Substrate access mechanism in a novel phospholipase A of *Pseudomonas aeruginosa*

Inaugural dissertation

for the attainment of the title of doctor

in the Faculty of Mathematics and Natural Sciences

at the Heinrich Heine University Düsseldorf

presented by

Sabahuddin Ahmad

from Allahabad, India

Düsseldorf, December 2021

from the Institute for Pharmaceutical and Medicinal Chemistry
at the Heinrich Heine University Düsseldorf

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for Collaborative Research Center 1208 (CRC 1208) “Identity and Dynamics of Membrane Systems - From Molecules to Cellular Functions”. The computing time was provided by the John von Neumann Institute for Computing (NIC) on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. Holger Gohlke
Co-supervisor: Prof. Dr. Karl-Erich Jaeger

Date of the oral examination: 08.04.2022

Affidavit

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the 'Principles for the Safeguarding Good Scientific Practice at Heinrich Heine University Düsseldorf'.

This dissertation has not been submitted in this or a similar form to any other Institution, and I have no unsuccessful defense attempts.

Düsseldorf, December 6th 2021

_____
(Sabahuddin Ahmad)

To my Supervisor(s) and my Parents…

“… He found you lost and guided (you)”

The Qur’an 93:7

Table of contents

I. PUBLICATIONS WITHIN THE SCOPE OF THE DOCTORAL PROGRAM . I

II. LIST OF ABBREVIATIONS II

III. ABSTRACT IV

IV. ZUSAMMENFASSUNG V

1. INTRODUCTION 1

2. BACKGROUND 5

2.1 *PSEUDOMONAS AERUGINOSA,* A MEDICALLY RELEVANT BACTERIUM 5
2.2 BACTERIAL PHOSPHOLIPASES IN PATHOGENESIS 8
2.3 PLAF IS A PLA_1, A VIRULENCE FACTOR OF *PSEUDOMONAS AERUGINOSA* 10
2.4 STRUCTURE OF PLAF, DIMERIZATION AND ITS ORIENTATION IN THE MEMBRANE 12
2.5 ROLE OF PROTEIN TUNNELS AND THEIR IDENTIFICATION TOOLS 15
2.6 STEERED MOLECULAR DYNAMICS SIMULATIONS 17
2.7 FREE ENERGY COMPUTATION FROM UMBRELLA SAMPLING SIMULATIONS 21
2.8 EXPERIMENTAL RELATION TO COMPUTATIONS 24
2.8.1 RELATING PMF TO EXPERIMENTS 24
2.8.2 ABSOLUTE BINDING FREE ENERGY FROM COMPUTED PMF 25
2.8.3 BLOCKING ACCESS OF PLAF SUBSTRATES 25

3. SCOPE OF THE THESIS 27

4. MATERIALS AND METHODS 29

4.1 IDENTIFICATION OF THE ACCESS TUNNELS 29
4.2 STARTING STRUCTURE PREPARATION 29
4.3 SIMULATED EXTRACTION OF SUBSTRATES FROM THE MEMBRANE 31
4.4 UMBRELLA SAMPLING AND POTENTIAL OF MEAN FORCE CALCULATIONS 33
4.5 ABSOLUTE BINDING FREE ENERGY FROM COMPUTED PMF 34
4.6 BLOCKING ACCESS OF THE PLAF SUBSTRATES 35
4.7 EGRESS OF PLAF PRODUCTS 35

5. RESULTS 37

5.1 ACCESS PATHWAYS TO THE CATALYTIC SITE IN PLAF 37
5.2 GPL AND LGPL SUBSTRATE EXTRACTION INTO SOLVENT AND ACYL CHAIN MOBILITY .40
5.3 ACCESS MODES OF GPL AND LGPL SUBSTRATES INTO PLAF 44
5.4 POTENTIALS OF MEAN FORCE OF DLPG ACCESS MODES 50
5.5 POTENTIALS OF MEAN FORCE FOR ACCESSES OF OTHER SUBSTRATES 56

Table of contents

5.6 Energetics of substrate access into T2 in dimeric PlaF 58
5.7 Computational costs to determine PMFs related to substrate access..... 60
5.8 Tryptophan substitutions in T2 hamper DLPG access 60
5.9 Potential egress pathways of PlaF products..... 65
6. DISCUSSION..... 69
7. CONCLUSION AND SIGNIFICANCE..... 74
8. PERSPECTIVES..... 76
9. ACKNOWLEDGEMENTS..... 78
10. CURRICULUM VITAE..... 79
11. BIBLIOGRAPHY..... 82

I. Publications within the scope of the doctoral program

A) The methods, corresponding results, and discussions written in this thesis are taken from the following peer-reviewed publication[†]:

Ahmad S., Strunk C.H., Schott-Verdugo S.N., Jaeger K.-E., Kovačić F., Gohlke H.; Substrate access mechanism in a novel membrane-bound phospholipase A of *Pseudomonas aeruginosa* concordant with specificity and regioselectivity.

Journal of Chemical Information and Modeling, 2021, 61(11), 5626-5643.

Impact factor reported for 2020: 4.956

[†]This publication forms the basis of chapters 4, 5, and 6 of this thesis. My contribution in this publication involves computational investigation, analysis, visualization, and writing.

B) Under MB Train (an integrated research training group of the Collaborative Research Centre 1280, HHU), the participation in the method training course "Plant Biophysics - Membrane Biophysics" resulted in the following peer-reviewed publication:

Dreyer I., Spitz O.[‡], Kanonenberg K.[‡], Montag K.[‡], Handrich M.R.[‡], **Ahmad S.**[‡], Schott-Verdugo S., Navarro-Retamal C., Rubio-Meléndez M.E., Gomez-Porras J.L., Riedelsberger J., Molina-Montenegro M.A., Succurro A., Zuccaro A., Gould S.B., Bauer P., Schmitt L., Gohlke H.; Nutrient exchange in arbuscular mycorrhizal symbiosis from a thermodynamic point of view.

New Phytologist, 2019, 222(2), 1043-1053.

Impact factor reported for 2020: 10.151

[‡]These authors contributed equally to this work.

II. List of abbreviations

UND Undecanoic acid
2LMG 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphoglycerol
AFM Atomic Force Microscopy
AIDS Acquired Immunodeficiency Syndrome
AMBER Assisted Model Building with Energy Refinement
APL Area-Per-Lipid
AsMD Adaptive biased Steered Molecular Dynamics
CF Cystic Fibrosis
CHARMM Chemistry at Harvard Macromolecular Mechanics
CMC Critical Micelle Concentration
COM Center of Mass
CPK Corey-Pauling-Koltun
di-PlaF PlaF as a dimer
DLPE 1,2-dilauroyl-*sn*-glycero-3-phosphorylethanolamine
DLPG 1,2-dilauroyl-*sn*-glycero-3-phosphoglycerol
DSPG 1,2-distearoyl-*sn*-glycero-3-phosphoglycerol
DSF Diffusible Signal Factor
FA Fatty Acid
GAFF General AMBER Force Field
GPL Glycerophospholipid
JM Juxta-Membrane
kcal 1 kcal ≈ 4.18 kJ
K_d Dissociation Constant
K_m Michaelis-Menten Constant
LysoPLA$_{1/2}$ Lysophospholipase A$_{1/2}$
LGPL Lysoglycerophospholipid
LAU Lauric acid
MD Molecular Dynamics
MYR Myristic acid
PDB Protein Data Bank
PE Phosphatidylethanolamine head group
PG Phosphatidylglycerol head group
PGR Phosphatidylglycerol from lysophospholipid
PLA$_1$ Phospholipase A$_1$
PLA$_2$ Phospholipase A$_2$
PLB Phospholipase B

List of abbreviations

PLC Phospholipase C
PLD Phospholipase D
PlaF$_{WT}$ PlaF wild-type
PlaF$_{A}$ Chain A of PlaF dimer
PlaF$_{B}$ Chain B of PlaF dimer
Δ*plaF* mutant without PlaF gene
PME Particle Mesh Ewald
PMF Potential of Mean Force
PPM Positioning of Proteins in Membranes
RESP Restrained Electrostatic Potential method
s-PlaF$_{A}$ Non-tilted, chain A of PlaF dimer
s-PlaF$_{B}$ Non-tilted, chain B of PlaF dimer
sMD Steered Molecular Dynamics
T2SS type II secretion system
t-PlaF$_{A}$ tilted, chain A of PlaF dimer
t-PlaF$_{B}$ tilted, chain B of PlaF dimer
TM Transmembrane
Trp Tryptophan
US Umbrella Sampling
VMD Visual Molecular Dynamics
WHAM Weighted Histogram Analysis Method
WHO World Health Organization

Abstract

III. Abstract

PlaF is an integral inner membrane phospholipase A_1 (PLA_1) from *Pseudomonas aeruginosa* that modulates the membrane glycerophospholipid (GPL) composition and thus promotes the virulence of *P. aeruginosa*. Interestingly, PlaF can exist as both monomer and dimer, but only the monomeric form shows the enzymatic activity. It is well established through experiments that the activity of PlaF is regulated by a dimer-to-monomer transition followed by monomer tilting in the membrane. However, how do substrates reach the active site for hydrolysis and how the characteristics of the tunnels connecting the active site govern the activity, specificity, and regioselectivity of PlaF for GPL substrates remains unknown. In this study, I combined all-atom molecular dynamics (MD) simulations together with configurational free energy calculations to identify access pathways of substrates to the active site of PlaF. Among the major outcomes, supported by experimental collaboration this study identifies a distinct tunnel that is involved in the PlaF activity. The preferable access of GPLs with the *sn*-1 acyl chain first is in excellent agreement with the demonstrated PLA_1 activity of PlaF. Further, the favorable free-energy of binding of the medium-chain GPLs is related to the acyl chain length specificity of PlaF, determined by structural features of the tunnels. Unbiased MD simulations revealed egress routes for fatty acid (FA) products, which transports them to the dimer interface, and likely lead to the fatty acid-triggered dimerization, explaining how products mediate feedback-regulation of PlaF. These atomistic-level studies enhance our understanding of the activity regulation of PlaF and its specificity toward GPL substrates. Taken together, this study opens up opportunities for developing potential drugs that inhibit PlaF to combat *P. aeruginosa* virulence during infection.

IV. Zusammenfassung

PlaF aus *Pseudomonas aeruginosa* ist eine integrale Innenmembran-Phospholipase A_1 (PLA_1), welche die Zusammensetzung der Membran-Glycerophospholipide (GPL) verändert und damit die Virulenz von *P. aeruginosa* fördert. Interessanterweise kann PlaF sowohl in monomeren als auch in dimeren Konfigurationen existieren, die Aktivität wurde jedoch nur im monomeren Zustand nachgewiesen. Es ist durch Experimente gut belegt, dass die PlaF-Aktivität durch einen Dimer-zu-Monomer-Übergang, gefolgt von einem Kippen des Monomers in der Membran, reguliert wird. Wie genau jedoch Substrate das aktive Zentrum erreichen und wie die Charakteristika der Tunnel die Aktivität, Spezifität und Regioselektivität von PlaF für natürliche GPL-Substrate bestimmen, blieb bisher ungeklärt. In dieser Studie habe ich *All-Atom*-Molekulardynamik-Simulationen (MD) mit Berechnungen der freien Konfigurationsenergie kombiniert, um Zugangswege von GPL-Substraten zum aktiven Zentrum von PlaF zu identifizieren. Als eines der wichtigsten Ergebnisse, unterstützt durch experimentelle Zusammenarbeit, beschreibt diese Studie einen Tunnel, der die Aktivität von PlaF mitbestimmt. Der bevorzugte Zugang von GPLs mit der *sn*-1 Acylkette zuerst steht in hervorragender Übereinstimmung mit der nachgewiesenen PLA_1-Aktivität von PlaF. Des Weiteren steht die günstige freie Bindungsenergie der mittelkettigen GPLs im Zusammenhang mit der Acylkettenlängenspezifität von PlaF, die durch strukturelle Merkmale der Tunnel bestimmt wird. Unbeeinflusste MD-Simulationen ergaben Austrittswege für Fettsäureprodukte, die sie zur Dimer-Grenzfläche transportieren und wahrscheinlich zur fettsäuregetriggerten Dimerisierung führen, was erklärt, wie die Produkte die Feedback-Regulation von PlaF vermitteln. Diese Studien auf atomistischer Ebene verbessern unser Verständnis der Aktivitätsregulation von PlaF und seiner Spezifität gegenüber GPL-Substraten. Insgesamt eröffnet diese Studie Möglichkeiten für die Entwicklung potenzieller Medikamente, die PlaF hemmen, um die Virulenz von *P. aeruginosa* während einer Infektion zu bekämpfen.

1. Introduction

The bacterium *Pseudomonas aeruginosa* is a frequent cause of nosocomial infections (1), affecting various subpopulations of immunocompromised patients. It is a versatile opportunistic pathogen and causes infections in both mammalian and non-mammalian hosts (2). Recently, the World Health Organization (WHO) included *P. aeruginosa* along with *Acinetobacter baumannii* and members of *Enterobacteriaceae* in the critical list of antibiotic-resistant "priority pathogens" (3) for which there is an immediate need for research and development initiatives related to new antibiotics. The pathogenicity of this bacterium mainly relies on cell-associated and extracellular virulence factors (4) (**chapter 2.1**). Among those virulence factors, phospholipases (5, 6) are particularly important as they contribute to an integral role in infections through hydrolysis of cellular glycerophospholipids (GPLs) (7). Depending on substrate's hydrolysis site, various classes of phospholipases are distinguished (8). Among these classes, the phospholipase A_1 (PLA_1) hydrolyze the GPLs at the *sn*-1 position into lysoglycerophospholipid (LGPL) and fatty acid (FA) (9, 10) (**chapter 2.2**).

By forming a bilayer, GPLs maintain the vital permeability barrier for cells and organelles (11), while membrane-bound LGPLs can impair membrane integrity in Gram-Negative bacteria (12, 13). GPLs (14) and LGPLs (15, 16) can regulate both the function and stability of the membrane proteins. Interestingly, biofilm formation as well as the growth phase transitions in *P. aeruginosa* results from the modification of membrane GPL composition (17, 18). FAs belong to the diffusible signal factor family (DSF) and are regarded as signal molecules because they can diffuse through cell membranes and contribute to the regulation of diverse biological functions in various Gram-Negative pathogens (19). In *P. aeruginosa,* DSFs promote biofilm formation and antibiotic resistance (20, 21). Moreover, relative amounts of GPLs are associated with FA composition in the

membrane, which allows them to thrive in a wide range of environments (17) (**chapter 2.2**). Increased tolerance of *P. aeruginosa* to antimicrobial peptides in the lungs of cystic fibrosis (CF) patients is a notable manifestation of such an interspecies signaling (20). Hence, enzymes with PLA_1 activity can be important therapeutic targets. Recently, an enzyme with this function has been identified in *P. aeruginosa.*

This enzyme is a membrane-associated α/β-hydrolase PA2949, also called PlaF, from *P. aeruginosa* (22). PlaF is a cytoplasmic, integral, inner membrane (IM) protein with the catalytic domain in the periplasmic space. PlaF exhibits PLA_1 activity that decreases with increasing length of the substrate's acyl chains, ranging from C12 to C18 (23). Interestingly, higher PLA_1 activity was observed for LGPL substrates compared to GPL substrates of the same acyl chain length (24). This indicates that not only the length of the acyl chains but also the number of acyl chains of the substrates influence the activity profile of PlaF. The knockout studies show a defined role of PlaF in the membrane GPL homeostasis, as it reduces biofilm formation (23) (**chapter 2.3**). This suggests PlaF as a virulence factor and a potential drug target of *P. aeruginosa* for which the regulation mechanism needs to be understood.

Interestingly, crosslinking and micro-scale thermophoresis experiments show that PlaF is active only in the monomeric configuration, even though it can exist as both monomer and dimer (23). The crystal structure of PlaF, Protein Data Bank (PDB) id: 6I8W (25, 26) reveals that the homodimer is formed by interactions between the transmembrane (TM) and juxtamembrane (JM) regions. The structure of the PlaF homodimer contains co-crystallized endogenous ligands, including the FAs myristic acid (MYR) and undecanoic acid (UND) (Figure 1A). Interestingly, these FAs are *in vivo* products co-purified with PlaF (23) and are found non-covalently bound to the active site cavity of the respective chains (26). In a recent collaboration, it was found that the active site of PlaF is connected to the surface with three distinct tunnels, which form a T-shaped cleft

(23). The structure was further studied and described to show how the activity of PlaF is regulated by product-feedback inhibition and a dimer-to-monomer transition followed by monomer tilting in the membrane (23). Interestingly, these tilting motion orients one of the tunnels of the active site cleft perpendicular to the membrane plane (Figure 1B), seemingly facilitating direct access of substrates into the active site cleft from the membrane (24) (see **chapter 2.4** for detailed information). By contrast, in the dimer configuration, the opening of the active site cleft is more than 5 Å away from the membrane interface (Figure 1A) (**chapter 2.4**).

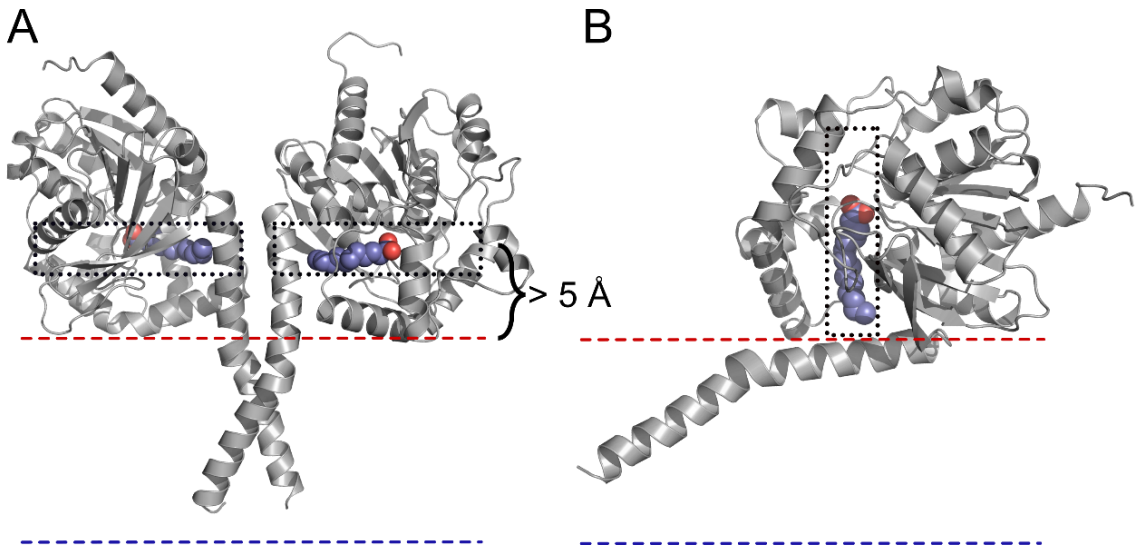

Figure 1: Orientation of PlaF in the inner membrane. A) In dimeric PlaF, the active site tunnel (black box) is located > 5 Å above the membrane. Blue spheres represent the co-crystallized products, myristic acid in chain A of dimeric PlaF (left) and undecanoic acid in chain B of dimeric PlaF (right) within the active site tunnel. B) Chain A of the dimeric PlaF in the tilted state; this configuration allows direct contact of the active site tunnel to the membrane interface. Blue spheres represent the co-crystallized PlaF product myristic acid. The upper leaflet of the membrane faces the periplasm (red), while the lower leaflet faces the cytoplasm (blue). Figure adapted from ref. (24).

From previous studies, however, it was not clear how substrates/products reach/leave the active site before/after hydrolysis and how the characteristics of the active site tunnels determine the activity and specificity of PlaF for medium-chain substrates. Therefore, in the presented study, I have first identified the main tunnels which connect the active site of PlaF to its surface (**chapter 2.5**). To

obtain substrate access pathways along the identified tunnels, I used steered molecular dynamics (sMD) simulations (**chapter 2.6**). To evaluate the most energetically favorable pathway for substrate access, umbrella sampling simulations (US) were performed along with potential of mean force (PMF) computations (**chapter 2.7**). The obtained results were used to calculate the binding free energy of the substrate-PlaF binding. The findings were further validated by site-directed mutagenesis experiments and the activity profile of PlaF was re-determined (**chapter 2.8**). To identify the role of different tunnels and possibly the egress route of PlaF products, a set of unbiased all-atom MD simulations was also performed.

The results map out a distinct tunnel in PlaF for substrate access and suggest egress routes for the products. The outcomes of this study shall enhance our understanding related to the specificity and activity of the novel PLA_1, PlaF.

2. Background

In this chapter, I will first describe the organism, *P. aeruginosa*. Then, I will review the enzyme class of phospholipases in general and discuss the clinical significance of bacterial phospholipases. Next, I will describe the biochemical and structural properties of a membrane-bound novel phospholipase, PlaF, as it forms the basis of my work. Later, I will introduce the methods and techniques which I considered to: 1) identify and analyze the tunnels connecting the active site to the surface, 2) simulate the substrate access along the identified pathways, and 3) compute the energetics of substrates access along the pathways of PlaF. In the last subsection, I will describe the methods that aided me to link my computations to experimental studies.

2.1 *Pseudomonas aeruginosa,* a medically relevant bacterium

Pseudomonas aeruginosa is a motile, rod-shaped Gram-Negative bacterium and was first described by pharmacist Carle Gessard in his study entitled 'On the blue and green coloration of bandages' in 1882 (27). This characteristic 'blue-green' pigmentation is attributed to a phenazine derivative pyocyanin, which is produced by *P. aeruginosa.* The name, *Pseudomonas* is derived from two Greek words: *Pseudo* (false) and *monas* (single unit); *aeruginosa* (greenish-blue) is derived from a Latin word, aerūgō which means 'rusted copper' (28).

P. aeruginosa commonly inhabits water, soil, and vegetation. It can be grown easily in a variety of conditions and temperatures. An optimal temperature for the growth of *P. aeruginosa* is 37 °C, however, it can survive a broad range of temperatures: from 4 °C to 42 °C (29). *P. aeruginosa* is capable of causing both acute and chronic infections in humans, often nosocomial, affecting primarily immune-compromised patients, including those with CF, cancer, diabetes, burn injuries, and acquired immunodeficiency syndrome (AIDS) (28, 30, 31). It is not only a frequent cause of pneumonia in hospitalized patients (32),

immunocompromised hosts, and patients with CF (33), but also a common cause of community-acquired infection (34) and is responsible for substantial healthcare costs and resource utilization (32, 35). *P. aeruginosa* is often resistant to many classes of antibiotics (30), posing a high risk of morbidity and mortality (36). This raised concern in the WHO for an immediate need of initiatives for novel therapeutics (3).

The pathogenicity of *P. aeruginosa* lies in its ability to produce cell-associated (flagellum, pili, alginate/biofilm, lipopolysaccharides) and extracellular (proteases, phospholipases, exotoxins, rhamnolipids, pyocyanin) virulence factors (4). The production and secretion of these virulence factors are to some extent regulated by a type of cell-to-cell signaling, popularly called “quorum sensing” (37, 38). With the dynamic production and secretion of these virulence factors, understanding the pathogenicity of *P. aeruginosa* seems to be a complex issue (39). Therefore, identifying novel anti-virulence strategies is needed for future treatment of *P. aeruginosa* infections (40).

One of the reasons for the increased survival of *P. aeruginosa* is the formation of multicellular biofilms (also referred to as aggregates). Biofilm formation generally involves four phases. In the first phase of development, bacteria adhere with the help of pili and flagella to the biotic surface or onto an abiotic surface like plastic, metal, or glass (Figure 2A). Type IV pili and flagella, being the first point of attachment, are essential components for the maturation of biofilms (41, 42). Gradually, the attachment of *P. aeruginosa* gets stronger by producing an extracellular matrix. The second phase starts within 24 hours, and microcolony formation occurs over multiple rounds of cell division, resulting in increased expression of pili and secretion of the components of the extracellular matrix (Figure 2B). This results in a strong association between the cells and the adherence of cells to the surface, which provides enhanced protection of cells from the environment. The third phase starts in the next 24 - 72 hours of attachment, where the growth of cells, together with the extracellular matrix,

continues, leading to the formation of a mature biofilm structure (Figure 2C). The fourth phase usually starts after 48 hours of attachment, and includes quorum-sensing, physical disruption force on pili (42). Eventually, the cells located on the outer surface of the biofilm colony become motile and disperse out (Figure 2D). This way, biofilms aid bacterial growth by protecting them against adverse environmental conditions, including physical and chemical stressors. For this reason, biofilm-specific therapies against *P. aeruginosa* infections are required (43).

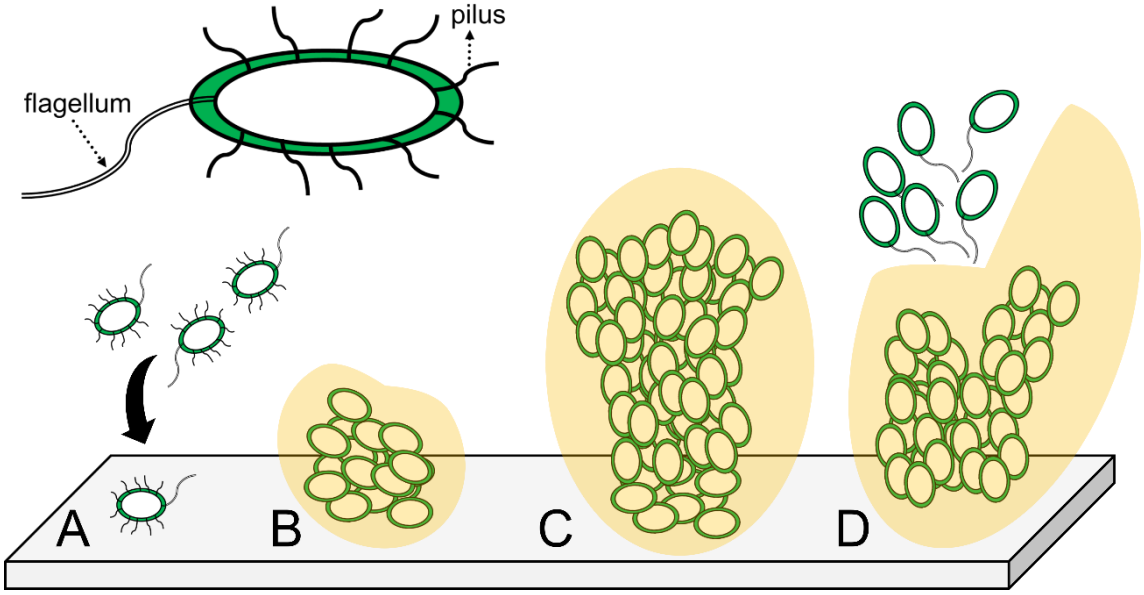

Figure 2: Steps involved in the formation of biofilm. A) Adhesion of *P. aeruginosa* on the surface with the help of the flagellum and pili, B) Increased numbers of pili and formation of extracellular matrix results in microcolony formation, C) Continued formation of extracellular matrix and growing bacterial cells leads to maturation of biofilm, D) External factors influence the dispersal of mature biofilm. Figure adapted from ref. (43).

The features of biofilm, like the extracellular matrix formation, strong adhesion to the surface, and the overall biofilm structure are important contributors to antibiotic resistance, but cannot be regarded as sole factors leading to it (44, 45). Notably, the outer membrane of *P.* aeruginosa has 12-100 times lower permeability than that of *Escherichia coli* (46, 47). Hence, its outer membrane acts as a selective barrier to the uptake of antibiotics, for example, β-lactams (47) and thus contributes to its intrinsic antibiotic resistance (30). Among

the other important factors involving the pathogenesis of *P. aeruginosa* are phospholipases, which I will discuss now.

2.2 Bacterial phospholipases in pathogenesis

Phospholipases are a class of enzymes that hydrolyze GPL substrates at specific ester bonds. Involved in GPL metabolism, phospholipases differ considerably in structure and function and can be classified according to the specific ester bond they cleave (Figure 3). The physiological relevance of phospholipases includes their role as a) digestive enzymes, b) in membrane maintenance and remodeling, and c) regulating cellular mechanisms, for example, creation of bioactive lipid molecules used in signal transduction (8). Phospholipases are classified as acyl hydrolases [PLA_1, phospholipase A_2 (PLA_2), phospholipase B (PLB), lysophospholipase $A_{1/2}$ ($LysoPLA_{1/2}$)], and as phosphodiesterases [phospholipase C (PLC) and phospholipase D (PLD)] (8).

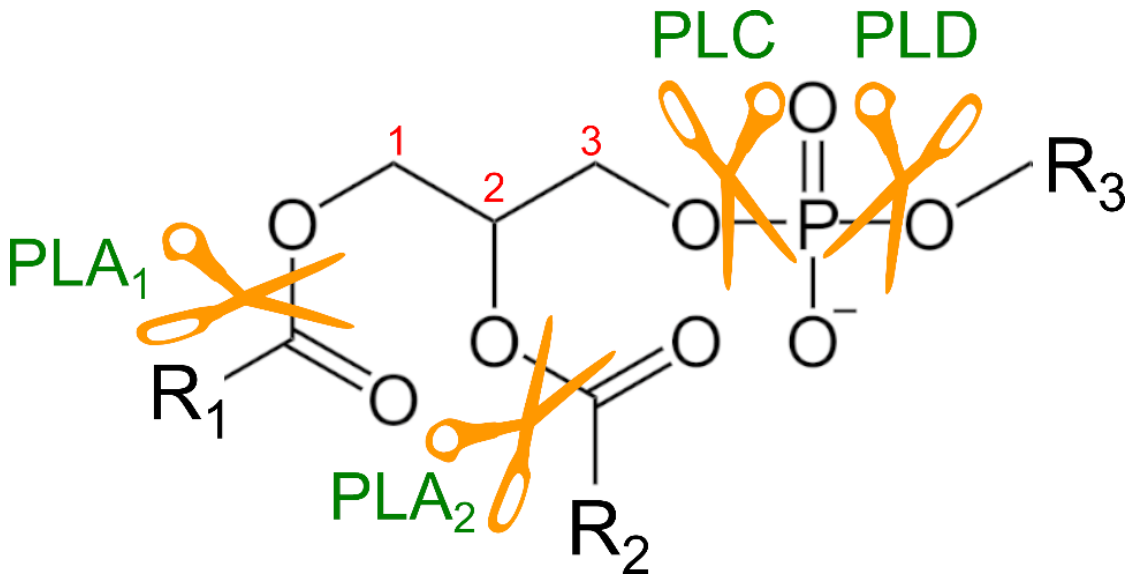

Figure 3: Specificity of the phospholipases. PLA_1, PLA_2, and PLC hydrolyze the ester bonds at the *sn*-1(1), *sn*-2(2), *sn*-3(3) sites, respectively. PLD hydrolyzes the other phosphodiester bond. PLB acts on both *sn*-1 and *sn*-2 ester bonds. LysoPLA can either cleave the ester bond at *sn*-1, *sn*-2, or both, when one or the other acyl chain is missing. R_1 and R_2 represent the acyl chain of the GPL, while R_3 represents the headgroup. Figure adapted from ref. (8).

In pathogenesis, phospholipases appear to be key virulence factors, by contributing to bacterial survival and interfering with the cellular signaling

cascades (7). Over time, scattered information on bacterial phospholipases and their putative role in disease continues to accumulate (48). Bacterial phospholipases have been reported to have a diverse role in disease, which includes triggering of bacterial entry, endosomal lysis, and cytolysis, to modulating the local immune response and stimulating cytokine secretion (7). Furthermore, bacteria are known to produce membrane-associated phospholipases (7), the hydrolysis products of which can modulate membrane fluidity, permeability, and overall membrane integrity (49, 50).

Compared to bacterial PLCs (48, 51), PLAs are less studied for their role in disease (7), but during the last decades, evidence on their involvement in bacterial invasion and pathogenesis has been significantly increased (49, 52-54). Particularly PLA_1 appears as an emerging therapeutic target because of its involvement in cellular functions associated with pathogenesis in humans (8). PLA_1 hydrolyze the GPLs at the *sn*-1 position to produce LGPL and FA (Figure 3, Figure 4).

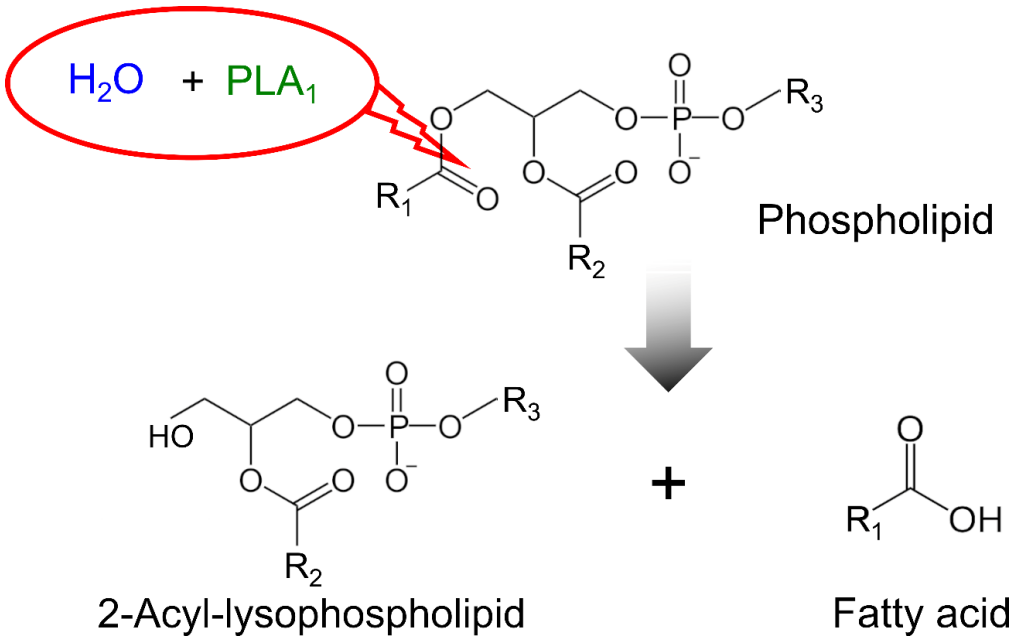

Figure 4: Hydrolysis products from PLA_1. On PLA_1 hydrolysis, a 2-acyl-lysophospholipid and a corresponding fatty acid is released from the phospholipid substrate.

