring of $PIM_1$, generating $PIM_2$.[21−23] $PIM_1$ and $PIM_2$ can be acylated with palmitate at position 6 of the Man*p* residue, transferred by PimA, by the acyltransferase PatA, a member of a large and important family of acyltransferases, to form $Ac_1PIM_1$ and $Ac_1PIM_2$, respectively.[24] The crystal structures of the mannosyltransferases PimA and PimB show the typical organization and catalytic machinery of GT-B glycosyltransferases.[25,26] More recently, the crystal structure of the acyltransferase PatA provided insight into the donor substrate recognition. However, elucidation of the molecular


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

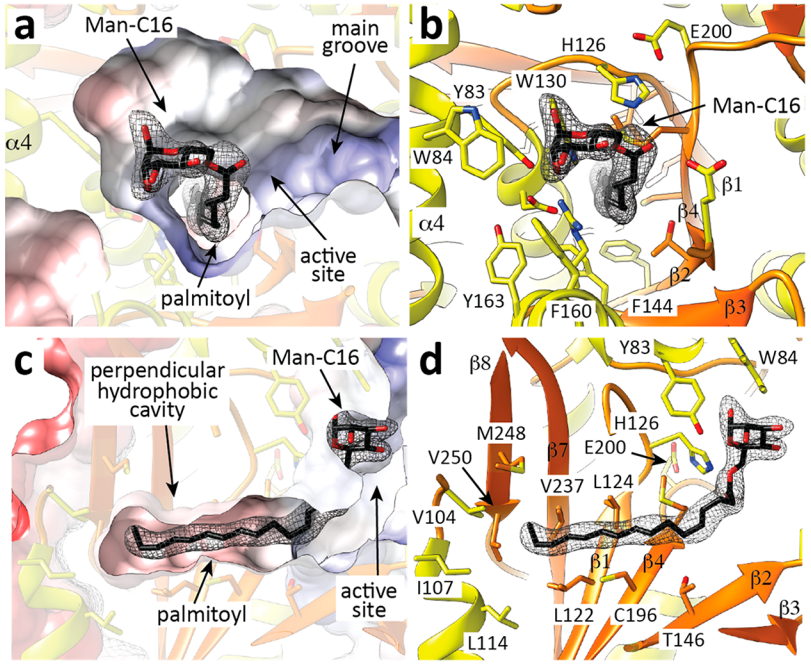



**Figure 3.** The acceptor binding site of PatA. (a–d) Four views of the PatA-Man-C16 crystal structure showing the palmitoyl chain deeply buried into the hydrophobic cavity and the mannose unit located in the acceptor binding site, at the end of the main groove. The final electron density maps ($2mF_o-DF_c$ contoured at $1\sigma$) of the 6-*O*-palmitoyl-$\alpha$-D-mannopyranoside is colored in gray.

mechanism of acceptor substrate recognition and catalysis remains a major challenge for PatA.[3,27,28]

This study describes a detailed investigation of the glycolipid acceptor binding site and the catalytic mechanism of PatA. Using a combination of X-ray crystallography, quantum-mechanics/molecular-mechanics (QM/MM) metadynamics, chemical synthesis, and enzymatic activity measurements, we propose a model for substrate recognition, catalysis, and enzyme inhibition. The implications of this model in the comprehension of the early steps of PIM biosynthesis and the mode of action of other members of the bacterial acyltransferase family of enzymes are discussed.

## RESULTS AND DISCUSSION

**Overall Structure of the PatA-6-O-palmitoyl-$\alpha$-D-mannopyranoside Complex.** Despite much effort, we were unable to produce recombinant PatA from *M. tuberculosis*. In contrast, recombinant PatA from *Mycobacterium smegmatis* was purified as previously described.[27] The amino-acid sequences of *M. tuberculosis* and *M. smegmatis* versions of PatA displayed 74% sequence identity and 84% sequence similarity. All residues that participate in both of the catalytic and substrate recognition mechanisms are strictly conserved between both proteins. To describe the architecture of the acceptor binding site, we first tried to crystallize PatA in the presence of $Ac_1PIM_1/Ac_2PIM_1$, $Ac_1PIM_2/Ac_2PIM_2$, or their deacylated analogs. However, these attempts were unsuccessful. Therefore, we decided to crystallize PatA in the presence of Manp (Table S1). The crystal structure of PatA from *M. smegmatis* was solved by using molecular replacement methods (Figure 2; Table S1). PatA crystallized in the $P2_1$ space group with four molecules (ca. 260 residues each) in the asymmetric unit and diffracted to a maximum resolution of 2.42 Å (Figures S1 and S2; Table S1). Careful inspection of the electron density maps revealed the presence of the product 6-*O*-palmitoyl-$\alpha$-D-mannopyranoside (Man-C16) in three of the four molecules (chains B, C, and D; PatA-Man-C16) of the asymmetric unit (Figures 2 and 3; Figures S1 and S2). The fourth molecule (chain A) showed (i) one molecule of free palmitate located in the hydrophobic pocket and (ii) one molecule of free $\beta$-D-mannopyranose, located in the acceptor binding cavity. Because palmitoyl-CoA was not present during the crystal screening experiments, we believe that the palmitate molecule is associated with the enzyme during the isolation and purification of PatA from *M. smegmatis* $mc^2155$, as previously described.[27] This suggests that PatA is mediating the formation of Man-C16 in three of the four molecules of PatA in the asymmetric unit. PatA crystallized as a monomer, displaying a conserved $\alpha/\beta$ architecture.[27] The central core comprises a six-stranded $\beta$-sheet ($\beta_1-\beta_2-\beta_3-\beta_4-\beta_7-\beta_8$ of with $\beta_8$ is antiparallel) bound by several $\alpha$-helices (Figure 2a). A surface-exposed, open groove with an overall size of 25 × 10 × 10 Å contains the active site. This long groove is flanked by several secondary structure elements including $\alpha_4$, $\beta_2$, $\alpha_8$, $\beta_4$, $\alpha_9$, and $\alpha_{10}$ and the connecting loops $\beta_1-\alpha_7$ (residues 124–130), $\beta_2-\alpha_8$ (residues 148–154), $\beta_3-\alpha_9$ (residues 173–181), $\beta_4-\beta_5$ (residues 198–208), $\beta_6-\alpha_{10}$ (residues 219–221), and $\alpha_{11}-\alpha_{12}$ (residues 282–291).[27] A deep hydrophobic cavity extends through the core of PatA, with the entrance located in the floor of the main groove and closed at the bottom by $\beta_8$, the connecting loop $\alpha_5-\alpha_6$, and part of the $\alpha_6$ (Figure 2b and c).

