

**Modeling of membrane proteins in a bilayer
context: from system construction to structural
prediction and dynamics**

Inaugural dissertation

for the attainment of the title of doctor

in the Faculty of Mathematics and Natural Sciences

at the Heinrich Heine University Düsseldorf

-&-

in the Faculty of Engineering

at the University of Talca

presented by

Stephan Nicolás Schott Verdugo

from

Valdivia, Chile

Düsseldorf, January 2020

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

and

the Center for Bioinformatics, Simulation and Modeling
at the University of Talca

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

Supervisor: Prof. Dr. Holger Gohlke

Co-Supervisor: Prof. Dr. Georg Groth

Date of the oral examination: 18th June 2020

Affidavit

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the „Rules on the Principles for 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, 30th of January, 2020

To Whom It May...
And Also To Whom It May Not Concern

“...At present this postulate rests only on suggestive evidence, most of it morphological. This situation brings us back to the necessity of obtaining detailed and -if possible-comprehensive data on the chemistry and function of the different membranes of the secretory pathway and on their interactions...”

GEORGE PALADE

Intracellular aspects of the process of protein secretion

Nobel Lecture, December 12, 1974

Table of Contents

List of Publications.....	ix
Abbreviations	x
Abstract	xiii
Zusammenfassung.....	xiv
Resumen	xvi
1 Introduction	1
2 Background	1
2.1 Membrane lipids and composition	1
2.2 Structure of membrane proteins	4
2.3 Computational methods to study membranes and membrane proteins	5
Molecular dynamics of membrane systems	5
Free energy umbrella sampling calculations	7
Structure prediction of membrane proteins	8
Protein residue-residue contact predictions.....	9
2.4 Ethylene-induced hormonal response and signaling pathway.....	12
Ethylene: the first identified gaseous hormone	12
Fruit ripening in economy: ethylene signaling as a biotechnological target.....	13
Ethylene receptors and the ethylene-signaling pathway	14
2.5 <i>Pseudomonas aeruginosa</i> , a nosocomial infection agent with multiple virulence factors	18
Medical relevance and virulence factors	18
(Phospho)lipases as a target of interest	19
2.6 Iron acquisition and CAR proteins in plants	22
Nutrient uptake by plant roots and the relevance of iron	22
Abscisic acid hormonal response and the modulation by CAR proteins	24
3 Scope of the Thesis	27
4 Publications	30

	Reprint permissions.....	30
4.1	PACKMOL-Memgen: A simple-to-use generalized workflow for membrane-protein/lipid-bilayer system building.....	31
	Background.....	31
	PACKMOL-Memgen Workflow.....	32
	Conclusions and significance.....	34
4.2	Structural model of the ETR1 ethylene receptor transmembrane sensor domain	
	36	
	Background.....	36
	Modeling of the ETR1 transmembrane domain.....	37
	MD simulations of free diffusing ethylene show binding in proximity of the putative copper binding site.....	40
	Conclusions and significance.....	42
4.3	Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence.....	44
	Background.....	44
	Molecular dynamics simulations show a preferred orientation of the PlaF monomer in a membrane context.....	45
	Free energy profiles of PlaF dimerization and tilting suggest a monomeric state <i>in vivo</i>	46
	Conclusions and significance.....	50
4.4	Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits <i>Arabidopsis</i> iron acquisition.....	51
	Background.....	51
	EHB1 binds to the membrane surface, fostering its interaction with IRT1.....	52
	Conclusions and significance.....	54
5	Summary and Perspectives.....	57
6	Acknowledgments.....	59

Publication I: PACKMOL-Memgen: A simple-to-use, generalized workflow for membrane-protein/lipid-bilayer system building	60
Publication II: Structural model of the ETR1 ethylene receptor transmembrane sensor domain	82
Publication III: Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence.....	111
Publication IV: Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits <i>Arabidopsis</i> iron acquisition.....	182
7 Curriculum vitae.....	212
Bibliography.....	216

List of Publications

This thesis is based on the following articles:

- I. **Schott-Verdugo, S. (85%)**, & Gohlke, H.
PACKMOL-Memgen: A Simple-To-Use, generalized workflow for membrane-protein/lipid-bilayer system building.
Journal of Chemical Information and Modeling, 2019, 59, 2522–2528.
2018 Journal Impact Factor: 3.966

- II. **Schott-Verdugo, S. (35%)**[§], Müller, L.[§], Classen, E., Gohlke, H.[‡], & Groth, G.[‡]
Structural model of the ETR1 ethylene receptor transmembrane sensor domain.
Scientific Reports, 2019, 9(1), 1–14.
2018 Journal Impact Factor: 4.011

- III. Bleffert, F., Granzin, J., Caliskan, M., **Schott-Verdugo, S. (15%)**, Siebers, M., Thiele, B., Rahme, L., Dörmann, P., Gohlke, H.[§], Batra-Safferling, R.[§], Jaeger K-E and Kovačić, F.[§]
Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence
Submitted to Proc. Natl. Acad. Sci. U. S. A., 2020

- IV. Khan, I., Gratz, R., Denezhkin, P., **Schott-Verdugo, S. N. (15%)**, Angrand, K., Genders, L., Basgaran, R.M., Fink-Straube, C., Brumbarova, T. Gohlke, H., Bauer, P. & Ivanov, R. (2019).
Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits *Arabidopsis* iron acquisition.
Plant Physiology, 2019, 180(3), 1564–1581.
2018 Journal Impact Factor: 6.305

^{§,‡} Authors contributed equally to this work.

Abbreviations

3'-UTR	3' untranslated region
ABA	Abscisic acid
ABC	ATP-binding cassette
ABI1	ABA INSENSITIVE 1
ACC	1-aminocyclopropane-1-carboxylic acid
AHA2	ARABIDOPSIS PLASMA MEMBRANE H(+) -ATPASE ISOFORM 2
AMBER	Assisted model building with energy refinement
APL	Area per lipid
BCA	Bicinchoninic acid
BiFC	Bimolecular fluorescence complementation
BLAST	Basic local alignment search tool
CAR	C2-DOMAIN ABA-RELATED
CASP	Critical assessment of protein structure prediction
CCH	COPPER CHAPERONE
CD	Catalytic domain
CHARMM	Chemistry at Harvard macromolecular mechanics
COM	Center of mass
CTR1	CONSTITUTIVE RESPONSE 1
DCA	Direct coupling analysis
DHp	Dimerization/histidine-phosphotransfer domain
DOPE	1,2-dioleoylphosphatidylethanolamine
DOPG	1,2-dioleoylphosphatidylglycerol
EBF	EIN3 BINDING F-BOX
EHB1	ENHANCED BENDING 1 (same as CAR6)
EIN	ETHYLENE INSENSITIVE
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
ERS	ETHYLENE RESPONSE SENSOR
ETP	EIN2 TARGETING PROTEIN
ETR	ETHYLENE RESISTANT / ethylene receptor
FA	Fatty acid

FRO2	FERRIC REDUCTION OXIDASE 2
GAF	A conserved domain found in mammalian cGMP-binding PDEs, <i>Anabaena</i> adenylyl cyclases and <i>Escherichia coli</i> Fh1A
GAFF	General AMBER force field
GROMOS	Groningen molecular simulation
IRT1	IRON-REGULATED TRANSPORTER 1
MC	Monte Carlo
MD	Molecular dynamics
ML	Machine learning
MSA	Multiple sequence alignment
MST	Microscale thermophoresis
NMR	Nuclear magnetic resonance
PA	Phosphatidic acid
PC	Phosphatidylcholine
PDB	Protein data bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PKC	Protein kinase C
PMF	Potential of mean force
PP2C	Protein phosphatases type 2C
PPM	Positioning of proteins in membranes
PS	Phosphatidylserine
PSICOV	Protein sparse inverse covariance estimation
PSIPRED	PSI-blast based secondary structure prediction
PtdIns	Phosphatidylinositol
PtdIns(4,5)P ₂	Phosphatidylinositol (4,5)-bisphosphate
PtdIns4P	Phosphatidylinositol 4-phosphate
PYL	PYR1 LIKE (same as RCAR)
PYR1	PYRABACTIN RESISTANT 1 (same as RCAR)
RCAR	REGULATORY COMPONENT OF THE ABA RECEPTOR (same as PYR/PYL)
RD	Receiver domain

RMSD	Root mean square deviation
SAM	<i>S</i> -adenosylmethionine
SNF1	Sucrose non-fermenting 1
SnRK2	SNF1-related kinase 2
TM	Transmembrane
TOM1	TRANSPORTER OF MUGINEIC ACID FAMILY PHYTOSIDEROPHORES 1
VMD	Visual molecular dynamics
ZIP	ZRT1/IRT1 like Protein
ZRT	ZINC-REGULATED TRANSPORTER

Abstract

Biological membranes are cell-defining structural components that establish essential boundaries between inner compartments and their environment in all living organisms. Their composition, dynamics and functions are highly complex, not only depending on the lipids that form part of the membrane, but also on proteins that are integral or peripheral to them. The lipid bilayer and membrane proteins regulate fluxes between the separated compartments, as well as the activities of membrane associated enzymes. In addition, the activity and structure of proteins in the membrane environment depend on the interactions with lipids that form the membrane and/or with other membrane proteins.

In the last years, the increase of computational power and the development of computational techniques aimed at investigating membrane systems has enabled the theoretical study of otherwise experimentally difficult targets. In this thesis, three relevant membrane-related systems of *Arabidopsis thaliana* and *Pseudomonas aeruginosa* are studied at the atomic level with state-of-the-art computational methods.

First, to help in the process of setting up membrane and membrane/protein systems for molecular dynamics simulations, PACKMOL-Memgen was developed. It consists of a Python workflow based on open-source programs, which enables to reduce the time dedicated to prepare complex membrane systems by automatically imposing *ad hoc* system dimensions. The program was used throughout this thesis, and has been included in the AmberTools suite of programs. Then, to unveil the structure of the transmembrane domain of the *A. thaliana* ETR1 ethylene receptor, *ab initio* modeling with coevolution-derived protein residue-residue contacts was used. The obtained model was used to assemble a dimeric structure, which in molecular dynamics simulations showed the binding of ethylene in the proximity of included copper (I) ions, exemplifying the first step of the receptor inhibition. Next, the dimer-to-monomer dynamics of PlaF, a novel phospholipase A₁ from *P. aeruginosa*, was studied. Free energy calculations suggest that *in vivo* the monomer tilts on the membrane surface facilitating catalytic access, while changes in the protein concentration and membrane dynamics could regulate its enzymatic activity. Lastly, the IRT1 inhibition by binding of *A. thaliana*'s EHB1 to the membrane surface was studied. Free diffusion of the protein showed a calcium-dependent binding perpendicular to the membrane surface and a subsequent tilting, suggesting a two-step binding process that results in the iron-uptake inhibition.

Zusammenfassung

Biologische Membranen sind Zellbestandteile, die in allen lebenden Organismen die Zelle definieren und von ihrer Umgebung abgrenzen. Die Zusammensetzung, Dynamik und Funktion von Zellmembranen ist kompliziert und hängt nicht nur von den Lipiden ab, die die Membran bilden, sondern auch von integralen und peripheren Membranproteinen. Das Zusammenspiel von Lipiddoppelschicht und Membranproteinen, regulieren den Stoffaustausch zwischen den angrenzenden Kompartimenten sowie die Aktivitäten von membranassoziierten Enzymen. Darüber hinaus hängt die Aktivität und Struktur von Proteinen in der Membranumgebung von den Wechselwirkungen mit Membranlipiden und/oder mit anderen Membranproteinen ab.

In den letzten Jahren haben die Zunahme der Rechenleistung und die Entwicklung neuer Techniken zur Simulation von Membransystemen die theoretische Erforschung experimentell schwierig zu untersuchender Targets ermöglicht. In dieser Arbeit werden drei relevante membranbezogene Systeme von *Arabidopsis thaliana* und *Pseudomonas aeruginosa* im atomaren Maßstab mit aktuelle Berechnungsmethoden untersucht.

Zunächst wurde PACKMOL-Memgen entwickelt, um beim Aufbau von Membran- und Membran/Protein-Systemen für molekulardynamische Simulationen zu helfen. Es besteht aus einem Python-Workflow, der auf Open-Source-Programmen basiert ist. Dadurch kann die Zeit für die Vorbereitung von komplexer Membransysteme reduziert werden, indem automatisch *Ad-hoc*-Systemdimensionen festgelegt werden. Das Programm wurde im Rahmen dieser Arbeit verwendet und in die AmberTools-Programmsuite aufgenommen. Um die Struktur der Transmembrandomäne des Ethylenrezeptors von *A. thaliana* ETR1 aufzuklären, wurde eine *Ab-initio*-Modellierung anhand von durch Koevolution hergeleiteten Aminosäureinteraktionen durchgeführt. Das erhaltene Modell wurde zu einem Dimer zusammengesetzt, das in Molekulardynamik-Simulationen die Bindung von Ethylen in der Nähe komplexierter Kupfer (I)-Ionen zeigte. Dies veranschaulicht den ersten Schritt der Rezeptorinhibierung. Weiterhin wurde die Dimer-zu-Monomer-Dynamik von PlaF untersucht, einer neuen Phospholipase A₁ von *P. aeruginosa*. Berechnungen der freien Energie legen nahe, dass das Monomer sich *in vivo* zur Membranoberfläche neigt, um den Zugang zum katalytischen Zentrum zu erleichtern, wobei Änderungen der Proteinkonzentration und der Membrandynamik die enzymatische Aktivität regulieren könnten. Abschließend wurde die IRT1-Inhibierung durch Bindung von EHB1 aus *A. thaliana* an die Membranoberfläche untersucht. Die freie Diffusion des Proteins zeigte eine calciumabhängige Bindung senkrecht

zur Membran und ein anschließendes Neigen zur Membranoberfläche, was auf einen zweistufigen Bindungsprozess hinweist, der zur Hemmung der Eisenaufnahme führt

Resumen

Las membranas biológicas son componentes que definen a la célula y establecen los límites entre compartimientos intracelulares y sus alrededores en todos los organismos vivos. Su composición, dinámica y funciones son altamente complejas, y no dependen solamente de los lípidos que forman parte de la membrana como tal, sino que también de proteínas que forman parte integral o periférica a ellas. La membrana y las proteínas que la forman regulan flujos entre los compartimientos que separan, como también la actividad de enzimas asociadas a ella. Adicionalmente, las actividades y estructuras de proteínas en la membrana dependen de interacciones con lípidos que forman a la membrana y/o con otras proteínas de membrana.

En los últimos años, el aumento en capacidad de cómputo y el desarrollo en técnicas computacionales dedicadas a la investigación de sistemas de membranas ha permitido el estudio teórico de blancos que serían difíciles de otra manera. En esta tesis, tres sistemas relevantes asociados a membranas de *Arabidopsis thaliana* y *Pseudomonas aeruginosa* son estudiados a nivel atómico con métodos computacionales recientes.

Primero, para ayudar en el proceso de preparación de sistemas de membrana y membrana/proteína para simulaciones de dinámica molecular, PACKMOL-Memgen fue desarrollado. Consiste en un programa basado en Python y programas de código abierto, permitiendo reducir el tiempo dedicado a preparar sistemas de membrana complejos imponiendo dimensiones *ad hoc* de forma automática. El programa fue utilizado a lo largo de esta tesis, y fue incluido en el paquete de programas AmberTools. Después, para revelar la estructura del dominio de transmembrana del receptor de etileno ETR1 de *A. thaliana*, modelado *ab initio* en conjunto con contactos residuo-residuo derivados de información coevolutiva fueron utilizados. El modelo obtenido fue utilizado para formar una estructura dimerica, la cual en simulaciones de dinámica molecular mostraron la unión de etileno en la proximidad de los incluidos iones de cobre (I), ejemplificando el primer paso para la inhibición del receptor. Posteriormente, la dinámica dímero-monómero de PlaF, una nueva fosfolipasa A₁ de *P. aeruginosa*, fue estudiada. Cálculos de energía libre sugieren que *in vivo* el monómero se inclina en la superficie de la membrana facilitando el acceso catalítico, mientras que cambios en la concentración y en la dinámica de la membrana pudieran controlar la actividad enzimática.

Por último, la inhibición de IRT1 causada por la unión de EHB1 de *A. thaliana* a la superficie de la membrana fue estudiada. Libre difusión de la proteína mostró la unión perpendicular dependiente de calcio a la superficie de la membrana y su posterior inclinación,

sugiriendo un proceso de unión en dos pasos que resulta en la inhibición de la absorción de hierro.

1 Introduction

Biological membranes are the essential barrier between the cytosol and the extracellular environment, but also between the lumen of intracellular organelles and the cytosol itself. Despite their evident relevance in defining the boundaries of the cell, only in the last 40 years we have had insights into an approximated structure and composition of the membrane bilayer [1]. Due to their physicochemical characteristics, they restrict the passage of highly polar and charged molecules, allowing to generate membrane potentials essential for cell homeostasis and signaling. They not only protect the cellular components from the external environment, but also compartmentalize essential and sometimes toxic biological reactions [2].

The composition of biological membranes is highly complex, dynamic and diverse, where small changes can have big implications in the cellular function [3]. The general structure of a cellular membrane is a bilayer of a mixture of different types of lipids, with their aliphatic tails generating a hydrophobic core, and their polar heads exposed to the water environment (**Fig.1**). Additional to the lipid components, there are different amounts of intercalated proteins, ranging from 18% in mass in myelin-insulated neurons, to 75% in the mitochondrial membrane and Gram-positive bacteria, with 50% being a representative number overall [4, 5]. Early on, it was noted that membrane proteins are radically different than their soluble counterparts [6]. While globular proteins have a hydrophobic moment that points towards the protein interior, membrane proteins have it pointing towards the protein surface [7].

Probably the most successful way to describe biological membranes is the fluid-mosaic model from Singer and Nicolson [8]. The model assumed a 2D homogeneous arrangement of lipids with integral or peripheral proteins attached to it, freely diffusing across the membrane following Brownian dynamics. This model has been recently updated to include more complex membrane structures and dynamics [9], like the formation of microdomains (e.g. lipid rafts), and interactions with the underlying cytoskeleton to form what has been described as “pickets and fences” [10], which limit the free diffusion of lipids and proteins (**Fig.1**). This highlights how the complexity with which membranes have been described has been increasing dramatically over the years: starting from a picture that represented membranes as rigid structures, and going to ever changing and evolving platforms, where interactions within the

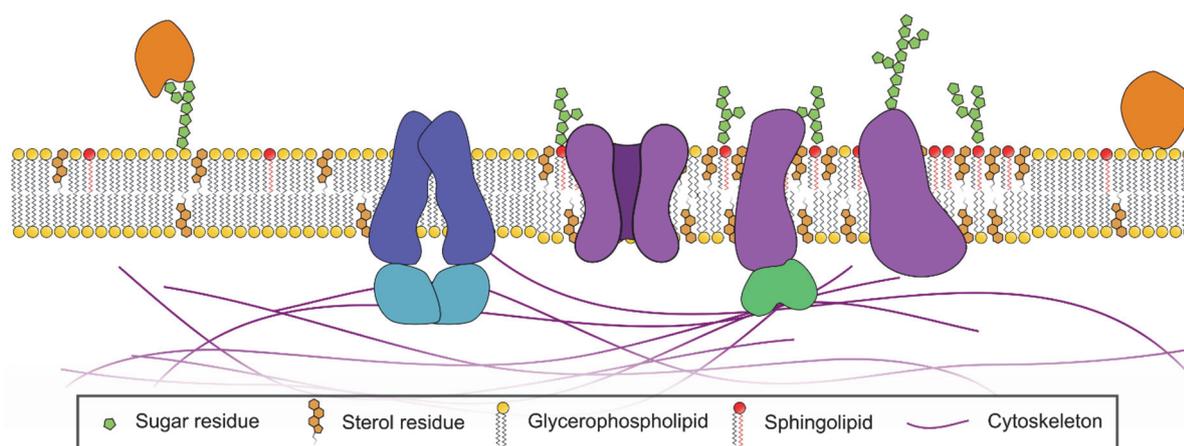


Figure 1. Schematic representation of a eukaryotic membrane bilayer. Membrane bilayers are mainly composed by glycerophospholipids, in some cases by glycolipids, and, in eukaryotes, sterols and sphingolipids. The distribution of lipid types on each leaflet is not necessarily homogeneous, depending on the membrane location. Membrane proteins either are an integral part of it (in blue and purple) or attached to its periphery (in orange). The lateral organization of the lipids in the bilayer is also not homogeneous, and microdomains like lipid rafts are formed, where a higher concentration of sterols, sphingolipids and certain type of proteins (in purple) are to be found. In addition to that, certain proteins (in green) interact with the cytoskeleton of the cell, generating a mesh that supports the membrane, but also restricts the free diffusion of both proteins and lipids (picket and fence model). Adapted from Nicolson [9].

membrane, between the membrane and membrane proteins, and between different membrane proteins are constantly being formed and disrupted.

As the idea of biological membranes as highly dynamic environments gained traction, and the functional relevance of the membrane in cell physiology was evident, methods for molecular modeling and simulations that describe this systems have been developed [11]. These methods allow to describe membrane systems at the atomic level, which is difficult if not impossible to achieve experimentally for complex membrane systems. Systems that are complex and that show dynamic association and dissociation of membrane proteins with other proteins [12] or with the membrane [13] are exactly the ones that require more thorough investigation and are good targets for computer simulations [14]. To help in this, in **Publication I**, a workflow to prepare molecular simulations of membrane or membrane/protein systems was developed. This software resulted in an essential step for the following studies, and has been incorporated into the well-known AmberTools suite of programs [15].

In the present thesis, three independent membrane-associated molecular systems will be presented. All of them have in common a biological membrane as their organizational center for their functions, with static structural information available in literature, but lacking information regarding their structure or structural dynamics in a membrane environment. In **Publication II**, the first model for the transmembrane domain of the ETR1 ethylene receptor from *Arabidopsis thaliana* was obtained. For this, *ab initio* modeling together with residue-

residue contacts were used. The model was experimentally validated, and the copper:receptor stoichiometry determined. Molecular simulations showed for the first time the binding of ethylene to a putative ethylene-binding site. In **Publication III**, the dimer-to-monomer dynamics of the novel phospholipase A₁ PlaF from *Pseudomonas aeruginosa* were studied, identifying through free energy calculations that the protein has a preferred monomeric state *in vivo*, but that it can dimerize upon changes in protein concentration and probably on membrane composition. Finally, in **Publication IV**, the dynamics of the interaction of the peripheral protein EHB1 with the membrane surface were studied. The protein seems to bind in a two-step manner, first responding to an increase in calcium, to then tilt on the membrane surface, where it can regulate the activity of the iron transporter IRT1.

2 Background

2.1 Membrane lipids and composition

The relative composition of different membrane bilayers change in Gram-positive or -negative bacteria and in every eukaryotic cell organelle [16], with more than 21,000 curated lipid species found in nature as of November 2019 [17, 18]. In the model Gram-negative bacteria *Escherichia coli*, 95% of the lipids in the inner membrane bilayer correspond to glycerophospholipids, with about 70% and 25% of them corresponding to phosphatidylethanolamine (PE) and phosphatidylglycerol (PG), respectively, with the rest being mainly cardiolipin (CL) [19]. In the Gram-positive *Bacillus subtilis*, on the other hand, ~85% of the lipids of the plasma membrane correspond to phospholipids [20], which is subdivided in 20% PE, 40% PG, 25% CL and 15% of lysyl-phosphatidylglycerol [21]. In general, glycerophospholipids are the biggest class of lipids present in both prokaryotic and eukaryotic species, while the latter have a higher complexity with additional sterols, sphingolipids and glycolipids (**Fig. 1**). Glycerophospholipids, as the name indicates, have a central moiety of glycerol. Its substitution sites follow the IUPAC “stereospecifically numbered” (*sn*) rule, with fatty acids in positions *sn*-1 and *sn*-2, and a phosphatidyl head group in position *sn*-3 (**Fig. 2A**). They represent the main structural lipids in biological membranes, where the size and level of unsaturation of the fatty acids, and the physicochemical characteristics of the phosphatidyl head group (**Fig. 3**), provide glycerophospholipids with the plethora of shapes and chemical features present in different biological membrane types [3]. From those head groups, phosphatidylcholine (PC) and PE are the most frequent lipids in general, with PC frequently used experimentally as a membrane model [22]. In this lipid class, cardiolipins and phosphatidylinositols highlight for being an essential component of bacteria and mitochondrial membranes related to oxidative phosphorylation [23], and for their relevant role in signaling [24], respectively.

A second major class of lipids are sphingolipids. They are based on a sphingoid base (usually sphingosine, **Fig 2B**), with a fatty acid tail and a head group, which can be a hydrogen, PC or PE, or a sugar, being called ceramides, sphingomyelins or glycosphingolipids, respectively. The chemical structure of the sphingoid base, its hydroxylation, and the usually longer saturated fatty acid chains [25, 26] enhance their lateral interactions with other sphingolipids and sterols (see below). They are a main component of the outer leaflet of the

eukaryotic plasma membrane, forming lipid rafts and potentially organizing and influencing the activity of multiple proteins [26].

A third class corresponds to sterols or steroid alcohols. They are a family of lipids based on a hydroxylated cyclopentanoperhydrophenanthrene (i.e. sterane) nucleus, a three six-carbon and one five-carbon fused ring system, with an aliphatic tail (**Fig.2C**). Similarly to sphingolipids, they are a main component of eukaryotic membranes, particularly in the plasma membrane. They can account for more than a fourth of the total composition [16] and interact with sphingolipids to form lipid rafts. Sterols tend to order and increase the lipid packing in the membrane and establish a liquid-ordered (L_o) phase, where the motion of the hydrophobic core is constrained [26].

There are smaller but relevant classes of lipids that do not fit in the previous groups. Worth mentioning are lipid A and hopanoids, which are essential for Gram-negative bacteria. Lipid A is a main component of the outer leaflet of the outer membrane bilayer. It is composed of at least four acyl chains covalently bound to two phosphorylated glucosamine units, which

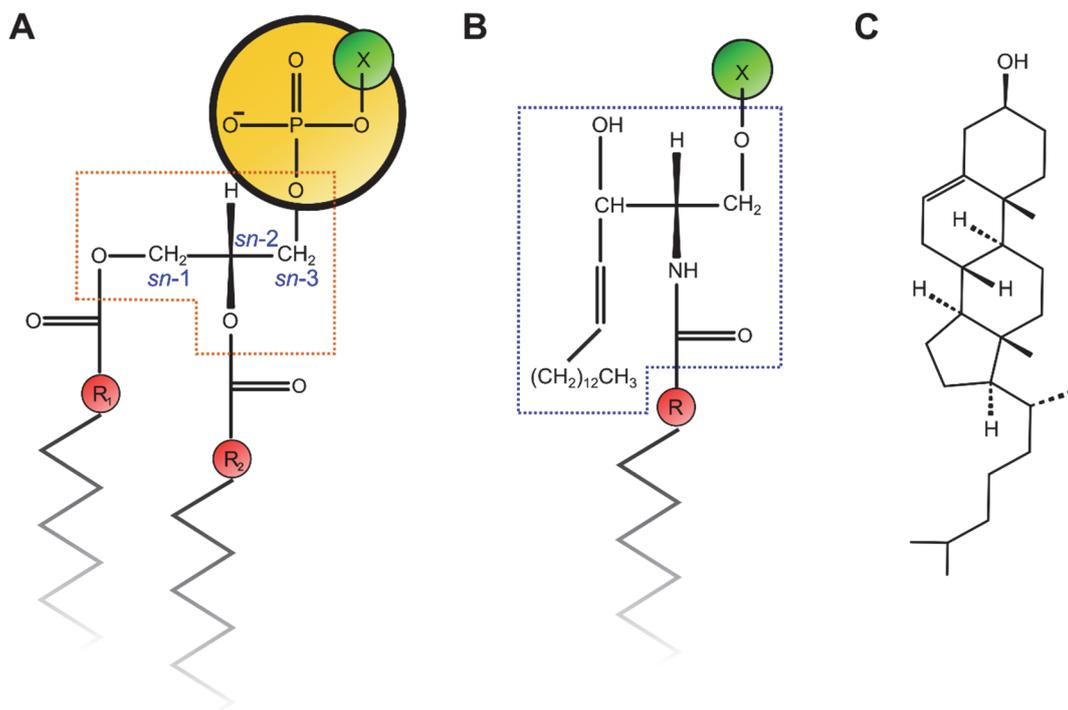


Figure 2. **A)** Glycerophospholipids are the main components of biological membranes. The general structure is a glycerol molecule (orange dashed box), which has fatty acid chains in positions *sn-1* and *sn-2* (R_1 and R_2), and a phosphatidyl group in position *sn-3*. As positions *sn-1* and *sn-2* can have different acyl chains, and there are multiple phosphatidyl head groups that can be added in position *sn-3* (green, X), glycerophospholipids can have very different structures. **B)** Sphingolipids are a main component of eukaryotic plasma membranes, usually formed by a sphingosine residue (blue dashed box), a fatty acid (R) and different head groups (X). Hydroxyl groups of the sphingosine residue establish hydrogen bonds with neighboring lipids, facilitating the formation of membrane microdomains (**Fig.1**) **C)** Sterols are a minor but essential component of eukaryotic membranes, which regulate the formation of microdomains and membrane rigidity. Cholesterol (in the Figure) is the main sterol, but changes in the alkyl tail (bottom) or in the unsaturations in the ring structures (top) are observed in other sterols.

anchors big oligo and polysaccharide structures (namely, the core oligosaccharide and the O antigen). The overall molecules are called lipopolysaccharides, which have been proposed to regulate molecule diffusion in and out of the cell, as well as protein folding and function in the outer membrane [27]. Interestingly, lipid A is also known as endotoxin, as it has a major role in the immune response of animal hosts during invasion [28].

Hopanoids, on the other hand, are not as well studied as the previous lipid classes. With sterols, in general, absent in prokaryotic species, hopanoids seem to fulfill an analogous ordering function in the bacterial environment. They contribute to form domains with lipid A in the outer membrane of Gram-negative bacteria in a comparable amount as if sterols would be added to it [29, 30], though there are reasons to believe the mechanism is different [31].

It is clear that the lipidic composition of biological membranes is complex, and although in different membranes it tends to stay static over time, lipids are in constant traffic through compartments of the cell, homeostatically responding to changes in the environment or maturity of individual cells [32]. The observation that bacteria require a complex network of enzymes to adapt the membrane composition under thermal or pH stress have made the membrane homeostasis pathways an interesting target for potential drug discovery [33]. Additionally, the proteic composition of membranes, as exposed above, is a major component of lipid

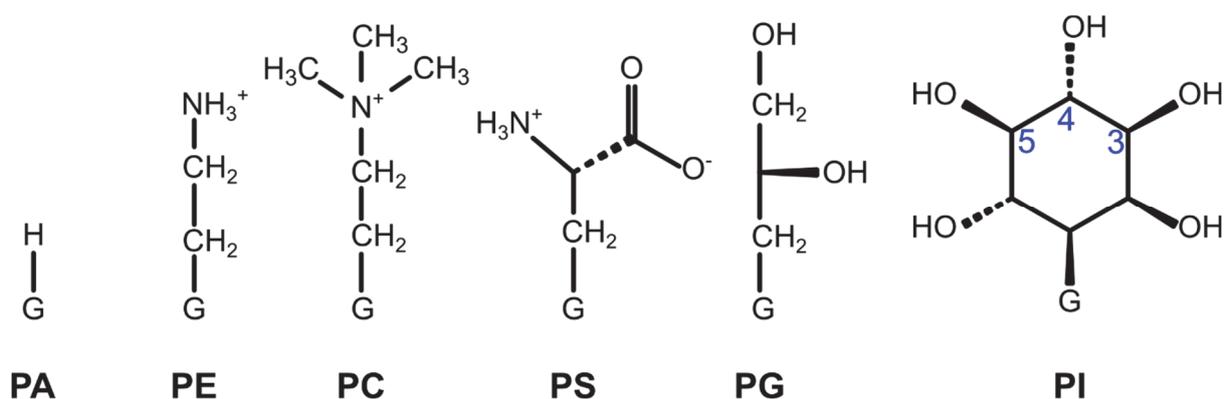


Figure 3. Common head groups of glycerophospholipids found in nature. G represents the phosphate group bound to the glycerol moiety shown in Fig. 2A. PA, phosphatidic acid; PE, phosphatidylethanolamine; PC, phosphatidylcholine; PS, phosphatidylserine; PG, phosphatidylglycerol; PI, phosphatidylinositol. PE and PC are the most common head groups, and have an overall neutral charge with the phosphate group. All the other head groups maintain a formal negative charge. PI can be additionally phosphorylated in positions 3, 4 and 5, adding additional charges relevant for molecular signaling [24].

membranes, adding additional complexity to it. In the next section, how proteins interact with the membrane environment will be described.

2.2 Structure of membrane proteins

The function of biological membranes, aside from being a structural component of the cell, depends on the subcellular compartment and the proteic composition. Different membranes can have functions related to selective permeability, cell metabolism, signal transduction and/or ionic transport [2, 34]. Their function is directly correlated to the proteins that form that particular membrane. It was not until the 1960s that the relevance of proteins in the functions of membranes was evidenced [1]. A good example is the brief publication of Rojas and Luxoro [35], showing that proteins are responsible for ion conduction in squid axons. Given the overall membrane structure described in the previous section, proteins can be placed through the bilayer, attached on the surface, or interact with other components located in the bilayer (**Fig.1**).

The interactions of proteins with the bilayer radically change their overall structural features compared to common water-soluble globular proteins. Where water soluble proteins tend to hide hydrophobic patches and expose polar surfaces to their surroundings, integral membrane proteins invert that relationship, exposing hydrophobic residues to the membrane core and hiding polar residues in the protein core. Additionally, due to the anisotropic structure of biological membranes and the need to reduce exposed polar groups in the membrane core, the amount of possible folds compared to globular proteins is greatly reduced, with in general two types of secondary structure folding cores: α -helical bundles and β -barrels [36]. It has been proposed that these defined secondary structure elements satisfy the majority of the possible backbone hydrogen bonds, reducing the unfavorable interactions with the hydrophobic lipid tails [34]. Single or multi-pass bundles of α -helical structures are ubiquitous in living organisms, while β -barrels are solely found in Gram-negative, mitochondrial and chloroplast membranes [37].

From a reductionist point of view, membrane proteins are not so different than globular proteins in the sense that their folding, as every other molecule, follows thermodynamic laws to reach its proper fold, with the clear difference that the interactions are arranged in a 2D plane rather than in 3D space [38]. Furthermore, not only does the thermodynamic environment of the membrane define the way membrane proteins are folded, but the membrane composition

can have a direct impact on their activity, modulating the fine structure of the protein, or acting as cofactors [39].

The extreme differences between the water and membrane environments make it difficult to use typical protein overexpression and purification protocols on membrane proteins. The hydrophobic regions of membrane proteins tend to aggregate in the cytosol of the expressing cell or during liquid chromatography steps of purification. This forced to include weak detergents and micelles as membrane mimetics, which allows to extract proteins from membranes and to stabilize the hydrophobic surfaces [34, 40]. Overall, solving the 3D structure of membrane proteins through experimental methods has been hampered in comparison to soluble proteins [34]. They are largely underrepresented in the structural databases available for the scientific community (< 1% of the Protein Data Bank (PDB), **Fig. 4**), with just about a thousand unique membrane proteins resolved up until October 2019. That is the case even when they represent 20 to 30% of the proteins in the eukaryotic genome, and they are essential in multiple areas of cell biology [41]. Interestingly, despite the aforementioned difficulties, it is estimated that more than half of the drug targets are integral membrane proteins[42]. This highlights their relevance in molecular pathways that control diverse cell processes and that are involved in disease, but also their importance for the pharmaceutical industry. Because of these reasons, the study of the structure and dynamics of membrane proteins has high importance for understanding cellular mechanisms, and for the development of new treatments for potential diseases.

2.3 Computational methods to study membranes and membrane proteins

In this section, computational tools used to describe and study membrane and membrane protein systems will be briefly described. These tools are the foundation for the studies and results obtained in this thesis.

Molecular dynamics of membrane systems

In the 1980s, along with the determination of the first atomic-resolution transmembrane protein structure [43] and with the overall notion of the Singer and Nicolson fluid-mosaic model [8], computational methods that derived from conventional molecular dynamics (MD) simulations were developed to study the physical environment of biological membranes. The molecular models have grown from small surfactant and lipid simulations with one lipid type and thousands of atoms in the picosecond (ps) range, to all-atom simulations of proteins

embedded in complex bilayers with 10^5 - 10^6 atoms for microseconds (μ s). This has been further extended to coarse-grained models that can describe tens of thousands of lipids for μ s of simulation [11]. MD simulations of lipid systems thus present an approach to study membrane environments and converge theory and experimental results. These simulations provide an atomistic view of experimental phenomena, and broaden what pure experiments can resolve, especially considering the difficulties of performing wet-lab experiments with membrane proteins and membranes themselves [11, 40].

For atomic MD simulations of membrane environments, proper parameters that describe the physical characteristics of the lipids and proteins involved have to be determined. Multiple force fields describing such systems have been developed throughout the years [22], where CHARMM [44, 45], Slipids [46, 47], GROMOS [48, 49] and AMBER [50, 51] are general biomolecular force fields worth mentioning. In general, the development of these force fields has evolved throughout the years, starting at the beginning of the 2000s with simulations that required constraints in the area of the membrane patch or on the surface tension to reproduce experimental data. In the last decade, with the inclusion of quantum mechanically determined charges and multiple corrections in the acyl chain torsion parameters, these force fields reached the point where simulations in the isothermic-isobaric (NPT) ensemble can be performed without the inclusion of additional constraints, as expected in membrane simulations [22, 52].

In particular, in the AMBER simulation package, the lipid force field development is a relatively new endeavor. Initial attempts to have lipid parameters in AMBER date from the 2000s, and were modifications of the small molecule General Amber Force Field (GAFF) [53, 54]. However, a more refined and modular parametrization was achieved in the beginning of the 2010s in Lipid11 [50] and GAFFlipid [55]. The lessons learned from these attempts were applied to develop the tensionless Lipid14 force field [51], and its extension in Lipid17 (not formally published as of 2020, [15]), where Lennard-Jones and torsion parameters, together with the atomic charges were revised to reproduce experimental data in the NPT ensemble. This family of force fields have shown to reproduce experimental data of membranes [56] and of membrane interactions with proteins on par or better than previously developed force fields [57].

Aside from the parameters, getting the initial structure of a reasonable membrane patch also has been a challenge, particularly with the increasing size of the simulation systems. Multiple methods allow to obtain a reasonable initial structure to perform MD simulations, with CHARMM-GUI being probably the most used option for all the simulation packages previously

described [58]. However, no dedicated option for the AMBER package existed until now, forcing researchers to depend on the web interface provided by CHARMM-GUI. A new software to generate initial structures of membrane systems for the AMBER package was developed in this thesis. This software enables the preparation of molecular simulations directly in the command line together with the AmberTools software package; this will be discussed further in this thesis (see **section 4.1, Publication I**).

Free energy umbrella sampling calculations

In molecular dynamics simulations, as the force fields (like the ones introduced above) describe the molecular interactions of the system, their statistics can be evaluated. As the probability of having a determined molecular structure is representative of the free energy of that structure, the sampling statistics should lead to the measurement and prediction of chemically relevant free energy values [59]. One of the caveats though, is that the sampled structures could not be representative of the process of interest. On one part, the sampling will be biased by the initial structure used in the molecular sampling, and on the other, the timescale in which the process under study occurs could be too long to be observed in the performed simulations. To study such long-term molecular processes, the molecular sampling can be enhanced by modifying the potential energy surface. If the free energy of the process is of interest, this changes in the potential energy are later taken into account to calculate the effective energy difference between the sampled structures [60].

One of such techniques is called umbrella sampling. In this method, structures that are representative of a so-called reaction coordinate (or of the transition between two end-points of interest) are sampled in harmonically restrained molecular dynamics simulations. The harmonic restraints allow to sample the reaction coordinate by applying an “umbrella”-shaped potential on top of the potential energy landscape. The objective of such a sampling approach is to obtain structures that reflect the potential energy landscape between the start and end point of interest. The distribution in the restrained umbrellas unveils information regarding the underlying potential energy surface, allowing to mathematically calculate a “potential of mean force” (PMF) [61]. This method has been used to calculate free energies of dimerization of proteins in a membrane environment [62, 63], and was used in this thesis to calculate the dimer-to-monomer dynamics of PlaF, a bacterial phospholipase A₁ (**section 4.3, Publication III**).

Structure prediction of membrane proteins

The huge increase in the amount of available protein sequences far exceeds the amount of resolved protein structures, which is even more dramatic for the membrane protein case (Fig. 4). As mentioned in the previous section, the experimental determination of membrane protein structures remains particularly challenging, with complications both in their purification and crystallization. As an alternative, the structural prediction of membrane proteins has gained an important role to study their structure and understand their functions [34, 40].

In general, there are two approaches for predicting a protein structure: I) comparative or homology modeling and II) *ab initio* (sometimes called *de novo*) structure prediction. Homology modeling is based on the structural conservation observed across proteins with diverse origins but a common ancestry, as long as their amino acid sequences have an identity $\geq 30\%$. In general, if a structure that fulfills that requirement has been already resolved in the PDB (Fig. 4), it can be used as a template and a proper model can be generated [64, 65]. Though these methods have been widely used to model soluble proteins, they have been validated for transmembrane proteins as well [66]. *Ab initio* methods, on the other hand, do not depend on previously determined protein structures directly, but rather on conformational searches of energy minima on a potential energy landscape that considers the native protein environment. These methods are especially interesting for the transmembrane protein case, given the limited

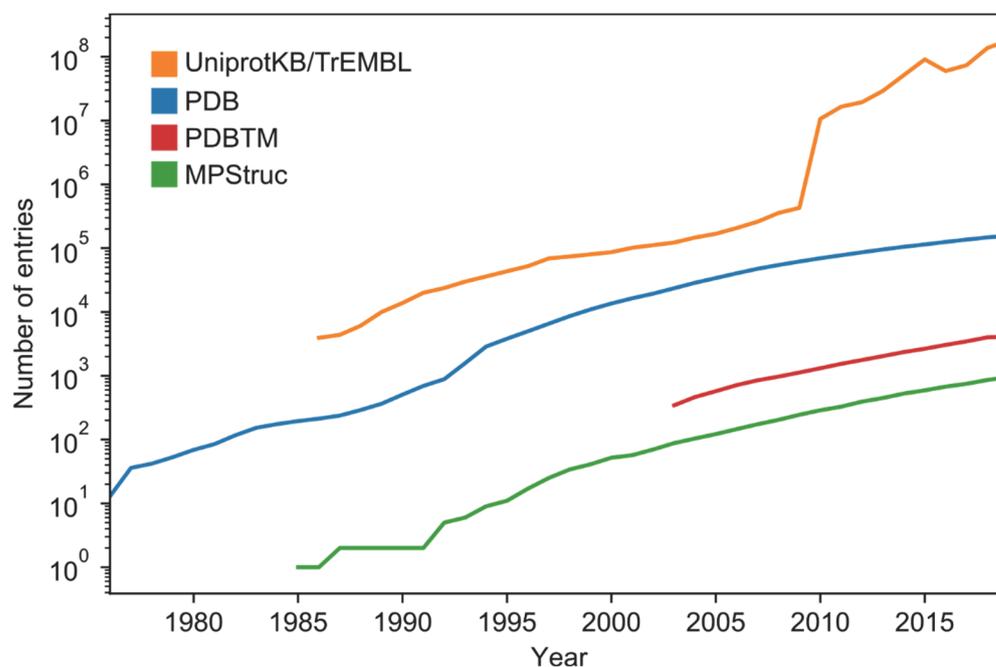


Figure 4. Number of entries in the automatically annotated protein sequence database (UniprotKB/TrEMBL), the protein structure database (protein data bank, PDB), the transmembrane PDB database (PDBTM) and the unique membrane protein structure database (MPStruc). The reduction observed in the number of sequences in UniprotKB/TrEMBL corresponds to the removal of redundant sequences of repeated entries or within groups of highly homologous organisms. The values correspond up until September 2019.

availability of proper structural templates (**Fig. 4**). The energy functions used to perform the structural search are either physics-based (as the ones used in MD simulations, see above), or specially developed knowledge-based energy functions. At the same time, these functions are usually associated with specific ways of performing the conformational search, from MD simulations to Monte Carlo (MC) protein-fragment assembly search. The former have had limited success due to the vast phase space of proteins, in general limiting its application to proteins of ≤ 100 residues and to the use of implicit solvent models to achieve enough sampling [67, 68]. Methods especially developed for membrane proteins have shown some success, particularly for small α -helical assemblies [69, 70].

Probably the most common *ab initio* alternative is the protein-fragment folding strategy used in FRAGFOLD [71] and in the Rosetta package [72], with the latter being widely used nowadays. There are multiple protocols that can be used within Rosetta, but a general approach is to search 3- and 9-mer fragments homologous to the target in the PDB, and use their secondary structure and side chain conformations to significantly reduce the search space, which at the same time reduces the computational cost of the folding process [73]. These fragments are later assembled using a MC search with the Rosetta knowledge-based potential energy function. This function can be a full-atom energy function, comparable to force fields used in MD [74], or coarse-grained per residue, reducing the ruggedness of the potential energy landscape and increasing the sampling speed even further [75]. As the protein folding can get stuck in multiple energy minima even with a coarse-grained potential, it is common that tens to hundreds of thousands of models (usually called decoys) are generated per target of interest, where one of them has to be selected as a representative model [73]. These methods have been further developed to model membrane proteins in particular, by including an artificial membrane slab, potential terms for the interactions of residues in the membrane environment, and additionally predicted information such as secondary structure, transmembrane topology and solvent accessibility [76, 77].

Protein residue-residue contact predictions

Though the *ab initio* methods have shown their usefulness during the last two decades of development, they are computationally expensive and impractical in cases of big proteins (> 300 residues) [68]. Early on, it was noted that just using native contact maps, the 3D structure of a protein could be recovered completely [78]. By combining the *ab initio* sampling strategy and including sparse experimental data, such as NMR or EPR restraints, the folding process

becomes not only feasible, but allows to reach atomic accuracy [73, 79]. Getting experimental measurements is not always possible though, as obtaining NMR or EPR restraints for transmembrane proteins share many of the difficulties exposed for purification and crystallization. Coevolutionarily determined residue-residue contact predictions, on the other hand, rely on new sequencing techniques and on the huge amount of accumulated protein sequences in the Uniprot database (**Fig. 4**). They make indirect use of experimental results, and can be applied to drive the search of a proper model within the generated decoys. The predicted contacts can be used analogously to sparse experimental restraints in *ab initio* structure predictions. They can restrict the sampling search space by imposing distance restraints between the predicted contacting residues, or to select the generated decoys that fulfill the predictions the best [40, 68]. Interestingly, it has been independently shown, that 8% of the essential native contacts or $\geq 30\%$ of randomly selected contacts are enough to fold a protein [80], and that one correct and nonlocal predicted contact every twelve residues is enough to obtain near to atomic protein structures [81].

The typical approach to predict residue-residue contacts consists in measuring coevolution signals throughout multiple sequence alignments (MSAs). The basic idea is that, as single positions in proteins mutate (i.e. evolve), evolutionary pressures against loss-of-function compensate for the change in the physicochemical features of the original residue by changing interacting residues to match the new physicochemical environment (**Fig. 5**) [82]. As the second change is a consequence of the first, they are correlated and can be measured directly from changes in the MSA [83]. Such pairs of co-evolving residues indicate positions where spatial proximity is likely. One caveat of this approach is that two correlated positions are not necessarily caused by a direct physical interaction, which is emphasized in transitive correlations, i.e., when a signal between positions A and C is obtained as a consequence of the interaction with a third intermediate position B (“chaining effect”, $A \leftrightarrow B \leftrightarrow C$) [84]. Put in a different way, these methods account for the probability of two positions affecting each other, neglecting the influence of the rest of the protein positions. Because of the noise produced by this lack of causation and the corresponding high rate of false positive signals in the predicted contacts, their usage was rather limited and unsuccessful to predict 3D protein structures *ab initio*. This changed in the last decade with the development of algorithms like DCA (direct contact analysis) entropy-maximization [85, 86] and PSICOV (protein sparse inverse covariance estimation) [87], which consider the whole protein statistics to derive the contact probabilities between two positions, effectively decoupling indirect residue-residue interactions

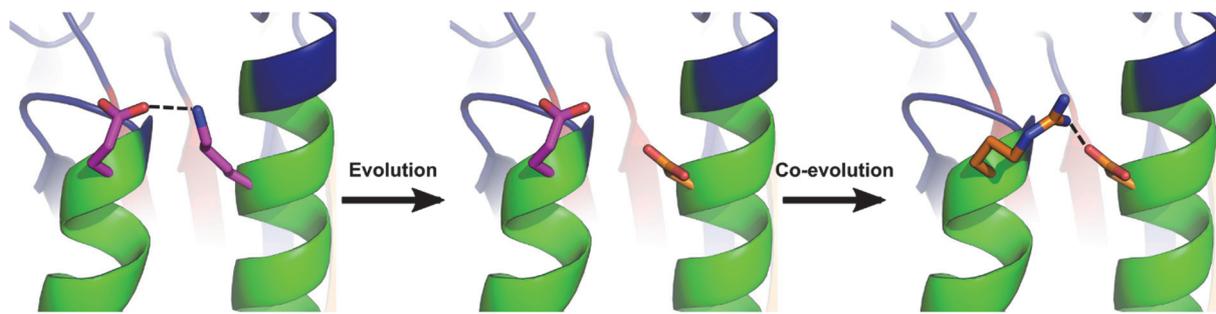


Figure 5. Schematic representation of coevolutionary changes in protein structures. With evolution, as single residue mutations occur, interactions of the original residue with the rest of the structure are affected, risking losing the overall protein function. Coevolutionarily, mutations that re-establish the original interaction network and maintain the overall protein fold are favored, generating identifiable correlations in multiple sequence alignments.

from the contact signals [84]. This resulted in multiple implementations based on these concepts to accurately predict protein contacts [85-90], not only intra- but also inter-protein [91]

In addition to these developments, the last four years have signified a quantum leap in protein contact predictions. The current rise of machine learning (ML) methods, the easier access to software and hardware for their implementation, and the application of such methods to contact predictions, have had an impact by doubling the top L/5 (where L is the length of the protein) prediction accuracies from the 11th to the 12th Critical Assessment of Structure Prediction (CASP) experiments [92]. The same trend was evidenced in CASP13, where the accuracies increased up to 70% in the same range [93]. The usage of ML has been determinant in being able to obtain a higher quality of contacts, especially in cases where the number of sequences in the MSA is subpar. The ML methods used include most notably deep neural networks, which with convolutional and residual neural network architectures have shown the best results to date [94-99].

Just to mention some examples, the tools here mentioned have been used to model the N-terminal portion of the human dopamine transporter [100] and the human alkylglycerol mono-oxygenase [101]. An interesting case has been the prediction with coevolutionary information of the structure for the human zinc transporter hZIP4 [102], which was later shown to be highly similar to a crystalized homolog [103]. A similar modeling approach was used in this thesis (**section 4.2, Publication II**). Interestingly, hZIP4 showed its relevance in a second project studied in this thesis, unrelated to the first case (**section 4.4, Publication IV**).

2.4 Ethylene-induced hormonal response and signaling pathway

Ethylene: the first identified gaseous hormone

At the beginning of the 1900s, Neljubow [104] and Crocker, Knight, Harvey and Doubt [105-108] showed that smoke and “illuminating gas” in laboratory conditions, and particularly the ethylene in it, acted as a phytohormone on plant seedlings. The phenotype produced by ethylene, known as the seedling triple response, is evidenced by modifications in the growth process, causing a shortening, curling, and thickening of the plant stem, a response that was proven to be specific to ethylene [109]. In addition to the triple response, Cousins reported in 1910 that fungi-damaged oranges emanate a gas that accelerates the banana ripening process, suggesting that they should be separated during transport [110, 111]. It was not until the 1930s that Kidd, West and Gane directly linked ethylene to the ripening process of some fruits, and suggested that they are able to generate ethylene themselves [112-114]. They were the first to describe that in some fruits the ripening process is accompanied by a rapid increase of the respiratory activity, coined by Kidd and West as the climacteric [114]. This term is currently used to group produces that are particularly sensitive to ethylene. Overall, the discovery of ethylene activity was controversial, considering that it was the first described gas with hormone features. This has the implication that its range of action is not restricted only to the synthesizing organism, but that it can diffuse and communicate signals between plants, and have the aforementioned effects [110].

The synthesis of ethylene is dependent on the methionine cycle, also known as the Yang cycle. *S*-adenosylmethionine (SAM), an intermediary commonly involved in methyl group transfer reactions, can be transformed into 1-aminocyclopropane-1-carboxylic acid (ACC) by the enzyme ACC synthase. This step is described as the first committed step to produce ethylene. One further oxidation step of ACC, catalyzed by ACC oxidase, yields ethylene [115]. Interestingly, the activity of both ACC synthase and oxidase has been linked to other hormones, to stress, and to the ripening responses of the plants, explaining to a large degree the effects observed on produces [115, 116]. More important, however, is that ethylene establishes a positive feedback loop in climacteric fruits, where a produce that ripens emits ethylene, triggers the ethylene signaling cascade and results in an enhanced expression of ethylene receptors and its biosynthetic enzymes [117].

Fruit ripening in economy: ethylene signaling as a biotechnological target

Fruit and vegetable ripening has a huge social and economic impact, making their storage and transport throughout the world a sensible point of the food supply chain. It is estimated that over 30% of all produced foods are wasted along the food supply chain. From that, 30 to 50% occurs postharvest or in the distribution stage. Proper access to energy sources has shown to be the main differentiating factor in the amount of postharvest losses [118]. On one side, better access to energy sources would reduce the wastage of multiple products, but on the other, an increase in energy production is usually associated with an increased generation of greenhouse contaminants. Taking the banana, the most traded horticultural product as an example, about 45% of the costs involved in placing the fruit into a supermarket is related to proper transport, storage, and ripening treatment [119]. “Cool chain management” is one of the reasons behind those high costs, i.e., transporting and keeping fruits at temperatures from 13 to 16 °C throughout the whole distribution chain, and in this way reduce the ripening of climacteric fruits [120]. It has been estimated that in Europe about 30% of the energy used in the food production sector corresponds to cooling and freezing, a remarkable amount compared to any other production area [121]. Low temperatures retard the ripening process of fruits, partially because it inhibits the production of ethylene [122-124], slowing down the onset of the whole ripening process.

Interestingly, an alternative to active refrigeration during transport are strategies that reduce the ethylene concentration in the air or inhibit its activity on fruits [120, 125]. For example, apples, kiwis, pears, strawberries and bananas display a relevant delay in the ripening process by maintaining a controlled atmosphere during transport and storage [126]. However, this implies expensive systems to either filter or maintain certain air composition. A different approach is to inhibit the production and/or the effects of ethylene on plants, where the latter have the advantage to additionally block effects of exogenous production from, for example, gas combustion. The main commercially available example for this is 1-methylcyclopropene (1-MCP) [127], which has been shown to work by competitively blocking the ethylene receptors. The response to treatments with 1-MCP in different commodities has been shown to delay the onset of the ripening process, showing improved responses on longer treatments or by reducing the storage temperature alongside its application [128]. The commercial usage of 1-MCP has implied a higher flexibility for cold storage temperatures during transport, and an increased shelf life of many commodities, though the response is not consistent on all climacteric fruits [129]. This highlights the relevance of understanding how the response

triggered by ethylene progresses at the molecular level. Interactions of the different components of the ethylene-signaling cascade can provide new strategies to maintain produces over long periods of time. The next subsection will describe the ethylene receptors and the information available up until now regarding the molecular responses to ethylene in plants.