The products of PLA-catalyzed hydrolysis play major roles in innumerable cellular processes. As of LGPLs, they are inverted cone-shaped molecules, share

the physical characteristics of detergents, and destabilize the membranes integrity, thus potentially contributing to the disruptive effects (13). LGPLs are also described as potent chemotactic molecules (55, 56). Moreover, increased LGPL-to-GPL ratios in ulcer patients could be attributed to the activity of an outer membrane PLA in *Helicobacter pylori*, OMPLA (12). The stimulating role of human PLA-derived LGPL has been prominent in prostate tumor development by prolonging the openings of TRPM8 ion channel (57). TRPM8 channel provides access to Ca^{2+} and, therefore, is necessary for prostate cancer cells to survive and grow (58). Overall, LGPLs have pathophysiological significance and are regarded as potential biomarkers for the diagnosis of certain diseases, including ovarian cancer (59-62), neurological (63), and cardiovascular (64, 65) disorders.

As of FAs, these were described as signal molecules of the DSF family for the first time in a plant pathogen, *Xanthomonas campestris* (66). Although the signaling cascade of the DSF family has been extensively studied in *X. campestris*, these signals are not limited to the genus *Xanthomonas* (67). Signaling molecules from the DSF family passively diffuse through cell membranes and can thus regulate diverse biological functions in a variety of Gram-negative bacteria (19). In *P. aeruginosa,* DSFs promote biofilm formation and antibiotic resistance (20, 21, 68). Consequently, DSF signaling interference provides new opportunities to control bacterial disease in both plants and animals (69).

Overall, it is well demonstrated that PLAs are attractive therapeutic targets.

2.3 PlaF is a PLA_1, a virulence factor of *Pseudomonas aeruginosa*

PlaF from *P. aeruginosa* is a PLA_1, and a virulence factor. Identified as α/β-hydrolase PA2949, PlaF was recently cloned, expressed, purified, and enzymatically characterized (22). Interestingly, PlaF has a profound role in regulating the membrane GPL homeostasis (23). While investigating the

hydrolytic function of PlaF on both artificial and natural GPL substrates, it was found that PlaF has PLA_1 activity and no PLA_2 activity (23). The substrate specificity of PlaF is similar to other phospholipases, and relies both on the head group as well as the acyl chains (7). The activity of PlaF decreases with increasing length of the substrate's acyl chains from C12 to C18 (23, 24), with a higher preference for LGPL, compared to the GPL substrates. In addition, the substrate's head groups also influences PlaF activity (23, 24).

Moreover, native PlaF was absent in the soluble fraction containing cytoplasmic and periplasmic proteins but was found co-localized with the membrane protein XcpQ (23). The XcpQ secretin is found in *P. aeruginosa* as an essential component of the type II secretion system (T2SS) (70, 71). PlaF was further found to be a cytoplasmic integral membrane protein of *P. aeruginosa*. A cellular localization model suggests that PlaF is anchored to the inner membrane with a single N-terminal TM helix, while having its catalytic C-terminal domain situated in the periplasmic space (23) (further structural details discussed in **chapter 2.4**).

A *Drosophila melanogaster* infection model was used for a comparative study of PlaF as a virulence factor of *P. aeruginosa*. Flies were infected using the needle-pricking method, as described previously (39, 72). The profiles of the host (*D. melanogaster*) response toward *P. aeruginosa* with PlaF wild-type ($PlaF_{WT}$) and *P. aeruginosa* lacking the entire *plaf* gene (Δ*plaf*) were determined. There was a remarkable difference in the survival rates of the flies. The majority (~80 %) of flies infected with $PlaF_{WT}$ died, while Δ*plaf* was almost avirulent. Moreover, these knockout studies revealed that Δ*plaf* reduces biofilm formation and impairs flagella-mediated swimming motility, suggesting PlaF as a virulence factor of *P. aeruginosa* (23). To understand and rationalize how the PlaF activity is regulated, the X-ray crystal structure was resolved.

2.4 Structure of PlaF, dimerization and its orientation in the membrane

Crosslinking and micro-scale thermophoresis experiments revealed PlaF to be only active in the monomeric form, although it exists as both a monomer and a dimer (23). Further analysis indicates that likely PlaF does not contribute to virulence by affecting the key metabolism of *P. aeruginosa*, but by hydrolysis of membrane GPLs (23). With PlaF being anchored to the cytoplasmic membrane, it is not in direct contact with the host cell, suggesting that there is some unknown mechanism for the PlaF-mediated virulence that needs to be investigated.

To understand the molecular mechanism of PlaF-mediated virulence of *P. aeruginosa*, the crystal structure of PlaF, PDB id: 6I8W (25, 26), was determined and refined at a resolution of 2.0 Å. The solved structure is an asymmetric homodimer with two chains, $PlaF_A$ and $PlaF_B$. During the crystallization process, MYR and UND are co-crystallized in $PlaF_A$ and $PlaF_B$, respectively (Figure 1). Interestingly, these FAs are natural products co-purified with PlaF (23) and are found non-covalently bound to the active site cavity of the respective chains (26). The structure reveals that the homodimer is stabilized by interactions of the N-terminal residues that form the TM (residues 5 – 27) and JM (residues 28 – 38) regions (Figure 5). The catalytic domain of the enzyme adopts a canonical α/β-hydrolase fold (73) and consists of eight-stranded β-sheets surrounded by eight α-helices, of which three α-helices form a distinct lid-like domain that covers the active site (Figure 5). Further analysis revealed that the deeply buried active site of PlaF is connected to the surface by three distinct tunnels, forming a T-shaped cleft that is compatible with the binding of bulky GPL substrates (23).

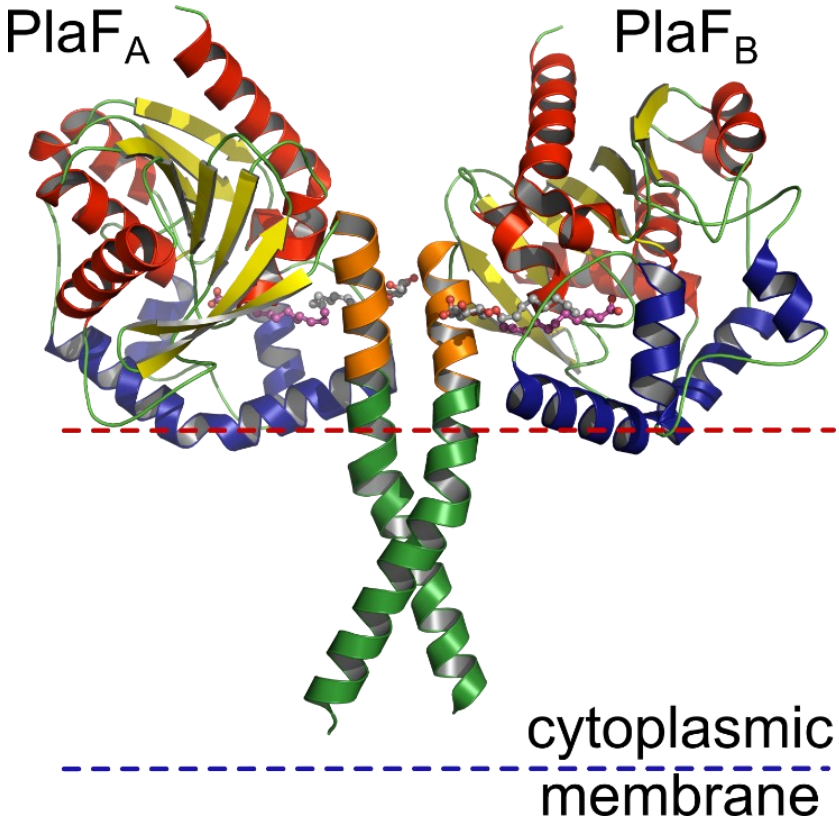

Figure 5: Crystal structure of PlaF as an asymmetric homodimer with bound endogenous FA products. The PlaF dimer crystal structure is composed of two asymmetric monomers (i.e., PlaF$_A$ and PlaF$_B$). The catalytic α/β-hydrolase domain of PlaF is situated in the periplasmic space and comprises eight α-helices (red) and eight β-sheets (yellow). The catalytic domain is anchored to the cytoplasmic membrane via a juxtamembrane helix (orange) and a transmembrane helix (green). Three α-helices form the lid domain (blue) and show putative interactions with the membrane interface. The endogenous FA (magenta spheres) and the exogenous ligands, octyl glucoside (gray spheres), are found in both chains of the PlaF dimer. Structure information retrieved from PDB id: 6I8W.

The structure information was further used to describe how PlaF activity is regulated by product-feedback inhibition. Together with biochemical analysis, it was interesting to note that the dimerization site of PlaF, V33 is connected to the catalytic residue, S137, via a MYR-mediated interaction network, suggesting that FA products can facilitate dimerization (23). Inhibition assays with FAs comprising 10 – 14 carbon atoms show strong inhibition of PlaF activity at mM concentrations, while the ones with shorter or longer FAs show moderate to weak inhibition. Further, crosslinking experiments revealed an increased number of dimeric PlaF (di-PlaF) in the presence of FAs (23). Taken together, these findings

suggest a putative role of FAs in dimerization and thus in the function of PlaF.

To investigate if the PlaF activity is regulated by the dimer-to-monomer transition, the activity was measured at varying concentrations of PlaF. Interestingly, the activity of PlaF at high concentrations is reduced by 96%, compared to that at low concentrations. These findings, together with the micro-scale thermophoresis measurements, strongly support that the PlaF activity is regulated by a reversible dimerization process at high concentrations of PlaF (23). Considering the energetics of monomer association, it was found that the PlaF in *P. aeruginosa* preferentially exist as t-PlaF, and the equilibrium shifts toward the di-PlaF on increasing the PlaF concentrations (23). Having the membrane-disruptive effect, an increase in concentration of PlaF in *P. aeruginosa* is not harmful to the cells, as it decreases the activity of PlaF (23).

Interestingly, the active sites of the di-PlaF already acquires catalytically active configurations, indicating that the structural re-arrangements of the active sites is not required for PlaF activation (23). Therefore, to investigate the structural dynamics on the time scale, unbiased MD simulations of various configurations of PlaF embedded in a membrane bilayer were performed. The results from 2 μs long simulations indicate that in all 20 replicas (i.e., 10 replicas per PlaF monomer), the tilted configurations of PlaF (t-PlaF$_{A/B}$) remained tilted, while non-tilted configurations (s-PlaF$_{A/B}$) tilted in 8/10 and 6/10 replicas for s-PlaF$_A$ and s-PlaF$_B$ respectively. Moreover, on calculating the energetics of monomer tilting, it was found that the tilted state of PlaF is energetically preferred over the non-tilted state (23). Orientating various configurations of PlaF in membrane reveals interesting information that reconciles with dimer-to-monomer transition in activity regulation. As a dimer, the active site cleft is not accessible to GPL substrates from the membrane (Figure 1A). In contrast to this, the monomer PlaF tilts, allowing a tunnel from the active site cleft to make direct contact with the membrane interface (Figure 1B). This likely favors the direct access of substrates from the membrane.

To summarize, from experiments we have identified two possible mechanisms of regulations to PlaF activity. 1) The kinetic inhibition studies with increasing concentration of FA, which reveals that FA products of PlaF regulate the function of PlaF by inducing dimerization. 2) The microscale thermophoresis experiments suggest that the activity of PlaF depends on the concentration of PlaF. At high concentrations, PlaF forms inactive dimers, but, at low concentrations, PlaF is active as the active site tunnel orients toward the membrane interface. The findings provide plausible support for the dimerization-based activity regulation in PlaF. However, the exact mechanism of substrate/product access/egress to/from the active site deserves further investigation.

2.5 Role of protein tunnels and their identification tools

Proteins are complex biomolecules that contain a variety of clefts, grooves, protrusions, and cavities in the interior (74). These cavities form a functional volume of the respective enzyme and can be used for the transport of water molecules, ions, and supply the substrates to, and evacuate the products from the active site of the enzyme (74). Such cavities are termed tunnels or channels, and their diverse roles and mechanisms in the transport of the aforementioned substances have been well-reviewed (75).

In the scientific literature, the terms tunnel and channel are frequently used interchangeable (76). However, a channel usually refers to a pathway that passes through the entire protein, has both sides open to the surrounding solvent, and is not interrupted by an internal cavity (76). In contrast, a tunnel is usually a pathway that connects the protein surface with one or more internal cavities (76). The geometry, physicochemical properties, and the dynamics of protein tunnels play an essential role in regulating the traffic to and from deeply buried active sites into the solvent in a large variety of proteins (74, 75, 77). Such tunnels thus 1) regulate the substrate specificity and selectivity, 2) prevent cellular damage by the uncontrolled release of toxic intermediates, and 3) control the enzymatic

reaction through hydration and dehydration of the active site (74).

Together with this, protein tunnels can perform several other functions, as shown by an analysis of pathways in 4,306 enzymes, which revealed that over 64 % of the studied enzymes contain on average two tunnels that connect the active site to the external environment and are longer than 15 Å (77). These tunnels vary in the amino acid residue composition (Figure 6) and hence facilitate enzyme-substrate recognition. This corroborates the essential role of such active site tunnels in the biological function (77).

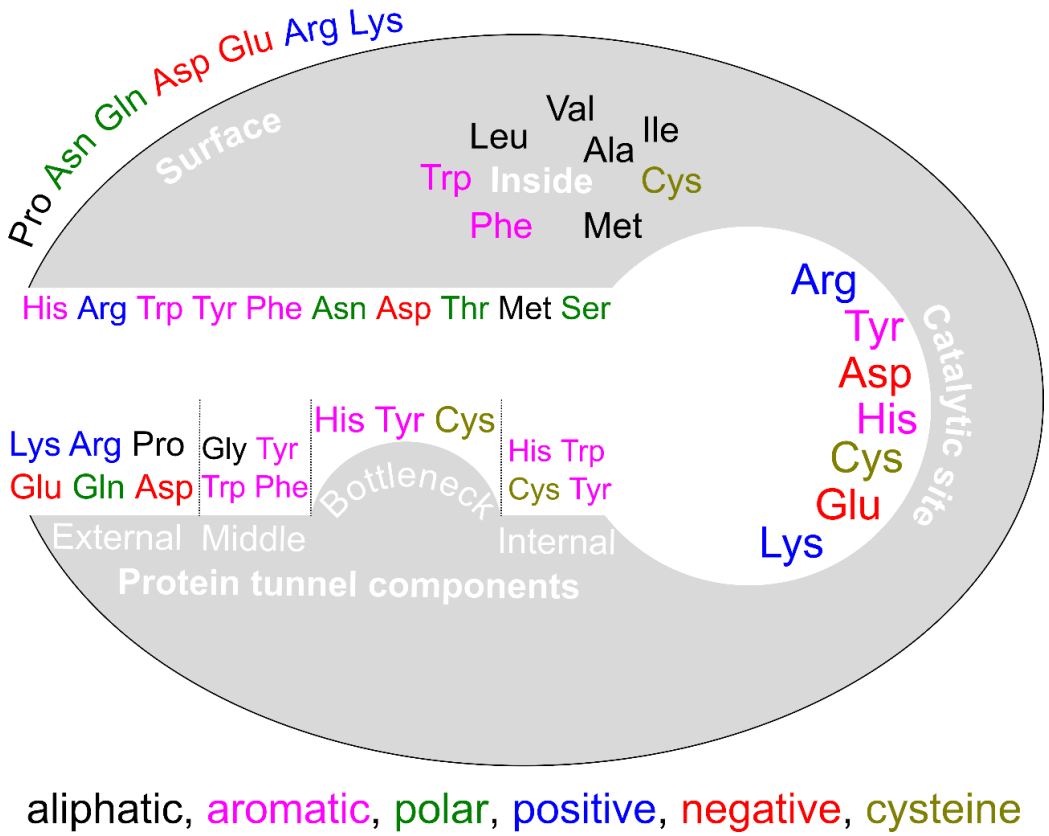

Figure 6: Variation in amino acid composition in different regions of the enzyme and tunnel. The residues found more often than average in different regions of an enzyme (gray) structure are depicted. Names of the residues are colored based on their side chain features (see the key). Figure adapted from ref. (77).

Characteristics of the protein tunnels have a significant role in enzyme activity. However, how do the tunnel characteristics influence the access of substrate or egress of products from PlaF remains uninvestigated. Since the protein dynamics and the features of the transport pathways may change significantly over time (78-81), a single static structure is not sufficient for tunnel

identification and analysis (76). Moreover, considering a static structure for the aforementioned purpose carries further pitfalls: transient tunnels might be overlooked or biologically irrelevant tunnels might be considered for closer analyses (74, 76, 82). To overcome this and obtain the characteristics of the pathways for a given protein (79-85), an ensemble of protein conformations obtained from NMR experiments, a set of crystal structures, and/or MD simulations can be used (74). Among the tools that allow the identification of pathways from MD ensembles are HOLE (v2.2) (86), MOLE (v2.0) (87), MolAxis (88), and CAVER (v3.0) (76). Among these, CAVER is particularly interesting, as it includes essential tunnel clustering together with a comprehensive analysis of tunnels from molecular ensembles (89). It is widely used for the identification and analysis of protein tunnels and is available in several implementations for users with varied experiences and expectations (74, 89).

To understand the loading mechanism of substrates into PlaF, first, I identified and analyzed the tunnels emerging from the active site of PlaF by using CAVER (v3.0), considering both static structures and the ensembles from unbiased MD simulations of PlaF from previous study (23), described in **chapter 2.4**. In the next step, I have identified different modes of substrate access in PlaF.

2.6 Steered molecular dynamics simulations

To identify different modes of substrate access across the tunnels of PlaF, I used sMD simulations (90). sMD is a type of biased simulation that allows exploring biological processes such as ligand binding and unbinding and the conformational dynamics of biomolecules on time scales that are not accessible to conventional MD simulations (90). One of the first studies to employ sMD provided an expanded understanding of receptor-ligand binding in the avidin-biotin complex in 1997 (91). Furthermore, sMD simulations have proven useful in studies on the activity of acyltransferases (92), the ligand unbinding from proteins (91, 93-95), monobody-protein interactions (96), DNA-histone interactions (97), extraction of lipids from the membrane (98), and the transport of sugar molecules

(99) or ions via membrane proteins (100, 101). Overall, sMD has been described as an important tool in rational drug design (102).

sMD allows one to study the transitions between equilibrium states in biological processes, as those rare events are associated with barrier crossings and are rarely observable at MD time scales (103). sMD closely resembles experimental techniques (104-106) that are based on the application of mechanical forces to molecules, for instance, atomic force microscopy (AFM), optical tweezers, bio-membrane force probes, surface force apparatus experiments to study the binding features, and mechanical properties of the biomolecules and their response to external stimuli (107).

In sMD simulations, time-dependent external forces are applied on one or more atoms, commonly referred as sMD atoms. These atoms can correspond to, e.g., the head group atoms of the GPL substrate being extracted from the membrane bilayer, or a molecule that has to be unbounded from the binding site of an enzyme, or can comprise a set of residues from a lid domain being analyzed for conformational dynamics of a particular protein. For sMD simulations, the direction of the applied force is chosen in advance (103). Generally, there are two ways in which one can set up the simulations, either with constant-velocity pulling or with constant-force pulling (103, 108). The principle of these two methods have been described in Figure 7.

In the constant-velocity method, sMD simulates the action of a moving AFM cantilever on a protein, and the sMD atom, or the center of mass (COM) of a group of atoms, is attached to a dummy atom with the help of a harmonic restraint (virtual spring), so that the dummy atom is moving at constant velocity (Figure 7) (103).

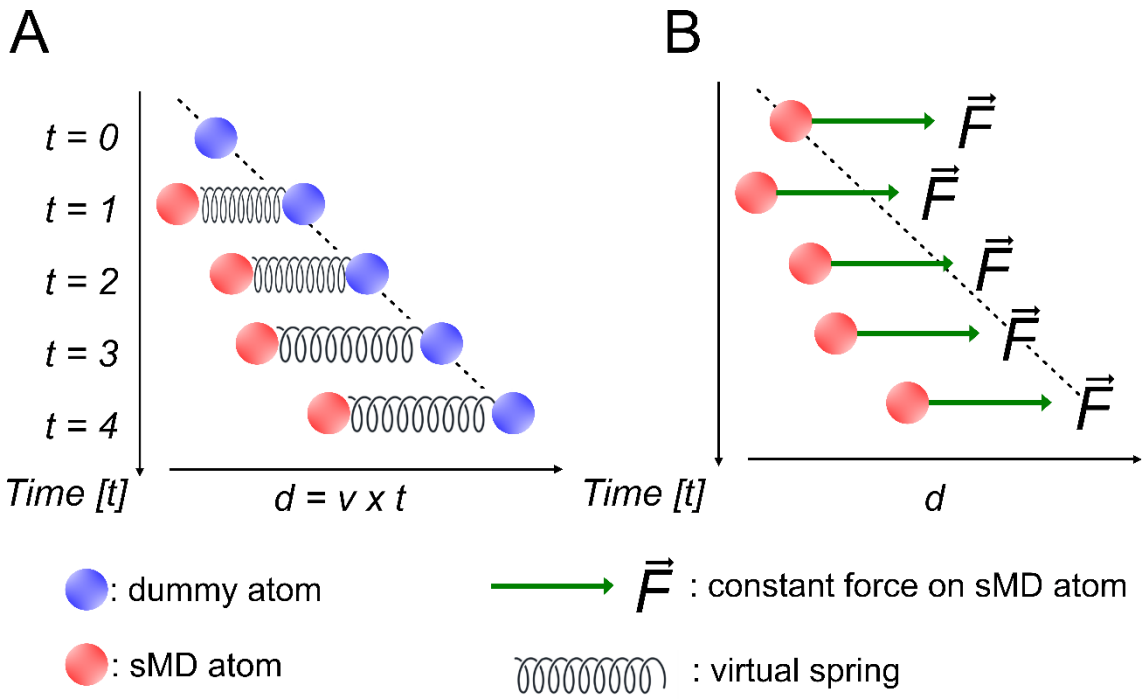

Figure 7: Principle of constant-velocity and constant-force sMD. A) In constant-velocity sMD, the sMD atom(s) (shown as a red sphere) is (are) linked to a dummy atom (blue sphere), with a virtual spring. The dummy atom has a linear motion as a function of the time *t,* with the virtual spring enforcing the sMD atom(s) to follow it. B) In constant-force sMD, there is no dummy atom. Constant force (green arrow) is applied to the sMD atom(s) in the same direction at each time step. Figure adapted from ref. (108).

During the process, the force experienced between the two atoms (i.e., the sMD atom and dummy atom) is defined using the Equation 1 and Equation 2:

$$\vec{F} = -\nabla U \qquad \text{(Equation 1)}$$

$$U = \frac{1}{2} k \left[vt - (\vec{r} - \vec{r}_o) \cdot \vec{n} \right]^2 \qquad \text{(Equation 2)}$$

Here, $\vec{F}$ represents the force, and U is the potential energy. The spring (force) constant k is to keep the group of atoms harmonically restrained. The time is represented as t, pulling velocity as v, and the pulling direction as $\vec{n}$. The $\vec{r}$ is the actual position of the sMD atom(s) and $\vec{r}_o$ is the original position of the sMD atom(s).

The choice of spring constant and pulling velocity has to be carefully considered (103). The spring constant should be high enough so that the local

unbinding potential is efficiently sampled, but must not be too high, since otherwise, the measured force will be dominated by noise (103). Likewise, the pulling velocity should be similar to experiments, but still computationally feasible (103). In other words, the chosen velocity for sMD should be the smallest velocity that allows to perform long enough simulations to see the desirable biological process, for instance, folding pathway or binding/unbinding of ligands (103). Compared to constant-velocity, in the constant-force method, there is no dummy atom or virtual spring. Instead, a fixed, constant force in the direction defined by the vector(s) is applied to the sMD atom(s) at each time step (Figure 7) (103).

Interestingly, sMD resembles the method of US (109-111), in a way that it enhances the sampling of a particular degree of freedom in a biomolecular system (103). Although sMD results in non-equilibrium simulations, the equilibrium properties, for example the free energy of the two states of a system, can be determined using Jarzynski's relation (112, 113) described in Equation 3.

$$\overline{e^{-W/k_BT}} = e^{-\Delta F/k_BT} \qquad \text{(Equation 3)}$$

Here, ΔF is the free energy difference between two states, which is connected to work W done on the system; k_B is the Boltzmann constant and T is the temperature of the system. As described in Jarzynski's relation, the work done is related to the free energy difference between starting and end states of the sMD simulation. The application of Jarzynski's relation is comparable to the US approach in terms of efficiency (113). The sMD method has several advantages over the US method and one of them is that the structural coordinates are generated in sMD that can be directly used for sampling simulations. In contrast, this is directly not feasible using US method, and the coordinates of the starting structures needs to be generated beforehand. Yet, US is a general method, and can be applied to a variety of systems (103).

Using Jarzynski's equality, sMD can be directly used to determine PMF (113). However, this method suffers from insufficient sampling convergence of the

membrane bilayer (114) and hence PMFs obtained for systems involving membrane are non-reliable. Interestingly, applying US along pathways identified by sMD simulations is an effective way of calculating PMF (115). Therefore, in this study, using sMD simulations, I have first determined low free energy pathways for substrate access to PlaF. Then, I have considered these pathways to define the reference points for subsequent US simulations.

2.7 Free energy computation from umbrella sampling simulations

US simulation (109) is a widely used method to compute the PMF and has demonstrated a great advantage over other free energy methods, like free energy perturbation, or thermodynamic integration simulations (110). US can be applied to a variety of systems (103), with some recent successful applications from our research group including the determination of energetics that leads to the opening-closing motion of nucleotide-binding domain of pyruvate phosphate dikinase (116), TM helix association in integrins (117), and the dimerization of G-protein coupled bile acid receptor TGR5 (118). In my studies using US simulations, I have calculated the energetics of substrate extraction from the membrane and their loading into the deeply buried hydrolysis site within PlaF.

A PMF describes the free energy F as a function of a reaction coordinate ξ according to Equation 4, below (119).

$$F(\xi) = -k_B T \ln[P(\xi)] + C \qquad \text{(Equation 4)}$$

Where k_B is the Boltzmann constant, T is the temperature, $P(\xi)$ is the probability of the system along the reaction coordinate, and C is a constant used for normalization. The reaction coordinate can be one-dimensional or multidimensional (110, 111) and represents conformational variables like distances or angles (120).

Background

In the first step, one needs to generate a series of configurations along the reaction coordinate (Figure 8A), of which some serve as the starting configurations for US windows (Figure 8B). These configurations characterize the intermediate steps involved in the process to be investigated. For each umbrella window, independent MD simulations are performed.

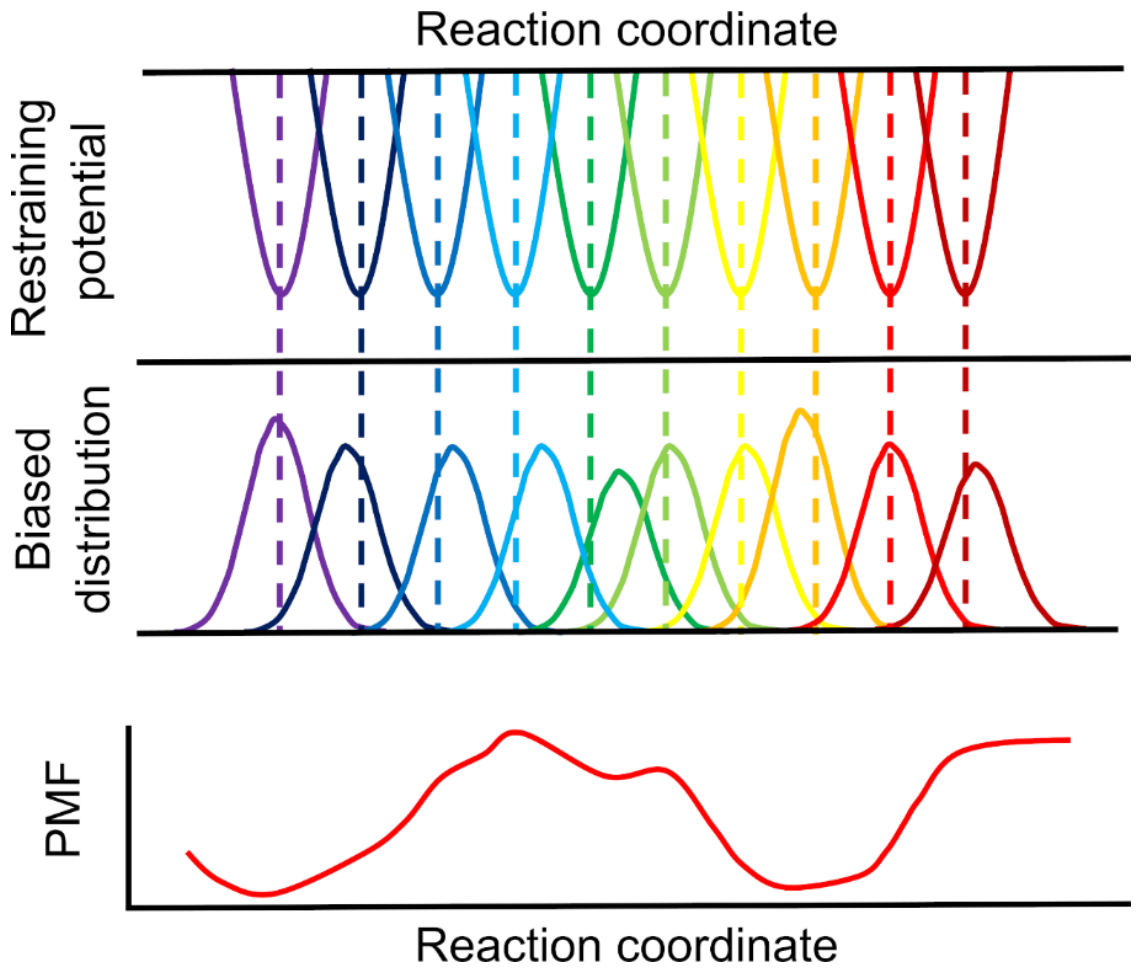

Figure 8: Umbrella sampling method for potential of mean force computations. In the first step, starting (reference) configurations (blue spheres) are generated along the reaction coordinate. During umbrella sampling simulations, selected configurations are restrained by harmonic potentials, allowing adjacent umbrella windows along the reaction coordinate to overlap. In the last step, the free energy profile (PMF) can be obtained by unbiasing and recombining the overlapping distributions using WHAM.

During the simulations, umbrella potentials allow the sampling of statistically improbable states, which are not accessible through conventional unbiased MD simulations. By restraining the umbrella windows to the reference points, the biasing potentials along the reaction coordinate thus drive the system from one

thermodynamic state to another (110); for instance, binding or unbinding of ligands from proteins, which can often reach time scales inaccessible to conventional MD simulations (103). For each window, the bias potential $\omega_i(\xi)$ keeps the system close to the reference points ξ_i^{ref} of the respective umbrella window i. These potentials can be in any functional form; however, for simplicity, harmonic bias potentials (Equation 5) of the strength (i.e., force constant) K are used (110, 120).

$$\omega_i(\xi) = K/2\ (\xi - \xi_i^{ref})^2 \qquad \text{(Equation 5)}$$

The obtained distributions from the sampling simulations are then used to calculate the change in the free energy profile for each umbrella window (110, 111, 120). However, the umbrella windows are needed to be combined and the biasing potential that was introduced in the previous step (Equation 5) has to be removed (109) (Figure 8C). For this, the weighted histogram analysis method (WHAM) (121, 122) is one of the most reliable approaches (111). To obtain unbiased distributions, WHAM uses Equation 6.

$$P(\xi)_i^{unbiased} = P(\xi)_i^{biased}\ exp\left(-\frac{\omega_i(\xi)}{k_B T}\right)\langle exp\left(\frac{\omega_i(\xi)}{k_B T}\right)\rangle \qquad \text{(Equation 6)}$$

where ⟨ ⟩ represents the ensemble average.