A structure homologue search using the DALI server revealed only two proteins with significant structural similarity, namely, the acyltransferase LpxM from *Acinetobacter baumannii* (*Ab*LpxM; PDB codes 5KN7 and 5KNK; Z-score of 26.7; RMSD value of 2.6 Å for 243 aligned residues; Figure S3),[29] and the glycerol-3-phosphate acyltransferase from *Cucurbita moschata* (*Cm*GPAT; PDB codes 1IUQ and 1K30; Z-score of 8.8; RMSD value of 3.9 Å for 164 aligned residues; Figure S3).[30,31] *Ab*LpxM catalyzes the transfer of two lauroyl (C12:0) groups to the *R*-3′- and *R*-2-hydroxymyristate positions of lipid A.[2,28] LpxM from *Escherichia coli* (*Ec*LpxM) prefers acyl–acyl-


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX–XXX

carrier protein (acyl−ACP) donors but can also function with acyl−CoA substrates.[2] *Cm*GPAT catalyzes the transfer of an acyl chain either from acyl-ACP or acyl-CoA, to the *sn*-1 position of glycerol-3-phosphate, to form 1-acylglycerol-3-phosphate.[32,33] *Cm*GPAT is able to use palmitoyl-CoA as a donor substrate, as PatA does, following a bibi—reactions involving two substrates and two products—ordered mechanism.[33] PatA and *Ab*LpxM are distantly related to *Cm*GPAT, and the lysophosphatidic acid (LPAAT), dihydroacetone phosphate (DHAPAT), and 2-acylglycero-phosphatidylethanolamine (LPEAT) families of acyltransferases.[3,27,34]

**The Acceptor Binding Site of PatA.** The Man*p* ring is located within a cavity located at the end of the main groove and comprising helices $\alpha_4$ and $\alpha_8$ and the connecting loops $\beta_1-\alpha_7$ (residues 83−90), $\beta_2-\alpha_8$ (residues 148−154), and $\alpha_{11}-\alpha_{12}$ (residues 282−291; Figure 3). The Man*p* ring is stabilized by van der Waals interactions with the side chains of four aromatic residues Tyr83, Trp84, Trp130, and Phe160. The O6 atom of the Man*p* ring makes a hydrogen bond with the NE2 atom of the imidazole moiety in His126, whereas the O2 atom makes a weak hydrogen bond with the $NH_2$ atom of Arg164 (Figure 3). It is worth noting that the Man*p* residue observed in the PatA-Man-C16 complex superimposed well with the corresponding Man*p* residue of the deacylated form of $PIM_2$ obtained by docking calculations.[27] The structural superposition of the PatA-ManC16 and S-C16CoA complexes gives a view of the acyl donor and acceptor substrate binding sites (Figure 4). The palmitoyl group is deeply buried into the hydrophobic cavity and oriented with the carboxyl group facing the groove and the acyl tail extending into the globular core of the monomer (Figures 3 and 4).[27] The 4-phosphopantetheinate moiety of S-C16CoA localizes in the opposite side, at the entrance of the main groove, flanked by the $\beta 2-\alpha 8$ (residues 149−153), $\beta 3-\alpha 9$ (residues 174−180), and $\beta 4-\beta 5$ (residues 199−207) loops, and two alpha helices, $\alpha 9$ (residues 181−190) and $\alpha 10$ (residues 221−230; Figure 4). The adenosine 3′,5′-diphosphate (3′,5′-ADP) moiety of S-C16CoA (disordered in three of the four protomers of the asymmetric unit) protrudes from the groove and is exposed to the bulk solvent, as observed in other acyl-CoA modifying enzymes.

**The Catalytic Mechanism of PatA.** To gain insight into the chemical reaction mechanism catalyzed by PatA, we modeled a PatA-$PIM_2$-C16CoA ternary complex, taking into account the experimental structural data available (Figure S4).[27] Molecular dynamics (MD) simulations (38 ns) were performed at RT to relax the system and sample all the accessible enzyme−substrate conformations (Figure S4). A snapshot of the equilibrated complex was used as a starting point for the modeling of the enzymatic reaction using QM/MM metadynamics.[35] Three collective variables (CVs, Figure S5) were selected to account for the main changes taking place during the reaction. The first collective variable (CV1) was defined as the difference of distances between O6−C16 and C16−S. This variable describes the nucleophilic attack of the O atom of the $CH_2OH$ group of one of the mannosyl residues of $PIM_2$ (Man1 in Figure S4) on the carbonyl of the thioester bond of C16-CoA. We named this variable "nucleophile attack." The second collective variable (CV2) was selected as the difference between the O6−H6 and $N_\varepsilon$−H distances. This variable accounts for the proton transfer between the $CH_2OH$ group of Man1 and His126; thus we named it "$CH_2OH$ proton transfer." The third collective variable (CV3), named "His proton transfer," was taken as the distance difference between

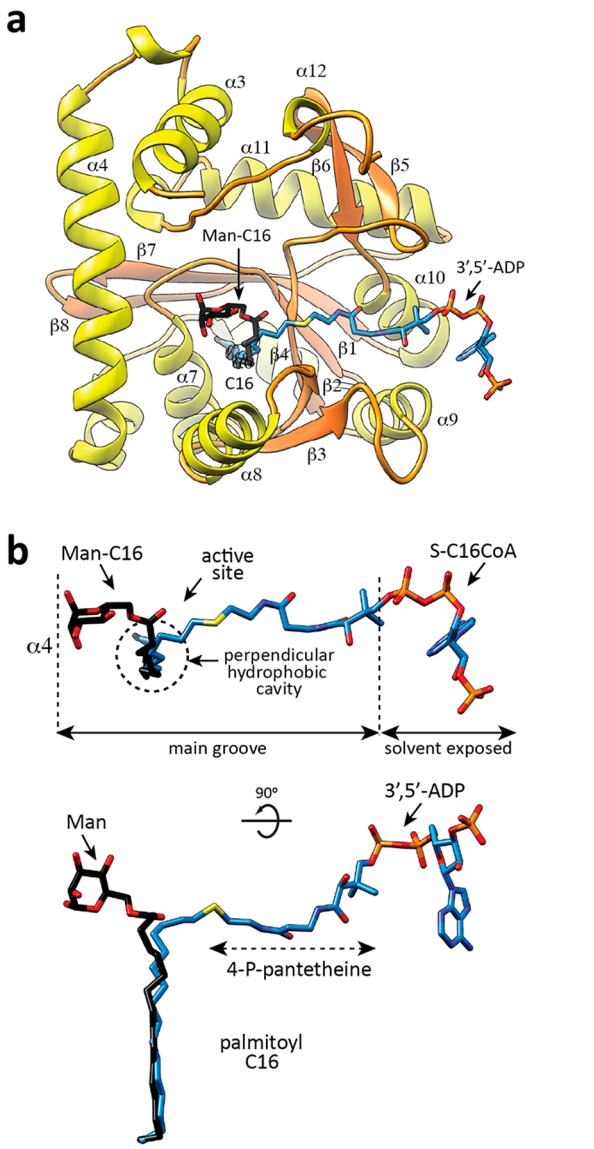



**Figure 4.** The donor and acceptor binding mechanism. (a) Cartoon representation showing the assembled active site and location of (i) Man-C16 and (ii) S-C16CoA into PatA. $\beta$-strands and $\alpha$-helices are shown in orange and yellow, respectively. (b,c) Two views showing the structural superposition of (i) Man-C16 and (ii) S−C16CoA.