Ethylene receptors and the ethylene-signaling pathway

More than 80 years after the first hints pointed towards ethylene as a plant hormone, Bleecker identified a gene which, when mutated, made *Arabidopsis thaliana* seedlings resistant to the treatment with the hormone and did not show the expected triple response [130]. The ETHYLENE RESISTANT 1 (also alternatively ETHYLENE RESPONSE) protein, widely known as ETR1, and later its homologs ETHYLENE RESPONSE SENSOR 1 (ERS1), ETHYLENE INSENSITIVE 4 (EIN4), ETR2, and ERS2 [131], were identified to be responsible for the physiological responses to ethylene in plants. The receptors form transmembrane dimers, fostered by disulfide bridges in the N-terminal region of the protein [132], followed by a cytosolic GAF domain [133]. Aside from the N-terminal and GAF domain regions, the protein resembles a two-component system histidine kinase (HK), similar to the ones widely described in bacteria [134], and can be subdivided in dimerization/histidine phosphotransfer (DHp), catalytic (CD) and receiver (RD) domains.

Though all receptors in *Arabidopsis thaliana* share the same overall structure, there are differences that allow to classify them in two subfamilies (**Fig. 6**) [135]. ETR1 and ERS1 correspond to subfamily 1, which has three transmembrane helices in the N-terminal binding domain and active histidine kinase catalytic domains. Interestingly though, the histidine kinase activity is not required for triggering the main signaling cascade [136]. ETR2, EIN4 and ERS2, on the other hand, correspond to subfamily 2, which has an additional transmembrane helix in the N-terminus, and a degenerate histidine kinase domain with serine kinase activity. It is still a matter of debate if the first transmembrane helix is a signal peptide or remains as a structural component in the active receptors [137-139]. It is relevant to mention that both ERS homologs lack the receiver domain altogether.

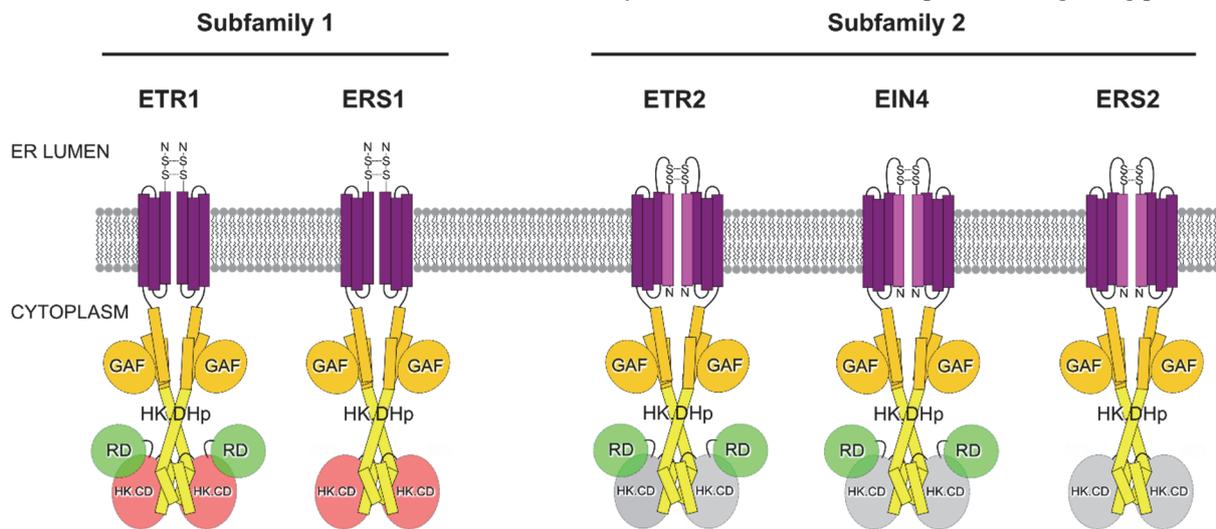


Figure 6. Schematic representation of the family of ethylene receptors found in *Arabidopsis thaliana*. The receptors are found in the membrane of the endoplasmic reticulum, with the two-component system domains exposed to the cytoplasm. All members of subfamily 2 have an additional signal peptide/transmembrane helix (magenta) and a catalytically inactive histidine kinase (HK) catalytic domain (CD). Both ERS1 and ERS2 receptors lack the receiver domain (RD). DHp, dimerization/histidine-phosphotransfer.

Regardless that all receptors respond to ethylene and have high similarity, they have both overlapping and divergent functions. They have been proposed to be differentially modulated by the receptor phosphorylation and homo-/hetero-oligomerization with other ethylene receptors and other ethylene cascade proteins [139-141]. The receptors are non-canonically localized in the endoplasmic reticulum (ER) membrane, instead of the usual hormone receptors in the plasma membrane [142]. Conveniently, as ethylene is a gaseous olefin, it is able to cross cellular membranes and reach to the receptor despite the intracellular localization.

Under conditions without ethylene, the receptors are constitutively active, which implies activating a downstream serine/threonine kinase, CONSTITUTIVE RESPONSE 1 (CTR1). CTR1 phosphorylates ETHYLENE INSENSITIVE 2 (EIN2), which labels the protein for binding and ubiquitination by the F-box proteins EIN2 TARGETING PROTEINS 1 and 2 (ETP1/2) and for the subsequent degradation by the ubiquitin proteasome pathway (**Fig. 7**, left) [143, 144]. Without EIN2, EIN3, a downstream signaling component, is targeted by the EIN3 BINDING F-BOX PROTEIN 1 and 2 (EBF1/2) and further degraded by the proteasome [145]. Under an ethylene atmosphere, on the other hand, the receptors are deactivated, inhibiting the kinase activity of CTR1. Without the phosphorylation of EIN2, a still unidentified protease cleaves the C-terminus of the protein, allowing its diffusion in the cytosol and its translocation to the nucleus, where it stabilizes EIN3 and triggers the main ethylene responses (**Fig. 7**, right). It has been shown that EIN2 localizes to P-bodies and binds to the 3'-UTR of EBF1/2,

repressing their translation [146], while also regulating the histone acetylation of regions in the chromatin that code for ethylene triggered responses [147].

Although it is known that the binding of ethylene triggers the inactivation of CTR1, the structural mechanism for the inactivation of the receptors remains unknown. It has been suggested that the binding of ethylene to the transmembrane domain of the ethylene receptors causes a conformational change independent of the receptors kinase activity. This would be transmitted to the cytosolic domain and result in the CTR1 kinase inactivation. Multiple cytosolic domains of the ethylene receptor have been resolved crystallographically [148, 149], but the transmembrane and GAF domains remain experimentally undetermined. In the latter case, homology modeling has resulted in a feasible structure, leaving the transmembrane domain and its connections to the GAF domain as the sole part of the protein without a structural model [150]. Binding of ethylene to the N-terminal transmembrane domain is supported by the copper cofactor in the binding site required by all ethylene receptors to perform their function. Interestingly, this was suggested many years before the discovery of the receptor proteins [151-

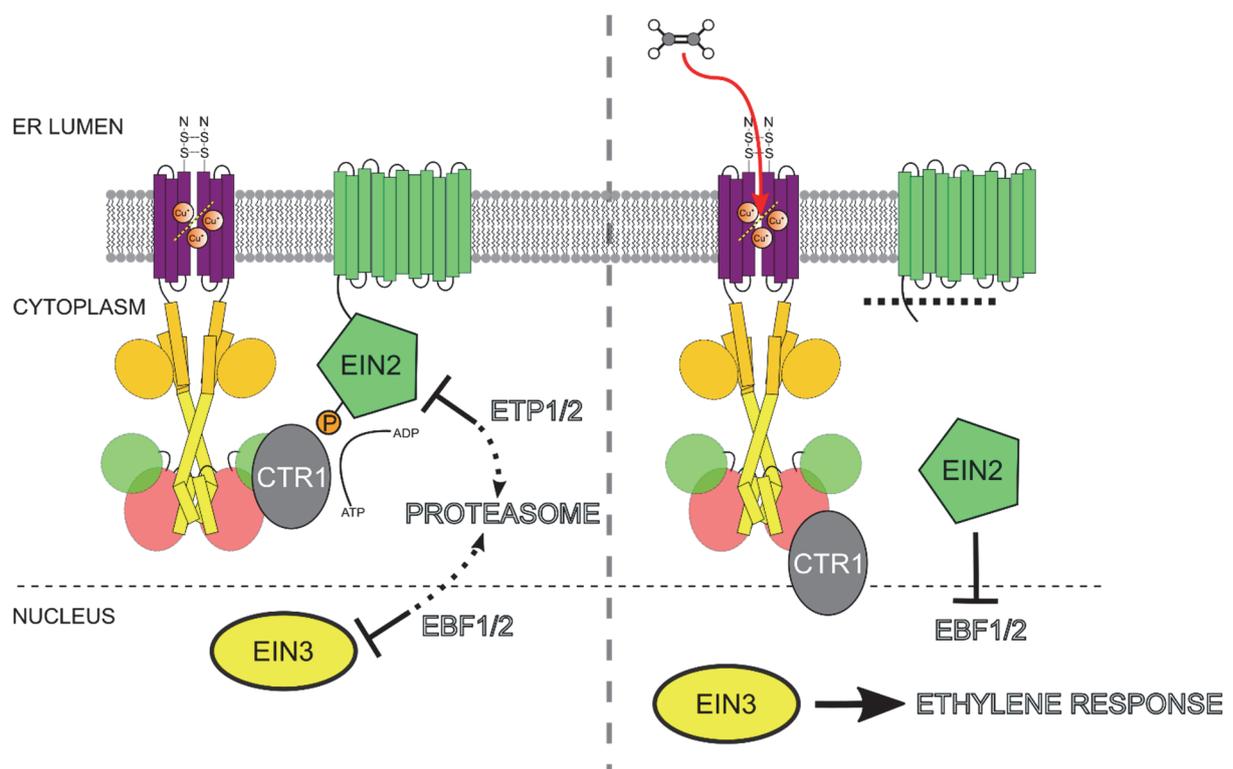


Figure 7. Inhibition of the ethylene receptors upon ethylene binding, and the subsequent activation of the ethylene-signaling cascade. The ethylene receptors are constitutively active, serving as a scaffold for CTR1, which phosphorylates EIN2. This triggers the degradation of EIN2 by the ubiquitin proteasome pathway. EIN3, lacking stabilization by EIN2 is also degraded. When ethylene binds to the receptor, the activity of CTR1 is inhibited. Without the phosphorylation, the C-terminus of EIN2 is cleaved by a protease and diffuses into the nucleus, protecting EIN3 from degradation by the proteasome and triggering the ethylene responses.

153]. The copper ion is expected to be in a +1 oxidation state and to interact with the ethylene double bond by a Dewar-Chat-Duncanson type of interaction [154], though whether one copper ion per dimer or one per monomer are required has remained a matter of discussion [151, 155, 156]. The requirement of the copper cofactor implies that the ethylene-signaling pathway depends on copper chaperones and transporters. Recently it was shown that, while well-described copper chaperones like ATX1 and CCH directly interact with ETR1, the P_{1B}-type ATPase RESPONSIVE TO ANTAGONIST 1 (RAN1) seems to be the main provider of copper for the ethylene receptor after the protein is loaded by the aforementioned chaperones itself [157].

Despite residues of the transmembrane domain have been related to the binding of copper or to be essential for the ethylene response, giving approximate information about the environment of the copper and ethylene-binding sites, no three dimensional structure of the transmembrane region of the receptor has been made available [137]. In this thesis, the first structural model for the transmembrane domain of the ethylene receptor ETR1 was generated (**section 4.2, Publication II**). This, together with the determination of the stoichiometry of copper binding and tryptophan mutants of the transmembrane domain, enabled us to perform molecular simulations of free diffusing ethylene into the protein, illustrating the initial steps for the receptor inactivation.

2.5 *Pseudomonas aeruginosa*, a nosocomial infection agent with multiple virulence factors

Medical relevance and virulence factors

Pseudomonas aeruginosa is an opportunistic Gram-negative pathogenic bacterium. It is versatile, being able to survive on soil and aquatic environments, and to colonize diverse multicellular eukaryote species such as plants and humans [158]. The bacterium causes local and systemic infections. It has become a frequent cause of nosocomial infections in immunocompromised patients, being responsible for high mortality rates, even after antibiotic treatment [159, 160]. For instance, the latest United States National Healthcare Safety Network report for healthcare-associated infections lists *Pseudomonas aeruginosa* with 8% of the reported cases as the 4th most prevalent infection, only behind *Escherichia coli*, *Staphylococcus aureus* and *Klebsiella* spp. [161]. The emergence of resistant and multi-drug resistant strains of *Pseudomonas aeruginosa* leads to a consequent increase in the average treatment cost, where the change in antibiotics alone can result in up to six times more expenditure in pharmacy [162]. For these reasons, as of 2017, *P. aeruginosa* has been classified by the World Health Organization as a desirable target for the development of new antibiotics [163], ranking only below *Acinetobacter baumannii* as the most concerning antibiotic-resistant pathogen [164].

The success of *P. aeruginosa* in surviving in different environments and its ability to infect multiple hosts resides in the diverse virulence factors available in different strains of the bacteria. The major types of virulence factors include flagella and pili, secretion systems like the Type 3 secretion system (3SS), biofilm and lipopolysaccharide formation, quorum sensing, and the synthesis of enzymes such as proteases and (phospho)lipases (**Fig. 8**) [158]. Additionally, the bacteria have resistance mechanisms such as efflux pumps and β -lactamases that confer multidrug-resistance. These factors can additionally be transferred between bacterial strains by movable DNA elements. Studying each of the previously mentioned virulence factors is an essential step in developing new potential treatments against pathogenic bacteria.

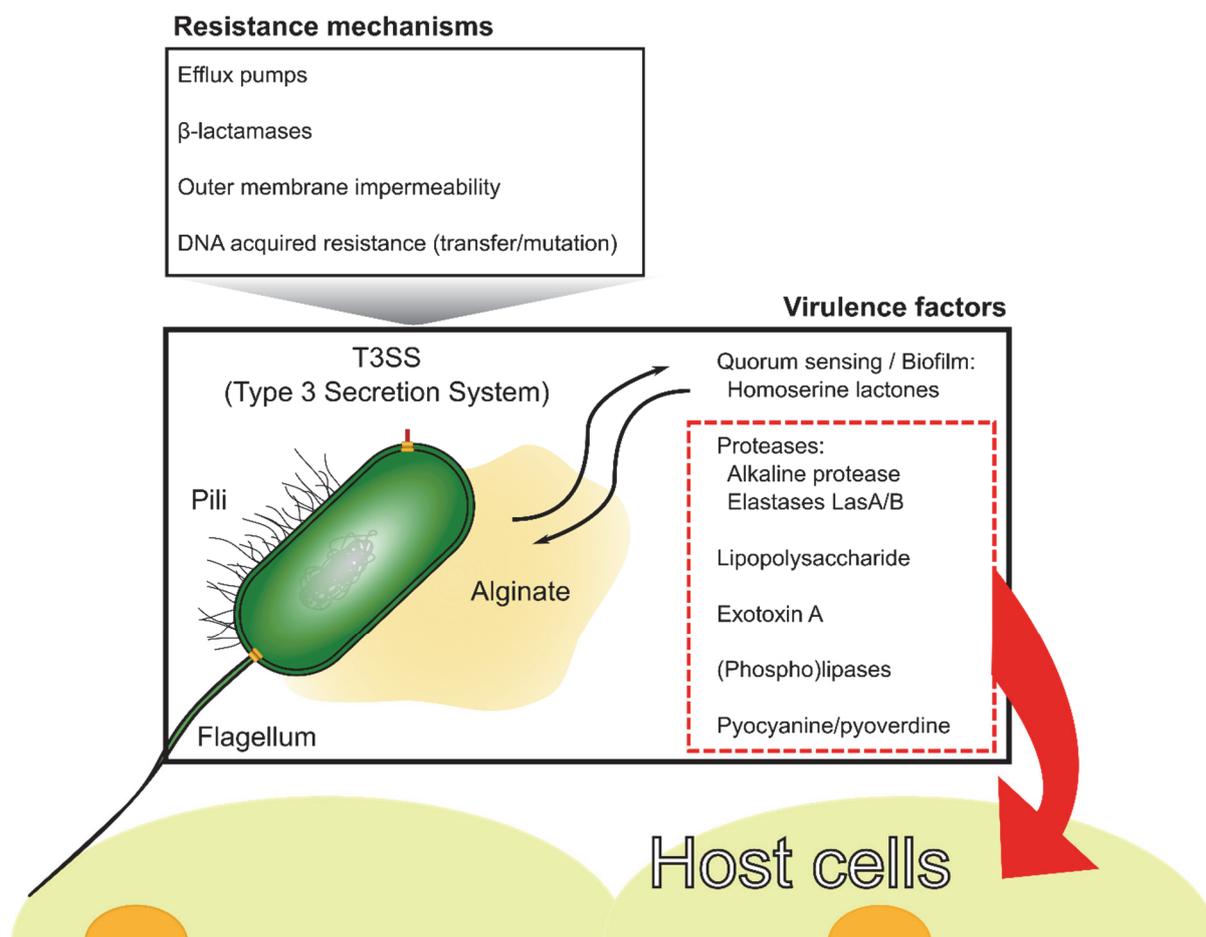


Figure 8. *Pseudomonas aeruginosa* resistance mechanisms and virulence factors. *Pseudomonas aeruginosa* is a versatile pathogen, mainly due to the flexibilities conferred by mechanisms to resist antibacterial treatments and multiple virulence factors. The bacteria can have specialized mechanisms that confer resistance towards antibiotic such as efflux pumps and β -lactamases, which can also be acquired through DNA transfer or modified by mutations, and a lower outer membrane permeability. In addition, *P. aeruginosa* can have mechanisms to inject toxins, like the type 3 secretion system, motility and adhesion structures, such as a flagellum or pili, and secrete biofilm thickening agents like alginate. The cells are able to adapt to surrounding bacteria and regulate biofilm formation by secreting and sensing homoserine lactones, but they are also capable of secreting molecules able to target and damage the host cells (red dotted box). They are able to generate proteases, modulate immune responses with lipopolysaccharides, block protein synthesis with exotoxin A, and directly damage host-cell membranes with lipases and phospholipases. On top of that, they can regulate iron homeostasis and oxidative responses with pyocyanine and pyoverdine. Figure adapted from Gellatly and Hancock [158].

(Phospho)lipases as a target of interest

From the previously mentioned virulence factors, lipases and phospholipases are carboxylesterases (EC 3.1.1) or phosphodiesterases (EC 3.1.4) that are common throughout diverse pathogenic organisms. They are usually involved in degrading the membrane of the host cells, causing general cellular damage, as well as blocking defense mechanisms from the host organism [165, 166]. They are not only interesting medically though [167], and applications in the generation of lysophospholipidic surfactants for the industry and in phospholipid composition analysis have been devised [168].

Phospholipases in particular can be classified depending on the ester bond they enzymatically cleave, where PLA₁ and PLA₂ cleave in positions *sn*-1 and *sn*-2, respectively, PLB cleaves both, and PLC and PLD cleave ester bonds from the phosphate group (**Fig. 9**) [165, 169]. From these, PLA₂, PLC and PLD are probably the most widely studied and understood, while for PLA₁ only a limited number of enzymes have been described. Until recently only secreted PLA₁ bacterial enzymes were characterized in *Serratia* spp. and *Yersenia enterocolitica*, all of which have a conserved lipase motif (GX₂SXG) [168]. Phospholipases fulfill, in general, three different functions: I) they can serve as digestive enzymes, II) they can remodel and regulate the membrane composition, and III) they can participate in relevant signaling cascades [168]. With respect to the latter, an interesting consequence of the activity of pathogenic phospholipases is that an important part of the damage to the host is independent of the direct damage caused on the host membranes, but rather as a consequence of second messengers that mediate inflammation responses, facilitating the host invasion [165].

A usually unexplored effect of phospholipases as virulence factors is their relevance in the adaptation of the pathogen to the host environment. Indeed, it has been described that *Pseudomonas aeruginosa* goes through important changes in the lipid composition of its plasma membrane compared to its planktonic state while forming biofilm colonies [170]. These changes could depend on remodeling the membrane by fine tuning the activity of lipases present in the bacteria. Additionally, recent publications have shown that, analogous to *Xanthomonas campestris*, fatty acids secreted by *Pseudomonas aeruginosa* (termed diffusible signal factors

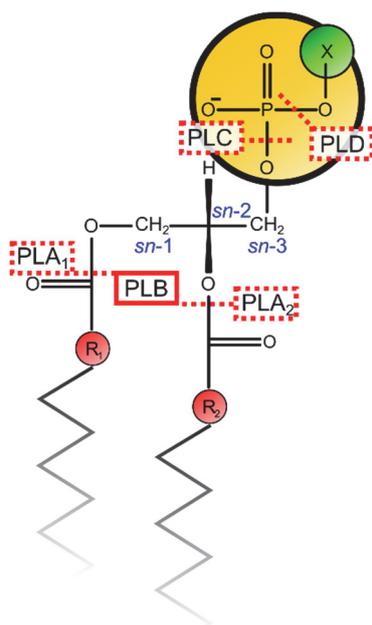


Figure 9. Phospholipases are classified according to the ester bond they are able to cleave. The color code corresponds to the one used in Figure 2. PLA₁ and PLA₂ are able to cleave ester bonds in positions *sn*-1 and *sn*-2, respectively. Enzymes that can cleave both bonds are classified as PLB. All of these enzymes are classified as acyl hydrolases. PLC and PLD on the other hand, cleave ester bonds from the phosphate group (phosphodiesterases), with the former resulting in a glycerol moiety without the phosphate group, while PLD cleaves the head group associated to the phosphate, resulting in a phosphatidic acid molecule.

or DSF), have been shown to regulate bacterial virulence and biofilm dispersion, essential for further infection spreading [171, 172]. Although little is known about the synthesis and sensing of these molecules, it seems reasonable to relate their activity to (phospho)lipase activities present in the pathogen.

Understanding the regulation of the pathogenic processes of the bacteria and its virulence factors, like the aforementioned phospholipases, has medical relevance. In 2010, Kovačić identified and purified a set of previously not described proteins with phospholipolytic activity [173], where TesA [174], PlpD [175] and PA2949 (PlaF) [176] have been further characterized. From those, PlaF has shown to be membrane bound, with features of a relevant virulence factor and was recently crystalized [177]. The exact physiological relevance of PlaF and its involvement in the virulence of the bacteria is not clear, however. In **section 4.3 (Publication III)**, the structure of the PLA₁ phospholipase PlaF from *Pseudomonas aeruginosa* will be described, with emphasis on the dimerization of the protein, a computational model of the energetics of the dimerization process itself, and its possible implications in regulating the activity of the enzyme.

2.6 Iron acquisition and CAR proteins in plants

Nutrient uptake by plant roots and the relevance of iron

Given the stationary nature of plants, nutrient acquisition from the surrounding environment and soil is essential to maintain their growth. Plants have to rely on extraction of the nutrients from the soil by their roots with the additional handicap that nutrients are heterogeneously distributed in space and time in their habitat [178]. In general, the most required nutrients obtained from the soil are the macronutrients nitrogen, phosphorous, potassium and magnesium, a lack thereof affects the ability to distribute carbon molecules and growth [179]. As a response, plants extend their root systems to patches of soil where the concentration of nutrients are higher [178, 180] and have a proper soil compaction and water supply [181]. One strategy used by plants to reach further soil space is to generate symbiosis with fungi to provide the proper nutrients. Up to 80% of plants, excluding the common plant model *Arabidopsis thaliana* [182], resort to arbuscular mycorrhizal symbiosis to obtain the right amounts of nutrients, in an interaction where fungi provide for nutrients like phosphorous and nitrogen in exchange for energy-rich molecules [183]. Taking into account the physicochemical descriptions of the transporters described until now, and of the periarbuscular space, we have recently described that the exchange between the plant and the fungus occurs as long as each part has enough nutrients to provide for the symbiosis partner, asymmetrically competing for the resources imported/exported to the periarbuscular space [184, 185].

In addition, plants require multiple micronutrients like Fe, Zn, Mn, Cu, Ni, B, Mo, and Cl. From those, iron is the micronutrient required in higher quantities, mainly due to its involvement in redox systems essential for photosynthesis and in enzymes required for nitrogen fixation [186]. Although iron is abundant in the earth crust, concentrations in the soil are close to 10^{-10} M, which is 10^2 - 10^5 fold lower than the optimal growth conditions for most plants [187]. Hence, iron tends to be a limiting growth factor, a lack of which causes decoloration or yellowing of the leaves, also known as iron chlorosis (see below). One factor that influences the low availability of iron in the soil is that under aerobic conditions, iron tends to get oxidized to Fe(III) and forms oxyhydrates, lowering its solubility at neutral and basic soil pH. On the other hand, anaerobic and acid conditions favor the accumulation of Fe(II), which can cause cellular damage due to the generation of reactive oxygen species [187, 188]. The essential nature of iron, its lower solubility and toxicity exemplify why having a proper homeostasis of iron is a crucial task for the root system of plants.

Even when reaching soil patches with proper nutrients, having the molecular mechanisms to fulfill the absorption is an essential part of the nutrient acquisition process. In general, plants have two different strategies for absorbing iron either as Fe(II) or Fe(III) (strategies I and II, respectively, see **Fig. 10A**) [192]. Though only strategy I is dependent on the IRON-REGULATED TRANSPORTER 1 (IRT1), many plants that use strategy II have an active IRT1 isoform, highlighting the broad relevance of the IRT1 transporter activity [192]. IRT1 was first described in 1996 by Eide *et al.* [196] as the major iron transporter under iron-deficient conditions, displaying high affinity for metal ions such as zinc, manganese, and cobalt. The knockout phenotype of the plant displays the previously described chlorotic characteristics

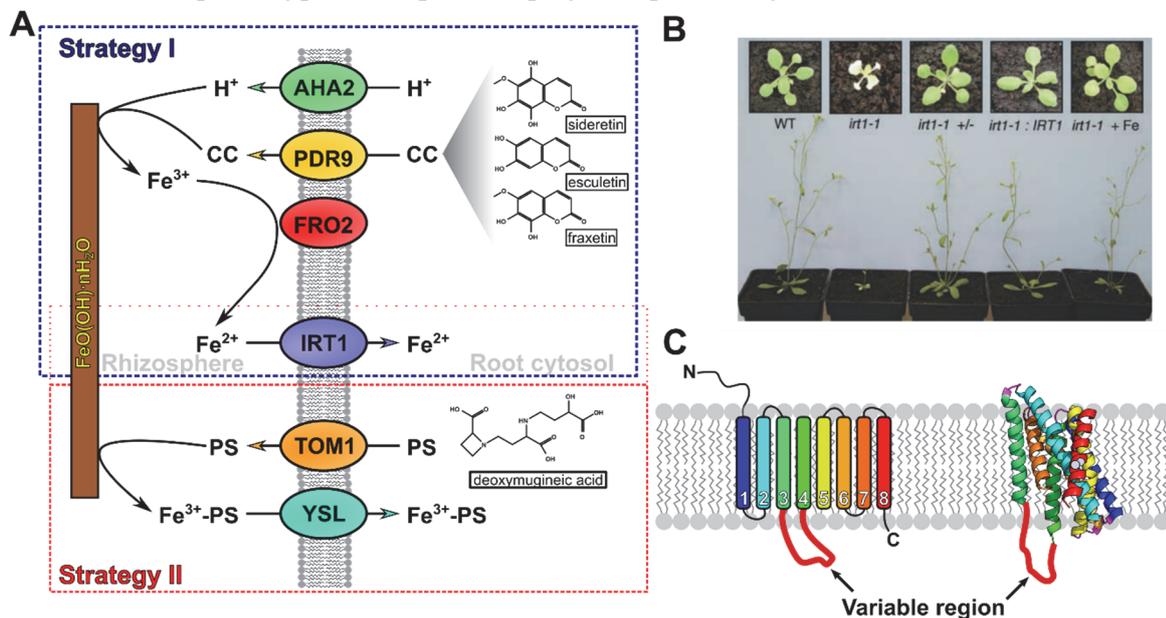


Figure 10. A) Root iron-uptake strategies in plants. Plants such as *Arabidopsis thaliana* and tomato follow strategy I, where Fe(III) is reduced to Fe(II) by FRO2 before transporting it with the IRT1 iron transporter. Fe(III) oxyhydrates are additionally solubilized by soil acidification by proton pumps (AHA2), and by catecholic coumarins (CC) such as sideretin, esculletin and fraxetin [189]. Grasses such as rice, maize and barley follow strategy II, and are able to uptake Fe(III) directly. For this, they secrete phytosiderophores (PS) such as deoxymugineic acid through TOM1, which after complexing Fe(III) are reabsorbed through the YSL/YS1 transporters [190, 191]. Interestingly though, many are able to transport Fe(II) through IRT1 transporters too. Figure adapted from Connorton *et al.* [192]. FRO2, FERRIC REDUCTION OXIDASE 2; IRT1, IRON-REGULATED TRANSPORTER 1; AHA2, ARABIDOPSIS PLASMA MEMBRANE H(+)-ATPASE ISOFORM 2; PDR9, PLEIOTROPIC DRUG RESISTANCE 9; TOM1, TRANSPORTER OF MUGINEIC ACID FAMILY PHYTOSIDEROPHORES 1; YSL, YELLOW STRIPE 1-LIKE. B) IRT1 is essential for iron uptake under low iron concentrations. Phenotype of *Arabidopsis thaliana* plants after 15 days (top) and 4 weeks (bottom) of growth. The lack of a functional IRT1 transporter causes chlorosis and lack of growth in the knockout plant (*irt1-1*). The wild type (WT) phenotype is maintained in heterozygous plants (*irt1-1 +/-*), and recovered when transformed with a plasmid containing the sequence of a functional IRT1 (*irt1-1:IRT1*). The underdeveloped phenotype is prevented when treated with water rich in iron (*irt1-1+Fe*). Figure licensed from Vert *et al.* [193]. C) The ZIP family of transporters consist of eight transmembrane helices and a variable region, rich in histidine residues, which is predicted to be unstructured and to be involved in binding metal ions ([194, 195], left). The only experimentally resolved structure for a ZIP transporter corresponds to an IRT1 homolog from *Bordetella bronchiseptica*, BbZIP (PDB: 5TSA [103], right). The structure confirms eight transmembrane helices in a 3+2+3 arrangement, where helices 1-3 are symmetrically related to helices 6-8, sandwiching helices 4 and 5. The unstructured variable region of BbZIP was not resolved in the crystal structure.

and growth defects while growing in soil [193] (**Fig. 10B**). Together with ZRT1 (ZINC-RELATED TRANSPORTER 1), IRT1 is the founding member of the ZIP (ZRT1/IRT1 like protein) family of metal transporters, which has representative homologs in all eukaryotic kingdoms [194]. ZIP proteins have been predicted to have eight membrane-spanning regions, with a histidine-rich cytosolic loop or “variable region” between transmembrane helices three and four (**Fig. 10C**). This variable region shows affinity for metal ions and has been proposed to be involved in regulating protein activity [197]. In IRT1, this loop has additionally been shown to be essential for posttranslational regulation of the activity of the transporter, mediating its ubiquitination [198] and internalization into the vacuole [199]. The internalization mechanism has in turn been shown to be regulated by the availability of other non-iron metal substrates of IRT1 [200]. Though an *ab initio* model of the human homolog hZIP4 was recently reported [102], only in 2017 the first structure of a transporter of the ZIP family, BbZIP from *Bordetella bronchiseptica*, was resolved [103] (**Fig. 10C**). The structure confirms the predicted eight transmembrane spanning helices, with a novel 3+2+3 TM arrangement, though the cytosolic loop remained unresolved. Though IRT1 and BbZIP belong to the same family, the sequence identity between them does not reach 20%, leaving room for important structural differences between the two.

Considering the relevance of IRT1 in iron uptake, having a better understanding of the regulation of this transporter is essential to comprehend plant responses to iron-deficient or iron-excess conditions. In the next subsection, the abscisic acid pathway will be described. Though IRT1 is not directly involved in this pathway, recent results exposed in this thesis suggest that a member of the C2-DOMAIN ABSCISIC ACID-RELATED (CAR) family of proteins is involved in regulating the transporter activity (**section 4.4, Publication IV**).

Abscisic acid hormonal response and the modulation by CAR proteins

Abscisic acid (ABA) is a hormone involved in responses to abiotic stress in plants. Responses related to this phytohormone physiologically prepare the plant and reduce growth to overcome situations with adverse temperatures, low water availability, high salinity and osmotic stress [201]. The mechanism used by plant cells to sense ABA remained obscure until recently. Under stress conditions, the biosynthesis of ABA is upregulated, leading to its cytoplasmic accumulation. Only in 2009 the receptor in charge of sensing the accumulated ABA was identified and described as the REGULATORY COMPONENT OF THE ABA RECEPTOR (RCAR) [202]. The receptor has a family composed of 14 proteins, all of which

show resistance to the synthetic ABA agonist pyrabactin in mutant strains. Because of that, RCAR proteins are also named PYRABACTIN RESISTANT 1 and PYRABACTIN RESISTANT LIKE proteins (PYR1/PYL) [203]. The complex of the receptor with the hormone then binds to type 2C protein phosphatases (PP2C) like ABA INSENSITIVE 1 (ABI1), inhibiting them (**Fig. 11A**) [202]. When inhibited, these phosphatases stop preventing the activation of proteins from the Sucrose non-fermenting 1 (SNF1)-related kinase 2 (SnRK2) family, enabling the responses commonly associated to ABA (**Fig. 11A**) [204].

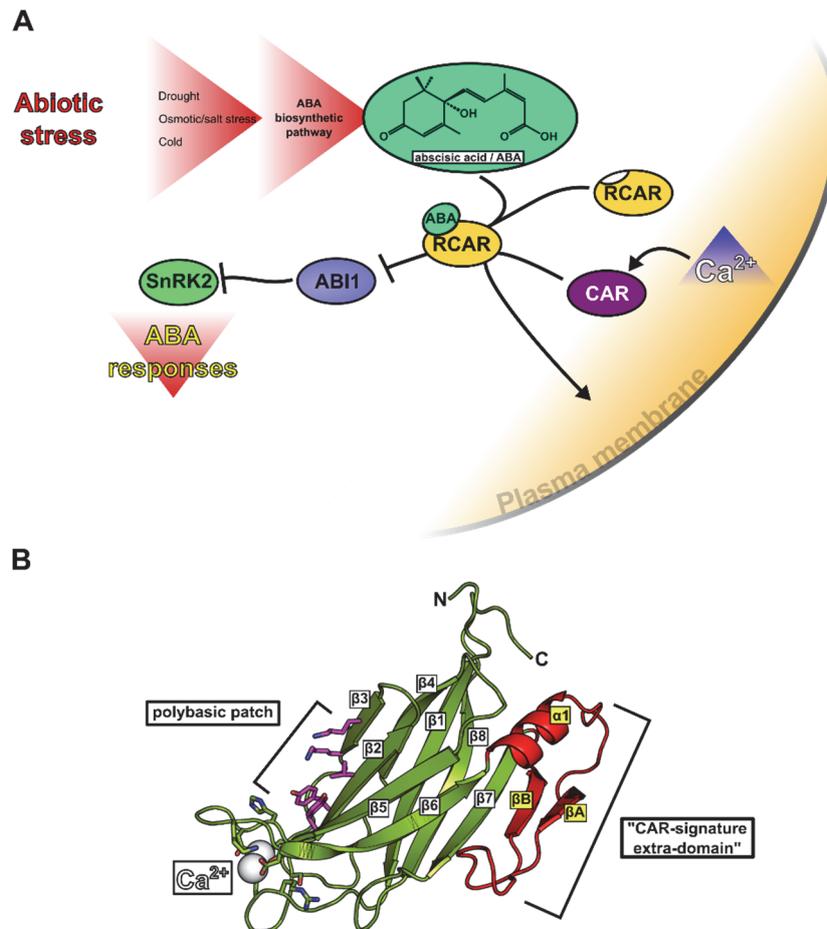


Figure 11. ABA signaling pathway and CAR proteins. **A)** ABA is an essential phytohormone that regulates the responses of plants to abiotic stress. Under these conditions, the *de novo* biosynthetic pathway of ABA is upregulated. In the cytoplasm, ABA is sensed by binding to its receptors of the RCAR (also known as PYR/PYL) family. Upon binding, the RCAR is able to bind and inhibit PP2C phosphatases like ABI1. Without the phosphatase activity, downstream kinases from the SnRK2 family remain phosphorylated, and trigger the ABA responses by phosphorylating other downstream effectors. Recently it has been shown that the action of the RCAR/ABA complex can be regulated by complementary calcium signals and C2-DOMAIN ABA-RELATED (CAR) proteins, which target the complex to the plasma membrane, where it can regulate the activity of additional effectors. **B)** The crystal structure of CAR4 shows the expected C2-domain structure, with eight β -strands forming a β -sandwich, with an occupied calcium-binding site (white spheres) and a polybasic patch (in magenta). The protein has an additional $\alpha 1\beta A\beta B$ “CAR-signature extra-domain”, present in all the members of the CAR family (highlighted in red).

Overall, though the activation of the RCAR receptor by ABA has been described, the cellular mechanisms used to target the complex to specific proteins and cell regions are unclear. Interestingly, and maybe to be expected, the responses to ABA and the calcium signaling pathways are convergent, with PP2Cs being essential in the signaling cross-talk [204]. An additional connecting point between both signaling pathways are the recently described CAR proteins (**Fig. 11A and B**). In 2014, Rodriguez *et al.* [205] identified the family of 10 CAR proteins, showing that multiple of its members are able to coimmunoprecipitate with different RCARs. PYL4 and CAR1 colocalize and bind to the plasma membrane in a calcium-dependent manner, showing the targeting of the receptor to potential protein interactors. Moreover, triple mutants for different CAR proteins showed reduced sensitivity to ABA, highlighting the involvement of the CAR family in the signaling pathway.

The structure of CAR4 was resolved, evidencing a “CAR-signature extra-domain” (or CAR-signature domain for short) formed by an additional $\alpha 1\beta A\beta B$ motif that distinguishes all members of the CAR family from a canonical C2-domain (**Fig. 11B**). C2-domains have for long shown to interact with membranes, both through their calcium-binding site [206] and with their polybasic patch, taking advantage of negatively charged lipids present in the membrane. The latter interaction has been shown to be specific for PtdIns(4,5)P₂ in canonical PKC- α C2-domains [207, 208]. Both interactions make proteins containing C2-domains interesting targets as membrane interacting hubs [209], where homologs of the canonical PKC domain have a wider variety of specificities towards PtdIns or affinities for calcium [210, 211].

From the members of the CAR family, CAR6 was already identified in 2011 by Knauer *et al.* [212]. The protein was initially named ENHANCED BENDING 1 (EHB1), and negatively regulates hypocotyl bending by gravitropism and under blue light conditions by forming a complex with phototropins [213]. However, not more is known on how this protein is connected to the ABA signaling pathway, or what other targets it could be regulating. In **section 4.4 (Publication IV)**, the experimental interaction of EHB1 with the membrane, as well as the binding and inhibition of IRT1 by EHB1 will be shown. This is a particularly interesting result, as it links the ABA pathway with the regulation of the uptake of iron and other metals, protecting the cells from potential metal toxicity and oxidative damage. With homology models of EHB1 and IRT1, we illustrate how the CAR protein could bind to the membrane in response to calcium, respond to changes in the PtdIns present in the membrane, and bind to IRT1.

3 Scope of the Thesis

Throughout the Introduction and Background chapters, the importance of cellular membranes and related systems of high-interest associated to membranes were presented: **I**) the ethylene-signaling pathway in plants (**section 2.4**), **II**) phospholipolytic enzymes associated with virulence in *Pseudomonas aeruginosa* (**section 2.5**), and **III**) the iron uptake from soil and the ABA/CAR regulation in the plant root (**section 2.6**). Though all of these systems are quite different from one another, being in different organisms and cellular environments, they all have in common using membranes as the central hubs to organize their corresponding pathways, relying on membrane, membrane/protein and/or protein/protein molecular interactions.

In general, biological systems associated with membranes are more difficult to approach, which might explain why these topics have only until recently been studied more thoroughly. The last two decades of investigation and development in membrane systems and techniques associated with their study have been accompanied by a constant increase in computational power and development of techniques related to the study of membrane environments. MD engines and their corresponding force fields are a good example for such a development (**section 2.3**). Related to this, the amount of time associated with preparing simulations of membrane bilayers with additional proteins or molecules of interest has increased dramatically, to the point where for particularly complex systems the time dedicated to setting up and preparing them can overcome the time required to reach the desired simulation time. To tackle this in this work, PACKMOL-Memgen, a software to assemble and help in the parametrization of common membrane bilayer systems, was developed (**section 4.1, Publication I**). We show that the generated systems are well-suited for performing MD simulations and that the software is generalizable to more complex systems. The program is included in the distribution of AmberTools version 18 and above.

To have a structural insight into the mechanism that triggers the ethylene-signaling pathway in *Arabidopsis thaliana* (**section 2.4**), the first model of the transmembrane domain of an ethylene receptor was obtained in this thesis. Using predicted contacts obtained from coevolutionary signals and *ab initio* modeling (**section 2.3**), a dimer structure of the transmembrane domain of ETR1 was obtained (**section 4.2, Publication II**). The structure was validated by previously reported single point mutants and by the effects of tryptophan mutants on the secondary structure of the receptor determined by circular dichroism[†]. Additionally, the copper/receptor stoichiometry was determined, resolving previously contradictory information

available in literature[†]. Molecular simulations of the modeled transmembrane domain in a fully explicit membrane environment and with high concentrations of ethylene showed binding of the ligand to the sensor domain. The binding occurs in proximity to the added copper ions and in accordance to the aforementioned reported mutations, illustrating the first steps of the ethylene-signaling pathway.

With respect to the discovered phospholipases of *Pseudomonas aeruginosa* (**section 2.5**), PlaF showed to be an inner membrane anchored enzyme with its catalytic activity exposed to the periplasmic space. The obtained crystalized structure of PlaF showed a dimer arrangement, while experimental measurements evidenced activity as a monomer and suggest a decrease in its activity as the protein dimerizes *in vitro* (**Publication III**)[‡]. In this work, and to rationalize the importance of the dimer for the protein activity, the dimerization free energy of the protein in a membrane bilayer context was calculated by umbrella sampling simulations. The calculations showed that the protein has a favored conformation as a dimer (**section 4.3**). Interestingly though, we showed that the monomer is able to interact with the membrane surface with the dimerization interface by performing a tilting motion, which also showed to be the favored monomeric conformation in free energy calculations. Overall, the dimerization and tilting free energies suggest that the protein activity might be controlled by tight regulations of the availability of the enzyme in the membrane surface, where an increase in the concentration would cause the dimer conformation to be favored and the activity of the enzyme to be reduced. As the activity of PlaF showed to be directly correlated with the virulence of the bacterial strain, this opens a new target for dealing with *Pseudomonas aeruginosa* and potential homolog infections.

Finally, EHB1, a recently identified member of the C2-DOMAIN ABA- RELATED (CAR) protein family of *Arabidopsis thaliana*, was identified as an interactor of the variable region of IRT1 by a yeast two-hybrid approach (**section 2.6**). This interaction with IRT1 regulates the activity of the iron transporter (**Publication IV**)[¶]. In this work, MD simulations were used to evaluate the binding of EHB1 to the membrane, evidencing that it is able to orient

[†]The experimental work was performed by Lena Müller and Elissa Classen in the Laboratory and under the guidance of Prof. Dr. Georg Groth, in the Institute for Biochemical Plant Physiology, at the Heinrich Heine University, Düsseldorf, Germany.

[‡] The experimental work was performed by Florian Bleffert, Joachim Granzin and Muttalip Caliskan under the supervision of Renu Batra-Safferling, Karl-Erich Jaeger and Filip Kovačić in the Institute for Molecular Enzyme Technology and the Institute of Complex Systems, Forschungszentrum Jülich, Jülich, Germany.

[¶] The experimental work was performed by Imran Kahn, Regina Gratz, Polina Denezhkin, Kalina Angrand, Lara Genders, Rubek Merina Basgaran, Claudia Fink-Straube, Tzvetina Brumbarova and Rumen Ivanov, under the supervision of Petra Bauer and Rumen Ivanov in the Institute of Botany, at the Heinrich Heine University, Düsseldorf, and at the Department of Biosciences-Plant Biology, Saarland University, Saarbrücken, Germany.

itself transiently parallel to the membrane plane, potentially enabling the interaction with the variable region of IRT1 (**section 4.2**). Interestingly, it was observed that EHB1 does not interact with PtdIns(4,5)P₂ as the canonical C2-domain from PKC, but rather with PtdIns and PtdIns(4)P. This could be explained in terms of changes in the polybasic patch, and suggested that this interaction could be facilitated by the transient tilting of the protein on the membrane surface.

The previously introduced works exemplify different levels of available experimental information: starting from coevolutionary information, where no structure was available (ETR1), with an X-ray determined set of atomic coordinates (PlaF), or with homolog structures with adequate sequence identity (EHB1). All of the aforementioned membrane systems have in common their dependence on membrane structures, and together with the computational approaches used to study them, form the main scope and framework on which this thesis work is funded. In the following Chapter, the computational work realized in this thesis, and how it helps to understand the behavior of the studied proteins, will be presented.

4 Publications

Reprint permissions

Publication I:

“PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein–Lipid-Bilayer System Building”

Stephan Schott-Verdugo and Holger Gohlke

Journal of Chemical Information and Modeling, 2019, 59 (6), 2522-2528

DOI: 10.1021/acs.jcim.9b00269

Reprinted with permission from the Journal of Chemical Information and Modeling. Copyright © 2019, American Chemical Society.

Publication II:

“Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain”

Stephan Schott-Verdugo, Lena Müller, Elisa Classen, Holger Gohlke & Georg Groth

Scientific Reports, 2019, 9(1), 1-14

DOI: 10.1038/s41598-019-45189-w

This is an open access article distributed under the terms of the Creative Commons CC BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Publication IV:

“Calcium-Promoted Interaction between the C2-Domain Protein EHB1 and Metal Transporter IRT1 Inhibits Arabidopsis Iron Acquisition”

Imran Khan, Regina Gratz, Polina Denezhkin, Stephan N. Schott-Verdugo, Kalina Angrand, Lara Genders, Rubek Merina Basgaran, Claudia Fink-Straube, Tzvetina Brumbarova, Holger Gohlke, Petra Bauer, Rumen Ivanov

Plant Physiology, 2019, 180 (3), 1564-1581

DOI: 10.1104/pp.19.00163

Reprinted with permission from Plant Physiology (www.plantphysiol.org). Copyright © American Society of Plant Biologists.

4.1 PACKMOL-Memgen: A simple-to-use generalized workflow for membrane-protein/lipid-bilayer system building

Schott-Verdugo, S., & Gohlke, H.

Journal of Chemical Information and Modeling, 2019, 59, 2522–2528.

For the original publication, see page 60.

Background

As mentioned in **section 2.3**, one of the challenges of studying membrane systems or membrane proteins by computational techniques is building a system that is representative of the *in vivo* conditions. The way lipids assemble to form a membrane and the placement of a protein within or at the membrane are not trivial due to, in part, the anisotropic nature of the lipid bilayer. To tackle the so-called “membrane-protein packing problem”, multiple tools have been developed to help in the process of generating such systems [214, 215], such as CHARMM-GUI[44], VMD (through the Membrane plugin) [216], Maestro [217], and Packmol [218]. CHARMM-GUI stands out for having multiple lipids available and a user-friendly web interface, making the packing process easier for newcomers but at the same time dependent on a web interface. This dependency turns into a challenge when a user needs to generate multiple systems with different proteins or different membrane compositions, or requires different starting configurations, as the process becomes time-consuming. Compared to CHARMM-GUI, Packmol stands out for having functionalities that allow building complex and intricate simulation systems directly in the command line [218]. These functionalities build upon solving the packing problem with the GENCAN optimizer, which minimizes an objective function that describes the molecular overlap [219]. To do so, appropriate PDB files for the individual system components and system-specific geometric constraints to restrict the packing are required as input. Together with the requirement of command line-usage, this makes applying Packmol difficult for non-expert users.

To conciliate the benefits from Packmol and user-friendliness, but still make a program that runs locally in the command line, we developed PACKMOL-Memgen. It is a Python-based program that uses Packmol as the packing engine but wraps the main procedures required to build complex membrane systems, such that only single-line commands of the user are required.

PACKMOL-Memgen Workflow

The general workflow of PACKMOL-Memgen follows similar principles as the CHARMM-GUI membrane builder [44], and is depicted in **Figure 12**. The user can decide to either pack a membrane-only system, or embed a protein structure into the membrane. For the latter, a PDB file of the protein structure following the conventional formatting is used as input (step 1). The next steps can be summarized in determining a default protonation for the input PDB with **reduce** [220] (step2), predicting the orientation of the protein in a membrane environment with **memembed** [221] (step3), to then determine the dimensions of a default packed box, given the size of the provided protein structure. To set the number of lipids that will be used, the area per lipid (APL) values reported in literature and the surface area of the membrane patch to be packed are considered. If lipid mixtures are used, the APL is computed as a weighted average according to the composition ratio. The volume occupied by the protein in each water slab, as well as in the membrane leaflets, is calculated by a grid approach derived from **pdbremix** [222] (step 4.2). This is later used to reduce the previously calculated number of molecules in the respective system sections. For this, the lipid molecular volume is obtained from literature or estimated based on the length of the acyl chains as described previously [223].

The user can select multiple lipids and define complex lipid mixtures at desired ratios, including different compositions per leaflet. In addition, multiple bilayers can be generated by calling the corresponding flags once per bilayer. This is useful if, for example, an electrochemical gradient is required to study a system as in the case of “computational electrophysiology”[224].

To initiate the molecule packing (step 5), the calculated number of molecules of each type and the geometrical constraints associated with each section of the system are written to a Packmol input file, and Packmol is started. If a protein is included, its position is considered fixed during the packing, and all other molecules will be packed such that they accommodate to the protein. The Packmol PDB file output is, by default, transformed into an AMBER-compatible file.

To show the validity of the generated systems, a membrane-only system with an inner-membrane-like composition of a Gram-negative bacteria of DOPE:DOPG 3:1 was simulated for 500 ns (**Fig. 13**, top). Parameters including the APL, the S_{CD} order parameter, membrane thickness, electron density, and the average lipid mean square displacement corresponded with values reported previously. Additionally, a membrane patch with the KcsA potassium channel (PDB 1BL8) was simulated (**Fig. 13**, bottom), showing well-behaved systems and replicating conformations previously reported in literature.

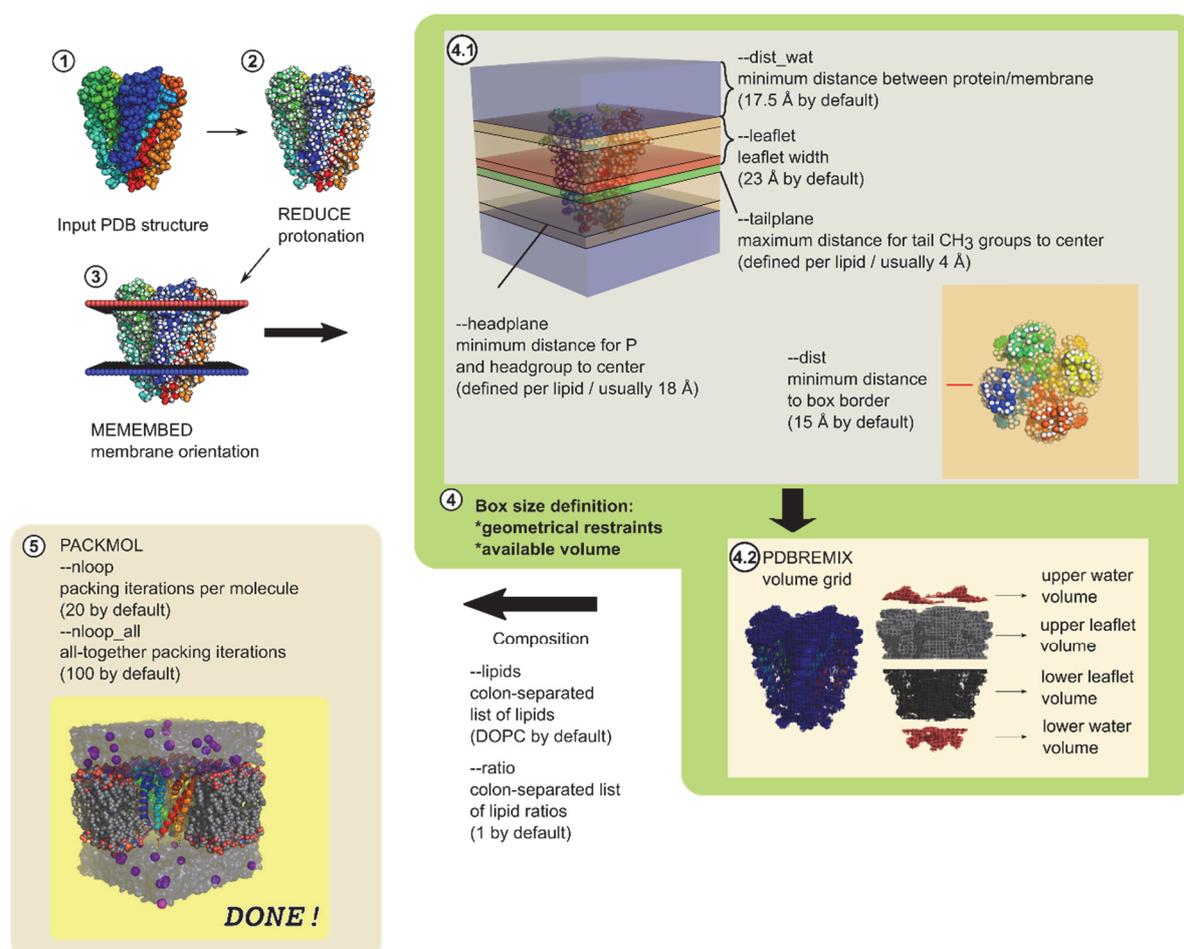


Figure 12. PACKMOL-Memgen workflow applied to the protein structure PDB ID 1BL8. The process comprises multiple, often optional, steps that are controlled by flags available in the software. The most important flags are mentioned in the Figure. Figure taken from Publication I, page 62.

Conclusions and significance

In this work, PACKMOL-Memgen was developed and evaluated. The program is a simple-to-use generalized workflow for automated building of membrane-protein/lipid-bilayer systems that relies on open-source tools only, including Packmol, memembed, pdbremix, and AmberTools. As evaluated in **Publication I**, the built systems are appropriate as starting point for MD simulations under periodic boundary conditions. There are more than one way of building a membrane system for performing computational studies, and the choices may depend on the molecular simulation package and the force field used, or the user preference. The development of PACKMOL-Memgen enabled users of the AMBER community to make a local packing of the system to be simulated. They were primarily dependent on webserver such as CHARMM-GUI to build membrane-protein/lipid-bilayer systems, which while user-friendly, becomes inconvenient if multiple systems need to be built. Interestingly, despite its recent release, it has already been used in scientific research from members of the AMBER community to study the folding of peptaiboil in a membrane context with accelerated MD [225] and a desensitization mechanism of the ATP-gated P2X2 receptor [226].