To obtain the global distribution, the potential, $P(\xi)_i^{unbiased}$ of the individual umbrella windows is then recombined with WHAM by calculating a weighted average of the distributions of the individual windows using Equation 7 (110).

$$P(\xi)^{unbiased} = \sum_i^{windows} P\,(\xi)_i^{unbiased}\ p_i(\xi) \qquad \text{(Equation 7)}$$

Where the weight $p_i(\xi)$ is fulfilling the condition $\sum p_i\,(\xi) = 1$ and calculated from Equation 8.

$$p_i(\xi) = N_i \, exp\,(-\frac{\omega_i(\xi) + F_i}{k_B T}) \quad \text{(Equation 8)}$$

Where N_i is the total number of steps sampled and, F_i is the free energy constant computed from Equation 9.

$$exp\,(-\frac{F_i}{k_B T}) = \int P(\xi)^{unbiased} \, exp\,(-\frac{\omega_i(\xi)}{k_B T})\, d\xi \quad \text{(Equation 9)}$$

As the global (unbiased) distribution $P(\xi)^{unbiased}$ enters Equation 9 and the free energy constant F_i enters Equation 7 via Equation 8, Equations 7-9 have to be iterated self-consistently until convergence is achieved (110).

2.8 Experimental relation to computations

On the one hand, experimental findings are the source of motivation to perform computer-based simulations, because they can extract such details that are not accessible to experiments. On the other hand, computer simulations are based on theoretical models, vary from experimental conditions and time scales, and therefore, it is essential to establish the connection with experiments. Such relations validate the computations and improve the understanding of the investigated system (123).

2.8.1 Relating PMF to experiments

Although the PMF cannot be measured experimentally, the free energy difference between the two states of the lipid in the solvent and in the membrane can be related to the experimentally measured critical micelle concentration (CMC) (124). The CMC can be related to an excess chemical potential using Equation 10.

$$\mu - \mu_0 = RT \, \ln\left(\frac{CMC}{55.5}\right) \quad \text{(Equation 10)}$$

$\mu - \mu_0$ is the excess chemical potential, *R* is the gas constant, *T* is the temperature, and CMC has been converted to mole fraction units. The CMC

values are obtained from the experimental studies.

2.8.2 Absolute binding free energy from computed PMF

To relate the computed PMFs with the experiments, I have also determined the absolute binding free energy of substrates to PlaF using an approach modified from Chen and Kuyucak (125). Here, the computed PMF from US simulations was integrated along the reaction coordinate to calculate an association (equilibrium) constant K_{eq} using Equation 11.

$$K_{eq} = \pi r^2 \int_{active\_site}^{membrane} e^{-E(\xi) / k_B T} \, d\xi \qquad \text{(Equation 11)}$$

Where the r in the factor πr^2 corresponds to the maximum bottleneck radius of the respective tunnel, determined by CAVER analysis; πr^2 is the cross-sectional area of the tunnel, $E(\xi)$ is the PMF at a specific value of the reaction coordinate, k_B is the Boltzmann constant, and *T* is the temperature at which the simulations were performed. K_{eq} was then transformed to the mole fraction scale K_x, taking the number of lipids N_L per membrane volume V into account via Equation 12.

$$K_x = K_{eq} \frac{N_L}{V} \qquad \text{(Equation 12)}$$

K_x was then used to determine the difference in the free energy between the bound and unbound state ($\Delta G^{\circ}_{\mathrm{b}}$) of a single substrate molecule using Equation 13.

$$\Delta G^{\circ}_{comp} = -RT \, ln(K_x) \qquad \text{(Equation 13)}$$

2.8.3 Blocking access of PlaF substrates

Transport of ligands from the outer environment to the deeply buried active site is an important event in the activity mechanism of enzymes (126). Naturally, such transportations are facilitated by the tunnels. Therefore, the characteristics of the tunnels, when modified, have a substantial impact on the function and activity of the enzyme (74, 126-132). Particularly, when bottleneck residues are

substituted with bulkier tryptophan residues, this reduces the radius of the tunnels, influencing the normal passage of ligands (128). On the positive side, such modifications can aid in understanding the catalytic mechanism of an enzyme.

Therefore, to validate my PMF computation and thus the identification of substrate access route in PlaF, I first identified substitution sites using the analysis from CAVER. The tunnel-lining residues corresponding to these substitution sites were then used for small-to-tryptophan substitutions using FoldX (133), and the stability of the PlaF variants was evaluated in terms of the change in free energy ($\Delta\Delta G$) with respect to the wild-type (134). Substitutions with an average $\Delta\Delta G > 3$ kcal mol^{-1} are considered destabilizing (135) and were not further pursued for biological evaluation. To further check if the proposed substitutions will block the substrate access, tunnels were re-calculated in the PlaF variants using CAVER.

3. Scope of the thesis

P. aeruginosa has been described as a medically relevant bacterium and its pathogenesis relies on the hydrolysis products from virulence factors like phospholipases (see **chapters 2.1, 2.2**). Recently, PlaF, a novel phospholipase bound to the cytoplasmic membrane was identified as a significant virulence factor from *P. aeruginosa*. Interestingly, the activity of PlaF is higher for single acyl-chained LGPL substrates than two-acyl chained GPL substrates. Moreover, the length of acyl chains of the substrates also influences the activity of PlaF. Experiments show that a high concentration of PlaF molecules results in dimerization, which is also confirmed by the crystal structure of PlaF as a homodimer. However, PlaF is only active in the monomeric form. Coupling the experiments with MD simulations revealed that the activity of the monomeric PlaF depends on its orientation to the membrane, which differs in the di-PlaF (see **chapters 2.3, 2.4**). Compared to di-PlaF, monomeric PlaF re-orients in the membrane such that the tunnel connecting the active site comes in direct contact to the membrane interface allowing access of substrates (Figure 9).

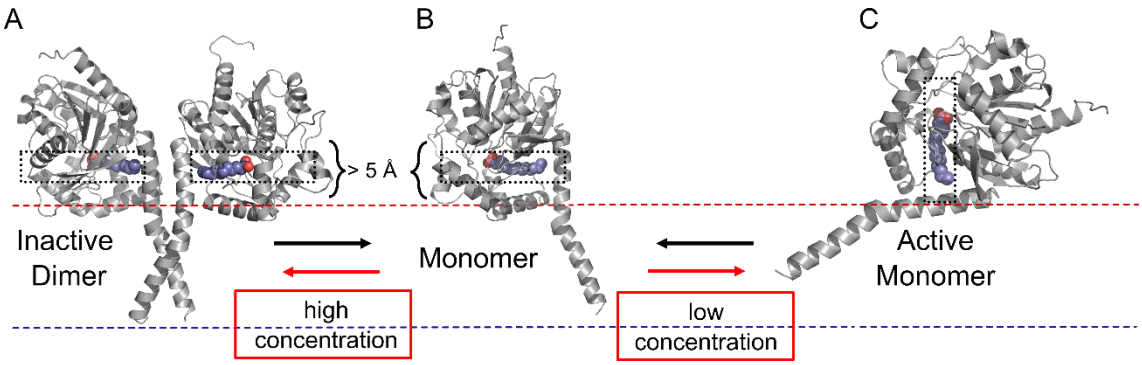

Figure 9: Dimer-monomer transition of PlaF. A) At high concentrations, PlaF molecules form a dimer. When a dimer is oriented to the membrane, the active site tunnel (black box) is located > 5 Å above the membrane interface. B) Chain A from the PlaF dimer, oriented as in the dimer. The entrance of the tunnel is located > 5 Å above the membrane interface. C) At low concentrations, the PlaF monomer re-orients and tilts toward the membrane; this configuration allows the direct contact of the active site tunnel to the membrane interface. Blue spheres represent the co-crystallized fatty acid molecules in respective chains of PlaF. Figure adapted from ref. (23).

Scope of the thesis

However, from the above studies it remains unclear:

1. How many tunnels connect the deeply buried active site of PlaF to the external environment?
2. How do substrates from the membrane reach the active site of PlaF?
3. How do the characteristics of the tunnels determine the activity and specificity of PlaF for medium-chain substrates?
4. How do products leave the active site after hydrolysis?
5. How is the PlaF activity regulated via monomer-dimer equilibrium?

To enhance our understanding of the mechanisms of PlaF activity, and find the answers to the above-mentioned questions, I have first identified the tunnels that connect the active site of PlaF to its surface. Then, I have identified preferred access pathways of substrates in PlaF using sMD simulations, US simulations, and PMF computations. Computed results were also related to experiments. Particularly, site-directed mutagenesis experiments performed by the Jaeger lab[†] support the computational findings of this study. To identify the preferred egress pathway of the PlaF products, I have performed unbiased MD simulations.

[†]The experimental work was performed by Christoph Heinrich Strunk under the supervision of Karl-Erich Jaeger and Filip Kovačić in the Institute of Molecular Enzyme Technology, Heinrich Heine University Düsseldorf, and the Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

4. Materials and methods

This chapter was taken from the following peer-reviewed publication:

"Substrate Access Mechanism in a Novel Membrane-Bound Phospholipase A of *Pseudomonas aeruginosa* Concordant with Specificity and Regioselectivity" Sabahuddin Ahmad, Christoph Heinrich Strunk, Stephan N. Schott-Verdugo, Karl-Erich Jaeger, Filip Kovacic, and Holger Gohlke
Journal of Chemical Information and Modeling, 2021, 61 (11), 5626-5643
DOI: 10.1021/acs.jcim.1c00973

Reprinted with permission from the Journal of Chemical Information and Modeling. Copyright © 2021, American Chemical Society.

4.1 Identification of the access tunnels

Tunnels emerging from the active site of PlaF were identified using CAVER 3.0 (76). The COM of the catalytic residues S137 and H286 was defined as the starting point of the search, from which the possible connections of the tunnels to the bulk solvent were identified. The catalytic residue D258 was not included in this search criteria since its side chain is distant from the catalytic site. Probe and shell radii of 2 Å and 6 Å were used, respectively. The probe radius of 2 Å is slightly larger than the van der Waals radius of a phosphorous atom (i.e., 1.8 Å), present in every PlaF substrate to be investigated.

4.2 Starting structure preparation

The crystal structure of PlaF is available from the PDB (25) (PDB id: 6I8W) (26). The last five residues of the C-terminus were missing in the structure and, hence, were added using MODELLER (136). The starting configuration of PlaF for MD simulations was prepared by embedding chain A of the PlaF dimer in a tilted configuration (t-PlaF$_A$) into a lipid bilayer membrane consisting of 75% 1,2-dilauroyl-*sn*-glycero-3-phosphorylethanolamine (DLPE) and 25% 1,2-dilauroyl-*sn*-glycero-3-phosphoglycerol (DLPG). The tilted configuration of PlaF embedded

in the membrane was predicted by the Positioning of Proteins in Membrane (PPM) method (137). The head group composition of the membrane closely resembles that of the inner membrane of Gram-Negative bacteria (11, 138, 139). The prepared structure was used to investigate the loading mechanism of DLPG or DLPE into t-PlaF$_A$. Furthermore, loading of 1,2-distearoyl-*sn*-glycero-3-phosphoglycerol (DSPG) and an LGPL, 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphoglycerol (2LMG), were also investigated. For that, t-PlaF$_A$ was embedded into a membrane consisting of ~10% of DSPG and 2LMG in the upper leaflet. The GPL composition in the lower leaflet of these systems is the same as that used for investigating DLPG and DLPE. The systems were prepared and solvated using CHARMM/GUI (140) or PACKMOL-Memgen (141). A distance of at least 15 Å between the protein or membrane and the solvent box boundaries was used. To obtain a neutral system, counter ions were added that replaced solvent molecules. The size of the resulting systems was ~140,000 atoms.

Systems excluding t-PlaF$_A$, but including one of the GPL substrates (i.e., DLPG) and one of the LGPL substrates (i.e., 2LMG), were also prepared to compare and decipher the energetics of lipid extraction from the membrane into solvent. Considering the orientation and position of t-PlaF$_A$ in the membrane, one can safely assume that only substrates located in one leaflet will contact the catalytic domain of t-PlaF$_A$ and, hence, have direct access. Therefore, the composition of one leaflet was slightly modified to reflect the inclusion of the selected substrate. For this, a ratio in the upper leaflet of 6:2:1 for DLPE, DLPG, and the respective substrate was used. Using PACKMOL-Memgen, the bilayer system was prepared, solvated, and necessary counter ions were added. The minimum water distance from the membrane surface to the solvent box boundaries was increased to 35 Å to leave enough space between the substrate and the membrane surface and avoid interactions with periodic images during the extraction. Box dimensions in the x and y axes were set to 70 Å, resulting in systems comprised of ~50,000 atoms.

4.3 Simulated extraction of substrates from the membrane

MD simulations were performed using the GPU implementation of the AMBER 16 molecular simulation package (142, 143), employing the ff14SB force field for the protein (144), the Lipid17 force field for the lipids (145-147), and the TIP3P water model (148). The SHAKE algorithm (149) was used to constrain bond lengths of hydrogen atoms to heavy atoms, enabling a time step of 2 fs. Long-range electrostatic interactions were considered using the Particle Mesh Ewald (PME) algorithm (150). The system was energy-minimized by three mixed steepest descent/conjugate gradient calculations with a maximum of 20,000 steps each. First, the initial positions of the protein and membrane were restrained, followed by a calculation with restraints on the protein atoms only, and finalizing with a minimization without restraints. The minimized system was then gradually thermalized in two stages. Initially, the temperature was increased from 0 K to 100 K under NVT conditions, then from 100 K to 300 K under NPT conditions at 1 bar, using a Langevin thermostat (151). The equilibration process continued for 5 ns, before starting with production simulations. As usual in membrane MD simulations, the NPT ensemble was used, allowing the membrane to accommodate along the trajectory (152). For US simulations, the pressure was maintained using an anisotropic Berendsen barostat (153), while for the rest of the simulations a semi-isotropic Berendsen barostat (153) was used, coupling the membrane (x-y) plane with the constant-surface-tension dynamics. All analyses were performed by using CPPTRAJ (154). Unless otherwise stated, molecular visualization was performed with PyMOL (155) and VMD (156). The Movie maker module within Visual Molecular Dynamics (VMD) was used to illustrate the acyl chain termini of lipids reaching the membrane interface and the access of substrates into PlaF.

To extract a substrate molecule from the membrane into one of the access tunnels, I selected the lipid that was closest to the entrance and pulled it from the membrane through the tunnel to the active site of PlaF, using constant-velocity

sMD simulations (see **chapter 2.6**). Pulling simulations at low velocities are recommended for small polar molecules (157) and large lipids (158) to calculate free energy profiles. At the lowest pulling rates, lipids have time to adapt to energetically favorable conformations during the extraction process (158). In a recent study investigating GPL binding to phospholipase A2 (PLA_2), a constant pulling velocity of 5 Å ns^{-1} was used (159). For the extraction of substrates, I considered all three possibilities by which a substrate may enter a tunnel: either the head group or one of the two tails. Depending on the type of head group, phosphatidylglycerol (PG) or phosphatidylethanolamine (PE), each substrate was pulled by its oxygen or nitrogen atoms at a constant velocity of 1 Å ns^{-1} using a force constant of 5 kcal mol^{-1} $Å^{-2}$. When pulling at the tail, the terminal carbon atom of the respective acyl chain was used.

Each tunnel was divided into several segments connected through virtual points formed by the COM of amino acids lining the respective tunnel. The number of virtual points depends on the length and shape of the respective tunnel. The virtual points guided the extraction of substrates such that the substrates followed the path of the respective tunnel. In addition, to obtain a low-energy pathway, an adaptive biased sMD (AsMD) protocol was implemented. For this, 50 replicas for each pulling simulation were carried out, and the work required was computed as a function of the reaction coordinate. The computed work was further related to free energy difference between two states of the pulling simulation applying Jarzynski's relation (Equation 3; see section 2.6) (112). The replica closest to the Jarzynski's average (112) was considered to describe the lowest free-energy pathway and provided the starting point for the next pulling stage. Trajectories further away from that pathway were removed. This procedure results in faster convergence of PMF profiles, decreasing the overall computation needed (160).

For the systems without t-$PlaF_A$, the substrates were extracted with the same pulling velocity and spring constant, as mentioned above. However, to

avoid edge effects, a substrate in the middle of the membrane was located. For this extraction process, the reaction coordinate was the distance between the head atom of the pulled substrate and COM of phosphorous atoms of the lipids in the opposite leaflet. Furthermore, to determine the free energy minimum of the phospholipids in the membrane more accurately, the substrate was first pulled into the membrane (~3 Å), before pulling it out of the membrane.

4.4 Umbrella sampling and potential of mean force calculations

To understand the substrate access mechanism in PlaF and to identify preferential substrate access tunnels, PMFs were computed based on US (109) (see **chapter 2.7**), taking structures from the sMD simulations as starting points. As a reaction coordinate, the COM distance of the three oxygen atoms of the glycerol moiety in the substrate to the COM of residues S137 and H286 (only Cα atoms) of the active site was used. This reaction coordinate was also taken for all other systems for it describes the essential aspects of the structural transformation during substrate access. Consecutive positions of the substrates from the membrane to the active site as determined by pulling simulations were considered reference points for US, with each position corresponding to one umbrella window. To achieve sufficient overlap between the umbrella windows, distances between reference points of ~1 Å were used along the reaction coordinate. The length of individual tunnels and the size of acyl chains for respective substrates vary. Therefore, for sampling the access of different substrates, different numbers of windows were required for each tunnel. Selected positions of the lipid in the tunnel were restrained by harmonic potentials, using a force constant of 5 kcal mol^{-1} $Å^{-2}$. To achieve sufficient convergence of the PMF profile, each window was sampled for 300 ns, of which the last 100 ns were used to calculate the PMF. Distance values were recorded every 2 ps and processed with WHAM (121, 122). To estimate the PMF error, the data was separated into blocks according to the maximum calculated autocorrelation time of 20 ns. The correlation time was obtained for the complete trajectory, excluding the first

20 ns of sampling data for equilibration. The last 100 ns of sampling data was split into five blocks of 20 ns each, a PMF profile was calculated for each block with WHAM, and the error at each PMF point was calculated as the standard error of the mean.

Similarly, for systems without t-PlaF$_A$, trajectories obtained by pulling simulations were used to set up US simulations. Umbrella windows were extracted at distances of 1 Å from the starting point of the pulling simulation until the substrate was not interacting with the membrane anymore. The selected positions of the lipid were restrained by harmonic potentials, using a force constant of 5 kcal mol^{-1} Å^{-2} and as the reaction coordinate the distance of the COM of the three oxygen atoms of the glycerol moiety of the substrate to the COM of phosphorous atoms of the lower membrane leaflet. Each window was simulated for 100 ns at constant pressure (1 bar) and temperature (300 K) conditions until convergence was achieved. The first 20 ns of simulation data was discarded. WHAM (121, 122) was used to calculate the PMF. The PMFs were evaluated for convergence by checking the change in the free energy profile with the increase in sampling time at every 10 ns. Furthermore, histograms of sampled configurations were visually inspected for sufficient overlap between the neighboring umbrella windows; otherwise, the iterative cycle in WHAM fails to converge and the free energy profiles have discontinuities.

4.5 Absolute binding free energy from computed PMF

The absolute binding free energy of substrates to PlaF was determined from the computed PMF using an approach modified from Chen and Kuyucak (125). The PMF was integrated along the reaction coordinate using equations 11,12,13 (see **chapter 2.8.2**). For the PMF integration, the maximum bottleneck radius of the respective tunnel was determined by a CAVER analysis.

4.6 Blocking access of the PlaF substrates

To corroborate predicted access tunnels for PlaF substrates, I intended to

block these by small-to-tryptophan substitutions of tunnel-lining residues. To do so, I have identified possible substitution sites from the previous CAVER analyses, taking into account the tunnels' bottleneck radii and lengths. For these analyses, the same trajectory used to search for tunnels in t-PlaF$_A$ was considered. Finally, 4-5 amino acids within each tunnel were selected for substitutions.

In the first step, all the amino acids except glycines and prolines within 3 Å of individual tunnels and oriented toward a tunnel were considered. In turn, residues with an outward orientation were disregarded as a substitution there will likely not block the tunnel. Furthermore, as the TM and JM helix was found to be important for both the dimerization and the activity of PlaF (23), residues of these helices were excluded. Finally, the catalytic residues S137, D258, and H286 and other residues of the active site were disregarded to avoid affecting the activity of PlaF.

The selected residues of each tunnel were substituted to tryptophan using FoldX (133), and the stability of the PlaF variants was evaluated in terms of the change in free energy ($\Delta\Delta G$) with respect to the wild-type (134). Single amino acid substitutions were performed 10 times for each proposed residue of each tunnel, and the results were averaged. If the average $\Delta\Delta G$ > 3 kcal mol^{-1}, the substitution is considered destabilizing (135) and was not further pursued. To check if the proposed substitutions will block the tunnel, the bottleneck radius of the variant tunnels was re-calculated using CAVER. As done earlier, the probe radius was set to 2 Å. If no tunnel was identified with this criterium, the probe radius for tunnel search was reduced until the tunnels started to appear again.

4.7 Egress of PlaF products

To determine the egress pathways of PlaF products, a system with a 2LMG substrate was considered. The final snapshot at 300 ns of the US simulations of the window with the substrate close to the active site was considered as the

starting structure for unbiased MD simulations. 2LMG was cleaved into the products: MYR and PGR (phosphatidylglycerol from lysophospholipid, 2LMG), without altering the orientation of each product within the tunnels. Atomic partial charges for the products were derived according to the restraint electrostatic potential fit (RESP) procedure (161), as implemented in Antechamber (162). Geometry optimizations and subsequent single-point calculations were performed with Gaussian (163) at the Hartree-Fock level with the 6-31G* basis set. Force field parameters for the products were taken from the general amber force field for organic molecules (GAFF, version 2) (164). The prepared system was then minimized, thermalized, and equilibrated using the protocol described for MD simulations (see **chapter 4.3**). 12 replicas of production MD simulations of 3 µs length each under NPT conditions were performed. The distance of the 2LMG products to the entrance of each tunnel was computed for each replica.

5. Results

This chapter was taken from the following peer-reviewed publication:

"Substrate Access Mechanism in a Novel Membrane-Bound Phospholipase A of *Pseudomonas aeruginosa* Concordant with Specificity and Regioselectivity" Sabahuddin Ahmad, Christoph Heinrich Strunk, Stephan N. Schott-Verdugo, Karl-Erich Jaeger, Filip Kovacic, and Holger Gohlke
Journal of Chemical Information and Modeling, 2021, 61 (11), 5626-5643
DOI: 10.1021/acs.jcim.1c00973

Reprinted with permission from the Journal of Chemical Information and Modeling. Copyright © 2021, American Chemical Society.

5.1 Access pathways to the catalytic site in PlaF

The crystal structure of PlaF revealed three pronounced tunnels, forming a large, T-shaped active site cleft. This cleft is compatible with binding bulky GPL substrates (23). However, the structural dynamics of biomolecules may lead to variations in the tunnel shape (165). Therefore, I have reanalyzed trajectories from 10 replicas of unbiased MD simulations of 2 µs length for each of the systems di-PlaF, $PlaF_A$, $PlaF_B$, and t-$PlaF_A$ from a previous work (23) using CAVER (76). CAVER is a program for analyzing and visualizing tunnels and channels in protein structures (89).

I have primarily focused on t-$PlaF_A$ because the tilted structure is likely the catalytically active form (23). Three tunnels were identified that connect the active site of t-$PlaF_A$ to its surface like in the crystal structure (Figure 10) (23): Tunnel 1 (T1) and tunnel 2 (T2) point toward the membrane, and tunnel 3 (T3) opens to the periplasmic space > 15 Å above the membrane (Figure 10). T1 and T2 converge close to the active site and connect to T3. In the crystal structure, T1 contains MYR (chain A) and UND (chain B), which are hydrolysis products of GPL substrates with C14 and C11 acyl chain(s), respectively.

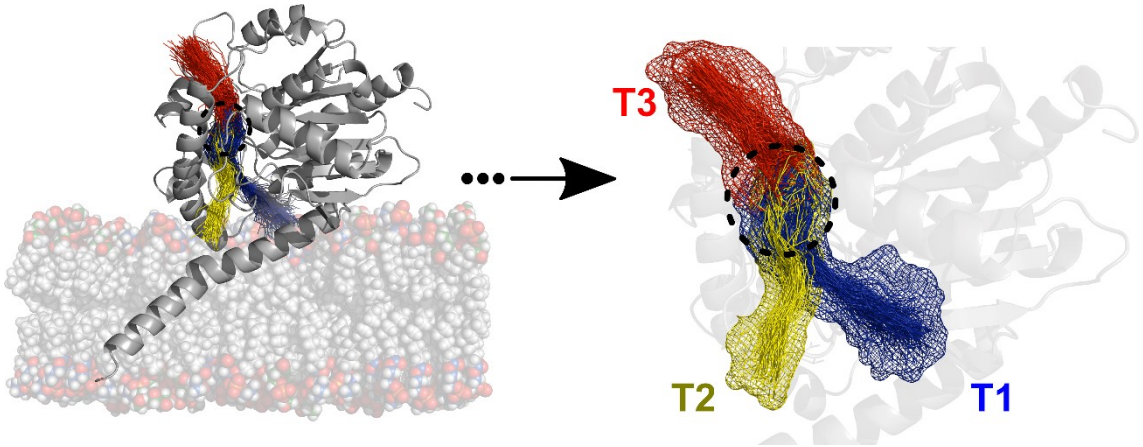

Figure 10: Clusters of tunnels identified in t-PlaF$_A$ ensembles. Three major tunnel clusters connect the catalytic site (black dashed circle) of PlaF to the protein surface. Tunnels T1 and T2 point toward the membrane; tunnel T3 is located > 15 Å above the membrane, with its opening pointing into the periplasmic space.

Table 1: Characteristics of tunnel clusters identified from unbiased MD simulations of t-PlaF$_A$ using CAVER.

Tunnel cluster	Occurrence [a,b]	Maximum bottleneck radius[c]	Average bottleneck radius[c]	Average length[c]
T1	30.45	3.18	2.28	27.08
T2	21.80	2.95	2.21	23.75
T3	27.75	3.13	2.29	15.16

[a] Snapshots in which the tunnel is identified with respect to the total number of snapshots, in %.

[b] Data calculated with a probe radius of 2.0 Å.

[c] In Å.

T1 is the longest tunnel (Table 1) and was open more often than the other two tunnels (Table 1). The tunnel radii fluctuate between 2 Å and 5 Å depending on the location in the tunnel and the simulation length (Figure 11). The average bottleneck (narrowest part of the tunnel) radius of all tunnels is 2.26 ± 0.02 Å (mean ± standard error of the mean), which is close to the radius of glycerol (2.74 Å) (166), an essential component of all GPLs, but smaller than the radius of DLPG (~4.4 Å) deduced from the lipid's area-per-lipid (APL) (167).

For comparison, tunnels in monomeric PlaF$_A$, PlaF$_B$ and the two chains of di-PlaF show open occurrences of ~20% to ~5% (Table 2), indicating no marked differences between monomeric and di-PlaF.

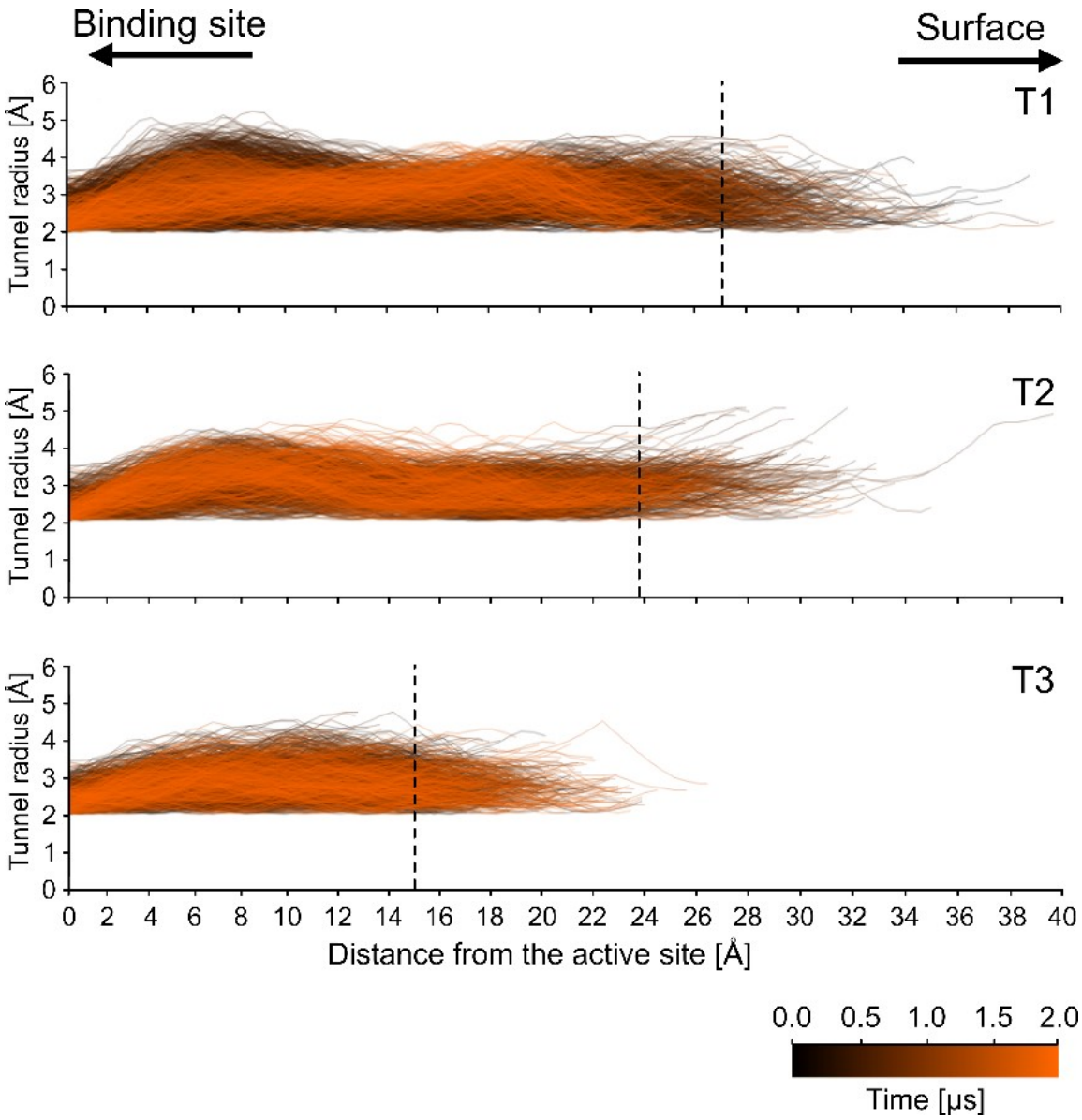

Figure 11: Properties of tunnels identified in t-PlaF$_A$ ensembles. Profiles of selected tunnel clusters T1-T3 were evaluated as to radius and distance from the active site during MD simulations of 2 µs length (see color scale). Each line represents the tunnel profile of a single snapshot. Black dashed lines mark the average length of tunnels in their respective cluster. T1 is the longest and T3 is the shortest of all the three tunnels.

Table 2: Characteristics of tunnels identified in PlaF$_A$, PlaF$_B$, and di-PlaF using CAVER.

Tunnel	Occurrence[a,b]				Average length[c]			
	Monomer		Dimer		Monomer		Dimer	
	PlaF$_A$	PlaF$_B$	di-PlaF$_A$	di-PlaF$_B$	PlaF$_A$	PlaF$_B$	di-PlaF$_A$	di-PlaF$_B$
T1	9.12	6.07	3.30	5.55	25.72	23.90	41.07	23.81
T2	18.35	5.07	18.45	5.02	23.97	22.35	23.89	22.20
T3	16.80	6.25	21.30	5.80	15.84	16.16	15.52	15.78

[a] Snapshots in which the tunnel is identified with respect to the total number of snapshots, in %.

[b] Data calculated with a probe radius of 2.0 Å.

[c] In Å.

As a conclusion from these results, the active site of PlaF is connected to its surface with three tunnels. In the t-PlaF$_A$ configuration, only T1 and T2 allow direct access of GPL or LGPL substrates from the membrane.

5.2 GPL and LGPL substrate extraction into solvent and acyl chain mobility

For probing the energetics of GPL and LGPL substrate extraction from the membrane into the solvent, I have first computed the free energy profile for DLPG and 2LMG extraction. This serves as reference points for the energetics that would be obtained for substrate access in PlaF. DLPG and 2LMG were chosen, as for these substrates, PlaF shows the highest activity (23, 24).

Applying sMD simulations, the substrates were pulled out of the membrane with their head group first, applying a constant velocity of 1 Å ns^{-1}, until the hydrophobic tails do not interact anymore with the membrane surface. Once in the solvent, the hydrophobic chains curl in order to minimize their exposure to water.

The differences in the PMF between the states in the solvent and in the membrane are ~13 ± 0.1 kcal mol^{-1} for DLPG and ~8 ± 0.3 kcal mol^{-1} for 2LMG (Figure 12A). Converged PMFs of extraction of the substrates were obtained after ~40 ns of US simulation time per window (Figure 12B), which also resulted in sufficient overlap between the reaction coordinate distributions of neighboring windows (Figure 12C).