$N_\varepsilon$−H and H−S; thus it accounts for the transfer of the His126 proton to the S atom of C16CoA. Similar ideas were previously used with success to model chemical reactions in enzymes.[36] It is important to note that the CVs used do not self-select any specific reaction pathway: they do not impose whether proton transfer takes place in concert with nucleophilic attack or is a later event along the reaction coordinate. Likewise, the CVs used do not impose whether the reaction is fully concerted or an intermediate species forms along the reaction coordinate.

The free energy landscape reconstructed from the QM/MM metadynamics simulation (Figure 5a) shows two minima connected by a unique transition state, suggesting a concerted reaction. However, analysis of the structures along the minimum energy pathway reveals that the minimum on the left-hand side of the free energy landscape does not correspond to the products of the reaction. Instead, the thioester bond is cleaved but the leaving group bears a negative charge (i.e., a thiolate leaving group). Additional analysis by free QM/MM


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

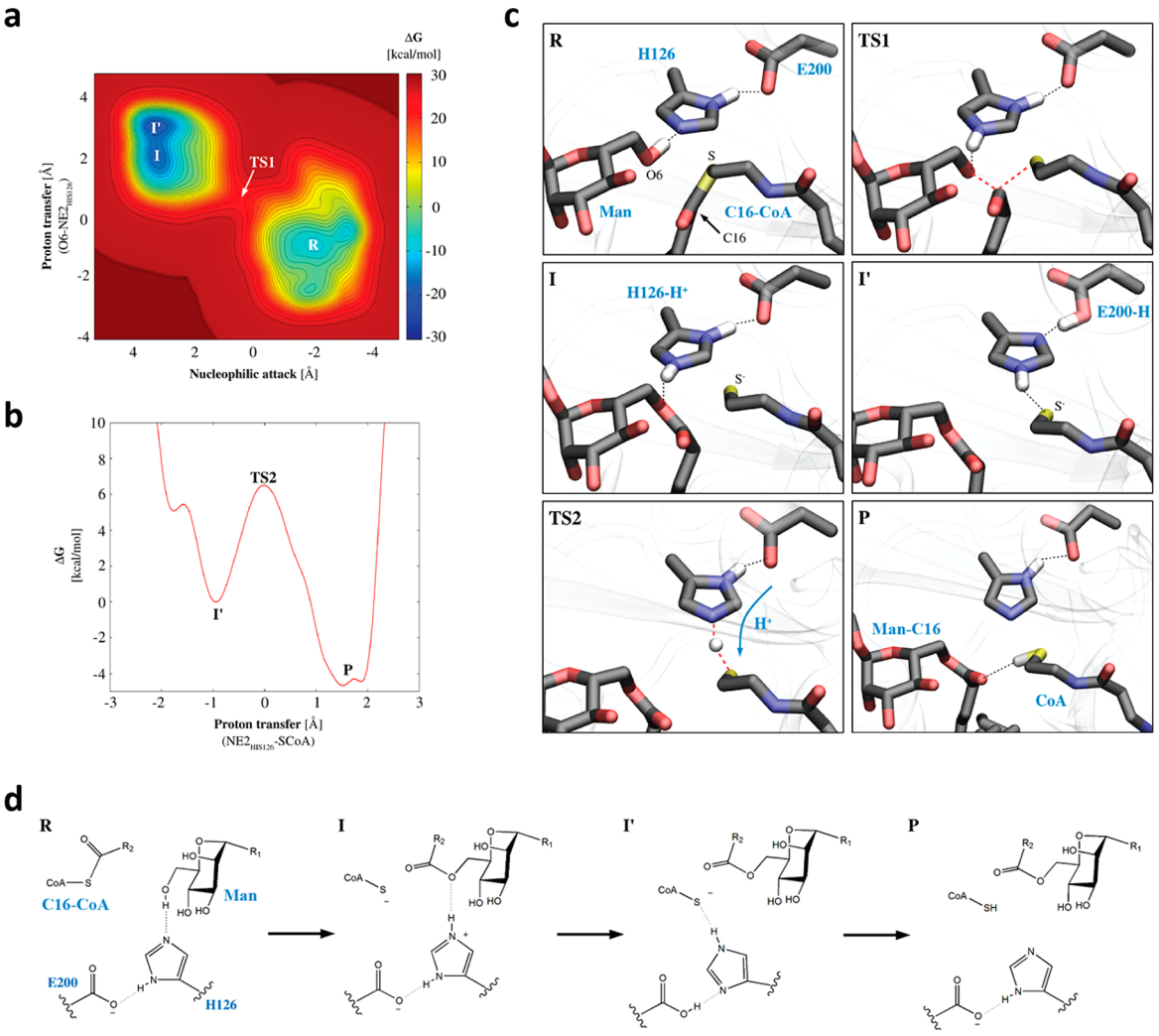



**Figure 5.** QM/MM studies on PatA mediated catalysis. (a) Free energy landscape reconstructed from the metadynamics simulation of the first step of the enzymatic reaction (projection on CV1 and CV2). (b) Free energy profile corresponding to the second reaction step. (c) Representative structures along the reaction coordinate, obtained from the minimum free energy reaction pathway. Red broken lines represent bonds that are being broken or formed, whereas dotted black lines represent relevant hydrogen bond interactions. (d) Proposed catalytic mechanism for PatA.

MD confirmed that this species is a stable reaction intermediate, thus the reaction is clearly stepwise. The reorientation of the thiolate after cleavage of the C−S bond, forming a strong interaction with Arg201 (Figure S6), makes proton transfer from His126 difficult. Therefore, we decided to model the second step of the reaction (protonation of the thiolate leaving group) separately, using as unique collective variable the distance difference between $N_\varepsilon$−H and H−S. Analysis of the corresponding free energy profile (Figure 5b) reveals a concerted proton transfer step, leading to the reaction products $Ac_1PIM_2$ and CoA.

To get a detailed picture of the atomic reorganization during the enzymatic reaction, we extracted configurations along the reaction coordinate of both reaction steps (Figure 5c). The reaction starts with the approach of the mannosyl $CH_2OH$ to the carbonyl of the thioester of Palm-CoA: the O6···C16 distance evolves from 4 to 2.8 Å (Figure S4). At this point, the C−S thioester bond starts to stretch as the catalytic histidine (His126) abstracts the proton from the mannosyl $CH_2OH$, leading to a transition state (TS; 32 kcal/mol above the reactants) in which the thioester carbon atom exhibits a tetrahedral coordination (Figure 5c). The transition state decays toward a reaction intermediate in which the C−S bond is broken and the mannosyl residue is acylated. It is worth noting that this intermediate is not the same as the one found for the hydrolysis of alkyl esters, for which the reaction mechanism is stepwise and passes through a tetrahedral intermediate.[37] There is a significant amount of evidence that shows that such an intermediate is generally not formed for esters bearing good leaving groups, such as aryl esters or thioesters.[38−42] The $N_\delta$−H of the catalytic histidine is involved in a low-barrier hydrogen bond with the carboxylate side chain of Glu200 (Figure 5c,d and Figure S7). In the second reaction step, a concerted double proton transfer between Glu200, His126, and the thiolate leaving group takes place (GluH···HisH···$S^-$ → $Glu^-$···HHis···HS), restoring the original tautomeric state and protonation of His126 and Glu200, respectively (Figure 5c,d). The free energy barrier of this reaction step is 6.5 kcal/mol, and the products state is 4.6 kcal/mol below the reaction intermediate.