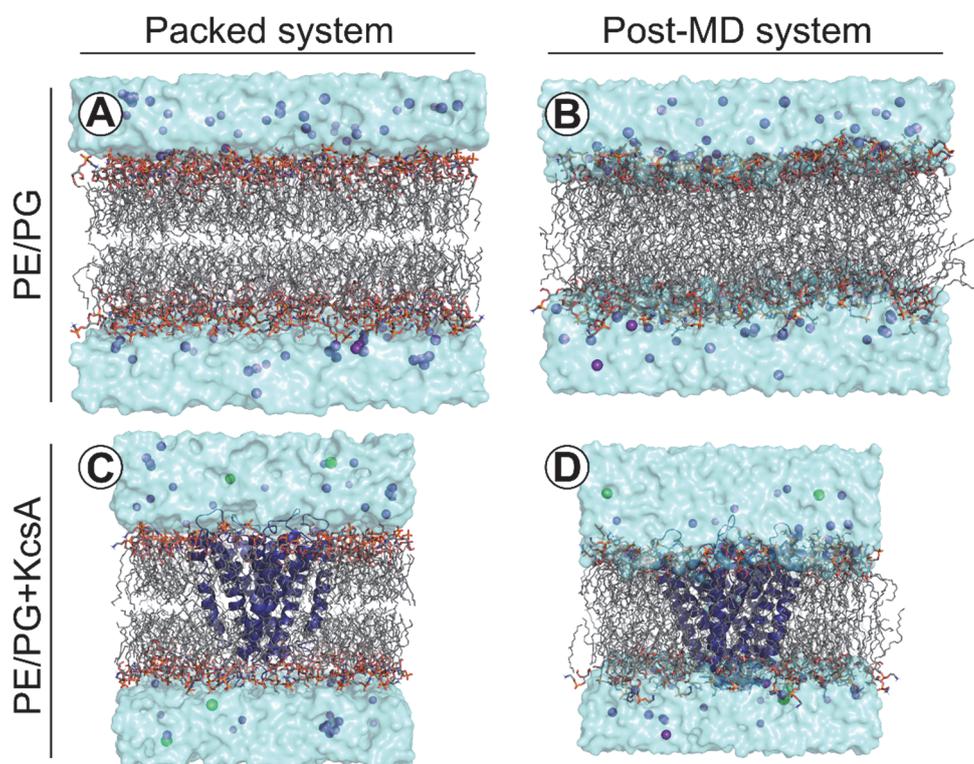


Figure 13. Representative structures for a membrane-only DOPE:DOPG 3:1 system (A, B) and the KcsA channel (PDB 1BL8) embedded in a membrane (C, D) generated as described in Publication I, immediately after packing with PACKMOL-Memgen (A and C) and after 500 ns of MD simulations (B and D). Figure taken from Publication I, page 64.

PACKMOL-Memgen was an important development for the workflow with computational membrane systems, and was essential in studying free diffusion dynamics of ethylene with the model of the transmembrane domain of ETR1 (**Publication II, section 4.2**), in measuring the dimerization free energy profile of the *Pseudomonas aeruginosa* phospholipase PlaF (**Publication III, section 4.3**), and in studying the interactions of the CAR protein EHB1 and model its mechanism of inhibition of the plant iron transporter IRT1 (**Publication IV, section 4.4**). On top of that, the development on the program has been continued, with recent incorporation of parameters for lysophospholipids, PtdIns, and cardiolipin, and incorporating functions to pack systems with curved membranes.

4.2 Structural model of the ETR1 ethylene receptor transmembrane sensor domain

Schott-Verdugo, S.[§], Müller, L.[§], Classen, E., Gohlke, H.[‡], & Groth, G.[‡]

Scientific Reports, 2019, 9(1), 1–14.

^{§,‡} Both authors contributed equally to this work.

For the original publication, see page 82.

Background

As presented in **section 2.4**, the ethylene pathway is essential in fruit ripening, senescence, and decay, with important economic implications. The ethylene receptors in *Arabidopsis thaliana* are a family of five homologs that resemble two-component systems from bacteria, form higher oligomeric assemblies at the ER membrane [227] and require copper ions in oxidation state +1 to exert their activity [151, 228, 229]. As mentioned above, the copper stoichiometry in the functional dimer is still a matter of discussion, with one Cu⁺ / monomer [155] and one Cu⁺ / dimer [151, 156] as the possible stoichiometries. On top of that, the lack of a protein structure for the whole receptor has hindered the understanding of the intra- and intermolecular mechanism used to transmit the ethylene signal in the ethylene receptors and to further downstream elements. A model of the whole cytosolic domain (residues 158 to 738) was published by Mayerhofer *et al.* [149], with the GAF, dimerization, catalytic, and receiver domains, but lacking the transmembrane domain and the connection to it. Most structural information of the transmembrane region comes from the study of Wang *et al.* [137], where three main classes of *loss-of-function* mutants were identified, showing different levels of ethylene binding, signal transmission, or intrinsic activity. Despite this, no structure of the transmembrane domain and the ethylene-binding region has been reported, which is aggravated by the fact that no homolog with an identity higher than 15% has been obtained until now.

The current development of *ab initio* protein folding tools allows one to predict the structure of such a domain (**section 2.3**). Using Rosetta to generate *ab initio* decoys, we generated the first monomeric structural models of the transmembrane sensor domain of the ethylene receptor ETR1. The search was refined by integrating coevolutionary information and in this way include all experimental information available. The selected representative model was used to generate an ETR1 dimeric structure of the transmembrane region in complex with copper (I) (ETR1_TMD/Cu) according to experimentally determined stoichiometries, and the

models were validated by tryptophan scanning mutagenesis. The obtained dimeric model was used to obtain free diffusion MD trajectories of ethylene into the receptor, exemplifying how ethylene can reach to the putative copper binding sites and potentially trigger the signaling cascade.

Modeling of the ETR1 transmembrane domain

To obtain a structure for the first 117 residues of ETR1, which includes the whole transmembrane domain, the *membrane_abinitio2* Rosetta protocol was used, including the secondary structure and transmembrane topology predictions from PSIPRED [230] and CCTOP [231], respectively. The 100,000 generated decoys were clustered in three groups, revealing two different handednesses for the obtained models (**Fig. 15**, top row). To filter the results and discriminate between the potential solutions, coevolutionary residue-residue contact predictions from MetaPSICOV [95] were used to rescore the models by a so-called Contact score. The Contact score was able to discriminate between models with similar Rosetta scores (**Fig. 14**, left panel), showing that the third largest cluster fulfills the predicted contacts best, as seen in the overlap of the predictions to the contact map (**Fig. 15**). Similar to the hZIP4 case, only models with a favorable z-score below -2 were kept for further analysis, which removed all right-handed helical models from the pool of 100,000 models and left 5,217 structures. Reclustering the remaining models shows conformations that only differ in the relative orientation of the helices and slightly in the positioning of the third helix with respect to the second (**Fig. 15**, bottom row conformations and contact map). The centroid of the most populated cluster (representing 40% of the filtered structures) was selected for further analysis. Interestingly, the remaining filtered decoys have a template modeling score (TM-score) to the selected model of 0.5 or higher, which has been shown to be indicative of having the same fold [232] (**Fig. 14**, right panel). The selected model was further refined with the Rosetta protocol *relax* to incorporate side chains [233], serving as the first available monomeric model of the transmembrane region of ETR1.

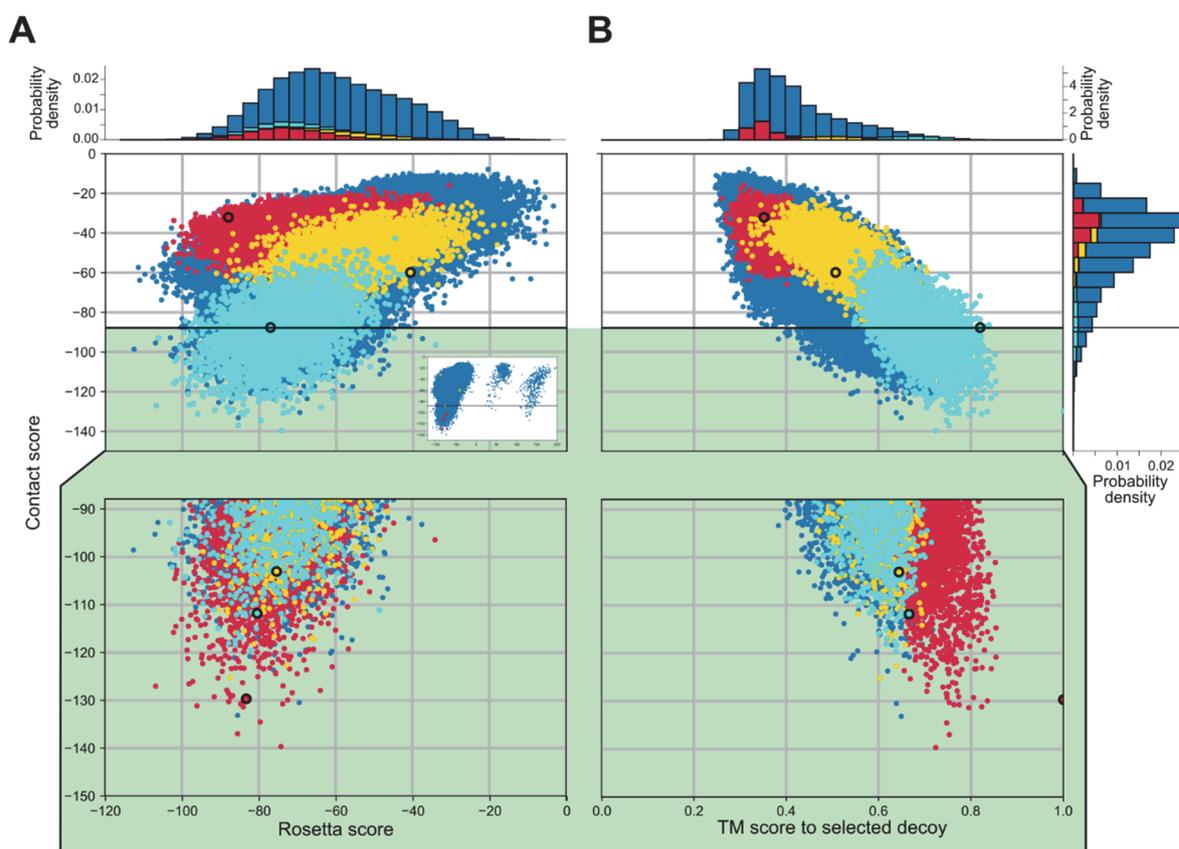


Figure 14. Distributions of Rosetta scores (membrane centroid score, **A**) and TM-scores with respect to the selected decoy (**B**) versus the calculated Contact score of the generated models. There is a positive correlation between the Rosetta score and the Contact score ($R^2 = 0.25$, $p < 0.001$) for models with a negative Rosetta score ($R^2 = 0.01$ for the complete distribution). The horizontal line demarks the -2 z-score threshold used with respect to the contact score to filter out the worst scoring models; selected structures are shown on green background as a zoom. The models were structurally clustered in three groups pre- and post-score filtering, with the structures corresponding to the first, second, and third cluster shown in red, yellow, and cyan, respectively. The centroid models of the identified clusters (**Fig. 15**) are highlighted with a thick outline. The inset in panel A shows the complete distribution, including outliers. Figure taken from Publication II, see page 84.

As the active form of the receptor is a dimer with copper ion cofactors, such a model had to be built from the monomeric structure. First, the copper stoichiometry was determined by loading the purified protein to saturation with copper (I) in reducing conditions, to later measure the amount of copper loaded with a bicinchoninic acid (BCA) assay. The ratio of loaded protein and BCA showed that one copper ion binds per protein monomer, solving the aforementioned unclarity regarding the requirement of copper. In addition, by titrating a solution of BCA-Cu(I) with the purified protein, an affinity constant of 1.3×10^{-15} M was obtained. Using this copper stoichiometry, the notion that the metal binding site should be shielded from the solvent [234], the monomeric model, the coevolutionary signals and information of low-lipophilicity regions, the interface formed by the first two helices was used to form a dimeric model by docking with HADDOCK [235] and refining with implicit solvent replica exchange MD simulations [236]. The representative dimer model (**Fig. 16**) consists of

an almost symmetrical arrangement of the previously modeled left-handed monomers, where the N-terminus includes the disulfide bonds in between chains formed by residues 4 and 6. The putative copper binding sites, composed of residues C65 and H69, are buried in the dimerization interface, as previously mentioned and suggested [137]. From the stoichiometry measurement and the residues involved according to Rodriguez *et al.* [151], two copper ions were added in proximity to S_γ of C65 and N_δ of H69. With the C-termini of the dimer at opposite sides of the dimeric configuration, there is enough room for residues connecting the transmembrane domain and the cytoplasmic GAF domain.

Interestingly, the model shows that the previously determined loss-of-function mutants for ethylene binding, signaling, and function form a “layered” spatial arrangement of residues, starting close to the putative copper binding site and with residues relevant to signal

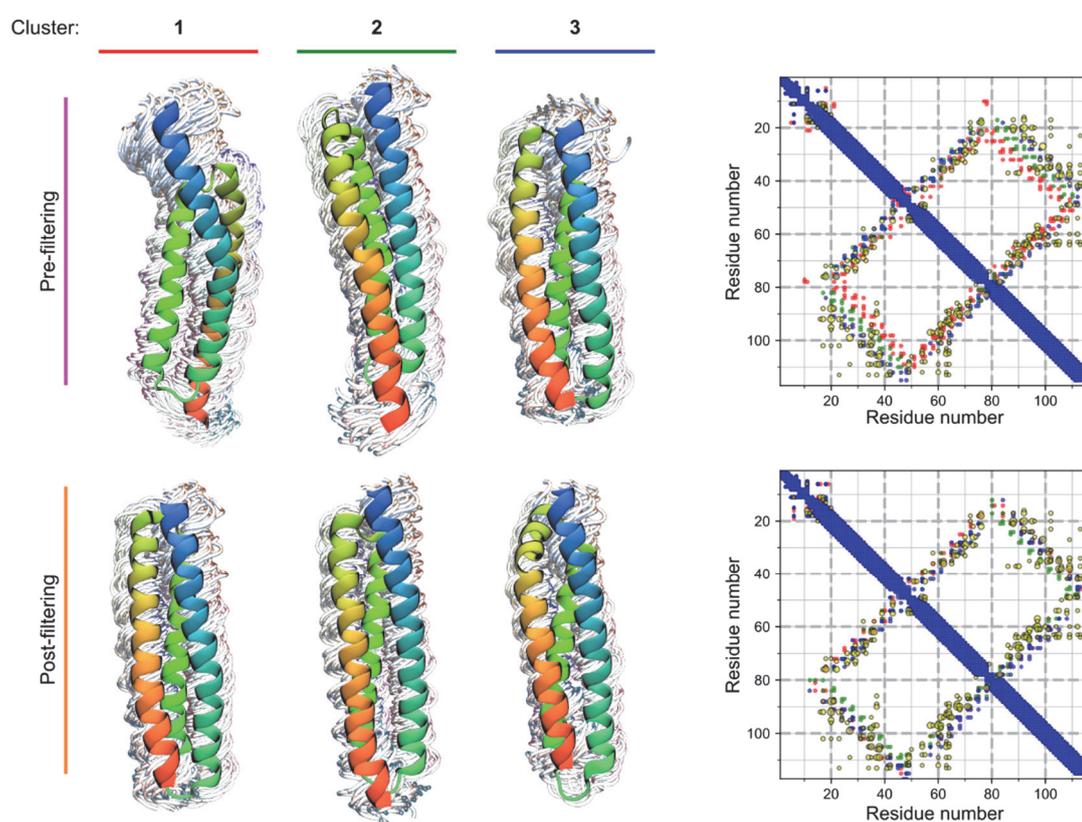


Figure 15. Clustered conformations and average contact maps for the generated decoys pre- (top) and post- (bottom) filtering according to the Contact Score. The centroid of each cluster is shown in a cartoon representation, overlaid over every tenth other structure in the cluster, shown as wires. The structures are colored blue to red, starting from the N-terminal portion. The contact map for each set of clustered conformations are overlaid on the right, following the color scheme shown on top of every cluster to the left. Additionally, the MetaPSICOV contact predictions are shown as yellow dots. The size of the dots reflects the confidence assigned by the method. Before filtering, the clusters show different orientations with respect to helix three, as visible in the different conformations and on the contact map (residues 20-40 contacting 80-100). Contacts of the right-handed configuration of cluster 1 deviate the most from the MetaPSICOV contact predictions, and such configurations are removed by the filtering. Figure taken from Publication II, see page 87.

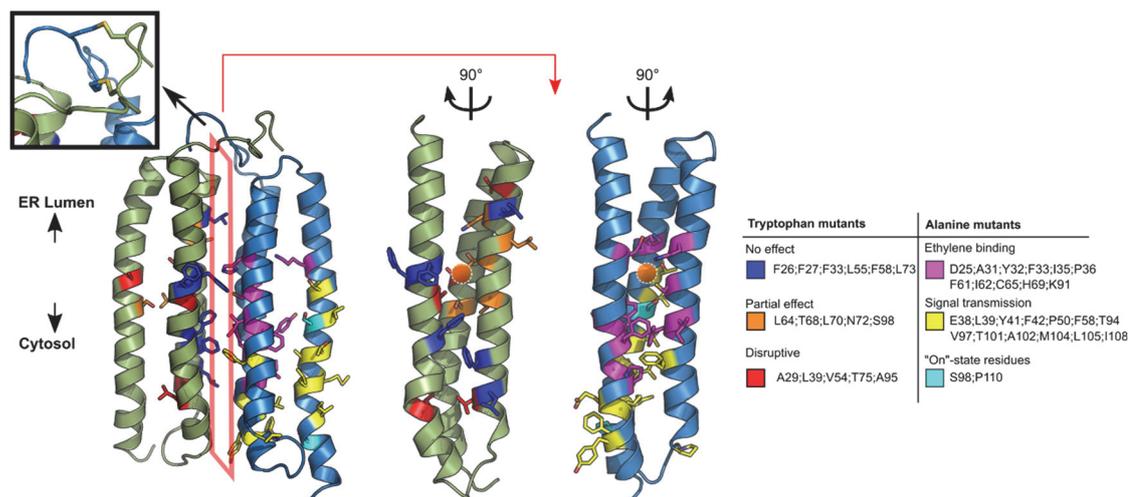


Figure 16. Dimer model of the ETR1 transmembrane domain. On the structure, the positions where tryptophan mutants were generated (left monomer) and where previous loss-of-function mutations were found ([137], right monomer) are denoted by the color code on the right, with a list of the residues found in each class. The upper left inset shows the disulfide bridges included during the refinement of the protein. The orange spheres show the putative copper binding site in proximity to residues C65 and H69. The orange polygon indicates the interface. The two structures on the right represent the interface in an “open book” representation. Tryptophan mutants of residues that are pointing towards the monomer bundle core are disruptive (red) or are partially disruptive (orange), while mutations of residues pointing towards the dimer interface of the model (blue) showed no effect on the alpha-helical content. Residues that are closer to the protein center and in proximity to the putative copper binding site are essential for ethylene binding (magenta), while residues farther away and closer to the cytosolic portion are responsible for the signal transmission (yellow). Residues shown to be relevant to maintain the protein in an active state are displayed in cyan. Figure taken from Publication II, see page 87.

transmission close to the third helix/cytoplasmic region. Additionally, S98 and P110 [137] which are essential for signal transmission, are pointing towards the copper binding site and generating a kink in the C-terminal portion of helix 3, respectively. From the dimeric model, and based on solvent accessibility measures, a set of residue positions were chosen to perform tryptophan mutants and secondary structure determination by circular dichroism. From this analysis, only residues that are not exposed and pointing towards the core of each monomer showed a significant reduction in the calculated alpha-helical content, while residues that point towards the dimer interface showed no significant change (Fig. 16).

MD simulations of free diffusing ethylene show binding in proximity of the putative copper binding site

To evaluate ethylene binding to the ETR1 transmembrane domain/copper dimer model, MD simulations of 1 μ s length with an explicit membrane environment were prepared with PACKMOL-Memgen (Publication I, section 4.1). Ten independent replicas were simulated, either containing 0.1 M ethylene in the water phase at the beginning of the simulation or not. The modelled structure showed only a moderate C α atom RMSD drift with respect to the

average structure, with all of the simulations reaching an apparent plateau after 500 ns. The copper ions remained bound to the sulfur atom of cysteine C65 of each chain, interacted with the non-protonated N δ of H69, and showed infrequent interactions with D25.

The positioning of ethylene in the simulation was measured as a 3D histogram, showing that ethylene has a higher propensity to bind within either monomer of the transmembrane domain than to freely diffuse in the membrane (**Fig. 17A**), further supported by a clustering analysis (**Fig. 17B**). Of particular interest is the binding of ethylene in proximity to the included copper ions, with three of the identified ethylene-binding sites at a distance < 5 Å. From those, two are located in the center of each monomer bundle (**Fig. 17**). A set of ten residues was identified in these binding sites, seven of which have been shown to be essential for ethylene binding to the protein (red labels in **Fig. 17**) [137]. Only three residues have been previously described and were not identified, but are in the immediate proximity to the centroids of the described clusters (I35, P36, and I62). These results are very promising, as no Dewar-Chatt-Duncanson type of interaction was included in the simulations, relying solely on the binding

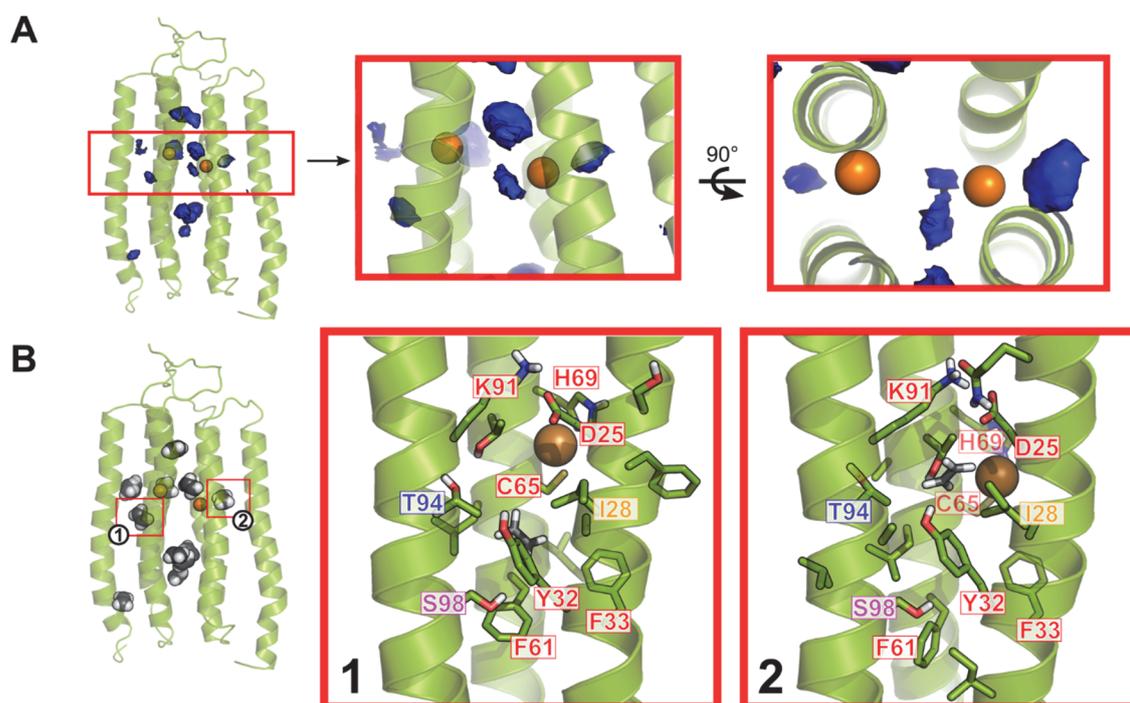


Figure 17. Putative ethylene-binding sites identified during MD simulations. The average structure of the ETR1_TMD dimer over all replicate simulations is shown. **A)** 3D propensity representation of the most probable locations of ethylene binding along all replicate simulations. The boxes highlight the portion close to the included copper ions, with a zoom on the blobs at a distance < 5 Å. **B)** Representative configurations of identified ethylene clusters. The clusters enclosed in the red boxes are at a distance < 5 Å to the copper ions, and are shown in detail on the right. The residues labeled in red have been shown to be essential for ethylene binding, S98 (magenta) has a moderate effect, while T94 (blue) has no effect on binding, but is relevant for signal transmission. I28 has not been tested. Mutation information from Wang *et al.* [137]. Figure taken from Publication II, see page 89.

site affinity and electrostatic interactions between the parametrized ethylene and copper ions. Inclusion of a more thorough model for this quantum effect would increase the affinity of ethylene for the copper ion and reduce their interaction distance, while also increasing the affinity for the ethylene receptor. This could give more insights into the effects of ethylene binding into the receptor structure and its signal transduction.

Conclusions and significance

In this study, the lack of a structural representation of the dimeric transmembrane domain of the ethylene receptor ETR1 of *Arabidopsis thaliana* was tackled. In order to model the protein, we used state-of-the-art computational methods such as *ab initio* modelling with coevolutionary predicted contacts and molecular dynamics simulations of explicit membranes, together with experimental determinations of the copper stoichiometry and affinity for ETR1.

The main results of this study are:

- The stoichiometry of copper (I) binding to the ethylene receptor ETR1 is one ion per monomer. The ions bind with a femtomolar affinity (1.3×10^{-15} M), which is in range with the copper affinities observed for copper chaperones in the plant [157].
- Extensive modelling of the transmembrane region of ETR1 with *ab initio* techniques, filtering with coevolutionarily predicted residue-residue contacts and refinement resulted in the first reported structure for the transmembrane region of an ethylene receptor. The resulting model relates well with previously determined loss-of-function mutants [137] and with tryptophan mutants generated in the study.
- MD simulations of the generated model in conjunction with free diffusing ethylene shows the binding of the ligand to the center of the monomer bundles in proximity of the included copper ions. This binding occurs next to residues that have been shown to be involved in ethylene binding, despite not including force-field terms to model the Dewar-Chatt-Duncanson olefin-copper interaction.

The obtained model not only sheds light into the structure of ETR1, but the same protocol and model could be used to study the whole family of the receptors. The inclusion of experimental information for the copper and ethylene binding sites could help in determining the validity of the obtained binding poses, and would help in setting up more complex simulations. Representing the copper-ethylene binding site at a quantum level could explain the

receptor changes while sensing the ligand. Additionally, the transmembrane region of the receptor leads to generate a full-length protein model. This could help in studying more complex regulations observed for the ethylene receptors, like receptor oligomerization.

4.3 Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence

Bleffert, F., Granzin, J., Caliskan, M., **Schott-Verdugo, S.**, Siebers, M., Thiele, B., Rahme, L., Dörmann, P., Gohlke, H. §, Batra-Safferling, R. §, Jaeger K-E and Kovačić, F. §

Submitted to *Proc. Natl. Acad. Sci. U. S. A.*, 2020

§ Authors contributed equally to this work.

For the original publication, see page 111.

Background

Pseudomonas aeruginosa is a relevant pathogenic bacteria, with multiple mechanism that increase its virulence, as described in **section 2.5**. The high number of infections caused by the bacteria has caused it to be classified as a high priority by the WHO in the development of antibiotics [163, 164]. As mentioned in the Introduction (**section 2.1**), the membrane composition is highly complex and requires regulation to maintain its structure, with recent studies suggesting its homeostasis as a potential target for pathogenic bacteria [237]. Since long it has been known that bacteria adapt their membrane composition in response to temperature changes [238, 239], with more recent descriptions of pH effects [240] and transitions from planktonic to biofilm of *Pseudomonas* [170]. Given the relevance of discovering new treatments for pathogens, studying pathways and enzymes involved in the lipid homeostasis in bacteria, particularly in pathogenic strains, has gained relevance in the last years [241].

Recently, a set of phospholipolytic enzymes from *Pseudomonas aeruginosa* have been discovered [174-177], from which PA2949 (now named PlaF) has shown promising evidence as a relevant virulence factor. In this study, the structure of PlaF is described and the mechanistic relevance of the dimerization in the regulation of its activity is analyzed. First, the cell localization and activity of the enzyme was studied in an overexpressing strain (*p-plaf*), showing that the protein is anchored in the inner membrane of the bacteria with its catalytic center exposed to the periplasmic space. A knockout for the enzyme ($\Delta plaf$) evidences a decreased virulence in infected *Drosophila melanogaster*, and reduced capability to swim and form biofilm, with 16 lipid species and presumable substrates of the enzyme significantly accumulated (showing preference for 35:1 lipids, but not particularly for head groups). The protein was crystalized as described previously [177], obtaining an asymmetric dimer with the cocrystalized fatty acids myristic and undecylic acid. The crystal dimer shows an interface with

the first α -helix of the protein, which comprises the single transmembrane helix present in the protein (see **Publication III**, page 122 for a detailed description). The dimerization was further analyzed by overexpressing the protein and crosslinking *in vivo* and by microscale thermophoresis (MST) of the purified protein, evidencing that the protein dimerizes in the bacteria, while also forming the complex in a protein-concentration dependent manner. Interestingly, the dimerization of the protein correlates with a decrease in the phospholipolytic activity and can be induced by incubation with fatty acids.

To analyze the monomer/dimer dynamics of PlaF, MD simulations in an explicit membrane environment were performed, starting from the protein dimer or monomers in different conformations. It was identified that the protein as a monomer has a preferred orientation in the membrane, while the dimer stays as such during the simulated time. Additionally, using biased MD simulations, we determined free energy profiles for the protein dimerization and for transition between the monomer orientations in the membrane, allowing to suggest a dimerization model for the protein *in vivo*.

Molecular dynamics simulations show a preferred orientation of the PlaF monomer in a membrane context

As PlaF was shown to be activated by monomerization, the dimer-to-monomer transition and the monomer dynamics in a membrane environment were studied with MD simulations. First, systems with the protein dimer as well as with the monomer were prepared. For the monomer simulations, we selected chain A from the asymmetric crystal of PlaF (PlaF_A) by removing PlaF_B from the crystal structure (di-PlaF). PlaF_B was used in control simulations. The orientation of the dimer and monomer in the membrane were predicted with the PPM method [242], resulting in the monomer having a tilted orientation towards the membrane, different than the monomer orientation in the dimer context. To study both orientations, we named the latter t-PlaF_A (**Fig. 19A**) and performed ten replicas of unbiased MD simulations of 2 μ s long starting for all the generated conformations in explicit membrane conditions. In all cases, the protein behaved well, with all dimer replicas keeping the inter dimer distance unchanged throughout the simulation time (**Fig. 19B**), suggesting the dimer is stable in the μ s timescale. In eight MD simulations that started with a monomer in a dimer-like orientation (PlaF_A), a transition to the tilted conformation was observed. In contrast, in all MD simulations starting from t-PlaF_A the protein remained in the tilted orientation (**Fig. 19C**). These results show that the tilted PlaF orientation is preferred over the orientation observed in the crystal

structure. The tilting of the protein on the membrane surface results particularly interesting, as the cocrystallized fatty acids in the crystal structure suggest that upon tilting, the catalytic tunnel would be perpendicular to the membrane plane, such that membrane substrates could access the catalytic site. In contrast, di-PlaF and PlaF_A have the residues lining the entrance of the active site tunnel at $> 5 \text{ \AA}$ above the membrane surface (**Fig. 19A**), hindering the access to the catalytic center.

Free energy profiles of PlaF dimerization and tilting suggest a monomeric state *in vivo*

Given that di-PlaF stays as a dimer during the unbiased simulations, and that the monomer transitions from the dimer-like (PlaF_A) to a tilted orientation (t-PlaF_A), the energetics of both processes were analyzed by computing the potential of mean force (PMF) using umbrella sampling and post-processing with the WHAM method [243]. All the simulation systems packed for the umbrella sampling simulations were generated using PACKMOL-Memgen, taking advantage of the easier handling of multiple systems in parallel (**Publication**

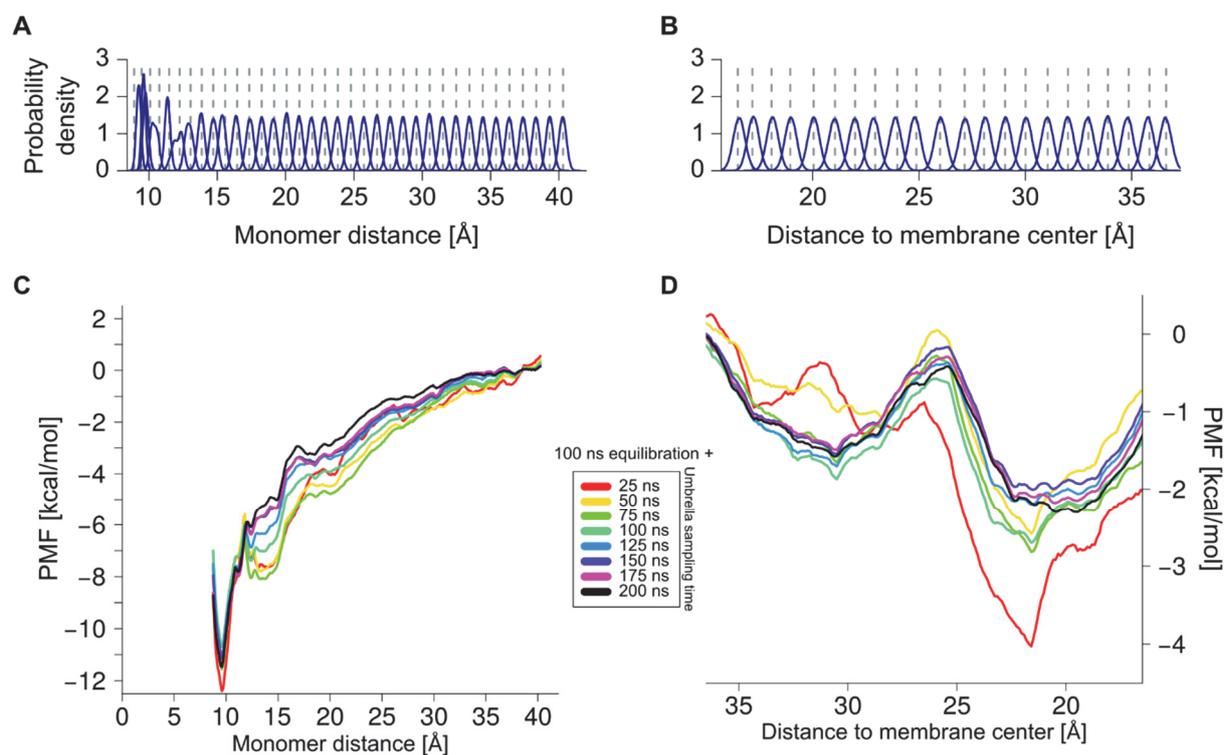


Figure 18. Distribution of reaction coordinate values obtained by umbrella sampling of dimer separation (**A**) and monomer tilting (**B**). The dashed lines represent the restrained distance used for each window. In both cases, a force constant of $4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was used, obtaining distributions with a median overlap of 8.2% and 8.6%, respectively. **C**) Convergence of the PMFs for dimer separation and **D**) monomer tilting. The plots show PMFs computed every 25 ns of umbrella simulations for each window; the first 100 ns of umbrella simulations were considered equilibration phase and removed. Figures taken from Publication III, see page 165.

I, section 4.1). As reaction coordinates, the distance between C α atoms of residues 25-38 of each chain (which in the crystal structure is 9.9 Å) and the distance between C α atoms of residues 33-37 and the center of the membrane slab were chosen, starting at 37 Å and until 15 Å (**Fig. 19E**, inset). In both cases overlapping distributions that permitted to perform a proper analysis were obtained (**Fig. 18A** and **18B**). The dimerization PMF was converged (**Fig. 18C**), revealing that di-PlaF is favored with a global minimum at 9.5 Å by -11.4 kcal mol⁻¹ over the monomer (**Fig. 19D**). The weakening of intermolecular interactions are observed from a distance of 15 Å onwards, disappearing beyond 27 Å. For the tilting process, an also converged PMF was obtained (**Fig. 18D**), revealing two minima at 19.6 and 30.6 Å and with t-PlaF_A favored over PlaF_A by 0.66 kcal mol⁻¹ (**Fig. 19E**). The rather small free energy barrier of ~1.2 kcal mol⁻¹ explains the rapid transition from PlaF_A to t-PlaF_A observed in the unbiased MD simulations.

To calculate the equilibrium constant of the protein dimerization in the membrane, the approach used by Provasi *et al.* [62] was implemented. The equilibrium constant for the protein tilting was calculated as the ratio of the partition functions bounded to the observed basins in the PMF (B1 and B2 in **Fig. 19E**) [244]. To have an idea of the overall monomerization and tilting process, the obtained dimerization constant ($K_a = 1.57 \times 10^7 \text{ Å}^2$, $\Delta G = -7.5 \pm 0.7 \text{ kcal mol}^{-1}$) can be combined with the tilting constant ($K_{\text{tilting}} = 3.35$, $\Delta G_{\text{tilting}} = -0.8 \pm 0.2 \text{ kcal mol}^{-1}$) to calculate the PlaF monomer-to-dimer dependency given the protein concentration in a membrane environment ($K = 1.4 \times 10^6 \text{ Å}^2$, **Fig. 19F**). Interestingly, at the experimental concentrations of PlaF in overexpressing conditions ($\sim 6.93 \times 10^{-6} \text{ PlaF Å}^{-2}$), the calculated amount of monomer is low, favoring a dimeric arrangement. This is in agreement with the crosslinking experiments performed in the study, and also explains the recovery of a crystal structure in a dimeric arrangement. At protein concentrations that are physiological (presumably 100-1000 fold less than in the overexpressing *p-plaf*), 86-98% of the protein is expected to be in a monomeric t-PlaF_A-like state. As the protein concentration in the bacteria lies at the calculated range, where relatively small changes in the effective concentration of the protein would change the dimer to monomer ratio, it is expected that by regulating the expression of the protein, or by modulating the membrane composition and dynamics, the activity of PlaF would be modulated. In this way, the bacteria would adapt its membrane, and have the observed consequences in virulence and biofilm formation. In addition, the inactivation of the protein at higher concentrations implies having a safeguarding mechanism,

in which the increase in the phospholipase would not imply a radical change in the membrane composition.

Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence

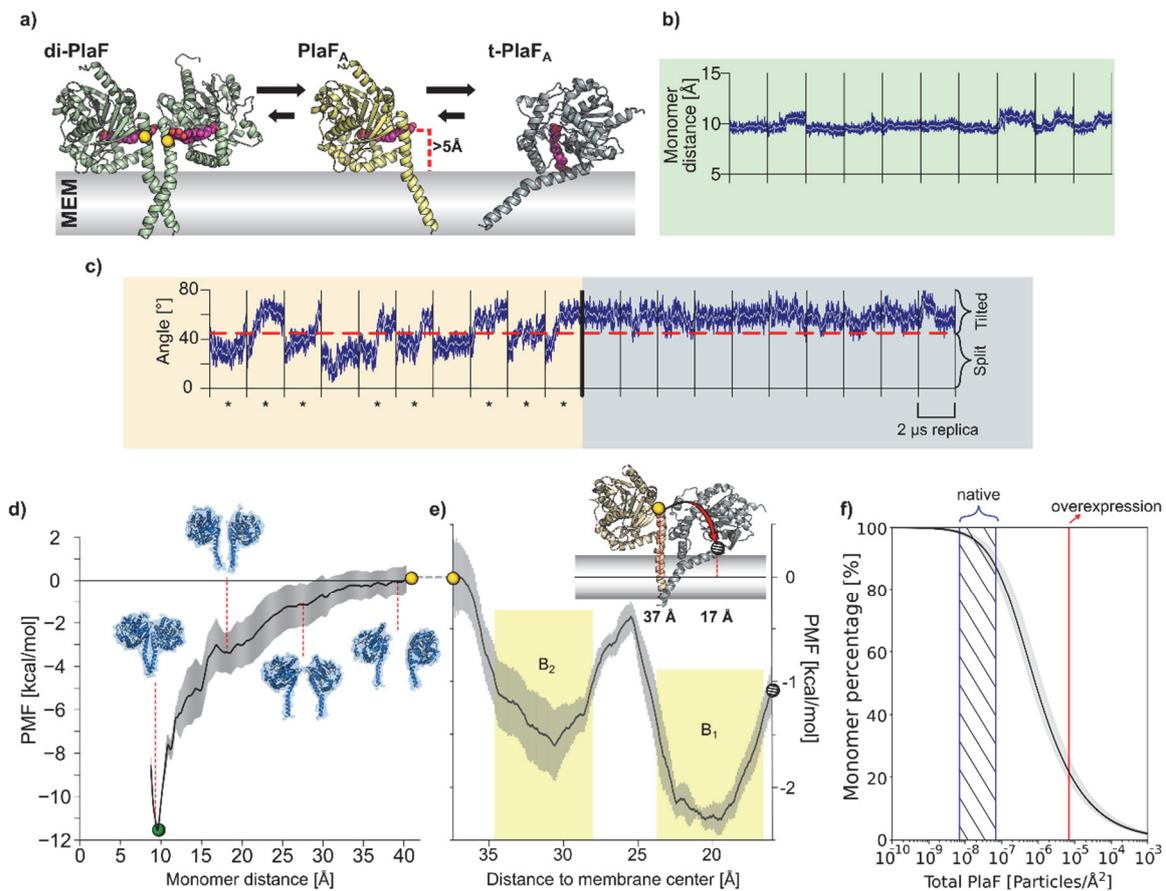


Figure 19. MD simulations and PMF computations for PlaF dimer and monomers. **A)** Structures used for MD simulations. di-PlaF: Crystal structure oriented in the membrane by the PPM method. Yellow spheres highlight the COM used to define the distance restraint for umbrella sampling simulations. PlaF_A: Chain A from PlaF dimer oriented as in the dimer. The entrance of the active site tunnel is more than 5 Å above the membrane bilayer surface. t-PlaF_A: Extracted monomer A was oriented using the PPM method. CocrySTALLIZED MYR and OG (not included in the simulations) are depicted in pink to highlight the orientation of the active site tunnel. Arrows between the structures reflect the predicted states of equilibria under physiological conditions in *P. aeruginosa*. **B)** Molecular dynamics simulations of di-PlaF. The dimer does not show a tendency to separate on the time scale of the MD simulations according to the distance between the COM of C_α atoms of residues 25 to 38 of each monomer (yellow spheres in panel A). **C)** Molecular dynamics simulations of s-PlaF and t-PlaF. Time course of the orientation of monomeric PlaF starting from the PlaF_A configuration (left) or t-PlaF_A configuration (right). In the former case, in eight out of ten trajectories the monomer ends in a tilted configuration (marked with *), while in the latter, no rearrangement occurs. This shows a significant tendency of the monomer to tilt, irrespective of the starting condition (McNemar's $X^2 = 6.125$, $p = 0.013$). **D)** Potential of mean force of dimer separation. The distance between the COM of C_α atoms of residues 25 to 38 of each chain (yellow spheres in panel A) was used as the reaction coordinate. The shaded area shows the standard error of the mean obtained by dividing the data into four independent parts of 50 ns each. Insets show representative structures at intermediate reaction coordinate values. **E)** Potential of mean force of monomer tilting. The distance between the COM of C_α atoms of residues 33 to 37 (yellow and gray spheres) and the COM of the C₁₈ of the oleic acid moieties of all lipids in the membrane (continuous horizontal line in the membrane slab) was used as a reaction coordinate. The gray shaded area shows the standard deviation of the mean obtained by dividing the data into four independent parts of 50 ns each. The yellow shaded regions correspond to the integration limits used to calculate K_{tilting} . The spheres in the PMF relate to monomer configurations shown in the inset. **F)** Percentage of PlaF monomer as a function of total PlaF concentration in the membrane according to the tilting model shown in A). The monomer percentage was computed according to the combined equilibrium constant of dimerization and tilting. The red line shows experimentally determined PlaF concentration under overexpressing conditions in *P. aeruginosa* p-plaF, while the blue dashed region shows an estimated span for the PlaF concentration in *P. aeruginosa* wild type. Figure taken from Publication III, see page 127.

Conclusions and significance

PlaF was characterized, a recently identified and crystalized PLA₁ enzyme of *Pseudomonas aeruginosa*, studying its enzymatic activity and its effects on the bacterial membrane. For this thesis, the PlaF monomer, dimer, and dimer to monomer dynamics with MD techniques were performed, identifying preferred states for the protein in explicit membranes. Calculated free energy profiles explain the protein dimerization and correlate with the obtained experimental information.

From this study, the most relevant results are:

- PlaF is an inner-membrane anchored protein, with its catalytic domain exposed to the periplasm. The activity of the enzyme is related with virulence, biofilm formation, swimming motility and with differential concentration of lipid species in the bacterial membrane.
- An asymmetric dimer of PlaF cocrystalized with fatty acids shows a single N-terminal transmembrane helix and the opening of the catalytic tunnel as the dimer interface. PlaF dimerizes *in vivo* and *in vitro* in crosslinking and microscale thermophoresis assays, respectively, showing a dimerization-dependent inactivation of the enzyme. The latter is enhanced by fatty acid products.
- Molecular dynamics simulations show that the dimer is stable as such in 2 μ s long simulations, while the monomer, if started from a dimer-like orientation, transitions to a tilted orientation with the opening of the catalytic tunnel pointing towards the membrane surface.
- Configurational free energy calculations for PlaF dimerization confirm an energy minima for the dimer structure, favoring it in about 11 kcal mol⁻¹. The tilted monomer, t-PlaF_A, is also favored by \sim 0.7 kcal mol⁻¹ with respect to the dimer like orientation, with an energy barrier of \sim 1.2 kcal mol⁻¹. The combination of both processes results in a predicted dimer to monomer equilibrium which transitions at concentrations close to those expected *in vivo*, suggesting that PlaF could be regulated by protein expression or membrane dynamics.

This novel regulation mechanism involves complex and tight regulation of the membrane environment from the bacteria, while also presenting the questions of how the substrates are loaded into the catalytic tunnel, and how the product is involved in enhancing the dimerization of PlaF.

4.4 Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits *Arabidopsis* iron acquisition

Khan, I., Gratz, R., Denezhkin, P., **Schott-Verdugo, S. N.**, Angrand, K., Genders, L., Basgaran, R.M., Fink-Straube, C., Brumbarova, T., Gohlke, H., Bauer, P. & Ivanov, R. (2019).

Plant Physiology, 2019, 180(3), 1564–1581.

For the original publication, see page 182.

Background

In **section 2.6**, the plants' difficulties in acquiring proper nutrients and in particular the root iron-uptake strategies were presented. Regulating the process that leads to iron absorption by the root is highly relevant, as a deficiency leads to reduced growth, while excessive amounts lead to toxicity and death [187, 188]. With this in mind, understanding the regulation and interactome of IRT1, the main transporter of iron in plant roots, is of utmost importance.

In this publication, a novel interactor of the IRT1 variable region (IRT1vr) was identified in yeast two-hybrid assays against a library of cDNA obtained from iron-deficient *Arabidopsis* roots. The identified protein, EHB1, corresponds to the recently described CAR protein family (**section 2.6**). As evaluated by BiFC, the interaction between IRT1vr and EHB1 is dependent on the CAR-signature domain (**Fig. 11B**). Coimmunoprecipitation showed that the interaction with IRT1 is calcium-dependent, and fluorescence colocalization showed that the interaction occurs in the plasma membrane. As shown for other C2-domain proteins [210], EHB1 is able to interact with membranes in a calcium-dependent manner on plant microsomes, showing binding to PtdIns and PtdIns4P but not to other lipids on membrane strips, and on PtdIns/PtdIns4P+PC liposomes; this interaction was not observed in pure PC liposomes. These results suggest the specific interaction with PtdIns and PtdIns4P in a calcium-dependent manner.

Interestingly, EHB1 is a negative regulator of iron uptake, as loss-of-function mutants for EHB1 show higher resistance to iron-deficient conditions, while overexpressing lines are sensitive and show signs of chlorosis. This regulation of the iron uptake depends on a direct interaction between EHB1 and IRT1, as iron-uptake deficient *fet3fet4* yeast double mutants are complemented by IRT1 or by IRT1 and EHB1 lacking the CAR-signature domain, but not when it is coexpressed with EHB1 wild type.

The obtained results beg the question of how the interaction between EHB1 and IRT1 occurs on the membrane. To unveil this, homology models of EHB1 and IRT1 were generated. MD simulations of EHB1 complexed with calcium in an explicit DOPC:DOPE:DOPG membrane were performed, allowing it to diffuse from the water phase onto the membrane surface. Binding of the protein “perpendicular” to the membrane plane through the calcium-binding site was observed, while also intermittently transitioning to a tilted orientation. The latter places the polybasic patch in proximity of the membrane surface, potentially enables the experimentally identified interaction with PtdIns, and places the CAR-signature domain in proximity of the membrane surface.

EHB1 binds to the membrane surface, fostering its interaction with IRT1

To obtain a model of the binding mode of EHB1 to the plant plasma membrane at the atomistic level, a homology model of EHB1 was generated with TopModel [245]. The model was based on the structures of the homologous proteins CAR1 and CAR4 (PDBs 5A52 and 4V29, respectively), and including calcium ions as found in CAR4 [209]. MD simulations in an explicit solvent/membrane environment were prepared by using PACKMOL-Memgen (**Publication I, section 4.1**), setting a composition of DOPC:DOPE:DOPG 4:4:1, which resembles the composition of a plasma membrane of a plant [246]. To allow the diffusion of EHB1, the starting configuration of the system had the protein at ~ 25 Å apart from the closest membrane surface. After ~ 1 μ s of simulation in five replicas, EHB1 bound to the membrane via the calcium-binding site in all cases, with four out of five replicas showing an additional potassium ion in the calcium-binding site (**Fig. 20A and B**). This is of particular interest, as the position is homologous to an additional calcium present in PKC (**Fig. 20B**) [208]. Though no PtdIns were included, EHB1 showed a strong tendency to tilt and to engage in interactions with the membrane with its sides, with one case in which the protein was tilted during the majority of the simulation (**Fig. 20C**). Surprisingly the interaction does not occur via the polybasic patch but rather through residues 51-62, 140-142, and 164-168, keeping the CAR-signature domain and the polybasic patch close to the membrane surface (**Fig. 20D**).

As described in **section 2.6**, it is known that C2-domains tend to interact with PtdInsP₂, with the canonical C2-domain of PKC interacting with PtdIns(4,5)P₂, so it was unexpected to see binding to PtdIns and PtdIns4P. Though differences in the polybasic binding site have been pointed out before, it was suggested that CAR proteins show unspecific binding to membranes [209]. When comparing the canonical polybasic patch with the one present in the EHB1 model,

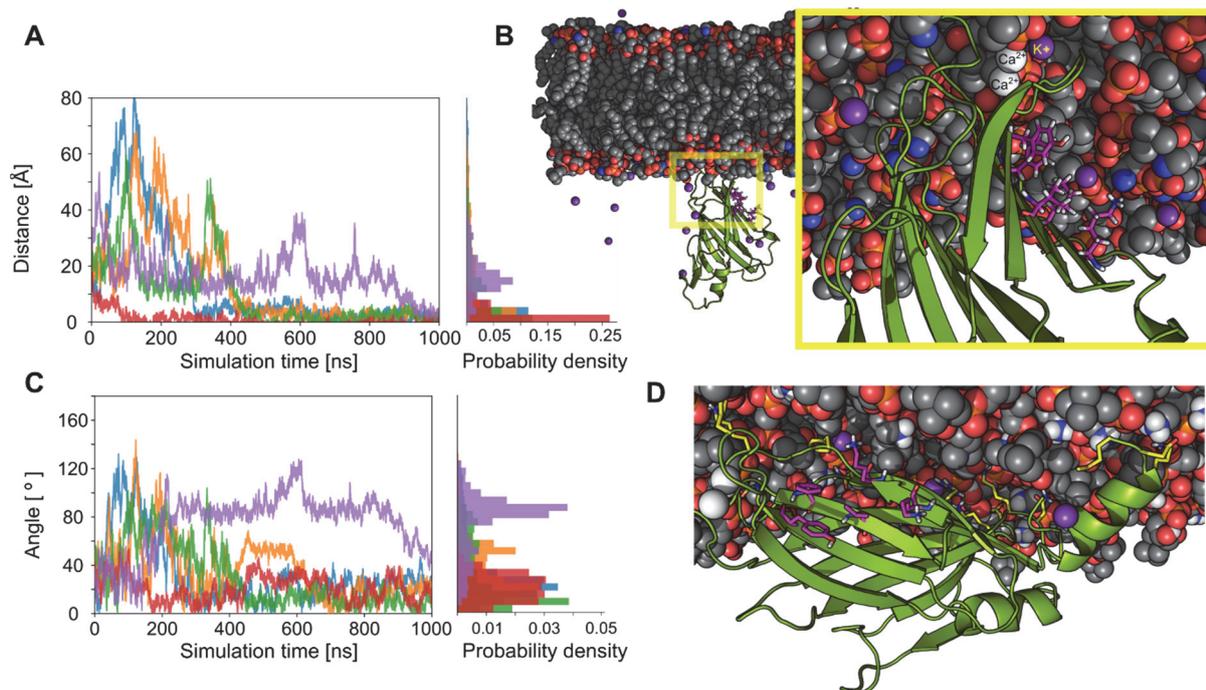


Figure 20. EHB1 membrane interactions through the calcium-binding site based on five independent MD simulations of 1- μ s length in the presence of a membrane bilayer with a 4:4:1 ratio of DOPC/DOPE/DOPG. **A)** Distance between the calcium ions bound to the protein and the center of mass of the phosphorous atoms of the phospholipids of the closest leaflet over the simulation time; in the starting configuration, EHB1 was placed 25 Å away from the membrane surface. In all replicas, EHB1 binds to the membrane surface through the calcium-binding site. **B)** Representative structure showing an EHB1 configuration (green) bound to the membrane (gray carbon atoms) through the calcium-binding site (Ca^{2+} : white spheres). An additional potassium ion was found to bind to an extra calcium-binding site (magenta). **C)** Tilt angle defined between the vectors formed by the center of mass of EHB1 with the complexed calcium ions and the membrane normal over the simulation time. A repeated, frequent, and, in one case, persistent tilting of EHB1 was observed in the replica simulations. **D)** Representative structure of the “parallel” configuration of EHB1 at the membrane. Residues that show direct interactions with the membrane are colored yellow; residues that correspond to the described polybasic patch are colored magenta. Figures taken from Publication IV, see page 189.

we noted that in PKC- α these amino acids can interact with all phosphate groups of PtdIns(4,5)P₂ (**Fig. 21A**), whereas in EHB1 only the negative charges of PtdIns4P (and PtdIns for that matter) would be properly compensated for by the binding site (**Fig. 21B** and **C**). Considering that in plants PtdIns4P is enriched in the plasma membrane [247], where EHB1 binds and IRT1 is located, the binding specificity relates well with the protein function. These variations in the binding site could explain the differences in lipid-binding specificity between other C2-domain proteins and EHB1 (**Fig. 21D**).

Overall, the obtained results allows us to propose a two-step model (**Fig. 21E**). First, an increase of cytosolic calcium triggers the binding of EHB1 to the membrane surface, “perpendicular” to the membrane plane. The simulations represent a condition where both calcium and the protein are at a high concentration in the proximity of the membrane surface.

The protein is then able to tilt over the membrane surface to a “parallel” orientation, probably stabilized by PtdIns4P *in vivo*, reducing the distance between the CAR-signature domain and the membrane surface. As the interaction occurs with the variable region, we generated a homology model of IRT1 and generated conformational sampling to have an idea of the placement of this region in a membrane context. The model, together with the attachment of EHB1 to the membrane surface, show that the interaction between them would be possible in the EHB1-tilted orientation, potentially causing the transporter inhibition by a mechanism to be described.

Conclusions and significance

In the search of novel interactors and regulators of the IRT1 iron transporter, EHB1 was identified, a protein previously associated with gravitropic and ABA responses [205, 212]. The protein negatively regulates IRT1 activity through a direct interaction between EHB1’s CAR-signature domain and IRT1_{vr}. Based on the experimental results, a binding mode of EHB1 to the membrane was simulated, and a binding model between the two proteins was generated.

In this study, the following were the most relevant results:

- Through yeast-two hybrid methods, EHB1 was identified as an IRT1_{vr} interactor. This interaction was further narrowed down to the CAR-signature domain of EHB1.
- EHB1 interacts with membranes and IRT1 in a calcium-dependent manner. The interaction with membranes showed specificity for PtdIns and PtdIns4P, and no interaction with pure PC liposomes.
- EHB1 negatively regulates the activity of IRT1, both *in planta* and in a yeast *fet3fet4* iron-uptake deficient double mutant. As the interaction between the proteins, this was dependent on the CAR-signature domain.
- Molecular dynamics simulations of a homology model of a EHB1/Ca²⁺ complex with an explicit membrane environment show the binding of the protein to the membrane surface “perpendicular” to the membrane plane.
- The bound protein is able to tilt on the membrane surface and to interact with charged residues on its sides, placing the polybasic batch and the CAR-signature domain in proximity of the membrane surface.

Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits Arabidopsis iron acquisition

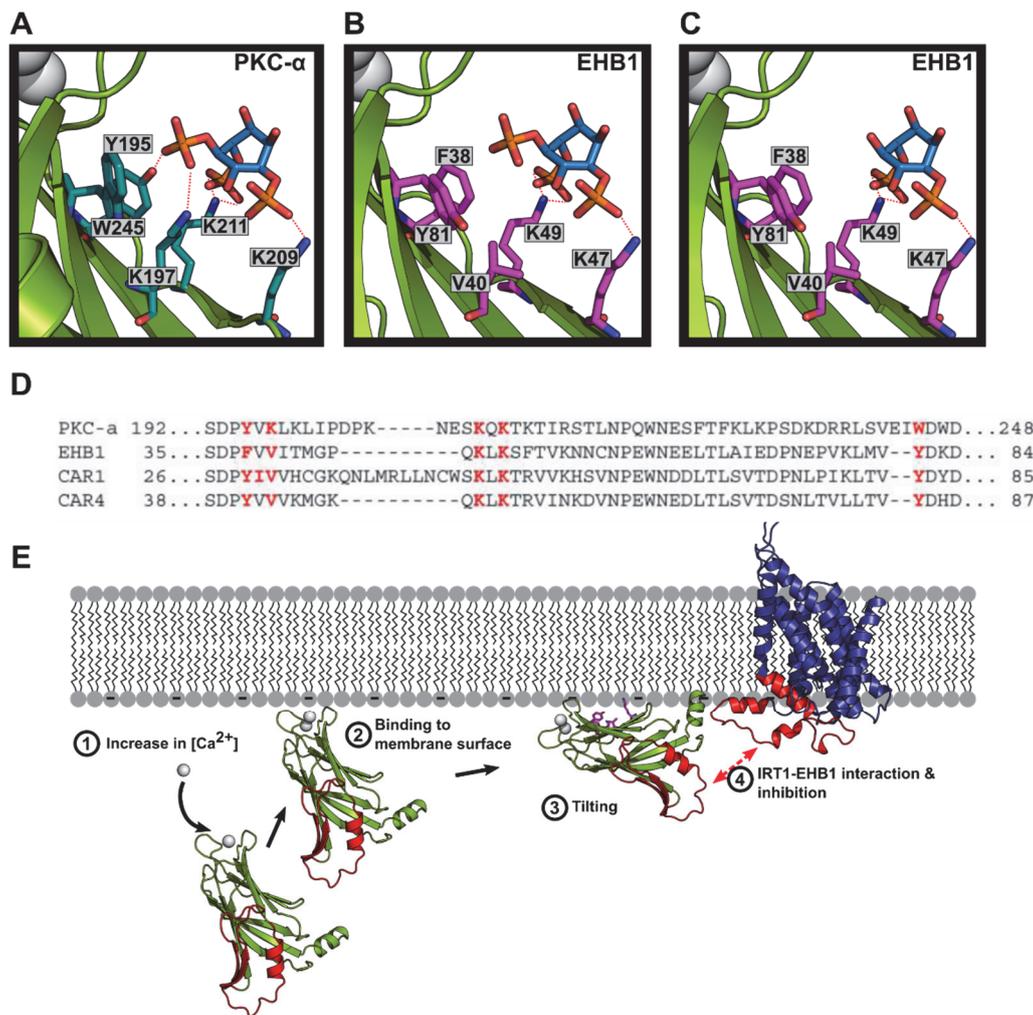


Figure 21. **A)** The C2-domain of PKC- α (PDB ID 3GPE) preferably binds to PtdIns(4,5)P₂ (blue carbon atoms) through its polybasic binding site (marine carbon atoms), which contains Tyr and Lys to interact with the phosphate at O5. **B)** In contrast, EHB1 has a Phe and Val at this site (magenta carbon atoms), which may explain the lack of EHB1 binding to PtdIns(4,5)P₂ in the strip analysis (Publication IV, see page 189). **C)** This site could bind PtdIns4P, as for this lipid no stabilization of the negative charge of the phosphate at O5 is required. The PtdIns(4,5)P₂ and PtdIns4P-bound structures were modeled based on PDB ID 3GPE. **D)** Alignment of the polybasic region of PKC- α , CAR1, CAR4 and EHB1. The amino acids described in **A)**-**C)** are highlighted in red. **E)** Hypothetical mechanism of the EHB1-IRT1 interaction using information from other C2 domain proteins, homology models for EHB1 and IRT1, and results from our MD simulations. Structures highlighted in red relate to protein regions identified here as relevant for the EHB1/IRT1 interaction. (1) An increase in the cytosolic calcium concentration leads to the occupation of the EHB1 calcium-binding sites. This increases EHB1's affinity for the plasma membrane surface (2), which contains negatively charged phospholipids. After binding in a "perpendicular" configuration, EHB1 would tilt to favor a "parallel" interaction (3). This would bring the CAR-signature domain in close proximity to the IRT1vr (4), allowing the two to interact, that way inhibiting iron uptake by a yet unknown mechanism. Figures taken from Publication IV, see pages 203 and 193.

The identification of EHB1 as an interactor with IRT1 and its regulation of iron uptake potentially connects its control to ABA stress-triggered responses (**section 2.6**). The data suggests that to prevent iron accumulation or absorption of other metals, EHB1 is recruited to the plasma membrane in a calcium-dependent manner to interact and inhibit IRT1, representing

Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1
inhibits Arabidopsis iron acquisition

a novel mechanism for the regulation of metal acquisition and for the prevention of metal
toxicity.

5 Summary and Perspectives

In the present thesis, three independent molecular systems that rely on intermolecular association and dissociation in a membrane environment to perform their function were investigated: **1)** the transmembrane domain of the *Arabidopsis thaliana* ethylene receptor ETR1, **2)** the membrane-anchored phospholipase A₁ PlaF of *Pseudomonas aeruginosa* and **3)** the membrane binding protein EHB1, which regulates the iron uptake activity of IRT1 in *Arabidopsis thaliana*.

First, for modeling the membrane environment of these systems, a software that generates atomic membrane/protein molecular structures was required. Until recently, researchers working in the AMBER/AmberTools package had to rely on setting up systems by hand or utilizing programs and web interfaces not included in the software suite. To facilitate the generation of MD-ready molecular systems, PACKMOL-Memgen was developed. The workflow was tested and validated on the KcsA potassium channel and on a Gram-negative inner-like membrane (**section 4.1, Publication I**). The program continues to be developed, including parameters for new lipid types and functions for packing more complex membrane systems.

Until now, no structure for the transmembrane domain of the ETR1 ethylene receptor has been obtained experimentally. Here, the first dimeric model for the transmembrane domain of the *Arabidopsis thaliana* ETR1 was generated, by using Rosetta membrane *ab initio* folding and residue-residue contacts predicted from coevolutionary information. The model was validated by the effects of a set of tryptophan mutants on the receptor α -helical content, and by rationalizing previously reported alanine mutants. With the determination of the 1:1 copper:monomer stoichiometry, molecular simulations showed binding of ethylene in proximity of the included copper ions. The ethylene binding sites correlate with previously loss-of-function mutants, and suggest a first step for the inhibition of the receptor (**section 4.2, Publication II**). The model opens the door to study more detailed receptor/copper/ethylene complexes, generate a full-receptor model and to study the regulation of the receptor by oligomerization.

Regarding the phospholipase A₁ PlaF from *Pseudomonas aeruginosa*, unbiased simulations showed that while the protein dimer stays as a dimer, the monomer orientation in the membrane tilts with respect to the crystal structure. Free energy profiles for the dimerization and tilting of the protein monomer in the membrane were calculated, confirming the favored

dimeric conformation and the tendency of the monomer to tilt. The latter is an interesting finding, as cocrystallized fatty acids indicate that the loading in such an orientation would be possible. From the obtained free energies, a membrane dimerization equilibrium was calculated, suggesting that *in vivo* the protein remains in a monomeric conformation. As the dimerization inhibits the enzyme, an increase in the protein concentration or probably changes in the membrane composition are able to regulate the protein dimerization and activity (**section 4.3, Publication III**).

Finally, molecular dynamics simulation of a homology model of the EHB1/Ca²⁺ complex in explicit membrane showed the free diffusion of the protein on the surface of a DOPC:DOPE:DOPG membrane. While the protein initially binds with the calcium binding site and perpendicular to the membrane plane, this binding mode allows the protein to tilt on the membrane surface. The tilted orientations would allow the binding of PtdIns4P to the polybasic patch and the interaction of the CAR-signature domain to the variable region of IRT1. This mechanism explains the experimental results where EHB1 inhibits IRT1 by a direct interaction, regulating the iron uptake (**section 4.4, Publication IV**).

As shown in this work, there are still methods that need to be generated in order to facilitate computational studies of membrane systems (PACKMOL-Memgen). Through the multiple systems studied, the present thesis shows how state-of-the-art structural bioinformatics and computational biophysics tools can be used to study complex membrane and membrane/protein systems. The available techniques allow to generate structures of membrane proteins with sparse residue-residue contact information (ETR1), to calculate thermodynamic free energies of dimerization and conformational changes of membrane proteins (PlaF), and to analyze the dynamics of association to membrane surfaces from peripheral membrane proteins (EHB1). These methods are extendable to systems with comparable problematics, especially considering the constant increase in computational power, providing atomic-scale simulations of systems that would otherwise be difficult if not impossible to study by conventional experimental approaches.

6 Acknowledgments

Acknowledgments are always bittersweet. There are so many people that influence what one's able to achieve, that it is impossible not to miss someone. All of this started in Talca, where thanks to Dr. Jans Alzate and the former Applied Sciences Program I could start my journey in Computational Biology. I would like to show my gratitude to my supervisor, Prof. Dr. Holger Gohlke. He was the one who convinced a series of people (including myself) to arrange and make possible my trip to Germany. For more than three years he has shown to somehow always have the time to listen and guide me if needed. I would also like to thank Prof. Dr. Ingo Dreyer. Thanks to his commitment, the collaboration between HHU and UTal has been possible.

In the HHU, I have had the pleasure to work with multiple people throughout these years. I would like to thank Prof. Dr. Georg Groth and Lena Müller for their collaboration during the work with ethylene receptors. That is what started it all. To Prof. Dr. Karl-Erich Jaeger and in particular to Dr. Filip Kovačić, for his passion regarding PlaF and its regulation. To Prof. Dr. Petra Bauer and to Dr. Rumen Ivanov, for their generosity while working in the EHB1 project. These collaborations have been possible thanks to the SFB1208, without which organizing this would have been extremely difficult.

All of the work done would not have been possible without the proper environment to work in, and I am very thankful of the human group that is CPCLab. I have had the chance to relate to people from different nationalities and completely different backgrounds, and still be able to call many of them my friends. Since my arrival, (in alphabetical order) Basti, Christina, Christoph, Christopher, Ciupie, Daniel, David, Denis, Giulia, Jonas, Lucas, Markus, Michele, Neha, Nicola, Sabah, Susanne and Tobi have been, at one moment or another, a great support for me. From them, I would like to thank in particular to the "Oberbilkers" for the nice times outside of the office, to Daniel Mulnaes, as his strong engagement has been very present all these years, and to Sabahuddin Ahmad, for being a great office partner and colleague. Thanks to David, Christopher, Christoph and Daniel for taking a look into some of these pages. Also, I would like to thank Peter, without whom a big part of our work would be impossible.

Finally, but not less important, I would like to thank to my family: my parents Eduardo and Any, my brother Eduardo, and to Francisca. They have supported me from the distance and understood the path I took, including my disconnection from the rest of the world. You all are really important to me.

Publication I

Publication I: PACKMOL-Memgen: A simple-to-use, generalized workflow for membrane-protein/lipid-bilayer system building

Stephan Schott-Verdugo and Holger Gohlke

Journal of Chemical Information and Modeling, 2019, 59, 2522–2528.

DOI: 10.1021/acs.jcim.9b00269

PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein–Lipid-Bilayer System Building

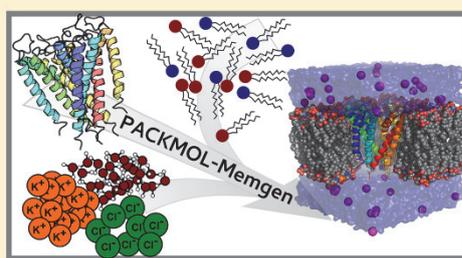
 Stephan Schott-Verdugo^{†,‡} and Holger Gohlke^{*,†,§}
[†]Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[‡]Centro de Bioinformática y Simulación Molecular (CBSM), Faculty of Engineering, Universidad de Talca, 1 Poniente 1141, Casilla 721, Talca, Chile

[§]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) & Institute for Complex Systems—Structural Biochemistry (ICS 6), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Supporting Information

ABSTRACT: We present PACKMOL-Memgen, a simple-to-use, generalized workflow for automated building of membrane-protein–lipid-bilayer systems based on open-source tools including Packmol, memembed, pdbremix, and AmberTools. Compared with web-interface-based related tools, PACKMOL-Memgen allows setup of multiple configurations of a system in a user-friendly and efficient manner within minutes. The generated systems are well-packed and thus well-suited as starting configurations in MD simulations under periodic boundary conditions, requiring only moderate equilibration times. PACKMOL-Memgen is distributed with AmberTools and runs on most computing platforms, and its output can also be used for CHARMM or adapted to other molecular-simulation packages.



INTRODUCTION

More than half of the current drug targets are integral membrane proteins, although they only represent about a third of the human proteome.¹ This highlights the functional importance of membrane proteins for physiology: membrane proteins participate in essential processes such as ion and molecule transport, signal transduction, and enzymatic catalysis.² These characteristics, together with an increase in available structures and computational resources, has motivated the proliferation of computational studies focusing on this type of protein³ and the lipid bilayer they are embedded in or associated with.⁴

One of the challenges of studying membrane systems or membrane proteins by computational techniques is building a system that is representative of the *in vivo* conditions. This is particularly true when one considers the anisotropic nature of the lipid bilayer, making the assembly of lipids to form a membrane and the placement of a protein within or at a membrane a nontrivial task (the membrane-protein-packing problem).^{5,6} Multiple tools have been developed to help in the process of generating such systems.^{5,6} In general, they can be subdivided into methods that pack the systems from scratch (self-assembly, grid, or geometry-based) and methods that use pre-equilibrated membrane systems to merge them with the components of interest (insertion or replacement methods).⁵ The former have the advantage of letting the user generate any composition they need, at the cost of requiring longer

simulation times in order to equilibrate the membrane system. This challenge is partially overcome by the latter methods, at the cost of having only a limited amount of pre-equilibrated membrane patches available, while still having to equilibrate the protein–membrane interface.

Popular tools that are widely used to solve the membrane-protein-packing problem are CHARMM-GUI,⁷ VMD (through the Membrane plugin),⁸ Maestro,⁹ and Packmol.¹⁰ Of these tools, CHARMM-GUI stands out for having multiple lipids available and a user-friendly web interface, making the packing process easier for newcomers but at the same time dependent on the web interface.⁷ This dependency turns into a challenge when a user needs to generate multiple systems with different proteins or different membrane compositions, or requires different starting configurations, as the process becomes time-consuming. Packmol, on the other hand, stands out for having functionalities that allow building complex and intricate simulation systems.¹⁰ These functionalities build upon solving the packing problem with the GENCAN optimizer, which minimizes an objective function that describes the molecular overlap.¹¹ To do so, appropriate PDB files for the individual system components and system-specific geometric constraints to restrict the packing are required as input.

Received: March 29, 2019

Published: May 23, 2019

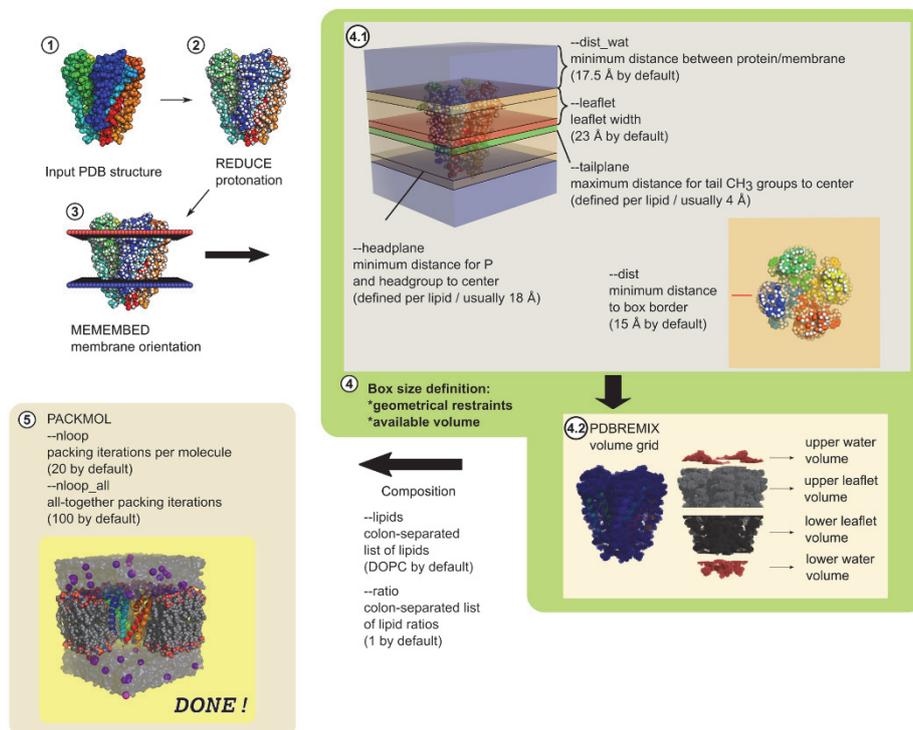


Figure 1. PACKMOL-Memgen workflow applied to the protein structure PDB ID 1BL8. The process comprises multiple (often optional) steps that are controlled by flags available in the software. The most important flags are mentioned in the figure. For a detailed description, see the [Workflow Description](#) section.

Together with the requirement of command line-usage, this makes applying Packmol difficult for nonexpert users.

Therefore, we developed PACKMOL-Memgen, a Python-based program that uses Packmol as the packing engine but wraps the main procedures required to build complex membrane systems, such that only single-line commands of the user are required. Further advantages of PACKMOL-Memgen are that its protocol is easily parallelizable when working with multiple proteins, membrane compositions, or starting configurations, the packing is achieved on the order of minutes for common system sizes, and it is user-friendly and easy to grasp. PACKMOL-Memgen generates formatted PDB files that can be used both with AMBER's LeAP with the Lipid17 force field^{12,13} and with VMD's psfgen with the CHARMM c36 force field.¹⁴

The program is distributed under a GPL license together with AmberTools¹³ (<http://ambermd.org/AmberTools.php>).

WORKFLOW DESCRIPTION

The general workflow of PACKMOL-Memgen is depicted in Figure 1. It follows similar principles to the CHARMM-GUI membrane builder.⁷ The user can decide to either pack a membrane-only system, or embed a protein structure into the membrane. For the latter, a PDB file of the protein structure

following the conventional formatting is used as input (step 1). The user needs to make sure that this PDB file represents the system of interest adequately (e.g., the protein configuration is representative of the process to be studied; ligands of interest are included, which can be kept for further processing with the flag `--keep`; and relevant protonation states are set). Next, unless otherwise specified, hydrogens are added to the input structure with "reduce",¹⁵ as this helps to correctly determine the protein volume in later steps (step 2). This step can be skipped by the user if the structure was previously protonated. To determine the placement of the protein within the membrane, the structure is oriented with "membed",¹⁶ which employs a knowledge-based potential function to mimic embedding of the protein in a membrane slab centered at the origin of the coordinate system (step 3). By default, the orientation is optimized over five cycles of the membed genetic algorithm. The automatic orientation can be skipped with the `--preoriented` flag.

From the placement of the protein, the size limits of the system to be packed are determined (step 4). Unless specified differently by the user, a distance of at least 15 Å to the boundaries in the *x-y* plane (`--dist`) and a water slab with a thickness of at least 17.5 Å above and below the membrane (`--dist_wat`) are used. The determined geometry is used

to prepare the placement of lipids, waters, ions, and possible additional solutes, which together with specified molecule concentrations (e.g., `--sol_con` or `--salt_con`) and lipid ratios (`--lipids`, `--ratio`) determines the number of molecules to be used. The placement of the center of the membrane slab is set to be on the *z*-coordinate origin, with the lipids oriented parallel to this axis and a leaflet width of 23 Å by default (`--leaflet`). To ensure the correct leaflet orientation, restricted volumes for (I) the phosphorus atom together with the terminal atom of the headgroup (`--headplane`) and (II) the carbon atoms of the terminal methyl groups of the tails (`--tailplane`) are set (step 4.1).

The default behavior of the program is to estimate the number of lipids per leaflet on the basis of the calculated leaflet area and the area per lipid (APL). APL values are obtained for systems at or close to 300 K from the literature,^{17–22} taken from the maximum value for lipids with the same headgroup, or set to 75 Å² otherwise. If lipid mixtures are used, the APL is computed as a weighted average according to the composition ratio. The number of water molecules in the upper and lower water slabs is determined on the basis of the respective volumes and the density of water. The volume occupied by the protein in each water slab, as well as in the membrane leaflets, is calculated by a grid approach derived from `pdbremix`²³ (step 4.2). The algorithm adds equidistant grid points (0.5 Å by default) within the radius of the atoms in the structure, representing the volume of the protein as a sum of cubes, with the grid spacing as the cube edge. This volume is used to reduce the previously calculated number of molecules in the different system sections by the numbers of lipids and waters that would occupy the protein volume. For this, the lipid molecular volume is obtained from the literature^{23,24–27} or estimated on the basis of the lengths of the acyl chains, as described previously.²⁸ If solutes are added, a similar approach is applied by using the `pdbremix`-computed volume to adapt the numbers of lipids or waters as appropriate. By default, potassium and chloride are added as counterions to reach electroneutrality with respect to the charges in the protein and the lipids. Extra ions can be included by specifying the `--salt` flag, and a concentration of 0.15 M in the water volume is used by default (`--salt_con`).

The user can select multiple lipids (`--lipids`) and define complex lipid mixtures at desired ratios (`--ratio`), including different compositions per leaflet. For this, corresponding colon-separated lists of lipids and ratios have to be provided, separated by two consecutive slashes (`//`), in case a different composition per leaflet is desired. For example, `--lipids POPE:POPC:POPS//POPC:POPE` `--ratio 3:1:1//4:1` would add a lower leaflet with POPE, POPC, and POPS in a 3:1:1 ratio and an upper leaflet with POPC and POPE in a 4:1 ratio. Table 1 lists lipids available in the current version, which can be extended with new lipids as long as parameters such as APL and molecular volume are provided. For example, parameters for phosphoinositides and lysophospholipids have been added to the AmberTools19 PACKMOL-Memgen release, extending the range of lipids in the Lipid17 force field toward those available in other lipid force fields²⁹ (Table S1).

In addition, multiple bilayers can be generated by calling the corresponding flags once per bilayer. This is useful if, for example, an electrochemical gradient is required to study a system, as in the case of computational electrophysiology.³⁰

Table 1. Lipids and Cholesterol Available in PACKMOL-Memgen^a

DAPA	DMPA	DPPA	PLPA	SDPA
DAPC	DMPC	DPPC	PLPC	SDPC
DAPE	DMPE	DPPE	PLPE	SDPE
DAPG	DMPG	DPPG	PLPG	SDPG
DAPS	DMPS	DPPS	PLPS	SDPS
DLPA	DOPA	DSPA	POPA	PSM
DLPC	DOPC	DSPC	POPC	SSM
DLPE	DOPE	DSPE	POPE	CHL1
DLPG	DOPG	DSPG	POPG	
DLPS	DOPS	DSPS	POPS	

^aAbbreviations used: DA, diarachidonoyl; DL, dilauroyl; DM, dimyristoyl; DO, dioleoyl; DP, dipalmitoyl; DS, distearyoyl; PL, palmitoyl lauroyl; PO, palmitoyl oleoyl; SD, stearyoyl docosahexaenoyl; PSM, palmitoyl sphingomyelin; SSM, stearyoyl sphingomyelin; CHL1, cholesterol; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PS, phosphatidylserine

To initiate molecule packing (step 5), the calculated number of molecules of each type and the geometrical constraints associated with each section of the system are written to a Packmol input file, and Packmol is started. If a protein is included, its position is considered fixed during the packing, and all other molecules will be packed such that they accommodate to the protein. By default, the system is packed with 20 iterations of the optimization algorithm per molecule section (or per entry in the Packmol input file) and 100 iterations system-wide. This is considerably less than the Packmol default value for iterations of 200 times the number of molecule sections but is usually more than enough to obtain the best possible solution from the GENCAN algorithm for the given input constraints. This is evidenced by the fact that, after the system-wide packing starts, usually only few cycles are required for the objective function of Packmol to reach a plateau (see Figure S1 for an example). Still, as the number of iterations required to pack a system depends on the system size and the complexity of the mixture, the number of iterations can be specified by the user if required (`--nloop` and `--nloop_all`).

The Packmol PDB file output is, by default, transformed into an AMBER-compatible file using an adapted version of `charmm lipid2amber.py` from Benjamin Madej;¹³ this step can be skipped if desired (`--charmm`). The resulting assembled system has the best possible packing given the input molecules and geometrical restraints but might not be an overlap-free solution. Thus, prior to using the system in MD simulations, usually a thorough energy minimization is required, which, if using the AMBER suite of programs, should be performed with the CPU code of `pmemd`.³¹ To simplify this step, functions to parametrize (`--parametrize`; using, by default, the ff14SB parameters³² for the protein, Lipid17 from Skjevik et al.^{12,13} for lipids, and the TIP3P water with AMBER 18 default ion parameters^{33–35}) and minimize (`--minimize`) simple protein-membrane systems after the packing are included.

Example Case 1: Bacteria-like Membrane. As a first example, the packing of a membrane bilayer composed of a mixture of DOPE–DOPG lipids with a 3:1 ratio is shown. This composition has been proposed as a representative model of the inner membrane of Gram-negative bacteria.³⁶ The packing of such a system can be performed with a single

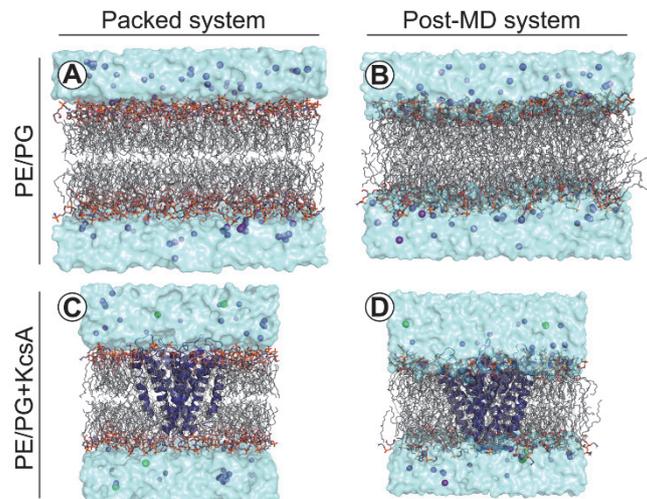


Figure 2. Representative structures for the membrane-only system (A,B) and the KcsA channel embedded in a membrane (C,D), generated as described in the main text, immediately after packing with PACKMOL-Memgen (A,C) and after 500 ns of MD simulations (B,D).

command line: packmol-memgen --ratio 3:1 --lipids DOPE:DOPG --distxy_fix 100.

Because no protein is included, a desired length in the x - y dimension has to be set (`--distxy_fix`, using 100 Å for this particular example). The resulting packed system is depicted in Figure 2A. The whole packing process took less than 13 min on a single i5-4590 CPU. In order to evaluate the stability of this system, MD simulations were performed as described in the Supporting Information, and a representative structure after 500 ns of production time is shown in Figure 2B. Parameters including the APL,³⁶ S_{CD} order parameter,^{72,37} membrane thickness,³⁸ electron density,^{38,39} and average lipid mean-square displacement¹² correspond with values reported previously (see Supplementary Methods and Figures S2–S4); this result demonstrates that the packed system is appropriate for performing all-atom MD simulations without requiring extensive equilibration times. Along these lines, the average x - y cross-sectional area is less than 5% smaller than the size of the system at the beginning of the production run, suggesting that the amount of lipids and geometrical constraints used for the packing are close to the equilibrium values.

Example Case 2: KcsA Potassium Channel. The packing is more challenging when a protein is included, because the other system components need to adapt to the protein during the packing itself and the equilibration phase of the MD simulations, and there may be different volumes in the membrane leaflets or water sections. As an example, we use KcsA, the potassium channel of *Streptomyces lividans*. The structure from PDB ID 1BL8 was capped at the termini, and a protonated glutamic acid in position 71 was used, as reported previously.^{40,41} As `--keep` was not used, crystallization ions were removed. The resulting packed system, including the protein, lipids (DOPE–DOPG, 3:1), water, and ions, is shown in Figure 2C. The packing of this system was performed with the following command line: packmol-memgen --pdb 1BL8.pdb --lipids DOPE:DOPG --ratio 3:1.

The packing time is comparable to the one of the membrane-only example. A script to pack this system is included with the software as a test case. Furthermore, animations of the packing process performed by Packmol are shown in Videos S1 and S2. MD simulations were performed to investigate whether the packed system is suitable as a starting structure. A representative structure after 500 ns of production is shown in Figure 2D. The protein structure shows slight deviations during the MD simulations, as indicated by the C_{α} -atom RMSD to the crystal structure (<2.5 Å) and the RMS average correlation reaching 1 Å within 200 ns (Figure S5). The membrane maintains its bilayer structure, as in the simulation with lipids only (Figure S4A,B). The KcsA channel has been investigated by MD simulations in detail before, particularly regarding the binding and flux of potassium ions into and through the selectivity filter.⁴² However, the exact mechanism is still a matter of debate.^{43,44} Although a thorough discussion of the dynamics of the protein and protein–potassium interactions is beyond the scope of the present work, the simulation results reveal that potassium stays bound mainly to the exit of the channel (Figures S4B and S6). These results resemble those of previously performed simulations of the protein started without potassium bound in the selectivity filter, where the protein stayed in an inactivated state.⁴⁵ Additionally, the selectivity filter adopts a conformation closer to a structure crystallized at low K^+ concentrations, where the selectivity filter is unoccupied¹⁶ (Figure S7). This finding is consistent with the observation that the channel becomes inactivated as a result of carbonyl repulsions in the selectivity filter and the hydration that follows.⁴⁵ Overall, the simulations show previously described structural dynamics for this protein, indicating that the packed system is a suitable starting structure.

DISCUSSION

In this study, we developed PACKMOL-Memgen, a simple-to-use, generalized workflow for automated building of membrane-protein–lipid-bilayer systems based on open-source tools including Packmol, memembed, pdbremix, and AmberTools. As demonstrated, the built systems are well-suited as starting configurations in MD simulations under periodic boundary conditions, requiring only moderate equilibration times. As discussed before,^{5,6} there is more than one way of building a membrane system for performing computational studies, and the choices may depend on the molecular-simulation package and the force field used, or the user preference. Until recently, users of the AMBER biomolecular-simulation programs were primarily dependent on webservers such as CHARMM-GUI to build membrane-protein–lipid-bilayer systems. Although user-friendly, providing support for many lipids, including glycolipids and lipopolysaccharides,⁴⁷ and yielding input for current molecular-simulation packages,^{48,49} the use of a web interface becomes inconvenient if many systems need to be built (e.g., when performing umbrella-sampling simulations), if different membrane compositions are to be tested, or if replicas are to be started from different configurations. A command line implementation helps to save time when such broad studies are performed, allowing scripting of system generation and, thereby, reducing the chances of making input errors. As shown in the example cases, the PACKMOL-Memgen workflow is both user-friendly and efficient, generating well-packed starting structures in minutes for typical simulation systems.

Although the workflow automatizes a large part of the building process, the user is still required to critically evaluate if the built system is representative from a molecular and physiological point of view. As an example, for packing a porin–lipid-bilayer system, one needs to consider the appropriate knowledge-based potential in memembed for β -barrel-type proteins (available through the `--barrel` flag). Furthermore, because the packing of lipids depends exclusively on reducing the molecular overlap, lipids may be inserted in the protein tunnel, which needs to be fixed manually prior to performing simulations.

PACKMOL-Memgen generates output with atom and residue names compatible with AMBER's Leap with the Lipid17 force field and with VMD's psfgen with the CHARMM c36 force field. Finally, the current implementation of the Windows Subsystem for Linux (WSL) in Windows 10 makes AmberTools and the included PACKMOL-Memgen workflow available on most computing platforms, including common Linux distributions and MacOS.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00269.

Supplementary methods on the molecular-dynamics simulations and lipid-diffusion calculation; figures showing changes in the target function during the packing process, plots of area per lipid and membrane thickness, lipid-order parameter, electron-density profile, RMSD and RAC of the performed MD simulations, 3D histograms of the potassium ions and water molecules, and structural comparison with crystal structures of the KcsA channel; and tables showing additional Lipid17

headgroups and the calculated lipid diffusion with periodic boundary corrections (PDF)

Video showing a top-down view of the packing process of the KcsA channel (MP4)

Video showing a side view of the packing process of the KcsA channel (MP4)

AUTHOR INFORMATION

Corresponding Author

*Tel.: (+49) 211 81 13662. Fax: (+49) 211 81 13847. E-mail: gohlke@uni-duesseldorf.de.

ORCID

Holger Gohlke: 0000-0001-8613-1447

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Projektnummer 267205415, SFB 1208 (project A03 to H.G.) and in part FOR 2518 (GO 1367/2-1). We are grateful for computational support and infrastructure provided by the “Zentrum für Informations- und Medientechnologie” (ZIM) at the Heinrich Heine University Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) to H.G. on the supercomputer JURECA at Jülich Supercomputing Centre (JSC, user IDs: HDD18 and HKF7). We acknowledge the developers of AMBER, Packmol, memembed, and pdbremix for making their software available to the scientific community

REFERENCES

- (1) Rask-Andersen, M.; Almen, M. S.; Schiöth, H. B. Trends in the exploitation of novel drug targets. *Nat. Rev. Drug Discovery* **2011**, *10*, 579–590.
- (2) Brown, D. G. Drug discovery strategies to outer membrane targets in Gram-negative pathogens. *Bioorg. Med. Chem.* **2016**, *24*, 6320–6331.
- (3) Almeida, J. G.; Preto, A. J.; Koukos, P. I.; Bonvin, A.; Moreira, I. S. Membrane proteins structures: A review on computational modeling tools. *Biochim. Biophys. Acta, Biomembr.* **2017**, *1859*, 2021–2039.
- (4) Stansfeld, P. J.; Sansom, M. S. Molecular simulation approaches to membrane proteins. *Structure* **2011**, *19*, 1562–1572.
- (5) Sommer, B. Membrane Packing Problems: A short Review on computational Membrane Modeling Methods and Tools. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302014.
- (6) Javanainen, M.; Martínez-Seara, H. Efficient preparation and analysis of membrane and membrane protein systems. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 2468–2482.
- (7) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.* **2009**, *97*, 50–58.
- (8) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (9) *Maestro*; Schrödinger, LLC: New York, NY, 2019.
- (10) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (11) Martínez, J. M.; Martínez, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *J. Comput. Chem.* **2003**, *24*, 819–825.

- (12) Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, L. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865–879.
- (13) Case, D.A.; Ben-Shalom, I. Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E., III; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; Gohlke, H.; Goetz, A.W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.J.; Merz, K.M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.C.; Wang, J.; Wei, H.; Wolf, R.M.; Wu, X.; Xiao, L.; York, D.M.; Kollman, P.A. AMBER 2018; University of California: San Francisco, 2018.
- (14) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; Mackerell, A. D., Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (15) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.
- (16) Nugent, T.; Jones, D. T. Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinf.* **2013**, *14*, 276.
- (17) Hung, W. C.; Lee, M. T.; Chen, F. Y.; Huang, H. W. The condensing effect of cholesterol in lipid bilayers. *Biophys. J.* **2007**, *92*, 3960–3967.
- (18) Petrache, H. I.; Dodd, S. W.; Brown, M. F. Area per Lipid and Acyl Length Distributions in Fluid Phosphatidylcholines Determined by 2H NMR Spectroscopy. *Biophys. J.* **2000**, *79*, 3172–3192.
- (19) Kucerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencier, J.; Jackson, A.; Katsaras, J. Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data. *Biophys. J.* **2008**, *95*, 2356–2367.
- (20) Hills, R. D.; McGlinchey, N. Model parameters for simulation of physiological lipids. *J. Comput. Chem.* **2016**, *37*, 1112–1118.
- (21) McIntosh, T. J.; Advani, S.; Burton, R. E.; Zhelev, D. V.; Needham, D.; Simon, S. a. Experimental tests for protrusion and undulation pressures in phospholipid bilayers. *Biochemistry* **1995**, *34*, 8520–8532.
- (22) Pan, J.; Heberle, F. A.; Tristram-Nagle, S.; Szymanski, M.; Koepfinger, M.; Katsaras, J.; Kucerka, N. Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering. *Biochim. Biophys. Acta, Biomembr.* **2012**, *1818*, 2135–2148.
- (23) Ho, B. pdbremix GitHub repository. <https://github.com/boscoh/pdbremix> (accessed June 3, 2018).
- (24) Murugova, T. N.; Balgavy, P. Molecular volumes of DOPC and DOPS in mixed bilayers of multilamellar vesicles. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18211–18216.
- (25) Greenwood, A. L.; Tristram-Nagle, S.; Nagle, J. F. Partial molecular volumes of lipids and cholesterol. *Chem. Phys. Lipids* **2006**, *143*, 1–10.
- (26) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *Biochim. Biophys. Acta, Rev. Biomembr.* **2000**, *1469*, 159–195.
- (27) Petrache, H. I.; Feller, S. E.; Nagle, J. F. Determination of component volumes of lipid bilayers from simulations. *Biophys. J.* **1997**, *72*, 2237–2242.
- (28) Koenig, B. W.; Gawrisch, K. Specific volumes of unsaturated phosphatidylcholines in the liquid crystalline lamellar phase. *Biochim. Biophys. Acta, Biomembr.* **2005**, *1715*, 65–70.
- (29) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Davila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35*, 1997–2004.
- (30) Kutzner, C.; Grubmüller, H.; de Groot, B. L.; Zachariae, U. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys. J.* **2011**, *101*, 809–17.
- (31) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.
- (32) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (34) Joung, I. S.; Cheatham, T. E., 3rd Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–41.
- (35) Li, P.; Song, L. F.; Merz, K. M. Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *J. Phys. Chem. B* **2015**, *119*, 883–895.
- (36) Murzyn, K.; Róg, T.; Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophys. J.* **2005**, *88*, 1091–1103.
- (37) Pluhackova, K.; Kirsch, S. A.; Han, J.; Sun, L.; Jiang, Z.; Unruh, T.; Bockmann, R. A. A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers. *J. Phys. Chem. B* **2016**, *120*, 3888–3903.
- (38) Hung, W. C.; Lee, M. T. The interaction of melittin with *E. coli* membrane: The role of cardiolipin. *Chinese J. Phys.* **2006**, *44*, 137–149.
- (39) Liu, Y.; Nagle, J. F. Diffuse scattering provides material parameters and electron density profiles of biomembranes. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2004**, *69*, No. 040901(R).
- (40) Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **1998**, *280*, 69–77.
- (41) Bhate, M. P.; McDermott, A. E. Protonation state of E71 in KcsA and its role for channel collapse and inactivation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15265–15270.
- (42) Fowler, P. W.; Tai, K.; Sansom, M. S. The selectivity of K⁺ ion channels: testing the hypotheses. *Biophys. J.* **2008**, *95*, 5062–5072.
- (43) Heer, F. T.; Posson, D. J.; Wojtas-Niziuski, W.; Nimigeon, C. M.; Bernèche, S. Mechanism of activation at the selectivity filter of the KcsA K⁺ channel. *eLife* **2017**, *6*, e25844.
- (44) Kopfer, D. A.; Song, C.; Gruene, T.; Sheldrick, G. M.; Zachariae, U.; de Groot, B. L. Ion permeation in K(+) channels occurs by direct Coulomb knock-on. *Science* **2014**, *346*, 352–355.
- (45) Boiteux, C.; Berneche, S. Absence of ion-binding affinity in the putatively inactivated low-[K⁺] structure of the KcsA potassium channel. *Structure* **2011**, *19*, 70–79.
- (46) Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* **2001**, *414*, 43–48.
- (47) Lee, J.; Patel, D. S.; Stahle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **2019**, *15*, 775–786.
- (48) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405–413.

(49) Jo, S.; Cheng, X.; Lee, J.; Kim, S.; Park, S. J.; Patel, D. S.; Beaven, A. H.; Lee, K. L.; Rui, H.; Park, S.; Lee, H. S.; Roux, B.; MacKerell, A. D., Jr.; Klauda, J. B.; Qi, Y.; Im, W. CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.* 2017, 38, 1114–1124.

Supporting Information

PACKMOL-Memgen: A simple-to-use generalized workflow for membrane- protein/lipid-bilayer system building

Stephan Schott-Verdugo^{1,2} and Holger Gohlke^{1,3*}

¹Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University
Düsseldorf, 40225 Düsseldorf, Germany

²Centro de Bioinformática y Simulación Molecular (CBSM), Faculty of Engineering,
University of Talca, 1 Poniente 1141, Casilla 721, Talca, Chile

³John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) &
Institute for Complex Systems - Structural Biochemistry (ICS 6), Forschungszentrum Jülich
GmbH, 52425 Jülich, Germany

*Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847

E-mail: gohlke@uni-duesseldorf.de

1 Supplementary Methods

Molecular dynamics simulations

The crystal structure of the KcsA potassium channel from *Streptomyces lividans* (PDB id. 1BL8) was embedded into a DOPE:DOPG 3:1 membrane, resembling the membrane of Gram-negative bacteria¹⁻², using the default parameters from Packmol-Memgen. The AMBER18 GPU PME implementation³⁻⁴ with ff14SB⁵, TIP3P⁶ and Lipid17 (⁷ and Skjevik *et al.*,⁸) parameters for the protein, water, and membrane lipids, respectively, were used. Five independent MD simulations of 0.5 μ s length were performed, constraining covalent bonds to hydrogens with the SHAKE algorithm⁹ and using a time step of 2 fs.

Five mixed steepest descent/conjugate gradient calculations with a maximum of 20,000 steps each were performed. Initially, the positions of the protein and membrane were restrained; later, restraints were kept on the protein atoms only and, finally, a minimization without restraints was performed. The temperature was maintained by using a Langevin thermostat¹⁰, with a friction coefficient of 1 ps⁻¹. The pressure, when required, was maintained using a semi-isotropic Berendsen barostat¹¹, coupling the membrane (x-y) plane. After minimization, the system was heated by gradually increasing the temperature from 10 to 100 K for 5 ps under NVT conditions, and from 100 to 300 K for 115 ps under NPT conditions at 1 bar. The thermalization process was continued for 5 ns under NPT conditions, after which production runs were started using the same conditions.

All simulations showed a stable membrane phase, as evidenced by electron density and order parameter calculations (SI Figures 3 and 4), with all double bonds in the unsaturated lipid chains staying in the corresponding *cis* isomer. The calculated area per lipid in all simulations was $63.2 \text{ \AA}^2 \pm 0.3$, similar to values reported previously².

Lipid in-plane diffusion

The in-plane diffusion of lipids was evaluated by measuring the average mean square displacement (MSD) of the phosphorous atoms of the phospholipids over time in the simulations without protein and using a linear fit to determine the diffusion coefficient D_{PBC} according to the Einstein relationship (eq. 1):

$$D_{PBC} = \lim_{t \rightarrow \infty} \left(\frac{1}{2n} \frac{d}{dt} \langle MSD(t)^2 \rangle \right) \quad \text{eq. 1,}$$

where t is the simulation time and n is the number of dimensions on which the displacement is evaluated (here, two for the membrane plane). Prior to calculating the MSD, the system was

centered with respect to the center of mass of the phospholipid molecules to remove overall translational motions and avoid diffusion overestimation¹². D_{PBC} is the diffusion coefficient of particles under periodic boundary conditions, which cannot be extrapolated to infinitely sized systems¹³⁻¹⁵. A correction was proposed by Yeh and Hummer¹³, which was recently modified by Vögele and Hummer¹⁵ to take into account the Saffman-Delbrück model for membrane diffusion (eq. 2):

$$D_0 \approx D_{PBC} - \frac{k_B T}{4\pi\eta_m} \frac{\ln\left(\frac{L}{L_{SD}}\right) - \ln\left(1 + \frac{\pi H}{2L_{SD}}\right) - e + 1}{1 + \frac{H}{L_{SD}}}, \quad L_{SD} = \frac{\eta_m}{2\eta_f}, \quad \eta_m = h \cdot \eta \quad \text{eq. 2,}$$

which is a fair approximation when $H < L$ and $H < L_{SD}$, and where k_B is the Boltzmann constant, T is the absolute temperature, η_m is the membrane surface viscosity, L is the lateral length of the system, H is half the water height between periodic membrane images, and L_{SD} is the Saffman-Delbrück length with η_f being the water viscosity. Experimental values of $\eta_m = 1.27$ nPa s m (with $h = 39.6$ Å [see SI Figure 2B] and $\eta = 320$ mPa s¹⁶) and $\eta_f = 0.85$ mPa s¹⁷ were used, resulting in $L_{SD} = 746$ nm within the range of reported values¹⁴. Every simulated leaflet was taken as an independent sample, resulting in $n = 5 * 2 = 10$, fitting on the last 300 ns of each replica. For the analysis, the NPT simulations using the Langevin thermostat were applied, considering that, with a moderate collision frequency of 1 ps⁻¹, about 90% of the self-diffusion of particles is recovered, even when the Langevin thermostat periodically affects the momentum of particles^{7, 18-19}.

Results are reported in Table S2. All analyses were performed with CPPTRAJ²⁰ and in-house Python scripts.

2 Supplementary Figures

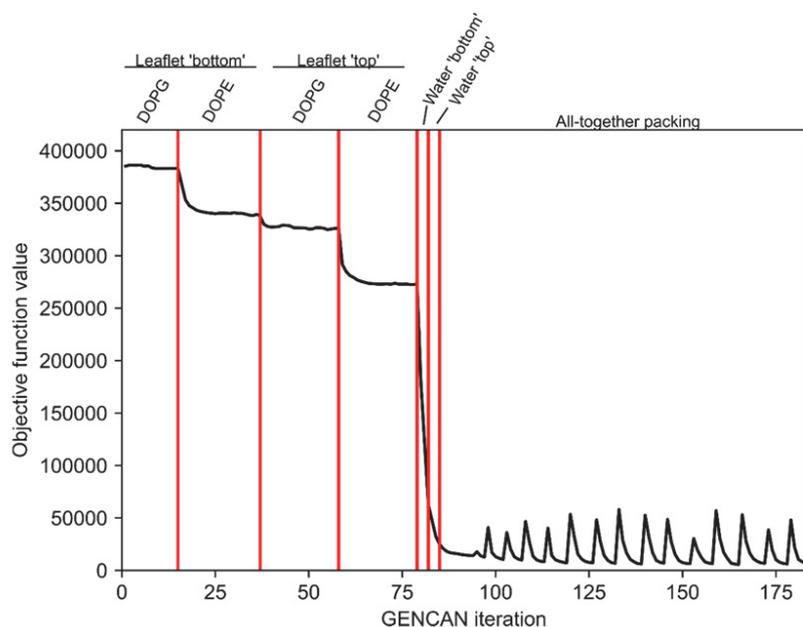


Figure S1. GENCAN objective function value for the packing of the KcsA potassium channel in a DOPE:DOPG 3:1 membrane. The red lines delimit the packing of individual molecular sections of the system. The all-together packing reaches a plateau, with spikes that correspond to molecular movements performed by Packmol in the search of a better solution. No overall packing of ions occurs, as their random initial placements remained a solution of the packing problem.

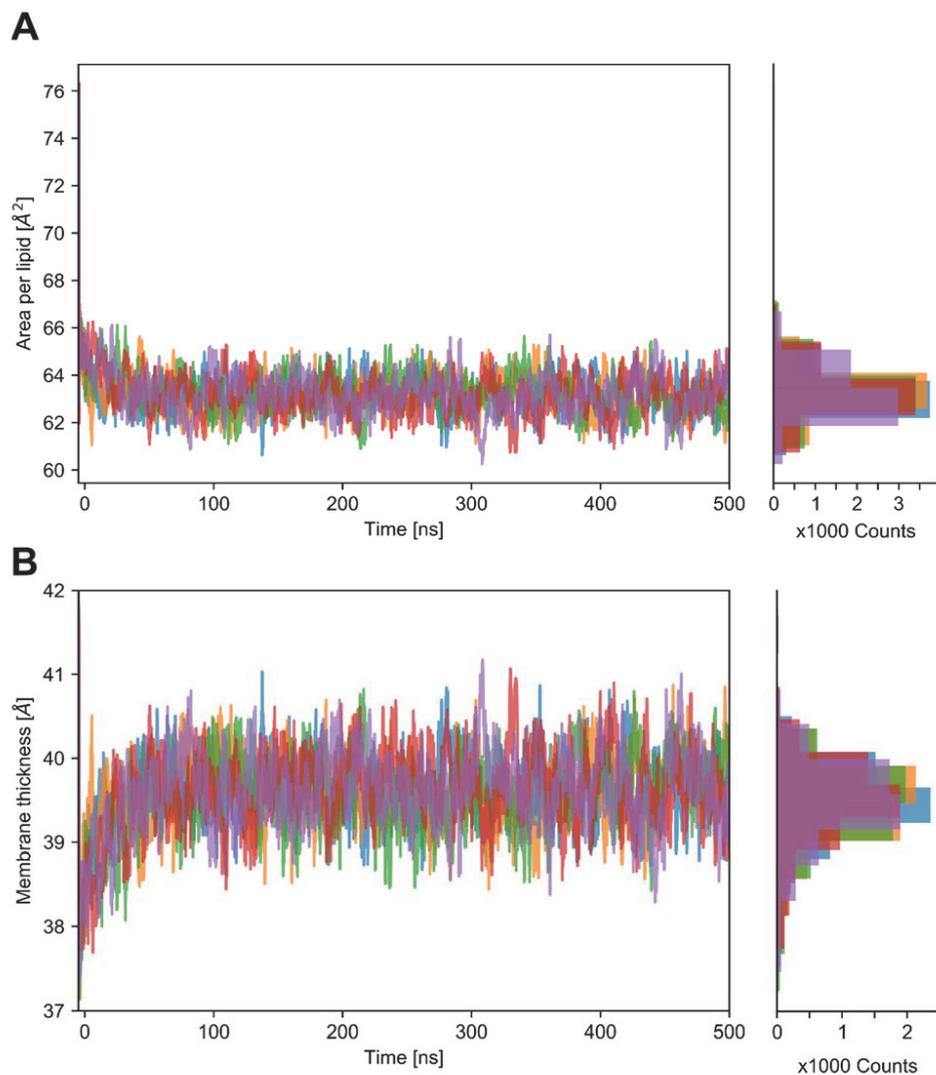


Figure S2. Average area per lipid (A) and membrane thickness (B) in simulations of a membrane of a Gram-negative organism. The individual values for the different replicas are shown as shaded colors. Both parameters reach a plateau after ~ 50 ns, in line with previously reported values²¹. The reported values before 0 ns correspond to the equilibration steps.

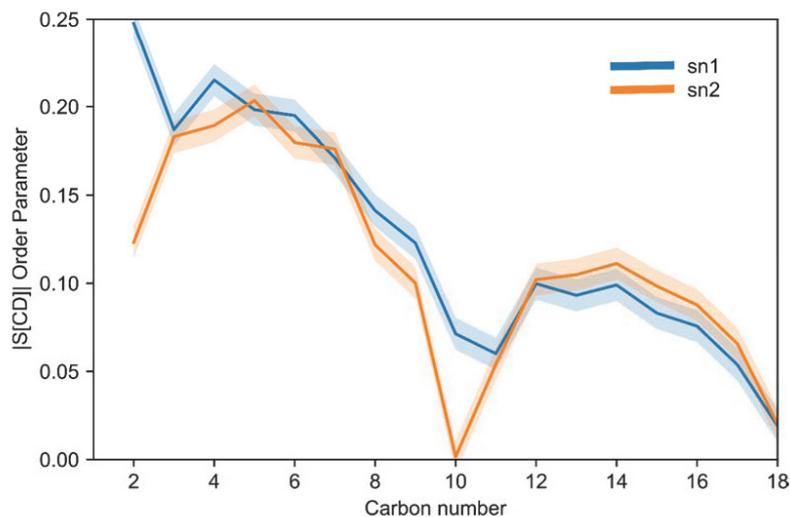


Figure S3. Lipid order parameter in the DOPE:DOPG 3:1 simulations. The bold line shows the average for all hydrogens of the *sn1* (blue) or *sn2* (orange) aliphatic tails over five replicas over 500 ns of simulation time. The shaded area shows the propagated standard error of the mean.

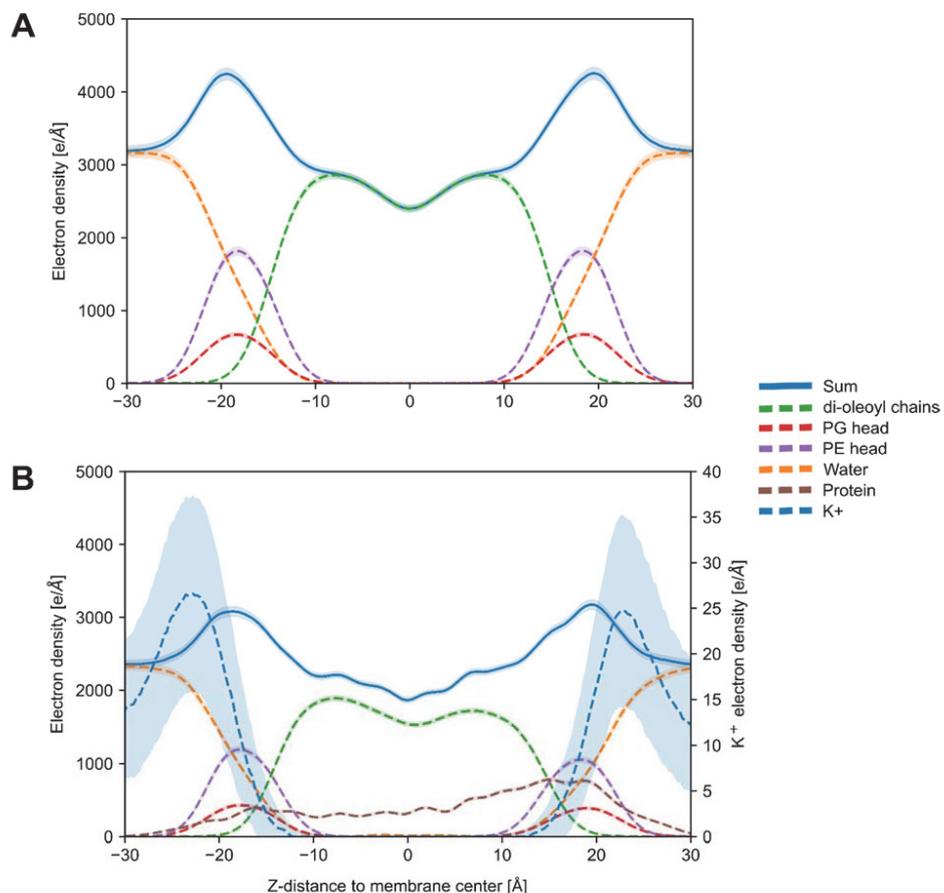


Figure S4. Electron density profiles of the DOPE:DOPG 3:1 bilayer in a membrane-only simulation (A) and with the KcsA potassium channel embedded (B). The electron density was calculated with CPPTRAJ in that the number of electrons over all atoms in a section across a desired axis of the simulation box is computed, thereby compensating for the charge of each included atom. The profiles show the average over five replicas of 500 ns simulation time, with the shaded area representing the propagated standard error of the mean.