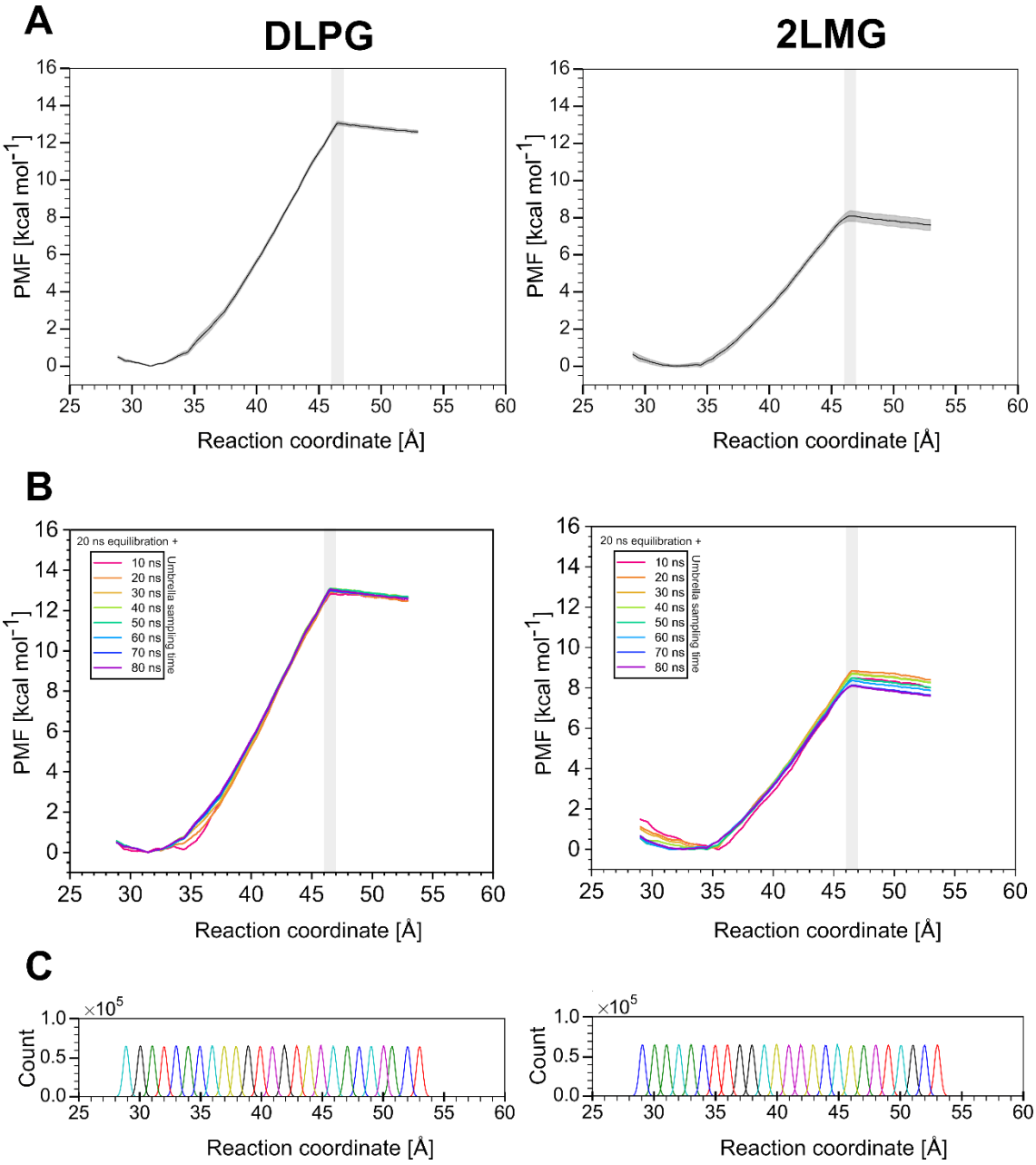

Figure 12: Extraction of PlaF substrates from the membrane. A) PMF profiles of selected substrates, DLPG (left) and 2LMG (right). B) Convergence plot indicates sufficient sampling time for the two substrates, DLPG (left) and 2LMG (right); convergence of profiles starts around 40 ns; at 100 ns, the PMF profiles are converged for both the substrates. C) Histograms indicate sufficient overlap among the umbrella windows of DLPG (left) and 2LMG (right) using a force constant of 5 kcal mol^{-1} $Å^{-2}$; the median overlap is 4.13% and 4.10% for DLPG and 2LMG, respectively. The gray box in A and B indicates the section of the reaction coordinate where the substrate loses interaction with the membrane surface.

Unfortunately, the PMF cannot be measured experimentally. However, the free energy difference between the two states of the lipid in the solvent and in the membrane can be related to the experimentally measured CMC values using

Equation 10, as discussed in section 2.8.1 (124). The CMC values were obtained from previous studies (168, 169). The determined excess chemical potential for DLPG and 2LMG is 11.69 kcal mol^{-1} and 7.60 kcal mol^{-1}, respectively. These values are within chemical accuracy (170, 171) to those obtained from the PMF.

For access to T3, substrates would need to leave the membrane and pass through the water phase, which makes this route energetically unfavorable. Hence, T3 was not considered for further analysis. As T1 and T2 are immersed in the hydrophilic membrane surface (Figure 13A), access of GPL and LGPL substrates to the tunnels via the head groups is plausible. However, the tunnels' diameters are much smaller than that of the GPL lipid (see above). To explore the possibility that lipids access via their acyl chain instead, I have probed how frequently the terminus of a GPL's acyl chain can reach the membrane interface. The probability distribution of GPL's acyl chains with respect to the coordinate perpendicular to the membrane (z-coordinate) was determined during the last 40 ns of 300 or 100 ns long MD simulations for membrane bilayers with or without t-$PlaF_A$, respectively (Figure 13A). Tails from both the upper and lower leaflet were considered. Positive *z*-coordinate values indicate that a tail moves toward the water-membrane interface of *its* leaflet; negative values indicate that it moves toward the interface of the *opposite* leaflet. The peak of the probability distributions is at $z \approx 2$ Å indicating the mobility of lipid termini within the leaflet (Figure 13A, see also movies M1 and M2). The interface of the simulated membrane is at $z \approx 10$ Å (Figure 13B). Notably, the cumulative probability of finding an acyl chain terminus at $z > 10$ Å is 1.5 % and 1.0 % for systems with or without PlaF, respectively. Hence, there is a finite likelihood that acyl chain termini can reach the entrances of T1 and T2. This result is also supported by the electron density profiles of the membrane components (Figure 13B).

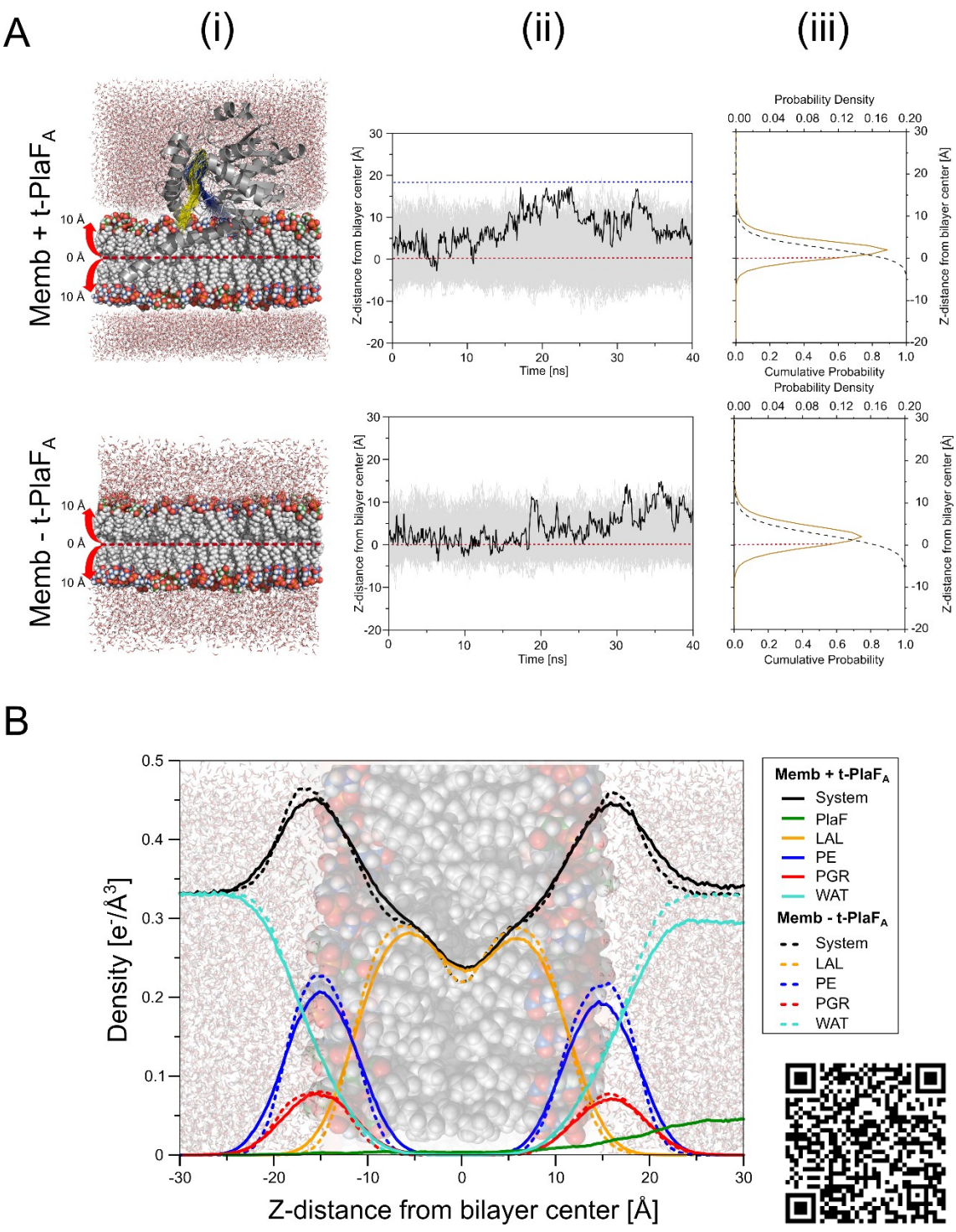

Figure 13: Distribution of acyl chain termini from GPLs and corresponding density profiles. A) To investigate the distribution of acyl chain termini during unbiased MD simulations, (i) a system with t-PlaF$_A$ (upper) and without t-PlaF$_A$ (lower) was considered. In t-PlaF$_A$, the tunnels T1 (blue) and T2 (yellow) are immersed into the head group region of the upper leaflet. (ii) With the bilayer center positioned at ≈ 0 Å (broken red line), the z-coordinate distance was measured for each acyl chain terminus during MD simulations of 40 ns. Tails from both upper as well as the lower leaflet were considered. For both systems, the acyl chain termini can reach the membrane interface, located around 10-15 Å from the bilayer center (please see section B for the electron density profile). For a system with t-PlaF$_A$, the tail termini go as high as 20 Å, close to the entrance

of access tunnels (broken blue lines) which are at z = 19.07 ± 1.42 Å for T1 and z = 17.94 ± 1.07 Å for T2. For system without t-PlaF$_A$, the termini reach 16 Å of the z-coordinate. Not only the acyl chain termini go up to the membrane surface, but can also go beyond the bilayer center along the negative z-coordinate. Black curve represents an example, where the acyl terminus of selected lipid reaches the membrane interface (please see the movies M1 and M2 corresponding to this event for the systems with and without t-PlaF$_A$ respectively; to access the movies, please scan the QR-code on the bottom right of the Figure). iii) Probability density plot (brown curve) shows that the distribution of acyl chains shifts toward the positive z-coordinate, indicating that tails of GPLs can reach the membrane interface. The cumulative probability (broken black curve) of finding an acyl chain terminus at z > 10 Å is 1.5 % and 1.0 % for system with and without t-PlaF$_A$ respectively. B) The electron density profile was measured and compared for the two systems. Differences in the profiles are due to GPLs and water replaced by t-PlaF$_A$.

As a conclusion from this result, for t-PlaF, the access route of substrates to T3 is energetically unfavorable. By contrast, acyl chain termini of GPL lipids can reach the entrances of T1 and T2 during the time scales of the MD simulations.

5.3 Access modes of GPL and LGPL substrates into PlaF

As a prerequisite to computing the energetics of substrate access to the active site of PlaF, I aimed to identify favorable access modes. I have applied sMD simulations (90) to pull the substrates inside T1 and T2 *via* head access first or tail access first (Figure 14). The closest substrate to the tunnel entrance was chosen for sMD simulations. The terminal oxygen and nitrogen atom of PG or PE head groups, respectively, was considered for head access pulling. For tail access, the terminal carbons of respective acyl chains were considered (Figure 14). Substrates from the membrane were initially pulled through consecutive virtual points in T1 or T2 using four or five steps, respectively (Figure 15A, Table 3). However, pulling with terminal atoms leaves the cleavage site of the substrate distant to the catalytic S137 (Figure 15B). Therefore, the substrates were further pulled into T3, using three additional steps (Figure 15A). Depending on the access mode, the *sn*-1 or *sn*-2 sites of respective substrates were further pulled toward the nucleophilic OH group of the catalytic S137 (Table 3). Finally, this resulted in pulling pathways subdivided into eight and nine steps for T1 and T2, respectively (Table 3).

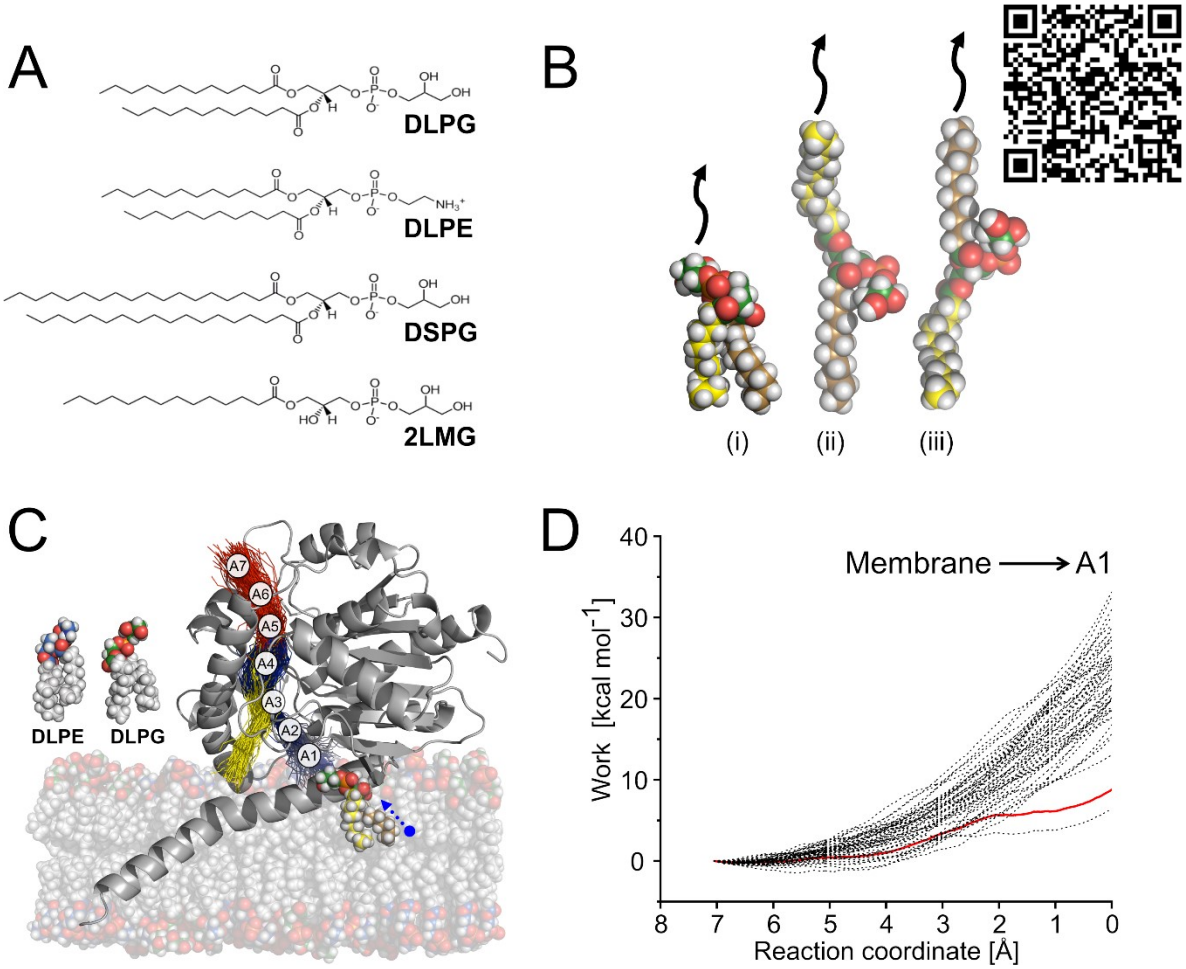

Figure 14: Illustration of the substrate access in t-PlaF$_A$. A) Investigated GPL substrates, 1,2-dilauroyl-*sn*-glycero-3-phosphoglycerol (DLPG), 1,2-dilauroyl-*sn*-glycero-3-phosphorylethanolamine (DLPE), 1,2-distearoyl-*sn*-glycero-3-phosphoglycerol (DSPG) and LGPL substrate, 1-myristoyl-2-hydroxy-*sn*-glycero-3-phosphoglycerol (2LMG). B) Possible modes by which a GPL can access a tunnel (indicated with black arrow): with its head (green spheres represent the C atoms) first (i), tail 1 (yellow spheres represent the C atoms) first (ii), or tail 2 (orange spheres represent the C atoms) first (iii). Please see the movies M3-M8 corresponding to the DLPG access in t-PlaF$_A$ via head, tail 1 and tail 2 access respectively; to access the movies, please scan the QR-code on the top right of the Figure. C) PlaF is embedded in a membrane consisting of DLPE (head group C atoms as blue spheres) and DLPG (head group C atoms as green spheres) at a ratio of 3:1. The DLPG closest to the entrance of T1 (acyl chains colored) is shown while being loaded by its head, in the direction marked with a blue arrow. A segmented path was considered for substrate access. T1 was segmented into four parts, and T3 into three parts, which are used as pulling points in sMD simulations. Depending on the access mode, in the last pulling step, the *sn*-1 or *sn*-2 site of the substrate is further pulled toward the nucleophilic OH group of the catalytic S137, resulting in total eight steps. A similar approach was used for T2 (Figure 15A). D) For the first segment of T1 (i.e., A1), the work done (black dotted lines) during 50 independent replicas of sMD simulations to pull the DLPG from the membrane is plotted against the reaction coordinate. The coordinates of the replica with the work-versus-reaction coordinate profile closest to the Jarzynski's average (red line) are considered for pulling in the next segment.

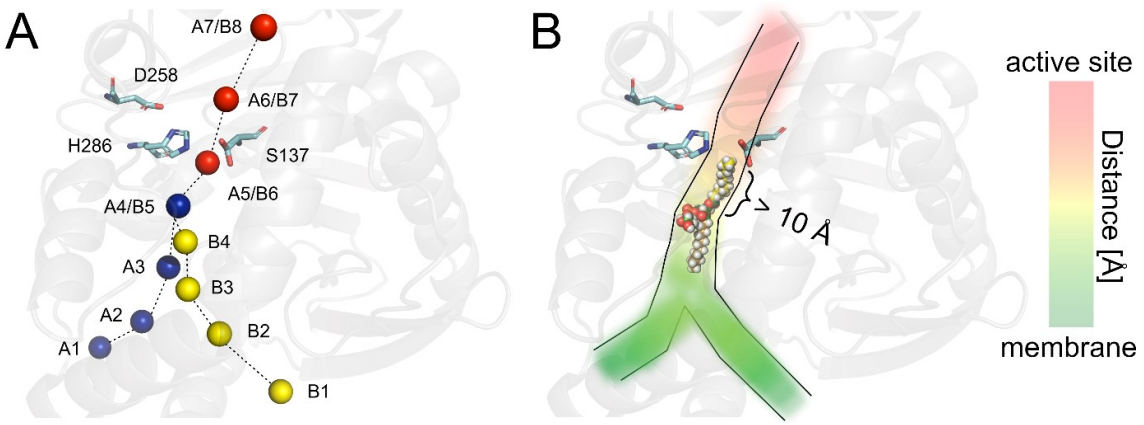

Figure 15: Substrate access pathway. A) Pulling points for substrate access in PlaF: using sMD simulations, substrates are first pulled out of the membrane to A1 (for T1, blue spheres) or B1 (for T2, yellow spheres). Red spheres correspond to pulling points lining T3. Substrate pulling through T1 involves points A1 to A7, while pulling through T2 involves points B1 to B8. T2 merges into T1 after A3; both follow a common path toward T3 across A4/B5. Catalytic residues are represented as cyan sticks. B) Requirement of T3 for substrate access: when pulled with terminal atoms, the *sn*-1 site of the substrate remains several Angstroms away from the catalytic S137 and, hence, needs to be further pulled into T3. Since the tunnels are almost straight, the reaction coordinate monotonically decreases as the substrate approaches the active site from the membrane.

Table 3: Pulling points across the tunnels for sMD simulations.

T1 pulling points[a,b]	Amino acid residues	T2 pulling points[a,c]	Amino acid residues
A1	V30, P205, L206	B1	A24, S102
A2	E34, F192	B2	L27, N225
A3	G72, L214, V287	B3	D76, F192, N225
A4	F71, D161, F192	B4	A73, V199, A221
A5	M138, L184, H286	B5	F71, D161, F192
A6	M138, F174	B6	M138, L184, H286
A7	K170, Q234, Y236	B7	M138, F174
A8[d]	S137	B8	K170, Q234, Y236
		B9[d]	S137

[a] Pulling points are COM of corresponding amino acid residues.

[b] For T1, A1-A4 are components of T1, and A5-A7 are components of T3.

[c] For T2, B1-B4 are components of T2, B5 is a component of T1, and B6-B8 are components of T3.

[d] For S137 of A8/B9, the OH group of the nucleophile was considered as a pulling point.

As a reaction coordinate, the distance between the pulled atom of a substrate and the consecutive virtual point was used. For each step, I repeated the pulling 50 times and computed the work done as a function of the reaction coordinate. By applying Jarzynski's relation (Equation 3) (112), the work was related to the free energy difference between the two states of the pulling simulation. The sMD trajectory whose work-versus-reaction coordinate profile is closest to the

Jarzynski average was considered most favorable (Figure 14D). Its endpoint provided the starting point for the sMD simulations in the next part of the pulling pathway. As a result, the access pathway is close to the lowest-free energy pathway of substrate access to the catalytic site. Overall, this approach is the reversed version of sampling unbinding trajectories of ligands from proteins before applying Jarzynski's relation (107, 160, 172) but uses piecewise sMD simulations along the pathway to account for the curvilinear tunnels. A total of ~27 µs of sMD simulation time was used for either tunnel in t-PlaF$_A$ (Table 4).

Table 4: Overview of sMD simulations for the substrate access through T1 and T2 in t-PlaF$_A$.

Tunnel	Substrate	Mode of access	Per access simulation time[a]	Number of replicas	Total simulation time[b]	Per substrate simulation time[b]
T1	DLPG	head	~46	50	~2.30	~8.80
		tail 1	~62	50	~3.10	
		tail 2	~67	50	~3.35	
	DLPE	tail 1	~60	50	~3.00	~6.00
		tail 2	~60	50	~3.00	
	DSPG	tail 1	~66	50	~3.30	~6.75
		tail 2	~69	50	~3.45	
	2LMG	head	~51	50	~2.55	~5.00
		tail 1	~49	50	~2.45	
T2	DLPG	head	~57	50	~2.85	~10.25
		tail 1	~70	50	~3.50	
		tail 2	~78	50	~3.90	
	DLPE	tail 1	~54	50	~2.70	~5.75
		tail 2	~61	50	~3.05	
	DSPG	tail 1	~63	50	~3.15	~6.15
		tail 2	~60	50	~3.00	
	2LMG	head	~52	50	~2.60	~5.30
		tail 1	~54	50	~2.70	

[a] In ns.

[b] In µs.

The activity of PlaF for GPL decreases with the increasing lengths of the acyl chain between C12 and C18, irrespective of the type of head group, PG or PE (23). In addition, the number of acyl chains in a substrate also influences the PlaF activity, with LGPLs yielding a higher activity than GPLs (24). Hence, I chose DLPG with which PlaF is most active (23), DLPE, DSPG, and 2LMG, for

generating access modes (Figure 14A). Figure 14 exemplarily shows illustrations of the three access types for DLPG. Work-versus-reaction coordinate profiles for all pulling simulations related to DLPG access are shown in Figure 16 for T1 and Figure 17 for T2. Based on the computed PMF to evaluate the energetics of the access modes (see the next chapter), only tail access was considered for sMD simulations of the other GPL substrates (Figure 18). For 2LMG, head and tail access were considered for sMD simulations.

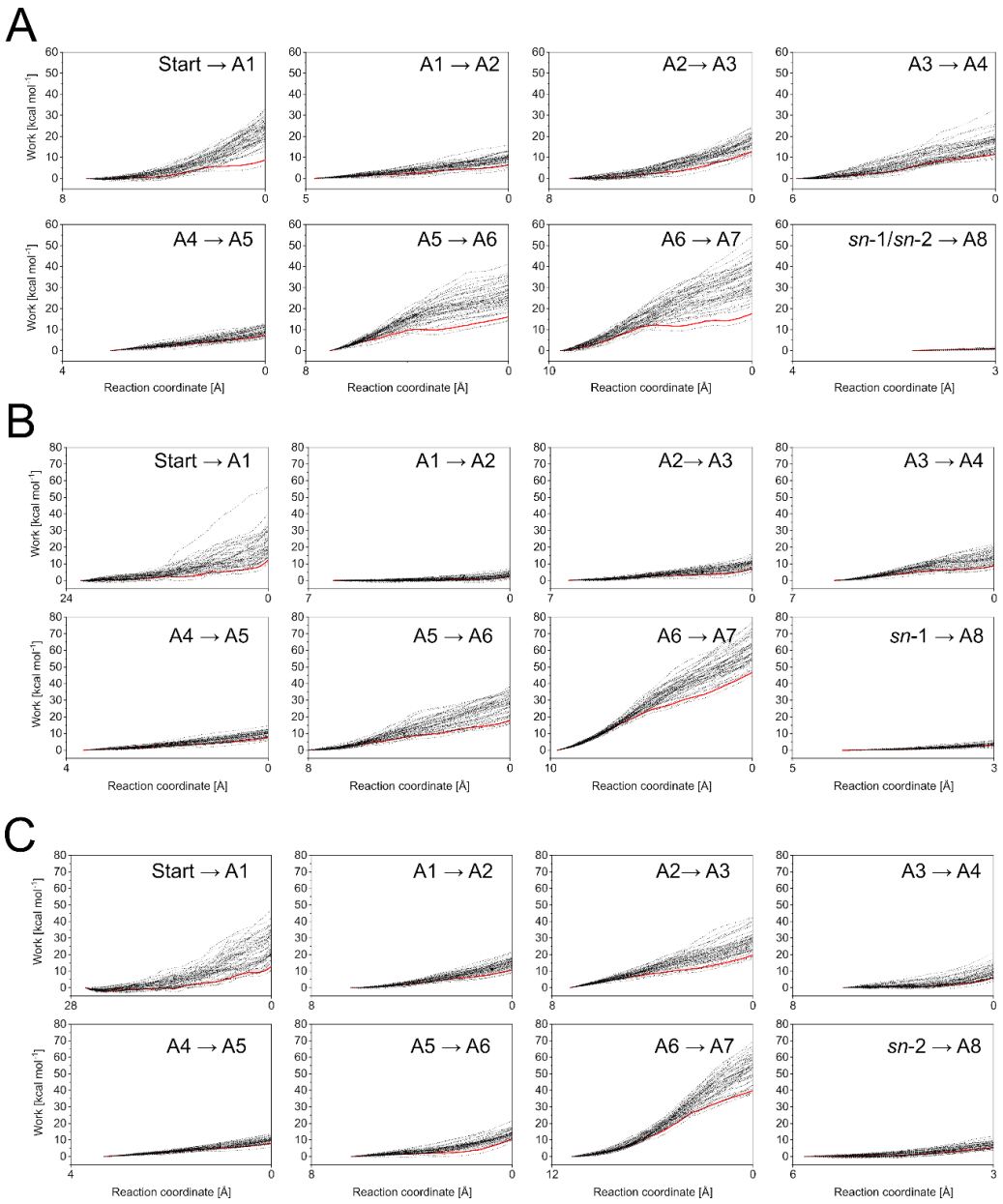

Figure 16: Work distributions (black lines) obtained from 50 replicas of sMD simulations to pull DLPG across T1 via (A) head, (B) tail 1, and (C) tail 2. For each mode of access, DLPG is first pulled out of the membrane to point A1. A replica closest to Jarzynski's average (red line) was

considered as the starting point for the next pulling, A1 → A2. This pulling continues until A7, after which the *sn*-1/*sn*-2 of DLPG is further pulled to the nucleophilic OH group of the catalytic S137. The reaction coordinate denotes the distance to the target point.

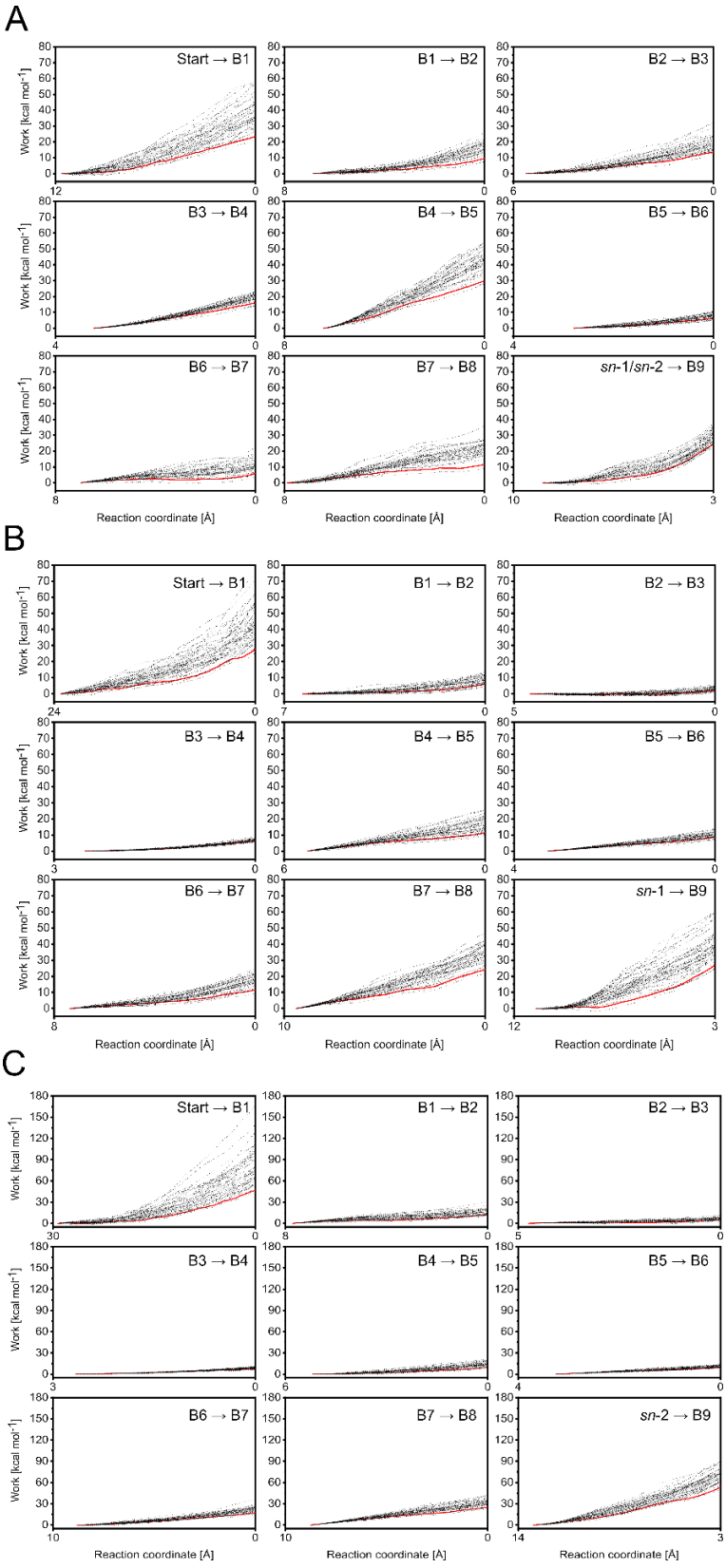

Figure 17: Work distributions (black lines) obtained from 50 replicas of sMD simulations to pull DLPG across T2 via (A) head, (B) tail 1, and (C) tail 2. For each mode of access, DLPG is first pulled out of the membrane to point B1. A replica closest to Jarzynski's average (red line) was considered as the starting point for the next pulling, B1 → B2. This pulling continues until the pulling point B8, after which the *sn*-1/*sn*-2 of DLPG is further pulled to the nucleophilic OH group of the catalytic S137. The reaction coordinate denotes the distance to the target point.