These calculations support a mechanism in which PatA is able to acylate $PIM_2$ in a two-step reaction mechanism in which His126 acts as a general base to deprotonate the hydroxyl group of a mannose ring in $PIM_2$, facilitating the nucleophilic


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

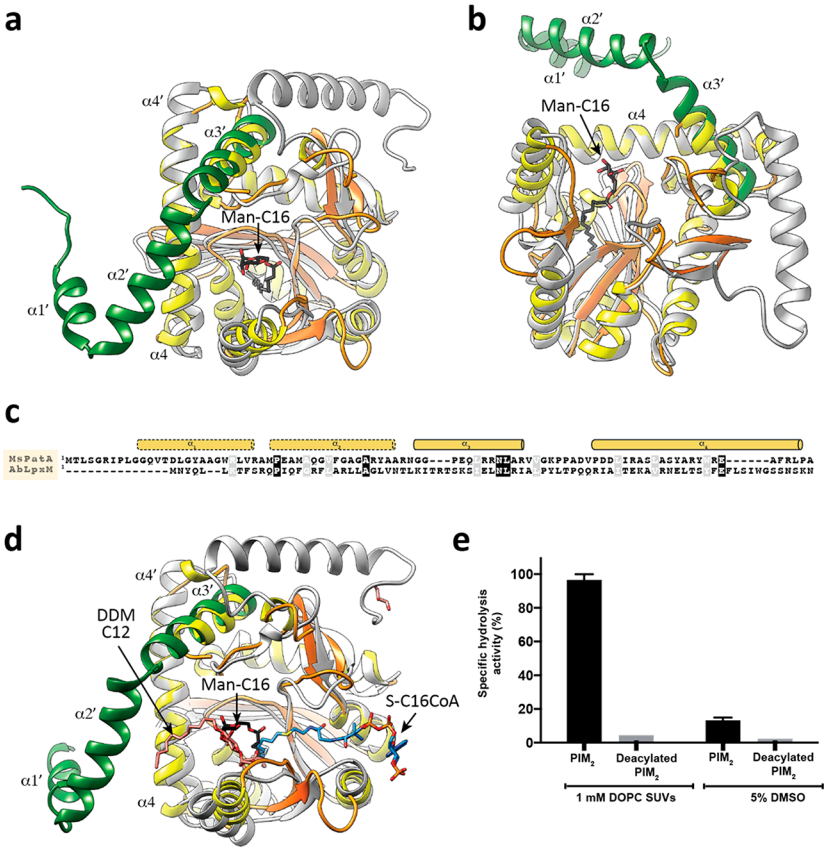



**Figure 6.** Structural similarities between PatA and LpxM. (A,B) Cartoon representation showing two views of the structure weighted sequence alignment of PatA with LpxM. PatA $\beta$-strands and $\alpha$-helices are shown in orange and yellow, respectively. The Man-C16 product is shown in black and red. The first three N-terminal $\alpha$-helices of LpxM are shown in green, whereas the remaining secondary structure elements are shown in gray. (C) Structure weighted sequence alignment of a selected region of PatA and LpxM, comprising residues that code for the four N-terminal $\alpha$-helices. (D) Cartoon representation showing the structural superposition of PatA and LpxM. Both structures are colored as in panel A. The Man-C16 product is shown in black and red. The S−C16CoA acyl donor substrate derivative is shown in light blue and red. The DDM molecule is shown in orange. (E) Enzymatic activity of full-length PatA against palmitoyl-CoA by using diacylated and deacylated $PIM_2$ as acceptors of the reaction and in the presence and absence of DOPC SUVs.

attack on the thioester of C16-CoA (Figure 5d). The reaction is assisted by Glu200, which modulates the p$K_a$ of the central histidine as a general base or acid during the catalytic cycle, similarly to the charge-relay system of serine proteases.[43] Our results also provide evidence of the formation of a tetrahedral-like transition state, as previously found for the closely related hotdog-fold acyl-CoA thioesterase.[42] These results highlight the relevance of the charge-relay glutamate, Glu200, and offer valuable detail on the reaction transition state that would serve for the design of new chemotherapeutic drugs.

**$PIM_2$ Acyl Chains Are Essential for PatA Enzymatic Activity.** A detailed comparison of the PatA-Man-C16 complex with that of *Ab*LpxM in complex with an *n*-dodecyl-$\beta$-D-maltoside provides insights into the molecular mechanism of acceptor recognition (Figure 6a−d). As depicted in Figure S3, the central $\beta$-sheet of *Ab*LpxM superimposes very well with that observed in PatA. Strikingly, (i) the main groove and (ii) the acyl-binding pocket identified in the crystal structures of PatA-Man-C16, PatA-S-C16CoA, and PatA-C16 complexes superimposed very well with the corresponding main groove and a hydrophobic cavity observed in *Ab*LpxM (Figure S8). On the basis of the experimental location of the S-C16CoA in PatA, and taking into account that *Ab*LpxM displayed a strong preference for lauroyl chains (C12) over other acyl chain lengths tested (C10, C14, and C16),[29] a lauroyl−CoA molecule was fitted into an equivalent cavity and subjected to energy minimization (Figure S8a,b). The lauroyl moiety accommodates into the hydrophobic cavity, which is shorter than that observed in the PatA-S-C16CoA complex, pointing to the occurrence of a hydrocarbon ruler mechanism.[29,44,45] The walls of the tunnel are covered with hydrophobic residues including Leu118 ($\beta$2), Val120 ($\beta$2−$\alpha$7 loop), Trp126 and Met129 ($\alpha$7), Met139 and Ile141($\beta$3), Val188 ($\beta$5), and Leu190 ($\beta$5−$\beta$6 loop), with Ile97 and Ile100 ($\beta$1), Phe106 ($\alpha$6), Trp132 ($\alpha$7), Phe136 ($\alpha$7−$\beta$3 loop), Leu229 ($\beta$8), and Ile243 and Met241 ($\beta$9) building the bottom cap.[29] The CoA-binding site observed in PatA is also conserved in *Ab*LpxM. Specifically, the 4-phosphopantetheinate moiety of S-C16CoA is located at the entrance of the main groove (Figure S8c,d). The two $\alpha$-helices $\alpha$9 and $\alpha$10, flanking the entrance of acyl−CoA to the main groove of PatA, are structurally equivalent to $\alpha$9 (residues 170−183) and $\alpha$10 (residues 213−223) in *Ab*LpxM, as also observed in *Cm*GPAT, pointing to a common binding mode (Figure S8c,d).[27,29] *Ab*LpxM crystallized with one molecule of *n*-dodecyl-$\beta$-D-maltoside, with the glucosyl residues deeply buried into a hydrophobic pocket located on one side of the main groove. Strikingly, the first glucose moiety of *n*-dodecyl $\beta$-D-maltoside (DDM) superimposed well with the Man residue of Man-C16 located in the acceptor binding site of PatA. The second glucose moiety of DDM superimposed well with the


