Title

Computational evaluation of acyl carrier proteins from Mycobacterium tuberculosis in the biosynthesis of mycolic acids

Permalink

https://escholarship.org/uc/item/6hj6m4q3

Author

Sanlley, Javier

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Computational evaluation of acyl carrier proteins from *Mycobacterium tuberculosis* in the biosynthesis of mycolic acids

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Chemistry

by

Javier O. Sanlley Hernandez

Committee in charge:

Professor Michael D. Burkart, Chair
Professor Rommie Amaro
Professor Nathan Romero

2022

The thesis of Javier O. Sanlley Hernandez is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

Dedication

To my parents, my brothers, my grandparents and the rest of my family
To my mentors, teachers, and peers that nurtured my curiosity
To all the friends that I have made along the way and the support they have provided
To that country in the world, situated right in the sun's path
And finally, to the divine order of all things in this life.

__Epigraph__

En el movimiento soy el mas completo,
Mi fe no mueve montañas, el pais completo.

El Alfa

Table of Contents

List of Abbreviations

TB: Tuberculosis

MTC: *Mycobacterium tuberculosis* complex

ACP: Acyl Carrier Protein

PPT: Phosphopantetheinyl transferase

ACC: Acetyl CoA Carboxylase

AT: Acyltransferase

KS: Ketosynthase

KR: Ketoreductase

DH: Dehydratase

ER: Enoylreductases

TE: Thioesterase

MA: Mycolic Acid

TMM: Trehalose monomycolate

TDM: Trehalose dimycolate

FAS: Fatty acid synthase

PKS: Polyketide synthase

NRPS: Non-ribosomal peptide synthase

CoA: Coenzyme A

PNS: Phosphopantetheinyl serine

Ppant: Phosphopantetheine

GaMD: Gaussian accelerated Molecular Dynamics

IDR: Intrinsically disordered regions

List of Figures

Acknowledgements

ABSTRACT OF THE THESIS

Computational evaluation of acyl carrier proteins from *Mycobacterium tuberculosis* in the biosynthesis of mycolic acids

by

Javier O. Sanlley Hernandez

Master of Science in Chemistry

University of California San Diego, 2022

Professor Mike Burkart, Chair

Mycolic acids are very long chain fatty acids (VLCFA) and are yet essential cell wall components of *Mycobacterium tuberculosis,* the causative agent of tuberculosis (TB) disease in humans. Central to the biosynthesis of mycolic acids is the acyl carrier protein, AcpM, responsible for transporting acyl intermediates between partner proteins that collectively produce mycolic acids. Although the structures of carrier proteins are

well conserved across life domains, AcpM contains an additional C-terminus extension that is speculated to play a role in its unique capacity to sequester acyl chains of up to 60 carbons.

To explore how AcpM can accommodate acyl chains during different stages of the fatty acid biosynthesis pathway, molecular dynamics (MD) simulations of AcpM in the *holo* and C16, C32, C64 acyl states were conducted in triplicates for 450ns. Key parameters from the averaged-out trajectories were analyzed such as RMSD, RMSF, radius of gyration, and distance between tail and phosphorus atom in the ligand and pocket volume. These results indicate that C64 AcpM is most likely to be hydrolyzed and not be bound to AcpM. C16 showed the highest dynamics reflected by the high number of interactions during the initial stages of this process. This work provides a theoretical foundation for further structural and computational studies of AcpM that help in understanding the substrate accommodation dynamics during long fatty acid biosynthesis.

Introduction

1.1 Overview of TB

Mycobacterium tuberculosis is the principal causative agent of tuberculosis (TB) disease, one of the most lethal bacterial infections ever known to man[1–3]. Together with other pathogenic strains from the genus *Mycobacteria* the *M. tuberculosis* complex (MTBC) is known to infect humans and other mammals[1]. Prior to the COVID-19 pandemic, TB stood as the leading cause of death by an infectious disease. It is currently estimated that a quarter of the world's population is infected with asymptomatic TB. In 2020, approximately 1.5 million people died from a TB infection[4].

Treatment of TB involves a combination of antibiotics that slow down bacterium growth and prevent TB from spreading to other tissues in the body[5]. While most TB cases can be effectively contained, complete eradication of the bacterium is not always guaranteed[6]. The observed persistence of drug-resistant strains of TB poses additional challenges for effective treatment strategies. Combating drug resistant *M. tuberculosis* infections requires potent antibiotics that can further compromise the patients' health and is an ongoing problem in the medical community[7].

The cellular envelope of *M. tuberculosis* is responsible for the properties that account for the unique robustness and is broken down into an outer layer, a mycomembrane, and plasma membrane (Figure 1)[,8]. The outer layer is a biofilm primarily composed of a heterogenous mixture of proteins, glycolipids, siderophores, and other small metabolites that are secreted to the extracellular space and interact with the host environment[9,10,11].
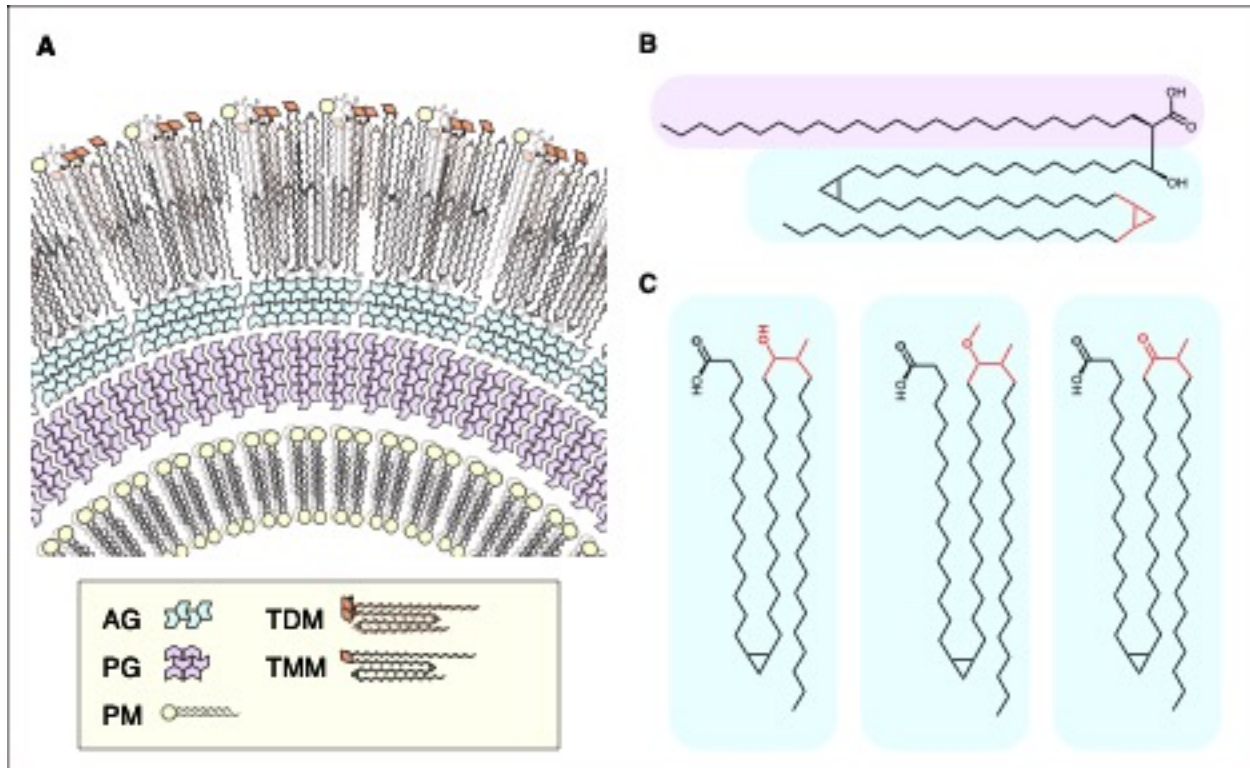
Figure 1: Cellular architecture of the cell wall in *M. tuberculosis* with a view of the outer layer, mycomembrane, and plasma membrane **(A)**. Structure of mycolic acids broken down into alpha (purple) and meromycolate (cyan) chains **(B)**. Structural diversity within meromycolates **(C).**

Below the outer layer, the mycomembrane acts as a highly impermeable barrier against most antibiotics[10,12,3]. The outer leaflet of the mycomembrane contains mycolic acids bound to trehalose units and exist as trehalose monomycolates (TMM) or trehalose dimycolates (TDM) that are known to trigger host inflammatory responses and allow them to bypass immune systems[10,11]. In contrast, the inner leaflet of the mycomembrane is primarily homogeneous and consists of MAs bound to arabinogalactans further attached to the peptidoglycan matrix encompassing the plasma membrane.

Mycolic acids, α-alkyl β-hydroxylated long chain fatty acids, have biochemical properties that similar to other lipids in nature have important roles in energy storage, heat

insulation, and structural integrity of the cell membrane[13,14]. Upon pyrolysis, mycolic acids break down into alpha and meromycolate chains. Historically, the length and different chemical groups in the meromycolate portion of mycolic acids have been used as taxonomic markers to distinguish between different species of mycobacteria[15].

Mycolic acids are widespread throughout the mycobacterial cell wall and account for the mechanical resistance to cell disruption, aversion to host immune system, impermeability to small molecules, and dispersion of inflammatory virulence factors[9,10]. At the same time, approximately 60% of the mycomembrane is composed of lipids, with mycolic acids being the most prevalent lipid[15]. In terms of combating TB, the enzymes involved in the biosynthesis of mycolic acids have therefore been the target of promising antitubercular compounds and continue to drive ongoing research for the development of novel solutions that can circumvent the cellular and molecular mechanisms that result in drug resistant strains[6].