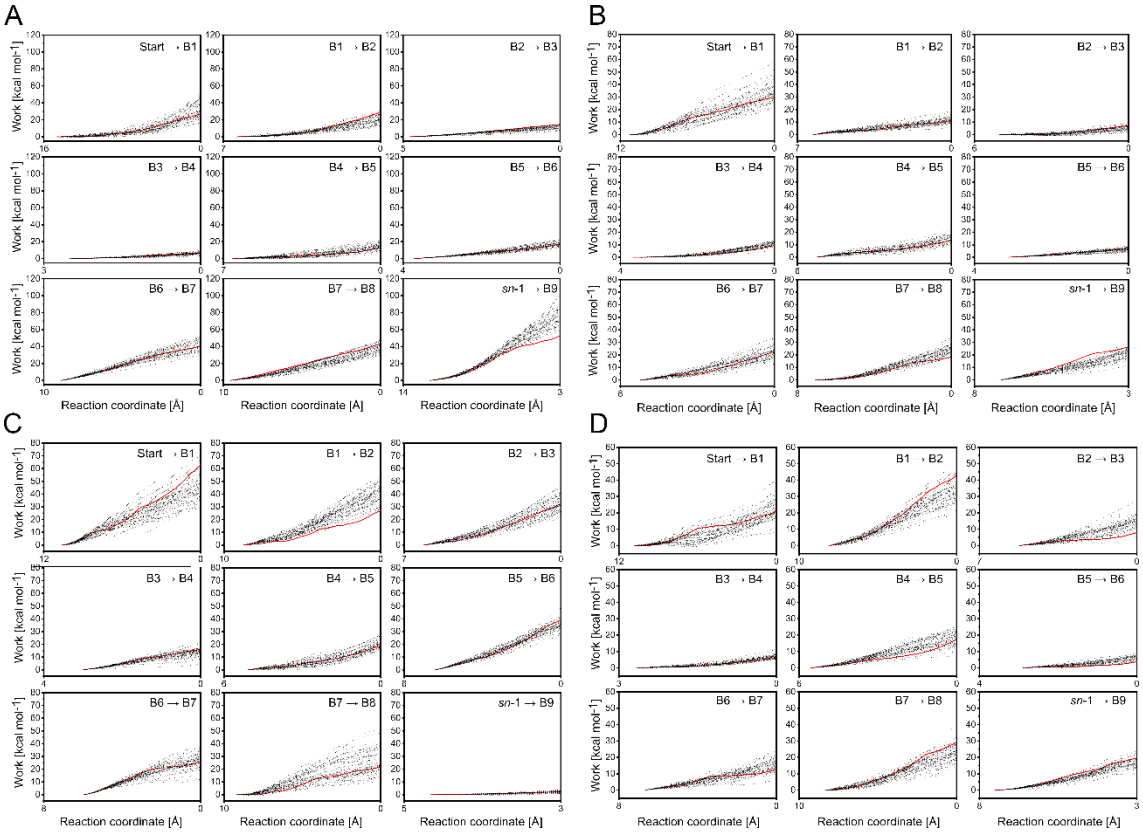

Figure 18: Work distributions (black lines) obtained from 50 replicas of sMD simulations to pull (A) DSPG via tail 1, (B) DLPE via tail 1, and (C) 2LMG via head (D) 2LMG via tail 1 across T2. Each substrate was first pulled out of the membrane to point B1. A replica closest to Jarzynski's average (red line) was considered as the starting point for the next pulling, B1 → B2. This pulling continues until the pulling point B8, after which the *sn*-1 of respective substrate is further pulled to the nucleophilic OH group of the catalytic S137. The reaction coordinate denotes the distance to the target point.

As a conclusion from this result, there are seven access modes of GPL and two of LGPL substrates into t-PlaF$_A$ that were generated for T1 and T2, resulting in 18 access modes in total.

5.4 Potentials of mean force of DLPG access modes

PMFs were computed from US simulations (109) and post-processing with WHAM (121, 122) to evaluate the energetics of substrate access for the access modes described in the previous chapter (Figure 14). As a reaction coordinate, the distance between the COM of the three oxygen atoms of the glycerol moiety in the substrate to the COM of C$_\alpha$ atoms of the catalytic residues S137 and H286 was used. Residue D258 was not included in the reaction coordinate, as its side

chain is distant from the active site (Figure 15A). As the tunnels are almost straight, the reaction coordinate monotonically decreases as the substrate approaches the active site from the membrane (Figure 15B). Initially, the focus was on the US simulations for the best PlaF substrate (23), DLPG. PMFs were calculated for the three access modes of DLPG across either tunnel, T1 and T2. The PMFs were evaluated for convergence, excluding the first 200 ns of 300 ns sampling time. PMFs were found converged by 300 ns, yielding a maximal difference of ~1 kcal mol^{-1} as to a PMF computed from 280 ns per window (Figure 19). The median overlap between the reaction coordinate distributions of neighboring windows was sufficient (≥ 4.8% and 3.5% for T1 and T2, respectively) Figure 20.

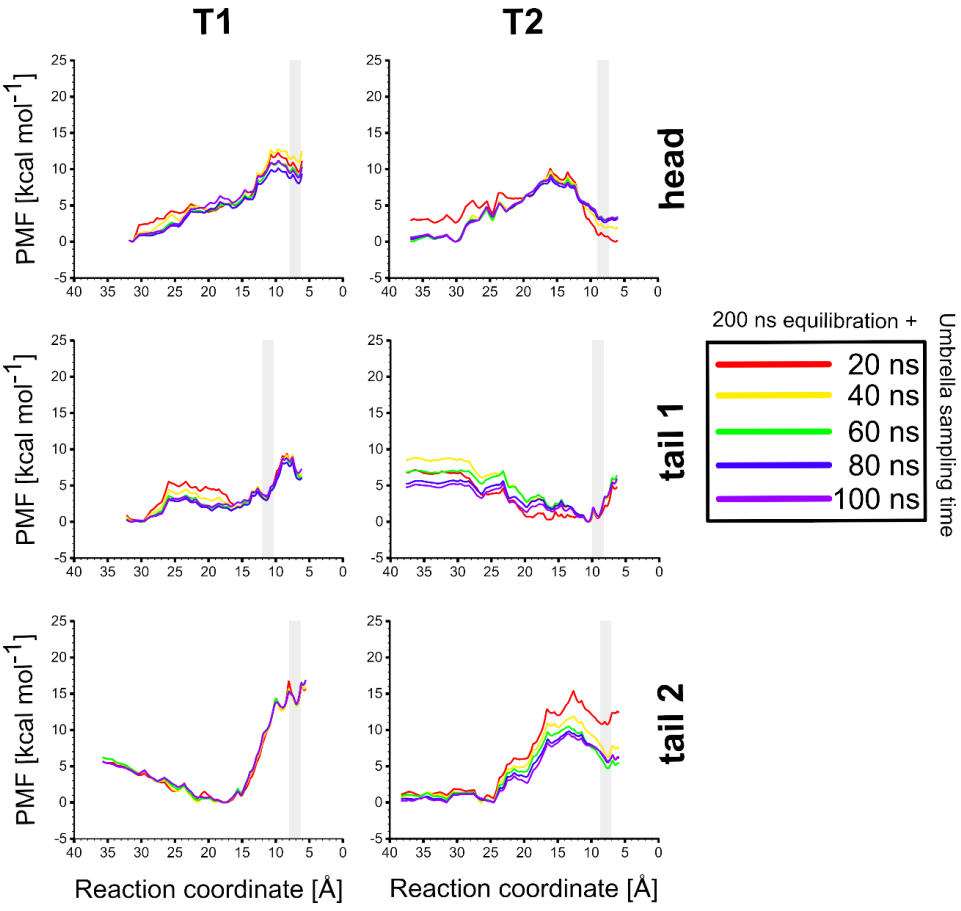

Figure 19: Convergence of PMFs for substrate access of DLPG. PMFs were computed every 20 ns for the range of 220-300 ns (see legend) of umbrella sampling simulations per window for T1 (left) and T2 (right). The first 200 ns of the sampling simulations were considered for equilibration and removed for every system. The gray box indicates the location of the active site. Overall, 300 ns of umbrella sampling per window are sufficient to achieve converged PMFs.

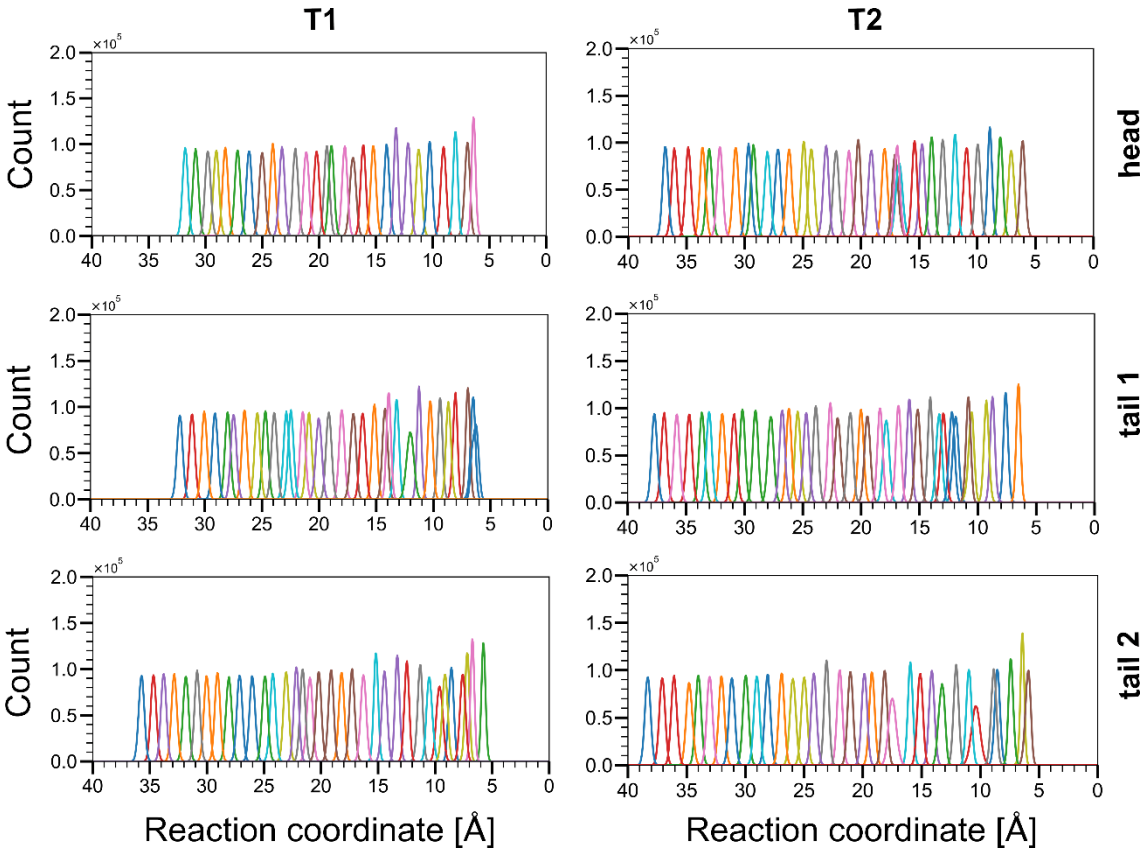

Figure 20: Distribution of reaction coordinate values obtained by umbrella sampling for DLPG access via T1 (left) and T2 (right). A force constant of 5 kcal mol^{-1} Å^{-2} was used to restrain the positions of DLPG to the reference point of an umbrella window, which resulted in distributions with a median overlap of at least 4.84% and 3.47% for T1 and T2, respectively.

The PMFs of DLPG access modes show marked differences (Figure 21A). Access with the head first is the least favorable for both T1 and T2, resulting in steep PMFs with free energy barriers of 11 and 9 kcal mol^{-1} (Figure 21A), in contrast to tail access. Most of the residues within a radius of 3 Å in T1 and T2 have either a neutral non-polar side chain, which likely facilitates tail access to the active site of PlaF. Furthermore, access with either one of the two tails first is more favorable in T2 than T1 (Figure 21A). Finally, access with tail 1 first in T2 is most favorable and results in no free energy barrier until the substrate reaches the active site (Figure 21A). As the two acyl chains of DLPG are identical, these results suggest that their connection with the glycerol moiety causes differences in how the lipid interacts with the tunnel, which may explain how PlaF achieves regioselectivity to exert its PLA_1 function.

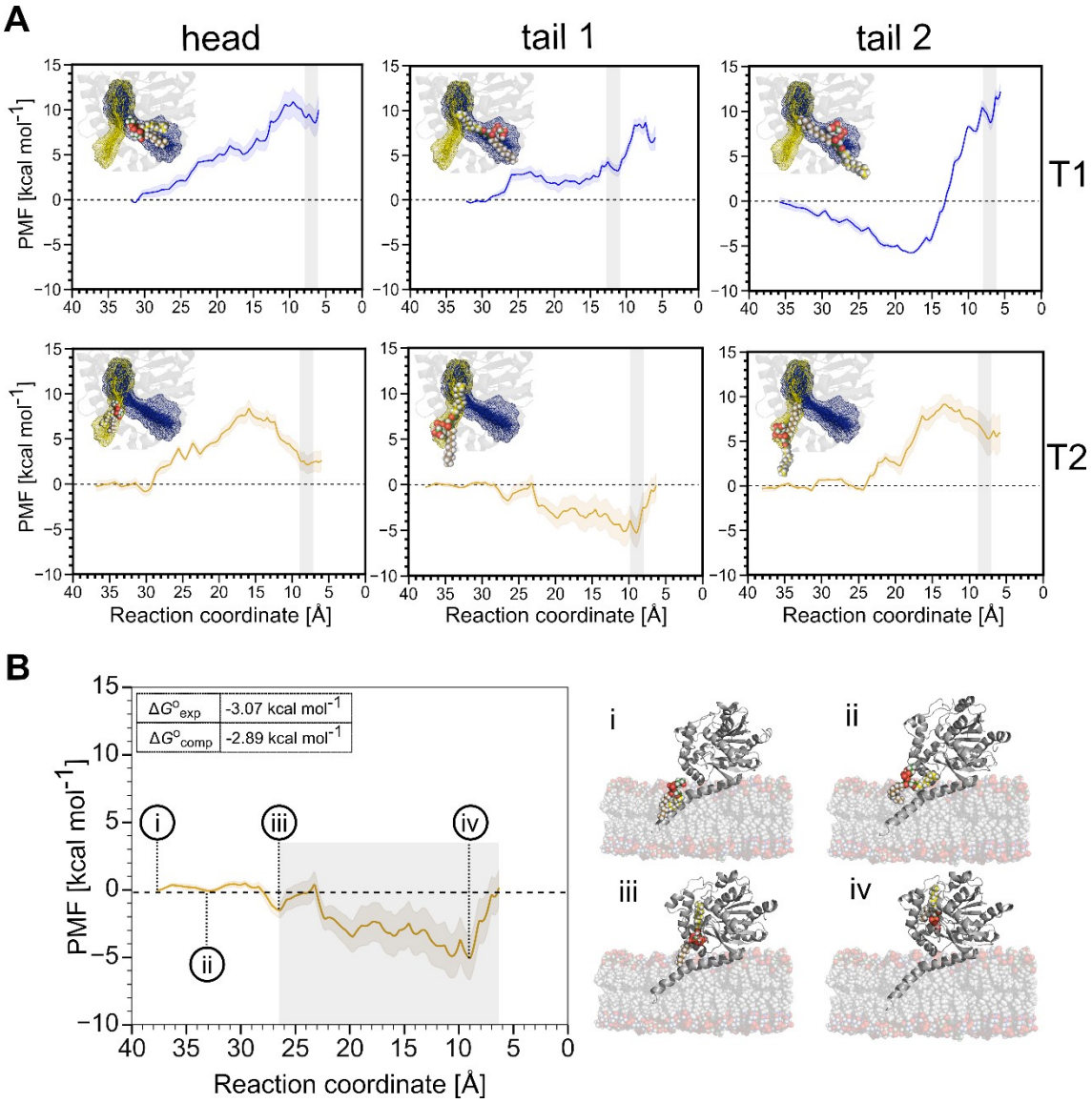

Figure 21: Potential of Mean Force profiles for DLPG access. (A) PMFs of three access modes (head, tail 1, tail 2; see Figure 14B) of DLPG in T1 (blue curve) and T2 (yellow curve). For both tunnels, access with tail 1 first yields the lowest free-energy barriers to reach the active site. Furthermore, DLPG access into T2 with tail 1 first is overall the most favorable. The catalytic site is marked with a gray box. Insets within the plots illustrate the different DLPG access modes into the respective tunnels. (B) States during DLPG access via tail 1 through T2, shown on the right, are marked in the PMF profile (left). The gray box corresponds to the integration limits used to calculate K_{eq} (Equation 11) to determine ΔG^{o}_{comp} (see inset). State i: The starting position of DLPG (in the membrane). State ii: Tail 1 reaches the surface of the membrane close to the entrance of T2. State iii: Tail 1 enters inside T2, while tail 2 remains within the membrane. State iv: *sn*-1 site of tail 1 reaches the catalytic site.

To validate the obtained results, I have then computed the absolute binding free energy of DLPG to PlaF from the PMF for tail 1 access in T2, ΔG°_{comp} =

-2.89 ± 1.46 kcal mol^{-1} as described in the section 2.8.2 using Equation 13. Assuming that product formation is slower than substrate dissociation from an enzyme, the Michaelis constant K_m is equal to the dissociation constant K_D of the enzyme-substrate complex (173, 174). Under this assumption, from K_m = 7.612 ± 1.907 mM for DLPG in PlaF (175), the experimental binding free energy ΔG°_{exp} = -3.07 ± 0.30 kcal mol^{-1} at T = 303 K is calculated, which is within chemical accuracy (176) of ΔG°_{comp}.

I have also computed ΔG°_{comp} for the other five access modes of DLPG as described in the section 2.8.2 using Equation 13. The lowest ΔG°_{comp} among all six modes was obtained for tail 2 access in T1 (Table 5). However, the PMF profile (Figure 21A) reveals that the configurational free energy minimum is not situated close to the active site but in the middle of T1. Here, one of the tails is still in the membrane, while the other is being loaded into the tunnel. If the PMF profile is integrated with two separate parts, first, a negative free energy for tail access into the tunnel results, followed by a positive free energy to reach the active site. This suggests that this access mode cannot yield a catalytically active configuration. For the other four access modes, ΔG°_{comp} > 0.96 kcal mol^{-1} (Table 5). These findings corroborate tail 1 access of DLPG in T2 as the most likely access mode.

Table 5: Overview of computed absolute binding free energy of DLPG to t-PlaF$_A$ from PMF.

System[a]	T1HG	T1T1	T1T2	T2HG	T2T1	T2T2
ΔG°_{comp}[b,c]	1.81 ± 0.25	1.31 ± 0.16	-4.01 ± 0.56	1.00 ± 0.48	-2.89 ± 1.46	0.96 ± 0.27

[a] Each system is represented as XXYY, where XX is name of tunnel (i.e., T1 or T2), and YY is the mode of substrate access (i.e., head (HG), tail 1 (T1), or tail 2 (T2)).

[b] In kcal mol^{-1}.

[c] Error estimation: For each system, the last 100 ns of sampling data was split into five independent blocks of 20 ns each. The PMF profiles obtained were used to determine the absolute binding free energy for each block, and the standard error of the mean was calculated.

Along the PMF of tail 1 access of DLPG in T2, four distinct states can be identified (Figure 21B). The two tails of DLPG are immersed in the membrane at a reaction coordinate

value of ~38 Å from the active site (state i). The PMF remains essentially unchanged if tail 1 approaches the surface of the membrane and the entrance of T2 (state ii). This is concordant with the tail distributions along the z-coordinate during unbiased MD simulations (Figure 13), indicating that tail termini can reach one of the access tunnels of t-PlaF$_A$ without a considerable energetic cost. Once tail 1 enters T2, the PMF becomes negative (state iii), indicating that way of DLPG access is favorable. Finally, at ~8 Å of the reaction coordinate, the PMF has a global minimum (state iv). There, tail 1 is located in T3, and the acyl moiety at the *sn*-1 position of DLPG is close to the catalytic S137 of PlaF (Figure 21 and Figure 22B) such that a nucleophilic attack can commence.

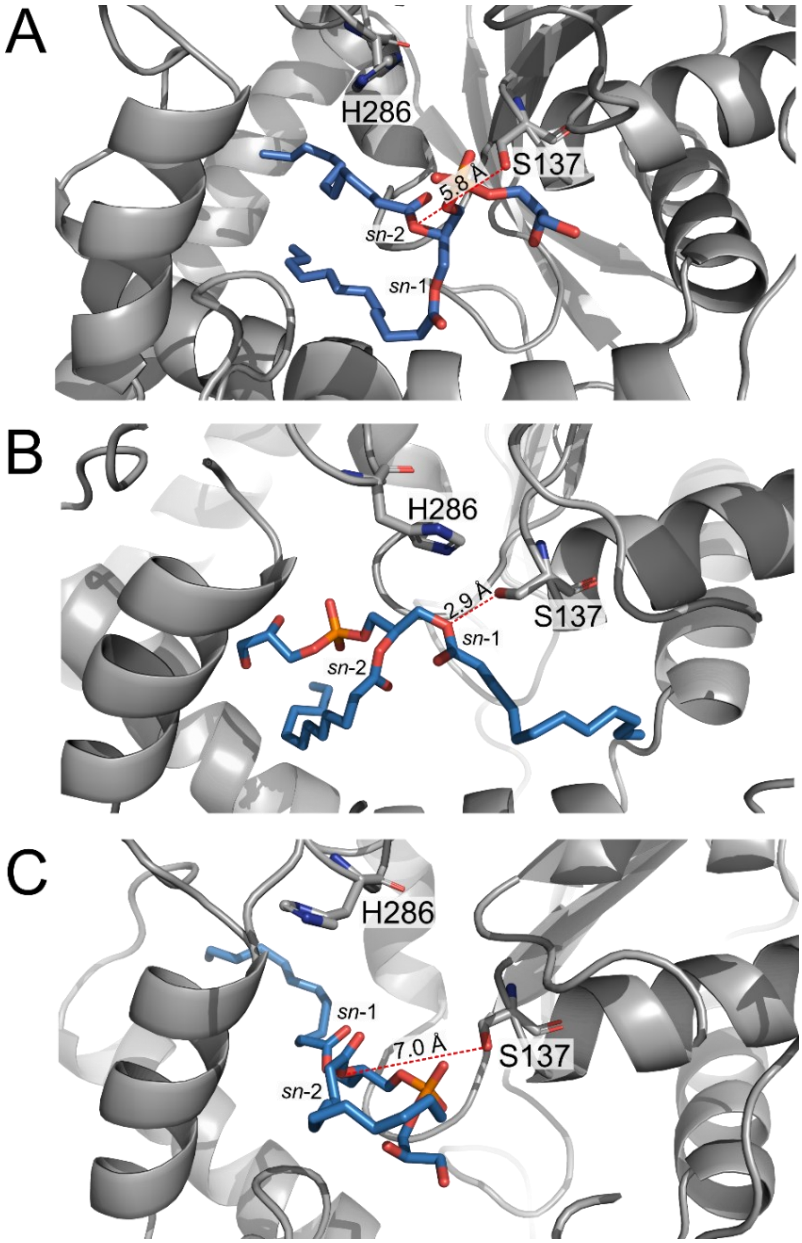

Figure 22: DLPG access to the catalytic site of t-PlaF$_A$. DLPG accessing t-PlaF$_A$ through T2 via (A) head first, (B) tail 1 first, and (C) tail 2 first. Snapshots were retrieved after 300 ns of umbrella sampling at the reference point where the substrate's cleavage site is closest to the active site of t-PlaF$_A$. For tail 1, the *sn*-1 site of DLPG comes closest to the catalytic residues (shown in sticks), compared to other access modes.

As a conclusion from this result, T2 has been identified as the preferred access tunnel for DLPG in PlaF. Access with tail 1 first is most favorable there. This is in line with PlaF being a PLA_1, which cleaves its substrates at the *sn*-1 position. As of T3, it is likely essential for substrate access by allowing to accommodate the substrate tail to be hydrolyzed by PlaF.

5.5 Potentials of mean force for accesses of other substrates

Considering the results for DLPG, I have also performed US simulations for DSPG and DLPE only for tail 1 access. For the LGPL substrate, it has remained undetermined if the head or tail access is energetically favorable; hence, I have performed US simulations for both access modes of 2LMG. As for DLPG, T2 is preferred over T1, regardless of the access modes (Figure 21A). Thus, I only considered T2 for computing PMFs for the other substrates. Similar to DLPG, the PMFs converged at 300 ns of sampling time, yielding a maximal difference of ~0.5 kcal mol^{-1} as to a PMF computed from 280 ns per window (Figure 23). Neighboring umbrella windows have a sufficient median overlap ≥ 3.2% (Figure 24).

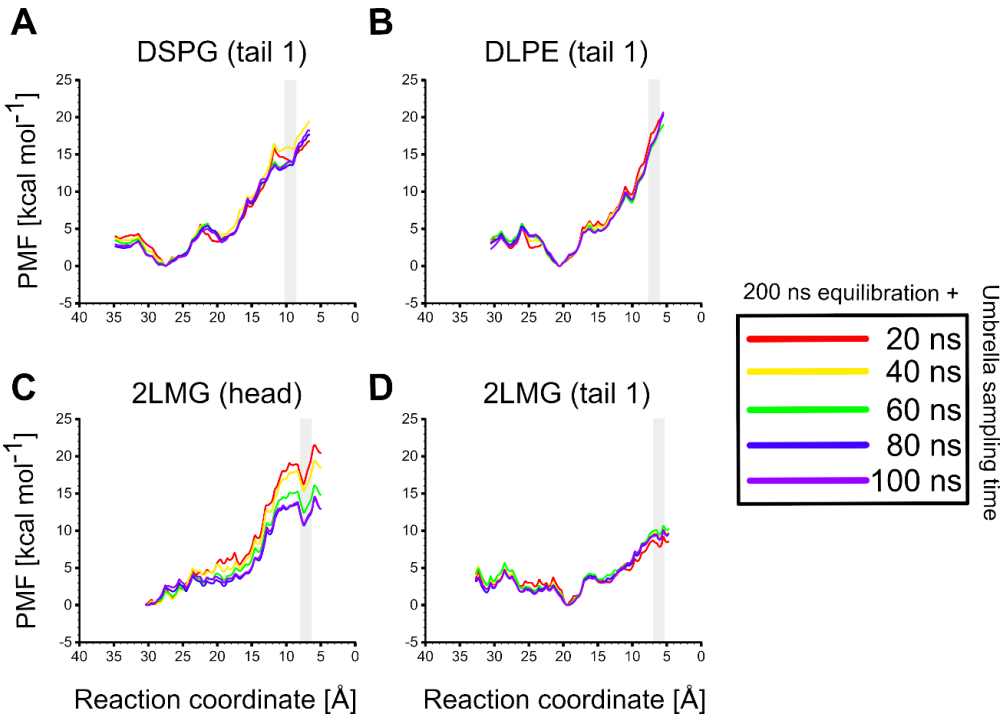

Figure 23: Convergence of PMFs for other substrate access via T2. PMFs were computed

every 20 ns for the range of 220-300 ns (see legend) of umbrella sampling simulations per window for DSPG-tail 1 access (A), DLPE-tail 1 access (B), 2LMG-head access (C), and 2LMG-tail 1 access (D). The first 200 ns of the sampling simulations were considered for equilibration and removed for every system. The location of the active site is indicated by a gray box. Overall, 300 ns of umbrella sampling per window are sufficient to achieve converged PMFs.

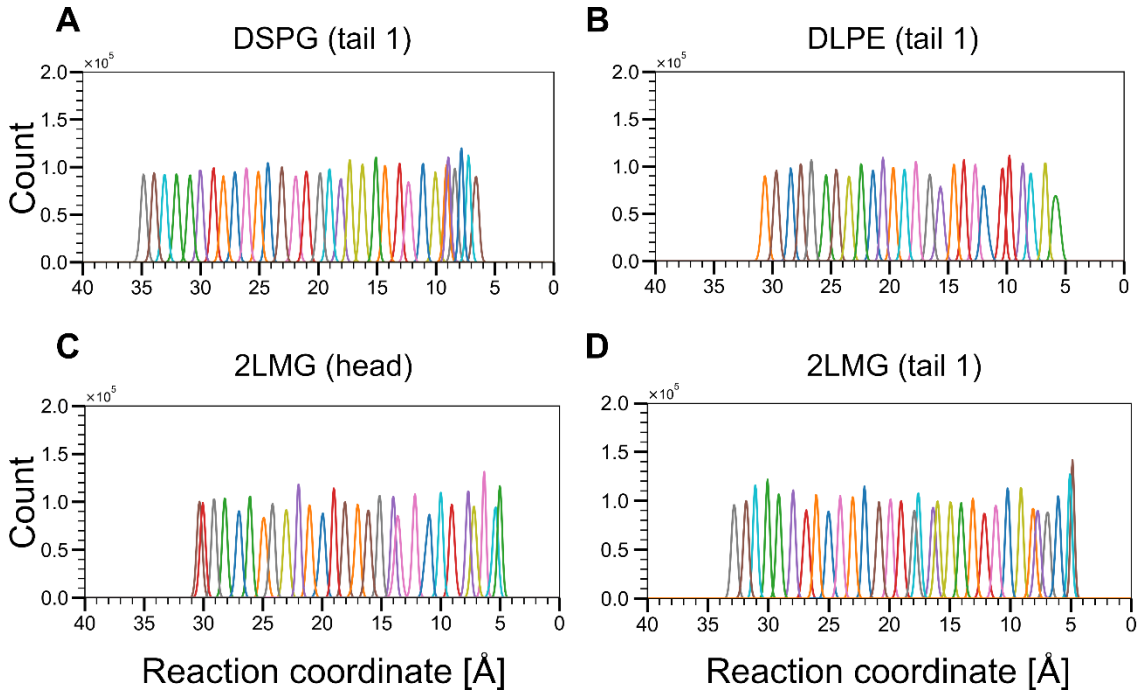

Figure 24: Distribution of reaction coordinate values obtained by umbrella sampling for loading of other substrates across T2. A force constant of 5 kcal mol^{-1} Å^{-2} was used to restrain the positions of substrate for (A) DSPG-head access, (B) DLPE-tail 1 access, (C) 2LMG-head access, and (D) 2LMG-tail 1 access to the reference point of an umbrella window, which resulted in distributions with a median overlap of at least 3.24%.

For DSPG and DLPE, access with tail 1 first in T2 results in pronounced free energy barriers of 11 and 14 kcal mol^{-1} (Figure 25A, B), in contrast to DLPG (0.5 kcal mol^{-1}). This finding indicates that a longer acyl chain or a neutral head group makes substrate access to PlaF disfavorable, which coincides with lower PlaF activities for such substrates (23). For 2LMG, access with the tail first is more favorable than with the head, as for DLPG (Figure 25C, D). Furthermore, tail access by 2LMG leads to a free energy barrier lower by ~6.5 kcal mol^{-1} than those for tail access by DSPG and DLPE (Figure 25A, B, D), which is concordant with the activity profile of PlaF (23).

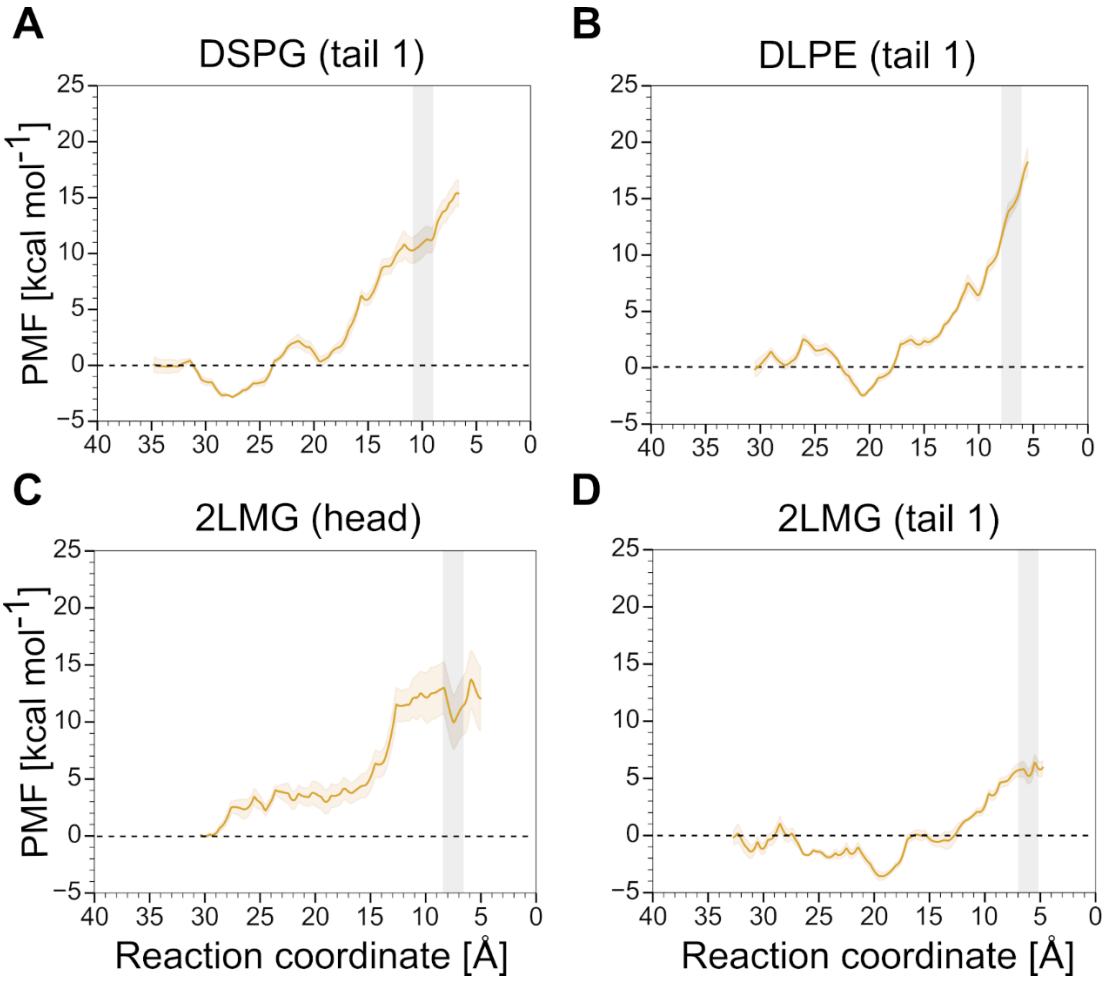

Figure 25: PMF profiles for other substrates across T2. Four systems were investigated to reveal the energetics of DSPG access via tail 1 (A), DLPE via tail 1 (B), 2LMG via head (C), and 2LMG via tail 1 (D). Among these substrates, access of 2LMG via tail 1 has the lowest free-energy barrier. The catalytic site is marked with a gray box.