7

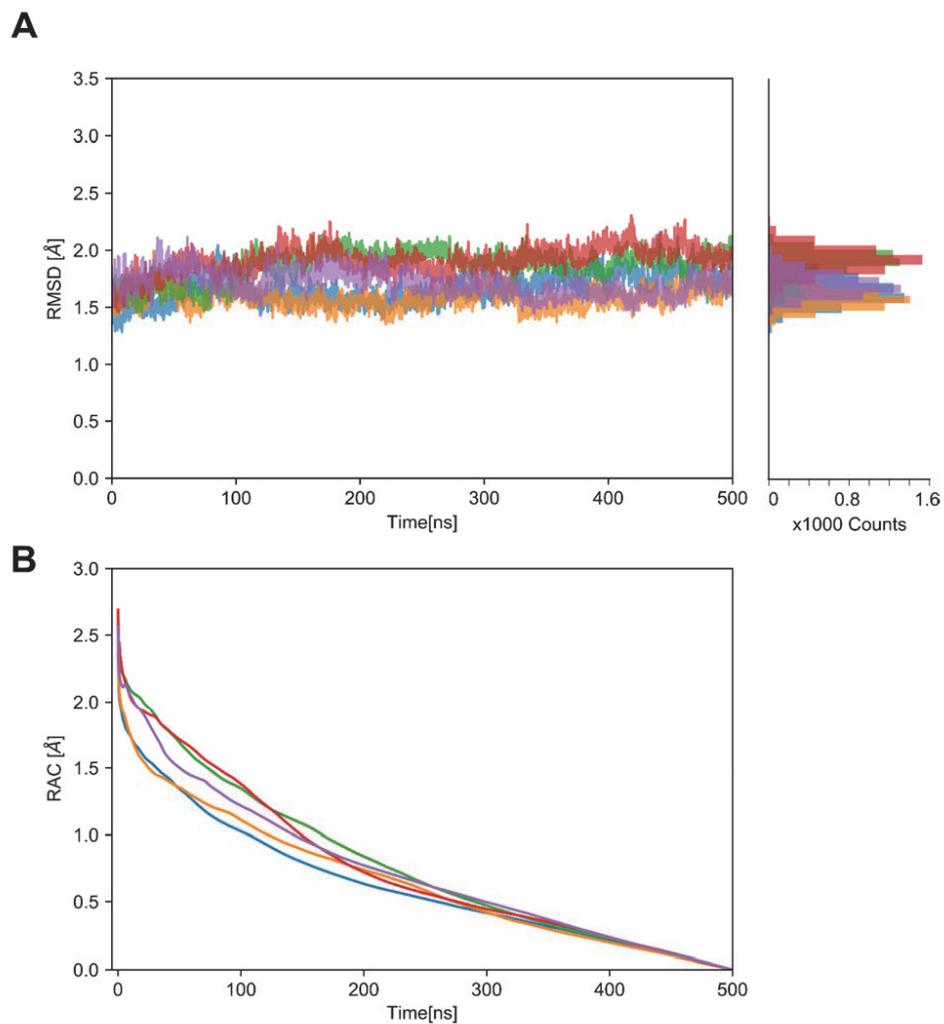


Figure S5. C_{α} root mean square deviation (RMSD, A) and RMS average correlation (RAC, B) of the simulated KcsA potassium channel along the trajectory. The RMSD is calculated with respect to the structure in the initially packed system. Both analyses were performed with CPPTRAJ.

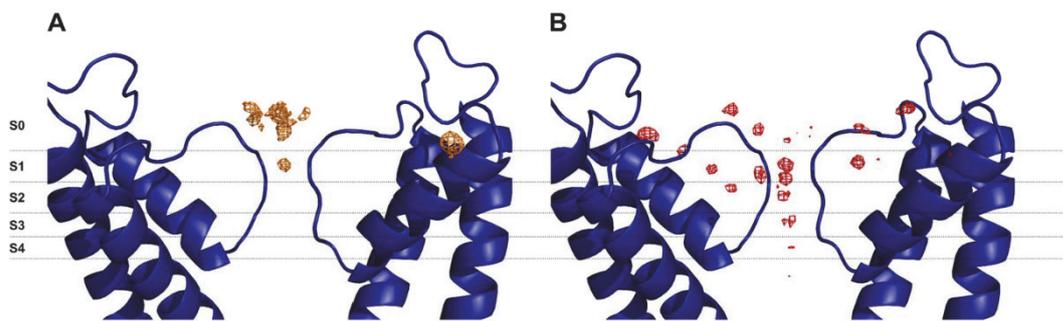


Figure S6. 3D histograms of the positions of potassium (A) and water (B) in the selectivity filter. Potassium ions remain mainly in the extracellular vestibule of the protein (S0), while also accessing position S1. Water is able to hydrate the unoccupied sites.

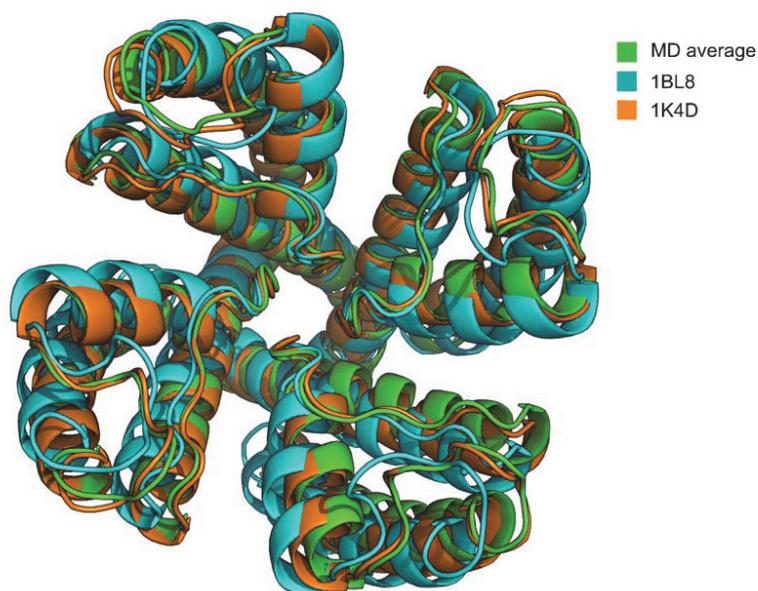


Figure S7. The selectivity filter assumes an inactivated structure during simulations started without K^+ bound. In green, the average structure for the largest cluster of all replicas is shown. In cyan and in orange, the structures of the channel in an activated (1BL8) and inactivated (1K4D) state are shown, respectively. The structures were superimposed with the selectivity filter (residues 75-79) as the reference. The RMSD between the selectivity filters of the MD average and 1BL8 and the MD average and 1K4D are 1.1 and 0.6 Å, respectively.

3 Supplementary Tables

Table S1. PACKMOL-Memgen lipid headgroups extending Lipid17.^a

Headgroup	Variants
Lyso-PC	Acyl chains in positions <i>sn</i> -1 or <i>sn</i> -2
Lyso-PE	Acyl chains in positions <i>sn</i> -1 or <i>sn</i> -2
Lyso-PG	Acyl chains in positions <i>sn</i> -1 or <i>sn</i> -2
Lyso-PS	Acyl chains in positions <i>sn</i> -1 or <i>sn</i> -2
PI3P	Phosphate group deprotonated or singly-protonated.
PI4P	Phosphate group deprotonated or singly-protonated.
PI5P	Phosphate group deprotonated or singly-protonated.
PI3,4P	Phosphate groups deprotonated or singly-protonated in 3, 4, or 3+4
PI3,5P	Phosphate groups deprotonated or singly-protonated in 3, 5, or 3+5
PI4,5P	Phosphate groups deprotonated or singly-protonated in 4, 5, or 4+5
PI3,4,5P	Phosphate groups deprotonated or singly-protonated in 3, 4, 5, 3+4, 3+5, 4+5 or 3+4+5

^a Abbreviations used are: PC = phosphatidylcholine; PE = phosphatidylethanolamine; PG = phosphatidylglycerol; PS = phosphatidylserine; PI = phosphatidylinositol; 3P = 3-phosphate; 4P = 4-phosphate; 5P = 5-phosphate; 3,4P = 3,4-diphosphate; 3,5P = 3,5-diphosphate; 4,5P = 4,5-diphosphate; 3,4,5P = 3,4,5-triphosphate

Table S2. Diffusion coefficients measured from NPT DOPE:DOPG 3:1 MD simulations.

Lipid	$D_{PBC}^{a,b}$	$D_0^{a,b}$	D_{exp}^a	Ref. ^c	Remarks ^d			
DOPG	4.5±0.4	6.0±0.4	5.6-8.2	22	NMR; range between 294 and 303 K at 0.15 M NaCl			
DOPE	3.8±0.3	5.4±0.3	~5.2	23	FRAP; data from plot; NBD-PE in <i>E. coli</i> extracts with 78:14 PE:PG at 25°C			
Mixture average	4.0±0.2	5.5±0.2				~8	24	NMR; data from plot; <i>E. coli</i> membrane lateral diffusion at 27°C.
						~6	25	DiD fluorescence correlation spectroscopy; data from plot; DOPE:DOPG:DOPC 2:1:1 at 20°C

^a In $\mu\text{m}^2 \text{s}^{-1}$.^b Mean \pm standard deviation.^c Reference from which experimental values were taken.^d Reported are the main experimental technique and measurement conditions.

4 Supplementary References

1. Conrad, R. S.; Gilleland, H. E., Jr., Lipid alterations in cell envelopes of polymyxin-resistant *Pseudomonas aeruginosa* isolates. *Journal of Bacteriology* **1981**, *148*, 487-497.
2. Murzyn, K.; Róg, T.; Pasenkiewicz-Gierula, M., Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophys. J.* **2005**, *88*, 1091-1103.
3. Le Grand, S.; Götz, A. W.; Walker, R. C., SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374-380.
4. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
5. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.
6. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935.
7. Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C., Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865-879.
8. D.A. Case, D. S. C., T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D.; Greene, N. H., S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo.; D. Mermelstein, K. M. M., G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A.; Roitberg, C. S., C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X.; Wu, L. X., D.M. York and P.A. Kollman, AMBER 2017 Reference Manual. University of California, 2017.
9. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
10. Quigley, D.; Probert, M. I., Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* **2004**, *120*, 11432-11441.
11. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
12. Pluhackova, K.; Kirsch, S. A.; Han, J.; Sun, L.; Jiang, Z.; Unruh, T.; Bockmann, R. A., A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers. *J. Phys. Chem. B* **2016**, *120*, 3888-3903.
13. Yeh, I.-C.; Hummer, G., System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873-15879.
14. Camley, B. A.; Lerner, M. G.; Pastor, R. W.; Brown, F. L. H., Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. *J. Chem. Phys.* **2015**, *143*, 1-12.
15. Vögele, M.; Hummer, G., Divergent Diffusion Coefficients in Simulations of Fluids and Lipid Membranes. *J. Phys. Chem. B* **2016**, *120*, 8722-8732.
16. Mika, J. T.; Thompson, A. J.; Dent, M. R.; Brooks, N. J.; Michiels, J.; Hofkens, J.; Kuimova, M. K., Measuring the Viscosity of the *Escherichia coli* Plasma Membrane Using Molecular Rotors. *Biophys. J.* **2016**, *111*, 1528-1540.

17. Cooper, J.; Dooley, R., Release of the IAPWS formulation 2008 for the viscosity of ordinary water substance. *The International Association for the Properties of Water and Steam* **2008**.
18. Ljubetic, A.; Urbancic, I.; Strancar, J., Recovering position-dependent diffusion from biased molecular dynamics simulations. *J. Chem. Phys.* **2014**, *140*, 1-11.
19. Basconi, J. E.; Shirts, M. R., Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2887-2899.
20. Roe, D. R.; Cheatham, T. E., 3rd, PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
21. Anézo, C.; de Vries, A. H.; Höltje, H.-D.; Tieleman, D. P.; Marrink, S.-J., Methodological Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B* **2003**, *107*, 9424-9433.
22. Filippov, A.; Oradd, G.; Lindblom, G., Effect of NaCl and CaCl₂ on the lateral diffusion of zwitterionic and anionic lipids in bilayers. *Chem. Phys. Lipids* **2009**, *159*, 81-87.
23. Jin, A. J.; Edidin, M.; Nossal, R.; Gershfeld, N. L., A singular state of membrane lipids at cell growth temperatures. *Biochemistry* **1999**, *38*, 13275-13278.
24. Lindblom, G.; Orädd, G.; Rilfors, L.; Morein, S., Regulation of lipid Composition in *Acholeplasma laidlawii* and *Escherichia coli* Membranes: NMR studies of lipid lateral diffusion at different growth temperatures. *Biochemistry* **2002**, *41*, 11512-11515.
25. Ramadurai, S.; Duurkens, R.; Krasnikov, V. V.; Poolman, B., Lateral diffusion of membrane proteins: consequences of hydrophobic mismatch and lipid composition. *Biophys. J.* **2010**, *99*, 1482-1489.

Publication II

Publication II: Structural model of the ETR1 ethylene receptor transmembrane sensor domain

Stephan Schott-Verdugo[§], Lena Müller[§], Elisa Classen, Holger Gohlke[‡] and Georg Groth[‡]

Scientific Reports, 2019, 9(1), 1–14.

DOI: 10.1038/s41598-019-45189-w

^{§,‡} Authors contributed equally to this work.

SCIENTIFIC REPORTS

OPEN

Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain

Stephan Schott-Verdugo^{1,2}, Lena Müller³, Elisa Classen^{3,4}, Holger Gohlke^{1,5,6} & Georg Groth^{3,6}

Received: 5 February 2019

Accepted: 3 June 2019

Published online: 20 June 2019

The structure, mechanism of action and copper stoichiometry of the transmembrane sensor domain of the plant ethylene receptor ETR1 and homologs have remained elusive, hampering the understanding on how the perception of the plant hormone ethylene is transformed into a downstream signal. We generated the first structural model of the transmembrane sensor domain of ETR1 by integrating *ab initio* structure prediction and coevolutionary information. To refine and independently validate the model, we determined protein-related copper stoichiometries on purified receptor preparations and explored the helix arrangement by tryptophan scanning mutagenesis. All-atom molecular dynamics simulations of the dimeric model reveal how ethylene can bind proximal to the copper ions in the receptor, illustrating the initial stages of the ethylene perception process.

Plant hormones regulate diverse aspects from growth and development to biotic and abiotic responses throughout the lifespan of plants. Signaling pathways and receptors have been identified to molecular detail for several of these signaling molecules, providing a basic integrated understanding on the mechanisms of hormone actions. High resolution structural information on receptors such as the brassinosteroid receptor BRI1^{1,2} or abscisic acid receptor PYR1/PYL1^{3,4} has been instrumental to understand receptor activation mechanisms^{5,6} or to identify small molecule mimetics activating receptors and related plant responses⁷. However, for most plant receptors such a level of structural and mechanistic understanding is still lacking.

Fruit ripening, senescence, and decay are highly relevant agronomical processes regulated by the plant hormone ethylene. Ethylene is perceived by a receptor family composed of ETR1, ERS1, ETR2, ERS2 and EIN4 in *Arabidopsis*⁸, which, in their functional state, form dimers and higher-molecular weight oligomers at the ER membrane⁹. These receptors show similarity to bacterial two-component histidine kinase receptors. All members of the receptor family have a similar overall modular structure composed of an N-terminal transmembrane sensor domain (TMD), a GAF domain in the middle portion, and a catalytic transmitter domain at the C-terminus. In receptors ETR1, ETR2, and EIN4, this basic structure is complemented by a C-terminal receiver domain. The presence of a copper cofactor in the TMD, likely in the +1 oxidation state¹⁰, was shown to be essential for ethylene binding and receptor function^{11,12}. Yet, the copper stoichiometry in the functional dimer is still a matter of discussion. Blecker, *et al.*¹³ consider one Cu⁺/monomer, while Rodriguez, *et al.*¹⁰ favor one Cu⁺/dimer. At present, the exact output of the receptors and mechanism of intramolecular and intermolecular signal transfer to further downstream elements in the ethylene signaling network is still obscure.

High-resolution structure information on the receptor is expected to solve this puzzle. Until today, crystal structures and low resolution SAXS models of the cytoplasmic part of ethylene receptors ERS1 and ETR1 excepting the GAF domain have been obtained^{14,15}. This has allowed to obtain a model of the whole cytosolic domain, including the GAF, dimerization, catalytic, and receiver domains (residues 158 to 738, see Fig. 1 in ref.¹⁴); yet, this

¹Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

²Centro de Bioinformática y Simulación Molecular (CBSM), Facultad de Ingeniería, Universidad de Talca, Talca, Chile.

³Institute of Biochemical Plant Physiology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ⁴Present address: Institute of Hydraulic Engineering and Water Resources Management, RWTH Aachen University, Aachen, Germany.

⁵John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) & Institute for Complex Systems - Structural Biochemistry (ICS 6), Forschungszentrum Jülich GmbH, Jülich, Germany.

⁶Bioeconomy Science Center, Forschungszentrum Jülich GmbH, Jülich, Germany. Stephan Schott-Verdugo and Lena Müller contributed equally. Correspondence and requests for materials should be addressed to H.G. (email: gohlke@uni-duesseldorf.de) or G.G. (email: georg.groth@hhu.de)

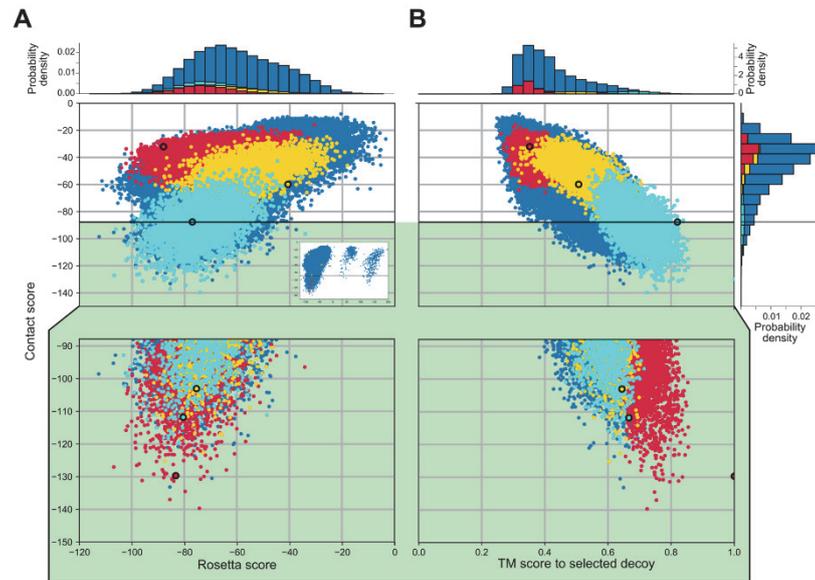


Figure 1. Distributions of Rosetta scores (membrane centroid score, **A**) and TM-scores with respect to the selected decoy (**B**) versus the calculated Contact score of the generated models. There is a positive correlation between the Rosetta score and the Contact score ($R^2 = 0.25$, $p < 0.001$) for models with a negative Rosetta score ($R^2 = 0.01$ for the complete distribution). The horizontal line demarks the -2 z-score threshold used with respect to the Contact score to filter out the worst scoring models; selected structures are shown on green background as a zoom. The models were structurally clustered in three groups pre- and post-score filtering, with the structures corresponding to the first, second, and third cluster shown in red, yellow, and cyan, respectively. The centroid models of the identified clusters (Fig. 2) are highlighted with a thick outline. The inset in panel A shows the complete distribution, including outliers.

model lacks the connection to the transmembrane domain, and the transmembrane domain itself. In addition, mutational studies have contributed to a structure-function understanding. In that respect, most structural information comes from the studies of Wang, *et al.*¹⁶. Three main classes of *loss-of-function* mutants were identified, showing different levels of ethylene binding, signal transmission, or intrinsic activity. However, the structure of the TMD bearing the ethylene binding site has remained elusive. The current development of *ab initio* protein folding tools allows one to predict the structure. Such tools have been used for modelling, e.g., the N-terminal portion of the human dopamine transporter¹⁷ and the human alkylglycerol mono-oxygenase¹⁸. A notable application is the structure prediction of the human zinc transporter hZIP4¹⁹, for which mainly coevolutionary information was used. The thus obtained structure was later shown to be highly similar to a crystalized homolog²⁰.

Following the same principles as for hZIP4, we generated the first structural model of the transmembrane sensor domain of ethylene receptor ETR1 (ETR1_TMD) and a ETR1_TMD/Cu dimer model by integrating *ab initio* structure prediction and coevolutionary information to drive the model selection process according to the available experimental information. In addition, to refine and independently validate the obtained structural model, we determined protein-related copper stoichiometries experimentally on purified receptor preparations and explored the helix arrangement in the TMD using tryptophan scanning mutagenesis. The models obtained by this pipeline allow deriving experimentally testable hypotheses as to how ethylene accesses the copper cofactor, how copper is transferred to the TMD, and how ethylene binding leads to downstream signaling.

Results

Structural model of the transmembrane sensor domain ETR1_TMD. Currently, no homologs of ETR1 with an experimentally resolved structure are available. A search for structural templates in the Protein Data Bank resulted in structures with sequence identities below 15%, which would likely result in an imprecise structural model of ETR1 by homology modeling (see Supplementary Results). Hence, the structure of the TMD of ETR1 was modeled using the *ab initio* Rosetta protocol *membrane_abinitio2*, and further validated by filtering resulting structural models with co-evolutionary methods. A flowchart of the whole process can be found in the Supplementary Information (SI Fig. 6). The secondary structure and transmembrane topology predictions from the methods PSIPRED and CCTOP, respectively, were used in the model building process, and they agree on three TM α -helices for the ETR1_TMD, as previously predicted¹⁶ and shown for the close homolog ERS1²¹

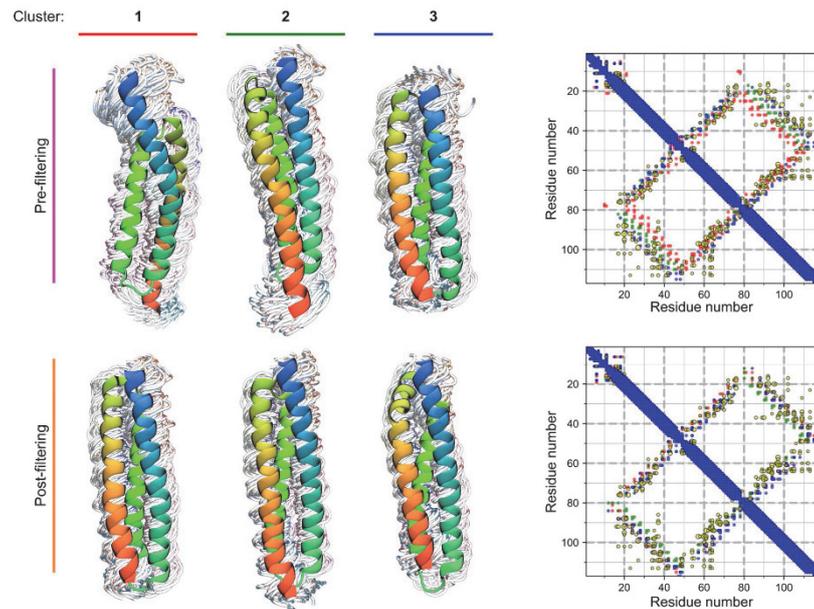


Figure 2. Clustered conformations and average contact maps for the generated decoys pre- (top) and post- (bottom) filtering according to the Contact score. The centroid of each cluster is shown in a cartoon representation, overlaid over every tenth other structure in the cluster, shown as wires. The structures are colored blue to red, starting from the N-terminal portion. The contact map for each set of clustered conformations are overlaid on the right, following the color scheme shown on top of every cluster to the left. Additionally, the MetaPSICOV contact predictions are shown as yellow dots. The size of the dots reflects the confidence assigned by the method. Before filtering, the clusters show different orientations with respect to helix three, as visible in the different conformations and on the contact map (residues 20–40 contacting 80–100). Contacts of the right-handed configuration of cluster 1 deviate the most from the MetaPSICOV contact predictions, and such configurations are removed by the filtering.

(SI Fig. 2). The distribution of the 100,000 models generated shows that > 99% of them have negative, i.e., favorable, Rosetta scores (Fig. 1). The models were clustered with *Calibur*²², with an estimated C_{α} atom RMSD of residues 15 to 117 ranging from 1.1 to 22 Å. All clustered models are located in the membrane, as evaluated by the orientation of each model with respect to the “MEM” coordinates printed by the score_jd2 protocol; the evaluation was done visually in that it was assured that all models are oriented along the membrane normal and are embedded in the membrane slab. The representative structures of the three clusters obtained (Fig. 2, top row) reveal two helix arrangements with different handednesses of the helix bundle. To discriminate between these two potential solutions, coevolutionary residue-residue contact predictions from MetaPSICOV²³ were used to rescore the models. The third largest cluster fulfills the predicted contacts best, as seen in the overlap of the predictions to the contact map (Fig. 2, top row, blue over yellow dots in the contact map) and given by a much more favorable average Contact score of -88 , compared to -36 and -49 for the most and the second most populated clusters, respectively. Filtering the generated models for structures that fulfill the contact predictions the best removes all configurations with a right-handed helical bundle from the pool of 100,000 models, leaving 5,217 structures that only differ in the relative orientation of the helices and slightly in the positioning of the third helix with respect to the second (Fig. 2, compare top and bottom conformations and contact maps).

After clustering of the filtered models, each cluster has a quadratic mean of the pairwise RMSD, $\langle \text{RMSD}^2 \rangle^{1/2}$, of 3.1 to 3.2 Å²⁴ and a distance in C_{α} atom RMSD of 2.8, 2.8, and 4.0 Å between centroids of clusters 1 and 2, 2 and 3, and 1 and 3, respectively. Both measures indicate the precision of our models²⁵. The centroid of the most populated cluster (representing 40% of the filtered structures) was selected for further analysis (Fig. 2, bottom left and SI Fig. 4A). Most of the other models that have a favorable Contact score also have a TM score to that selected model of 0.5 or higher; the TM score is a measure analogous to RMSD, but less sensitive to local variations, and is bound between 0 and 1, where 1 relates to an identical 3D structure, indicating that these models have the same fold²⁶ (Fig. 1). The selected model (SI Fig. 4A) was further refined with the Rosetta protocol *relax* to incorporate side chains²⁷.

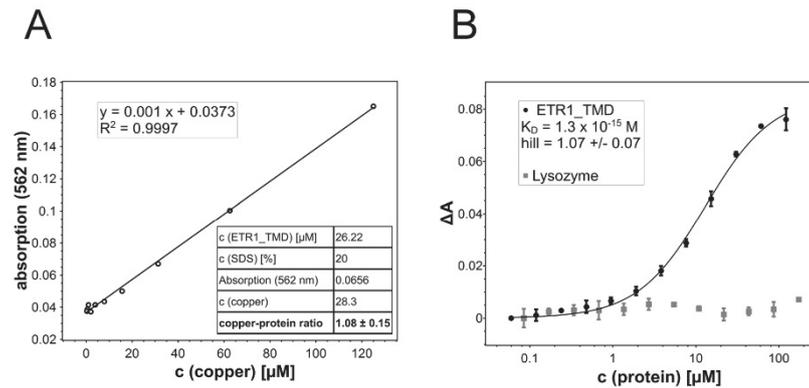


Figure 3. Copper-binding stoichiometry (A) and binding of copper(I) by the ETR1 transmembrane sensor domain (B). (A) Calibration curve of different $\text{BCA}_2\text{-Cu(I)}$ concentrations used to determine the copper concentration released from the protein ($28.3 \mu\text{M}$). Stoichiometry of copper-loaded ETR1 was determined by denaturing purified ETR1_TMD ($26.22 \mu\text{M}$) previously saturated with copper(I) by adding the detergent SDS at 20% (w/v) and heating the sample to 95°C . The chromophoric copper chelator BCA (2 mM) was added, and absorption at 562 nm monitored. The table shown in the inset summarizes protein and copper concentrations of the experiment corresponding to a copper:protein molar ratio of $1.08:1$. (B) Purified ETR1_TMD was titrated to $\text{BCA}_2\text{-Cu(I)}$ complex at concentrations from $122\text{--}0.06 \mu\text{M}$. Binding of the metal ion was monitored spectrophotometrically by measuring absorbance of the purple $\text{BCA}_2\text{-Cu(I)}$ complex at 562 nm (black points). From the binding curve a dissociation constant $K_D = 1.3 \times 10^{-15} \text{ M}$ and a Hill coefficient $h = 1.07 \pm 0.07$ were calculated for copper binding to ETR1_TMD. The non-copper binding protein lysozyme (grey squares) was used as negative control. All measurements were run in triplicates.

Stoichiometry and affinity of copper(I) binding to ETR1_TMD. To generate an ETR1_TMD/Cu dimer model (see below), the stoichiometry of ETR1_TMD with respect to the copper cofactor needs to be established. Previous studies have shown that the receptors sensing the plant hormone ethylene require monovalent copper ions in their transmembrane sensor domain in order to bind their ligand. Still, the number of copper ions per dimer forming the basic functional unit of the receptors is not clear. To clarify the stoichiometry of metal binding, purified ETR1_TMD was saturated with monovalent copper. To this end the purified sensor domain was incubated with saturating concentrations of the chromophoric copper chelator BCA, a water-soluble ligand that stabilizes the metal ion in the monovalent oxidation state. Transfer of the copper ion from the intensely purple-colored $\text{BCA}_2\text{-Cu(I)}$ chelate complex to the purified ETR1 sensor domain was monitored directly by absorption spectroscopy. Complete saturation of the copper binding sites in the receptor was achieved when no further change in the deeply purple-colored solution was detectable. Then, excess $\text{BCA}_2\text{-Cu(I)}$ chelate was removed by gel filtration, and the copper bound to the ETR1_TMD was released from the protein by chemical and thermal denaturation using the harsh detergent and amphipathic surfactant sodium dodecyl sulfate (SDS). The released metal ions were recomplexed by adding BCA, and the related copper concentration of the solution was determined by comparing the observed absorption at 562 nm to standard concentrations of $\text{BCA}_2\text{-Cu(I)}$. Figure 3A shows the calibration curve used for quantification. From these studies a metal-to-protein stoichiometry of one copper ion per ETR1_TMD was obtained (see table inserted in Fig. 3A).

Since monovalent copper promotes highly reactive oxygen species, the free copper concentrations in biological systems is kept extremely low due to high affinity binding to specific proteins, in order to avoid toxic effects. To determine the *in vitro* copper-binding affinity of ETR1_TMD, the purified sensor domain was titrated with Cu(I). As for the experiments resolving the copper stoichiometry of ETR1, the metal ion was stabilized in the monovalent oxidation state by BCA. Reduction in absorbance at 562 nm observed at increasing protein concentration was used to follow concentration-dependent copper binding to the receptor and to determine copper affinity of the purified TMD. Figure 3B shows the corresponding absorption changes as a function of the protein concentration added. The dissociation constant $K_D^{1/2} = 1.3 \times 10^{-15} \text{ M}$ (eq. 4) obtained from the metal-titration curve confirms the capacity of the purified ETR1 sensor domain to bind copper at very low concentrations, as previously reported for copper chaperones in living cells²⁶. Under the same conditions, no changes in absorption and, thus, no copper binding was observed with the non-copper(I) binding protein lysozyme upon titration with the monovalent cation (also see Fig. 3B).

To confirm our stoichiometry analysis, titration data was fitted to the Hill binding model²⁹. In this model, the stoichiometric coefficient h provides a measure for the cooperativity between binding sites and, in the extreme case of strict cooperativity, i.e., if all ligands bind at once (all-or-nothing binding), reflects the number of ligand binding sites on a protein. In this context, the Hill coefficient $h = 1.07 \pm 0.07$ obtained from a fit of the titration data shown in Fig. 3B implies simple, non-cooperative copper binding at the ETR1 transmembrane

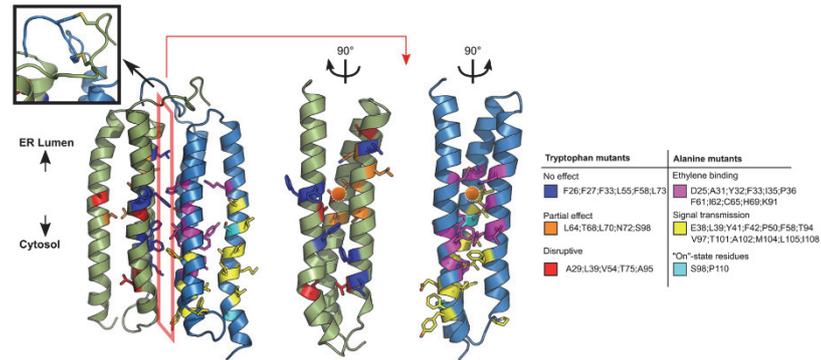


Figure 4. Dimer model of the ETR1 TMD. On the structure, the positions where tryptophan mutants were generated (left monomer) and where previous loss-of-function mutations were found¹⁶ (right monomer) are denoted by the color code on the right, with a list of the residues found in each class. The upper left inset shows the disulfide bridges included during the refinement of the protein. The orange spheres show the putative copper binding site in proximity to residues C65 and H69. The orange polygon indicates the interface. The two structures on the right represent the interface in an “open book” representation. Tryptophan mutants of residues that are pointing towards the monomer bundle core are disruptive (red) or are partially disruptive (orange), while mutations of residues pointing towards the dimer interface of the model (blue) showed no effect on the alpha-helical content, as shown in Fig. 5. Residues that are closer to the protein center and in proximity to the putative copper binding site are essential for ethylene binding (magenta), while residues farther away and closer to the cytosolic portion are responsible for the signal transmission (yellow). Residues shown to be relevant to maintain the protein in an active state are displayed in cyan.

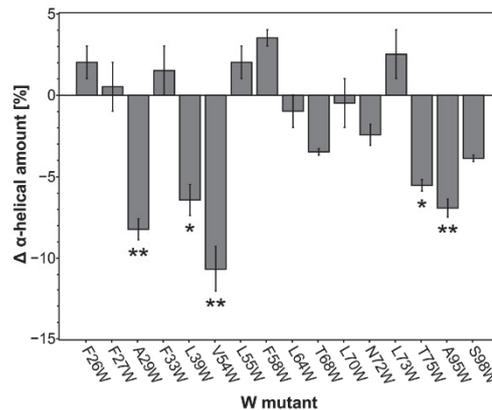


Figure 5. Changes in overall α -helical content in ETR1 by tryptophan substitution. Secondary structure content of purified full-length ETR1 tryptophan mutants was determined by CD spectroscopy. Changes in α -helical content related to individual tryptophan mutants are shown as deviation from the tryptophan-free background mutant ETR1^{W7X}. All measurements were taken in triplicates. Mean and standard deviation are shown (* $P \leq 0.05$; ** $P \leq 0.02$).

sensor domain (lower bound of potential sites). In the case of strict binding cooperativity, this figure reflects a metal-to-protein stoichiometry of one copper ion per monomer (upper bound of potential sites) as obtained by direct analysis of the purified ETR1 sensor domain (Fig. 3B).

Structural model of the ETR1 TMD/Cu dimer. From the generated ETR1 TMD model, and the determined copper stoichiometry, a dimeric model of ETR1 TMD/Cu can now be generated. For generating this model, the TMD model was docked using HADDOCK. Coevolutionary signals (Fig. 2 and SI Fig. 3), knowledge

about low-lipophilicity regions (SI Fig. 2B), and the notion that the metal binding site should be shielded from the solvent³⁰ were used as information to select the interacting interface; all this data suggests that the interface involves H2, while H3 is more membrane-exposed, indicating H1/H2 as the proper interface. The representative dimer model obtained from clustering after refinement by replica exchange MD simulations is shown in Fig. 4. The structure consists of an almost symmetrical arrangement of the previously modelled left-handed monomers. The N-terminal part displays the disulfide bonds in-between chains, formed by residues 4 and 6, respectively. The putative copper binding sites, composed of residues C65 and H69, are buried in the dimerization interface, as previously mentioned and suggested¹⁶; in the monomer, they would be exposed to the membrane environment. Based on the experimentally determined metal-to-protein stoichiometry of ETR1 (Fig. 3), two copper ions were added in subsequent steps in proximity to S_γ of C65 and N_δ of H69, considering that these residues are relevant for copper binding (Rodríguez, *et al.*¹⁹). The C-termini of the dimer are at opposite sides of the dimeric configuration, which would leave enough room for residues connecting the transmembrane domain and the cytoplasmic GAF domain. The obtained dimer model is supported by loss-of-function mutations according to which mainly H1 and H2 are involved in the binding of ethylene¹⁶ (for further details see Discussion section). These mutations form a “layered” spatial arrangement of residues, with relevant ethylene-binding residues close to the putative copper binding site and residues relevant to signal transmission close to the third helix/cytoplasmic region. Interestingly, the model shows residues S98 and P110, both of which seem to be essential for the correct transmission of the ethylene binding signal¹⁶, with the former pointing towards the copper binding site, while the latter generates a kink in the C-terminal portion of helix 3.

Based on the structural model, positions 26, 27, 29, 33, 39, 54, 55, 58, 64, 68, 70, 72, 73, 75, 95, and 98 were considered non-solvent/membrane exposed and good candidates for tryptophan scanning experiments; their mutation to tryptophan is expected to distort the structure as the relative solvent accessibility is reduced in the model ensemble (see also below).

Experimental validation of the structural model by tryptophan scanning mutagenesis. To validate the ETR1_TMD/Cu dimer model, tryptophan variants were generated. In the past tryptophan substitution mutagenesis has been applied as a useful tool to identify the relative arrangement and orientation of transmembrane helices in membrane proteins^{31,32}. In this approach the large and moderately hydrophobic tryptophan side-chain is introduced at different positions of a transmembrane helix. Substitutions are tolerated at positions facing the membrane only, whereas introduction of tryptophan residues at helix-helix interfaces show disruptive effects on protein function or structure. Making use of this approach, ETR1 tryptophan mutants were generated based on predictions from the generated structural model of the ETR1_TMD/Cu dimer, and CD spectra of the variants were measured. From there, the α -helical amount of each mutant was computed. The tryptophan-free version of the full length ETR1 (ETR1^{WT}) shows an α -helical amount of 33%. Some of the other mutants show similar values, e.g., F27W (34%), F33W (35%), L64W (32%), and L70W (33%) (see Fig. 5, Supplemental Information Table 2, and Fig. 4). However, exchange of amino acids at positions 29, 39, 54, 75, and 95 to tryptophan show a significant decrease in α -helical amount to 22–27%. A lower amount in this value compared to the ETR1^{WT} was used as an indicator for changes in protein structure due to specific tryptophan insertion.

Molecular dynamics simulations of free ethylene diffusion show preferential occurrence of ethylene in the proximity of copper in the dimer model. In order to elucidate how ethylene can bind to the ETR1_TMD/Cu dimer model, molecular dynamics (MD) simulations of 1 μ s length with an explicit membrane environment were performed (SI Fig. 4B). Ten independent replicates each were simulated, either containing 0.1 M ethylene in the water phase at the beginning of the simulation or not. Experimentally identified disulfide bridges between the protomers were included in the N-terminal region³³. The trajectories show a moderate C _{α} atom RMSD drift with respect to the average structure (distribution maximum between 2 to 4 Å in all but one case (SI Fig. 5)), with all of the simulations reaching an apparent plateau after 500 ns. The copper ion was included with a 12-6-4 Lennard Jones non-covalent model, which has been shown to yield good coordination geometries with surrounding amino acids without additional restraints³⁴. The ions remained bound to the sulfur atom of cysteinate C65 of each chain, interacted also with the non-protonated N_δ of H69, and showed infrequent interactions with D25 (SI Fig. 5). The overall coordination number between the protein and the copper ions fluctuates between 2 and 3, while there are 2 to 1 sites occupied by water molecules, yielding a total of 4 coordination sites. While the number of coordination sites may be biased by the used non-covalent model³⁴, copper is coordinated in a similar manner in copper chaperones³⁵.

According to the preparation of the simulation system, ethylene starts in the water phase but rapidly enters the membrane, and from there the ETR1_TMD (see also below), yielding a distribution with the maximum located in the central plane of the membrane (SI Fig. 5). A 3D histogram shows that ethylene has a higher propensity to bind within the TMD than the membrane, and it does so in particular regions of the TMD (Fig. 6A); these results are confirmed by a cluster analysis (Fig. 6B). Ethylene binding and unbinding occurs from the center of the membrane to the accessible binding site(s) in the protein; an example path is shown in SI Fig. 6B and C. That ethylene binding to the protein is almost in equilibrium is suggested by rapid and repeated binding/unbinding events (SI Fig. 6A) and the fact that ethylene binds similarly to both monomers (Fig. 6B), although no symmetry restraints were imposed during the MD simulations.

Three of the identified ethylene binding sites are at a distance < 5 Å to the copper ions, with one being in between the two ions, while the other two are located in the center of each monomer bundle (Fig. 6). These putative ethylene binding regions are close to the central plane of the membrane, as are the locations of the copper ions (SI Fig. 5). Residues lining ethylene binding sites close to the copper ions (Fig. 6B) are conserved (SI Fig. 2). Of these residues, seven have been shown to be essential for ethylene binding to the protein (red labels; (Fig. 6B)), and S98 has been shown to affect the binding moderately (magenta label; (Fig. 6B))¹⁶. In turn, of the

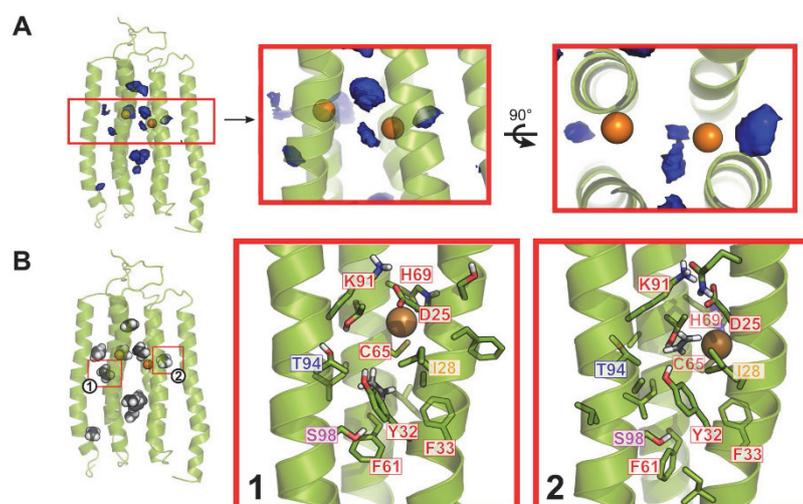


Figure 6. Putative ethylene binding sites identified during MD simulations. The average structure of the ETR1_TMD dimer over all replicate simulations is shown. (A) 3D propensity representation of the most probable locations of ethylene binding along all replicate simulations. The boxes highlight the portion close to the included copper ions, with a zoom on the blobs at a distance $<5 \text{ \AA}$. (B) Representative configurations of identified ethylene clusters. The clusters enclosed in the red boxes are at a distance $<5 \text{ \AA}$ to the copper ions, and are shown in detail on the right. The residues labeled in red have been shown to be essential for ethylene binding, S98 (magenta) has a moderate effect, while T94 (blue) has no effect on binding, but is relevant for signal transmission. I28 has not been tested. Mutation information from Wang, *et al.*¹⁶.

residues found to be essential for ethylene binding aside from C65 and H69, which also have been linked with copper binding, only I35, P36, and I62 were not identified, but are in immediate proximity to the centroids of the described clusters. In some cases, distances between a copper ion and the center of mass of ethylene $< 3 \text{ \AA}$ were observed, which is in the range of the sum of van der Waals radii for carbon and copper and only $\sim 1 \text{ \AA}$ larger than distances observed in organometallic complexes³⁶. While this clearly demonstrates a close approach of ethylene to copper, smaller distances cannot be expected, because the force field approach used here to model ethylene-copper interactions does not incorporate electron exchange between the binding partners, as expected from the Dewar-Chart-Duncanson model of ethylene binding³⁷.

Discussion

In this study, we describe the first structural model of the transmembrane sensor domain of the plant ethylene receptor ETR1 at the molecular level. The structure reflects the functional dimeric state of the receptor with the bound copper cofactor, which was previously shown to be essential for ethylene binding and perception. Experimental analysis of purified ETR1 revealed a metal-to-protein stoichiometry of 1:1 and a dissociation constant ($K_D^{1/2}$, eq. 4) of $1.3 \times 10^{-15} \text{ M}$ for the monovalent copper cofactor.

Evidence for the 1:1 stoichiometry was obtained in two ways, using in both cases BCA as reporter and copper(I)-chelating system. First, the ETR1 sensor domain was saturated with the metal ion. Monovalent copper bound to the receptor then was released by denaturation of the protein, recomplexed with BCA, and determined spectrophotometrically. The measured copper concentration was then related to the concentration determined for the purified ETR1_TMD to obtain the molecular metal-to-protein ratio. In a second approach, the stoichiometry was deduced from the Hill coefficient of the binding isotherm obtained by titration of the soluble $\text{BCA}_2\text{-Cu(I)}$ chelate to increasing concentrations of the purified ETR1 sensor domain, assuming that the system shows strict cooperativity (all-or-nothing binding). Such behavior has been reported previously for various protein-ligand systems such as receptor interactions in tyrosine kinase ErbB-3/ErbB-2³⁸, proton trapping in chloroplast ATPase³⁹, or proton release in violaxanthin deepoxidase⁴⁰. At first glance, the experimentally derived stoichiometry of one copper per monomer found in this work is contrary to earlier studies, which, based on the binding of radio-labeled ethylene in the presence of copper, proposed a stoichiometry of one copper per receptor dimer¹⁰. However, later work on this topic⁴¹ conceded that potentially not all receptors were active and capable of binding ethylene in this earlier study. Based on this boundary, the authors proposed an alternative model where each ETR1 receptor contains more than one copper ion per active receptor dimer - each one capable of binding ethylene⁴¹. The experimentally derived stoichiometry and molecular model of our study fit well to this idea.

The femtomolar affinity of copper towards the purified ETR1 sensor domain conforms well to the high metal affinities of 1.6 to $4.5 \times 10^{-11} \text{ M}$ reported for the metal binding domains of Menkes copper-transporting P_{1B} -type

ATPases - a homolog of plant RAN1⁴² - or the soluble chaperone Atox1²⁸, which may be involved in copper transfer to the receptor at cellular conditions. Moreover, the higher copper affinity of the ETR1 sensor domain and the thermodynamic gradient to the above-mentioned cellular copper delivery systems may assure copper routing from the chaperons to the receptor target and thus an efficient net copper flux among these sites. Note, however, that these values are apparent affinities of purified proteins, which may change for membrane-integrated receptor dimers or receptor clusters at physiological conditions.

Generating a structural model of the transmembrane domain of ethylene receptors has been hampered so far by the lack of close homologs with known structure (see Supplementary Information). As an alternative, the Rosetta protocol for membrane protein modeling has been widely used for *ab initio* structure prediction^{17,18}. The accuracy of protein structural models obtained by *ab initio* techniques is markedly improved by including co-evolutionary information into the model process¹⁵. A notable case was the structural model obtained for the human zinc transporter hZIP4¹⁹, which was later shown to be highly similar to a crystalized homolog²⁰. Using the same principles as reported there, we resorted to the Rosetta membrane *ab initio* protocols and coevolutionary contacts to generate a structural model of the ETR1_TMD. The resulting large ensemble of structural models was screened with respect to coevolutionary residue-residue contact predictions. About 5% of the initial ensemble fulfilled the Contact score to values below $-2 Z$ -score, showing an enrichment of models that favor the predicted contacts compared to the reduction shown for the hZIP4 modeling case¹⁹. The validity of the finally chosen model is further supported by the fact that measures of structural variability within a cluster of filtered models ($< \text{RMSD}^2 >^{1,2}$) as well as between clusters of suitable models (C_{α} atom RMSD between centroid structures) are small and very similar (~ 3 – 4 Å). This indicates, first, a precision of the model comparable to structural deviations observed during molecular dynamics simulations started from crystal structures and, second, that no grossly different structural models have been generated with a likewise favorable Contact score.

For generating a dimer model from the left-handed helical bundle of three TM helices, we again used co-evolutionary information as restraints for protein-protein docking in addition to experimental knowledge on the shieldedness of the copper binding site³⁰, information on the roles of C65 and H69 as copper-interacting and putatively copper-interacting residues, respectively^{10,44,45}, and our own data on the copper:ETR1 monomer stoichiometry. The finding that there is one copper(I) per monomer suggests that there is one ethylene binding site per monomer as well, as the Dewar-Chatt-Duncanson interaction model involves one olefin per copper ion^{11,12,36}. This also implies that the interplay between the monomers is probably necessary to transmit the signal to the C-terminal part of the receptor, rather than to provide a joint binding interface for ethylene.

The dimer model was validated by mutational analysis, both based on *loss-of-function* mutations previously described¹⁶ and tryptophan scanning mutagenesis performed here at sites identified from the dimer model. As to the former, residues involved in *loss-of-function* mutations form a layered arrangement around the copper binding site, which, first, indicates that in the structural dimer model, functionally relevant residues are in close proximity and, second, provides a suggestion on the mechanism how the receptor transmits the signal of ethylene binding to the cytosolic domain. The inner-center layer involves *loss-of-function* mutations on helices 1 and 2 (H1 and H2) relevant for (assisting in) binding of copper(I) and ethylene, while *loss-of-function* mutations on H1 and helix 3 (H3) closer to the cytosol are mainly involved in the transmission of the inhibitory signal to the rest of the protein. This suggests that a possible conformational change triggered by the binding of ethylene to the receptor is transmitted through H3 as a signal to the rest of the protein, thereby inhibiting it, while H1 and H2 provide support for the core of the binding site and might constitute the structural foundation against which the conformational change of H3 occurs. As to tryptophan scanning mutagenesis, tryptophan mutations that point to the hydrophilic core of the monomer according to our dimer model affected the helical content the most, as expected^{31,32}. In contrast, mutations with a lower effect are either pointing to the membrane environment, again as expected, or the dimer interface. The latter seems surprising, but the incorporation of bulky residues at the interface might only affect the dimerization efficiency¹⁹ and not the overall α -helical content used here as a read-out for structural distortions.

We finally used the dimer model in all-atom explicit solvent/membrane MD simulations of free ethylene diffusion to probe if and how ethylene can access the copper cofactor. The MD simulations are converged with respect to ethylene binding from the central membrane slab to the ETR1_TMD dimer, as demonstrated by repeated and frequent binding and unbinding events and an almost symmetrical ethylene propensity within the protein. The MD simulations revealed one dominant binding site per monomer in the proximity of the copper ion. Notably, the location of this binding site is also in line with previously reported *loss-of-function* mutations¹⁶. The MD simulations furthermore reveal a prominent binding region between the monomers. However, accumulation of ethylene there may be due to a high concentration of ethylene used in the MD simulations and/or a lack of sufficiently high attraction between copper and ethylene. The latter results from missing charge transfer effects in the classical force field representation of intermolecular interactions used here^{37,46}. In turn, this fact implicates that ETR1 itself provides sufficient affinity for ethylene, such that ethylene is attracted from the membrane already before the interaction with copper occurs. In line with this implication, the MD simulations also reveal that predominant paths of ethylene access and egress run parallel to the membrane and roughly at the center of the membrane slab (SI Figs 5C and 6).

Due to the only approximate representation of copper-ethylene interactions and the fact that no small-molecule model compound has been described yet in which cysteinate and histidine mimics, resembling the sidechains of C65 and H69, chelate Cu(I), we are currently unable to provide detailed insights into the exact binding geometry of ethylene at the copper ion in ETR1, the related electronic structure, and/or further influences of neighboring residues³⁶. Providing such insights may also help answering the question how ETR1 manages to avoid unspecific reactions with dioxygen species³⁷.

Still, the derived information can be used to speculate on the mechanism underlying the inhibition of ETR1 by ethylene. Residues S98 and P110 in H3 have been shown to be relevant to maintain the activity of the receptor,

as mutations to alanine resulted in a permanently semi-inactivated state¹⁶. Our model shows that S98 is close to the putative ethylene binding site, while P110 generates a kink in the C-terminal region of H3. We now speculate that S98 is relevant to maintain H3 in a “receptor-on” conformation, which is perturbed due to the binding of ethylene, causing the inhibition of the receptor. The kink caused by P110, on the other hand, might be structurally required to correctly connect the transmembrane domain of ETR1 to the C-terminal part of the protein. The strain imposed by the lack of a helix breaking residue at this position might be enough to inhibit the whole receptor. Finally, binding of ethylene to copper might displace one of the coordinating residues and/or lead to a change in the coordination geometry of these residues. This change would be transmitted to the cytosolic domain. It has been previously shown that two component systems rely on amplifying the signal to transmit the message through the protein⁴⁷. The binding of one ethylene molecule to each monomer might thus cause an even stronger signal, making a small coordination effect significant for the whole structure.

In summary, we have identified that ETR1 binds one copper(I) ion per monomer, and based on this information and coevolutionary signals, we built a dimeric *ab initio* model of the transmembrane sensor domain of ETR1 including copper(I). The structural model is supported by alanine and tryptophan scanning mutagenesis studies and reveals how ethylene accesses the center of each monomer bundle and can bind proximal to the copper(I) ions. From the model, we propose that ethylene binding perturbs interactions of H3 with the remaining receptor part, which leads to a signal towards the cytoplasmic domains that switches off the receptor.

Materials and Methods

Ab initio modeling with Rosetta. For modeling the transmembrane sensor domain of ETR1 of *Arabidopsis thaliana*, residues 1–117 (Uniprot code P49333) were used. The selection was based on the transmembrane topology prediction obtained by the meta approach CCTOP⁴⁸ and on the secondary structure prediction from PSIPRED v4⁴⁹ (both shown in SI Fig. 2A). The former was chosen as it averages via a weighted Hidden Markov Model results of ten state-of-the-art transmembrane topology predictors and has shown a significantly higher accuracy⁴⁸. A search of protein templates was performed, resulting in a maximum sequence identity of 11.3% (Supplementary Information). As a consequence, the protein was modelled *ab initio*. For this, the exposure of the transmembrane portion of each predicted helix towards the membrane environment was evaluated with the LIPS algorithm⁵⁰. Based on the secondary structure prediction, 3- and 9-mer fragments were generated by using the *make_fragments.pl* script included in Rosetta 3.6. These initial predictions were included into the modelling process, which was performed using the RosettaMembrane *membrane_abinitio2* protocol (Supplementary Methods), also included in Rosetta^{27,51}. The modelling followed the developer recommendations⁵², generating 100,000 centroid models (structures without explicit side-chains). To evaluate the orientation of the models in the membrane with respect to the “MEM” coordinates printed by the score_jd2 protocol, it was assured visually that all models are oriented along the membrane normal and are embedded in the membrane slab. Contact predictions were calculated using MetaPSICOV with the default protocol²³, generating an alignment of 447 homologs with an alignment depth of 3.8, which is in line with previously shown modelling scenarios^{43,53}. As a result, 397 predicted contacts with a confidence > 0.1 were obtained. The contact predictions were transformed into Rosetta constraint format between the C_β atoms (C_α if glycine) to score the obtained models, using the following sigmoidal function:

$$\text{Contact score } f(c, r) = c \left(\frac{1}{1 + e^{-3(r-8)}} - 1 \right) \quad (1)$$

where c corresponds to the prediction confidence used as a weighting factor and r is the corresponding distance between C_β atoms in Ångströms. This is the suggested way to score models in the GREMLIN contact web-server⁵⁴. The sigmoid was centered at a distance of 8 Å, following the contact prediction convention. The models were scored during the folding process using a weight of 4, and were filtered using a z -score ≤ -2 with respect to the contact score, similar to what has been used before¹⁹, reducing the number of structures to 5217. To identify a representative structure of the ensemble, a clustering step was performed by using *Calibur*²². Contact maps for the decoys corresponding to each cluster were calculated and averaged by using a 8 Å threshold. The centroid of the largest identified cluster was used for further analysis and *relaxed* using the corresponding Rosetta protocol (Supplementary Methods) with membrane scoring and the RosettaMembrane *membrane_highres* scoring weights, generating 100 decoys with different side-chain configurations. The per-residue solvent accessibility of the models in the selected cluster was evaluated using FreeSASA⁵⁵, and residues with low average accessibility were considered for tryptophan mutation analysis. To have a measure of the sampled model space and convergence of the filtered models, the final selected model was compared with the rest by using the template modeling score (TM score), a measure of structural similarity analogous to RMSD, but less sensitive to local variations, which is bounded between 0 and 1; values > 0.5 have been shown to reflect similar folds⁵⁶.