Advancements in molecular biology and bioinformatics have played a major role in discovering new alternatives to combat drug resistant TB. The sequencing of genomes from different tubercular strains[16], coupled with high resolution information of cellular structures and proteins important to *M .tuberculosis* metabolism lay the foundation to many of the underlying processes that contribute to understanding and effectively intercepting *M. tuberculosis*[5,9]. Additionally, computational advances, such as molecular dynamics simulations provide an alternative platform to model the structure and dynamics of proteins that are more difficult to access experimentally.

1.2 Overview of fatty acid biosynthesis

In almost every living organism, the fatty acid synthase (FAS) involves the collective action of multiple enzymes that together generate phospholipids, cofactors, waxes as well as other important natural products with widespread applications[9,13]. Enzymes in FAS are classified as either type I or type II FAS based on the arrangement of catalytic domains in space[17]. Type I FAS are expressed as one large, multi-domain, multi-functional complex where all the catalytic domains exist in the same polypeptide chain (Figure 2).
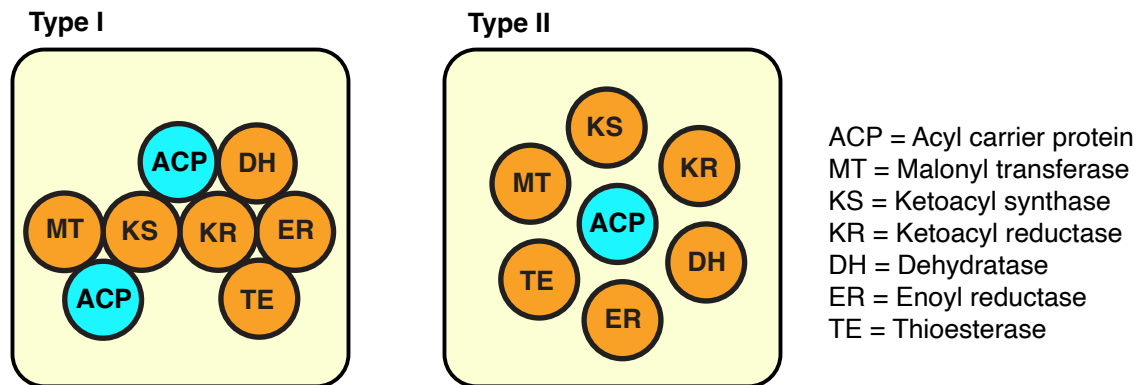
Figure 2: Structural organization of type I and type II FAS systems. Type I FAS systems exist as large multi-modular complexes expressed as one polypeptide. Type II FAS systems have the same domains but are expressed as standalone domains.

The catalytic domains in type II FAS are expressed separately and operate as standalone domains[18]. Enzymes in type I FAS are common in fungi and mammals, while type II FAS systems predominates in plants, bacteria, and algae[13]. Shorter chains produced by FAS can be used as building blocks for important metabolites such as biotin and lipoic acids[9,13]. Although most lipids in nature are produced by either type I or type II

FAS synthases, the biosynthesis of mycolic acids involves the action of both FAS I and FAS II.
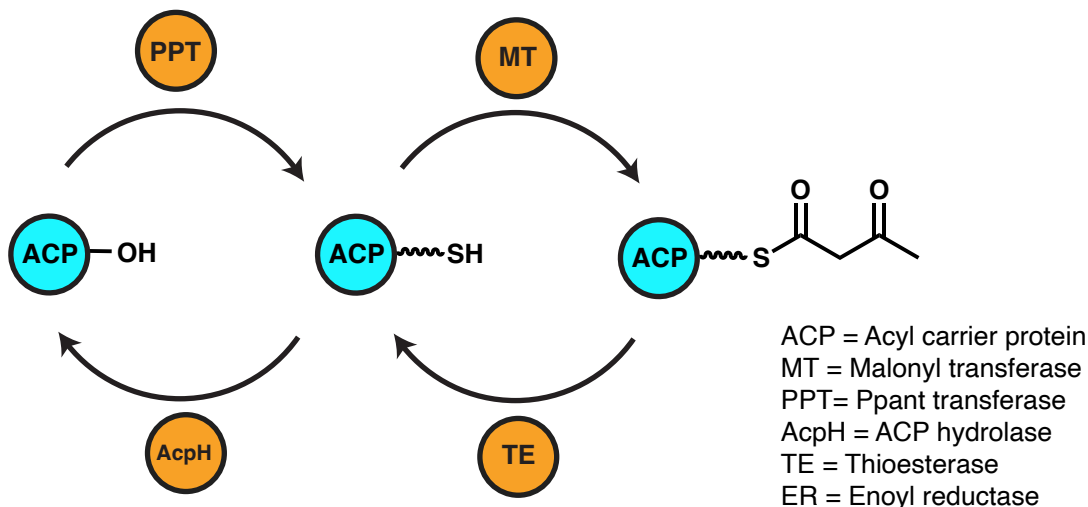
Figure 3: Activation and acylation reactions for CoA and ACP that activate precursors towards mycolic acid biosynthesis by loading and reloading of phosphopantetheine to acyl carrier proteins, as well as acylating and de-acylation of ACP.

The biosynthesis of fatty acids typically begins with acetyl-CoA as two carbon precursors that are condensed over multiple cycles to form fully or partially reduced acyl chains of varying lengths[19]. Acetyl-CoA is involved in many cellular processes, it consists of a coenzyme A (CoA) group bound to acyl isomers that participate in different metabolic pathways[17]. The carboxylation of acetyl-CoA by carboxylases are one of the first steps in fatty acid biosynthesis that commits the acyl groups bound to coenzyme A towards anabolic pathways[20]. This reaction is catalyzed by the acetyl-CoA carboxylases (ACC) that form a malonyl-CoA molecule from acetyl-CoA. The malonyl group in CoA is then loaded onto the phosphopantetheine thiol group of an acyl carrier protein (ACP) by an acyltransferase (AT) domain, forming malonyl-ACP[21]

Carrier proteins bound to acyl chains are known to be in the acyl state and participate in a wide variety of reactions with different partner proteins that depend on the identity of the acyl group attached to them. Outside of the class of proteins that act on carrier proteins bound to acyl intermediates, there are a series of reactions that activate and modify carrier proteins for them to be able to carry, load, and unload acyl chains (Figure 3).
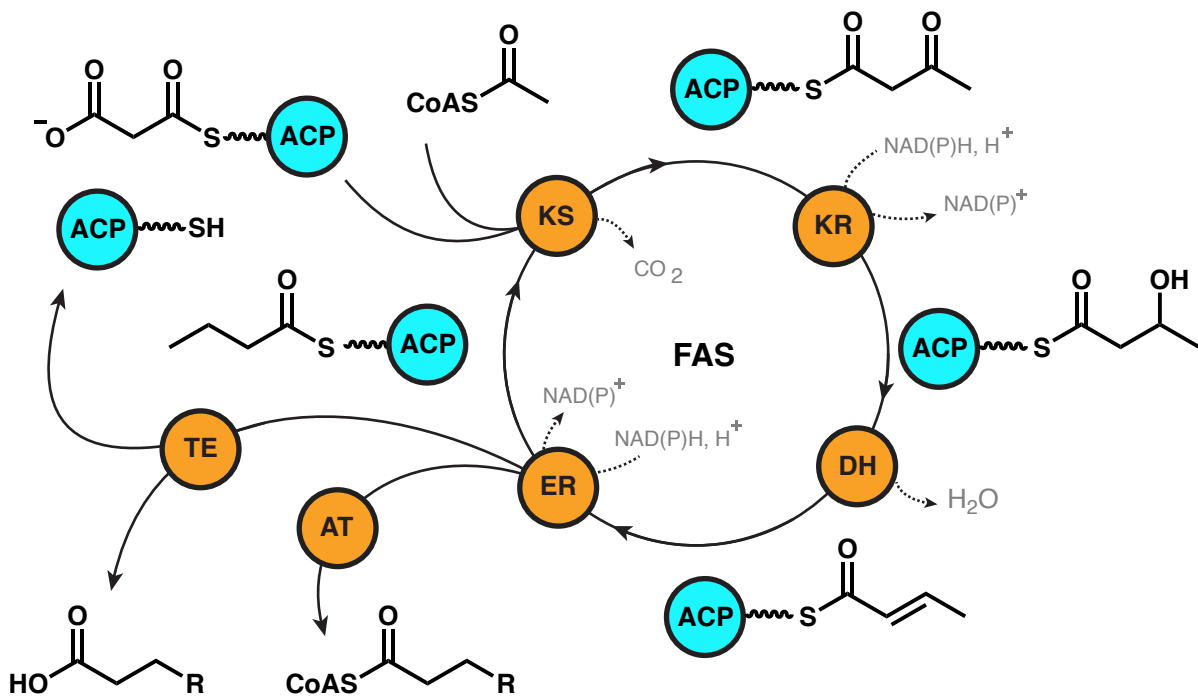
Figure 4: Processing of extended acyl chain bound to carrier protein during FAS II by the enzymatic stepwise reduction of the beta keto group produced after the condensation reaction catalyzed by either initiating or elongating ketosynthase