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

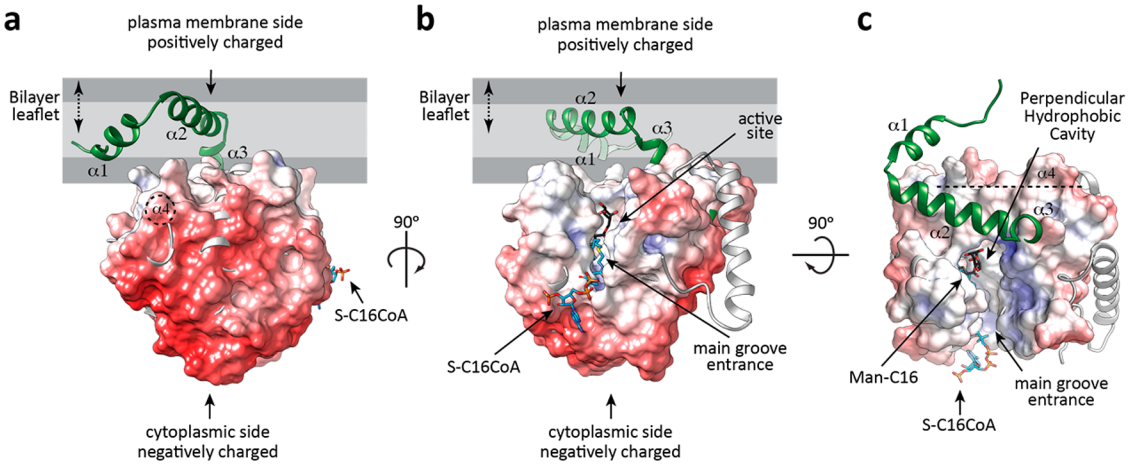



**Figure 7.** A model of action for PatA. (a) Electrostatic surface representation of PatA. PatA is a polar protein, with hydrophobic and positively charged residues oriented toward the plasma membrane. Negatively charged residues are facing the cytosol. The Poisson−Boltzmann electrostatic color map spans from red to blue, ranging from −10 to +10 kcal/mol. The first three N-terminal $\alpha$-helices of LpxM are shown in green, whereas the remaining secondary structure elements are shown in gray. (b,c) Two views of the active site of PatA and the location of the acyl donor and acceptor binding sites.

inositol residue of $PIM_2$.[8,27] Interestingly, the C12 acyl chain of DDM localizes in a hydrophobic cavity formed by $\alpha$1, $\alpha$2, and $\alpha$4 in *Ab*LpxM, conserved in PatA, suggesting a binding mode for the acyl chains of the lipid acceptor substrates.[29]

To further understand how PatA recognizes its lipid acceptor substrate, we explored the contribution of the fatty acid moieties of $PIM_2$ in the enzymatic activity. It is worth noting that previous reports indicate that $PIM_2$ is the preferred substrate of PatA *in vitro*.[21,46] Consequently, we chemically synthesized $PIM_2$ containing two palmitoyl groups as fatty acid chains and compared its acceptor capacity with that of the deacylated version of $PIM_2$. We studied the specific hydrolysis activity of full-length PatA against palmitoyl-CoA by using diacylated and deacylated $PIM_2$ as acceptors of the reaction and in the presence and absence of artificial model lipid membrane DOPC SUVs. We observed only residual specific hydrolysis activity (up to 5%) of PatA against deacylated $PIM_2$, similar to that observed in the absence of any acceptor. We also observed maximum hydrolysis activity in the presence of DOPC SUVs and diacylated $PIM_2$ (0.09 ± 0.004 $\mu$mol CoA/min/mg PatA), about 10-fold higher than without DOPC SUVs. Thus, the enzymatic activity assays indicate that (i) the presence of the $PIM_2$ acyl groups and (ii) an artificial model lipid membrane (DOPC SUVs) are essential for the enzymatic activity of PatA, strongly supporting the notion that the acyl chains of the acceptor glycolipid play a prominent structural role in the arrangement of its sugar moiety/moieties into the catalytic center and the generation of a competent active site (Figure 6e).

**A Model of Action for PatA.** Biological membranes are primarily used as physicochemical barriers allowing cells to be functionally constituted and differentiated from the environment. Biological membranes also play critical roles in all living organisms, mediating the selective transport of molecules across the cell, the modulation of the cellular response, as well as the occurrence of a diverse and important set of biochemical reactions.[47] Many of the enzymes that work in biological membranes are required to interact/recognize both hydrophobic and hydrophilic substrates, located in the membrane and the aqueous environment of the cytosol/organelles, respectively.[4] To perform their biochemical functions, these enzymes interact with membranes by two different mechanisms. Whereas integral membrane enzymes are permanently attached (*e.g.*, through hydrophobic $\alpha$-helices and transmembrane $\beta$-sheets), peripheral proteins temporarily associate with membranes by (i) a cluster of solvent-exposed hydrophobic residues, (ii) electropositive surface patches that interact with acidic phospholipids (*e.g.*, amphipathic $\alpha$-helices), and (iii) protein−protein interactions and/or post-translational modifications. PatA catalyzes a critical step in the biosynthesis of PIMs in mycobacteria.[8,12] The enzyme transfers a palmitoyl moiety from a water-soluble substrate donor palmitoyl−CoA, to the 6-position of the Man*p* ring linked to the 2-position of *myo*-inositol of the membrane glycolipids $PIM_1$ or $PIM_2$.[24,27] Therefore, a close interaction of the enzyme with the cytosolic face of the mycobacterial plasma membrane might be a strict requirement for $PIM_1$ or $PIM_2$ modification by PatA. Interestingly, the electrostatic surface potential of PatA revealed several hydrophobic/aromatic residues interspersed with positively charged residues in the vicinity of the major groove (Figure 7).[27] This region comprises the amphipathic helices $\alpha$3, $\alpha$4, $\alpha$8, and $\alpha$12 and the connecting loop $\beta$2−$\alpha$8 (residues 148−154). Importantly, the opposite side of PatA displays a negatively charged surface, certainly contributing to determine the correct orientation of the enzyme into the membrane (Figure 7). Moreover, the structural comparison of the PatA and LpxM crystal structures suggest that $\alpha$1 and $\alpha$2 N-terminal helices are deeply inserted into the lipid bilayer. Altogether, the structural data support a model in which PatA is permanently associated with the membrane (Figure 7).