As a conclusion from this result, the tail 1 access in T2 of GPL substrates with longer acyl chains or neutral head groups is disfavorable compared to DLPG access, in line with PlaF's substrate specificity. For the LGPL substrate 2LMG, tail 1 access is also favored over head access and more favorable than DSPG and DLPE access.

5.6 Energetics of substrate access into T2 in dimeric PlaF

In di-PlaF, the orientation of the tunnels with respect to the membrane changes (Figure 26A), and the tunnel entrances are higher above the membrane interface. In this configuration, T2 is closest of all tunnels to the membrane interface with a distance of 7.4 ± 1.5 Å (Figure 26A). Therefore, I have computed

the PMF for substrate access across T2 in di-PlaF. I followed the same steps considered for the substrate access in monomeric PlaF (t-PlaF$_A$) (see above). The tail 1 access of DLPG in di-PlaF revealed a free energy barrier of 13 kcal mol^{-1} (Figure 26B), compared to no energy barrier in t-PlaF$_A$ (Figure 21B). The PMF was found converged after 300 ns yielding a maximal difference of ~0.5 kcal mol^{-1} as to a PMF computed from 280 ns per window (Figure 26C), and neighboring umbrella windows have a sufficient median overlap of 4.2% (Figure 26C). These results indicate that substrate access via tail 1 across T2 is disfavorable in di-PlaF compared to that in t-PlaF, and may explain why PlaF is inactive in the dimeric configuration (23).

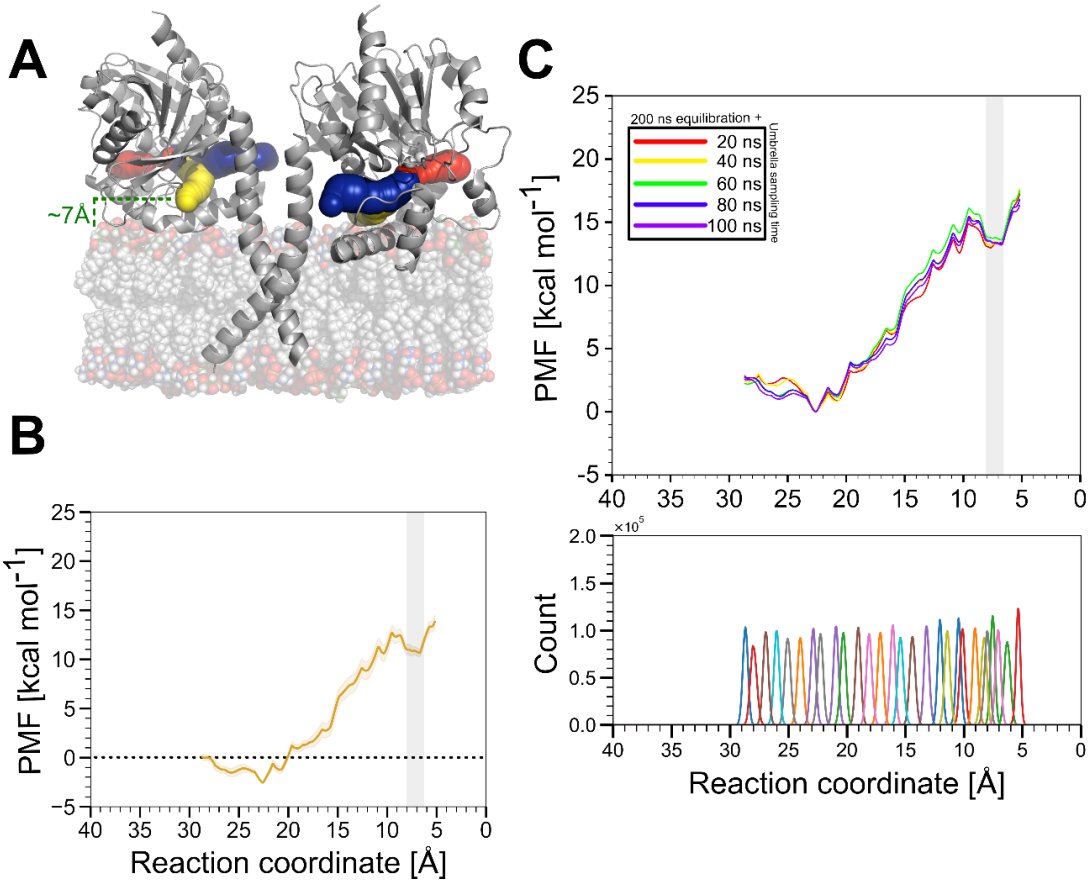

Figure 26: Substrate access in dimeric PlaF. A) Among the three tunnels, T2 (yellow spheres) is closest to and its entrance situated at ~7 Å from the membrane interface. The entrances of T1 (blue sphere) and T3 (red sphere) are much farther away (≥ ~12 Å) from the membrane interface, making substrate access into them energetically unfavorable in di-PlaF. B) The PMF of DLPG access via tail 1 across T2 shows a barrier height of 13 kcal mol^{-1} on approaching the catalytic site (gray box). C) Convergence plot (top) indicates sufficient sampling time; the profile converges at 300 ns yielding a maximum difference of ~0.5 kcal mol^{-1} compared to the PMF computed at

280 ns. The histograms (bottom) indicate sufficient overlap among the umbrella windows using a force constant of 5 kcal mol^{-1} $Å^{-2}$; the median overlap is 4.2%.

As a conclusion from this result, PlaF is inactive in the di-PlaF configuration, because the substrate from the membrane has to pass through the solvent by ~7 Å to enter into the T2. The T1 and T3 are much farther from the membrane interface than T2.

5.7 Computational costs to determine PMFs related to substrate access

Considering the different substrates and their modes of access across the tunnels T1 and T2, there are 11 systems for which the PMFs were calculated using US simulations. Depending on the properties of the substrates, their position in the membrane, their access mode, and the tunnel, each system required a different number of umbrella windows and, hence, different amounts of computation costs. In total, all computations add up to ~104 µs of sampling simulations for the substrates investigated in this study (Table 6).

Table 6: Setup of umbrella sampling simulations for the substrate access through T1 and T2.

Substrate	Tunnel	Mode of access	No. of windows	Sampling length[a]	Total sampling length[a]
DLPG[b]	1	head	28	0.3	8.4
DLPG[b]	1	tail 1	32	0.3	9.6
DLPG[b]	1	tail 2	35	0.3	10.5
DLPG[b]	2	head	34	0.3	10.2
DLPG[b]	2	tail 1	37	0.3	11.1
DLPG[b]	2	tail 2	34	0.3	10.2
DSPG[b]	2	tail 1	32	0.3	9.6
DLPE[b]	2	tail 1	26	0.3	7.8
2LMG[b]	2	head	28	0.3	8.4
2LMG[b]	2	tail 1	32	0.3	9.6
DLPG[c]	2	tail 1	28	0.3	8.4

[a] In µs.

[b] Considering t-$PlaF_A$.

[c] Considering di-PlaF.

5.8 Tryptophan substitutions in T2 hamper DLPG access

To validate the prediction that T2 is the preferred access pathway, I have identified residue positions in all identified tunnels that, when substituted with

tryptophan (Trp), should constrict the tunnel and, thus, block substrate access. Earlier, this strategy has been used to block tunnels of a dehalogenase and influence its activity by limiting the rate of product release (128). In the case of PlaF, the products are less bulky than the substrates, such that product release should be less impacted than substrate access due to constricted tunnels.

PlaF variants were predicted subject to minimizing the structural destabilization due to the Trp substitution and preferring sites within the tunnels that influence its geometric characteristics (Table 7). I have predicted four Trp substitutions for T1 and five for T2 and T3 each (Table 7). With any one of these substitutions in place, the impacted tunnel could not be identified anymore by CAVER applying the previously used probe radius of 2 Å, but with a smaller probe radius of 1.2 Å (Figure 27). This indicates their constriction, also displayed by the time evolution of the tunnel profiles of the PlaF variants compared to $PlaF_{WT}$ (Figure 28).

Table 7: Structural stability of proposed tunnel variants of PlaF determined using FoldX and corresponding influence on tunnel characteristics calculated with CAVER.

Tunnel	PlaF variant	ΔΔG[a,b]	Average bottleneck radius[c,d]	Average length[d]
T1	N77W	-0.48	2.21	27.40
	R80W	0.68	2.42	26.08
	L214W	1.13	2.15	26.84
	V290W	-0.65	1.83	26.06
T2	D74W	-0.69	1.63	30.85
	R217W	-0.10	1.87	24.30
	A218W	0.40	1.86	23.67
	A221W	-0.11	1.82	28.26
	N225W	-1.20	1.80	29.77
T3	M166W	0.15	2.11	14.72
	L177W	0.48	2.24	14.25
	F229W	-0.28	2.23	14.18
	R233W	-0.50	2.16	18.54
	Y236W	-0.18	2.10	13.82

[a] $\Delta\Delta G = \Delta G_{variant} - \Delta G_{wild\ type}$.

[b] In kcal mol^{-1}.

[c] Data calculated with a probe radius of 1.2 Å.

[d] In Å.

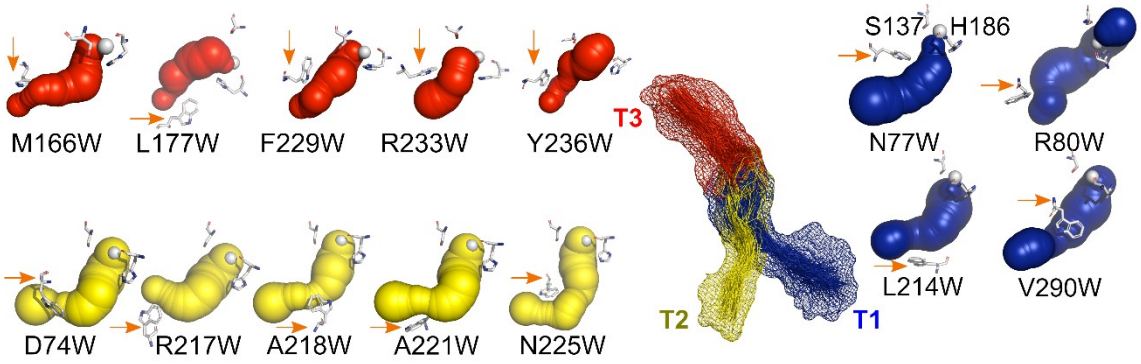

Figure 27: Influence of tryptophan substitutions on the radius of PlaF tunnels. The tunnels T1 (blue), T2 (yellow), and T3 (red) are identified by CAVER with a reduced probe radius of 1.2 Å, instead of 2 Å used otherwise (Figure 10), showing that tryptophan substitutions (orange arrows) narrow the tunnels. White spheres, wherever visible, represent the origin of the search defined by the COM of the catalytic residues S137 and H286.

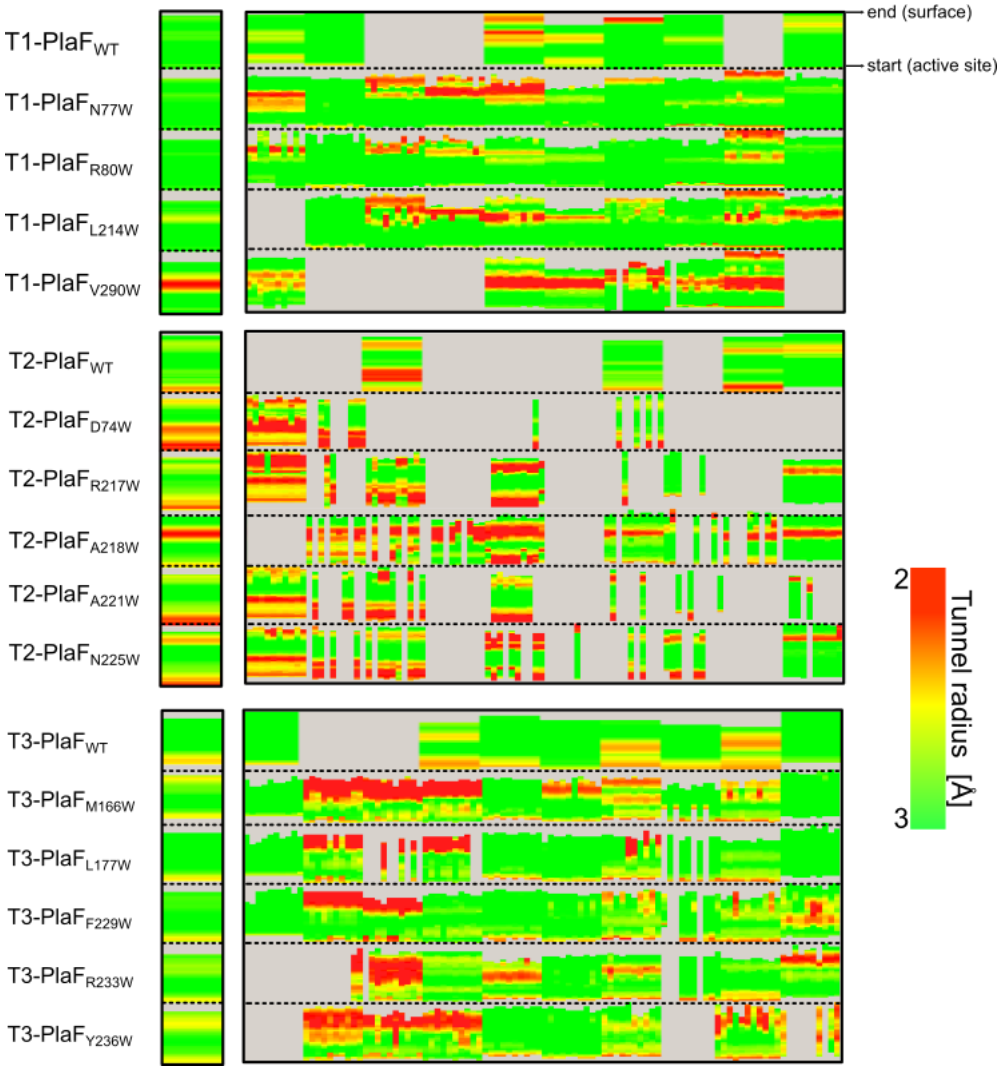

Figure 28: Heat map visualizing the time evolution of the tunnel profile for the proposed PlaF variants with tryptophan substitutions in T1-T3. The first row represents the tunnel profile

for $PlaF_{WT}$, evaluated from 10 snapshots, obtained at every 200 ns of 2 μs long unbiased MD simulations. Corresponding PlaF variants (second row onwards) were modeled 10 times for each of the 10 snapshots, resulting in 100 snapshots. (A) Average profile for each variant. (B) Time evolution of each variant, with each column corresponding to one snapshot. For $PlaF_{WT}$, continuous snapshots correspond to the increasing time scale of 2 μs in steps of 200 ns. For other variants, every 10 snapshots represent a block of 10 individual profiles of models, obtained from a single snapshot of $PlaF_{WT}$. Each block appears in the increasing time scale of 2 μs. A gray column indicates that the given tunnel was not identified in that particular snapshot. The color scale depicts the tunnel radius.

The above described fourteen Trp variants were expressed, purified, and evaluated for activity at the Jaeger lab. The specific activity of each PlaF variant was compared with that of $PlaF_{WT}$ by measuring the hydrolysis of small (*p*-nitrophenyl butyrate, *p*-NPB) and large (DLPG) substrates (Figure 29A). Interestingly, all five T2 variants had a significantly lower activity with *p*-NPB and DLPG than $PlaF_{WT}$. In contrast, the activities of T1 and T3 variants with both *p*-NPB and DLPG were similar to that of $PlaF_{WT}$ (Figure 29A). In another set of experiments at the Jaeger lab, thermostability of the PlaF variants was determined. Among all the variants, none showed a drastically reduced stability (Figure 29B). However, two T2 variants were more stable than the $PlaF_{WT}$, and three variants were slightly less stable (2.7 - 4.2ºC), confirming that the proposed Trp substitutions do not largely impact the thermostability of PlaF. The observation that PlaF activities with DLPG and *p*-NPB predominantly decreased with Trp substitutions in T2 and that the DLPG activities decreased more (53 – 68%) than *p*-NPB activities (16 – 31%) (Figure 29A) indicate that the bulky Trp substitutions impact passage through T2 (Figure 29C).

As a conclusion, the biochemical studies indicate that among the proposed Trp substitutions introduced in the three tunnels, only those in T2 reduced lipolytic activity of PlaF. These results confirm that T2 is the main route for substrate access from the membrane to the catalytic site.

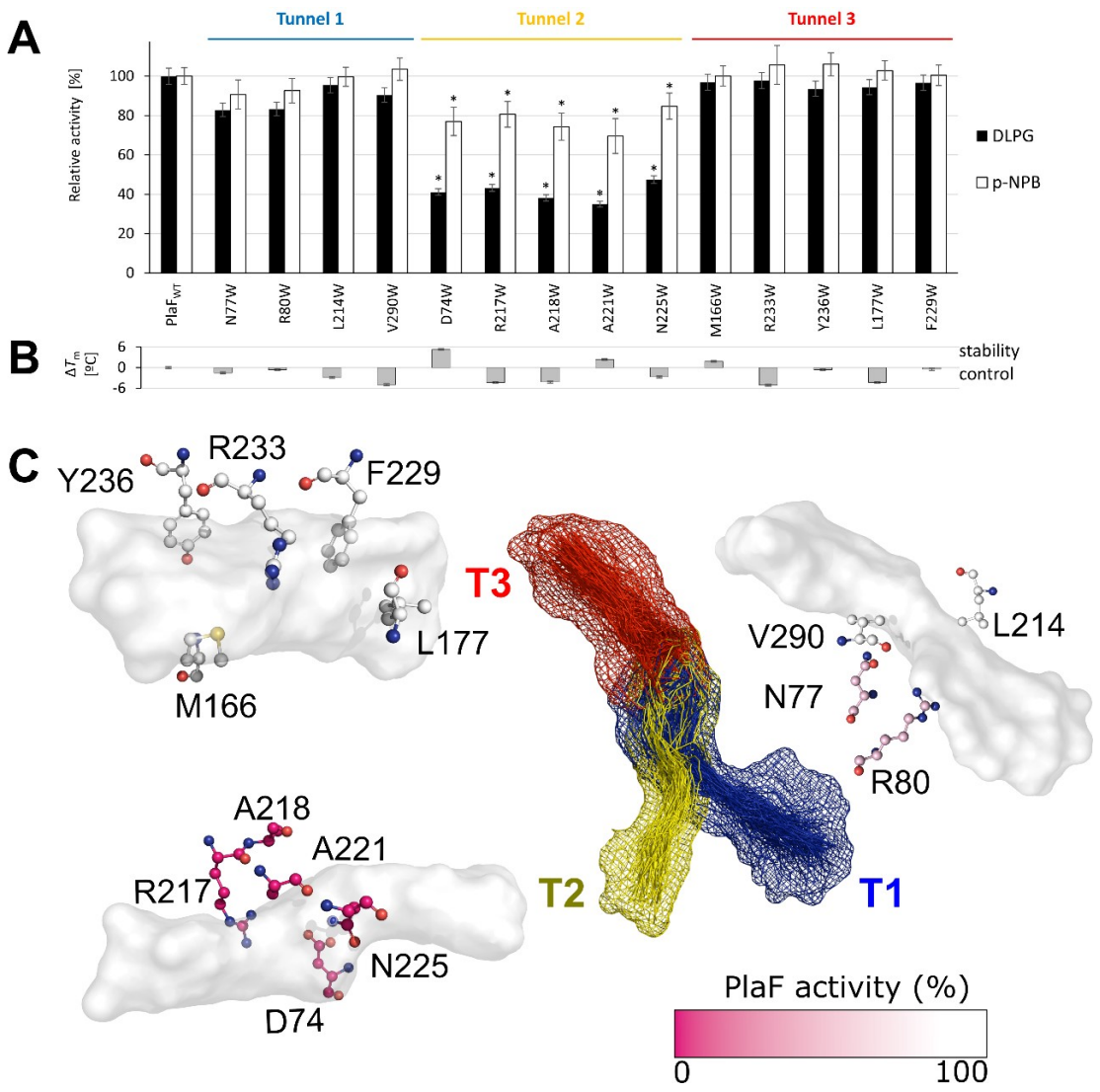

Figure 29: Lipolytic activity of PlaF and variants with Trp substitutions in T1-T3. A) Enzyme activities of purified PlaF$_{WT}$ and variants carrying respective substitutions measured with DLPG and *p*-NPB. Activities are normalized to the activity of PlaF$_{WT}$, which was set as 100%. Results are means ± standard deviation of three independent measurements. Statistical analysis of was performed using the T-test (* $p < 0.001$) of normally distributed values for DLPG ($n = 8$) and *p*-NPB ($n = 9$) measurements. B) The thermal stabilities of purified PlaF$_{WT}$ and variants were measured. Results are shown as a difference in the melting temperatures (ΔT_m) of the respective PlaF variant and PlaF$_{WT}$, which was 57.3 ± 0.2 ºC. Results are means ± standard deviation of three independent measurements, each performed with three samples. C) The tunnels, T1 - T3 (mesh view in the center) are represented as white surfaces. The investigated amino acids are shown in ball-and-stick representation. Corey-Pauling-Koltun (CPK) coloring scheme was used to color all the atoms of residue, except for carbon atoms, which vary from pink to white and relates to the PlaF activity for DLPG after substituting the corresponding residue for a tryptophan. The PlaF activity is reduced the most if Trp substitutions involve T2. The experimental data and corresponding figures A, B were provided by Jaeger lab at Forschungszentrum Jülich, Germany.

5.9 Potential egress pathways of PlaF products

Next, I aimed at identifying potential egress pathways for products of PlaF-catalyzed hydrolysis. I have performed a set of unbiased MD simulations starting from a hydrolyzed 2LMG in t-PlaF$_A$. The starting coordinates were taken from the last snapshot of the US simulations of 2LMG with tail 1 access through T2, considering the umbrella window where the *sn*-1 position of 2LMG was closest to the catalytic site. Then, 2LMG was cleaved into the respective products without changing their orientation in the tunnels (Figure 30A). This led to MYR being in T3 at the beginning of the simulations and the PGR (phosphatidylglycerol from LGPL, 2LMG) moiety pointing toward T2 (Figure 30B).

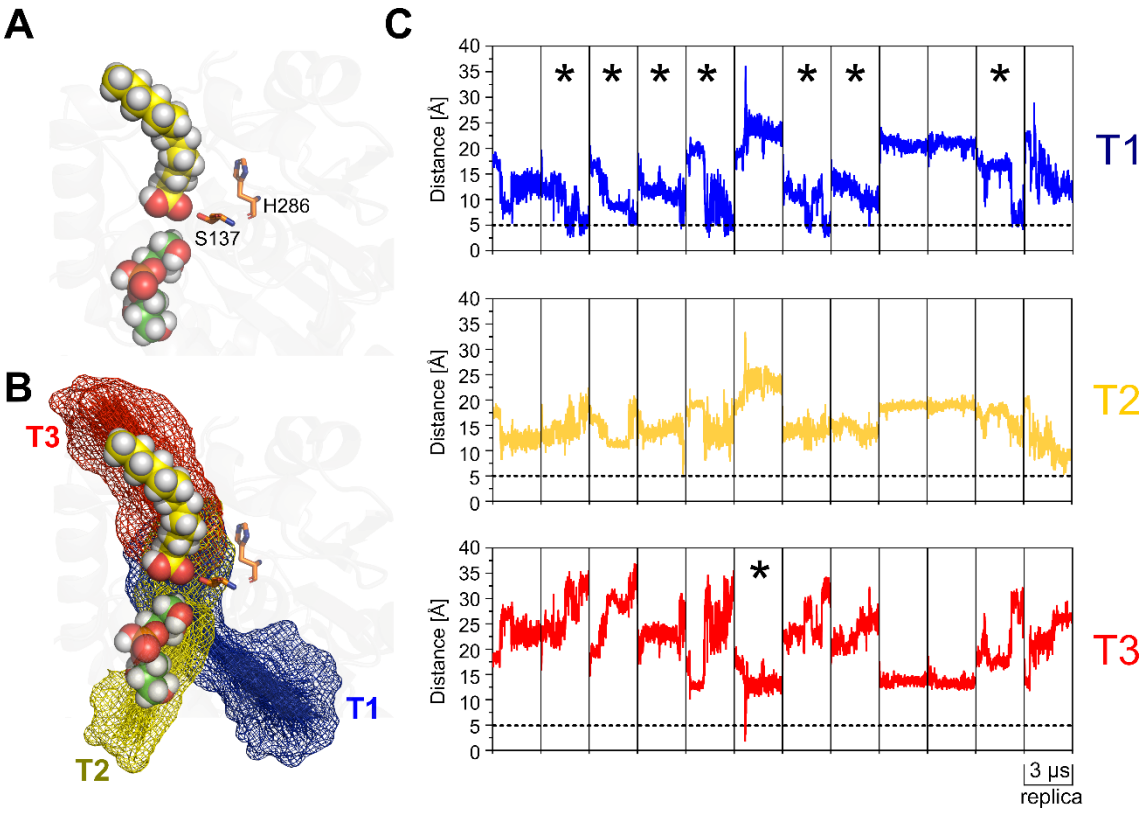

Figure 30: Unbiased MD simulations of t-PlaF$_A$ with bound hydrolysis products. (A) Starting configuration of the 2LMG products in t-PlaF$_A$. MYR is represented with yellow spheres, and PGR with green spheres. The catalytic S137, and H286 are shown as orange sticks. (B) The products are mapped over the respective tunnels. (C) The distance of MYR to the entrances of T1-T3 during 12 replicas of unbiased MD simulations of 3 µs. The dashed black line depicts the chosen cutoff of 5 Å, with replicas that reach this cutoff marked with an asterisk. MYR reaches a distance ≤ 5 Å to the entrance of T1 in 7 replicas, in 1 replica for T3, and in none for T2. Note that in chain A of the PlaF crystal structure, MYR is found in T1.

Results

In 12 replicas of 3 µs length each, the products relocated within the tunnels, sometimes even diffusing into the solvent (PGR moiety in 2/12 replicas via T1 and 3/12 replicas via T2; Figure 31). MYR relocated from its original position in T3 and approached the other tunnels of PlaF during the course of the MD simulations (Figure 30C). To deduce the displacement of MYR, I measured the distance of the carboxyl carbon to the entrance of each tunnel. A cutoff of 5 Å, according to previous studies (177-179), was used to identify those replicas where MYR reaches close to the tunnel entrance. MYR moved in 7/12 replicas to the entrance of T1 and in 1/12 replicas to the entrance of T3; the entrance of T2 was not reached (Figure 30C). Interestingly, the instance of MYR reaching the entrance of T3 flips within T3 such that the carboxyl group points to the entrance, rather than to the active site as after hydrolysis (Figure 30B). Altogether, MYR reaches the entrance of T1 significantly more frequently than T2 (p = 0.0008), and T3 (p = 0.0047) (Figure 32, and Table 8).

Table 8: Statistical test[a] to determine the tendency of MYR reaching the entrance of tunnels T1-T3 in 12 replicas.

Tunnels	T3	T2
T1	2.5981 (p = 0.0047)	3.1436 (p = 0.0008)
T2	-1.0215 (p = 0.1539)	

[a] The z-score for two population proportions related to two tunnels was calculated (180-182). A cutoff of 5 Å was chosen to identify the tunnels where MYR reaches the entrance during 12 independent replicas of 3 µs long unbiased MD simulations. In 7 replicas MYR reaches T1, in 1 replica T3, but it does not reach the T2 entrance (Figure 32). The tendency of MYR reaching the entrance of T1 is significantly higher (at $p < 0.05$, considering a one-tailed z-score test) than reaching the entrance of T2 or T3.

As a conclusion from this result, hydrolysis products of 2LMG diffuse within PlaF during time scales of 3 µs, sometimes also between tunnels. T1 and, to a lower extent, T3 are the most likely egress pathways of FAs from PlaF, although more sampling is required to observe actual egress.

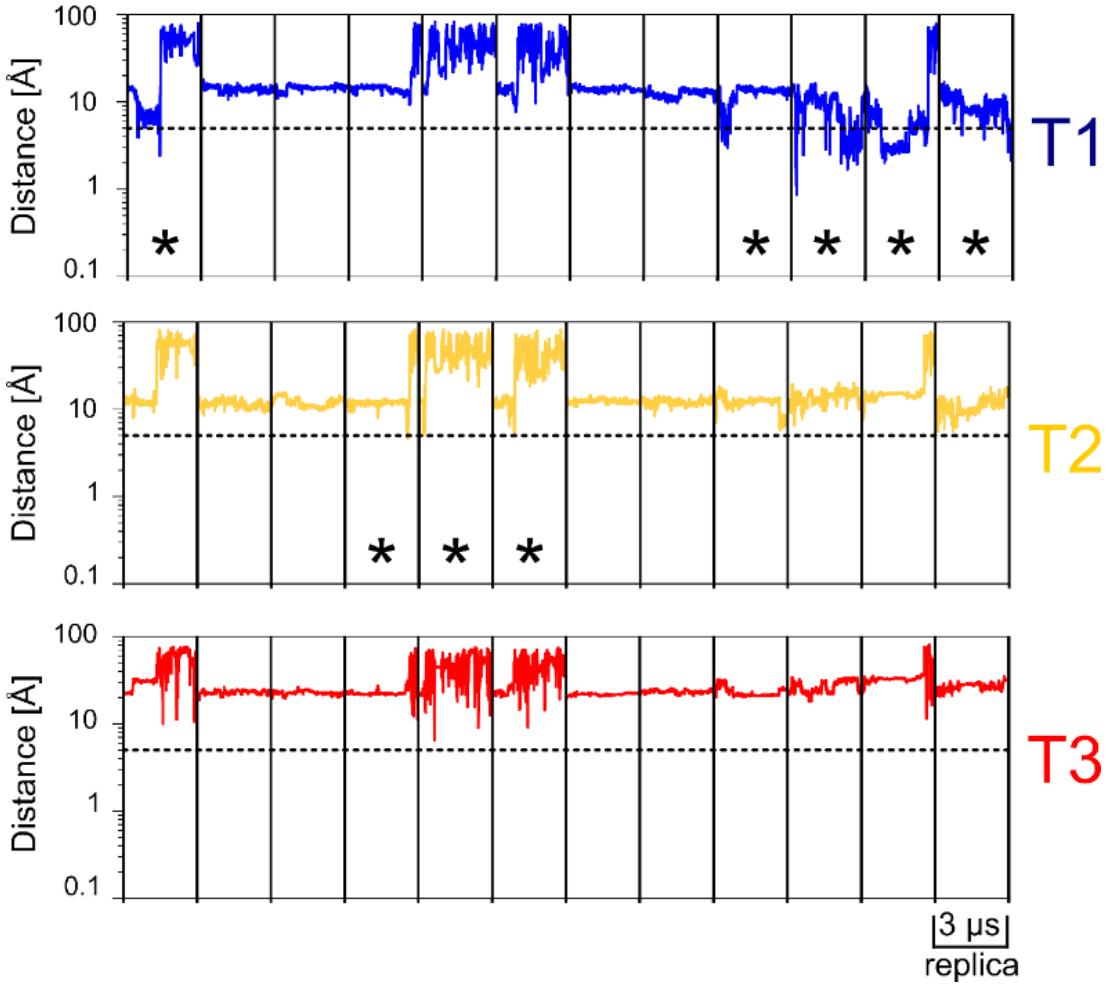

Figure 31: Unbiased MD simulations of hydrolysis products inside of t-PlaF$_A$. The distance (in $\log_{10}$ scale) of PGR to the entrance of each of T1-T3 during 12 replicas of 3 µs unbiased MD simulations is plotted, considering the phosphorous atom of PGR. The dashed black line depicts the chosen cutoff of 5 Å, with replicas that reach this cutoff marked with an asterisk. PGR reaches a distance ≤ 5 Å to the entrance of T1 in 5 replicas, including 2 replicas where PGR ultimately leaves T1 to enter into solvent. In 3 replicas, PGR comes close to the T2 entrance and exits into solvent.