Carrier proteins bound to malonyl groups usually serve as the starting point for the reactions that serve to extend the acyl chains (Figure 4). In the case of malonyl-ACP, an initiating ketosynthase (KS) condenses the malonyl-ACP with an acetate group from either acetyl-CoA or an acyl-ACP via a decarboxylative Claisen-like condensation. The beta-

acetoacetyl ACP formed by this reaction is then subjected to a series of partner proteins that sequentially act on and reduce the beta-keto group bound to ACP. The process by which the acyl intermediate is exposed to the active site of the partner protein upon binding is known as chain flipping[13]. The beta-keto group in beta- acetoacetyl ACP is first converted to a beta-hydroxy ACP by the NAD(P)H dependent ketoreductase (KR). Following this reaction, ACP reacts with a dehydratase (DH) domain will carry out a dehydration reaction on the beta-hydroxyl group that results in enoyl-ACP that is finally reduced by an NAD(P)H dependent enoyl-reductase (ER) domain.
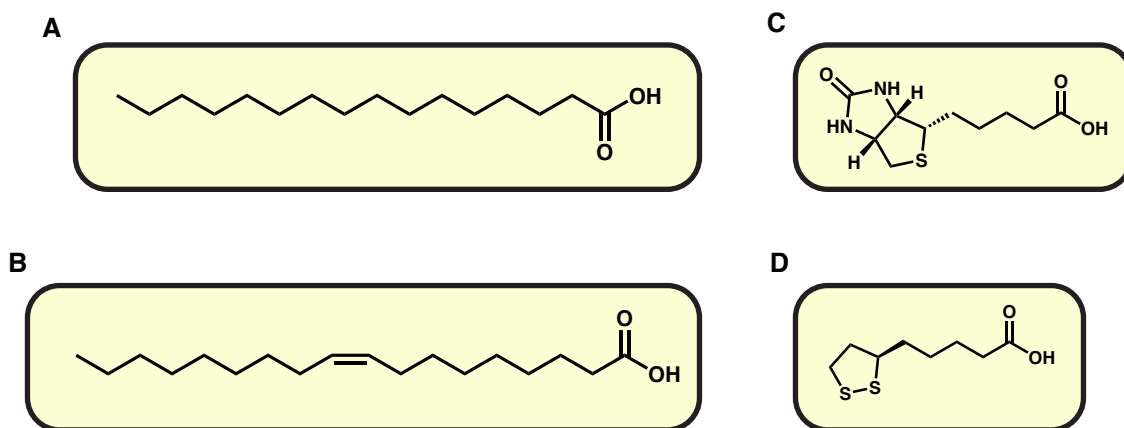
Figure 5: Common products from the fatty acid biosynthesis pathway: palmitic acid **(A)**, oleic acid **(B)**, biotin **(C)** . lipoic acid **(D)**.

After multiple cycles of elongation and reductions, the chain bound to acyl-ACP is hydrolyzed by either thioesterase (TE) or acyltransferase (AT) domains at varying lengths that depend on the fate of the acyl chain. For instance, the palmitic (16:0) and oleic (18:1) acids are some of the principal lipids produced by these pathways (Figure 5). However, shorter chains such as octanoic (8:0) ACP are extracted from the cycle by acyltransferases

and loaded onto a CoA moiety for the biosynthesis of essential cofactors such as biotin and lipoic acids[14,22].

E. coli ACP	*S. oleracea* C18-ACP	*M. tuberculosis* ACP
PDB: 2FAD	PDB: 2FVA	PDB: 2FAD
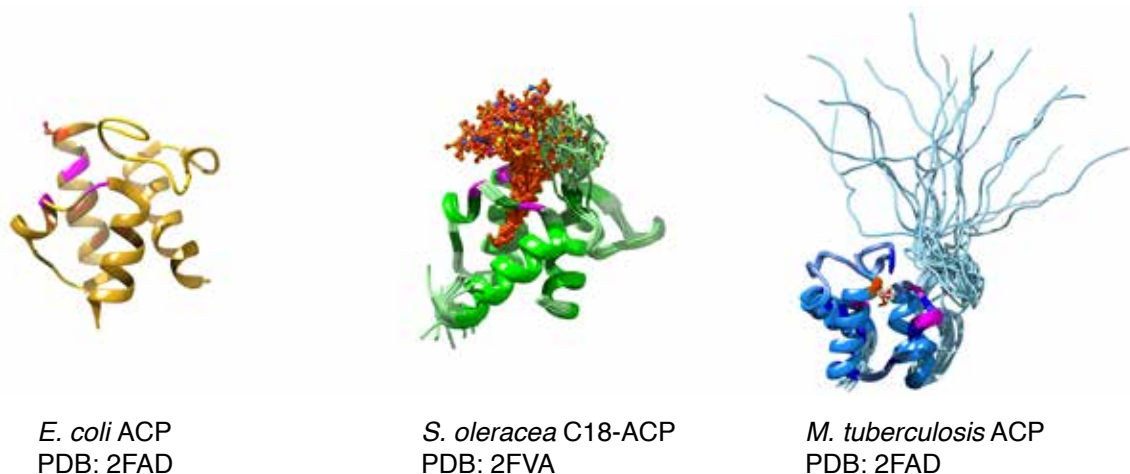
Figure 6: Structures of model acyl carrier proteins from different organisms. Regions colored in pink line the hydrophobic tunnel entrance and regions in a darker shade are known to accommodate acyl chains through hydrophobic interactions.

Central to both type I and type II FAS is the acyl carrier protein (ACP): a small four alpha helical protein that is responsible for the transport of intermediates between different FAS partner proteins (Figure 6). Carrier proteins operate on other synthases such as polyketide (PKS), non-ribosomal peptide synthases (NRPS)[17,23] by stabilizing substrates against unwanted reactions through allosteric changes that simultaneously inform catalytic partners about the state of their cargo[24,25,26].The secondary structure of carrier proteins is generally well conserved. ACP contains a conserved serine residue that is modified by a phosphopantetheine transferase (PPTase) forming *holo*-ACP. This reaction introduces a phosphopantetheine moiety that allows ACP to form reversible thioester bonds with acyl substrates and effectively transport them between different catalytic partners during their biosynthesis.
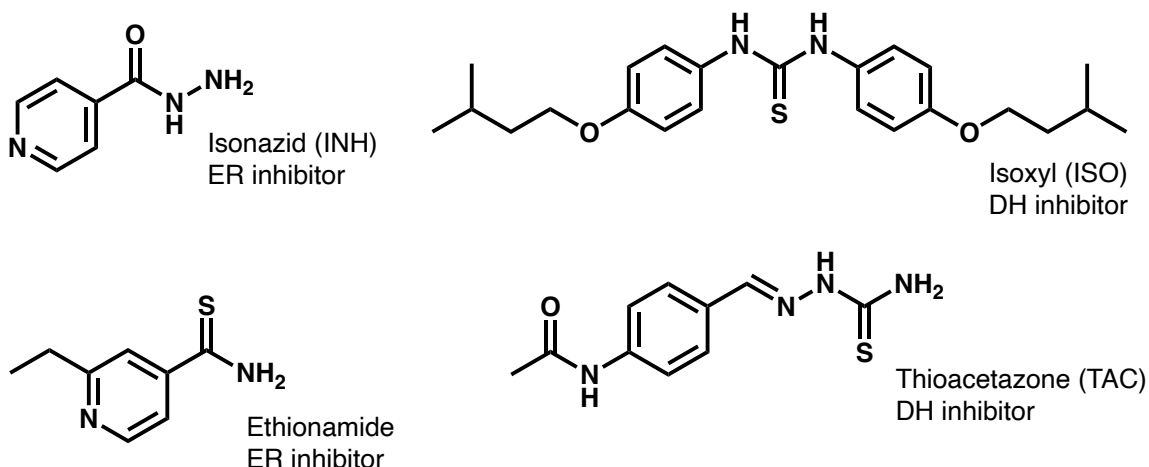
Figure 7: FDA approved first line antitubercular drugs.

Over the years, enzymes involved in the biosynthesis of fatty acids have been the target of multiple first line inhibitors used to combat *M. tuberculosis* infections. Some of these inhibitors, such as Isoniazid (ISO), Ethionamide (ETH) and Thioacetazone (TAC) bind to and inhibit FAS related enzymes as prodrugs that require chemical activation by other proteins (Figure 7). Other proteins in this pathway continue to emerge as attractive candidates that drive progress to combat drug resistant strains. However, an understanding of how these proteins bind to each other can open venues to developments in the field of small peptide inhibitors that can halt key interactions in mycolic acid biosynthesis[27].

1.3 Mycolic acid biosynthesis in *M. tuberculosis*

Mycobacteria contain both type I and type II FAS systems that together produce the meromycolate (C40-C60) and alpha (C22-C26) portions of mycolic acids[10,15]. Generally, type I FAS in mycobacteria result in a bimodal distribution of C16-C18 and C24-C26 fatty acids. Before being condensed with the C24-C26 acyl-CoA produced by FAS I,

the C16-C18 chain is loaded onto the acyl carrier protein in *M. tuberculosis,* AcpM, and is