Despite the fact that acyl transfer is catalyzed by PatA between the palmitoyl moiety of palmitoyl-CoA and the mannosyl group of $PIM_1$ or $PIM_2$, our experimental data indicate the enzyme displays an absolute necessity for both fatty acid chains of PI for the transfer reaction to take place, pointing to a mechanism requiring a lipid−water interface (Figure 6e). The fatty acid chains of the acceptor substrate seem to play an important structural role in the generation of a competent active site, assisting/guiding the glycosylated moiety to the appropriate environment in the catalytic center. We thus propose a model of interfacial catalysis in which PatA recognizes the fully acylated $PIM_1$ or $PIM_2$ substrates with


DOI: 10.1021/acschembio.7b00578
*ACS Chem. Biol.* XXXX, XXX, XXX−XXX

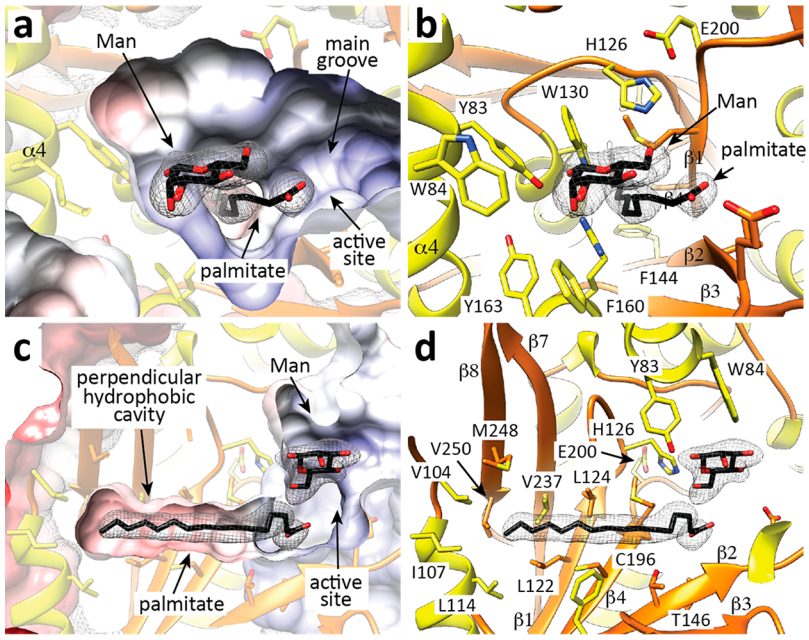



**Figure 8.** The crystal structure of PatA-Man-C16 suggesting a possible mechanism of enzyme inhibition. (A–D) Four views of PatA-Man-C16 crystal structure showing the palmitate chain deeply buried into the hydrophobic cavity and the $\beta$-Man residue located in the acceptor binding site, at the end of the main groove.

their polar heads within the acceptor binding pocket located in the main groove and the fatty acid moieties are likely in close contact with $\alpha$1, $\alpha$2, and $\alpha$4. The positioning of the first two $\alpha$-helices $\alpha$1 and $\alpha$2 in a membrane environment remains intriguing. The palmitoyl CoA binding site involves two grooves (Figures 2, 4, and 6d). The palmitoyl chain is deeply buried into the hydrophobic pocket that runs perpendicular to the main groove, in the core of the enzyme. The 4-phosphopantetheinate moiety is located at the entrance of the main groove, in close contact with a highly conserved region flanked by the $\beta$2–$\alpha$8 (residues 149–153), $\beta$3– $\alpha$9 (residues 174–180), and $\beta$4–$\beta$5 (residues 199–207) loops and two alpha helices, $\alpha$9 (residues 181–190) and $\alpha$10 (residues 221–230). The adenosine 3′,5′-diphosphate (3′,5′-ADP) moiety of the ligand (disordered in three of the four protomers of the asymmetric unit) protrudes from the globular core and is exposed to the bulk solvent. This structural orientation of the acyl–CoA binding site seems to be preserved in LpxM and *Cm*GPAT, supporting a common binding mode for acyl–CoAs.[29,31] Interestingly, LpxM and *Cm*GPAT prefer acyl-ACP as donor substrates; nevertheless, they can also function with acyl-CoA derivatives,[2,33] suggesting the enzymes preserve common structural features to recognize both acyl donor substrates, which might involve important conformational changes.

**Inhibition Mechanism of PatA.** *patA* (*Rv2611c*) is the third gene of a cluster of five ORFs[20,24,48] potentially organized as a single transcriptional unit and likely to be involved in the synthesis of PIMs. The second and fourth ORFs encode the phosphatidylinositol phosphate synthase PgsA (*Rv2612c*)[48,49] and the mannosyltransferase PimA, respectively (*Rv2610c*).[20,21] In addition, the mannosyltransferase PimB is encoded by the *Rv2188c* gene.[22,23] Interestingly, the four enzymes involved in the biosynthesis of the metabolic end product $Ac_1PIM_2$, PgsA1, PimA, PimB, and PatA, are essential for the growth of *M. smegmatis* and/or *M. tuberculosis*.[20,21,24,48,50] Specifically, (i) a *patA* mutant of *M. smegmatis* $mc^2155$ exhibited severe growth defects,[24] while (ii) *patA* was confirmed to be an essential gene in *M. tuberculosis* H37Rv, by a combination of transposon high-density mutagenesis and deep-sequencing.[51]

The crystal structure of PatA-Man-C16 provides one possible mechanism of enzyme inhibition. The PatA-Man-C16 complex revealed the presence of the product Man-C16 in three of the four molecules (chains B, C, and D) of the asymmetric unit, as a consequence of the chemical reaction between the palmitate and $\alpha$-D-mannopyranose. Careful inspection of the electron density maps revealed that in the fourth molecule of the asymmetric unit, the reaction between palmitate and D-mannopyranose did not take place, showing (i) one molecule of free palmitate located in the hydrophobic pocket and (ii) one molecule of free $\beta$-D-mannopyranose, located in the acceptor binding cavity (Figure 8a,b). The structural comparison of the four molecules of the asymmetric unit revealed subtle but important conformational change in the active site (Figure 8c,d). The carboxylate moiety moves away from the catalytic center as observed in the PatA-S-C16CoA complex, in close contact with the main chain of residues Val147 and Glu149.[27] Importantly, the Man O1 in $\beta$ configuration is now stabilized by new interactions with the side chains of Tyr83, Trp130, and Arg164 residues (Figure 8b and d). The Man O4 makes a hydrogen bond with a water molecule, which in turn interacts with the carboxylate moiety of Glu149. Similarly, the Man O6 makes a hydrogen bond with the side-chain of His126 (Figure 8b). Thus, the structural data presented here offer exciting possibilities for inhibitor design and the discovery of chemotherapeutic agents against this major human pathogen.

## METHODS

Details on methods, including (i) purification of PatA from *M. smegmatis*, (ii) N-terminal sequence analyses, (iii) PatA-Man-C16 complex crystallization and data collection, (iv) PatA-Man-C16 structure determination and refinement, (v) quantum mechanics/molecular mechanics molecular dynamics simulations, (vi) chemical synthesis of $PIM_2$, (vii) preparation of deacylated $PIM_2$, (viii)


DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX–XXX

preparation of small unilamellar vesicles (SUVs), (ix) PatA enzymatic activity assay, and (x) structural analysis and sequence alignment, can be found in the Supporting Information Experimental Procedures.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acschembio.7b00578.

Figures S1−S9, Table S1, and additional experimental methods and references (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: mrcguerin@cicbiogune.es.