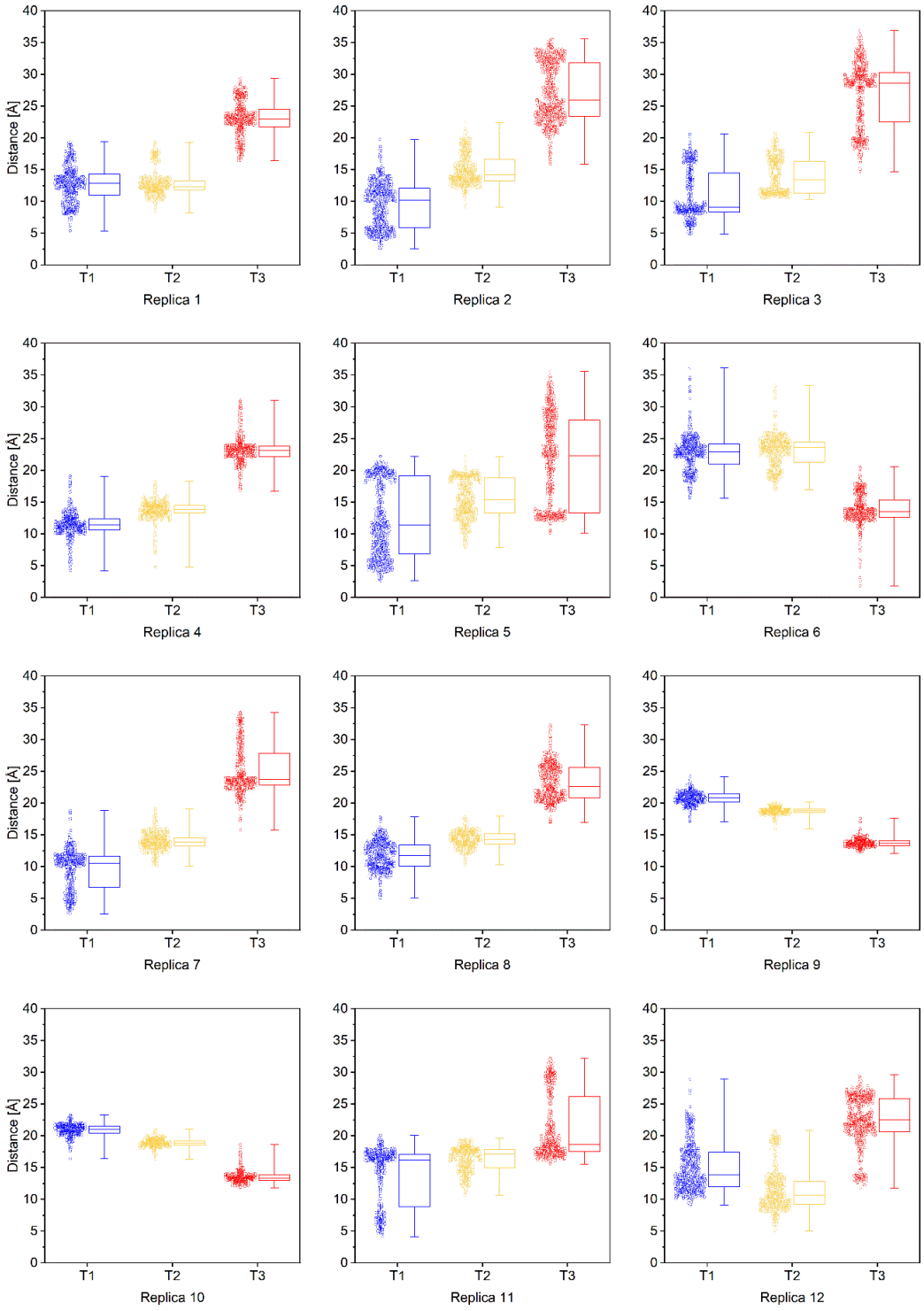

Figure 32: Unbiased MD simulations of hydrolysis products within t-PlaF$_A$. For each replica, three box plots represent the distance of MYR to the entrance of tunnels, T1 (blue), T2 (yellow), and T3 (red), during 3 µs long simulations. The corresponding distribution of data is plotted on the left of each box. Particularly in replica 2, 3, 4, 5, 7, 8, and 11, MYR comes within ~5 Å distance to the entrance of T1, compared to replica 6 where it reaches T3. MYR does not reach T2 at the selected cutoff of 5 Å. These observations indicate that the tendency of MYR reaching the entrance of T1 is significantly higher than for the other two tunnels (Table 8) and likely the egress route for fatty acid products from t-PlaF$_A$.

6. Discussion

This chapter was taken from the following peer-reviewed publication:

"Substrate Access Mechanism in a Novel Membrane-Bound Phospholipase A of *Pseudomonas aeruginosa* Concordant with Specificity and Regioselectivity" Sabahuddin Ahmad, Christoph Heinrich Strunk, Stephan N. Schott-Verdugo, Karl-Erich Jaeger, Filip Kovacic, and Holger Gohlke

Journal of Chemical Information and Modeling, 2021, 61 (11), 5626-5643
DOI: 10.1021/acs.jcim.1c00973

Reprinted with permission from the Journal of Chemical Information and Modeling. Copyright © 2021, American Chemical Society.

Dimer-monomer transitions regulate the activity of several membrane-bound phospholipases, including PLA_1, and PLA_2 (183-190). Previously, it was shown that PlaF becomes active due to a dimer-to-monomer transition followed by tilting of the monomer in the membrane, resulting in t-$PlaF_A$ being the active configuration of PlaF (23). Here, I have addressed the questions of how membrane-bound substrates reach the active site of $PlaF_A$ and how the characteristics of the active site tunnels determine the activity, specificity, and regioselectivity of PlaF for medium-chain substrates. I performed unbiased and biased MD simulations and showed by configurational free energy computations and mutational and enzymatic studies for t-$PlaF_A$ that A) access of the two main PlaF substrates DLPG and 2LMG occurs most likely through tunnel T2 in a tail first mode, B) access of substrates with longer acyl chains or neutral head groups is less favorable, C) tail 1 access of DLPG and 2MLG in T2 is more favorable than tail 2 access, D) T3 accommodates the substrate tail to be hydrolyzed, and E) T1 and T3 are potential product egress pathways.

Previous studies indicated that the characteristics of substrate access tunnels can have a decisive influence on enzyme-substrate specificity and activity (126, 131, 132, 191). In t-$PlaF_A$, I focused on T1 and T2 because only these two allow

direct access of GPL or LGPL substrates from the membrane in the t-PlaF$_A$ configuration. By contrast, to enter into T3, substrates would need to pass through the solvent, which is energetically unfavorable. In di-PlaF, T2 is closest to the membrane with a distance of 7.4 ± 1.5 Å but T1 and T3 are at a distance >12 Å (Figure 26A). Hence, I also investigated substrate access to T2 in di-PlaF.

For assessing the energetics of substrate access, first, I generated 18 pathways, considering GPL and LGPL as substrates in T1 and T2 using sMD simulations. By relating the work along the reaction coordinate to the free energy difference between two states of the pulling simulations via Jarzynski's relation and considering the endpoint of the sMD trajectory closest to the Jarzynski average as the starting point for the next sMD simulation, I obtained low free energy pathways of substrate access to the catalytic site. sMD simulations have been widely used to explore similar biological processes such as the loading of GPL substrates into human phospholipase A_2 (PLA_2) (159) or recognition of arachidonic acid by cytochrome P450 2E1 across the access channel (192). The pathways served for defining reference points for subsequent US simulations, such that distributions of sampled states sufficiently overlapped, which is essential to yield accurate results in PMF computations (193). Applying US along pathways identified by sMD simulations (115) or targeted simulations (116, 194) has been shown to be an effective method of computing PMF. Moreover, the choice of an appropriate reaction coordinate is essential for this approach (195-197). I probed for the convergence of the PMFs by comparing PMFs generated from increasing lengths of US simulations and found that US times of ~300 ns are needed to yield PMF differences below chemical accuracy (118). Finally, the PMF computations were validated by comparing the computed absolute binding free energy of DLPG to PlaF for the most preferred access mode to an estimate of the experimental binding free energy.

The PMFs revealed that tail first access through T2 is most preferred for

DLPG and 2LMG. This finding is in line with the geometric analysis of T2, which revealed a tunnel bottleneck radius about half as large as the radius of DLPG deduced from the lipid's APL, which can explain why a headgroup-first access is disfavorable for steric reasons. Furthermore, I have shown that acyl chains of lipids embedded in a membrane can reach the interface region in unbiased MD simulations and, thus, can interact with the tunnel entrance. Such protrusions of lipid tails occur on a timescale of approximately 100 ns depending on the extent of solvent exposure (198). Tail first access of GPLs into the active site has also been found for Cyclopropane Fatty Acid Synthase (199). Tail first access through T2 is favored because of the predominant hydrophobic nature of the tunnel walls. By contrast, T1 contains a higher number of charged Asp and Arg residues and fewer neutral residues than T2, which makes tail first access there less favorable. In particular, the side chain of R80 protrudes into T1 at the tunnel entrance, which is reflected in an energy barrier of ~3 kcal mol^{-1} found there for tail first access.

Modifications in tunnels that connect a buried active site to the bulk solvent have been shown to affect ligand binding and unbinding (128). Tunnel residues situated away from the active site are suitable targets for mutagenesis, as their replacement should not lead to a loss of the functionality of the active site (129). Considering this, I introduced Trp substitutions to each of the three tunnels of PlaF and measured the activity of these PlaF variants. The Trp substitutions decreased PlaF's lipolytic activity for small and large substrates only when introduced in T2, which suggests that T2 is involved in substrate access. However, from such steady-state experiments, it cannot be excluded that the Trp substitutions influence product egress, too (126).

Among the investigated substrates, higher energy barriers for access to the active site were found for those with longer acyl chains and neutral head groups, concordant with PlaF's activity profile (23). This finding may be explained with differences in the energetics of GPL self-assembly, which is influenced by the hydrocarbon chain length and the polarity of the head group (168): Longer

hydrocarbon chains and less polar head groups foster self-assembly, which would lead to higher energy barriers for leaving this equilibrium state (124) and entering into PlaF. These results indicate that the energetics of access of a membrane GPL substrate to the active site through tunnel T2 contributes to the substrate specificity of PlaF.

Furthermore, of the two constitutopic acyl chains in DLPG, access via tail 1 in T2 is energetically preferred over tail 2 access. If tail 1 enters first, the carbonyl oxy group at C1 of the glycero moiety can come closer to the nucleophilic S137 than if tail 2 enters first (Figure 22), leading to preferential hydrolysis of the carboxylic ester bond at C1. Likewise, the regioselectivity of human 5-lipoxygenase is determined by the head/tail first type orientation of its main substrate arachidonic acid in the active site (200): The arachidonic acid can be positioned in the holoenzyme active site with both head first and tail first orientation, but only the tail first orientation results in a configuration that yields 5-lipoxygenating activity. These results indicate that the tail first access mode of a diacyl GPL substrate determines the regioselectivity of PlaF for hydrolysis of the acyl chain bound to the *sn*-1 position.

As T3 is oriented to the membrane neither in the monomeric nor in the di-PlaF configuration, it likely does not contribute to substrate access. The suggested role of T3 is to accommodate the acyl chain of substrates before and products after hydrolysis. T3, with a length of ~15 Å, provides adequate space for substrates with medium-lengths of acyl chains and, thus, may affect the specificity of PlaF. Substrate tunnels that accommodate acyl chains hydrolyzed from their respective precursors have also been described for cholesterol acyltransferases (201). Likewise, lipid phosphate phosphatases harbor such a cavity, accommodating the substrate's acyl chain for optimal catalysis (202). Site-directed mutagenesis in *Candida rugosa* lipase 1 revealed the role of such tunnels in determining the acyl chain length specificity (127).

As to di-PlaF, tail 1 access of DLPG across T2 revealed a free energy barrier of ~13 kcal mol^{-1} (Figure 26B), in contrast to no free energy barrier in t-$PlaF_A$ (Figure 21B). This high barrier may arise because of the location of T2 in di-PlaF, ~7 Å above the membrane. Thus, substrates would need to pass through the solvent to enter T2. These findings indicate that di-PlaF is catalytically inactive, as determined experimentally (23), because of energetically unfavorable substrate access.

The results from unbiased MD simulations of products suggest that T1 and, to a lower extent, T3 are egress pathways of FAs. As to T1, this suggestion is in agreement with the crystal structure of PlaF, where FAs are found in T1 (23). In the tilted orientation of PlaF, FAs egressing via T1 would interact with the membrane interface and could diffuse into it. FAs in a membrane can affect its fluidity and permeability and protein-lipid interactions, thereby regulating important cell processes including signal transduction, motility, and biofilm formation (203, 204). Via T3, they would egress into the periplasmic space. Anchored to the cytoplasmic membrane, PlaF is not a toxin targeting the host cell membrane but it has a direct influence on virulence adaptation of *P. aeruginosa* by modulating the membrane GPL composition (23). However, it is unknown if FAs released from GPLs by PlaF are targeted to the external environment as for example diffusible FAs involved in cell-to-cell signaling (20, 21). In this case, egress of FAs *via* T3 to the periplasm and their further passive diffusion or active transport would be possible (205).

7. Conclusion and significance

PlaF is a novel integral inner membrane protein from *P. aeruginosa*, that shows PLA_1 activity, and preferably hydrolyzes endogenous GPLs with medium-length acyl chains; the activity of PlaF is regulated by the dimer-to-monomer transition. For this thesis, I have investigated with the help of (un)biased MD simulations, and free energy computations, the preferred access modes and routes for the GPL substrates into PlaF. The computations reveal a substrate access mechanism in good agreement with the experimental studies.

The most relevant results are:

1. The deeply buried active site of PlaF is connected to the surface, primarily with three tunnels T1-T3. Among these, only T1 and T2 are in direct contact with the membrane interface in the tilted configuration, while T3 opens in the periplasmic space.
2. Free energy computations indicated that T2 is the preferred route for substrate access from the membrane. This was confirmed by site-directed mutagenesis experiments, where Trp substitution within the tunnels T1 and T3 do not influence the activity. In contrast, such substitutions in T2 grossly impact and reduce the lipolytic activity of PlaF by ~70%.
3. Among different modes of access of GPL substrates, the access with tail 1 has almost no energy barrier to the active site across T2 and is hence deemed to be the most favorable access mode of the substrate. This is in agreement with PlaF being PLA_1, as it hydrolyzes its substrates at position *sn*-1. The neutral, non-polar nature of residues lining the T2 favor the tail first access into PlaF.
4. T3 spans ~15 Å and has an important role in the activity of PlaF by accommodating the medium-length acyl chains. Together with T2, T3 determines the specificity and regioselectivity of PlaF substrates.
5. Unbiased simulations at the time scale of 3 μs indicate that FA products likely

egress from T1. Interestingly, the X-ray crystal structure of PlaF harbors endogenous FAs in T1.

The mechanism of the substrate access into PlaF and the activity regulation has been summarized in the schematic model below (Figure 33). The described substrate access mechanism suggests an important role of the tunnels in determining the activity, specificity, and regioselectivity of PlaF for its substrates. However, it remains to be characterized how the PlaF-produced FAs regulate the PlaF activity by inducing dimer-to-monomer transitions, and how do these FAs reach the external environment.

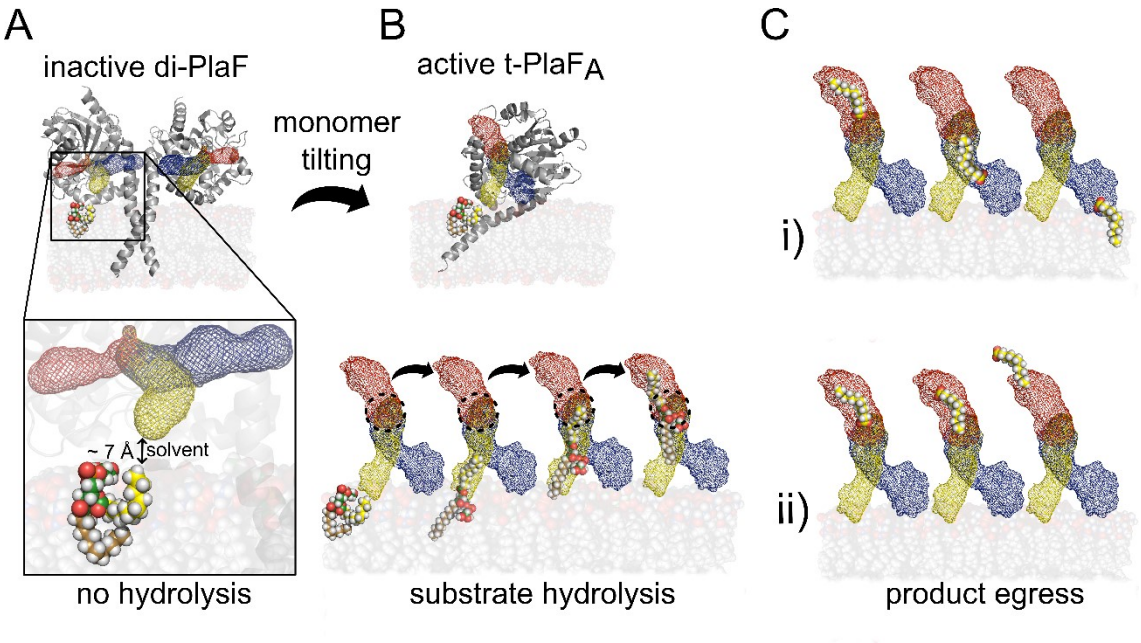

Figure 33: Schematic model of the mechanism of PlaF activity regulation. A) A higher concentration of PlaF results in the formation of an inactive dimer (23). In di-PlaF, T2 is closer to the membrane interface than the other tunnels at ~7 Å, which requires the substrate to pass through the water solvent. Hence, this configuration leads to inactive PlaF. B) At low concentrations, monomeric PlaF shows PLA_1 activity and adopts a tilted configuration (23). In the tilted configuration, PlaF orients such that T1 and T2 come close to the membrane interface. Substrate access occurs *via* tail 1 into T2. The acyl chain reaches the active site (dashed black circle) of PlaF, but the *sn*-1 cleavage site is still away from the active site. Further loading of the acyl chain requires it to enter into T3, and the substrate is hydrolyzed. C) After hydrolysis, the resulting FA is in T3. i) Now, either the FA relocates into T1, with its carboxyl group toward the entrance, where it can interact with the membrane interface and diffuse into it. ii) Alternatively, the FA can flip around, such that the carboxyl group faces the T3 entrance, from where it can exit into the periplasmic space.

8. Perspectives

In this thesis, I described a potential mechanism of substrate access in the phospholipase PlaF from *P. aeruginosa*. However, there are still a few points that remain open and need further investigation.

First, from the available steady state kinetics experiments, it remains unknown if the role of T2 is explicit for the substrate access or if it also has a role in product egress. The PlaF products are less bulky than the substrates, and therefore the constrictions from Trp substitutions should not grossly impact the product's egress. Yet, pre-steady state kinetics experiments is worth considering to dissect the explicit role of T2 in substrate access.

At the available time scale of product egress simulations, the actual egress route of the PlaF products is not clear; therefore, an extension of these simulations is required. Moreover, the simulations in this study represent only the products from LGPL substrates, while PlaF shows PLA_1 activity for both LGPL as well as GPL substrates. Therefore, it is of great interest to set up simulations to understand the egress of products from GPL substrates.

Available data from the product egress simulations suggest that FAs egress through T1 and to a lesser extent T3. Via T3, FAs can exit into the periplasmic space. However, if the exit route is T1, FAs can diffuse in the membrane and influence its integrity by altering the composition. Therefore, it is worth investigating, if this change has an impact on the energetics of PlaF dimer-to-monomer transition and the tilted state of the monomer.

At last, PlaF is anchored to the cytoplasmic membrane of *P. aeruginosa*, with no direct contact to target the host cell membrane. Yet, diffusible FAs can participate in cell-to-cell signaling and biofilm formation by migrating to the external environment. For this, FAs can egress via T3 into the periplasmic space, from where they can be further transported across the membrane. However, the exact mechanism involved remains unknown.

Perspectives

The atomistic-level findings of this study provide an understanding of the unique structural features, showing that PlaF's function is dependent on monomerization followed by a global reorientation of the single-pass TM protein at the membrane. These results open up opportunities for developing drugs that inhibit PlaF and potentially decrease *P. aeruginosa* virulence during infections.

9. Acknowledgments

I sincerely express my gratitude to Prof. Dr. Holger Gohlke for his generous supervision, continuous support, guidance throughout my doctorate work. I will admit, such supervision, and weekly topic group meetings have built-in me better scientific research capabilities.

I would like to thank my co-supervisor, Prof. Dr. Karl-Erich Jaeger, his team leader Dr. Filip Kovačić for fruitful scientific discussions about the PlaF project. In particular, I appreciate the enthusiastic collaboration by Dr. Christoph Heinrich Strunk. Thanks to CRC1208, for providing a necessary platform in terms of meetings, conferences, and workshops that built my academic profile during my doctoral studies. Many thanks to JUNO-HHU, especially Dr. Wegener-Feldbrügge, and Han Dang-Klein for their support from the first day of the paper work at HHU.

Thanks to Dr. Daniel Ciupka, Dr. Markus Dick, and Michele Bonus, for spending their valuable time in setting me up in my very first days at the university. My heartiest thanks to Dr. Stephan Schott. Not only as an office partner, CRC1208 colleague, but also as an unofficial mentor, a friend; he was readily available to help me whenever I got stuck with my project, professional or personal life.

I would like to thank the group “CPClab” headed by Prof. Gohlke; CPClab is very special. It is rich with multinational, genius and expert minds. As a plus, the colleagues are both social, supportive, and they made my stay quite enjoyable. To name (in alphabetical order), Annika, Bastian, (Dr.) Benedikt, Birte, (Dr.) Christina, (Dr.) Christoph, (Dr.) Christopher, Daniel B, (Dr.) Denis, (Dr.) Fathoni, Filip, Giulia, Jesko, Jonas, Korana, Rocco, Till, (Dr.) Tobias, (Dr.) Stefanie, are very nice personalities whom I met here. Countless thanks to (Dr.) Benoit, David, (Dr.) Daniel M, Lucas, (Dr.) Neha, Nicola, Pablo, Pegah, and Susanne for sharing memorable moments outside the office hours. (Dr.) Mohan and (Dr.) Kumaresan for their unconditional support, and a feel like India, even at the CPClab. There was a perfect balance of work and social life. I am grateful to (Dr.) Christoph, Michele, (Dr.) Stephan, Lucas, for their time in going through part(s) of this thesis. Thanks to Mr. Sippel and Mrs. Hübsch, for their support at the workplace.

Last but not the least, I would like to thank my father, Prof. Dr. Rafi Ullah Ansari, my siblings (Dr.) Shahnaz, Salahuddin, Umamah, cousin (Dr.) Laeeq, my in-laws, and other relatives for their love, care, and motivation from a distance of several thousand kilometers. Special thanks to Naghma for her constant support, even in her toughest time, and to Ziauddin for bringing the change in my life.

10. Curriculum Vitae

Sabahuddin Ahmad

Contact: ahmads@hhu.de

Date of birth: August 19th,1989

Place of birth: Allahabad, India

Education:

2017 – 21	**Heinrich Heine University, Düsseldorf, Germany** (www.hhu.de) PhD student (Computational Biology)
2011 - 13	**Jamia Millia Islamia (A Central University), New Delhi, India** (www.jmi.ac.in) Master of Science (M.Sc.) in Bioinformatics
2008 - 11	**Aligarh M University, Aligarh, India** (www.amu.ac.in) Bachelor of Science (B.Sc. Hons.) in Botany

Research experience:

2021 – now	**Research Assistant** at Forschungszentrum, Jülich, Germany
2017 – 21	**Research Assistant** at Heinrich Heine University, Düsseldorf, Germany
2015 - 17	**Project Research Fellow** at CSIR-Central Drug Research Institute, Lucknow, India
2014	**Short-term Visiting Scholar** at Stritch School of Medicine, Chicago, USA

Teaching experience:

2017 - 21	**Teaching Assistant** of the practical course **"**Molecular modeling" at Department of Pharmacy, Heinrich Heine University, Düsseldorf, Germany

Award:

2014	International Summer Research Award (ISRA) was conferred by Sir Syed Educational Society of North America (SSESA). Webpage: www.ssesa.us

Curriculum Vitae

Skills, proficiency and activities:

Skills:

1. *Molecular modeling:*
   a) (macro) Molecule visualization via PyMOL, Chimera
   b) Homology modeling via MODELLER
   c) Mutant modeling of proteins via FOLDX
   d) Analyzing tunnels/cavities in proteins via CAVER

2. *Computer aided drug design, virtual screening:*
   a) Molecular docking via GLIDE (Schrödinger, Inc.), Surflex-Dock (Tripos, Inc.), AutoDock
   b) Protein-protein/peptide docking via ZDOCK (Accelrys, Inc.)
   c) Structure/Ligand-based pharmacophore model generation

3. *Molecular dynamics (MD) simulations:*
   a) (membrane) Proteins via GROMACS, and AMBER
   b) Steered MD simulations to evaluate the ligand binding/unbinding
   c) Umbrella Sampling simulations to compute potential of mean force (PMF)
   d) Visualizing, animating the MD simulation results via VMD

4. *Computer-based skills:*
   a) High-performance computing at Jülich Supercomputing Centre, Germany (2018 - 2021)
   b) Shell scripting with Bash
   c) Plotting data with Origin Pro, Xmgrace
   d) Vim (text) editor
   e) Microsoft Office (Word, PowerPoint, Excel)
   f) Computer maintenance and repairing

Languages: English (Proficient); German (A2 level); Hindi and Urdu (Native languages)

Workshops:

1. Two-day iGRAD workshops at Heinrich Heine University, Düsseldorf, Germany
   a) Career Planning in Business - How to shape your future (2019)
   b) Writing paper and Theses in the Life Sciences (2018)
   c) Get into teaching (2017); Presenting (in) Science (2017)
2. Two-day workshop in association with Schrödinger, Inc. on Molecular Docking, Virtual Screening and Computational Biology at CSIR-Central Drug Research Institute, Lucknow, India (2016)

E-courses:

1. Cyber-Security, Heinrich Heine University, Düsseldorf, Germany (2020)
2. Basic training in data protection, Heinrich Heine University, Düsseldorf, Germany (2020)

Conferences:

1. Oral presentation at 34th Molecular Modeling Workshop 2020, Erlangen, Germany (2020)
2. Poster presentation at John von Neumann Institute for Computing (NIC) Symposium, Forschungszentrum Jülich, Germany (2020, 2018)

3. Poster presentation at CRC 1208 Conference 'Dynamic of Membrane Systems', Düsseldorf, Germany (2019, 2017)

Selected list of peer-reviewed publications[§]:

1) **Ahmad S.**, Strunk C.H., Schott-Verdugo S.N., Jaeger K.-E., Kovačić F., Gohlke H.; Substrate access mechanism in a novel membrane-bound phospholipase A of *Pseudomonas aeruginosa* concordant with specificity and regioselectivity. Journal of Chemical Information and Modeling. 61 (11): 5626-5643; 2021
2) **Ahmad S.**[‡], Bhagwati S.[‡], Kumar S., Banerjee D. and Siddiqi M.I.; Molecular modeling assisted identification and biological evaluation of potent cathepsin S inhibitors; Journal of Molecular Graphics and Modelling. 96: 107512; 2020
3) Zafar A., Singh S., **Ahmad S.**, Siddiqi M.I., and Naseem I.; Interaction of C20-substituted derivative of pregnenolone acetate with copper (II) leads to ROS generation, DNA cleavage and apoptosis in cervical cancer cells: Therapeutic potential of copper chelation for cancer treatment; Bioorganic Chemistry. 87: 276-290; 2019
4) Dreyer I., Spitz O.[⊥], Kanonenberg K.[⊥], Montag K.[⊥], Handrich M.R.[⊥], **Ahmad S.**[⊥] et al.; Nutrient exchange in arbuscular mycorrhizal symbiosis from a thermodynamic point of view; New Phytologist. 222 (2): 1043-1053; 2019
5) Tandon A., Pathak M., Harioudh M.K., **Ahmad S.**, Sayeed M., et al.; A TLR4-derived non-cytotoxic, self-assembling peptide functions as a vaccine adjuvant in mice; Journal of Biological Chemistry. 293 (51): 19874-19885; 2018
6) **Ahmad S.** and Siddiqi M.I.; Molecular modeling-based insights into the selective inhibition of cathepsin S by its inhibitor; Journal of Molecular Modeling. 23 (3): 92; 2017
7) Khan S., Ahmad S., Siddiqi M.I. and Bano B.; Physico-chemical and in silico analysis of a phytocystatin purified from *Brassica juncea* cultivar RoAgro 5444; Biochemistry and Cell Biology. 94 (6): 584-596; 2016
8) Zafar A.[‡], **Ahmad S.**[‡] and Naseem I.; Insight into the structural stability of coumestrol with human estrogen receptor α and β subtypes: A combined approach involving docking and molecular dynamics simulation studies; RSC Advances. 5 (99): 81295-81312; 2015
9) Zafar A.[‡], **Ahmad S.**[‡], Rizvi A. and Ahmad M.; Novel non-peptide inhibitors against SmCL1 of *Schistosoma mansoni:* in silico elucidation, implications and evaluation via knowledge based drug discovery; PLoS ONE. 10 (5): 1-33; 2015

[§] *A complete list of publication is available at https://orcid.org/0000-0002-8762-4670*

[‡] *First authors*

[⊥] *Authors with equal contribution*

11. Bibliography

1. Obritsch MD, Fish DN, MacLaren R, Jung R. Nosocomial infections due to multidrug-resistant *Pseudomonas aeruginosa*: Epidemiology and treatment options. Pharmacotherapy 2005;25(10):1353-1364.

2. Mahajan-Miklos S, Rahme LG, Ausubel FM. Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts. Mol Microbiol. 2000;37(5):981-988.

3. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. Lancet Infect Dis. 2018;18(3):318-327.

4. Van Delden C, Iglewski BH. Cell-to-cell signaling and *Pseudomonas aeruginosa* infections. Emerg Infect Dis. 1998;4(4):551-560.

5. da Mata Madeira PV, Zouhir S, Basso P, Neves D, Laubier A, Salacha R, et al. Structural basis of lipid targeting and destruction by the Type V secretion system of *Pseudomonas aeruginosa*. J Mol Biol. 2016;428(9 Pt A):1790-1803.

6. Salacha R, Kovačić F, Brochier-Armanet C, Wilhelm S, Tommassen J, Filloux A, et al. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. Environ Microbiol. 2010;12(6):1498-1512.

7. Schmiel DH, Miller VL. Bacterial phospholipases and pathogenesis. Microbes and Infection. 1999;1(13):1103-1112.

8. Richmond GS, Smith TK. Phospholipases A(1). Int J Mol Sci. 2011;12(1):588-612.

9. Fozo EM, Rucks EA. The making and taking of lipids: the role of bacterial lipid synthesis and the harnessing of host lipids in bacterial pathogenesis. Adv Microb Physiol. 2016;69:51-155.

10. Housley NA, Winkler HH, Audia JP. The *Rickettsia prowazekii* ExoU homologue possesses phospholipase A1 (PLA1), PLA2, and lyso-PLA2 activities and can function in the absence of any eukaryotic cofactors in vitro. J Bacteriol. 2011;193(18):4634-4642.

11. Dowhan W. Molecular basis for membrane phospholipid diversity: why are there so many lipids? Annu Rev Biochem. 1997;66:199-232.

12. Tannaes T, Bukholm IK, Bukholm G. High relative content of lysophospholipids of *Helicobacter pylori* mediates increased risk for ulcer disease. FEMS Immunol Med Microbiol. 2005;44(1):17-23.

13. Zheng L, Lin Y, Lu S, Zhang J, Bogdanov M. Biogenesis, transport and remodeling of lysophospholipids in gram-negative bacteria. Biochim Biophys Acta Mol Cell Biol Lipids. 2017;1862(11):1404-1413.

14. Corradi V, Sejdiu BI, Mesa-Galloso H, Abdizadeh H, Noskov SY, Marrink SJ, et al. Emerging diversity in lipid-protein interactions. Chem Rev. 2019;119(9):5775-5848.

15. Koehler J, Sulistijo ES, Sakakura M, Kim HJ, Ellis CD, Sanders CR. Lysophospholipid micelles sustain the stability and catalytic activity of diacylglycerol kinase in the absence of lipids. Biochemistry 2010;49(33):7089-7099.

16. Nomura T, Cranfield CG, Deplazes E, Owen DM, Macmillan A, Battle AR, et al. Differential effects of lipids and lyso-lipids on the mechanosensitivity of the mechanosensitive channels MscL and MscS. Proc Natl Acad Sci U S A. 2012;109(22):8770-8775.

17. Le Senechal C, Crouzet M, Costaglioli P, Barthe C, Bure C, Vilain S. Phospholipid content of *Pseudomonas aeruginosa* PAO1 is modulated by the growth phase rather than the immobilization state. Lipids 2019;54(9):519-529.

18. Benamara H, Rihouey C, Abbes I, Ben Mlouka MA, Hardouin J, Jouenne T, et al. Characterization of membrane lipidome changes in *Pseudomonas aeruginosa* during biofilm growth on glass wool. PLoS One. 2014;9(9):e108478.