extended by the type II FAS until a given length that ranges from C40 to C64.
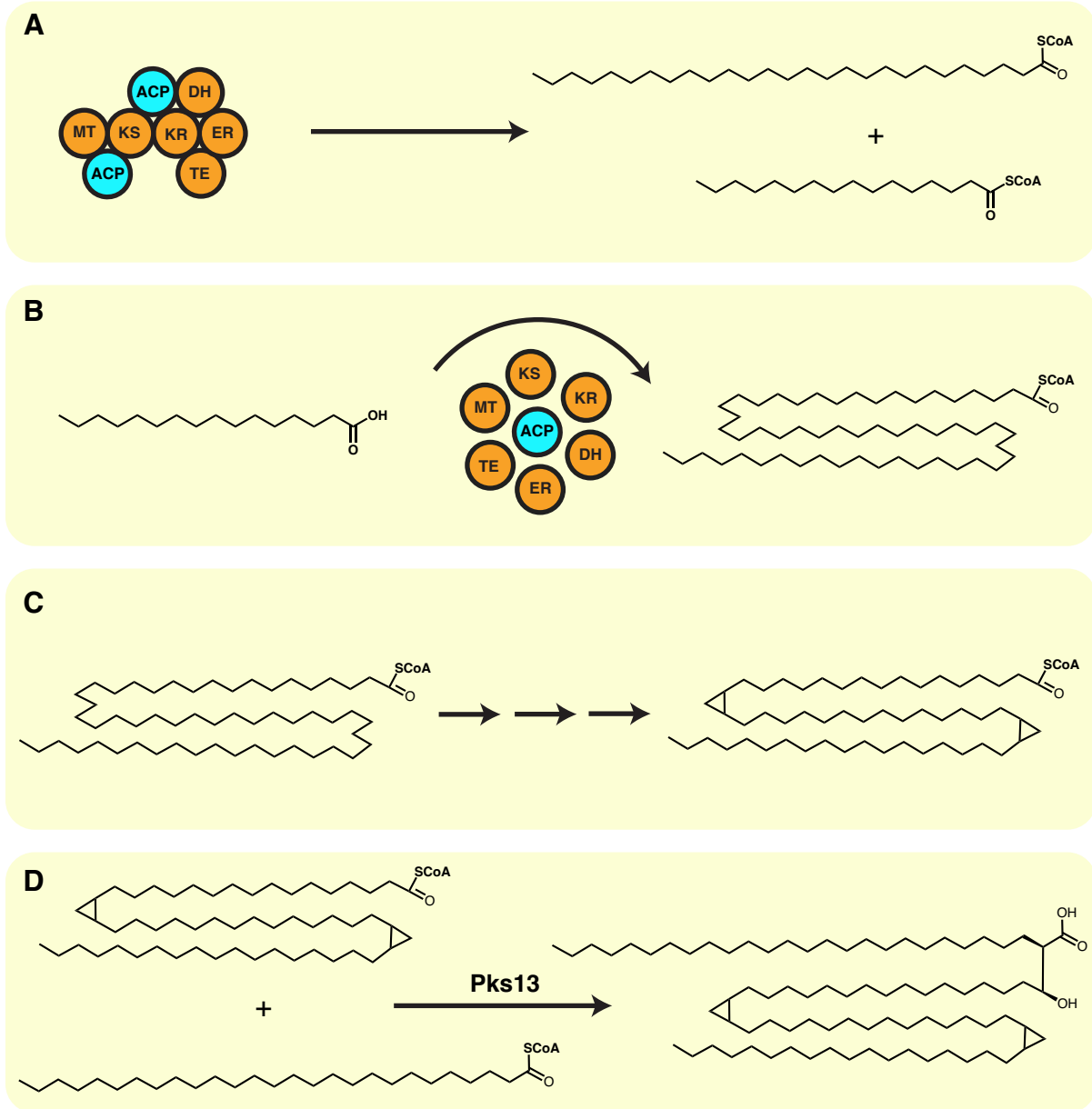
Figure 8: Mycolic acid biosynthesis involves type I FAS that produces the alpha chain and a meromycolate precursor **(A)** that is extended by type II FAS systems **(B)**.The extended meromycolate chain is modified by tailoring enzymes **(C)** and condensed and localized to the mycomembrane **(D)**.

Following the extension by type II FAS enzymes, the meromycolate chain is bound to CoA by an acyltransferase domain before being acted upon by tailoring enzymes that introduce different chemical groups such as double bonds, hydroxyl, methoxy, epoxide, or carbonyl groups at specific locations of the meromycolate (Figure 8)[15]. These modification occurs specifically at carbons C18-C22 of the meromycolate chain, and play an important role in many of the structural and pathological properties attributed to TB disease[12].Ultimately, the extended and modified C62-C64 CoA is condensed with the C24-C26 CoA chain produced by FAS I to form the a-alkyl beta-hydroxy linkage typical of mycolic acids that is acted on by other enzymes that glycosylate and localize the mycolic acid to the mycomembrane[15].

In terms of the elongation of C16-C18 acyl chains to form the meromycolate portion of mycolic acids, AcpM, the acyl carrier protein from *M. tuberculosis,* is crucial for the interactions between different partners during the type II FAS[28]. However, several questions arise in terms of the structural dynamics associated with ACP in terms of pushing the limits of its capacity to sequester long acyl intermediates during type II FAS despite the higher probability of hydrolysis[29]. Although the secondary structure of carrier proteins is generally conserved across life domains, mycobacterial ACP possess a unique intrinsically disordered region (IDR) of about 34 amino acids at their C-terminus[28]. IDR's are generally challenging to study given their capacity to undergo conformational changes *in vivo*, and the disordered tail in ACP has been speculated to play a role in stabilizing long chain fatty acids, as well as facilitating protein-protein interactions[30]. At the same time, it

is not clear whether the tailoring enzymes that modify the C18-C22 carbons of the meromycolate chain act on AcpM or CoA bound intermediates.

Experimental efforts that aim to understand how the structure of AcpM can accommodate long chain fatty acids are hampered by the difficulties involved in working with loading long fatty acids to AcpM, as well as the C-terminus domain in AcpM which makes it difficult to purify. However, the wealth of structural and dynamics information obtained from structures of carrier proteins from different organisms in different states provide a starting point to understand how they are able to sequester and transport long fatty acids[28]. By studying how the structure of AcpM can accommodate acyl substrates of 1.5 differing lengths, we can observe how allosteric changes in the core helices and the tail can have an impact on how AcpM interacts with different partner proteins during the FAS II stage of mycolic acid biosynthesis.

1.4 Molecular dynamics (MD) simulations of AcpM

Molecular dynamics (MD) simulations study the motions of biomolecules by recording their trajectories at relevant timescales to better understand the structural mechanisms associated with important biological processes such as protein folding, drug binding, and changes to a protein structures. MD simulations also take into consideration the chemical, electronic and physical properties of molecules at physiologically relevant conditions and effectively recapitulate the behavior of proteins *in-silico* providing a theoretical understanding of their dynamics that agrees with experimentally derived data. Because of the varying size and timescale associated with a given biological process, MD simulations can become computationally expensive especially for complex systems i.e.,

membrane embedded proteins. Despite these limitations, advances in computational hardware allow trajectories to be processed in parallel using multiple graphics processing units (GPU) that significantly diminish computation time.

MD simulations are based on classical energy functions, known as force fields for modelling molecules in a system after fitting data from experimental results or quantum mechanical calculations, that aim to describe the bonded and non-bonded interactions of molecules[31]. The Amber simulation package, a successful group of MD software for different simulation purposes uses different force fields that describe the bonded and nonbonded interactions between atoms for different classes of biomolecules[32]. Bonded interactions include the bonds (1-2 interactions), angles (1-3 interactions) and dihedrals (1-4 interactions) formed between atoms, while the non-bonded terms include Coulombic interactions and Lennard-Jones potentials, also known as "van der Waals" interactions[31]. Collectively, these terms comprise the amber force field and define the potential energy terms for a molecule in different environments (explicit solvent, implicit solvent, vacuum). It is imperative that force fields can properly capture the behavior of molecules, which is why they are continuously with improved physics that can be validated by experimental data[33].

The potential energy landscape of a protein represents the different states that a protein's structure can adopt according to the different bonded and nonbonded energy terms associated with their geometric and electronic properties[34]. In the context of protein folding, the potential energy landscape can be thought of as a funnel with rugged edges, where proteins can be "stuck" in one of these edges as it scans the landscape for a

configuration that represents its global "minima". Catalytically relevant configurations, such as those observed during the accommodation and processing of intermediates at the active site of a protein, can be thought of as peaks.

In the case of the carrier protein from *M. tuberculosis,* there are several structural and functional components that currently limit their accessibility for structural determination experiments. For instance, the disordered extension at the C-terminus domain is prone to aggregation and can complicate purification strategies. At the same time, the acylated configurations of interest involve acyl chains that are hard to work with experimentally especially when trying to attach them to AcpM through either chemical or chemoenzymatic methods. It is for these reasons that computational approaches to AcpM can provide a reasonable understanding of how their behavior and dynamics when bound to acyl chains of lengths relevant to their processing during the type II FAS stage of mycolic acid biosynthesis.

1.6 Hypothesis and Aims

The conformational changes to the structure of carrier proteins caused by the elongation of acyl chain intermediates within determine the structural factors that guide specific protein-protein interactions during the FAS type II stage of mycolic acid biosynthesis. AcpM, the carrier protein in *M. tuberculosis* is capable of sequestering one of the longest acyl chains produced by nature despite the observed increased rate of thioester bond hydrolysis for longer acyl chains. Additionally, AcpM possesses an IDR extension at its C-terminus extension that is strongly speculated to not only interact with partner proteins during FAS II. MD simulations of AcpM bound to acyl chains of different

lengths can provide insight to the dynamics involved in the accommodation of different substrates, as well as elucidate the role of the disordered tail in stabilizing longer acyl chains.

Methods

The NMR solution structure of AcpM in the *apo* state (PDB:1KIP) and *S. oleracea* bound to a C18 chain (PDB: 2FVA) were obtained from the RCSB. AcpM in the *holo,* C16, C32 and C64 served as starting structures for the simulations and were manually prepared using Pymol while AcpM was aligned with the carrier protein from spinach. Both structures share a conserved sequence identity of 37.9% and were aligned to each other with a final RMSD of approximately 3.3Å[35]. The phosphopantetheinyl (Ser-Ppant) ligand bound to the catalytically conserved serine residue in spinach ACP was used to build the holo, C16, C32, and C64 AcpM starting structures. Protein hydrogens were calculated using the H++ server to account for the protonation states of titratable residues at neutral pH[36].