Dimer model generation. The decoy with the highest ProQM⁵⁷ and high QMEANbrane⁵⁸ scores was selected from the relaxed configurations. Both scores are model quality assessment scores for membrane proteins, which allow evaluating protein structures without knowing the target structure, and are derived from top-ranked methods in the CASP12 competition⁵⁹. To generate a dimer structure, two copies of this decoy were docked together using HADDOCK⁶⁰. Residues of the TM1/TM2 interface were selected to fulfil restraints between monomers. The rationale for selecting this interface was twofold: I) it has been proposed that copper ions bind to residues C65 and H69 on TM2¹⁰ and that residues chelating metal ions are less solvent accessible³⁰; II) TM1 and TM2 have more helical surfaces with a low LIPS score and stronger coevolutionary signals between them considering homodimeric symmetry (SI Fig. 2). Both I) and II) have been shown to be predictive of interactions between monomers⁶¹ and, thus, were used to guide the docking with HADDOCK. Additional unambiguous restraints between residues E15 and K45 were included to promote a near-parallel and membrane-like orientation

of the monomers. The first 15 unstructured aminoacids were removed to avoid sampling restrictions during the rigid docking phase, the DMSO model was used for the “water” refinement stage, and a C_2 symmetry was enforced throughout the docking process. The generated models were scored with QMEANbr as before, and the N-terminal residues were modelled with MODELLER⁶², including restraints between residues 4 and 6 of monomers, respectively, to represent experimentally shown disulfide bonds³³.

The relative orientation between the transmembrane helices in the dimer was refined by using a replica exchange molecular dynamics protocol (REMD) in a GBSW⁶³ implicit membrane model. The sampling of the relative orientation of the helices was enforced by imposing restraints on dihedral angles in portions of the protein with predicted secondary structure. It has been shown previously that refinement of models through molecular dynamics (MD) simulations is well suited for local structure refinement but can drift the model away from the real structure if no restraints are used⁶⁴. The simulations were performed using charmm 41b2 with the included CHARMM22 GBSW parameters and run using the MMTSB Tool Set⁶⁵ *aarex.pi* script. For this, 16 exponentially spaced replicas between 300 and 460 K were used, ensuring an exchange probability between 12 and 15% throughout the simulations. An exchange was tried every picosecond of simulation time, and each replica was run for 5 ns, making a total of 80 ns. The temperature was maintained with a Langevin thermostat with a friction coefficient of 5 ps^{-1} . A membrane thickness of 30 Å, a switching length of 2.5 Å, and a surface tension coefficient of $0.005 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ were used. The replica running at 300 K was clustered by using the DBSCAN clustering algorithm in CPPTRAJ⁶⁶. The representative structure of the biggest cluster was selected for further MD simulations.

Molecular dynamics simulations of the ETR1_TMD/Cu dimer model in the absence and presence of ethylene. The refined model was embedded into a DOPC:DOPE 3:1 membrane, resembling the major components of the plant endoplasmic reticulum membrane⁶⁷, by using PACKMOL-Memgen⁶⁸. A Cu^+ ion was included per subunit (see below with respect to the stoichiometry) in between residues C65 and H69, and K^+ and Cl^- were added as counterions in the solvation box. Ions were treated with a 12-6-4 non-bonded model³⁴. The GPU particle mesh Ewald implementation from the AMBER18 molecular simulation suite⁶⁹ with the ff14SB⁷⁰ and Lipid17⁷¹ and Skjevik *et al.*⁷² parameters for the protein and the membrane lipids, respectively, were used; water molecules were added using the TIP3P model⁷³. Ten independent MD simulations of 1 μs length were performed. Covalent bonds to hydrogens were constrained with the SHAKE algorithm⁷⁴ in all simulations, allowing to use a time step of 2 fs. All analyses were performed using CPPTRAJ⁶⁶.

Initially, the total potential energy was minimized by three mixed steepest descent/conjugate gradient minimizations 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, finally, a minimization without restraints. The temperature was maintained by using Langevin dynamics⁷⁵, with a friction coefficient of 1 ps^{-1} . The pressure, when required, was maintained using a semi-isotropic Berendsen barostat⁷⁶, coupling the membrane (x-y) plane. The equilibration started from the minimized structure, which was heated by gradually increasing the temperature from 10 to 100 K for 5 ps under NVT conditions, and from 100 to 300 K for 115 ps under NPT conditions at 1 bar. This was continued for 5 ns under NPT conditions, after which production runs were started using the same setup. Ethylene was parametrized with the general Amber force field (GAFF)⁷⁷, and RESP charges⁷⁸ were obtained from electrostatic potentials computed at the HF/6-31 G* level of theory with Gaussian 09⁷⁹ and fitting with *antechamber*⁸⁰. Ethylene was included at a concentration of 0.1 M in the water volume, and its diffusion was observed in an unbiased manner^{81,82}.

Cloning of ETR1 tryptophan mutants and ETR1_TMD. DNA coding tryptophan-free ETR1 (ETR1^{WT}) from plasmid pET16b-ETR1(ΔW)⁸³ was cloned in vector pET15b. The resulting vector carrying ampicillin resistance and an N-terminal Hexahistidine-tag was used for cloning a set of ETR1 tryptophan substitution mutants. Mutants were either cloned by using mega primer built on mutagenesis primer or by round-the-horn site-directed mutagenesis⁸⁴. pET16b_ETR1_TMD was cloned from pET16b-ETR1⁸⁵ which differs from vector pET15b by an N-terminal Decahistidine-tag by round-the-horn site-directed mutagenesis. Oligonucleotides used by round-the-horn site-directed mutagenesis were phosphorylated at their 5' ends. A table with all oligonucleotides used can be found in the supplemental information.

Expression of ETR1 tryptophan mutants and ETR1_TMD in *E. coli* C43. Expression of ETR1 protein constructs was performed in 2YT media (1.6% (w/v) peptone, 1% (w/v) yeast extract, and 0.5% (w/v) NaCl) containing 100 $\mu\text{g/ml}$ ampicillin for selection. For agar plates 1.5% (w/v) agar was added. Vectors pET15b or pET16b carrying the DNA sequence for ETR1 tryptophan mutants or ETR1_TMD, respectively, were individually transformed into the expression host *E. coli* C43 and incubated on agar plates at 37 °C. The main culture was inoculated with a preculture to an $\text{OD}_{600} = 0.1$. For ETR1 tryptophan mutant expression, cells were grown at 30 °C and 180 rpm, and expression was induced by adding 0.5 mM IPTG at an $\text{OD}_{600} = 0.8$. After 5 hours, cells were harvested by centrifugation at 7,000 g and 4 °C for 15 min, flash frozen in liquid nitrogen and stored at -20 °C. For expression of ETR1_TMD, 2% (v/v) ethanol was added to the media, and cells were grown at 30 °C and 180 rpm until $\text{OD}_{600} = 0.4$ was reached. Temperature was reduced to 16 °C, and protein expression was induced with 0.5 mM IPTG at an $\text{OD}_{600} = 0.6$. Cells were incubated for 20 h and harvested as described for ETR1 tryptophan mutant expression.

Sequential fractionation and isolation of full-length ETR1 and ETR1_TMD. Cells were resuspended in buffer M (PBS, 10% (w/v) glycerol, 0.002% (w/v) PMSF, and 10 $\mu\text{g}/\mu\text{L}$ DNaseI) and broken in a Constants Cell Disruption System (Constant Systems, Daventry, United Kingdom) at 2.4 kbar and 4 °C. The lysate was centrifuged at 14,000 g and 4 °C for 30 min. The supernatant was centrifuged again at 40,000 g and 4 °C for

30 min. The resulting pellet was resuspended in buffer M and centrifuged at 34,000 g and 4 °C for 30 min. Finally, the pellet was flash frozen in liquid nitrogen and stored at −80 °C.

Denatured purification and renaturation of ETR1 tryptophan mutants. Membrane fractions of ETR1 tryptophan mutants were resuspended in buffer D (50 mM TRIS/HCl pH 8, 100 mM NaCl and 8 M urea) and stirred for 2 h at 37 °C before centrifugation at 100,000 g and RT for 30 min. The supernatant was loaded to a buffer D-equilibrated HisTrap FF column operated by an ÄKTAprius plus (both GE Healthcare Life Sciences) and purified by immobilized metal-ion affinity chromatography (IMAC) at 4 °C. The column was washed with buffer D containing 50 mM imidazole, and the protein was eluted with buffer D containing 500 mM imidazole. Protein fractions were pooled and concentrated to 1.5 ml, refilled to 15 ml with buffer D and again concentrated to a final concentration of 0.8 mg/ml. For renaturation 100 mM DTT was added to 0.5 ml protein solution (0.8 mg/mL), mixed with 10 ml buffer R (55 mM TRIS/HCl pH 8, 264 mM NaCl, 11 mM KCl, 0.1% (w/v) DDM, 1.1 mM EDTA, 10 mM DTT, and 0.002% (w/v) PMSF) and centrifuged at 229,600 g and 4 °C for 30 min. The protein solution was concentrated to 500 µl, buffer was changed to buffer C (50 mM potassium phosphate pH 7.5 and 0.05% (w/v) DDM), concentrated to a final concentration of 0.1–0.3 mg/mL for circular dichroism spectroscopy studies.

Solubilization and purification of ETR1_TMD. Membrane fraction of ETR1_TMD was resuspended in buffer S (50 mM TRIS/HCl pH 8, 200 mM NaCl, 1.2% (w/v) FosCholine-16, and 0.002% (w/v) PMSF), stirred for 1 h at RT, and centrifuged for 30 min at 229,600 rpm and 4 °C. Protein was purified by IMAC. To this end, the supernatant was loaded to a HisTrap FF column equilibrated in buffer A (50 mM TRIS/HCl pH 8, 200 mM NaCl, 0.015% (w/v) FosCholine-16, and 0.002% (w/v) PMSF). After washing with 20 column volumes with buffer ATP (50 mM TRIS/HCl pH 8, 200 mM NaCl, 50 mM KCl, 20 mM MgCl₂, 10 mM ATP, and 0.002% (w/v) PMSF), the column was washed with buffer A containing 50 mM imidazole. ETR1_TMD was eluted from the column with buffer A containing 250 mM imidazole and fractions containing the purified protein were concentrated to 2.5 ml. Buffer was changed to buffer A and protein was loaded to a Superdex 200 Increase 10/300 GL column previously equilibrated with buffer A (GE Healthcare Life Sciences) and further purified by size exclusion chromatography. Protein fractions containing the purified ETR1_TMD dimers were pooled, concentrated to a minimum final concentration of 200 µM and used for copper binding studies.

Circular dichroism spectroscopy of ETR1 tryptophan mutants. CD measurements of ETR1 tryptophan mutants were performed using a *Jasco-715* spectropolarimeter (Jasco GmbH, Gross-Umstadt, Germany) and a cylindrical quartz cuvette (Hellma GmbH & Co. KG, Muellheim) with a path length of 1 mm and a volume of 200 µl. All measurements were performed at room temperature in buffer C at a protein concentration of 0.1–0.3 mg/ml. Spectra were recorded from 260–195 nm with a step resolution of 1 nm and a bandwidth of 2 nm. The scan speed was set to 50 nm/min, and 10 scans were accumulated. Secondary structure content of purified proteins was calculated from the spectra by Selcon3 and CONTINLL.

Copper binding studies on purified ETR1_TMD. Protein-related copper stoichiometries were determined from ETR1_TMD samples saturated with BCA₂-Cu(I). First, these were loaded on PD10 mini columns (GE Healthcare Life Sciences) to remove excess BCA₂-Cu(I). Subsequently, protein concentration of the samples was determined by measuring absorbance at 280 nm in a TECAN plate reader Infinite M200 PRO using a Nano Quant plate (Tecan, Männedorf, Schweiz). For determination of protein-related copper stoichiometries, ETR1_TMD was denatured with SDS at a final concentration of 20% (w/v) and heated at 95 °C for 10 min resulting in the release of copper cofactor bound to the protein. To complex the copper released by the protein, the solution then was incubated for 10 min with 2 mM BCA. After incubation, absorbance at 562 nm was measured, and copper concentration in the sample was calculated from a standard curve obtained by measuring the absorbance of different concentrations of BCA₂-Cu(I) at 562 nm⁸⁶. All samples were run in triplicates.

For copper binding studies, ETR1_TMD was serially diluted from 122 µM to 60 nM in a transparent 96 well plate (Sarstedt/Nümbrecht) at a volume of 25 µl for each concentration. Then, an equivalent volume of 25 µl of a 1:10 dilution of BCA₂-Cu(I) reagent (50 mM TRIS/HCl pH 7.5, 200 mM NaCl, 2.5 mM BCA, 1 mM CuCl, and 20 mM ascorbate) was mixed with each of the ETR1_TMD samples. Absorption of the purple BCA₂-Cu(I) complex at 562 nm was measured in a TECAN plate reader Infinite M200 PRO (Tecan, Männedorf, Schweiz). Lysozyme - a non-copper(I) binding protein - was used as negative control in these measurements⁸⁷. Based on previous studies recognizing homodimers as the minimum functional unit of the ethylene receptor family⁸⁸ titration data of the purified TMD with BCA₂-Cu(I) was fitted to a binding isotherm reflecting the following equilibrium:



which consists of the two partial reactions



and



The equilibrium constant for eq. 2 then is

$$\begin{aligned}
 K_R &= \frac{[(ETR1_TMD)_2 - Cu(I)_2][BCA]^4}{[(ETR1_TMD)_2][(BCA)_2 - Cu(I)]^2} \\
 &= \frac{[(ETR1_TMD)_2][Cu]^2[BCA]^4}{K_{D(ETR1_TMD)_2-Cu(I)}[(ETR1_TMD)_2][(BCA)_2 - Cu(I)]^2} \\
 &= \frac{1}{K_{D(ETR1_TMD)_2-Cu(I)}\beta_2[(BCA)_2 - Cu(I)]^2} \quad (3)
 \end{aligned}$$

From eq. 3, the dissociation constant for eq. 2b, $K_{D(ETR1_TMD)_2-Cu(I)}$, is calculated from K_R and previous estimates of the formation constant of $BCA_2-Cu(I)$, $\beta_2 = 2 \times 10^{17} M^{-2.89}$.

$$K_{D(ETR1_TMD)_2-Cu(I)} = \frac{1}{K_R \beta_2 [(BCA)_2 - Cu(I)]^2} = \frac{1}{1.5 \times 10^{-5} M^2 \times 4 \times 10^{34} M^{-4}} = 1.7 \times 10^{-30} M^2 \quad (4)$$

This value corresponds to a dissociation constant $K_{D(ETR1_TMD)_2-Cu(I)}^{v2}$ of $1.3 \times 10^{-15} M$ per copper equivalent.

References

- Hothorn, M. *et al.* Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* **474**, 467–471, <https://doi.org/10.1038/nature10153> (2011).
- She, J. *et al.* Structural insight into brassinosteroid perception by BRI1. *Nature* **474**, 472–U496, <https://doi.org/10.1038/nature10178> (2011).
- Miyazono, K. *et al.* Structural basis of abscisic acid signalling. *Nature* **462**, 609–614, <https://doi.org/10.1038/nature08583> (2009).
- Nishimura, N. *et al.* Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science* **326**, 1373–1379, <https://doi.org/10.1126/science.1181829> (2009).
- Santiago, J., Henzler, C. & Hothorn, M. Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* **341**, 889–892, <https://doi.org/10.1126/science.1242468> (2013).
- Zhang, H. *et al.* SERK Family Receptor-like Kinases Function as Co-receptors with PXY for Plant Vascular Development. *Mol Plant* **9**, 1406–1414, <https://doi.org/10.1016/j.molp.2016.07.004> (2016).
- Cao, M. *et al.* An ABA-mimicking ligand that reduces water loss and promotes drought resistance in plants. *Cell Res.* **23**, 1043–1054, <https://doi.org/10.1038/cr.2013.95> (2013).
- O'Malley, R. C. *et al.* Ethylene-binding activity, gene expression levels, and receptor system output for ethylene receptor family members from Arabidopsis and tomato. *Plant J.* **41**, 651–659, <https://doi.org/10.1111/j.1365-313X.2004.02331.x> (2005).
- Chen, Y. F. *et al.* Ethylene receptors function as components of high-molecular-mass protein complexes in Arabidopsis. *PLoS One* **5**, e8640, <https://doi.org/10.1371/journal.pone.008640> (2010).
- Rodríguez, F. I. *et al.* A copper cofactor for the ethylene receptor ETR1 from Arabidopsis. *Science (New York, N.Y.)* **283**, 996–998, <https://doi.org/10.1126/science.283.5404.996> (1999).
- Suenaga, Y., Ping, W. L., Kuroda-sowa, T., Munakata, M. & Maekawa, M. Structure and 1H NMR study of copper(I) complex with ethylene and tetramethylethylenediamine. *Polyhedron* **16**, 67–70, [https://doi.org/10.1016/0277-5387\(96\)00266-5](https://doi.org/10.1016/0277-5387(96)00266-5) (1997).
- Hirsch, J. *et al.* Raman and Extended X-ray Absorption Fine Structure Characterization of a Sulfur-Ligated Cu (I) Ethylene Complex: Modeling the Proposed Ethylene Binding Site of Arabidopsis thaliana. 2439–2441 (2001).
- Bleeker, A. B., Esch, J. J., Hall, A. E., Rodríguez, F. I. & Binder, B. M. The ethylene-receptor family from Arabidopsis: structure and function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353**, 1405–1412, <https://doi.org/10.1098/rstb.1998.0295> (1998).
- Mayerhofer, H. *et al.* Structural Model of the Cytosolic Domain of the Plant Ethylene Receptor 1 (ETR1). *Journal of Biological Chemistry* **290**, 2644–2658, <https://doi.org/10.1074/jbc.M114.587667> (2015).
- Müller-Dieckmann, H.-I., Grantz, A. A. & Kim, S.-H. The structure of the signal receiver domain of the Arabidopsis thaliana ethylene receptor ETR1. *Structure* **7**, 1547–1556, [https://doi.org/10.1016/s0969-2126\(00\)88345-8](https://doi.org/10.1016/s0969-2126(00)88345-8) (1999).
- Wang, W. *et al.* Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis. *The Plant cell* **18**, 3429–3442, <https://doi.org/10.1105/tpc.106.044537> (2006).
- Khelashvili, G. *et al.* Computational modeling of the N-terminus of the human dopamine transporter and its interaction with PIP2-containing membranes. *Proteins* **83**, 952–969, <https://doi.org/10.1002/prot.24792> (2015).
- Watschinger, K. *et al.* Catalytic residues and a predicted structure of tetrahydrobiopterin-dependent alkylglycerol mono-oxygenase. *Biochem. J.* **443**, 279–286, <https://doi.org/10.1042/BJ20111509> (2012).
- Antala, S., Ovchinnikov, S., Kamisetty, H., Baker, D. & Dempski, R. E. Computation and functional studies provide a model for the structure of the zinc transporter hZIP4. *J. Biol. Chem.* **290**, 17796–17805, <https://doi.org/10.1074/jbc.M114.617613> (2015).
- Zhang, T. *et al.* Crystal structures of a ZIP zinc transporter reveal a binuclear metal center in the transport pathway. *Science Advances* **3**, e1700344, <https://doi.org/10.1126/sciadv.1700344> (2017).
- Ma, B. *et al.* Subcellular localization and membrane topology of the melon ethylene receptor CmERS1. *Plant Physiol.* **141**, 587–597, <https://doi.org/10.1104/pp.106.080523> (2006).
- Li, S. C. & Ng, Y. K. Calibur: a tool for clustering large numbers of protein decoys. *BMC Bioinform.* **11**, 25, <https://doi.org/10.1186/1471-2105-11-25> (2010).
- Jones, D. T., Singh, T., Kosciółek, T. & Tetchner, S. MetaPSICOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics* **31**, 999–1006, <https://doi.org/10.1093/bioinformatics/btu791> (2015).
- Kuzmanic, A. & Zagrovic, B. Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys J* **98**, 861–871, <https://doi.org/10.1016/j.bpj.2009.11.011> (2010).
- Dimura, M. *et al.* Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. *Curr Opin Struct Biol* **40**, 163–185, <https://doi.org/10.1016/j.sbi.2016.11.012> (2016).
- Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302–2309, <https://doi.org/10.1093/nar/gki524> (2005).
- Barth, P., Wallner, B. & Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc. Natl. Acad. Sci. USA* **106**, 1409–1414, <https://doi.org/10.1073/pnas.0808323106> (2009).
- Yatsunyk, L. A. & Rosenzweig, A. C. Cu(I) Binding and Transfer by the N Terminus of the Wilson Disease Protein. *The Journal of Biological Chemistry* **282**, 8622–8631, <https://doi.org/10.1074/jbc.M609533200> (2007).
- Hill, A. V. The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. I. *Biochem. J.* **7**, 471–480 (1913).
- Dutta, A. & Bahar, I. Metal-binding sites are designed to achieve optimal mechanical and signaling properties. *Structure* **18**, 1140–1148, <https://doi.org/10.1016/j.str.2010.06.013> (2010).

31. De Feo, C. J., Mootien, S. & Unger, V. M. Tryptophan scanning analysis of the membrane domain of CTR-copper transporters. *J. Membr. Biol.* **234**, 113–123, <https://doi.org/10.1007/s00232-010-9239-4> (2010).
32. Guzman, G. R. *et al.* Tryptophan scanning mutagenesis in the alphaM3 transmembrane domain of the Torpedo californica acetylcholine receptor: functional and structural implications. *Biochemistry* **42**, 12243–12250, <https://doi.org/10.1021/bi034764d> (2003).
33. Schaller, G. E., Ladd, A. N., Lanahan, M. B., Spanbauer, J. M. & Bleecker, A. B. The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer. *J. Biol. Chem.* **270**, 12526–12530, <https://doi.org/10.1074/jbc.270.21.12526> (1995).
34. Li, P., Song, L. F. & Merz, K. M. Jr. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **11**, 1645–1657, <https://doi.org/10.1021/ct500918t> (2015).
35. Ansbacher, T. & Shurki, A. Predicting the coordination number within copper chaperones: Atox1 as case study. *J. Phys. Chem. B* **116**, 4425–4432, <https://doi.org/10.1021/jp210678n> (2012).
36. Geri, J. B., Pernicone, N. C. & York, J. T. Comparing the impact of different supporting ligands on copper(I)-ethylene interactions. *Polyhedron* **52**, 207–215, <https://doi.org/10.1016/j.poly.2012.09.046> (2013).
37. Light, K. M., Wisniewski, J. A., Vinyard, W. A. & Kieber-Emmons, M. T. Perception of the plant hormone ethylene: known-knowns and known-unknowns. *J. Biol. Inorg. Chem.* **21**, 715–728, <https://doi.org/10.1007/s00775-016-1378-3> (2016).
38. Pinkas-Kramarski, R. *et al.* ErbB tyrosine kinases and the two neuregulin families constitute a ligand-receptor network. *Mol. Cell. Biol.* **18**, 6090–6101 (1998).
39. Junge, W., Hong, Y. Q., Qian, L. P. & Viale, A. Cooperative transient trapping of photosystem II protons by the integral membrane portion (CF0) of chloroplast ATP-synthase after mild extraction of the four-subunit catalytic part (CF1). *Proc. Natl. Acad. Sci. USA* **81**, 3078–3082 (1984).
40. Delrieu, M. J. Regulation of thermal dissipation of absorbed excitation energy and violaxanthin deepoxidation in the thylakoids of *lactuca sativa*. Photoprotective mechanism of a population of photosystem II centers. *Biochim. Biophys. Acta* **1363**, 157–173, [https://doi.org/10.1016/S0005-2728\(97\)00097-2](https://doi.org/10.1016/S0005-2728(97)00097-2) (1998).
41. McDaniel, B. K. & Binder, B. M. Ethylene receptor 1 (ETR1) is sufficient and has the predominant role in mediating inhibition of ethylene responses by silver in Arabidopsis thaliana. *J. Biol. Chem.* **287**, 26094–26103, <https://doi.org/10.1074/jbc.M112.383034> (2012).
42. Hirayama, T. *et al.* Responsive-To-Antagonist1, a Menkes/Wilson Disease-Related Copper Transporter, Is Required for Ethylene Signaling in Arabidopsis. *Cell* **97**, 383–393, [https://doi.org/10.1016/S0092-8674\(00\)80747-3](https://doi.org/10.1016/S0092-8674(00)80747-3) (1999).
43. Schaarshmidt, J., Monastyrskyy, B., Kryshchak, A. & Bonvin, A. Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. *Proteins* **86**(Suppl 1), 51–66, <https://doi.org/10.1002/prot.25407> (2018).
44. Rubino, J. T., Chenkin, M. P., Keller, M., Riggs-Gelasco, P. & Franz, K. J. A comparison of methionine, histidine and cysteine in copper(I)-binding peptides reveals differences relevant to copper uptake by organisms in diverse environments. *Metalomics* **3**, 61–73, <https://doi.org/10.1039/C0MT00044B> (2011).
45. Rubino, J. T. & Franz, K. J. Coordination chemistry of copper proteins: How nature handles a toxic cargo for essential function. *Journal of Inorganic Biochemistry* **107**, 129–143, <https://doi.org/10.1016/j.jinorgbio.2011.11.024> (2012).
46. Hæffner, F., Brinck, T., Hæberlein, M. & Moberg, C. Force field parameterization of copper(I)-olefin systems from density functional calculations. *Journal of Molecular Structure: THEOCHEM* **397**, 39–50, [https://doi.org/10.1016/S0166-1280\(96\)04978-0](https://doi.org/10.1016/S0166-1280(96)04978-0) (1997).
47. Bhate, M. P., Molnar, K. S., Goullan, M. & DeGrado, W. F. Signal transduction in histidine kinases: insights from new structures. *Structure* **23**, 981–994, <https://doi.org/10.1016/j.str.2015.04.002> (2015).
48. Dobson, L., Reményi, I. & Tusnady, G. E. CCTOP: A Consensus Constrained TOPOlogy prediction web server. *Nucleic Acids Res.* **43**, W408–W412, <https://doi.org/10.1093/nar/gkv451> (2015).
49. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202, <https://doi.org/10.1006/jmbi.1999.3091> (1999).
50. Adamian, L. & Liang, J. Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC Struct. Biol.* **6**, 13, <https://doi.org/10.1186/1472-6807-6-13> (2006).
51. Yarov-Yarovoy, V., Schonbrun, J. & Baker, D. Multipass membrane protein structure prediction using Rosetta. *Proteins* **62**, 1010–1025, <https://doi.org/10.1002/prot.20817> (2006).
52. Yarov-Yarovoy, V., Schonbrun, J., Barth, P. & Wallner, B. Membrane Abinitio, https://www.rosettacommons.org/docs/latest/application_documentation/structure_prediction/membrane_abinitio.
53. Kamisetty, H., Ovchinnikov, S. & Baker, D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. *Proc. Natl. Acad. Sci. USA* **110**, 15674–15679, <https://doi.org/10.1073/pnas.1314045110> (2013).
54. Ovchinnikov, S., Kamisetty, H. & Baker, D. OPENSEQ.org, <http://openseq.org> (2019).
55. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *P1000Res* **5**, 189, <https://doi.org/10.12688/f1000research.7931.1> (2016).
56. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895, <https://doi.org/10.1093/bioinformatics/btq066> (2010).
57. Wallner, B. PQM-resample: Improved Model Quality Assessment for Membrane Proteins by Limited Conformational Sampling. *Bioinformatics (Oxford, England)* **30**, 2221–2223, <https://doi.org/10.1093/bioinformatics/btu187> (2014).
58. Studer, G., Biasini, M. & Schwede, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* **30**, 505–511, <https://doi.org/10.1093/bioinformatics/btu457> (2014).
59. Kryshchak, A., Monastyrskyy, B., Fidells, K., Schwede, T. & Tramontano, A. Assessment of model accuracy estimations in CASP12. *Proteins* **86**(Suppl 1), 345–360, <https://doi.org/10.1002/prot.25371> (2018).
60. Dominguez, C., Boelens, R. & Bonvin, A. M. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society* **125**, 1731–1737, <https://doi.org/10.1021/ja026939x> (2003).
61. Wang, Y. & Barth, P. Evolutionary-guided de novo structure prediction of self-associated transmembrane helical proteins with near-atomic accuracy. *Nature Communications* **6**, 7196, <https://doi.org/10.1038/ncomms8196> (2015).
62. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815, <https://doi.org/10.1006/jmbi.1993.1626> (1993).
63. Im, W., Lee, M. S. & Brooks, C. L. Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **24**, 1691–1702, <https://doi.org/10.1002/jcc.10321> (2003).
64. Feig, M. Computational protein structure refinement: almost there, yet still so far to go. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, e1307, <https://doi.org/10.1002/wcms.1307> (2017).
65. Feig, M., Karanicolas, I. & Brooks, C. L. 3rd MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **22**, 377–395, <https://doi.org/10.1016/j.jmgn.2003.12.005> (2004).
66. Roe, D. R. & Cheatham, T. E. 3rd (PTRA) and CPPTRA]: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **9**, 3084–3095, <https://doi.org/10.1021/ct400341p> (2013).
67. Coughlan, S. J., Hastings, C. & Winfrey, R. J. Molecular characterisation of plant endoplasmic reticulum - Identification of protein disulfide-isomerase as the major reticuloplasm. *European Journal of Biochemistry* **235**, 215–224, <https://doi.org/10.1111/j.1432-1033.1996.00215.x> (1996).

68. Schott-Verdugo, S. & Gohlke, H. PACKMOL-Mengen: A simple-to-use generalized workflow for membrane-protein/lipid-bilayer system building. *Journal of Chemical Information and Modeling* (2019).
69. Le Grand, S., Götz, A. W. & Walker, R. C. SPPF: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380, <https://doi.org/10.1016/j.cpc.2012.09.022> (2013).
70. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713, <https://doi.org/10.1021/acs.jctc.5b00255> (2015).
71. Dickson, C. I. *et al.* Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **10**, 865–879, <https://doi.org/10.1021/ct4010307> (2014).
72. AMBER 2018 (University of California, San Francisco, 2018).
73. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926–935, <https://doi.org/10.1063/1.445869> (1983).
74. Ryckaert, J.-P., Ciccolini, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341, [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5) (1977).
75. Quigley, D. & Probert, M. I. Langevin dynamics in constant pressure extended systems. *J. Chem. Phys.* **120**, 11432–11441, <https://doi.org/10.1063/1.1755657> (2004).
76. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **81**, 3684–3690, <https://doi.org/10.1063/1.448118> (1984).
77. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174, <https://doi.org/10.1002/jcc.20035> (2004).
78. Bayly, C. L., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280, <https://doi.org/10.1021/j100142a004> (1993).
79. Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009).
80. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260, <https://doi.org/10.1016/j.jmgn.2005.12.005> (2006).
81. Gohlke, H. *et al.* Binding Region of Alanine Dehydrogenase Predicted by Unbiased Molecular Dynamics Simulations of Ligand Diffusion. *J. Chem. Inf. Model.* **53**, 2493–2498, <https://doi.org/10.1021/ci400370y> (2013).
82. Bhatia, S. *et al.* Targeting HSP90 dimerization via the C terminus is effective in imatinib-resistant CML and lacks the heat shock response. *Blood* **132**, 307–320, <https://doi.org/10.1182/blood-2017-10-810986> (2018).
83. Bisson, M. M., Bleckmann, A., Allekotte, S. & Groth, G. EIN2, the central regulator of ethylene signalling, is localized at the ER membrane where it interacts with the ethylene receptor ETR1. *Biochem. J.* **424**, 1–6, <https://doi.org/10.1042/BJ20091102> (2009).
84. Folio, C. & Isidoro, C. A fast and simple method for simultaneous mixed site-specific mutagenesis of a wide coding sequence. *Biotechnol. Appl. Biochem.* **49**, 175–183, <https://doi.org/10.1042/BA20070045> (2008).
85. Voel-van-Vormizeele, J. & Groth, G. High-level expression of the Arabidopsis thaliana ethylene receptor protein ETR1 in Escherichia coli and purification of the recombinant protein. *Protein Expr. Purif.* **32**, 89–94, [https://doi.org/10.1016/S1046-5928\(03\)00215-8](https://doi.org/10.1016/S1046-5928(03)00215-8) (2003).
86. Drees, S. L. & Lübben, M. Analytical Gel Filtration for Probing Heavy Metal Transfer between Proteins. *Bio-protocol* **6**(15), e1888, <https://doi.org/10.21769/BioProtoc.1888> (2016).
87. Wermimont, A. K., Yatsunyk, L. A. & Rosenzweig, A. C. Binding of copper(I) by the Wilson disease protein and its copper chaperone. *J. Biol. Chem.* **279**, 12269–12276, <https://doi.org/10.1074/jbc.M311213200> (2004).
88. Shakeel, S. N., Wang, X., Binder, B. M. & Schaller, G. E. Mechanisms of signal transduction by ethylene: overlapping and non-overlapping signalling roles in a receptor family. *AoB Plants* **5**, pl010–pl010, <https://doi.org/10.1093/aobpla/pl010> (2013).
89. Xiao, Z., Donnelly, P. S., Zimmermann, M. & Wedd, A. G. Transfer of copper between bis(thiosemicarbazone) ligands and intracellular copper-binding proteins: insights into mechanisms of copper uptake and hypoxia selectivity. *Inorg. Chem.* **47**, 4338–4347, <https://doi.org/10.1021/ic702440e> (2008).

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 267205415 – SFB 1208 (project B06 to G.G. and A03 to H.G.). H.G. is grateful to the Jülich Supercomputing Centre at the Forschungszentrum Jülich for computing time on the supercomputer JURECA and JUWELS (NIC project ID: HKF7) and to the “Zentrum für Informations- und Medientechnologie” (ZIM) at the Heinrich Heine University Düsseldorf for computational support. The authors declare no conflict of interests.

Author Contributions

H.G. and G.G. designed the research. H.G. and S.S.-V. designed molecular modeling and simulation work. G.G., E.C. and L.M. designed wet-lab experiments. S.S.-V., L.M. and E.C. performed research. S.S.-V., L.M., G.G. and H.G. analyzed data. All authors wrote the article. H.G. and G.G. acquired funding.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-45189-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Supporting Information

Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain

Stephan Schott-Verdugo^{1,2,†}, Lena Müller^{3,†}, Elisa Classen^{3,4},

Holger Gohlke^{1,5,6,*}, Georg Groth^{3,6,*}

¹Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University
Düsseldorf, Düsseldorf, Germany

²Centro de Bioinformática y Simulación Molecular (CBSM), Facultad de Ingeniería,
Universidad de Talca, 2 Norte 685, CL-3460000 Talca, Chile

³Institute of Biochemical Plant Physiology, Heinrich Heine University Düsseldorf,
Düsseldorf, Germany

⁴Present address: Institute of Hydraulic Engineering and Water Resources Management,
RWTH Aachen University, Aachen, Germany

⁵John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) &
Institute for Complex Systems - Structural Biochemistry (ICS 6), Forschungszentrum Jülich
GmbH, Jülich, Germany

⁶Bioeconomy Science Center, Forschungszentrum Jülich, Jülich, Germany

*Corresponding Authors: Holger Gohlke; Georg Groth

Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: (+49) 211 81 13662 or 12822

E-mail: gohlke@uni-duesseldorf.de, groth@uni-duesseldorf.de

1 Supplementary Methods

a) **membrane_abinitio2:**

The *ab initio* folding was performed in batches of 100 structures across multiple computers using the following flags (bold text represents options/paths that have to be replaced appropriately):

```
-in:file:fasta FASTA_FILE
-seed_offset RANDOM_NUMBER
-in:file:spanfile SPAN_FILE
-in:file:lipofile LIPS4_FILE
-in:file:frag3 3MERS_FILE
-in:file:frag9 9MERS_FILE
-in:path:database DATABASE_PATH
-abinitio:membrane
-score:find_neighbors_3dgrid
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
-out:file:silent SILENT_FILE
-out:sf SCORE_FILE
-out:nstruct 100
-out:file:silent_struct_type binary
-mute all

-constraints
  -cst_file CONSTRAINT_FILE
  -cst_weight 4
```

b) relax:

The selected centroid model was relaxed to incorporate side chains using the following flags:

```
-in:file:s PDB_FILE
-in:file:spanfile SPAN_FILE
-in:file:lipofile LIPS4_FILE
-in:path:database DATABASE_PATH
-relax:thorough
-out:file:silent_struct_type binary
-membrane
-membrane:Membed_init
-score:weights membrane_highres.wts
-nstruct 100
-out:file:silent SILENT_FILE
```

c) score_jd2:

The predicted membrane orientation was obtained by rescoring with the following flags:

```
-in:file:silent SILENT_FILE
-in:file:spanfile SPAN_FILE
-in:file:lipofile LIPS4_FILE
-in:path:database DATABASE_PATH

-score:weights score_membrane
-score:find_neighbors_3dgrid
-membrane:no_interpolate_Mpair
-membrane:Menv_penalties
```

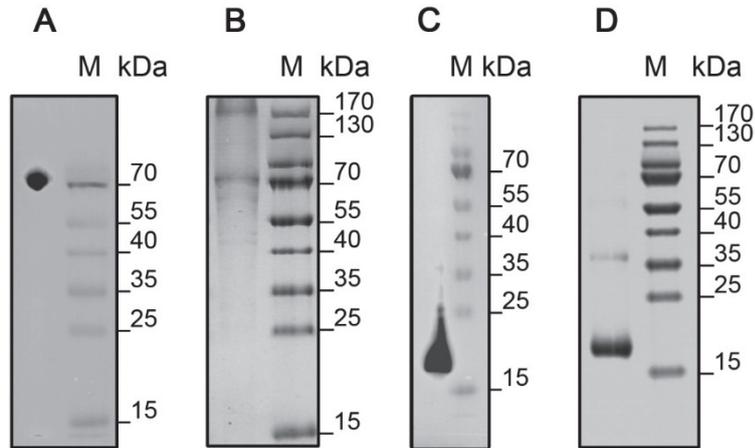
2 Supplementary Results

ETR1 transmembrane domain homolog search. The homolog search for the transmembrane domain of ETR1 was performed using the same N-terminal 117 residues as for *ab initio* modeling. For this, the threading module of TopModel was used ¹. The top three identified homologues were (sequence identity and protein in parenthesis) 2C12_A (11.3%, nitroalkane oxidase), 1EGD_A (9.7%, acyl-CoA dehydrogenase) and 1RX0_A (7.3%, isobutyryl-CoA dehydrogenase), none of which corresponds to transmembrane proteins. Even when a transmembrane protein could be modeled based on the hydrophobic core of a soluble protein, the additional low sequence identity² and the lack of homology urged us to sample the fold space by *ab initio* modelling.

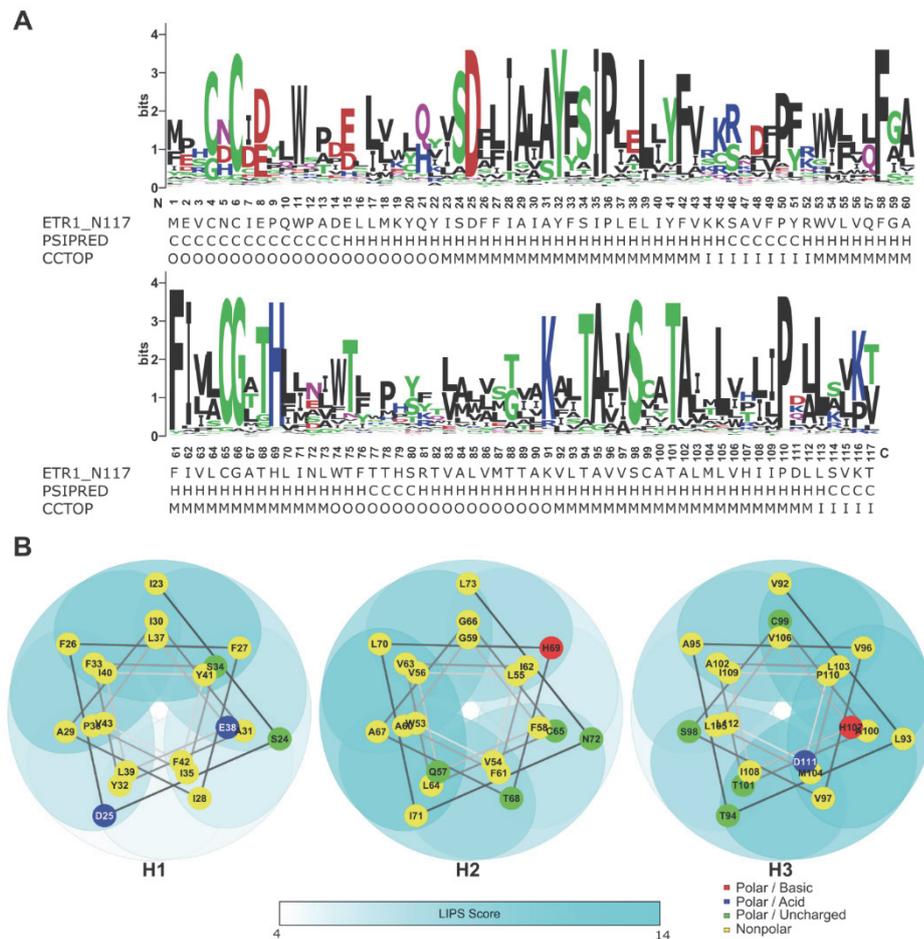
MetaPSICOV and Rosetta constraint files. Files containing both constraints are accessible as separate files at <https://uni-duesseldorf.sciebo.de/s/i0s6IKIZmw8v5Ef>.

PDB file of the ETR1_TMD/Cu dimer. Coordinates of the dimer model depicted in Figure 4 are accessible as a separate PDB file at <https://uni-duesseldorf.sciebo.de/s/i0s6IKIZmw8v5Ef>.

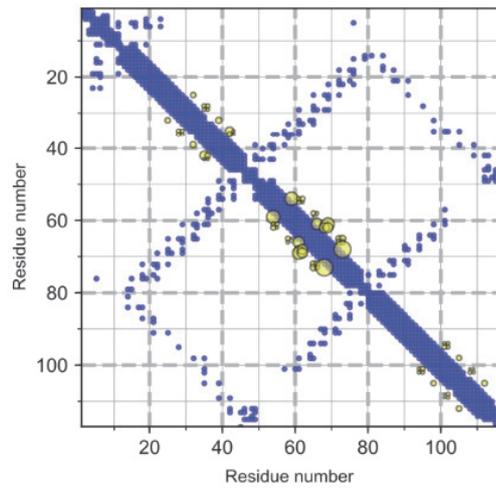
3 Supplementary Figures



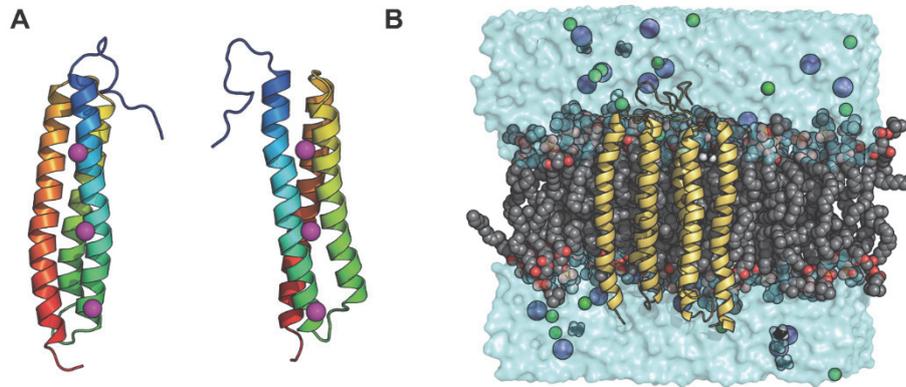
SI Figure 1. Expression and purification of ETR1 tryptophan mutants and ETR1_TMD. *E. coli* C43 (DE3) was transformed with a pET16b vector carrying the DNA sequence of ETR1 tryptophan mutants and ETR1_TMD, respectively. ETR1 tryptophan mutants (A) were expressed for 5 h at 30 °C and ETR1_TMD (C) was expressed for 20 h at 16 °C. For both expressions host cell extract was analyzed by western blotting. For protein detection an Anti-His antibody targeting the proteins deca-histidine tag was used. For purification from the bacterial host, ETR1 tryptophan mutants (B) were purified by IMAC and ETR1_TMD (D) was purified by IMAC and SEC. Analysis was performed by coomassie stained SDS-PAGE.



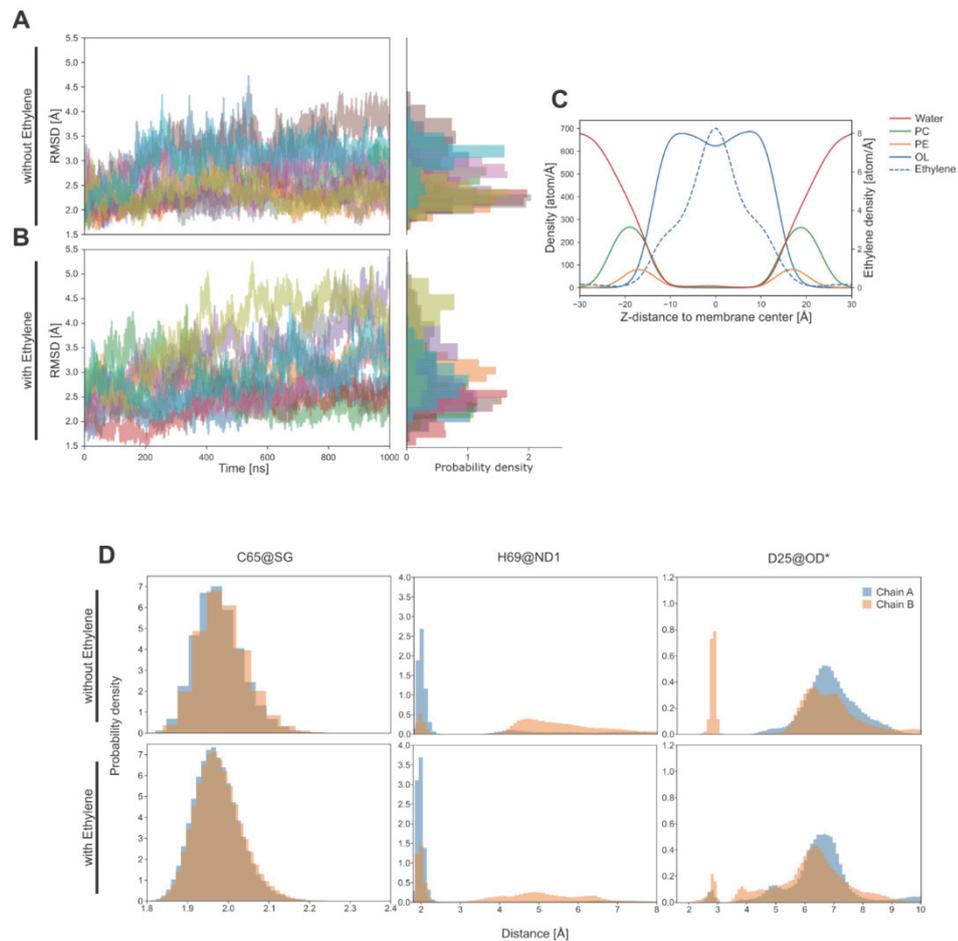
SI Figure 2. Analysis of modeled ETR1_TMD structure. A) Weblogo³, PSIPRED secondary structure, and CCTOP transmembrane topology predictions for the modeled structure. C = coil; H = helix; O = extra-cytosolic region; M = transmembrane region; I = cytosolic region. B) Schematic representation of the LIPS score of the surfaces of the predicted helical transmembrane helices H1, H2 and H3. A higher LIPS score implies a higher propensity to be exposed to the membrane environment. The helical wheel representations were modified from NetWheels⁴.



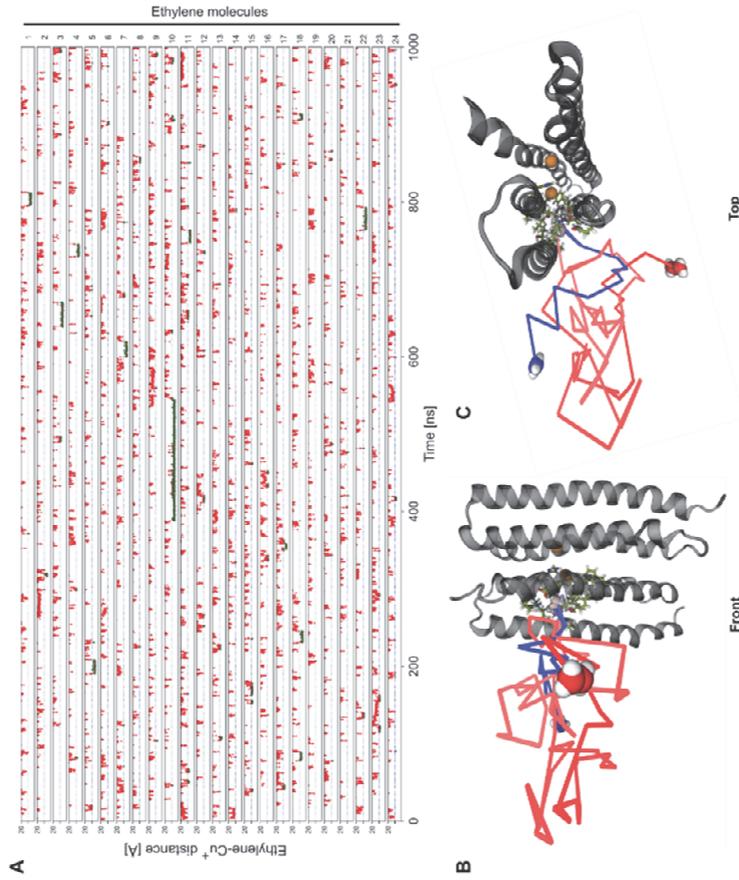
SI Figure 3. Coevolutionary signals on transmembrane helix surfaces with high LIPS score and at a sequence distance of 8 or less residues (yellow). These signals have been proposed as being indicative of dimerization interface. The signals are shown with a radius 30 times bigger than those shown in Figure 2 in the main text. The contact map of the selected monomer model is shown for reference (blue).



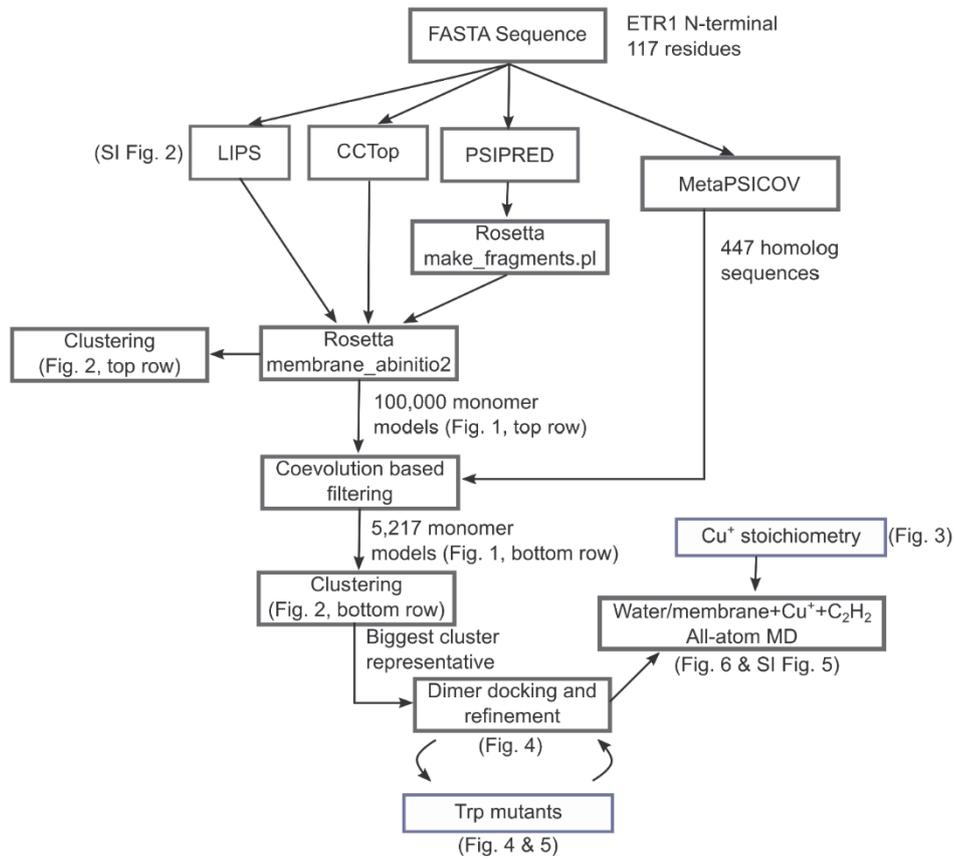
SI Figure 4. ETR1 membrane orientation. A) Selected monomer model, colored blue to red from the N-terminal to the C-terminal region, shown from two different orientations. The membrane orientation (“MEM” residue) obtained from Rosetta is superimposed in magenta spheres, where the central sphere represents the center of the membrane, and the external spheres represent the predicted upper and lower water-membrane interfaces. B) Representative snapshot of the dimer model in the molecular dynamics setup. All atoms in front of the protein were removed to allow a direct view onto the protein. The protein is shown in yellow, chloride ions in green, potassium ions in purple, and water as a transparent cyan surface. Ethylene and the aliphatic tails of the lipids are shown in black and grey, respectively. Hydrogens of the latter were omitted for clarity.



SI Figure 5. Analysis of MD simulations. A) & B) C_{α} RMSD for the simulated systems without (A) or with (B) ethylene with respect to the average structure. Residues 15-117 of both chains were considered for the calculation. Different colors represent different trajectories. On the right, normalized distributions of the time plots are shown. C) Decomposed density profile along the membrane axis. The plot shows the proportion of every component at a given distance of the membrane center. The dashed line shows the density for ethylene, with a peak at the center of the membrane slab. The ethylene molecules were added to the water bulk at the start and initially diffuse into the membrane during the simulations, from which they bind to the embedded dimer model of the ETR1 TMD. D) Histograms of distances between chelating atoms of C65, H69, or D25 and the Cu^{+} ion of the corresponding chain. Interactions with C65 prevail along the whole trajectory.



SI Figure 6. Ethylene binding during MD simulations of free ligand diffusion. A) Distance between the center of mass of the different ethylene molecules and one of the copper ions during the simulations. Only one representative replica and distances below 20 Å are shown. Distances below 10 Å are depicted in green to highlight binding events. B) and C) show the path of one representative binding event, with the panel in C rotated by 90 degrees with respect to the one in B). The coloring goes from red to blue, according to the simulation time. Binding site residues are depicted in green, while the Cu⁺ ions are depicted as orange spheres.



SI Figure 7. Summary of the modelling (black boxes) and experimental validation (blue boxes) process followed in this study. The main modelling tasks are depicted as a flowchart, with references to corresponding figures in the article. For details, refer to the main text.

4 Supplementary Tables

SI Table 1. Oligonucleotides used for cloning of ETR1 tryptophan mutants.

Oligonucleotides were used for cloning of tryptophan mutants building megaprimer or by round-the-horn site-directed mutagenesis. Oligonucleotides used for round-the-horn site-directed mutagenesis were 5 prime phosphorylated.