### ORCID

Carme Rovira: 0000-0003-1477-5010

Marcelo E. Guerin: 0000-0001-9524-3184


### Author Contributions

These authors equally contributed to this work

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the European Commission Contract HEALTH-F3-2011-260872, the Spanish Ministry of Economy and Competitiveness Contracts BIO2013-49022-C2-2-R and BFU2016-77427-C2-2-R (to M.E.G.), Juan de la Cierva Program IJCI-2014-19206 (to B.T.), CTQ2014-55174-P (to C.R.), and Centre National de la Recherche Scientifique (to M.G.). We gratefully acknowledge A. Marina (CIC bioGUNE, Spain) for technical assistance. We acknowledge Diamond Light Source (proposals mx15304) for providing access to synchrotron radiation facilities.

## ■ REFERENCES


(1) Varki, A., Cummings, R., Esko, J., Stanley, P., Hart, G., Aebi, M., Darvill, A., Kinoshita, T., Packer, N. H., Prestegard, J. J., Schnaar, R. L., and Seeberger, P. H. (2017) *Essentials of Glycobiology*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(2) Raetz, C. R., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007) Lipid A modification systems in gram-negative bacteria. *Annu. Rev. Biochem. 76*, 295−329.

(3) Röttig, A., and Steinbüchel, A. (2013) Acyltransferases in bacteria. *Microbiol Mol. Biol. Rev. 77*, 277−321.

(4) Forneris, F., and Mattevi, A. (2008) Enzymes without borders: mobilizing substrates, delivering products. *Science 321*, 213−216.

(5) Jackson, M., McNeil, M. R., and Brennan, P. J. (2013) Progress in targeting cell envelope biogenesis in *Mycobacterium tuberculosis*. *Future Microbiol. 8*, 855−875.

(6) Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H., and Jackson, M. (2014) The cell envelope glycoconjugates of *Mycobacterium tuberculosis*. *Crit. Rev. Biochem. Mol. Biol. 49*, 361−399.

(7) Jankute, M., Cox, J. A., Harrison, J., and Besra, G. S. (2015) Assembly of the mycobacterial cell wall. *Annu. Rev. Microbiol. 69*, 405−423.

(8) Sancho-Vaello, E., Albesa-Jové, D., Rodrigo-Unzueta, A., and Guerin, M. E. (2017) Structural basis of phosphatidyl-*myo*-inositol mannosides biosynthesis in mycobacteria. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids 1862*, 1355−1367.

(9) Gilleron, M., Jackson, M., Nigou, J., and Puzo, G. (2008) Structure, Activities and Biosynthesis of the Phosphatidyl-myo-inositol-based Lipoglycans, in *The Mycobacterial Cell Envelope* (Daffé, M., and Reyrat, J.-M., Eds.), pp 75−105, American Society for Microbiology, Washington, DC.

(10) Jackson, M. (2014) The mycobacterial cell envelope-lipids. *Cold Spring Harbor Perspect. Med. 4*, a021105.

(11) Ishikawa, E., Mori, D., and Yamasaki, S. (2017) Recognition of Mycobacterial Lipids by Immune Receptors. *Trends Immunol. 38*, 66−76.

(12) Guerin, M. E., Korduláková, J., Alzari, P. M., Brennan, P. J., and Jackson, M. (2010) Molecular basis of phosphatidylinositol mannosides biosynthesis and regulation in mycobacteria. *J. Biol. Chem. 285*, 33577−33583.

(13) Morita, Y. S., Fukuda, T., Sena, C. B., Yamaryo-Botte, Y., McConville, M. J., and Kinoshita, T. (2011) Inositol lipid metabolism in mycobacteria: biosynthesis and regulatory mechanisms. *Biochim. Biophys. Acta, Gen. Subj. 1810*, 630−641.

(14) Gilleron, M., Ronet, C., Mempel, M., Monsarrat, B., Gachelin, G., and Puzo, G. (2001) Acylation state of the phosphatidylinositol mannosides from *Mycobacterium bovis* bacillus Calmette Guérin and ability to induce granuloma and recruit natural killer T cells. *J. Biol. Chem. 276*, 34896−34904.

(15) Gilleron, M., Quesniaux, V. F. J., and Puzo, G. (2003) Acylation state of the phosphatidylinositol hexamannosides from *Mycobacterium bovis* Bacillus Calmette Guérin and *Mycobacterium tuberculosis* H37Rv and its implication in toll-like receptor response. *J. Biol. Chem. 278*, 29880−29889.

(16) Chatterjee, D., Hunter, S. W., McNeil, M., and Brennan, P. J. (1992) Lipoarabinomannan. Multiglycosylated form of the mycobacterial mannosylphosphatidylinositols. *J. Biol. Chem. 267*, 6228−6233.

(17) Khoo, K. H., Dell, A., Morris, H. R., Breman, P. J., and Chatterjee, D. (1995) Structural definition of acylated phosphatidylinositol mannosides from *Mycobacterium tuberculosis*: definition of a common anchor for lipomannan and lipoarabinomannan. *Glycobiology 5*, 117−127.

(18) Berg, S., Kaur, D., Jackson, M., and Brennan, P. J. (2007) The glycosyltransferases of *Mycobacterium tuberculosis*: roles in the synthesis of arabinogalactan, lipoarabinomannan, and other glycoconjugates. *Glycobiology 17*, 35R−56R.

(19) Albesa-Jové, D., and Guerin, M. E. (2016) The conformational plasticity of glycosyltransferases. *Curr. Opin. Struct. Biol. 40*, 23−32.

(20) Korduláková, J., Gilleron, M., Mikusová, K., Puzo, G., Brennan, P. J., Gicquel, B., and Jackson, M. (2002) Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. *J. Biol. Chem. 277*, 31335−31344.

(21) Guerin, M. E., Kaur, D., Somashekar, B. S., Gibbs, S., Gest, P., Chatterjee, D., Brennan, P. J., and Jackson, M. (2009) Jackson, M. New insights into the early steps of phosphatidylinositol mannoside biosynthesis in mycobacteria: PimB′ is an essential enzyme of *Mycobacterium smegmatis*. *J. Biol. Chem. 284*, 25687−25696.

(22) Lea-Smith, D. J., Martin, K. L., Pyke, J. S., Tull, D., McConville, M. J., Coppel, R. L., and Crellin, P. K. (2008) Analysis of a new mannosyltransferase required for the synthesis of phosphatidylinositol mannosides and lipoarbinomannan reveals two lipomannan pools in Corynebacterineae. *J. Biol. Chem. 283*, 6773−6782.

(23) Mishra, A. K., Batt, S., Krumbach, K., Eggeling, L., and Besra, G. S. (2009) Characterization of the *Corynebacterium glutamicum* deltapimB′ deltamgtA double deletion mutant and the role of *Mycobacterium tuberculosis* orthologues Rv2188c and Rv0557 in glycolipid biosynthesis. *J. Bacteriol. 191*, 4465−4472.

(24) Korduláková, J., Gilleron, M., Puzo, G., Brennan, P. J., Gicquel, B., Mikusová, K., and Jackson, M. (2003) Identification of the required acyltransferase step in the biosynthesis of the phosphatidylinositol mannosides of mycobacterium species. *J. Biol. Chem. 278*, 36285−36295.