To account for the additional atoms in the modified serine during the simulations, the phosphopantetheinyl-serine bound to acyl chains for holo, C16, C32 and C64 were individually parametrized to obtain the necessary force field parameters required for MD simulations. The serine residue bound to the rest of the ligand was extracted and capped with acetyl and N-methyl groups at the N and C termini of the serine residue, respectively. The electronic structure of the modified serine ligands was then calculated using semi-empirical AM1 level of theory while atomic charges were determined using the BCC charge method using the antechamber module from the AMBER package.

Force field parameters obtained from the previous steps were used to parametrize and solvate *holo,* C16, C32, and C64 AcpM with ff19SB force fields and OPC water models using the LEaP software from the AMBER package. The system was solvated in explicit

solvent with a truncated octahedron geometry with a 0.150M salt concentration. Additionally, protein charges were neutralized with the appropriate counterions on average, the solvated protein model consisted of approximately 65,000 atoms, x water molecules and an average volume of Å^3.

Minimization of the system was carried out using a steepest descent algorithm over 5,000 steps. The system was then subjected to a constant volume (NVT) restrained heating to a target temperature of 310K over 2.5ns and was then allowed to equilibrate for 25 ns to acquire the necessary boost potential parameters for running GaMD. A 25 kcal/mol*Å^2 harmonic force constant was imposed on the solvent during restrained minimization and heating cycles. A Langevin thermostat and Berendsen barostat were used to maintain the temperature and pressures at 310K and 1atm, respectively. The SHAKE algorithm was used to restrain hydrogens during the simulations.

Using the pmemd.cuda extension from the AMBER package, three 450ns conventional molecular dynamics production runs were conducted for *apo*, *holo*, 16:0 , 32:0 and 64:0 ACP starting with randomized velocities and a timestep of 2 fs. For the starting point in each run, the velocities were assigned randomly to obtain a more accurate representation of the sampled conformational space. The same preparation steps were applied to spinach acyl carrier protein loaded with C16 acyl chain parametrized to obtain a control of the ligand dynamics.

The 450ns trajectories in triplicates of holo, C16, C32, and C64 AcpM were obtained and using CPPTRAJ the averaged out values for the RMSD, RMSF, radius of gyration and distance values for the core (helix I-IV) tail (residues 81-115) and chain

(residue 41) regions in AcpM[37]. RMSD values were calculated based on the deviations

from the backbone alpha carbons. Qualitative analysis of structural changes in AcpM

were based on ensembles of every 45ns interval aligned to each other for a total of ten

structures per simulation.

Results and Discussion

To validate the accuracy of the force field parameters calculated in this study, the acyl chain from the solution NMR structure of C18 spinach ACP was truncated by two carbons to run short MD simulations using the same parameters calculated for the C16 ligand (Figure 9). A short 150ns simulation was compared against the solution structure, as well as parameters obtained from a recent publication that developed phosphopantetheine ligand parameters with high theoretical accuracy[38].
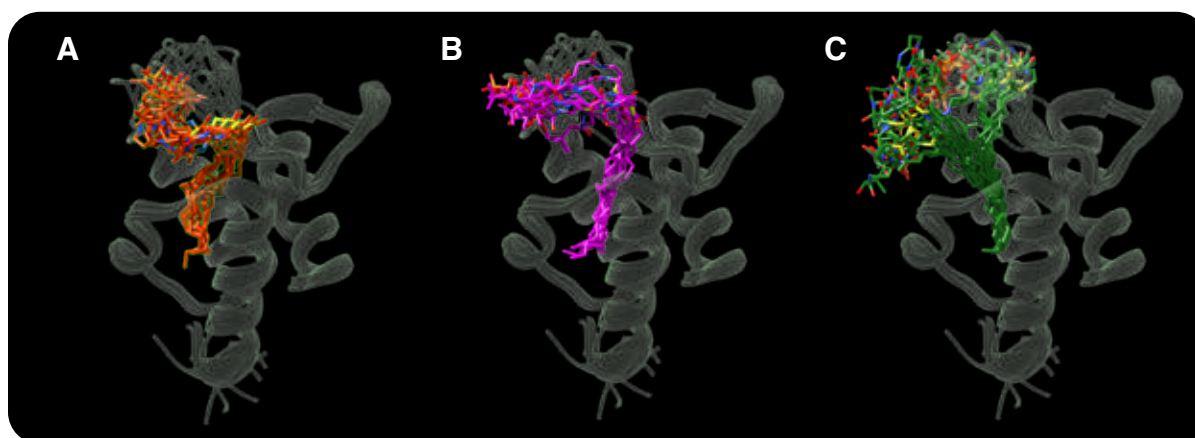
Figure 9: Comparison of simulation results for acyl chain control using *S. oleracea* ACP modeled with C16 acyl chains parametrized with AM1-BCC method from this work **(A)**, using published parameters (Luo, 2021) **(B)** and referenced to experimentally derived dynamics **(C)**.

In this way, the theoretically derived parameters from this worm were able to be compared against the ligand dynamics from experimentally derived data. One of the drawbacks of this method is the fact that the spinach ACP structure was solved while bound to a C18 chain, which may have different ligand dynamics when compared to a C16 solution structure [25]. At the same time, the theoretically derived parameters for

phosphopantetheine obtained from published results only reach C16 acyl chains, which

is why all ligands used in this work had to be parametrized in the same manner in order

to achieve consistency across simulations[38].
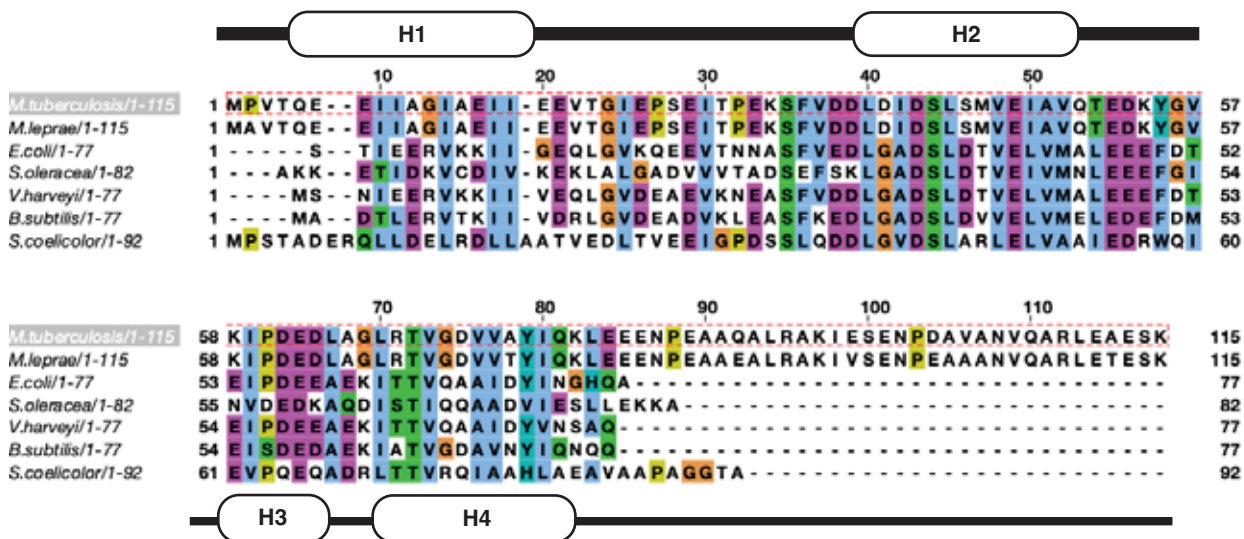
Figure 10: Sequence alignment of carrier proteins from commonly studied acyl carrier proteins from different model organisms.

The structure of carrier proteins are widely conserved throughout nature and the

way in which they accommodate and introduce their sequestered acyl chains to partner

proteins inform these specific interactions[26,24] (Figure 10). Out of the four alpha helices

encountered in carrier proteins, helix II (residues 21-40 in AcpM), known as the

"recognition helix", is known for its role in interacting with partner proteins by forming

transient interfaces[39]. These interfaces ultimately depend on allosteric changes caused

by the accommodation of substrates with varying lengths[26]. Hydrophobic residues in

helix II, as well as residues in helices I and IV interact with each other to stabilize the

acyl chain. Other hydrophobic residues in AcpM line the pocket where the acyl chain is

accommodated. At the top of helix II, a catalytically conserved serine residue serves as attachment point for phosphopantetheine located next to a loop region involved in the formation of transient interfaces with partner proteins. From the NMR solution structure (PDB: 1KLP), tail residues in AcpM have no defined secondary structures. However, at the end of the simulations, secondary structures at the tail region are observed in all structures as alpha helices. Future benchmark experiments should address the consistency of these secondary structures to properly account for their formation as dependent on the choice of water model and force field used.

The alignment of amino acid sequences from *E. coli*, *S. oleracea*, and *M. tuberculosis* acyl carrier proteins provide a starting point for comparing the similarities and differences between different homologues (Figure 10). The acyl carrier protein from *E. coli* fatty acid biosynthesis is a well-studied model protein given the extensive amount of deposited structural and sequence data is a starting point for making comparisons with other carrier proteins, such as AcpM. At the same time, the solution structure of *S. oleracea* ACP was used as a starting model for generating acylated models of AcpM as it is one of the few carrier proteins solved by NMR methods that provide insight to the chain dynamics. Generally, the structure of carrier proteins can be divided into regions that are responsible for accommodating acyl substrates, stabilizing alpha helices and attaching phosphopantetheine to a conserved serine.