ETR1 tryptophan mutant	direction	Oligonucleotide
ETR1 ^{W7X} _F26W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-AATCGCAATGAACCAATCGGAGATGTA-3'
ETR1 ^{W7X} _F27W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-CGCAATCGCAATCCAGAAAATCGGAGAT-3'
ETR1 ^{W7X} _A29W	for	5'-ATTGCGTATTTTCGATTCTCTTGAGTTG-3'
	rev	5'-CCAAAATGAAGAAAATCGGAGATGTTGGTA-3'
ETR1 ^{W7X} _F33W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-AAGAGGAATCGACCAATACGCAATCGC-3'
ETR1 ^{W7X} _L39W	for	5'-GATTTACTTTGTGAAGAAAATCAGCCGTGTT-3'
	rev	5'-CACTCAAGAGGAATCGAAAAATACGCAATC-3'
ETR1 ^{W7X} _V54W	for	5'-CTTGTTCAGTTTGGTGCTTTTATCGTTCTT-3'
	rev	5'-CCAAAATCTATACGGAAAACCGGCTGATT-3'
ETR1 ^{W7X} _L55W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-ACCAAATGAACCCATACAAATCTATA-3'
ETR1 ^{W7X} _F58W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-GATAAAAGCACCCACTGAACAAGTAC-3'
ETR1 ^{W7X} _L64W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-AGTTGCTCCACACCAAACGATAAAAAGC-3'
ETR1 ^{W7X} _T68W	for	5'-TGGCATCTTATTAACCTATTTACTTTCACT-3'
	rev	5'-TGCTCCACAAAGAACGATAAAAAGC-3'
ETR1 ^{W7X} _L70W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-AAATAAGTTAATCCAATGAGTTGCTCC-3'
ETR1 ^{W7X} _N72W	for	5'-TGGTTATTTACTTTCACTACGCATTGCGA-3'
	rev	5'-AATAAGATGAGTTGCTCCACAAAGAACG-3'
ETR1 ^{W7X} _L73W	for	5'-GTAATTCATATGGAAGTCTGCAATTG-3'
	rev	5'-AGTGAAAGTAAACCAGTTAATAAGATG-3'
ETR1 ^{W7X} _T75W	for	5'-ATTAACCTATTTTGGTTCACTACGCATTGCG-3'
	rev	5'-AAGATGAGTTGCTCCACAAAGAACG-3'
ETR1 ^{W7X} _A95W	for	5'-GTTGTCTCGTGTGCTACTGCGTTGAT-3'
	rev	5'-CCAGGTTAACACCTTCGCGGTAGTC-3'
ETR1 ^{W7X} _S98W	for	5'-GCTGTTGCTGGTGTGCTACTGC-3'
	rev	5'-GGTTAACACCTTCGCGGTAGTCAT-3'
ETR1_TMD	for	5'-TAAGGATCCGGCTGCTAACA-3'
	rev	5'-TAAAGTGCTTCTAATCTCATGAGTC-3'

SI Table 2. Amounts of secondary structure in ETR1 tryptophan mutants determined by CD spectroscopy. The composition of structural elements in all constructed ETR1 tryptophan mutants were investigated by CD spectroscopy.

ETR1 tryptophan mutant	α-helices [%]	β-strands [%]	β-turn [%]	Random coil [%]
ETR1 ^{W7X}	33	18	21	28
ETR1 ^{W7X} _F26W	35	17	19	29
ETR1 ^{W7X} _F27W	34	18	20	28
ETR1 ^{W7X} _A29W	25	25	22	28
ETR1 ^{W7X} _F33W	35	18	19	28
ETR1 ^{W7X} _L39W	27	23	22	28
ETR1 ^{W7X} _V54W	22	27	23	28
ETR1 ^{W7X} _L55W	35	17	20	28
ETR1 ^{W7X} _F58W	36	18	19	27
ETR1 ^{W7X} _L64W	32	19	21	28
ETR1 ^{W7X} _T68W	30	20	23	27
ETR1 ^{W7X} _L70W	33	18	21	28
ETR1 ^{W7X} _N72W	31	19	21	29
ETR1 ^{W7X} _L73W	35	17	20	28
ETR1 ^{W7X} _T75W	27	21	23	29
ETR1 ^{W7X} _A95W	26	24	23	27
ETR1 ^{W7X} _S98W	29	20	22	29

5 Supplementary References

- 1 Mulnaes, D. & Gohlke, H. TopScore: Using Deep Neural Networks and Large Diverse Data Sets for Accurate Protein Model Quality Assessment. *J Chem Theory Comput* **14**, 6117-6126, doi:10.1021/acs.jctc.8b00690 (2018).
- 2 Forrest, L. R., Tang, C. L. & Honig, B. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys J* **91**, 508-517, doi:10.1529/biophysj.106.082313 (2006).
- 3 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
- 4 Mol, A. R., S. Castro, M. & Fontes, W. NetWheels: A web application to create high quality peptide helical wheel and net projections. *bioRxiv*, doi:10.1101/416347 (2018).

Publication III

Publication III: Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence

Florian Bleffert, Joachim Granzin, Muttalip Caliskan, Stephan Schott-Verdugo, Meike Siebers, Björn Thiele, Laurence Rahme, Peter Dörmann, Holger Gohlke[§], Renu Batra-Safferling[§], Karl-Erich Jaeger and Filip Kovačić[§]

Submitted to Proc. Natl. Acad. Sci. U. S. A., 2020

[§] Authors contributed equally to this work.

Structural and mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence

Florian Bleffert¹, Joachim Granzin², Muttalip Caliskan¹, Stephan N. Schott-Verdugo^{3,4}, Meike Siebers^{5,6}, Björn Thiele⁷, Laurence Rahme⁸, Peter Dörmann⁵, Holger Gohlke^{2,3,9,*}, Renu Batra-Safferling^{2,*}, Karl-Erich Jaeger^{1,10} and Filip Kovačič^{1,*}

1. Institut für Molekulare Enzymtechnologie, Heinrich-Heine Universität Düsseldorf, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
2. Institute of Complex Systems - Structural Biochemistry (ICS-6), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
3. Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany
4. Centro de Bioinformática y Simulación Molecular (CBSM), Faculty of Engineering, University of Talca, 1 Poniente 1141, Casilla 721, Talca, Chile
5. Institute of Molecular Physiology and Biotechnology of Plants (IMBIO), University of Bonn, D-53115 Bonn, Germany
6. Institute for Plant Genetics, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf, Germany
7. Institute of Bio- and Geosciences, Plant Sciences (IBG-2) and Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
8. Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA, USA
9. John von Neumann Institute for Computing (NIC) & Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
10. Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

* *Corresponding authors*

Abstract

Cells are steadily adapting their membrane glycerophospholipid (GPL) composition to changing environmental and developmental conditions. Although the mechanism to maintain membrane homeostasis *via* GPL biosynthesis is known, the mechanism underlying the controlled degradation of endogenous GPLs is unknown. Here, we show that the cytoplasmic membrane-bound phospholipase A1 from *Pseudomonas aeruginosa* (PlaF) alters the membrane GPL composition, which leads to attenuated virulence of this human pathogen. We present a 2.0-Å-resolution crystal structure of a homodimer of PlaF, the first structure of a single-span transmembrane protein in full-length. The structure reveals homodimerization *via* the transmembrane and juxtamembrane regions and an intricate ligand-mediated interaction network that bridges the catalytic and dimerization sites. We used the structure to rationalize how PlaF activity is regulated by product-feedback inhibition and a dimer-to-monomer transition followed by tilting of the monomer in the membrane. These data provide a framework for understanding PLA-mediated membrane GPL degradation and pave the way for the development of novel therapeutics against *P. aeruginosa* infections.

Keywords: *Pseudomonas aeruginosa*, crystal structure, dimer, juxtamembrane region, transmembrane helix, fatty acid, virulence factor

Biological membranes are steadily changing and adapting to environmental and developmental conditions.^{1,2} These changes affect processes indispensable for cell life such as nutrient uptake,³ chemical signaling,⁴ protein secretion,⁵ and folding,⁶ interaction with hosts,⁷ antibiotic resistance⁸. An important mechanism to maintain membrane functionality in bacteria is the alteration of lipid composition.^{9,10,11} One of the best understood and probably the most important mechanism of membrane lipid homeostasis is the adjustment of the FA composition in glycerophospholipids (GPL) upon thermal adaptation.^{12,13} Furthermore, bacteria maintain membrane functionality under stress conditions by chemical modifications of FAs in GPLs^{14,15} and during the transition to stationary growth phase or biofilm lifestyle by tuning the proportions of lipid classes and polar head groups in GPLs,^{16,17} although the latter is poorly understood.

Alteration of membrane lipid composition is also thought to affect virulence properties of the human pathogen *Pseudomonas aeruginosa*,^{11,17,18} which is listed as a critical pathogen due to difficult-to-treat infections.¹⁹ The extraordinary pathogenicity of *P. aeruginosa* is associated with the production of a broad spectrum of cell-associated and secreted virulence factors.^{20,21} Among them are several phospholipases,²²⁻²⁴ belonging to the diverse group of enzymes that cleave ester (phospholipases A1 and A2) or phosphoester (phospholipases C and D) bonds in GPLs.²⁵ Their hydrolytic activity toward GPLs is commonly related to host tissue destruction by cell lysis²⁶ or to immunomodulatory effects in hosts achieved through the release of bioactive compounds.²⁷ These modes of action require that phospholipases are secreted into the medium or directly into the host cells.^{24,27} On the other hand, some bacterial phospholipases support bacterial survival and adaptation without harming the host.^{28,29} Whether and how bacteria utilize intracellular phospholipases A to maintain their membrane homeostasis is not known. We previously identified several lipolytic enzymes of *P. aeruginosa* with putative roles for virulence.³⁰⁻³² Here, we describe PlaF as a novel phospholipase A1 from *P. aeruginosa*, which is anchored *via* a single N-terminal TM helix to the cytoplasmic membrane where it degrades endogenous GPLs.

Notably, all hitherto known cytoplasmic membrane-spanning protein domains adopt α -helical structures;³³ these structures govern vital biological processes in bacteria and eukaryotes through dimerization.^{27,28} Prominent examples are proteins of the receptor tyrosine kinase family proteins that contain a single-pass TM α -helix followed by a JM region that links the TM with a soluble kinase domain and dimerize *via* the TM α -helices to maintain the kinase in a

catalytically active state.³⁴ Another example is the ToxR-family of transcriptional regulators, widespread among bacterial pathogens, where a single-pass TM α -helix is linked to a periplasmic sensory domain and a cytoplasmic JM region followed by a DNA binding domain³⁵; homodimerization *via* the TM α -helix and the JM region is required to achieve the DNA binding-competent state.³⁶ However, while the role TM helix-mediated dimerization for the functional regulation of single-span membrane proteins is undisputed, the exact structural mechanism has remained enigmatic due to the lack of a high-resolution structure of a homodimer of a full-length single-span membrane protein.

Here, we thus set out to resolve the X-ray crystal structure of the homodimer of PlaF as a basis to provide mechanistic insights into phospholipase A-mediated membrane phospholipid degradation related to the bacterial virulence.

Results

PlaF is a cytoplasmic membrane-bound phospholipase A that alters membrane glycerophospholipid composition. In our previous studies, we purified PlaF from the membrane fractions of *E. coli* and *P. aeruginosa* overexpressing *plaF*.^{37,38} Here, we show that native PlaF co-localizes with the known membrane protein XcpQ to the membranes of the *P. aeruginosa* strain carrying the *p-plaF* plasmid, while it was absent in the soluble fraction containing cytoplasmic and periplasmic proteins (Fig. 1a). We next investigated if PlaF is an integral or peripheral membrane protein by treating the *P. aeruginosa p-plaF* membranes with Na_2CO_3 and urea, which destabilize weak interactions between peripheral proteins and the membrane. Under these conditions, PlaF remained membrane-associated, indicating its integral membrane localization (Fig. 1b). Consistent with our previous study,³⁷ PlaF was released from membranes after treatment with Triton X-100 (Fig. 1b). Cytoplasmic and outer membranes of *P. aeruginosa p-plaF* were separated by ultracentrifugation in a sucrose density gradient, which was confirmed by immunodetection of the cytoplasmic membrane protein SecG³⁹ and the outer membrane-associated lipid A⁴⁰. The majority of PlaF esterase activity was detected in the low-sucrose fractions (# 9-13) where SecG was detected (Fig. 1c), whereas Lipid A was detected in the high-sucrose fractions that showed negligible esterase activity (Fig. 1c), indicating that PlaF is a cytoplasmic integral membrane protein. The sequence analysis of PlaF indicated the presence of a transmembrane domain at the N-terminus.³⁷ We performed proteolysis experiments in which *P. aeruginosa p-plaF* cells with a chemically permeabilized

outer membrane were treated with trypsin. The results revealed a time-dependent degradation of PlaF (Fig. 1d) supporting a cellular localization model in which PlaF is anchored in the cytoplasmic membrane of *P. aeruginosa* with its N-terminal transmembrane helix, while its catalytic C-terminal domain protrudes into the periplasm.

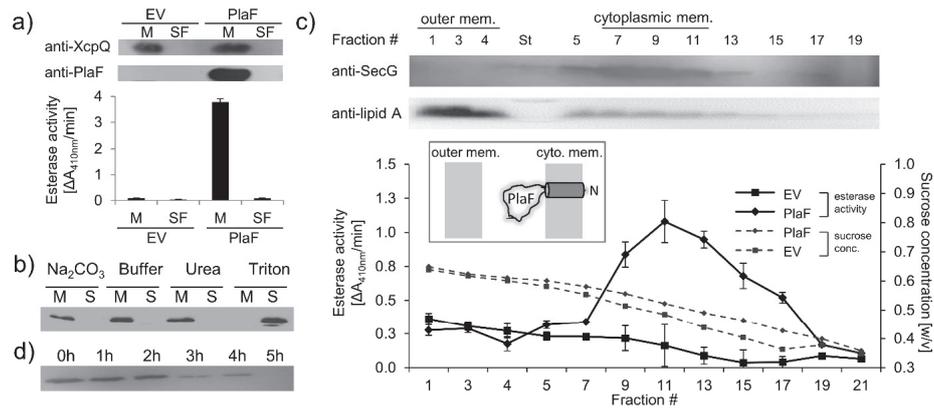


Fig. 1: Subcellular localization of PlaF. a) PlaF is a membrane protein of *P. aeruginosa*. The membrane (M) and soluble fractions (SF) of cell extracts from *P. aeruginosa* p-plaF and the empty vector control strain (EV) were separated, analyzed by immunodetection with anti-His₆-tag antibodies and by esterase activity assay with p-NPC₆ substrate. The membrane protein marker *P. aeruginosa* XcpQ was detected with anti-XcpQ antibodies. **b) PlaF is an integral membrane protein of *P. aeruginosa*.** The crude membranes of *P. aeruginosa* p-plaF were treated with sodium carbonate (Na₂CO₃), urea, Triton X-100, or MES buffer control followed by ultracentrifugation (S, supernatant; M, membrane proteins). PlaF was detected as in panel a. **c) PlaF is a cytoplasmic-membrane protein of *P. aeruginosa*.** The membrane fractions of *P. aeruginosa* p-plaF and the EV strains were isolated and separated by ultracentrifugation in a sucrose density gradient. The esterase activity was assayed as in panel a. *P. aeruginosa* SecG and outer membrane lipid A were used as markers for cytoplasmic and outer membranes and detected by Western blotting using anti-SecG and anti-Lipid A antibodies, respectively. Inlet: A model of PlaF cellular localization. All values are mean \pm standard deviation (S.D.) of three independent experiments measured in triplicates. **d) The catalytic domain of PlaF is exposed to the periplasm.** *P. aeruginosa* p-plaF cells with permeabilized outer membrane were treated with trypsin for the indicated time periods and PlaF was detected as described in figure 1a.

We next analyzed the hydrolytic function of PlaF against a spectrum of artificial and natural GPLs. PlaF purified with DDM as described previously³⁷ showed PLA1 activity yet no PLA2

activity towards the artificial substrates specific to each of these two phospholipase families (Fig. 2a). The substrate profile of PlaF against natural di-acyl GPLs commonly occurring in *P. aeruginosa* membranes¹⁷ was determined with a spectrum of substrates (see legend to Fig. 2b). Purified PlaF *in vitro* hydrolyzed GPLs containing medium-chain FAs with ethanolamine, glycerol and choline head groups to a similar extent (Fig. 2b).

To address the role of membrane-bound PlaF in the regulation of the membrane GPL content *in vivo*, we constructed by homologous recombination the *P. aeruginosa* deletion mutant Δ plaF which is missing the entire *plaF* gene (Figs. S1a-c). The PLA1 activity assay showed 60% loss of activity indicating that PlaF is a major but not the only PLA1 in *P. aeruginosa* (Fig. S1c). The Q-TOF-MS/MS analysis of total GPL content isolated from four biological replicates (n = 4) of *P. aeruginosa* wild-type and Δ plaF revealed significant differences in membrane GPL composition (Fig. 2c, Table S1). Statistical analysis of 334 GPL molecular species identified 16 significantly accumulated phospholipids ($p < 0.05$) in *P. aeruginosa* Δ plaF. These GPLs are candidates for native substrates (Table S2). The most significantly accumulated GPLs were monounsaturated phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and phosphatidylcholine (PC) species with two FA containing 35 carbon atoms. We also identified 19 GPLs significantly depleted in *P. aeruginosa* Δ plaF, in line with the finding that the total GPL contents of the wild-type and the mutant strain were not significantly ($p = 0.91$) different (Table S3). These results suggest that *P. aeruginosa* Δ plaF responds to unbalanced membrane GPL content by a yet unidentified pathway to maintain the net GPL content at physiological levels. Taken together, quantitative lipidomics revealed that significantly accumulated GPLs in the Δ plaF strain account for > 12% of the total GPL content (Table S3), indicating the profound function of PlaF in membrane GPL homeostasis.

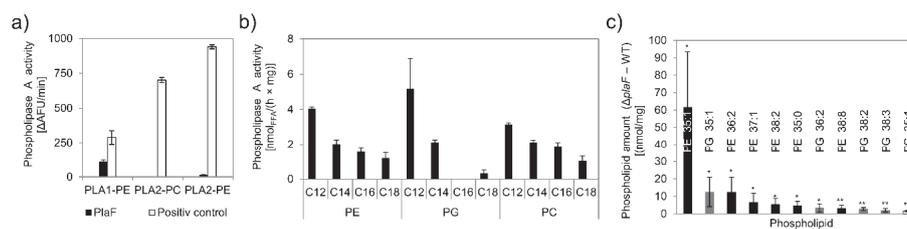


Fig. 2: Phospholipolytic activity profiling of PlaF. a) PlaF is a phospholipase A1. Enzyme activities of PlaF were measured fluorimetrically using artificial PLA1 and PLA2 substrates containing either ethanolamine (PE) or choline (PC) head groups. The control enzymes were PLA1 of *Thermomyces* 6

lanuginosus and PLA2 of *Naja mocambique*. Results are means \pm S.D. of three independent measurements performed with at least three samples. **b) PlaF releases FAs from naturally occurring bacterial GPLs.** PLA activity of PlaF was measured by quantification of released FAs after incubation of PE, PG and PC substrates containing FAs with different chain lengths (C12-C18). **c) GPL composition of *P. aeruginosa* wild-type and Δ plaF membrane differs.** Crude lipids extracted from *P. aeruginosa* wild-type and Δ plaF membranes were quantified by Q-TOF-MS/MS. For quantification, an internal standard mixture of GPLs was used. The GPL amount (nmol) was normalized to mg of crude lipids (Table S4). FA composition of GPL is depicted as XX:Y, where XX defines the number of carbon atoms and Y defines the number of double bonds in FAs bound to GPL. Results are mean \pm S.D. of four biological replicates.

PlaF is a virulence factor of *P. aeruginosa* affecting swimming motility and biofilm formation.

Bacterial membranes and GPLs regulate virulence.²⁵ Therefore, we addressed the question of whether PlaF contributes to the virulence of *P. aeruginosa*. In a comparative study using a *D. melanogaster* infection model, a remarkable difference in the survival of flies infected with *P. aeruginosa* wild-type or Δ plaF was observed. While Δ plaF was almost avirulent during 30 h of infection, the majority of the flies (~80 %) infected with the wild-type did not survive (Fig. 3a). These results suggest that PlaF is a novel virulence factor of *P. aeruginosa*. Comparison of growth of *P. aeruginosa* wild-type and Δ plaF in rich medium (Fig. S2a) did not reveal significant differences indicating that PlaF most likely does not contribute to virulence by affecting the primary metabolism of *P. aeruginosa*. As PlaF is anchored to the cytoplasmic membrane and is not in direct contact with the host cell, PlaF-mediated virulence is likely different from host membrane disruption mechanisms commonly observed for bacterial PLAs.²² The analysis of typical virulence-associated phenotypes revealed significantly impaired flagella-mediated swimming motility of *P. aeruginosa* Δ plaF compared to the wild-type strain (Fig. 3b), but no effects on swarming and twitching motilities (Figs. S2b-c). Quantification of the bacterial adherence to a plastic surface showed 40-60 % less biofilm production by *P. aeruginosa* Δ plaF compared to the wild-type (n = 4, p < 0.05) after 16 and 24 h of growth (Fig. 3c). To conclude, the results indicate that PlaF substantially contributes to the virulence properties of *P. aeruginosa* in the early phase of biofilm development, including the irreversible attachment of cells to a surface.⁴¹

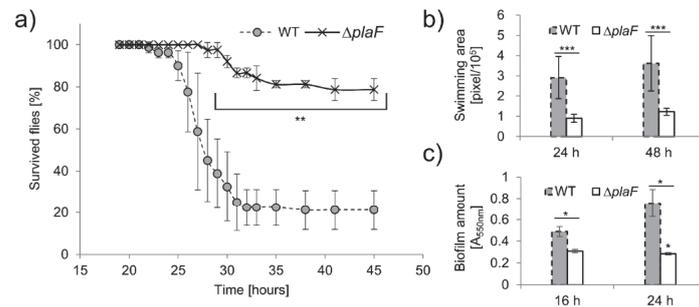


Fig. 3: PlaF is a novel virulence factor of *P. aeruginosa* PAO1. a) *P. aeruginosa* $\Delta plaF$ strain is less virulent than the respective wild-type strain in a *D. melanogaster* virulence assay. The results are mean \pm S.D. of at least three experiments with 20 flies, ** $p < 0.01$. b) The *P. aeruginosa* $\Delta plaF$ mutant strain is impaired in swimming motility. The growth of the strains on the LB agar swimming plate was quantified using ImageJ and expressed as swimming area, *** $p < 0.00005$, $n = 10$. c) *P. aeruginosa* $\Delta plaF$ produces a reduced amount of biofilm as compared to the wild-type strain. Cultures were grown in a microtiter plate and attached biofilm was quantified with the crystal violet. The results are mean \pm S.D. of three biological replicates each measured four times.

Crystal structure analysis demonstrates that PlaF forms dimers. To reveal the molecular mechanism of PlaF-mediated virulence through hydrolysis of membrane GPLs, PlaF was purified from *P. aeruginosa* *p-plaF* membranes after solubilization in OG, crystallized as described previously,³⁸ the structure was determined and refined at a resolution of up to 2.0 Å (Table 1). The final atomic model in the asymmetric unit consists of two protein molecules (PlaF_A and PlaF_B), which are related by improper 2-fold non-crystallographic symmetry (Fig. 4a). Each monomer is composed of residues 1-310 and the ligands (myristic acid (MYR), OG and isopropyl alcohol (IPA) in PlaF_A; undecyclic acid (11A), OG and IPA in PlaF_B) non-covalently bound in the active site cavity (Fig. 4a, Table S5). The N-terminal 38 amino acids comprise the putative transmembrane helix (α TM1, residues 5 – 27, all of which are hydrophobic) and the juxtamembrane helix (α JM1, residues 28 – 38). No ordered water molecules were found in the vicinity of α TM1 (Fig. 4b), which is bent at residue Y20 flanked at both sides by charged and polar residues. α JM1 extends α TM1, giving the TM-JM domain an appearance of a long kinked helix, and connects the membrane and the catalytic domain (Fig. 4a). No relevant match to the TM-JM domain was found using the Dali server (Fig. S3, Tables S6-8). Dimerization is mediated primarily *via* hydrophobic interactions of symmetry-nonrelated residues comprising the TM-JM domains of two monomers (Fig. 4c, Table S9). The TM-JM helices adopt a unique

coiled-coil-like conformation (Fig. 4d), where the α TM1 crosses its counterpart at about V14 to form an elongated X-shaped dimer interface with a large contact area of 656 Å² per monomer as determined by PISA.⁴² Superposition of PlaF_A and PlaF_B shows mainly differences in kinked conformations of TM-JM helix (Fig. S4a), likely caused by crystal packing effects where interactions of TM-JM helices from two PlaF dimers (di-PlaF) interact to form a four-helix bundle (Fig. S4b).

The catalytic domain of PlaF adopts a typical α/β -hydrolase fold,⁴³ consisting of the canonical eight-stranded β -sheet surrounded by eight α -helices (Fig. 4a). Three of these α -helices form a distinct lid-like domain, which covers the active site (Fig. 4a). Despite high homology of the catalytic PlaF domain with other lipolytic enzymes, the lid-like domain varies significantly among PlaF homologs (Fig. S5), as observed for other lipolytic enzymes previously.⁴⁴ The lid-like domain shows a less ordered structure compared to the catalytic domain, as suggested from the analysis of anisotropic B factors (Fig. S6a) using the translation-libration-screw-rotation (TLS) model (Fig. S6b). This is likely a consequence of the absence of stabilizing interactions of the charged residue-rich lid-like domain with the hydrophilic head groups of membrane phospholipids. The TLS model also revealed a less ordered structure of the TM-JM domains presumably caused due to the missing interactions with the hydrophobic membrane core (Fig. S6b).

The active site of PlaF comprises the typical serine-hydrolase catalytic triad with S137, D258 and H286 interacting through H-bonds⁴⁵ (Table S10). In both PlaF molecules, S137 shows two side-chain conformations, where one conformer is within hydrogen bond distance with the MYR and IPA in PlaF_A and PlaF_B, respectively (Fig. 4e, Tables S5 and S10), as is expected for PLA cleavage products. The presence of FA in the active sites of both monomers in the PlaF structure was unexpected, as they were not added exogenously during crystallization. These FAs are the natural products co-purified with PlaF from the homologous organism *P. aeruginosa* as confirmed by GC-MS analysis of organic solvent extracts of purified PlaF (Fig. S7). Furthermore, binding of these FAs is in line with the demonstrated specificity of PlaF for short and medium-chain FA esters.³⁸ Additionally, S137 forms H-bonds with residues I160, D161 and A163 located in the loop of the lid-like domain. Superposition of PlaF and the crystal structures of lipases from *P. aeruginosa* (PAL)⁴⁶ and *B. cepacia* (BCL)⁴⁷, both with triacylglycerol substrate analogs covalently bound to the active site, reveals a conserved α/β -hydrolase domain, catalytic triad and oxyanion hole, which is formed by F71 and M138 in PlaF

(Fig. S8). Two water molecules at H-bond distances to the backbone nitrogen atoms of F71 and M138 and the side-chain oxygen atom of S137 are likely involved in the catalytic cycle (Fig. 4e, Table S10).

The active site cleft in PlaF is formed by residues from the helix α M1, the α/β -hydrolase and the lid-like domains (Fig. 4e, Table S11). In contrast to the structures of other lipolytic enzymes, the cleft in PlaF has a T-shape and three openings. The pyranose ring of the OG molecule is located on the opening, between two monomers, made by α M1 and the loop preceding α F of the lid-like domain (Fig. 4d). A second opening lined with residues from the loops preceding α E and α F is located on the opposite end of the cleft, close to the catalytic S137. In the middle of the T-shaped cleft, likely pointing towards the membrane is a third opening lined mostly with polar residues of the loops preceding α B and α F. The amphiphilic and large T-shaped active site cleft is compatible with binding amphiphilic, bulky GPL substrates. The alkyl chains of OG and MYR bound in the active site cleft are stabilized *via* hydrophobic interactions (Fig. 4e). The pyranose ring of OG interacts, among others, with the residue V33 of PlaF_A, which, in turn, is involved in dimerization through interactions with V33 and T32 of PlaF_B (Fig. 4e). The H-bond interaction of the catalytic S137 with the carboxyl group of MYR completes an intricate ligand-mediated interaction network that bridges the catalytic (S137) and dimerization (V33) sites in PlaF (Fig. 4e). In summary, the crystal structure suggested a role of dimerization and FAs for the regulation of the PlaF function, which was subsequently analyzed biochemically.

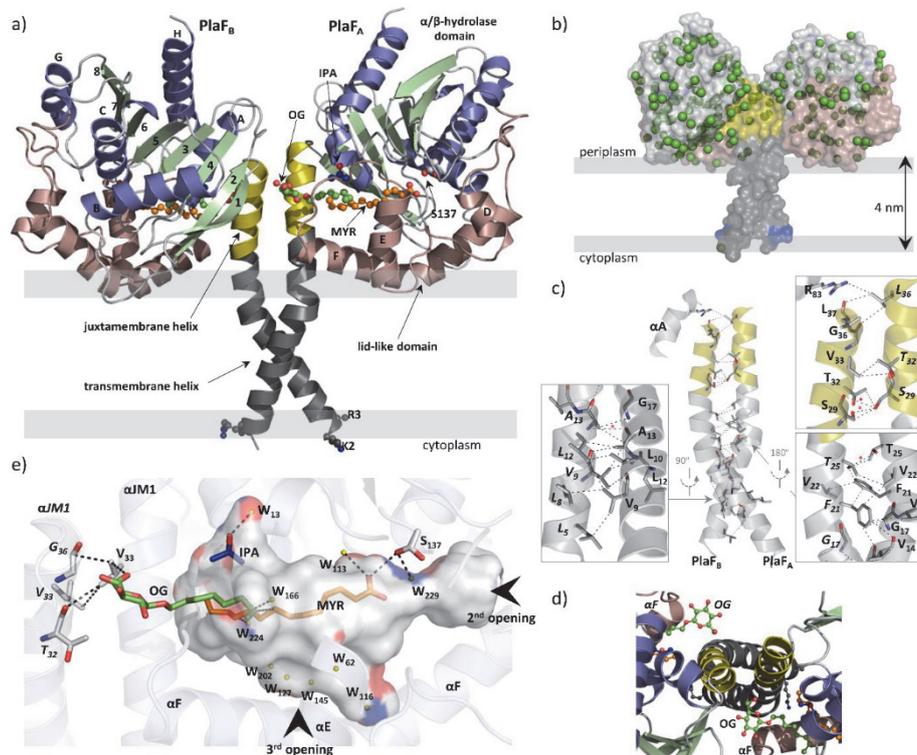


Fig. 4: Overall structure of dimeric PlaF with bound endogenous FA ligands. **a)** A unique N-terminal helix comprising a putative transmembrane helix (α TM1, grey) flanked by charged residues (K2, R3) on one side (R3 side chain is not completely resolved) and, on the another side, the juxtamembrane helix (α JM1, yellow) which links the α TM1 with the periplasmic catalytic domain which consists of an α/β -hydrolase (colored as follows: blue, α -helices; green, β -strands and grey, loops) and a lid-like domain (brown). Ligands bound in the active site cleft of PlaF are represented as ball-and-sticks and colored as follows: oxygen in red; carbon of OG, MYR and IPA in green, orange and blue, respectively. Thick grey lines roughly depict the membrane. **b)** A model for PlaF orientation in the membrane as suggested from the distribution of charged/polar (K2, R3, S29, R31, T32, E34) and hydrophobic residues and the water molecules (green spheres). The transparent surface of PlaF was colored as in Fig. 4a. PlaF is rotated by 180° along the normal to the membrane. **c)** The interaction interface between two PlaF molecules involving TM-JM helices is predominantly hydrophobic with four weak H-bonds (indicated by a red asterisk) were detected mostly in the JM helix (S29) and close to it (T25). The only interaction with residues outside of the JM-TM helix involves R₈₃ from α A. Residues of the PlaF_B molecule are indicated in italics. A detailed list of interactions is provided in table S9. **d)** Enlarged section of PlaF viewed from the periplasmic side indicating the coiled-coil organization of the TM-JM helix and OG

ligands placed between JM and the loop preceding α F in both PlaF monomers. Elements of the PlaF_B molecule are indicated in italics. **e)** Interaction network within T-shaped ligand-binding cleft of PlaF_A. MYR is linked *via* H-bond with the catalytic S137 and *via* hydrophobic interactions with OG. The sugar moiety of OG from PlaF_A forms H-bonds with V33 of PlaF_A which is interacting with V33 and G36 of PlaF_B. The part of the cleft in direction of the opening 2 is occupied by several water molecules (W, yellow spheres). The cleft accommodates one IPA molecule bound to the water. Arrows indicate the openings one and two not visible in this orientation. The cleft was calculated using the Pymol software (<http://www.pymol.org>) and colored by elements: carbon, gray; oxygen, red; nitrogen, blue.

The monomer-dimer ratio of PlaF affects enzyme activity. Purified PlaF elutes as a monomer during size exclusion chromatography (Fig. S9). Yet, the crystal structure and the PISA analysis⁴² performed on it reveal a dimer. To investigate the oligomeric state of PlaF *in vivo*, we performed cross-linking experiments using the bi-functional cross-linking reagent dimethyl pimelimidate (DMP), which was added to intact *P. aeruginosa* p-*plaF* cells. As expected, Western blot results revealed monomeric PlaF in all samples. PlaF dimers migrating at ~70 kDa and a faint band corresponding to higher apparent molecular weight of ~120 kDa (PlaF*) were detected in DMP treated cells (Figs. 5a and S10a). The latter is likely due to a cross-linking product between PlaF and an unknown protein interaction partner. *In vitro* time-resolved cross-linking experiments by incubating purified PlaF with DMP revealed spontaneous dimerization (Figs. 5b and S10b). The presence of PlaF and di-PlaF was confirmed by MALDI mass spectrometric fingerprint analysis of the respective protein bands after SDS-PAGE and trypsin digestion (data not shown). To quantify spontaneous PlaF dimerization, we performed microscale thermophoresis (MST) measurements with purified PlaF solubilized with OG. Titration of the fluorescence-labeled PlaF with an equimolar concentration of non-labeled PlaF revealed a sigmoidal binding curve, with a dissociation constant $K_D = 637.9 \pm 109.4$ nM (Fig. 5c). Next, we measured the esterase activity of PlaF as a function of protein concentration using the slowly hydrolyzable substrate 4-methylumbelliferyl palmitate, which allows reliable activity measurements even at high PlaF concentrations where dimerization occurs. The specific activity of PlaF at concentrations above 100 μ M dropped to 4 % (Fig. 5c) compared to the low-concentrated sample (30 nM) containing presumably monomeric PlaF. The activation constant $K_{act} = 916.9 \pm 72.4$ nM is similar to the dissociation constant of di-PlaF. These data strongly support the regulation of PlaF activity through reversible dimerization *in vitro* and likely also in *P. aeruginosa* cells.

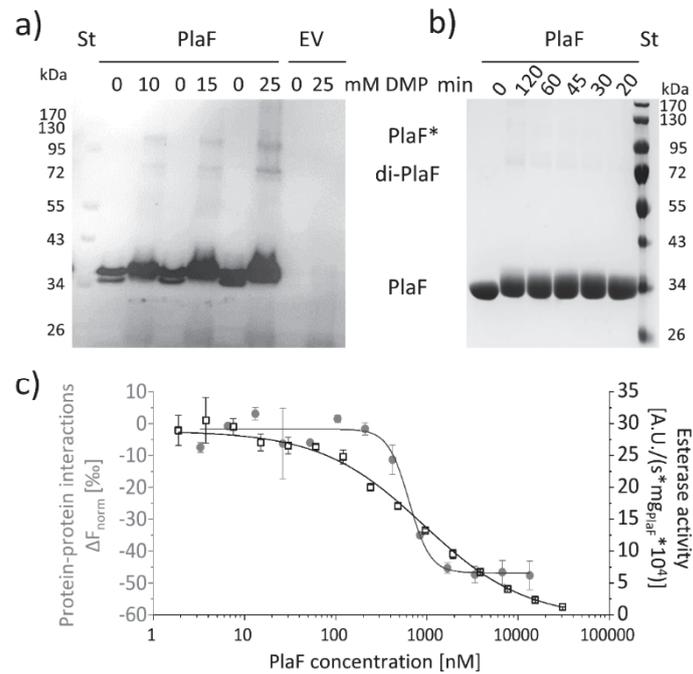


Fig. 5: PlaF oligomeric states and their enzymatic activity. **a)** PlaF forms dimers in cell membranes. *In vivo* cross-linking experiments were performed by incubating *P. aeruginosa p-plaF* or the empty vector control (EV) cells with different concentrations of DMP cross-linker followed by immunodetection of PlaF with anti-PlaF antiserum. **b)** *In vitro* cross-linking of purified PlaF. Purified PlaF was incubated for up to 120 min with DMP and the samples were analyzed by SDS-PAGE. Molecular weights of protein standard in kDa are indicated. **c)** PlaF homodimerization and activity are concentration-dependent. Protein-protein interactions of purified PlaF were monitored by measuring the changes in thermophoresis (ΔF_{norm} , grey circles) using the MST method. The MST results are mean \pm S.D. of two independent experiments with PlaF purified with OG. Esterase activity (black squares) of PlaF was measured in three independent experiments using 4-methylumbelliferyl palmitate substrate. Dissociation (K_D) and activation (K_{act}) constants were calculated using a logistic fit of sigmoidal curves.

Fatty acid-induced PlaF dimerization regulates PlaF activity. Inhibition assays with several FAs of different chain lengths revealed strong inhibition of PlaF activity with FAs with 10 to 14 carbon atoms and at mM concentrations (Fig. 6a). Shorter and longer FAs showed only moderate or weak inhibition (Fig. 6a). To explore the underlying mechanism of regulation, we performed inhibition kinetics studies with decanoic acid as the inhibitor and *p*-NPB as the

substrate. The results showed typical properties of mixed-inhibition kinetics with lowered maximal hydrolysis rates (v_{max}) and elevated binding constants (K_m) in the presence of increasing concentrations of decanoic acid (Fig. 6b, Table S12). In section 3.4, we showed that the enzyme activity of di-PlaF is strongly reduced compared to monomeric PlaF. To test whether FA inhibition is related to dimerization, we performed DMP crosslinking experiments with PlaF in the presence of C10, C11 and C12 FAs. The results revealed an increasing amount of di-PlaF in FA treated samples (Fig. 6c). This effect of FAs on the dimerization and the Michaelis-Menten constants suggest a potential regulatory role of FAs on PlaF activity *via* FA-induced dimerization.

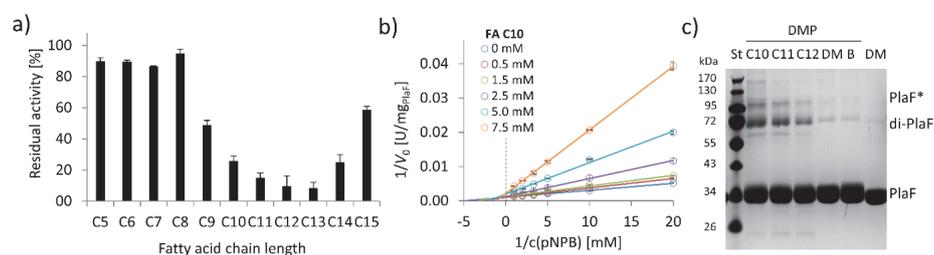


Fig. 6: FAs exert an inhibitory effect on PlaF and trigger dimerization. a) Inhibition of PlaF with FAs. Esterase activity of PlaF was measured in the presence of 7.5 mM FA (C5 – C15); a untreated PlaF sample was set as 100 %. The results are mean \pm S.D. of three experiments with three samples each. **b) Kinetic studies with FA C10 show evidence of mixed-inhibition.** Double-reciprocal plots of initial reaction velocities measured with the *p*-NPB substrate and FA C10 inhibitor at concentrations in a range 0-7.5 mM. **c) The effect of FAs on PlaF dimerization.** PlaF samples incubated with FAs (C10, C11, C12), dimethyl sulfoxide (DM, DMSO used to dissolve FAs) and purification buffer (B, dilution control) were cross-linked with dimethyl pimelimidate (DMP).

The tilt of the monomeric PlaF in a lipid bilayer permits direct GPL access to the active site.

To examine possible structural changes associated with the dimer-monomer transition in relation to the experimentally suggested activation of PlaF through monomerization, we performed MD simulations on PlaF. First, we selected PlaF_A by removing PlaF_B from the crystal structure of di-PlaF; by the same procedure, PlaF_B was generated. The spatial orientation of both PlaF_A and PlaF_B in the membrane were predicted with the PPM method.⁴⁸ This resulted in configuration with the dimerization interface tilted towards the membrane, which we named t-PlaF_A and t-PlaF_B, respectively (Fig. 7a). A set of ten independent, unbiased 2 μ s long

MD simulations starting from either di-PlaF, PlaF_A, PlaF_B, t-PlaF_A or t-PlaF_B embedded in an explicit membrane revealed only moderate intramolecular changes compared to the initial structure (RMSD_{all atom} < 4.0 Å), as well as among each other (RMSD_{all atom} < 4.3 Å) (Table S13, Fig S12a). No spontaneous monomerization of di-PlaF was observed during the MD simulations (Fig. 7b), which is expected due to the high viscosity of the lipid bilayer. In most MD simulations of the monomeric PlaF starting from a dimeric orientation (eight for PlaF_A and six for PlaF_B), a transition to the tilted configuration was observed, while in all MD simulations starting from either t-PlaF_A or t-PlaF_B the protein remained in the tilted orientation (Figs. 7c and S12b). These results suggest that the tilted PlaF orientation is preferred over the respective configurations of monomers observed in the crystal structure. Strikingly, the tilting orients the active site cleft perpendicular to the membrane plane such that GPL substrates can have direct access to the cleft. In contrast, in the cases of dimeric and monomeric configurations observed in the crystal structure, the residues lining the opening of the active site cleft located between the monomers in the crystal structure are > 5 Å above the membrane surface (Fig. 7a). Diffusion of a GPL from the membrane bilayer to the cleft entrance in this configuration is essentially thermodynamically unfavorable.

As no spontaneous separation of the di-PlaF was observed, we explored the energetics of monomerization and tilting of the monomer by computing the PMF for the dimer-to-monomer transition and for the tilting of monomeric PlaF from PlaF_A to t-PlaF_A, using umbrella sampling and post-processing with the WHAM method. As reaction coordinates, we chose the distance (r) between C_α atoms of residues 25-38 of the two molecules (in the crystal structure: $r = 9.9$ Å) and the distance (d) between C_α atoms of residues 33-37 and the center of the membrane slab ($d(\text{PlaF}_A) \approx 37$ Å and $d(\text{t-PlaF}_A) \approx 15$ Å). The converged (Fig. S12c, left) and precise (SEM < 1.4 kcal mol⁻¹) PMF revealed that di-PlaF at the global minimum with $r = 9.5$ Å is favored by -11.4 kcal mol⁻¹ over the monomer (Fig. 7d). The weakening of intermolecular interactions are observed from $r \approx 15$ Å onwards, disappearing beyond $r \approx 27$ Å. For the tilting process, the converged and precise (Fig. S12c, right; SEM < 0.4 kcal mol⁻¹) PMF revealed two minima at $d = 19.6$ and 30.6 Å, with t-PlaF_A favored over PlaF_A by 0.66 kcal mol⁻¹ (Fig. 7e). The free energy barrier of ~1.2 kcal mol⁻¹ explains the rapid transition from PlaF_A to t-PlaF_A observed in the unbiased MD simulations.

The equilibrium constants and free energies of PlaF dimerization (eqs. 1-3; $K_a = 1.57 \times 10^7$ Å², $K_x = 2.58 \times 10^5$; $\Delta G = -7.5 \pm 0.7$ kcal mol⁻¹) and tilting (eq. 5,6; $K_{\text{tilting}} = 3.35$,

$\Delta G_{\text{tilting}} = -0.8 \pm 0.2 \text{ kcal mol}^{-1}$) were computed, taking into consideration that K_x and ΔG relate to a state of one PlaF dimer in a membrane of 764 lipids, according to our simulation setup. This ratio is presumably higher by a factor of 100-1000 than for *P. aeruginosa* under non-overproducing, physiological conditions (see SI). Under such physiological conditions and considering that the equilibria for dimer-to-monomer transition and titling are coupled (Fig. 7f and eq. 11), between 86 and 98 % (for an overproduction factor of 100 to 1000) of the PlaF molecules are predicted to be in a monomeric, tilted state in *P. aeruginosa*. This result is in agreement with cross-linking results (Fig. 5a) showing that the equilibrium between monomeric and dimeric PlaF in *P. aeruginosa* cells is shifted towards the monomer as well as that *in vitro* at low concentrations PlaF exists predominantly as the monomer (Fig. S9). Taken together, the molecular simulations suggest that monomeric t-PlaF is favored under physiological conditions in *P. aeruginosa*. This configuration might allow direct access of GPL substrates to the active site.

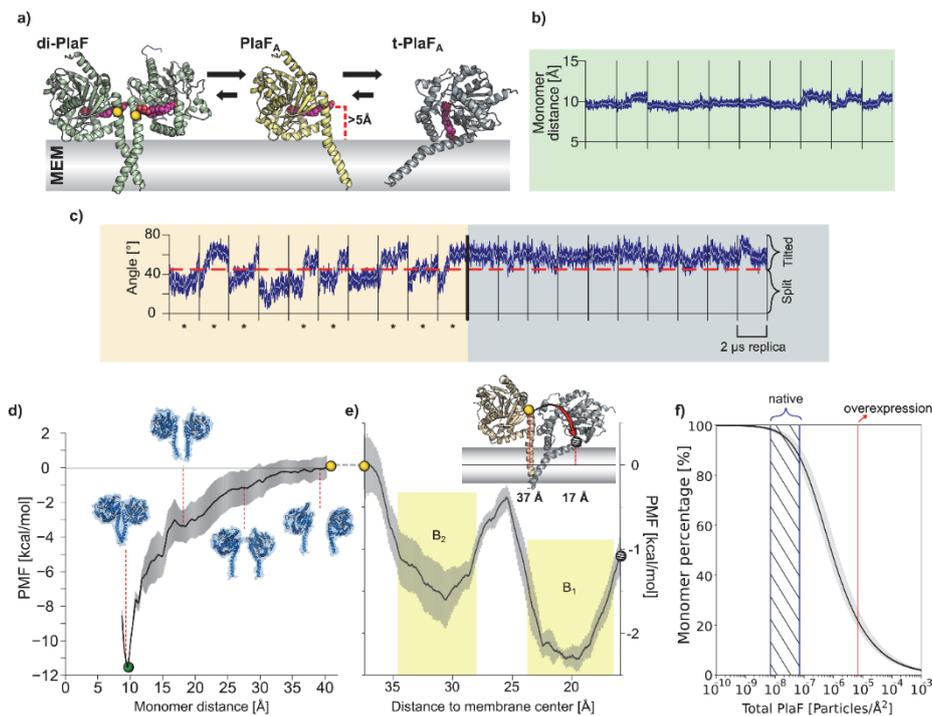


Fig. 7: MD simulations and PMF computations for PlaF dimer and monomers. a) Structures used for MD simulations. di-PlaF: Crystal structure oriented in the membrane by the PPM method. Yellow

spheres highlight the COM used to define the distance restraint for umbrella sampling simulations. **PlaF_A**: Chain A from PlaF dimer oriented as in the dimer. The entrance of the active site cleft is more than 5 Å above the membrane bilayer surface. **t-PlaF_A**: Extracted monomer A was oriented using the PPM method. CocrySTALLIZED MYR, 11A and OG (not included in the simulations) are depicted in pink to highlight the orientation of the active site cleft. Arrows between the structures reflect the predicted states of equilibria under physiological conditions in *P. aeruginosa*. **b) Molecular dynamics simulations of di-PlaF**. The dimer does not show a tendency to separate on the time scale of the MD simulations according to the distance between the COM of C_α atoms of residues 25 to 38 of each monomer (yellow spheres in panel a). **c) Molecular dynamics simulations of monomeric PlaF**. Time course of the orientation of monomeric PlaF with respect to the membrane starting from the PlaF_A configuration (left) or t-PlaF_A configuration (right). In the former case, in 8 out of 10 trajectories the monomer ends in a tilted configuration (marked with *), while in the latter, no rearrangement occurs. This shows a significant tendency of the monomer to tilt, irrespective of the starting condition (McNemar's $\chi^2 = 6.125$, $p = 0.013$). **d) Potential of mean force of dimer separation**. The distance between the COM of C_α atoms of residues 25 to 38 of each chain (yellow spheres in panel a) was used as the reaction coordinate. The shaded area shows the standard error of the mean obtained by dividing the data into four independent parts of 50 ns each. Insets show representative structures at intermediate reaction coordinate values. **e) Potential of mean force of monomer tilting**. The distance between the COM of C_α atoms of residues 33 to 37 (yellow and grey spheres) and the COM of the C₁₈ of the oleic acid moieties of all lipids in the membrane (continuous horizontal line in the membrane slab) was used as a reaction coordinate. The gray shaded area shows the S.D. of the mean obtained by dividing the data into four independent parts of 50 ns each. The yellow shaded regions correspond to the integration limits used to calculate K_{tilting} (eq. 5). The spheres in the PMF relate to monomer configurations shown in the inset. **f) Percentage of PlaF monomer as a function of total PlaF concentration in the membrane according to the equilibria shown in Fig. 7a**. The monomer percentage was computed according to eqs. 7-11. The red line shows experimentally determined PlaF concentration under overexpressing conditions in *P. aeruginosa p-plaF*, while the blue-dashed region shows an estimated span for the PlaF concentration in *P. aeruginosa* wild-type (see methods for details).

Discussion

Cells continuously adapt the phospholipid composition of their membranes to changing functional requirements. Thus, it is plausible to assume that phospholipases are involved in this process; however, the role of phospholipase A-mediated membrane GPL degradation for membrane homeostasis in bacteria is unknown. In this study, we presented the first high-resolution structure of a homodimer of a single-span membrane protein in full-length, where

protomer association occurs *via* the TM and JM regions. We used this structure to rationalize how a dimer-to-monomer transition is associated with PlaF switching from an enzymatically inactive to an active state. These results provide the basis for understanding how PlaF acts as a virulence factor of *P. aeruginosa* by releasing FA from the *sn*1 position in membrane GPLs, thereby affecting swimming motility and biofilm formation. Interestingly, recent lipidomic studies revealed that the transition of *P. aeruginosa* from a planktonic to a sessile lifestyle upon biofilm formation is accompanied by the alteration of membrane GPL composition.^{17,49} Furthermore, changes in membrane GPLs were observed in a *P. aeruginosa* small colony variant that shows low virulence and enhanced biofilm phenotypes.⁵⁰ Such a regulation of bacterial virulence by tuning membrane GPL composition should rely on both pathways for the biosynthesis and controlled degradation of GPLs.

Enzymatic properties of PlaF (Fig. 2b) suggest that it functions in the GPL-degradation pathway. This hypothesis is supported by cellular localization results that demonstrated that PlaF is tethered to the cytoplasmic membrane (Fig. 1), that way being in direct contact with putative GPL substrates. Despite the astounding diversity of lipids in membranes⁵¹, a global lipidomic analysis identified a small subset of sixteen GPLs accumulating in *P. aeruginosa* Δ *plaF* (Fig. 2c, Table S2) that are presumably native PlaF substrates. Among them were PG, PC and PE predominantly containing unsaturated FAs (15 of 16 GPLs). This result is concordant with previous data showing that PlaF does not show specificity for GPL head groups (Fig. 2b) and can hydrolyze unsaturated lipids.³⁸ PG, PC and PE 35:1 account for > 65 % of the accumulated GPLs highlighting these molecular species as main PlaF substrates. These results reveal an unprecedented *in vivo* specificity of a GPL-degrading enzyme. Interestingly, changes in the concentration of these three GPL species result in a large change (~8 %) of total membrane GPL composition. Thus, PlaF rather specifically tunes the membrane GPL composition thereby exerting a profound effect on membrane properties, with potential consequences for functional protein-lipid interactions.⁵¹

The observations that the net GPL amount was not significantly different (Table S4) between *P. aeruginosa* wild-type and Δ *plaF* cells and that nineteen GPL species were depleted in Δ *plaF* suggest that a PlaF-mediated GPL modulation is coupled to GPL biosynthesis or another still unknown GPL degrading pathway of *P. aeruginosa*. Although the cellular pathway in which PlaF is active is unknown, it is tempting to speculate that lysophospholipids produced by PlaF-

mediated release of FAs from GPLs will be acylated to di-acyl GPLs by the *de novo* GPL biosynthetic pathway^{9,52} or the recently described GPL repair pathway⁵³.

The unique crystal structure of PlaF revealed several features that underlined the role of PlaF for degradation of membrane GPLs. First, FA ligands identified in the crystal structure (C14 and C11) and bound to the active site (Fig. 4a) showed inhibitory activity on PlaF (Fig. 6a). These FAs were also copurified with PlaF from *P. aeruginosa* (Fig. S7). As PlaF can *in vitro* (Fig. 2b) and *in vivo* (Fig. 2c) produce such FAs, it is reasonable to assume that their cellular function is related to the regulation of PlaF activity by product-feedback inhibition. This phenomenon is well known for lipolytic^{54,55} and other central metabolic enzymes^{56,57}. A recent global analysis of metabolic enzyme inhibition by metabolites revealed that lipids account for most inhibitions.⁵⁸ An additional indication for a negative feedback regulation by FA products is the finding that bacteria intracellularly accumulate FAs up to the mM range,⁵⁹ which is the same range in which inhibitory activity on PlaF was observed (0.5 – 7.5 mM) (Fig. 6b and Table S12). Inhibition kinetics experiments revealed an alteration of the substrate-binding constant (K_m) by FAs, indicating that substrate (GPL) and product (FA) both compete for the active site. Yet, the same experiments showed that the maximal velocity of PlaF reaction (v_{max}) was lower. This indicates that FAs also allosterically alter the catalytic properties of PlaF by changing either its structure or oligomerization state. The crystal structure of the PlaF:FA complex revealed a dimer, prompting us to test the relationship between dimerization and FA inhibition. Indeed, *in vitro* crosslinking of PlaF revealed that the presence of FA enhances PlaF dimerization (Fig. 6c), suggesting that the activity of dimeric PlaF is reduced compared to monomeric PlaF. Concentration-dependent enzymatic measurements of PlaF activity and MST analysis of homodimerization revealed a coupled sigmoidal decrease of activity and an increase of dimerization as a function of increasing PlaF concentration (Fig. 6c). The finding that the activation ($K_{act} = 916.9$ nM) and dimerization ($K_D = 637.9$ nM) constants are similar strengthens our hypothesis that PlaF inhibition is dimerization-mediated.