19. Zhou L, Zhang LH, Camara M, He YW. The DSF family of quorum sensing signals: diversity, biosynthesis, and turnover. Trends Microbiol. 2017;25(4):293-303.

20. Twomey KB, O'Connell OJ, McCarthy Y, Dow JM, O'Toole GA, Plant BJ, et al. Bacterial cis-2-unsaturated fatty acids found in the cystic fibrosis airway modulate virulence and persistence of *Pseudomonas aeruginosa*. Isme J. 2012;6(5):939-950.

21. Ryan RP, Fouhy Y, Garcia BF, Watt SA, Niehaus K, Yang L, et al. Interspecies signalling via the *Stenotrophomonas maltophilia* diffusible signal

factor influences biofilm formation and polymyxin tolerance in *Pseudomonas aeruginosa*. Mol Microbiol. 2008;68(1):75-86.

22. Kovačić F, Bleffert F, Caliskan M, Wilhelm S, Granzin J, Batra-Safferling R, et al. A membrane-bound esterase PA2949 from *Pseudomonas aeruginosa* is expressed and purified from Escherichia coli. FEBS Open Bio. 2016;6(5):484-493.

23. Bleffert F, Granzin J, Caliskan M, Schott-Verdugo S, Siebers M, Thiele B, et al. Evidence for a bacterial Lands cycle phospholipase A: Structural and mechanistic insights into membrane phospholipid remodeling. bioRxiv. 2021:2021.06.22.448587.

24. Ahmad S, Strunk CH, Schott-Verdugo SN, Jaeger K-E, Kovacic F, Gohlke H. Substrate access mechanism in a novel membrane-bound phospholipase A of *Pseudomonas aeruginosa* concordant with specificity and regioselectivity. J Chem Inf Model. 2021;61(11):5626–5643.

25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-242.

26. Bleffert F, Granzin, J., Caliskan, M., Schott-Verdugo, S.N., Rahme, L., Siebers, M., Thiele, B., Doermann, P., Gohlke, H., Batra-Safferling, R., Kovacic, F., Jaeger, K.-E. Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation associated with bacterial virulence. 2019. doi: 10.2210/pdb6I8W/pdb (accessed on 21 June 2021).

27. Gessard C. Sur les colorations bleue et verte des linges a pansements. Compt Rend Acad Sci. 1882;94:536-568.

28. Diggle SP, Whiteley M. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. Microbiology (Reading) 2020;166(1):30-33.

29. LaBauve AE, Wargo MJ. Growth and laboratory maintenance of *Pseudomonas aeruginosa*. Curr Protoc Microbiol. 2012;Chapter 6:Unit 6E 1.

30. Breidenstein EB, de la Fuente-Nunez C, Hancock RE. *Pseudomonas aeruginosa*: all roads lead to resistance. Trends Microbiol. 2011;19(8):419-426.

31. Gellatly SL, Hancock REW. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. Pathog Dis. 2013;67(3):159-173.

32. Kyaw MH, Kern DM, Zhou S, Tunceli O, Jafri HS, Falloon J. Healthcare utilization and costs associated with *S. aureus* and *P. aeruginosa* pneumonia in

the intensive care unit: a retrospective observational cohort study in a US claims database. BMC Health Serv Res. 2015;15:241.

33. Driscoll JA, Brody SL, Kollef MH. The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections. Drugs. 2007;67(3):351-368.

34. Arancibia F, Bauer TT, Ewig S, Mensa J, Gonzalez J, Niederman MS, et al. Community-acquired pneumonia due to gram-negative bacteria and *Pseudomonas aeruginosa*: incidence, risk, and prognosis. Arch Intern Med. 2002;162(16):1849-1858.

35. Kaier K, Heister T, Gotting T, Wolkewitz M, Mutters NT. Measuring the in-hospital costs of *Pseudomonas aeruginosa* pneumonia: methodology and results from a German teaching hospital. BMC Infect Dis. 2019;19(1):1028.

36. Kang CI, Kim SH, Kim HB, Park SW, Choe YJ, Oh MD, et al. *Pseudomonas aeruginosa* bacteremia: risk factors for mortality and influence of delayed receipt of effective antimicrobial therapy on clinical outcome. Clin Infect Dis. 2003;37(6):745-751.

37. Girard G, Bloemberg GV. Central role of quorum sensing in regulating the production of pathogenicity factors in *Pseudomonas aeruginosa*. Future Microbiol. 2008;3(1):97-106.

38. Newman JW, Floyd RV, Fothergill JL. The contribution of *Pseudomonas aeruginosa* virulence factors and host factors in the establishment of urinary tract infections. FEMS Microbiol Lett. 2017;364(15).

39. Kovačić F. Novel phospholipases A of *Pseudomonas aeruginosa* biochemical characterisation and cellular localisation. Düsseldorf, Germany: PhD Thesis, Heinrich Heine University; 2010

40. Rasko DA, Sperandio V. Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. 2010;9(2):117-128.

41. O'Toole GA, Kolter R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. Mol Microbiol. 1998;30(2):295-304.

42. Craig L, Forest KT, Maier B. Type IV pili: dynamics, biophysics and functional consequences. Nat Rev Microbiol. 2019;17(7):429-440.

Bibliography

43. Taylor PK, Yeung AT, Hancock RE. Antibiotic resistance in *Pseudomonas aeruginosa* biofilms: towards the development of novel anti-biofilm therapies. J Biotechnol. 2014;191:121-130.

44. Anderl JN, Franklin MJ, Stewart PS. Role of antibiotic penetration limitation in *Klebsiella pneumoniae* biofilm resistance to ampicillin and ciprofloxacin. Antimicrob Agents Chemother. 2000;44(7):1818-1824.

45. Walters MC, 3rd, Roe F, Bugnicourt A, Franklin MJ, Stewart PS. Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of *Pseudomonas aeruginosa* biofilms to ciprofloxacin and tobramycin. Antimicrob Agents Chemother. 2003;47(1):317-323.

46. Nicas TI, Hancock RE. *Pseudomonas aeruginosa* outer membrane permeability: isolation of a porin protein F-deficient mutant. J Bacteriol. 1983;153(1):281-285.

47. Hancock RE. Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermentative gram-negative bacteria. Clin Infect Dis. 1998;27 Suppl 1:S93-S99.

48. Songer JG. Bacterial phospholipases and their role in virulence. Trends Microbiol. 1997;5(4):156-161.

49. Hinkel LA, Wargo MJ. Participation of bacterial lipases, sphingomyelinases, and phospholipases in gram-negative bacterial pathogenesis. In: Goldfine H, editor. Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids. Handbook of Hydrocarbon and Lipid Microbiology. Cham, Switzerland: Springer International Publishing; 2020.

50. Snijder HJ, Dijkstra BW. Bacterial phospholipase A: structure and function of an integral membrane phospholipase. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids. 2000;1488(1-2):91-101.

51. Titball RW. Bacterial phospholipases. J Appl Microbiol. 1998;84:127S–137S.

52. Gonzalez-Bullon D, Uribe KB, Martin C, Ostolaza H. Phospholipase A activity of adenylate cyclase toxin mediates translocation of its adenylate cyclase domain. Proc Natl Acad Sci U S A. 2017;114(33):E6784-E6793.

53. Istivan TS, Coloe PJ. Phospholipase A in Gram-negative bacteria and its role in pathogenesis. Microbiology (Reading) 2006;152(Pt 5):1263-1274.

54. Schmiel DH, Wagar E, Karamanou L, Weeks D, Miller VL. Phospholipase A of *Yersinia enterocolitica* contributes to pathogenesis in a mouse model. Infect Immun. 1998;66(8):3941-3951.

55. Monet M, Gkika D, Lehen'kyi V, Pourtier A, Vanden Abeele F, Bidaux G, et al. Lysophospholipids stimulate prostate cancer cell migration via TRPV2 channel activation. Biochim Biophys Acta. 2009;1793(3):528-539.

56. Whetton AD, Lu Y, Pierce A, Carney L, Spooncer E. Lysophospholipids synergistically promote primitive hematopoietic cell chemotaxis via a mechanism involving Vav 1. Blood 2003;102(8):2798-2802.

57. Vanden Abeele F, Zholos A, Bidaux G, Shuba Y, Thebault S, Beck B, et al. Ca2+-independent phospholipase A2-dependent gating of TRPM8 by lysophospholipids. J Biol Chem. 2006;281(52):40174-40182.

58. Zhang L, Barritt GJ. Evidence that TRPM8 is an androgen-dependent Ca2+ channel required for the survival of prostate cancer cells. Cancer Res. 2004;64(22):8365-8373.

59. Mills GB, Eder A, Fang X, Hasegawa Y, Mao M, Lu Y, et al. Critical role of lysophospholipids in the pathophysiology, diagnosis, and management of ovarian cancer. Cancer Treat Res. 2002;107:259-283.

60. Murph M, Tanaka T, Pang J, Felix E, Liu S, Trost R, et al. Liquid chromatography mass spectrometry for quantifying plasma lysophospholipids: potential biomarkers for cancer diagnosis. Methods Enzymol. 2007;433:1-25.

61. Sutphen R, Xu Y, Wilbanks GD, Fiorica J, Grendys EC, Jr., LaPolla JP, et al. Lysophospholipids are potential biomarkers of ovarian cancer. Cancer Epidemiol Biomarkers Prev. 2004;13(7):1185-1191.

62. Xu Y. Lysophospholipid signaling in the epithelial ovarian cancer tumor microenvironment. Cancers (Basel). 2018; 10(7):227.

63. Hayakawa K, Kurano M, Ohya J, Oichi T, Kano K, Nishikawa M, et al. Lysophosphatidic acids and their substrate lysophospholipids in cerebrospinal fluid as objective biomarkers for evaluating the severity of lumbar spinal stenosis. Sci Rep. 2019; 9(1):9144.

64. Yan YF, Du ZF, Chen C, Li JX, Xiong X, Zhang Y, et al. Lysophospholipid profiles of apolipoprotein E-deficient mice reveal potential lipid biomarkers associated with atherosclerosis progression using validated UPLC-QTRAP-MS/MS-based lipidomics approach. J Pharm Biomed Anal. 2019;171:148-157.

Bibliography

65. Zhou XS, Wang RP, Zhang T, Liu F, Zhang W, Wang GL, et al. Identification of lysophosphatidylcholines and sphingolipids as potential biomarkers for acute aortic dissection via serum metabolomics. Eur J Vasc Endovasc. 2019;57(3):434-441.

66. Barber CE, Tang JL, Feng JX, Pan MQ, Wilson TJG, Slater H, et al. A novel regulatory system required for pathogenicity of *Xanthomonas campestris* is mediated by a small diffusible signal molecule. Mol Microbiol. 1997;24(3):555-566.

67. Jimenez PN, Koch G, Thompson JA, Xavier KB, Cool RH, Quax WJ. The multiple signaling systems regulating virulence in *Pseudomonas aeruginosa*. Microbiol Mol Biol Rev. 2012;76(1):46-65.

68. Ryan RP, Dow JM. Diffusible signals and interspecies communication in bacteria. Microbiology 2008;154(Pt 7):1845-1858.

69. Dow JM. Diffusible signal factor-dependent quorum sensing in pathogenic bacteria and its exploitation for disease control. J Appl Microbiol. 2017;122(1):2-11.

70. Douzi B, Trinh NTT, Michel-Souzy S, Desmyter A, Ball G, Barbier P, et al. Unraveling the self-assembly of the *Pseudomonas aeruginosa* XcpQ secretin periplasmic domain provides new molecular insights into Type II secretion system secreton architecture and dynamics. mBio. 2017;8(5):e01185-17.

71. Hay ID, Belousoff MJ, Lithgow T. Structural basis of type 2 secretion system engagement between the inner and outer bacterial membranes. mBio. 2017;8(5): e01344-17.

72. Apidianakis Y, Rahme LG. Drosophila melanogaster as a model host for studying *Pseudomonas aeruginosa* infection. Nat Protoc. 2009;4(9):1285-1294.

73. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, et al. The alpha/beta hydrolase fold. Protein Eng. 1992;5(3):197-211.

74. Brezovsky J, Kozlikova B, Damborsky J. Computational analysis of protein tunnels and channels. Methods Mol Biol. 2018;1685:25-42.

75. Gora A, Brezovsky J, Damborsky J. Gates of enzymes. Chem Rev. 2013;113(8):5871-5923.

76. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol. 2012;8(10):e1002708.

77. Pravda L, Berka K, Svobodova Varekova R, Sehnal D, Banas P, Laskowski RA, et al. Anatomy of enzyme channels. BMC Bioinformatics. 2014;15:379.

78. Arroyo-Manez P, Bikiel DE, Boechi L, Capece L, Di Lella S, Estrin DA, et al. Protein dynamics and ligand migration interplay as studied by computer simulation. Biochim Biophys Acta. 2011;1814(8):1054-1064.

79. Ho FM. Uncovering channels in photosystem II by computer modelling: current progress, future prospects, and lessons from analogous systems. Photosynth Res. 2008;98(1-3):503-522.

80. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol. 2002;9(9):646-652.

81. Shen T, Tai K, Henchman RH, McCammon JA. Molecular dynamics of acetylcholinesterase. Acc Chem Res. 2002;35(6):332-340.

82. Kingsley LJ, Lill MA. Ensemble generation and the influence of protein flexibility on geometric tunnel prediction in cytochrome P450 enzymes. PLoS One. 2014;9(6):e99408.

83. Klvana M, Pavlova M, Koudelakova T, Chaloupkova R, Dvorak P, Prokop Z, et al. Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. J Mol Biol. 2009;392(5):1339-1356.

84. Li W, Shen J, Liu G, Tang Y, Hoshino T. Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations. Proteins. 2011;79(1):271-281.

85. Otyepka M, Skopalik J, Anzenbacherova E, Anzenbacher P. What common structural features and variations of mammalian P450s are known to date? Biochim Biophys Acta. 2007;1770(3):376-389.

86. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MSP. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph. 1996;14(6):354-360.

Bibliography

87. Sehnal D, Svobodova Varekova R, Berka K, Pravda L, Navratilova V, Banas P, et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. J Cheminform. 2013;5(1):39.

88. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R. MolAxis: efficient and accurate identification of channels in macromolecules. Proteins. 2008;73(1):72-86.

89. Brezovsky J, Chovancova E, Gora A, Pavelka A, Biedermannova L, Damborsky J. Software tools for identification, visualization and analysis of protein tunnels and channels. Biotechnol Adv. 2013;31(1):38-49.

90. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. Steered Molecular Dynamics. Deuflhard P, Hermans J, Leimkuhler B, Mark AE, Reich S, Skeel R.D. (Eds), Computational molecular dynamics: challenges, methods, ideas. Lecture notes in computational science and engineering. 1999; vol 4. Springer, Berlin, Heidelberg, Germany.

91. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular dynamics study of unbinding of the avidin-biotin complex. Biophys J. 1997;72(4):1568-1581.

92. Niemela A, Koivuniemi A. Positive allosteric modulators of lecithin: Cholesterol acyltransferase adjust the orientation of the membrane-binding domain and alter its spatial free energy profile. PLoS Comput Biol. 2021;17(3):e1008426.

93. Lo R, Chandar NB, Ghosh S, Ganguly B. The reactivation of tabun-inhibited mutant AChE with Ortho-7: steered molecular dynamics and quantum chemical studies. Mol Biosyst. 2016;12(4):1224-1231.

94. Ludemann SK, Lounnas V, Wade RC. How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. J Mol Biol. 2000;303(5):813-830.

95. Martinez L, Webb P, Polikarpov I, Skaf MS. Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands. J Med Chem. 2006;49(1):23-26.

96. Cheung LS, Shea DJ, Nicholes N, Date A, Ostermeier M, Konstantopoulos K. Characterization of monobody scaffold interactions with ligand via force spectroscopy and steered molecular dynamics. Sci Rep. 2015;5:8247.

97. Ettig R, Kepper N, Stehr R, Wedemann G, Rippe K. Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. Biophys J. 2011;101(8):1999-2008.

98. Stepaniants S, Schulten K, Izrailev S. Extraction of lipids from phospholipid membranes by steered molecular dynamics. J Mol Model. 1997;3(12):473-475.

99. Jensen MO, Yin Y, Tajkhorshid E, Schulten K. Sugar transport across lactose permease probed by steered molecular dynamics. Biophys J. 2007;93(1):92-102.

100. Monticelli L, Robertson KM, MacCallum JL, Tieleman DP. Computer simulation of the KvAP voltage-gated potassium channel: steered molecular dynamics of the voltage sensor. FEBS Letters. 2004;564(3):325-332.

101. Liu Z, Xu Y, Tang P. Steered molecular dynamics simulations of Na+ permeation across the gramicidin A channel. J Phys Chem B. 2006;110(25):12789-12795.

102. Do PC, Lee EH, Le L. Steered molecular dynamics simulation in rational drug design. J Chem Inf Model. 2018;58(8):1473-1482.

103. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781-1802.

104. Binnig G, Quate CF, Gerber C. Atomic force microscope. Phys Rev Lett. 1986;56(9):930-933.

105. Svoboda K, Block SM. Biological applications of optical forces. Annu Rev Biophys Biomol Struct. 1994;23:247-285.

106. Evans E, Ritchie K, Merkel R. Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. Biophys J. 1995;68(6):2580-2587.

107. Izrailev S, Crofts AR, Berry EA, Schulten K. Steered molecular dynamics simulation of the Rieske subunit motion in the cytochrome bc(1) complex. Biophys J. 1999;77(4):1753-1768.

108. Celerse F, Lagardere L, Derat E, Piquemal JP. Massively Parallel Implementation of Steered Molecular Dynamics in Tinker-HP: Comparisons of

polarizable and non-polarizable simulations of realistic systems. J Chem Theory Comput. 2019;15(6):3694-3709.

109. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J Comput Phys. 1977;23(2):187-199.

110. Kästner J. Umbrella sampling. Wiley Interdiscip Rev Comput Mol Sci. 2011;1(6):932-942.

111. Roux B. The calculation of the potential of mean force using computer simulations. Comput Phys Commun. 1995;91(1-3):275-282.

112. Jarzynski C. Nonequilibrium equality for free energy differences. Phys Rev Lett. 1997;78(14):2690-2693.

113. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. J Chem Phys. 2003;119(6):3559-3566.

114. Noh SY, Notman R. Comparison of umbrella sampling and steered molecular dynamics methods for computing free energy profiles of aromatic substrates through phospholipid bilayers. J Chem Phys. 2020;153(3):034115.

115. Yang Y, Pan L, Lightstone FC, Merz KM, Jr. The role of molecular dynamics potential of mean force calculations in the investigation of enzyme catalysis. Methods Enzymol. 2016;577:1-29.

116. Ciupka D, Gohlke H. On the potential alternate binding change mechanism in a dimeric structure of Pyruvate Phosphate Dikinase. Sci Rep. 2017;7(1):8020.

117. Pagani G, Gohlke H. On the contributing role of the transmembrane domain for subunit-specific sensitivity of integrin activation. Sci Rep. 2018;8(1):5733.

118. Waschenbach L, Gertzen CGW, Keitel V, Gohlke H. Dimerization energetics of the G-protein coupled bile acid receptor TGR5 from all-atom simulations. J Comput Chem. 2020;41(9):874-884.

119. van Gunsteren WF, Daura X, Mark AE. Computation of free energy. Helvetica Chimica Acta. 2002;85(10):3113-3129.

120. Mills M, Andricioaei I. An experimentally guided umbrella sampling protocol for biomolecules. J Chem Phys. 2008;129(11):114101.

121. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules.1. The method. J Comput Chem. 1992;13(8):1011-1021.

122. Grossfield A. WHAM: the weighted histogram analysis method, version 2.0.9 Available online: http://membrane.urmc.rochester.edu/content/wham (accessed on 21 June 2021)

123. Feig M, Nawrocki G, Yu I, Wang PH, Sugita Y. Challenges and opportunities in connecting simulations with experiments via molecular dynamics of cellular environments. J Phys Conf Ser. 2018;1036:012010.

124. Tieleman DP, Marrink SJ. Lipids out of equilibrium: energetics of desorption and pore mediated flip-flop. J Am Chem Soc. 2006;128(38):12462-12467.

125. Chen PC, Kuyucak S. Accurate determination of the binding free energy for KcsA-charybdotoxin complex from the potential of mean force calculations with restraints. Biophys J. 2011;100(10):2466-2474.

126. Kokkonen P, Slanska M, Dockalova V, Pinto GP, Sanchez-Carnerero EM, Damborsky J, et al. The impact of tunnel mutations on enzymatic catalysis depends on the tunnel-substrate complementarity and the rate-limiting step. Comput Struct Biotechnol J. 2020;18:805-813.

127. Schmitt J, Brocca S, Schmid RD, Pleiss J. Blocking the tunnel: engineering of *Candida rugosa* lipase mutants with short chain length specificity. Protein Eng. 2002;15(7):595-601.

128. Biedermannova L, Prokop Z, Gora A, Chovancova E, Kovacs M, Damborsky J, et al. A single mutation in a tunnel to the active site changes the mechanism and kinetics of product release in haloalkane dehalogenase LinB. J Biol Chem. 2012;287(34):29062-29074.

129. Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, et al. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. Nat Chem Biol. 2009;5(10):727-733.

130. Gustafsson C, Vassiliev S, Kurten C, Syren PO, Brinck T. MD simulations reveal complex water paths in squalene-hopene cyclase: tunnel-obstructing mutations increase the flow of water in the active site. ACS Omega. 2017;2(11):8495-8506.

Bibliography

131. Lu Z, Li X, Zhang R, Yi L, Ma Y, Zhang G. Tunnel engineering to accelerate product release for better biomass-degrading abilities in lignocellulolytic enzymes. Biotechnol Biofuels. 2019;12:275.

132. Kokkonen P, Beier A, Mazurenko S, Damborsky J, Bednar D, Prokop Z. Substrate inhibition by the blockage of product release and its control by tunnel engineering. RSC Chem Biol. 2021;2(2):645-655.

133. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005;33(Web Server issue):W382-W388.

134. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. J Mol Biol. 2002;320(2):369-387.

135. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The stability effects of protein mutations appear to be universally distributed. J Mol Biol. 2007;369(5):1318-1332.

136. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993;234(3):779-815.

137. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 2012;40(Database issue):D370-D376.

138. Conrad RS, Gilleland HE, Jr. Lipid alterations in cell envelopes of polymyxin-resistant *Pseudomonas aeruginosa* isolates. J Bacteriol. 1981;148(2):487-497.

139. Murzyn K RT, Pasenkiewicz-Gierula M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. Biophys J. 2005:1091-1103.

140. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. Biophys J. 2009;97(1):50-58.

141. Schott-Verdugo S, Gohlke H. PACKMOL-Memgen: A simple-to-use, generalized workflow for membrane-protein-lipid-bilayer system building. J Chem Inf Model. 2019;59(6):2522-2528.

142. Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., et al. The Amber biomolecular simulation programs. J Comput Chem. 2005;26(16):1668-1688.

143. Le Grand S, Gotz AW, Walker RC. SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations. Comput Phys Commun. 2013;184(2):374-380.

144. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput. 2015;11(8):3696-3713.

145. D.A. Case DSC, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D.Greene,N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. AMBER 2017 Reference Manual, University of California, San Francisco. 2017. Available from: https://ambermd.org/doc12/Amber17.pdf.

146. Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR, et al. Lipid14: The Amber lipid force field. J Chem Theory Comput. 2014;10(2):865-879.

147. Skjevik AA, Madej BD, Dickson CJ, Lin C, Teigen K, Walker RC, et al. Simulation of lipid bilayer self-assembly using all-atom lipid force fields. Phys Chem Chem Phys. 2016;18(15):10573-10584.

148. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926-935.

149. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-Alkanes. J Comput Phys. 1977;23(3):327-341.

150. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys. 1993;98(12):10089-10092.

151. Quigley D, Probert MI. Langevin dynamics in constant pressure extended systems. J Chem Phys. 2004;120(24):11432-11441.

Bibliography

152. Berger O, Edholm O, Jahnig F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. Biophys J. 1997;72(5):2002-2013.

153. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684-3690.

154. Roe DR, Cheatham TE, 3rd. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput. 2013;9(7):3084-3095.

155. Schrodinger, LLC. The PyMOL molecular graphics system, Version 1.8. 2015.

156. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J. Mol. Graph. 1996;14(1):33-38.

157. Paloncyova M, Berka K, Otyepka M. Convergence of free energy profile of coumarin in lipid bilayer. J Chem Theory Comput. 2012;8(4):1200-1211.

158. Marrink S-J, Berger O, Tieleman P, Jähnig F. Adhesion forces of lipids in a phospholipid membrane studied by molecular dynamics simulations. Biophys J. 1998;74(2):931-943.

159. Mouchlis VD, Bucher D, McCammon JA, Dennis EA. Membranes serve as allosteric activators of phospholipase A2, enabling it to extract, bind, and hydrolyze phospholipid substrates. Proc Natl Acad Sci U S A. 2015;112(6):E516-E525.

160. Bureau HR, Merz DR, Jr., Hershkovits E, Quirk S, Hernandez R. Constrained unfolding of a helical peptide: implicit versus explicit solvents. PLoS One. 2015;10(5):e0127034.

161. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem. 1993;97(40):10269-10280.

162. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. 2006;25(2):247-260.

163. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 16 Revision A.03. Gaussian, Inc., Wallingford, CT, 2016.

164. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004;25(9):1157-1174.

165. Schmidt D, Boehm M, McClendon CL, Torella R, Gohlke H. Cosolvent-enhanced eampling and unbiased identification of cryptic pockets suitable for structure-based drug design. J Chem Theory Comput. 2019;15(5):3331-3343.

166. Goldstein DA, Solomon AK. Determination of equivalent pore radius for human red cells by osmotic pressure measurement. J Gen Physiol. 1960;44:1-17.

167. Pan J, Marquardt D, Heberle FA, Kucerka N, Katsaras J. Revisiting the bilayer structures of fluid phase phosphatidylglycerol lipids: Accounting for exchangeable hydrogens. Biochim Biophys Acta. 2014;1838(11):2966-2969.

168. King MD, Marsh D. Head group and chain length dependence of phospholipid self-assembly studied by spin-label electron spin resonance. Biochemistry 1987;26(5):1224-1231.

169. Stafford RE, Fanni T, Dennis EA. Interfacial properties and critical micelle concentration of lysophospholipids. Biochemistry 1989;28(12):5113-5120.

170. Stoica I, Sadiq SK, Coveney PV. Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. J Am Chem Soc. 2008;130(8):2639-2648.

171. Homeyer N, Gohlke H. FEW: a workflow tool for free energy calculations of ligand binding. J Comput Chem. 2013;34(11):965-973.

172. Homeyer N, Stoll F, Hillisch A, Gohlke H. Binding free energy calculations for lead optimization: assessment of their accuracy in an industrial drug design context. J Chem Theory Comput. 2014;10(8):3331-3344.

173. Bisswanger H. Enzyme kinetics: principles and methods. 3rd edn. Weinheim, Germany: Wiley-VCH; 2017.

174. Nelson DL, Lehninger AL, Cox MM. Lehninger principles of biochemistry. 5th edn. New York, USA: W.H. Freeman; 2008.

175. Strunk CH. Biochemische charakterisierung der phospholipase PlaF aus *Pseudomonas aeruginosa* und ihres humanen Homologs ABHD6. Düsseldorf, Germany: PhD Thesis, Heinrich Heine University; 2020.

Bibliography

176. Michel J, Essex JW. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. J Comput Aided Mol Des. 2010;24(8):639-658.

177. Friedman R, Nachliel E, Gutman M. Molecular dynamics of a protein surface: ion-residues interactions. Biophys J. 2005;89(2):768-781.

178. Salamanca Viloria J, Allega MF, Lambrughi M, Papaleo E. An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass. Sci Rep. 2017;7(1):2838.

179. Cornell CE, Black RA, Xue M, Litz HE, Ramsay A, Gordon M, et al. Prebiotic amino acids bind to and stabilize prebiotic fatty acid membranes. Proc Natl Acad Sci U S A. 2019;116(35):17239-17244.

180. Glen S. Z-test: definition & two proportion Z-test—statistics how to. Available online: https://www.statisticshowto.com/z-test/ (accessed on 21 June 2021)

181. McKillup S. Statistics explained: an introductory guide for life scientists. Cambridge, United Kingdom: Cambridge University Press; 2011.

182. Sprinthall RC. Basic Statistical Analysis. 9th edn. London, United Kingdom: Pearson Education; 2011.

183. Awasthi M, Batra J, Kateriya S. Disulphide bridges of phospholipase C of *Chlamydomonas reinhardtii* modulates lipid interaction and dimer stability. PLoS One. 2012;7(6):e39258.

184. de Oliveira AH, Giglio JR, Andriao-Escarso SH, Ito AS, Ward RJ. A pH-induced dissociation of the dimeric form of a lysine 49-phospholipase A2 abolishes Ca2+-independent membrane damaging activity. Biochemistry 2001;40(23):6912-6920.

185. Dekker N, Tommassen J, Lustig A, Rosenbusch JP, Verheij HM. Dimerization regulates the enzymatic activity of *Escherichia coli* outer membrane phospholipase A. J Biol Chem. 1997;272(6):3179-3184.

186. Gomes AAS, Cardoso FF, Souza MF, Oliveira CLP, Perahia D, Magro AJ, et al. The allosteric activation mechanism of a phospholipase A2-like toxin from *Bothrops jararacussu* venom: a dynamic description. Sci Rep. 2020;10(1):16252.

187. Kingma RL, Egmond MR. Activation of a covalent outer membrane phospholipase A dimer. Eur J Biochem. 2002;269(8):2178-2185.

188. Malley KR, Koroleva O, Miller I, Sanishvili R, Jenkins CM, Gross RW, et al. The structure of iPLA2beta reveals dimeric active sites and suggests mechanisms of regulation and localization. Nat Commun. 2018;9(1):765.

189. Snijder HJ, Ubarretxena-Belandia I, Blaauw M, Kalk KH, Verheij HM, Egmond MR, et al. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. Nature 1999;401(6754):717-721.

190. Tomasselli AG, Hui J, Fisher J, Zürcher-Neely H, Reardon IM, Oriaku E, et al. Dimerization and activation of porcine pancreatic phospholipase A2 via substrate level acylation of lysine 56. J Biol Chem. 1989;264(17):10041-10047.

191. Wullich SC, Wijma HJ, Janssen DB, Fetzner S. Stabilizing AqdC, a *Pseudomonas* quinolone signal-cleaving dioxygenase from Mycobacteria, by FRESCO-based protein engineering. Chembiochem. 2021;22(4):733-742.

192. Cui YL, Zheng QC, Zhang JL, Zhang HX. Molecular basis of the recognition of arachidonic acid by cytochrome P450 2E1 along major access tunnel. Biopolymers 2015;103(1):53-66.

193. Wu D. An efficient umbrella potential for the accurate calculation of free energies by molecular simulation. J Chem Phys. 2010;133(4):044115.

194. Ahmed A, Rippmann F, Barnickel G, Gohlke H. A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins. J Chem Inf Model. 2011;51(7):1604-1622.

195. Domanski J, Hedger G, Best RB, Stansfeld PJ, Sansom MSP. Convergence and sampling in determining free energy landscapes for membrane protein association. J Phys Chem B. 2017;121(15):3364-3375.

196. Rosta E, Woodcock HL, Brooks BR, Hummer G. Artificial reaction coordinate "tunneling" in free-energy calculations: the catalytic reaction of RNase H. J Comput Chem. 2009;30(11):1634-1641.

197. Neale C, Pomes R. Sampling errors in free energy simulations of small molecules in lipid bilayers. Biochim Biophys Acta. 2016;1858(10):2539-2548.

198. Tahir MA, Van Lehn RC, Choi SH, Alexander-Katz A. Solvent-exposed lipid tail protrusions depend on lipid membrane composition and curvature. Biochim Biophys Acta. 2016;1858(6):1207-1215.

Bibliography

199. Hari SB, Grant RA, Sauer RT. Structural and functional analysis of *E. coli* cyclopropane fatty acid synthase. Structure 2018;26(9):1251-1258.e3.

200. Saura P, Marechal JD, Masgrau L, Lluch JM, Gonzalez-Lafont A. Computational insight into the catalytic implication of head/tail-first orientation of arachidonic acid in human 5-lipoxygenase: consequences for the positional specificity of oxygenation. Phys Chem Chem Phys. 2016;18(33):23017-23035.

201. Qian H, Zhao X, Yan R, Yao X, Gao S, Sun X, et al. Structural basis for catalysis and substrate specificity of human ACAT1. Nature 2020;581(7808):333-338.

202. Tong S, Lin Y, Lu S, Wang M, Bogdanov M, Zheng L. Structural insight into substrate selection and catalysis of lipid phosphate phosphatase PgpB in the cell membrane. J Biol Chem. 2016;291(35):18342-18352.

203. Ibarguren M, Lopez DJ, Escriba PV. The effect of natural and synthetic fatty acids on membrane structure, microdomain organization, cellular functions and human health. Biochim Biophys Acta. 2014;1838(6):1518-1528.

204. Desbois AP, Smith VJ. Antibacterial free fatty acids: activities, mechanisms of action and biotechnological potential. Appl Microbiol Biotechnol. 2010;85(6):1629-1642.

205. Liu H, Yu C, Feng D, Cheng T, Meng X, Liu W, et al. Production of extracellular fatty acid using engineered *Escherichia coli*. Microb Cell Fact. 2012;11:41.