In AcpM, 41S is the catalytically conserved serine that is modified by phosphopantetheinyl transferases, while the disordered C-terminus domain is comprised by residues 81-115. Residues 12I, 16I, 33F, 34V, 44M, 47I, 48A, 51T, 55Y, 59I, 64L,

65A, 67L and 68R in AcpM are responsible for lining the hydrophobic cavity. From this region, residues such as 33F, located in loop two, have been postulated to act as a cover for the hydrophobic tunnel and are highly conserved across all carrier protein sequences while 68R, located at loop three, is unique to AcpM since in other carrier proteins this residue exists as a serine or threonine (Figure 11). 68R has been observed to form hydrogen bonds with the carbonyl oxygens from phosphopantetheine in holo AcpM. This interaction is instead observed as hydrogen bonds with S97 from the tail region in C16, C32 and C64 AcpM.

A

B
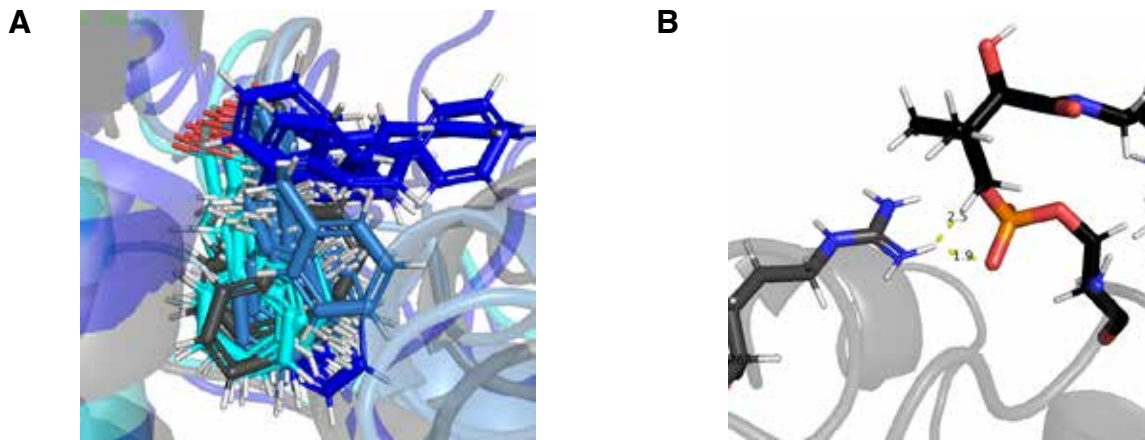
Figure 11: 33F dynamics at the entrance of the hydrophobic tunnel in *holo,* C16, C32 and C64 colored in gray, blue pale blue and cyan respectively **(A)**. Hydrogen bonding observed for residue 68R and phosphopantetheine for *holo* AcpM **(B)**.
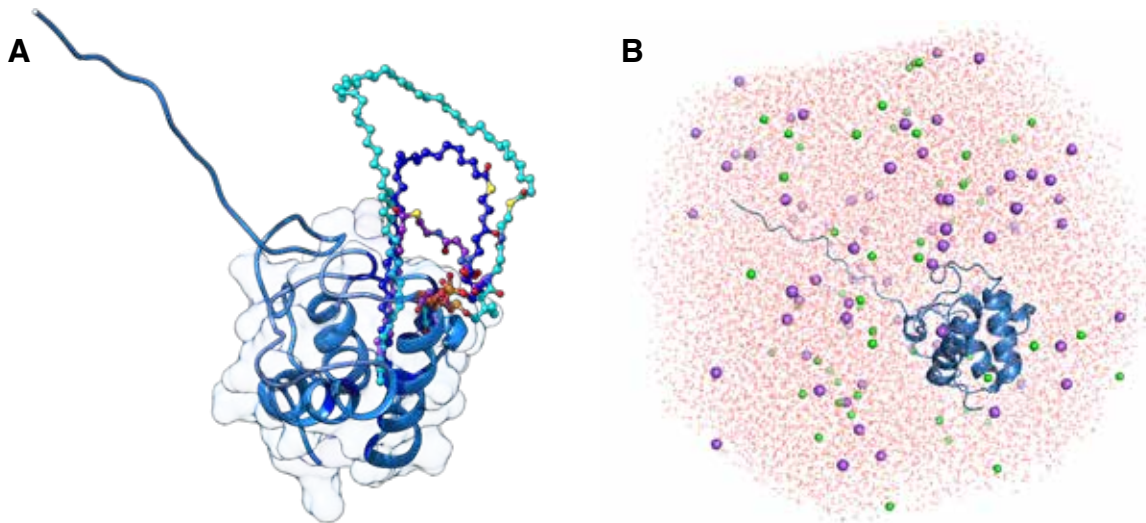
Figure 12: Model structure of overlapped AcpM structures in the, C16, C32 and C64 forms used at the beginning of the simulations colored in purple, blue and cyan, respectively **(A)**. AcpM in explicit solvent box with a truncated octahedron geometry **(B)**.

While preparing the model systems, all acyl chains bound to phosphopantetheine were manually positioned in a conformation where the omega carbon of the chain remained in the same position inside the helices, and the rest of it extended outside of the protein in a hairpin conformation[26] (Figure 12).

Following the unrestrained equilibration, however, sequestration of the acyl chain was observed for C16 and C32 AcpM, while the phosphopantetheine group in holo remained outside of the pocket due to the number of nitrogen and oxygen atoms that interact more with the aqueous environment. In the case of C64 AcpM, there was minimal to no sequestration observed as the C64 acyl chain was found to be in a bundled configuration outside of the cavity (Figure 13).
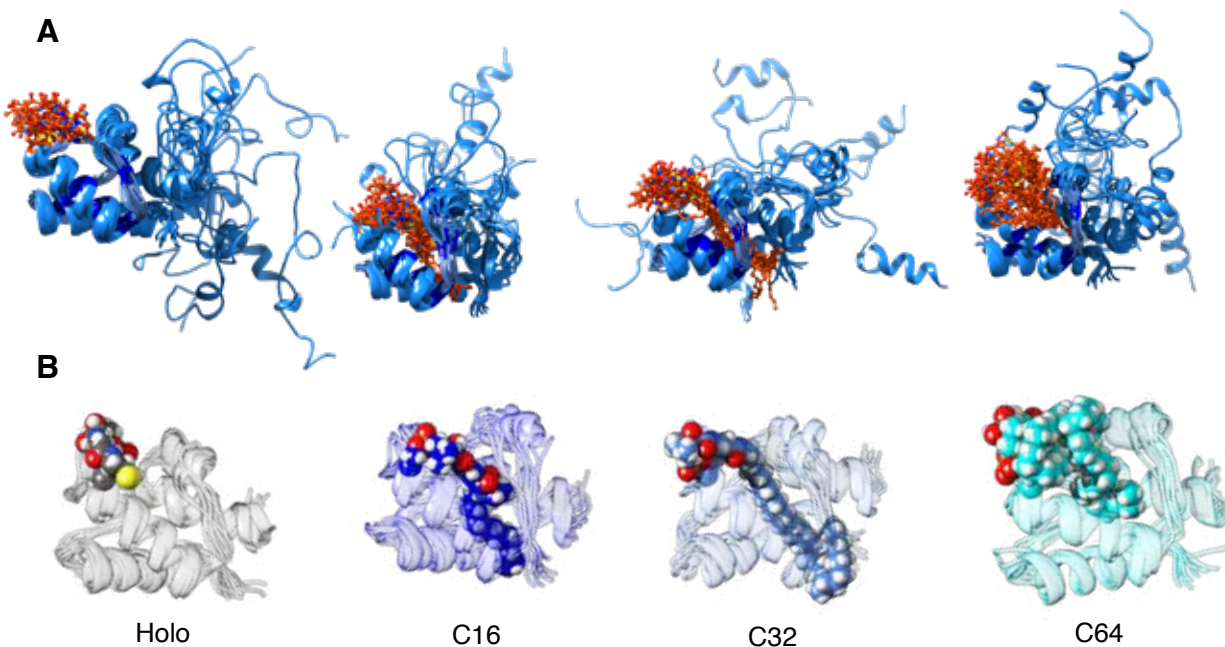
Figure 13: Structural ensemble of AcpM in the *holo*, C16, C32, and C64 states showing the unique modes of substrate accommodation throughout the simulations **(A)**. Space filling model for averaged structures of AcpM in different conformational states **(B)**.

Before loading a C16 acyl chain onto AcpM which marks the initiation of FAS II stage of mycolic acid biosynthesis AcpM exists in the *holo* state, and its trajectory serves as a control that helps understand its dynamics in this premature state. Compared to C16 and C32 acyl AcpM, *holo* AcpM has the lowest RMSD (Figure 14) and radius of gyration (Figure 17) core residues indicating that these residues are more rigid when not bound to acyl chains. In terms of distance, *holo* AcpM had the longest distance between the tail residues' center of mass and the phosphorus atom in phosphopantetheine. This is largely because phosphopantetheine contains polar atoms that largely favor hydrogen bonding interactions with the solvent environment, as opposed to the hydrophobic environment within the AcpM tunnel.
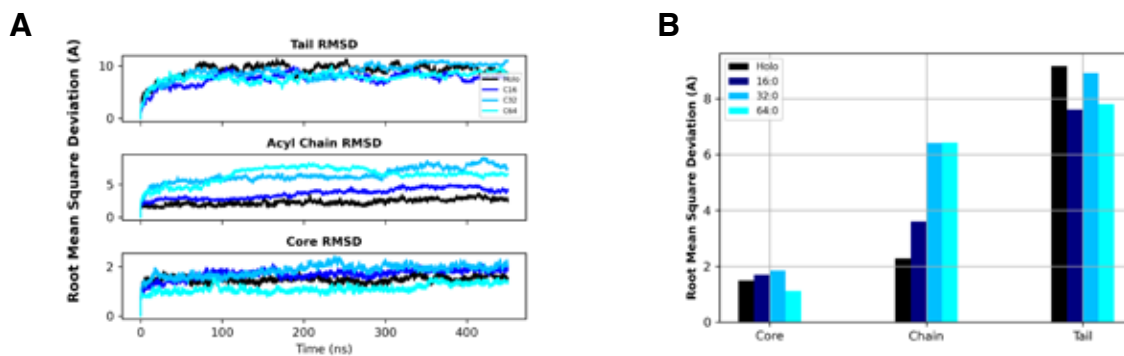
Figure 14*: Average root mean square deviation (RMSD) for holo, 16:0, 32:0, and 64:0 acyl-AcpM for triplicate runs over the course of 450ns broken down into tail, chain, and core regions of AcpM **(A)**. RMSD bar plot broken down into tail, core, and chain regions **(B)**.