(25) Giganti, D., Albesa-Jové, D., Urresti, S., Rodrigo-Unzueta, A., Martínez, M. A., Comino, N., Barilone, N., Bellinzoni, M., Chenal, A., Guerin, M. E., and Alzari, P. M. (2014) Secondary structure reshuffling modulates glycosyltransferase function at the membrane. *Nat. Chem. Biol. 11*, 16−18.




DOI: 10.1021/acschembio.7b00578
ACS Chem. Biol. XXXX, XXX, XXX−XXX

(26) Batt, S. M., Jabeen, T., Mishra, A. K., Veerapen, N., Krumbach, K., Eggeling, L., Besra, G. S., and Futterer, K. (2010) Acceptor substrate discrimination in phosphatidyl-*myo*-inositol mannoside synthesis: structural and mutational analysis of mannosyltransferase *Corynebacterium glutamicum* PimB′. *J. Biol. Chem. 285*, 37741−37752.
(27) Albesa-Jové, D., Svetlíková, Z., Tersa, M., Sancho-Vaello, E., Carreras-González, A., Bonnet, P., Arrasate, P., Eguskiza, A., Angala, S. K., Cifuente, J. O., Korduláková, J., Jackson, M., Mikušová, K., and Guerin, M. E. (2016) Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nat. Commun. 7*, 10906.
(28) Whitfield, C., and Trent, M. S. (2014) Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem. 83*, 99−128.
(29) Dovala, D., Rath, C. M., Hu, Q., Sawyer, W. S., Shia, S., Elling, R. A., Knapp, M. S., and Metzger, L. E., 4th (2016) Structure-guided enzymology of the lipid A acyltransferase LpxM reveals a dual activity mechanism. *Proc. Natl. Acad. Sci. U. S. A. 113*, E6064−E6071.
(30) Turnbull, A. P., Rafferty, J. B., Sedelnikova, S. E., Slabas, A. R., Schierer, T. P., Kroon, J. T., Simon, J. W., Fawcett, T., Nishida, I., Murata, N., and Rice, D. W. (2001) Analysis of the structure, substrate specificity, and mechanism of squash glycerol-3-phosphate (1)-acyltransferase. *Structure 9*, 347−353.
(31) Tamada, T., Feese, M. D., Ferri, S. R., Kato, Y., Yajima, R., Toguri, T., and Kuroki, R. (2004) Substrate recognition and selectivity of plant glycerol-3-phosphate acyltransferases (GPATs) from *Cucurbita moscata* and *Spinacea oleracea*. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 13−21.
(32) Murata, N., and Tasaka, Y. (1997) Glycerol-3-phosphate acyltransferase in plants. *Biochim. Biophys. Acta, Lipids Lipid Metab. 1348*, 10−16.
(33) Hayman, M. W., Fawcett, T., and Slabas, A. R. (2002) Kinetic mechanism and order of substrate binding for sn-glycerol-3-phosphate acyltransferase from squash (*Cucurbita moschata*). *FEBS Lett. 514*, 281−284.
(34) Lewin, T. M., Wang, P., and Coleman, R. A. (1999) Analysis of amino acid motifs diagnostic for the sn-glycerol-3-phosphate acyltransferase reaction. *Biochemistry 38*, 5764−5771.
(35) Barducci, A., Bonomi, M., and Parrinello, M. (2011) Metadynamics. *WIREs Comput. Mol. Sci. 1*, 826−843.
(36) Ardèvol, A., and Rovira, C. (2015) Reaction mechanisms in Carbohydrate-Active eZYmes: glycoside hydrolases and glycosyltransferases. Insights from *ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc. 137*, 7528−7547.
(37) Fukuda, E. K., and McIver, R. T. (1979) Effect of solvation upon carbonyl substitution reactions. *J. Am. Chem. Soc. 101*, 2498−2499.
(38) Hengge, A. C., and Hess, R. A. (1994) Concerted or Stepwise Mechanisms for Acyl Transfer Reactions of p-Nitrophenyl Acetate? Transition State Structures from Isotope Effects. *J. Am. Chem. Soc. 116*, 11256−11263.
(39) Hess, R. A., Hengge, A. C., and Cleland, W. W. (1998) Isotope Effects on Enzyme-Catalyzed Acyl Transfer from p-Nitrophenyl Acetate: Concerted Mechanisms and Increased Hyperconjugation in the Transition State. *J. Am. Chem. Soc. 120*, 2703−2709.
(40) Yang, W., and Drueckhammer, D. G. (2001) Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *J. Am. Chem. Soc. 123*, 11004−11009.
(41) Wang, C., Guo, Q.-X., and Fu, Y. (2011) Theoretical analysis of the detailed mechanism of native chemical ligation reactions. *Chem. - Asian J. 6*, 1241−1251.
(42) Cantu, D. C., Ardevol, A., Rovira, C., and Reilly, P. J. (2014) Molecular mechanism of a hotdog-fold acyl-CoA thioesterase. *Chem. - Eur. J. 20*, 9045−9051.
(43) Botos, I., and Wlodawer, A. (2007) The expanding diversity of serine hydrolases. *Curr. Opin. Struct. Biol. 17*, 683−690.
(44) Ahn, V. E., Lo, E. I., Engel, C. K., Chen, L., Hwang, P. M., Kay, L. E., Bishop, R. E., and Privé, G. G. (2004) A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J. 23*, 2931−2941.
(45) Cuesta-Seijo, J. A., Neale, C., Khan, M. A., Moktar, J., Tran, C. D., Bishop, R. E., Pomès, R., and Privé, G. G. (2010) PagP crystallized from SDS/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure 18*, 1210−1219.
(46) Svetlíková, Z., Baráth, P., Jackson, M., Korduláková, J., and Mikušová, K. (2014) Purification and characterization of the acyltransferase involved in biosynthesis of the major mycobacterial cell envelope glycolipid–monoacylated phosphatidylinositol dimannoside. *Protein Expression Purif. 100*, 33−39.
(47) McMahon, H. T., and Gallop, J. L. (2005) Membrane curvature and mechanisms of dynamic cell membrane remodeling. *Nature 438*, 590−596.
(48) Jackson, M., Crick, D. C., and Brennan, P. J. (2000) Phosphatidylinositol is an essential phospholipid of mycobacteria. *J. Biol. Chem. 275*, 30092−30099.
(49) Morii, H., Ogawa, M., Fukuda, K., Taniguchi, H., and Koga, Y. (2010) A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. *J. Biochem. 148*, 593−602.
(50) Boldrin, F., Ventura, M., Degiacomi, G., Ravishankar, S., Sala, C., Svetlikova, Z., Ambady, A., Dhar, N., Kordulakova, J., Zhang, M., Serafini, A., Vishwas, K. G., Kolly, G. S., Kumar, N., Palù, G., Guerin, M. E., Mikusova, K., Cole, S. T., and Manganelli, R. (2014) The phosphatidyl-*myo*-inositol mannosyltransferase PimA is essential for *Mycobacterium tuberculosis* growth *in vitro* and *in vivo*. *J. Bacteriol. 196*, 3441−3451.
(51) Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol. 48*, 77−84.


DOI: 10.1021/acschembio.7b00578
*ACS Chem. Biol.* XXXX, XXX, XXX−XXX