C16 AcpM shows the most interesting dynamics compared to the other AcpM models. Over the course of the simulation, it has the smallest distance between the tail's center of mass and the the phosphorus atom in phosphopantetheine. The tail residues in C16 AcpM are also observed to have the lowest RMSD values across all models. The individual RMSF results for C16 are also the highest of all models. Together, the increased dynamics observed in the RMSF could account for the high number of protein-protein interactions during FAS II (Figure 15).

Generally, the trends for substrate accommodation in C64 AcpM behaved in a way that was not in accordance with the expected increase in size when compared to *holo,* C16 and C32 AcpM. For instance, in terms of pocket calculations (Figure 16), the cavity volume of C64 appeared to have the second smallest volume after holo AcpM when it should have the largest volume of all other AcpM systems. Since MD simulations do not consider the breakage or formation of chemical bonds, the thioester bond that connects phosphopantetheine to the acyl chain is exposed to the solvent and most likely

prone to hydrolysis which is known to be the case for longer acyl chains[29]. The lack of sequestration as well as the relatively smaller pocket volume for C64 AcpM support this claim and would most likely be hydrolyzed. Furthermore, the absence of substrate sequestered conformations indicates that before reaching C64 the acyl chain will most likely already exist bound to CoA and modified by other enzymes that act independently of AcpM. The tail regions in C64 AcpM show the lowest RMSF amongst all forms of AcpM. This may be due by the hydrophobic patch created by the aggregation of C64 outside of AcpM that contributes to stabilizing interactions from the tail residues, although direct interactions between the tail and C64 were not visualized.

Figure 15: Average root mean square fluctuations (RMSF) for AcpM **(A)** broken down into core **(B)** and tail **(C)** regions. Reference image of AcpM **(D).**

The root mean square fluctuation (RMSF) measures the deviation for every amino acid in a protein from their average position over the course of a simulation (Figure 15). This parameter helps evaluate the protein dynamics at the residue level. A general increasing trend is observed for tail residues in all forms of AcpM given the large dynamics associated with the disordered tail region. Although the curvature in the RMSF graph are representatives of the secondary structures and appear similar for all states, AcpM in the *holo* and C32 states have the lowest magnitudes of RMSF values, while C16 and C64 possess the RMSF values with the highest magnitudes. Discarding C64 as an outlier due to its inability to sequester its acyl chain, the high magnitude of RMSF values observed in C16 AcpM reflects the increased dynamics that are necessary to produce the specific protein-protein interactions that are essential for FAS II. On the other hand, one explanation for the increased rigidity observed in the low RMSF values for *holo* and C32 could be due to these states happening before and towards the end of protein-protein interactions that could require more dynamics.
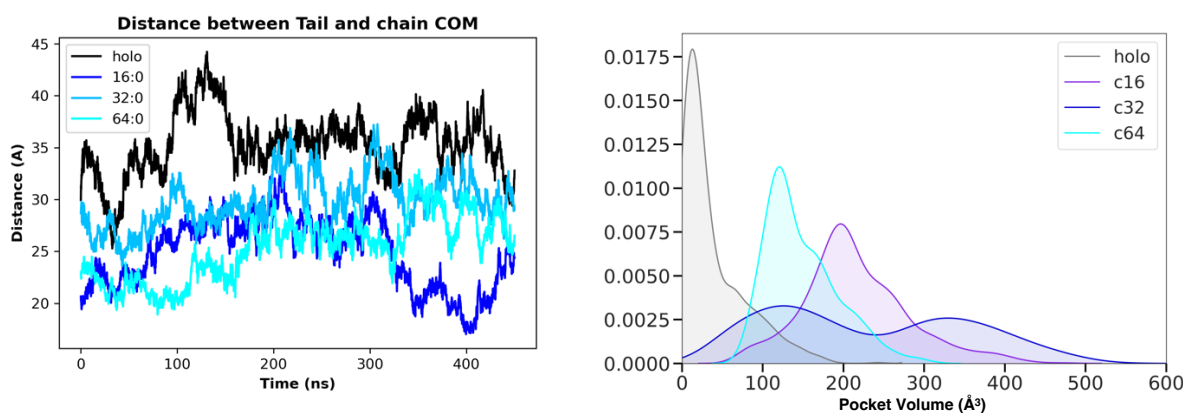

Figure 16: Average distance between the tail (81-115) residues center of mass and phosphorus atom in phosphopantetheine **(A)**. Pocket volume histogram for different acylation states in AcpM **(B)**.

In terms of the RMSD and radius of gyration values for the acyl chains there is an increasing trend that agrees with their expected behavior due to their increasing degrees of freedom with size. A negligible difference is observed between RMSD values for C32 and C64 AcpM which further supports the claim that C64 AcpM acts as the upper limit of sequestered acyl intermediates. In terms of the pocket volume distribution for the different forms of AcpM, C32 shows a bimodal distributions of values around 100 and 310Å^3, while C16 is midway between these two values with an average pocket volume of 200Å^3. The bimodal distribution observed for the pocket volumes in C32 AcpM could account for two distinct modes of accommodation (Figure 16). For instance, the larger pocket volume in C32 AcpM represents a conformation where the acyl chain is threaded through, such as the case of C16. On the other hand, the smaller pocket volume could account for a conformation that more closely resembles C64 AcpM where the chain resides outside of the pocket and is not accommodated by the alpha helices.

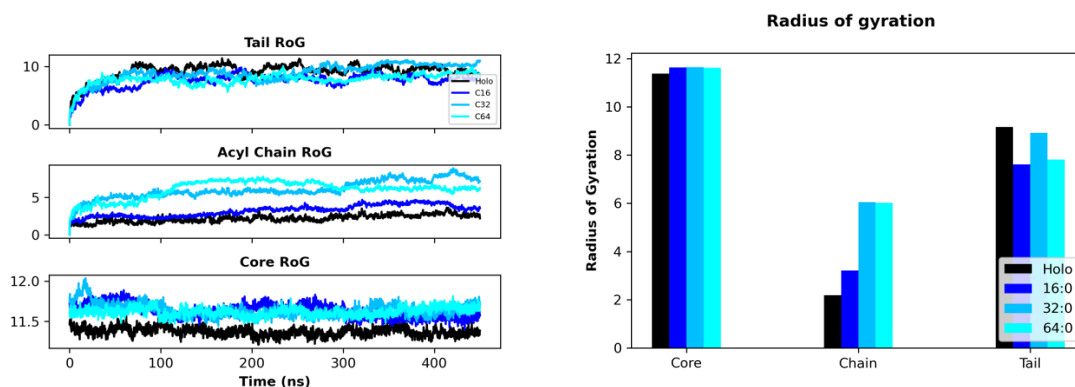**Figure 17:** Averaged out radius of gyration for triplicate runs highlighting the different regions in AcpM containing acyl chains of different lengths.

By observing and comparing the dynamics from the different acylated forms of AcpM in explicit solvent we begin to observe interesting patterns that help us infer how the meromycolate chain portion of mycolic acids are assembled. At the same time, the

different conformational dynamics form the starting points for interface studies, where the

residues in helix II can be docked to the carrier protein recognition domains observed in

partner proteins. Other partner proteins from FAS II have subdomains that probe the

chain length, as they only act on carrier proteins bound to longer structures. The

structural modifications to the meromycolate chain, however, were not accounted for in

this study. Although these modifications have not been confirmed to be carrier protein

dependent, their influence on acyl chain geometry cannot be neglected and would be

interesting to see in future studies.

References

(1) Brown III, J. W. The International Journal of Tuberculosis and Lung Disease. *JAMA J. Am. Med. Assoc.* **1998**, *280* (13), 1200-a-1200. https://doi.org/10.1001/jama.280.13.1200-a.

(2) Sukhithasri, V.; Vinod, V.; Varma, S.; Biswas, R. Mycobacterium Tuberculosis Treatment Modalities and Recent Insights. *Curr. Drug Deliv.* **2014**, *11* (6), 744–752. https://doi.org/10.2174/1567201811666140619121728.

(3) Nataraj, V.; Varela, C.; Javid, A.; Singh, A.; Besra, G. S.; Bhatt, A. Mycolic Acids: Deciphering and Targeting the A Chilles' Heel of the Tubercle Bacillus. *Mol. Microbiol.* **2015**, *98* (1), 7–16. https://doi.org/10.1111/mmi.13101.

(4) Shyam, M.; Shilkar, D.; Verma, H.; Dev, A.; Sinha, B. N.; Brucoli, F.; Bhakta, S.; Jayaprakash, V. The Mycobactin Biosynthesis Pathway: A Prospective Therapeutic Target in the Battle against Tuberculosis. *J. Med. Chem.* **2021**, *64* (1), 71–100. https://doi.org/10.1021/acs.jmedchem.0c01176.

(5) Smith, I. *Mycobacterium Tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clin. Microbiol. Rev.* **2003**, *16* (3), 463–496. https://doi.org/10.1128/CMR.16.3.463-496.2003.

(6) Pontali, E.; Raviglione, M. C.; Migliori, G. B. Regimens to Treat Multidrug-Resistant Tuberculosis: Past, Present and Future Perspectives. *Eur. Respir. Rev.* **2019**, *28* (152), 190035. https://doi.org/10.1183/16000617.0035-2019.

(7) Ballinger, E.; Mosior, J.; Hartman, T.; Burns-Huang, K.; Gold, B.; Morris, R.; Goullieux, L.; Blanc, I.; Vaubourgeix, J.; Lagrange, S.; Fraisse, L.; Sans, S.; Couturier, C.; Bacqué, E.; Rhee, K.; Scarry, S. M.; Aubé, J.; Yang, G.; Ouerfelli, O.; Schnappinger, D.; Ioerger, T. R.; Engelhart, C. A.; McConnell, J. A.; McAulay, K.; Fay, A.; Roubert, C.; Sacchettini, J.; Nathan, C. Opposing Reactions in Coenzyme A Metabolism Sensitize *Mycobacterium Tuberculosis* to Enzyme Inhibition. *Science* **2019**, *363* (6426), eaau8959. https://doi.org/10.1126/science.aau8959.

(8) Vincent, A. T.; Nyongesa, S.; Morneau, I.; Reed, M. B.; Tocheva, E. I.; Veyrier, F. J. The Mycobacterial Cell Envelope: A Relict From the Past or the Result of Recent Evolution? *Front. Microbiol.* **2018**, *9*, 2341. https://doi.org/10.3389/fmicb.2018.02341.

(9) Kolattukudy, P. E.; Fernandes, N. D.; Azad, A. K.; Fitzmaurice, A. M.; Sirakova, T. D. Biochemistry and Molecular Genetics of Cell-wall Lipid Biosynthesis in Mycobacteria. *Mol. Microbiol.* **1997**, *24* (2), 263–270. https://doi.org/10.1046/j.1365-2958.1997.3361705.x.

(10) Vander Beken, S.; Al Dulayymi, J. R.; Naessens, T.; Koza, G.; Maza-Iglesias, M.; Rowles, R.; Theunissen, C.; De Medts, J.; Lanckacker, E.; Baird, M. S.; Grooten, J. Molecular Structure of the Mycobacterium Tuberculosis Virulence Factor, Mycolic Acid, Determines the Elicited Inflammatory Pattern. *Eur. J. Immunol.* **2011**, *41* (2), 450–460. https://doi.org/10.1002/eji.201040719.

(11) Ryll, R.; Kumazawa, Y.; Yano, I. Immunological Properties of Trehalose Dimycolate (Cord Factor) and Other Mycotic Acid-Containing Glycolipids--A Review. *Microbiol. Immunol.* **2001**, *45* (12), 801–811. https://doi.org/10.1111/j.1348-0421.2001.tb01319.x.

(12) Daffé, M.; Draper, P. The Envelope Layers of Mycobacteria with Reference to Their Pathogenicity. In *Advances in Microbial Physiology*; Elsevier, 1997; Vol. 39, pp 131–203. https://doi.org/10.1016/S0065-2911(08)60016-8.

(13) Beld, J.; Lee, D. J.; Burkart, M. D. Fatty Acid Biosynthesis Revisited: Structure Elucidation and Metabolic Engineering. *Mol. Biosyst.* **2015**, *11* (1), 38–59. https://doi.org/10.1039/C4MB00443D.

(14) Mindrebo, J. T.; Misson, L. E.; Johnson, C.; Noel, J. P.; Burkart, M. D. Activity Mapping the Acyl Carrier Protein: Elongating Ketosynthase Interaction in Fatty Acid Biosynthesis. *Biochemistry* **2020**, *59* (38), 3626–3638. https://doi.org/10.1021/acs.biochem.0c00605.

(15) Marrakchi, H.; Lanéelle, M.-A.; Daffé, M. Mycolic Acids: Structures, Biosynthesis, and Beyond. *Chem. Biol.* **2014**, *21* (1), 67–85. https://doi.org/10.1016/j.chembiol.2013.11.011.

(16) Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. V.; Eiglmeier, K.; Gas, S.; Iii, C. E. B.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M. A. Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence. **1998**, *393*, 7.

(17) Chen, A.; Re, R. N.; Burkart, M. D. Type II Fatty Acid and Polyketide Synthases: Deciphering Protein–Protein and Protein–Substrate Interactions. *Nat. Prod. Rep.* **2018**, *35* (10), 1029–1045. https://doi.org/10.1039/C8NP00040A.

(18) Elad, N.; Baron, S.; Peleg, Y.; Albeck, S.; Grunwald, J.; Raviv, G.; Shakked, Z.; Zimhony, O.; Diskin, R. Structure of Type-I Mycobacterium Tuberculosis Fatty Acid Synthase at 3.3 Å Resolution. *Nat. Commun.* **2018**, *9* (1), 3886. https://doi.org/10.1038/s41467-018-06440-6.

(19) Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.; McCammon, J. A.; Opella, S. J.; Tsai, S.-C.; Burkart, M. D. Trapping the Dynamic Acyl Carrier Protein in Fatty Acid Biosynthesis. *Nature* **2014**, *505* (7483), 427–431. https://doi.org/10.1038/nature12810.

(20) Choi, J. W.; Da Silva, N. A. Improving Polyketide and Fatty Acid Synthesis by Engineering of the Yeast Acetyl-CoA Carboxylase. *J. Biotechnol.* **2014**, *187*, 56–59. https://doi.org/10.1016/j.jbiotec.2014.07.430.

(21) Yan, Q.; Cordell, W. T.; Jindra, M. A.; Courtney, D. K.; Kuckuk, M. K.; Chen, X.; Pfleger, B. F. Metabolic Engineering Strategies to Produce Medium-Chain Oleochemicals via Acyl-ACP:CoA Transacylase Activity. *Nat. Commun.* **2022**, *13* (1), 1619. https://doi.org/10.1038/s41467-022-29218-3.

(22) Beld, J.; Blatti, J. L.; Behnke, C.; Mendez, M.; Burkart, M. D. Evolution of Acyl-ACP Thioesterases and β-Ketoacyl-ACP Synthases Revealed by Protein–Protein Interactions. *J. Appl. Phycol.* **2014**, *26* (4), 1619–1629. https://doi.org/10.1007/s10811-013-0203-4.

(23) Jaremko, M. J.; Davis, T. D.; Corpuz, J. C.; Burkart, M. D. Type II Non-Ribosomal Peptide Synthetase Proteins: Structure, Mechanism, and Protein–Protein Interactions. *Nat. Prod. Rep.* **2020**, *37* (3), 355–379. https://doi.org/10.1039/C9NP00047J.

(24) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. *Nat Prod Rep* **2014**, *31* (1), 61–108. https://doi.org/10.1039/C3NP70054B.

(25) Sztain, T.; Patel, A.; Lee, D. J.; Davis, T. D.; McCammon, J. A.; Burkart, M. D. Modifying the Thioester Linkage Affects the Structure of the Acyl Carrier Protein. *Angew. Chem. Int. Ed.* **2019**, *58* (32), 10888–10892. https://doi.org/10.1002/anie.201903815.

(26) Sztain, T.; Bartholow, T. G.; Lee, D. J.; Casalino, L.; Mitchell, A.; Young, M. A.; Wang, J.; McCammon, J. A.; Burkart, M. D. Decoding Allosteric Regulation by the Acyl Carrier Protein. *Proc. Natl. Acad. Sci.* **2021**, *118* (16), e2025597118. https://doi.org/10.1073/pnas.2025597118.

(27) Wang, X.; Ni, D.; Liu, Y.; Lu, S. Rational Design of Peptide-Based Inhibitors Disrupting Protein-Protein Interactions. *Front. Chem.* **2021**, *9*, 682675. https://doi.org/10.3389/fchem.2021.682675.

(28) Wong, H. C.; Liu, G.; Zhang, Y.-M.; Rock, C. O.; Zheng, J. The Solution Structure of Acyl Carrier Protein from Mycobacterium Tuberculosis. *J. Biol. Chem.* **2002**, *277* (18), 15874–15880. https://doi.org/10.1074/jbc.M112300200.

(29) Zornetzer, G. A.; Fox, B. G.; Markley, J. L. Solution Structures of Spinach Acyl Carrier Protein with Decanoate and Stearate ,. *Biochemistry* **2006**, *45* (16), 5217–5227. https://doi.org/10.1021/bi052062d.

(30) Graether, S. P. Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Front. Mol. Biosci.* **2019**, *5*, 118. https://doi.org/10.3389/fmolb.2018.00118.

(31) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. In *Advances in Protein Chemistry*; Elsevier, 2003; Vol. 66, pp 27–85. https://doi.org/10.1016/S0065-3233(03)66002-X.

(32) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package: Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2), 198–210. https://doi.org/10.1002/wcms.1121.

(33) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

(34) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267* (5612), 585–590. https://doi.org/10.1038/267585a0.

(35) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539. https://doi.org/10.1038/msb.2011.75.

(36) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537–W541. https://doi.org/10.1093/nar/gks375.

(37) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(38) Zhao, S.; Schaub, A. J.; Tsai, S.-C.; Luo, R. Development of a Pantetheine Force Field Library for Molecular Modeling. *J. Chem. Inf. Model.* **2021**, *61* (2), 856–868. https://doi.org/10.1021/acs.jcim.0c01384.

(39) Bunkoczi, G.; Pasta, S.; Joshi, A.; Wu, X.; Kavanagh, K. L.; Smith, S.; Oppermann, U. Mechanism and Substrate Recognition of Human Holo ACP Synthase. *Chem. Biol.* **2007**, *14* (11), 1243–1253. https://doi.org/10.1016/j.chembiol.2007.10.013.