Specific homo- and heterodimerization is commonly underlying control of protein function.⁶⁰ Among single TM spanning proteins, this concept is poorly understood at an atomistic level due to the lack of full-length dimeric structures.⁶¹ The structure of di-PlaF with its membrane-spanning domain is unique and without homology to any protein listed in the PDB (Table S6). This structure revealed unprecedented details of interactions between the membrane-spanning domains. First, TM and JM domains are not distinct but form a single kinked α -helix

(Fig. 4a). Sole interactions of two TM-JM helices, which fold into a coiled-coil motif, stabilize the PlaF dimer (Fig. 4c). In contrast, the catalytic domains of two PlaF monomers do not interact with each other. di-PlaF was also detected by crosslinking experiments on *P. aeruginosa* p-*plaF* cells (Fig. 5a). The predominantly hydrophobic interaction interface of TM-JM domains with a few weak H-bonds detected in the JM region (Fig. 4c) seems to be designed to specifically binds amphipathic lipid effectors. Ligands (FAs and OG) bound to the PlaF active site cleft form an intricate interaction network connecting the catalytic (S137) with the dimerization site (S29, T32 and V33) located in the JM domain (Fig. 4e). Assuming that the OG molecule is replaced by a natural ligand in the cell, this interaction network provides a possible explanation how FAs can mediate dimerization between the two TM-JM helices (Fig. 6c), thereby inhibiting PlaF activity (Fig. 6b). Furthermore, it is feasible that interactions of lipids from the hydrophobic membrane core with the TM helix fine-tune PlaF activity by counteracting or enhancing the TM-JM dimerization.^{60,61}

How does a PlaF dimer-to-monomer transition then lead to enzymatic activity? The active sites in the crystal structure of di-PlaF already adopt catalytically active conformations (Fig. 4a). Hence, activation of PlaF most likely does not involve structural rearrangements of the active site. To unravel a possible effect of the structural dynamics of PlaF in the membrane on the enzyme's regulation by dimerization, we performed extensive MD simulations and configurational free energy computations on di-PlaF and monomeric PlaF embedded into a GPL bilayer mimicking the bacterial cytoplasmic membrane. While structural changes within di-PlaF and monomeric PlaF were moderate (Table S13), monomeric PlaF spontaneously tilted as a whole towards the membrane, constraining the PlaF in a configuration with the opening of the active site cleft immersed into the GPL bilayer (t-PlaF, Figs. 7a and 7c). A configuration similar to t-PlaF was observed for monomeric *Saccharomyces cerevisiae* lanosterol 14 α -demethylase, a single TM spanning protein acting on a membrane-bound substrate.⁶² In t-PlaF, GPL can likely access the active site cleft directly from the membrane, in contrast to di-PlaF, in which the opening of the active site cleft is > 5 Å above the membrane (Fig. 7a). There, a GPL would need to transition from the bilayer to the water milieu prior to entering the active site cleft, which is thermodynamically unfavorable.

Taking together our experimental evidence, we therefore propose a hitherto undescribed mechanism, by which the transition of PlaF between a dimeric, not-tilted *versus* a monomeric, tilted configuration is intimately linked to the modulation of PlaF's functional state. This

mechanism, to the best of our knowledge, expands the general understanding of mechanisms of inactivation of integral single-pass TM proteins and differs from suggested allosteric mechanisms implying structural rearrangements (even folding), mostly in the JM domain, upon ligand binding as underlying causes for functional regulation.⁶¹ Rather, for PlaF, monomerization followed by a global reorientation of the single-pass TM protein in the membrane is the central, function-determining element.

Based on computed free energies of association and tilting, and taking into account the concentration range of PlaF in *P. aeruginosa*, PlaF preferentially exists as t-PlaF in the cytoplasmic membrane (Fig. 7f). Increasing the PlaF concentration in the membrane will thus shift the equilibrium towards the di-PlaF. This result can provide an explanation for the observations that PlaF, an enzyme with membrane-disruptive activity, is found in only very low amounts (Fig. S11) in wild type *P. aeruginosa* cells and why overproduction of PlaF in *P. aeruginosa* is not harmful to the cells.

In summary, our biochemical, structural and computational data uncover a unique structural feature of PlaF in that its function is dependent on monomerization followed by global reorientation of the single-pass TM protein at the membrane. The dimer-to-monomer transition is furthermore inhibited by FAs, the products of PlaF activity on membrane GPLs. We suggest that PlaF-mediated alteration of membrane GPLs might regulate the function of cytoplasmic membrane-embedded and virulence-related proteins in *P. aeruginosa*, such as two-component signaling systems indispensable for biofilm and swimming motility²¹. These results open up novel avenues for the development of potential drugs to extenuate *P. aeruginosa* virulence during infections through inhibition of PlaF.

Acknowledgments

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) CRC 1208 grant to FK, KEJ and HG. We are grateful to the beamline scientists at the European Synchrotron Radiation Facility (Grenoble, France) for assisting with the use of beamline ID29. We thank R. Voulhoux (CNRS AMU LCB, Marseille) for providing anti-XcpQ antiserum; P. Dollinger (HHU Düsseldorf) for help with MST measurements; R. Molitor (HHU Düsseldorf) for help with bacterial motility assays; M. Dick (HHU Düsseldorf) for assistance in setting up biased MD simulations. We are grateful for computational support by the “Zentrum für Informations und Medientechnologie” at the Heinrich-Heine-Universität Düsseldorf and the computing time provided by the John von Neumann Institute for Computing (NIC) to HG on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC) (user IDs: HKF7; HDD18).

Author contributions

F.K. conceptualization, supervision, analysis, visualization, writing; K.-E.J. supervision, writing; H.G. conceptualization, supervision, analysis, writing; F.B. and M.C. investigation, analysis, visualization; S.N.S.V. investigation, analysis, visualization, writing; B.T. analysis, visualization; P.D. and M.S., analysis; R.B.S. conceptualization, supervision, analysis, visualization, writing; and J.G. investigation, analysis, visualization, writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

Tables

Table 1: Data collection and refinement statistics on PlaF

X ray-data

Beamline/Detector	ID29, ESRF (Grenoble, France)/DECTRIS PILATUS 6M
Wavelength (Å)/Monochromator	$\lambda=0.96863$ /channel-cut silicon monochromator, Si(111)
Resolution range (Å)	47.33 - 2.0 (2.05 - 2.0)**
Space group	I 4 ₁ 2 2
Unit cell (a=b), c (Å); $\alpha=\beta=\gamma$	a=133.87 c=212.36; 90°
Total reflections	669964 (47385)
Unique reflections	65113 (4527)
Multiplicity	10.3 (10.5)
Completeness (%)	100.0 (100.0)
Mean I/sigma (I)	24.6 (2.5)
Wilson B-factor (Å ²)	38.3
R-merge %	5.3 (91.3)
R-meas %	5.6 (100.6)

Refinement

R-work %	16.3 (23.15)(2.071 - 2.0)**
R-free %	18.57 (27.81)
Number of atoms	5187
macromolecules	4831
ligands	123
water	233
Protein residues	620
RMS (bonds)	0.008
RMS (angles)	1.07
Ramachandran favored (%)	99
Ramachandran outliers (%)	0
Clashscore	3.14
Average B-factor (Å ²)	49.1
macromolecules (Å ²)	48.8
ligands (Å ²)	79.2
solvent (Å ²)	47.9

**Values in parentheses are for the highest resolution shell

References

1. Eickhoff, M.J. & Bassler, B.L. SnapShot: bacterial quorum sensing. *Cell* **174**, 1328-1328 (2018).
2. Parsons, J.B. & Rock, C.O. Bacterial lipids: metabolism and membrane homeostasis. *Prog. Lipid Res.* **52**, 249-276 (2013).
3. Higgins, C.F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8**, 67-113 (1992).
4. Venturi, V. & Fuqua, C. Chemical signaling between plants and plant-pathogenic bacteria. *Annu. Rev. Phytopathol.* **51**, 17-37 (2013).
5. Krampen, L. et al. Revealing the mechanisms of membrane protein export by virulence-associated bacterial secretion systems. *Nat. Commun.* **9**, 3467 (2018).
6. Mackenzie, K.R. Folding and stability of alpha-helical integral membrane proteins. *Chem. Rev.* **106**, 1931-1977 (2006).
7. Baxter, A.A., Hulett, M.D. & Poon, I.K. The phospholipid code: a key component of dying cell recognition, tumor progression and host-microbe interactions. *Cell Death. Differ.* **22**, 1893-1905 (2015).
8. Garcia-Fernandez, E. et al. Membrane microdomain disassembly inhibits MRSA antibiotic resistance. *Cell* **171**, 1354-1367 (2017).
9. Zhang, Y.M. & Rock, C.O. Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* **6**, 222-233 (2008).
10. Rowlett, V.W. et al. Impact of membrane phospholipid alterations in *Escherichia coli* on cellular function and bacterial stress adaptation. *J. Bacteriol.* **199**, e00849-16 (2017).
11. Schniederjans, M., Koska, M. & Haussler, S. Transcriptional and mutational profiling of an aminoglycoside-resistant *Pseudomonas aeruginosa* small-colony variant. *Antimicrob. Agents Chemother.* **61**, e01178-17 (2017).
12. Sinensky, M. Homeoviscous adaptation - a homeostatic process that regulates the viscosity of membrane lipids in *Escherichia coli*. *Proc. Natl. Acad. Sci. U S A* **71**, 522-525 (1974).
13. Cossins, A.R. Temperature adaptation of biological membranes: proceedings of the meeting held in Cambridge under the auspices of the Society for Experimental Biology in conjunction with its US/Canadian counterparts, NC: Portland Press, **14**, 227 (1994).
14. Marr, A.G. & Ingraham, J.L. Effect of temperature on the composition of fatty acids in *Escherichia coli*. *J. Bacteriol.* **84**, 1260-1267 (1962).
15. Wang, A.Y. & Cronan, J.E., Jr. The growth phase-dependent synthesis of cyclopropane fatty acids in *Escherichia coli* is the result of an RpoS(KatF)-dependent promoter plus enzyme instability. *Mol. Microbiol.* **11**, 1009-17 (1994).
16. Fozo, E.M. & Rucks, E.A. The making and taking of lipids: the role of bacterial lipid synthesis and the harnessing of host lipids in bacterial pathogenesis. in *Adv. Microb. Physiol.*, Vol. Volume 69 (ed. Robert, K.P.) 51-155 (Academic Press, 2016).
17. Benamara, H. et al. Characterization of membrane lipidome changes in *Pseudomonas aeruginosa* during biofilm growth on glass wool. *PLoS One* **9**, e108478 (2014).
18. El Khoury, M. et al. Targeting bacterial cardiolipin enriched microdomains: an antimicrobial strategy used by amphiphilic aminoglycoside antibiotics. *Sci. Rep.* **7**, 10697 (2017).
19. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet. Infect. Dis.* **18**, 318-327 (2018).
20. Van Delden, C. & Iglewski, B.H. Cell-to-cell signaling and *Pseudomonas aeruginosa* infections. *Emerg. Infect. Dis.* **4**, 551-560 (1998).
21. Moradali, M.F., Ghods, S. & Rehm, B.H. *Pseudomonas aeruginosa* lifestyle: a paradigm for adaptation, survival, and persistence. *Front. Cell Infect. Microbiol.* **7**, 39 (2017).
22. Cota-Gomez, A. et al. PlcR1 and PlcR2 are putative calcium-binding proteins required for secretion of the hemolytic phospholipase C of *Pseudomonas aeruginosa*. *Infect. Immun.* **65**, 2904-2913 (1997).
23. Adam, P.B. et al. A novel extracellular phospholipase C of *Pseudomonas aeruginosa* is required for phospholipid chemotaxis. *Mol. Microbiol.* **53**, 1089-1098 (2004).

24. Russell, A.B. et al. Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**, 508-512 (2013).
25. Flores-Diaz, M., Monturiol-Gross, L., Naylor, C., Alape-Giron, A. & Flieger, A. Bacterial sphingomyelinases and phospholipases as virulence factors. *Microbiol. Mol. Biol. Rev.* **80**, 597-628 (2016).
26. Schmiel, D.H. & Miller, V.L. Bacterial phospholipases and pathogenesis. *Microbes. Infect.* **1**, 1103-1112 (1999).
27. Istivan, T.S. & Coloe, P.J. Phospholipase A in Gram-negative bacteria and its role in pathogenesis. *Microbiology* **152**, 1263-1274 (2006).
28. Lang, C. & Flieger, A. Characterisation of *Legionella pneumophila* phospholipases and their impact on host cells. *Eur. J. Cell. Biol.* **90**, 903-912 (2011).
29. Schunder, E. et al. Phospholipase PlaB is a new virulence factor of *Legionella pneumophila*. *Int. J. Med. Microbiol.* **300**, 313-323 (2010).
30. Wilhelm, S., Gdynia, A., Tielen, P., Rosenau, F. & Jaeger, K.E. The autotransporter esterase EstA of *Pseudomonas aeruginosa* is required for rhamnolipid production, cell motility, and biofilm formation. *J. Bacteriol.* **189**, 6695-6703 (2007).
31. Kovacic, F. et al. Structural and functional characterisation of TesA - a novel lysophospholipase A from *Pseudomonas aeruginosa*. *PLoS One* **8**, e69125 (2013).
32. Salacha, R. et al. The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environ. Microbiol.* **12**, 1498-1512 (2010).
33. Koebnik, R., Locher, K.P. & Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* **37**, 239-253 (2000).
34. Li, E. & Hristova, K. Receptor tyrosine kinase transmembrane domains: function, dimer structure and dimerization energetics. *Cell. Adh. Migr.* **4**, 249-254 (2010).
35. Schlundt, A. et al. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Sci. Rep.* **7**, 1051 (2017).
36. Buchner, S., Schlundt, A., Lassak, J., Sattler, M. & Jung, K. Structural and functional analysis of the signal-transducing linker in the pH-responsive one-component system CadC of *Escherichia coli*. *J. Mol. Biol.* **427**, 2548-2561 (2015).
37. Kovacic, F. et al. A membrane-bound esterase PA2949 from *Pseudomonas aeruginosa* is expressed and purified from *Escherichia coli*. *FEBS Open Bio* **6**, 484-493 (2016).
38. Bleffert, F. et al. *Pseudomonas aeruginosa* esterase PA2949, a bacterial homolog of the human membrane esterase ABHD6: expression, purification and crystallization. *Acta Crystallogr. F* **75**, 270-277 (2019).
39. Bleves, S., Lazdunski, A. & Filloux, A. Membrane topology of three Xcp proteins involved in exoprotein transport by *Pseudomonas aeruginosa*. *J. Bacteriol.* **178**, 4297-4300 (1996).
40. Matsushita, K., Adachi, O., Shinagawa, E. & Ameyama, M. Isolation and characterization of outer and inner membranes from *Pseudomonas aeruginosa* and effect of EDTA on the membranes. *J. Biochem.* **83**, 171-181 (1978).
41. Coffey, B. & Anderson, G. Biofilm formation in the 96-well microtiter plate. In *Pseudomonas Methods and Protocols*, eds. Filloux, A. & Ramos, J.-L. Springer New York, **1149**, 631-641 (2014).
42. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797 (2007).
43. Ollis, D.L. et al. The alpha/beta hydrolase fold. *Protein Eng.* **5**, 197-211 (1992).
44. Chow, J. et al. The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS One* **7**, e47665 (2012).
45. Jaeger, K.E. et al. Bacterial lipases. *FEMS Microbiol. Rev.* **15**, 29-63 (1994).
46. Nardini, M., Lang, D.A., Liebeton, K., Jaeger, K.E. & Dijkstra, B.W. Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation. The prototype for family I.1 of bacterial lipases. *J. Biol. Chem.* **275**, 31219-31225 (2000).
47. Lang, D.A., Mannesse, M.L.M., De Haas, G.H., Verheij, H.M. & Dijkstra, B.W. Structural basis of the chiral selectivity of *Pseudomonas cepacia* lipase. *Eur. J. Biochem.* **254**, 333-340 (1998).

48. Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I. & Lomize, A.L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370-6 (2012).
49. Le Senechal, C. et al. Phospholipid content of *Pseudomonas aeruginosa* PAO1 is modulated by the growth phase rather than the immobilization state. *Lipids* (2019).
50. Blanka, A. et al. Constitutive production of c-di-GMP is associated with mutations in a variant of *Pseudomonas aeruginosa* with altered membrane composition. *Sci. Signal.* **8**, ra36 (2015).
51. Coskun, Ü. & Simons, K. Cell membranes: the lipid perspective. *Structure* **19**, 1543-1548 (2011).
52. Zhang, Y.M. & Rock, C.O. Thematic review series: glycerolipids. Acyltransferases in bacterial glycerophospholipid synthesis. *J. Lipid. Res.* **49**, 1867-1874 (2008).
53. Lin, Y. et al. The phospholipid-repair system LplT/Aas in Gram-negative bacteria protects the bacterial membrane envelope from host phospholipase A2 attack. *J. Biol. Chem.* **293**, 3386-3398 (2018).
54. Ruiz, C. et al. Activation and inhibition of *Candida rugosa* and *Bacillus*-related lipases by saturated fatty acids, evaluated by a new colorimetric microassay. *Biochim. Biophys. Acta* **1672**, 184-191 (2004).
55. Markweg-Hanke, M., Lang, S. & Wagner, F. Dodecanoic acid inhibition of a lipase from *Acinetobacter* sp. OPA 55. *Enzyme Microb. Technol.* **17**, 512-516 (1995).
56. Rose, I.A. Regulation of human red cell glycolysis: a review. *Exp. Eye Res.* **11**, 264-272 (1971).
57. Van Schaftingen, E. & Hers, H.G. Inhibition of fructose-1,6-bisphosphatase by fructose 2,6-biphosphate. *Proc. Natl. Acad. Sci. U S A* **78**, 2861-2863 (1981).
58. Alam, M.T. et al. The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nat. Commun.* **8**, 16018 (2017).
59. Lennen, R.M., Politz, M.G., Kruziki, M.A. & Pflieger, B.F. Identification of transport proteins involved in free fatty acid efflux in *Escherichia coli*. *J. Bacteriol.* **195**, 135-144 (2013).
60. Marianayagam, N.J., Sunde, M. & Matthews, J.M. The power of two: protein dimerization in biology. *Trends. Biochem. Sci.* **29**, 618-625 (2004).
61. Bocharov, E.V. et al. Helix-helix interactions in membrane domains of bitopic proteins: specificity and role of lipid environment. *Biochim. Biophys. Acta. Biomembr.* **1859**, 561-576 (2017).
62. Monk, B.C. et al. Architecture of a single membrane spanning cytochrome P450 suggests constraints that orient the catalytic domain relative to a bilayer. *Proc. Natl. Acad. Sci. U S A* **111**, 3865-3870 (2014).

Online methods

Cloning, protein expression and purification

Molecular biology methods, DNA purification and analysis by electrophoresis were performed as described previously.¹ For the expression of PlaF, *P. aeruginosa* PAO1 (wild-type) cells transformed² with plasmid pBBR-*pa2949*,¹ here abbreviated as p-*plaF*, or pBBR1mcs-3 used as empty vector were grown overnight at 37°C in *lysogeny broth* (LB) medium supplemented with tetracycline (100 µg/ml).³ The total membrane fraction of each strain was obtained by ultracentrifugation and proteins were purified as described previously.³ To reconstitute PlaF in octyl β-D-glucopyranoside (OG) or *n*-dodecyl β-D-maltoside (DDM) micelles, protein eluted from a Ni-NTA column was transferred into Tris-HCl buffer (100 mM, pH 8) supplemented with 30 mM or 0.22 mM detergent, respectively.

SDS-PAGE, zymography and immunodetection

The protein analysis by electrophoresis under denaturation conditions,⁴ in-gel esterase activity (zymography) and immunodetection by Western blotting were performed as described previously¹. The protein concentration was determined by UV spectrometry using a theoretical extinction coefficient of PlaF_{H6} of 22,920 M⁻¹ cm⁻¹.³

Enzyme activity assays, inhibition and enzyme kinetic studies

Esterase activity assays with *p*-nitrophenol fatty acid esters as substrates were performed in 96-well microplates as described previously.¹ Phospholipid and lysophospholipid substrates purchased from Avanti Polar Lipids (Alabaster, USA) were prepared for enzyme activity assays (25 µL enzyme + 25 µL substrate) as described previously.⁵ The amount of free fatty acids (FAs) released by (phospho)lipases was determined using the NEFA-HR(2) kit (Wako Chemicals, Richmond, USA).³ PLA1 and PLA2 activities of PlaF were measured using fluorescent substrates purchased from Thermo Fisher Scientific Inc. (Waltham, Massachusetts, USA): PLA1-PE, [N-((6-(2,4-DNP)amino)hexanoyl)-1-(BODIPY®FL C5)-2-hexyl-sn-glycero-3-phosphoethanolamine]; PLA2-PC, 1-O-(6-BODIPY®558/568-amino)hexyl)-2-BODIPY®FL C5-Sn-glycero-3-phosphocholine; and PLA2-PE, N-((6-(2,4-dinitrophenyl)amino)hexanoyl)-2-(4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-pentanoyl)-1-hexadecanoyl-sn-glycero-3-phosphoethanolamine triethylammonium salt as described by da Mata Madeira and coworkers.⁶ Measurements were performed using a plate reader in 96-well plates at 25°C by combining 50 µL of the substrate with 50 µL PlaF (0.7 µg/mL), or control enzymes, the PLA1 of

Thermomyces lanuginosus (5 U/mL) and the PLA2 or *Naja mocambique mocambique* (0.7 U/mL).

Inhibition. The inhibition of PlaF by FAs was assayed by combining FA dissolved in DMSO (20 fold stock solution) with *p*-NPB substrate solution followed by the addition of PlaF sample (8 nmol) and spectrophotometric enzyme activity measurement using *p*-NPB substrate.⁷ In control experiments, all compounds except FA were combined to assess PlaF stability in DMSO. Inhibition constants were calculated by fitting enzyme kinetic parameters obtained by varying FA concentration (0, 0.5, 1.5, 2.5, 5 and 7.5 mM) for different substrate concentrations (0.05, 0.1, 0.2, 0.3, 0.5 and 1 mM).⁸

Subcellular localization

Membranes from *P. aeruginosa* wild-type and overexpressing strain *p-plaF* were isolated as described previously.¹ To differentiate integral from peripheral membrane proteins, total cell membranes were incubated for 30 min at room temperature with three different reagents: 10 mM Na₂CO₃ (pH = 11), 4 M urea (in 20 mM MES buffer pH = 6.5) and 2 % (w/v) Triton X-100 (in 20 mM MES buffer pH = 6.5). After the incubation, the samples were centrifuged for 30 min at 180,000 *g* to separate integral proteins (pellet) from peripheral proteins (soluble fraction).

The separation of the inner and outer membrane was performed with a discontinuous sucrose gradient by ultracentrifugation at 180,000 *g* for 72 h and 4°C.⁹ The sucrose gradient consists of 1.5 ml fractions with 35, 42, 46, 50, 54, 58, 62 and 65 % (w/v) sucrose in 100 mM Tris-HCl pH 7.4. Isolated membranes from *P. aeruginosa* wild-type were resuspended in buffer containing 35 % sucrose and loaded on the top of the discontinuous sucrose gradient. Fractions were collected from the bottom (pierced tube) and sucrose concentration was determined with a refractometer (OPTEC, Optimal Technology, Baldock UK). To determine the orientation of catalytic PlaF domain *P. aeruginosa p-plaF* cells (10 ml culture with optical density (OD_{580nm}) 1 grown in LB medium at 37°C) were harvested by centrifugation (4,000 *g*, 4°C, 5 min) and suspended in 1 ml Tris-HCl buffer (50 mM, pH 7.5, 10 % sucrose (w/v)) followed by shock freezing with liquid nitrogen.¹⁰ Cells were thawed to room temperature and centrifuged (4,000 *g*, 4°C, 5 min) followed by incubation of the pellet for one hour on ice in Tris-HCl buffer (30 mM, pH 8.1, sucrose 20 % (w/v) EDTA 10 mM). Trypsin (20 µL, 1 mg/ml) was added to the suspension containing the cells with the permeabilized outer membrane and incubated at room temperature up to 5 h. The proteolytic reaction was stopped with SDS-

PAGE sample buffer and by incubation for 10 min at 99°C. Immunodetection of SecG and lipid A was performed as described above for PlaF using the respective antisera at 1/5,000, 1/2,000 and 1/1,000 dilutions.

Cross-linking assays

In vitro cross-linking using the bifunctional cross-linking reagent dimethyl pimelimidate (DMP) was performed as previously described¹¹ with the following modifications. PlaF (10 µL, 21.5 µM) purified with DDM was incubated with 5 µL freshly prepared DMP (150 mM in 100 mM Tris-HCl pH 8.4) for up to 2 h. The cross-linking reaction was terminated with 5 µL stop solution (50 mM Tris-HCl, 1M glycine, NaCl 150 mM, pH 8.3). For *in vivo* cross-linking, *P. aeruginosa* *p-plaF* and EV strains were grown in LB medium at 37°C until OD_{580nm} = 1. Cells were harvested by centrifugation (10 min, 4,000 g, 4°C), suspended in 1/20 volume of Tris-HCl (pH 8.3, 100 mM, NaCl 150 mM) and treated with the same volume of freshly prepared cross-linking reagent DMP (0, 20, 30 and 50 mM in Tris-HCl buffer 100 mM, pH 8.4) for 2 h. The cross-linking reaction was terminated with the same volume of stop solution (50 mM Tris-HCl, 1 M glycine, NaCl 150 mM, pH 8.3).

Analysis of concentration-dependent dimerization

Purified PlaF (20 µL, 50-60 µM) was transferred from purification buffer into labeling buffer (Na-PO₄ 20 mM, pH 8.3) supplemented with OG (30 mM) or DDM (0.22 mM) using PD10 columns (GE Healthcare, Solingen, Germany) according to the manufacturer's protocol. Labeling was performed by incubating PlaF with 15 µL dye (440 µM stock solution) for 2.5 h using the NHS labeling Kit (NanoTemper, Munich, Germany). PlaF was again transferred into a purification buffer using PD10 columns. Non-labeled PlaF was diluted with the same buffer in 16 steps by combining the same volume of the protein and buffer yielding samples with concentrations from 26.9 µM to 1.6 nM. Samples containing 100 nM labeled PlaF were incubated for 16 h at room temperature in the dark and microscale thermophoresis (MST) experiments were performed using the Monolith® NT.115 device (NanoTemper, Munich, Germany) with the following set up: MST power, 60 %; excitation power 20 %; excitation type, red; 25°C. Constants were calculated according to the four-parameter logistic, nonlinear regression model using Origin Pro 2018 software.

The activity of PlaF purified with OG and used for MST analysis was assayed by combining 15 µL of enzyme solution with 15 µL 4-methylumbelliferyl palmitate (4-MUP, 2 mM) dissolved in

purification buffer containing 10 % (v/v) propan-2-ol. Fluorescence was measured during 10 min (5 seconds steps) using plate reader in black 96-well plates at 30°C.

Construction of a *P. aeruginosa* Δ plaF mutant

P. aeruginosa PAO1 (wild-type) cells were transformed with the pEMG- Δ plaF mutagenesis vector and *P. aeruginosa* Δ plaF mutant strain was generated by homologous recombination.¹²

***Drosophila melanogaster* virulence model**

P. aeruginosa wild-type and a Δ plaF strains were cultivated in LB medium at 37°C until they reached the stationary phase ($OD_{580nm} = 3$). Twenty flies per experiment were locally infected by the pricking method with 100 bacteria and the live flies were counted throughout 45 h.¹³ The results are means of 3 and 4 experiments for each of the Δ plaF and the wild-type strain, respectively.

Growth curves, biofilm and motility assays

The growth of *P. aeruginosa* wild-type and Δ plaF cultures in Erlenmeyer flasks (agitation at 160 rpm) was monitored by measuring OD_{580nm} during 24 h. OD_{580nm} was converted to colony-forming units (CFU) by multiplying with the factor 8×10^8 experimentally determined for *P. aeruginosa* wild-type strain used in our laboratory. The biofilm assay was performed as described elsewhere¹⁴ by growing the *P. aeruginosa* cultures (100 μ l) for 48 h at 37°C in LB medium in plastic 96-well microtiter-plates. Cells attached to the plastic surface after removing the planktonic cells were stained with 0.1 % (w/v) crystal violet solution for 10 min, solubilized with acetic acid (30 % v/v) and quantified spectrophotometrically. Swimming, swarming and twitching motility assays were performed as described previously.¹⁵ The swimming area on the LB agar plate was measured using the program ImageJ 1.51j8.¹⁶

Lipidomic analysis of glycerophospholipids extracted from cell membranes

The cells of *P. aeruginosa* wild-type and Δ plaF cultures grown overnight in LB medium (15 ml) at 37°C were harvested by centrifugation at 4,000 *g* and 4°C for 15 min and suspended in 2 ml ddH₂O followed by boiling for 10 min to inactivate (phospho)lipases. Cells were harvested by centrifugation (4,000 *g*, 4°C, 15 min) and total lipids were extracted from the cell pellet.¹⁷ Briefly, after boiling the water was removed by centrifugation (4,000 *g*, 4°C, 15 min). Lipids were extracted with CHCl₃ : CH₃OH = 1 : 2 (v/v) and the organic phase collected. The extraction was repeated with CHCl₃ : CH₃OH = 2 : 1 (v/v) and the organic phases were combined. One volume of CHCl₃ and 0.75 volumes of an aqueous solution containing 1 M KCl and 0.2 M H₃PO₄ were added to the combined chloroform/methanol extracts. Samples were vortexed and

centrifuged (2,000 *g*, 5 min). The organic phase was withdrawn and the solvent of the lipid extract was evaporated under a stream of N₂. Total lipids were dissolved in CHCl₃ : CH₃OH = 2:1 (v/v). GPLs were quantified by Q-TOF mass spectrometry (Q-TOF 6530; Agilent Technologies, Böblingen, Germany) as described elsewhere.¹⁷ Statistical analysis of the GPL amount was performed using the T-test and the Shapiro-Wilk method to determine significant changes of normally distributed values obtained from four *P. aeruginosa* wild-type lipidome and four Δ *plaF* samples.

Gas chromatography-mass spectrometric (GC-MS) analysis of FA extracted from PlaF

FAs were extracted from PlaF purified from 13 g *P. aeruginosa* *p-plaF* cells with OG using four parts of organic solvent (CHCl₃ : CH₃OH = 2 : 1). Extraction was repeated three times, the chloroform extracts were combined, chloroform was evaporated and FAs were dissolved in 200 μ L chloroform. The chloroform extract was mixed with ten volumes of acetonitrile and filtered through a 0.2 μ m pore size filter. For GC-MS analysis, FA extracts and standards (C10-, C11-, C14-, C15-, C16- and C18-fatty acid; C16-, C18- and C20-primary fatty alcohol) were derivatized to their trimethylsilylester and trimethylsilylether, respectively. 900 μ L of the sample or standard solution (CHCl₃ : acetonitrile = 1 : 5) was mixed with 100 μ L N-methyl-N-(trimethylsilyl) trifluoroacetamide and heated to 80°C for 1 h. The GC-MS system consisted of a Trace GC Ultra gas chromatograph, TriPlus autosampler and an ITQ 900 mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA). Analytes were separated on a Zebtron-5-HT Inferno column (60 m x 0.25 mm i.d., 0.25 μ m film thickness, Phenomenex, USA). Helium was used as carrier gas at a constant gas flow of 1.0 ml/min. The oven temperature program employed for analysis of silylated fatty acids and alcohols was as follows: 80°C; 5°C/min to 340°C, held for 5 min. The injector temperature was held at 290°C, and all injections (1 μ L) were made in the split mode (1:10). The mass spectrometer was used in the electron impact (EI, 70 eV) mode and scanned over the range *m/z* 25 - 450 with an acquisition rate of 3 microscans. The transfer line and ion source were both kept at 280°C. Data processing was performed by use of the software XCalibur 2.0.7 (ThermoFisher Scientific). Fatty acids from PlaF sample were identified by comparison of their retention times and mass spectra with those of fatty acid standards.

Crystallization, data collection, structure determination and analysis

PlaF purified with OG was crystallized as described previously.³ The X-ray diffraction data were recorded at beamline ID29 of the European Synchrotron Radiation Facility (ESRF, Grenoble,

France) and processed as described.³ The structure was determined by molecular replacement using the automated pipeline “MrBUMP” from the CCP4 package.¹⁸ In detail, a combination of PHASER,¹⁹ REFMAC,²⁰ BUCCANEER,²¹ and SHELXE²² resulted in an interpretable electron density map to expand the placed model by molecular replacement using the model built with HsaD from *Mycobacterium tuberculosis* (PDB code: 2VF2).²³ Phase improvement was achieved by running several cycles of automated model building (ARP/wARP, CCP4) and refinement using the PHENIX²⁴ package. The model was further corrected by manual rebuilding using the program COOT.²⁵ Detailed statistics on data collection and refinement are provided in table 1. None of the residues are in disallowed regions according to Ramachandran plots generated with MolProbity (PHENIX).²⁶ The secondary structure was defined according to Kabsch and Sander.²⁷ Coordinates and structure factors for PlaF have been deposited in the Protein Data Bank under accession code 6I8W.

Identification of structural homologs of PlaF

PlaF structural homologs were defined as protein structures from a non-redundant subset of PDB structures with less than 90 % sequence identity to each other (PDB90 database, 12.10.2019) with a Z-score higher than 2 according to the DALI server²⁸. Sequence alignment (Table S6) based on structural superimposition of all 357 homologs (Table S7) of PlaF_B (340 homologs (Table S8) of PlaF_A were among PlaF_B homologs) was used to identify proteins with homology in TM-JM helix of PlaF (residues 1-38). To evaluate homology, thirty-nine 3D structures with partial conservation of TM-JM helix were superimposed onto the PlaF structure (Fig. S3).

Molecular dynamics simulations of dimer and monomers

The crystal structure of the PlaF dimer was used as the starting point for building the systems for MD simulations. Five missing C-terminal residues on both chains were added by using MODELLER.²⁹ The dimer was oriented into the membrane using the PPM server from the OPM webpage.³⁰ From the so oriented dimer structure, chain B was deleted to obtain a PlaF_A monomer in a dimer-oriented configuration; in the same way, chain A was deleted to keep PlaF_B. Additionally, the PlaF_A and PlaF_B monomers were oriented by themselves using the PPM server, yielding tilted configurations (t-PlaF_A and t-PlaF_B). These five starting configurations, di-PlaF, PlaF_A, and PlaF_B, t-PlaF_A and t-PlaF_B were embedded into a DOPE : DOPG = 3 : 1 membrane with CHARMM-GUI v1.9,³¹ resembling the native inner membrane of Gram-negative bacteria,^{32,33} ensuring a distance of at least 15 Å between the protein or membrane

and the solvation box boundaries. KCl at a concentration of 0.15 M was included in the solvation box to obtain a neutral system. The GPU particle mesh Ewald implementation from the AMBER16 molecular simulation suite^{34,35} with the ff14SB³⁶ and Lipid17^{37,38,39} parameters for the protein and the membrane lipids, respectively, were used; water molecules were added using the TIP3P model⁴⁰. For each protein configuration, ten independent molecular dynamics (MD) simulations of 2 μ s length were performed. Covalent bonds to hydrogens were constrained with the SHAKE algorithm⁴¹ in all simulations, allowing to use a time step of 2 fs.

Thermalization of simulation systems

Initially, systems were 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 finally, a minimization without restraints. The temperature was maintained by using a Langevin thermostat⁴², with a friction coefficient of 1 ps⁻¹. The pressure, when required, was maintained using a semi-isotropic Berendsen barostat⁴³, coupling the membrane (x-y) plane. The thermalization was started from the minimized structure, which was heated by gradually increasing the temperature from 10 to 100 K for 5 ps under NVT conditions, and from 100 to 300 K for 115 ps under NPT conditions at 1 bar. The equilibration process was continued for 5 ns under NPT conditions, after which production runs were started using the same conditions.

Structural analysis of MD trajectories

All simulations showed stable protein structures (Figs. S12a and S12b) and membrane phases, evidenced by electron density and order parameter calculations (Figs. S12d and S12e). The area per lipid through all simulations calculated for the leaflet opposite to the one where PlaF was embedded was $61.3 \pm 0.13 \text{ \AA}^2$ (mean \pm SEM), similar to values reported previously³³. All analyses were performed by using CPPTRAJ⁴⁴. The distance between the centers of mass (COM) of residues 25 to 38 C α atoms of the chains in the dimer structure was evaluated; this residue range corresponds to the solvent-accessible half of helix TM-JM (Figure 8a). The distance between the COM of C α atoms of the transmembrane portion (residues 1 to 25; the first half of TM-JM helix) of each chain was also evaluated. For the monomer structures, the angle with respect to the membrane normal was assessed. For this, the angle between the membrane normal and the vector between the COM of C α atoms of residues 21 to 25 and residues 35 to 38 was calculated.

Potential of mean force calculations of dimer dissociation

For calculating a configurational free energy profile (potential of mean force, PMF) of the process of dimer dissociation, 36 intermediate states were generated by separating one chain of the dimer along the membrane plane at 1 Å steps, resulting in a minimum and maximum distance between the chain centers of mass (COM) of 40.8 and 68 Å, respectively. Each intermediate state was embedded into a membrane of approximately $157 \times 157 \text{ Å}^2$ by using PACKMOL-Memgen,⁴⁵ and independent MD simulations of 300 ns length each were performed using the same conditions as in the unbiased simulations, resulting in a total simulation time of 10.8 μs. Umbrella sampling simulations were performed by restraining the initial distance between chains in every window with a harmonic potential, using a force constant of $4 \text{ kcal mol}^{-1} \text{ Å}^{-2}$;⁴⁶ the distance between the COM of C_α atoms of residues 25 to 38 of each monomer was used as a reaction coordinate. The values for the distances were recorded every 2 ps and post-processed with the Weighted Histogram Analysis Method implementation of A. Grossfield (WHAM 2.0.9),^{47,48} removing the first 100 ns as an equilibration of the system. The kernel densities showed a median overlap of 8.2 % between contiguous windows (Fig. S12f, upper panel), well suited for PMF calculations.⁴⁹ The error was estimated by separating the last 200 ns of data in four independent parts of 50 ns each and then calculating the standard error of the mean of the independently determined energy profiles.

Estimation of association free energy from PMF

The association free energy was estimated from the obtained PMF following the membrane two-body derivation from Johnston *et al.* (2012)⁵⁰ and our previous work⁵¹. In details, the PMF of dimer association is integrated along the reaction coordinate to calculate an association constant (K_a), which is transformed to the mole fraction scale (K_x) taking into account the number of lipids N_L per surface area A , and this value is used to calculate the difference in free energy between dimer and monomers (ΔG), according to eqs. 1-3:

$$K_a = \frac{|\Omega|}{(2\pi)^2} \int_0^D r e^{\frac{-w(r)}{k_B T}} dr \quad K_x = K_a \frac{N_L}{A} \quad \Delta G = -RT \ln(K_x) \quad (\text{eqs. 1-3})$$

where r is the value of the reaction coordinate, $w(r)$ is the PMF at value r , D is the maximum distance at which the protein is still considered a dimer, k_B is the Boltzmann constant, and T is the temperature at which the simulations were performed. The factor $\frac{|\Omega|}{(2\pi)^2}$ considers the

restriction of the configurational space of the monomers upon dimer formation in terms of the sampled angle between the two chains in the dimeric state (eq. 4) and the accessible space for the monomers, $(2\pi)^2$.

$$||\Omega|| = [\max(\theta_a) - \min(\theta_a)] * [\max(\theta_b) - \min(\theta_b)] \quad (\text{eq. 4})$$

In eq. 4, the angle θ_a is defined as the angle formed between the vectors connecting the COM of chain b with the COM of the chain a and with the COM of residues 25 to 38 of the latter chain; θ_b is defined analogously starting from the COM of chain a . A value for $||\Omega||$ of 0.55 computed from eq. 4 indicates the fraction of the accessible space that the PlaF monomers have in the dimeric state compared to when both chains rotate independently $[(2\pi)^2]$.

Potential of mean force calculations of monomer tilting

The initial conformations used in every window for calculating the PMF of the monomer tilting were obtained from a representative trajectory of an unbiased MD simulation where spontaneous tilting occurred (first microsecond of MD simulations of replica 10, starting configuration as in di-PlaF crystal structure). The distance d along the z-axis between the COM of C_α atoms of residues 33 to 37 of the monomer with the COM formed by atom C_{18} of the membrane phospholipids was used. d significantly correlates ($R^2 = 0.997$, $p < 0.001$) with the angle formed by the second half of helix α M1 of the monomer (residues 25 to 38) and the normal vector of the membrane (Fig. S12g). Twenty-two conformations were extracted from the representative trajectory, taking the respective snapshots where d and the angle showed the least absolute deviation to the average value obtained by binning d in windows of 2 Å width and with an evenly distributed separation of 1 Å. The initial distance for every configuration was restrained by a harmonic potential with a force constant of 4 kcal mol⁻¹ Å⁻², and sampling was performed for 300 ns per window. The data were obtained every 2 ps and analyzed as described above, resulting in 8.6 % of median overlap between kernel densities of contiguous windows (Fig. S1f, lower panel). The error was estimated in the same way as for the dimerization.

Estimation of tilting free energy

For calculating the free energy difference between the obtained basins, the PMF of monomer tilting was integrated using eq. 5 and 6⁵²:

$$K_{tilting} = \frac{\int_{B_1} e^{-\frac{w(d)}{k_B T}} dr}{\int_{B_2} e^{-\frac{w(d)}{k_B T}} dr} \quad \Delta G_{tilting} = -RT \ln K_{tilting} \quad (\text{eqs. 5,6})$$

where d is defined as above, $w(d)$ is the value of the PMF at that distance, and B_1 and B_2 represent the basins for the tilted and split configurations, respectively. The integration limits B_1 and B_2 included each basin portion below half of the value between the basin minimum and the energy barrier separating the basins, respectively (Fig. 7e, yellow shaded regions).

PlaF dimer versus monomer proportion under *in vivo* conditions

Based on the association constant computed according to eq. 1, $K_a = [D] / [M]^2 = 1.57 \times 10^7 \text{ \AA}^2$, with $[D]$ and $[M]$ as area concentrations of dimer and monomer, respectively, the proportion of PlaF dimer versus monomer in a live cell of *P. aeruginosa* can be computed. Experimentally, from 1 L ($OD_{580nm} \sim 1$) of *P. aeruginosa* p-plaF ~ 1 mg PlaF was purified³ while from 15 mL of *P. aeruginosa* wild-type culture ($OD_{580nm} \sim 5$) ~ 3 mg of lipids were obtained (Table S4). Considering the molecular weight of PlaF of 35.5 kDa and assuming 750 Da as the average molecular weight of membrane GPL, this relates to a concentration under overexpressing conditions of $\sim 4.23 \times 10^{-4}$ PlaF monomers per lipid. Under non-overexpressing conditions, the concentration of PlaF monomers is estimated to be at least 100-1000 fold lower, i.e., 4.23×10^{-6} - 4.23×10^{-7} PlaF monomers per lipid. Considering that the area per lipid in a PE : PG = 3 : 1 membrane at 300 K is approximately 61 \AA^2 (computed in this work and ref.⁵³), the total area concentration of PlaF molecules then is

$$T = 2[D] + [M] = [6.93 \times 10^{-9}, 6.93 \times 10^{-8}] \frac{\text{PlaF}}{\text{\AA}^2} \quad (\text{eq. 7}).$$

Expressing the association constant in terms of the monomer concentration using eq. 1 yields

$$K_a = \frac{T - [M]}{[M]^2} \Leftrightarrow 2K_a[M]^2 + [M] - T = 0 \quad (\text{eq. 8}),$$

and solving the quadratic equation then results in

$$[M] = \frac{-1 + \sqrt{1 + 8K_a T}}{4K_a} = [5.85 \times 10^{-9}, 3.37 \times 10^{-8}] \frac{\text{PlaF}}{\text{\AA}^2} \quad (\text{eq. 9})$$

and

$$[D] = \frac{T - [M]}{2} = [5.38 \times 10^{-10}, 1.78 \times 10^{-8}] \frac{\text{PlaF dimer}}{\text{\AA}^2} \quad (\text{eq. 10}),$$

These results show that in live cells the fraction of PlaF in the monomeric (dimeric) state is between 49 and 85 % (51 and 15 %), where the PlaF monomer is considered to be in the “split” configuration with respect to the membrane normal.

As the tilting of the PlaF monomer is energetically favorable compared to the “split” configuration and, hence, depletes the concentration of “split” PlaF monomers, the dimeric PlaF concentration will decrease (Figure 8a). To quantitatively consider the effect of the tilting, we express the overall equilibrium constant for the processes shown in Figure 8a as

$$K = K_a K_{\text{tilting}}^{-2} = \frac{[D]}{[M_{\text{tilted}}]^2} \quad (\text{eq. 11}),$$

where

$$K_{\text{tilting}} = \frac{[M_{\text{tilted}}]}{[M]} = 3.35, \text{ equivalent to } \Delta G_{\text{tilting}} = -0.72 \frac{\text{kcal}}{\text{mol}}, \text{ computed according to eq. 5.}$$

Following the same procedure as before then yields

$$[M_{\text{tilted}}] = [6.8 \times 10^{-9}, 5.94 \times 10^{-8}] \frac{\text{PlaF}}{\text{\AA}^2}$$

$$[D] = [6.5 \times 10^{-11}, 4.96 \times 10^{-9}] \frac{\text{PlaF dimer}}{\text{\AA}^2},$$

showing that in live cells the fraction of PlaF in the tilted monomeric (dimeric) state is between 86 and 98 % (14 and 2 %). A graphical representation of the percentage of protein as a tilted monomer with respect to the protein concentration in the membrane is shown in figure 7f.

References

1. Kovacic, F. et al. A membrane-bound esterase PA2949 from *Pseudomonas aeruginosa* is expressed and purified from *Escherichia coli*. *FEBS Open Bio* **6**, 484-493 (2016).
2. Choi, K.H., Kumar, A. & Schweizer, H.P. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J. Microbiol. Methods*. **64**, 391-397 (2006).
3. Bleffert, F. et al. *Pseudomonas aeruginosa* esterase PA2949, a bacterial homolog of the human membrane esterase ABHD6: expression, purification and crystallization. *Acta Crystallogr. F* **75**, 270-277 (2019).
4. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
5. Jaeger, K.E. & Kovacic, F. Determination of lipolytic enzyme activities. *Methods Mol. Biol.* **1149**, 111-134 (2014).
6. da Mata Madeira, P.V. et al. Structural basis of lipid targeting and destruction by the type V secretion system of *Pseudomonas aeruginosa*. *J. Mol. Biol.* **428**, 1790-1803 (2016).
7. Tian, W.X. & Tsou, C.L. Determination of the rate constant of enzyme modification by measuring the substrate reaction in the presence of the modifier. *Biochemistry* **21**, 1028-1032 (1982).
8. Kenakin, T.P. Enzymes as Drug Targets. in *Pharmacology in Drug Discovery*, Academic Press, 105-124 (2012).
9. Viarre, V. et al. HxcQ liposecretin is self-piloted to the outer membrane by its N-terminal lipid anchor. *J. Biol. Chem.* **284**, 33815-33823 (2009).
10. Eichler, J. & Wickner, W. The SecA subunit of *Escherichia coli* preprotein translocase is exposed to the periplasm. *J. Bacteriol.* **180**, 5776-5779 (1998).
11. de Jong, L. et al. In-culture cross-linking of bacterial cells reveals large-scale dynamic protein-protein interactions at the peptide level. *J. Proteome. Res.* **16**, 2457-2471 (2017).
12. Martinez-Garcia, E. & de Lorenzo, V. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environ. Microbiol.* **13**, 2702-2716 (2011).
13. Apidianakis, Y. & Rahme, L.G. *Drosophila melanogaster* as a model host for studying *Pseudomonas aeruginosa* infection. *Nat. Protoc.* **4**, 1285-1294 (2009).
14. Coffey, B. & Anderson, G. Biofilm formation in the 96-well microtiter plate. in *Pseudomonas Methods and Protocols*, eds. Filloux, A. & Ramos, J.-L., Springer New York, **1149**, 631-641 (2014).
15. Ha, D.-G., Kuchma, S.L. & O'Toole, G.A. Plate-based assay for swimming motility in *Pseudomonas aeruginosa*. in *Pseudomonas Methods and Protocols*, eds. Filloux, A. & Ramos, J.-L. Springer New York, 59-65 (2014).
16. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).
17. Gasulla, F. et al. The role of lipid metabolism in the acquisition of desiccation tolerance in *Craterostigma plantagineum*: a comparative approach. *Plant J.* **75**, 726-741 (2013).
18. Keegan, R.M. et al. Evaluating the solution from MrBUMP and BALBES. *Acta Crystallogr. D* **67**, 313-323 (2011).
19. McCoy, A.J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
20. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240-255 (1997).
21. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002-1011 (2006).
22. Hubschle, C.B., Sheldrick, G.M. & Dittrich, B. ShelXle: a Qt graphical user interface for SHELXL. *J. Appl. Crystallogr.* **44**, 1281-1284 (2011).

23. Lack, N.A. et al. Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *J. Biol. Chem.* **285**, 434-443 (2010).
24. Adams, P.D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94-106 (2011).
25. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126-2132 (2004).
26. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948-1954 (2002).
27. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
28. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-549 (2010).
29. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).
30. Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I. & Lomize, A.L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370-376 (2012).
31. Jo, S., Lim, J.B., Klauda, J.B. & Im, W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50-58 (2009).
32. Benamara, H. et al. Characterization of membrane lipidome changes in *Pseudomonas aeruginosa* during biofilm growth on glass wool. *PLoS One* **9**, e108478 (2014).
33. Murzyn, K., Rog, T. & Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophys. J.* **88**, 1091-1103 (2005).
34. Le Grand, S., Götz, A.W. & Walker, R.C. SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374-380 (2013).
35. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089-10092 (1993).
36. Maier, J.A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696-3713 (2015).
37. Dickson, C.J. et al. Lipid14: the Amber lipid force field. *J. Chem. Theory Comput.* **10**, 865-879 (2014).
38. Skjevik, A.A. et al. Simulation of lipid bilayer self-assembly using all-atom lipid force fields. *Phys. Chem. Chem. Phys.* **18**, 10573-10584 (2016).
39. al., D.A.C.e. AMBER 2017 Reference Manual. University of California (2017).
40. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926-935 (1983).
41. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).
42. Quigley, D. & Probert, M.I. Langevin dynamics in constant pressure extended systems. *J Chem Phys* **120**, 11432-11441 (2004).
43. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984).
44. Roe, D.R. & Cheatham, T.E., 3rd. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084-3095 (2013).
45. Schott-Verdugo, S. & Gohlke, H. PACKMOL-Memgen: a simple-to-use, generalized workflow for membrane-protein-lipid-bilayer system building. *J. Chem. Inf. Model.* **59**, 2522-2528 (2019).
46. Torrie, G.M. & Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* **23**, 187-199 (1977).

47. Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H. & Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011-1021 (1992).
48. Grossfield, A. An implementation of WHAM: the Weighted Histogram Analysis Method, <http://membrane.urmc.rochester.edu/content/wham> (2019).
49. Chen, P.C. & Kuyucak, S. Accurate determination of the binding free energy for KcsA-charybdotoxin complex from the potential of mean force calculations with restraints. *Biophys. J.* **100**, 2466-2474 (2011).
50. Johnston, J.M., Wang, H., Provasi, D. & Filizola, M. Assessing the relative stability of dimer interfaces in G protein-coupled receptors. *PLoS Comput. Biol.* **8**, e1002649 (2012).
51. Pagani, G. & Gohlke, H. On the contributing role of the transmembrane domain for subunit-specific sensitivity of integrin activation. *Scientific Reports* **8**, 5733 (2018).
52. Doudou, S., Burton, N.A. & Henchman, R.H. Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *Journal of Chemical Theory and Computation* **5**, 909-918 (2009).
53. Murzyn, K., Rog, T. & Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophysical Journal* **88**, 1091-1103 (2005).

1 Supplementary material for the article:

2

3 **Structural and mechanistic insights into phospholipase A-mediated membrane**
4 **phospholipid degradation related to *Pseudomonas aeruginosa* virulence**

5

6 Florian Bleffert¹, Joachim Granzin², Muttalip Caliskan¹, Stephan N. Schott-Verdugo^{3,4}, Meike
7 Siebers^{5,6}, Björn Thiele⁷, Laurence Rahme⁸, Peter Dörmann⁵, Holger Gohlke^{2,3,9,*}, Renu Batra-

8

Safferling^{2,*}, Karl-Erich Jaeger^{1,10} and Filip Kovačič^{1,*}

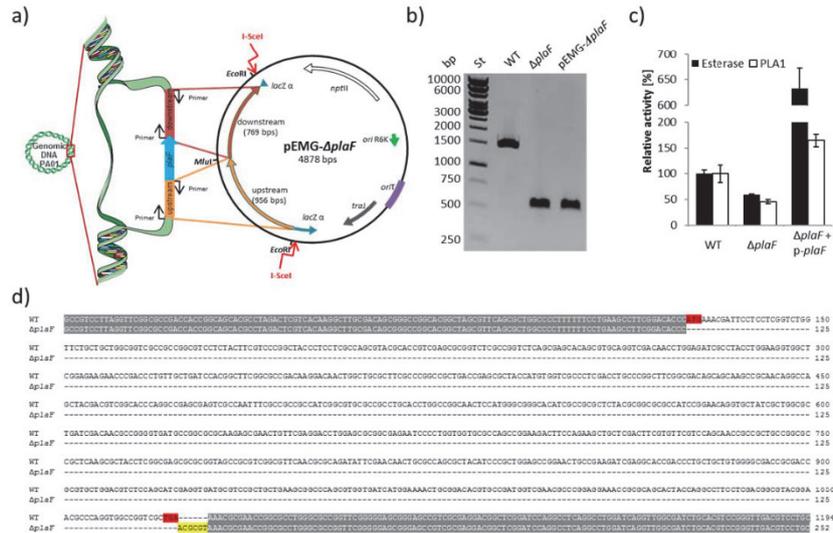
9

- 10 1. Institute of Molecular Enzyme Technology, Heinrich Heine University Düsseldorf,
11 Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
- 12 2. Institute of Complex Systems - Structural Biochemistry (ICS-6), Forschungszentrum
13 Jülich GmbH, D-52425 Jülich, Germany
- 14 3. Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University
15 Düsseldorf, D-40225 Düsseldorf, Germany
- 16 4. Centro de Bioinformática y Simulación Molecular (CBSM), Faculty of Engineering,
17 University of Talca, 1 Poniente 1141, Casilla 721, Talca, Chile
- 18 5. Institute of Molecular Physiology and Biotechnology of Plants (IMBIO), University of
19 Bonn, D-53115 Bonn, Germany
- 20 6. Institute for Plant Genetics, Heinrich Heine University Düsseldorf, D-40225 Düsseldorf,
21 Germany
- 22 7. Institute of Bio- and Geosciences, Plant Sciences (IBG-2) and Agrosphere (IBG-3),
23 Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
- 24 8. Department of Microbiology and Immunobiology, Harvard Medical School, Boston,
25 MA, USA
- 26 9. John von Neumann Institute for Computing (NIC) & Jülich Supercomputing Centre
27 (JSC), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
- 28 10. Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich
29 GmbH, D-52425 Jülich, Germany

30

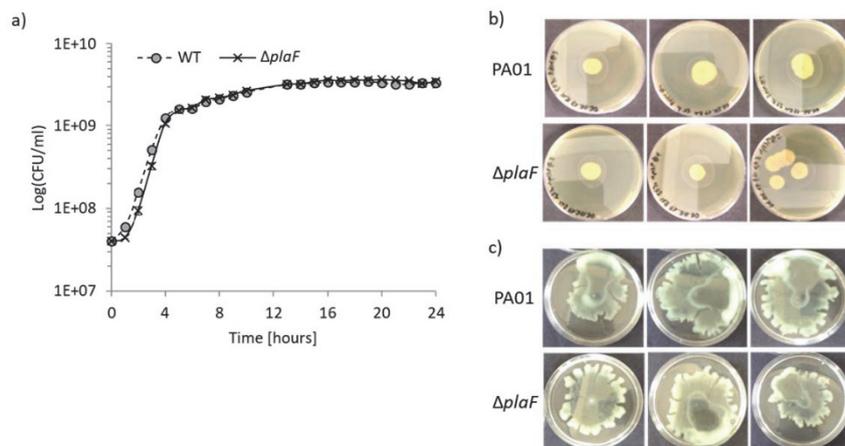
31 * *Corresponding authors*

32 **Supplementary figures**



33
 34 **Fig. S1: Generation of *P. aeruginosa* Δ plaF deletion mutant strain.** a) Generation of the mutagenesis
 35 vector pEMG- Δ plaF and homologous recombination with a chromosome of *P. aeruginosa* PAO1.¹
 36 Upstream and downstream regions of *plaF* gene were amplified by standard PCR using Phusion DNA
 37 polymerase, a genomic DNA of *P. aeruginosa* PAO1 as a template, and primer pairs 5'-
 38 ATATATGAATTCTGTCTGCGCGGAAACGACAGCGP-3'/5'-
 39 ATATATACGCGTGGGTGTCCGAAGGCTTCAGGAAAAAGGGG-3' and 5'-
 40 ATATATACGCGTAAACGGAACCGGCGCCTGGG-3'/5'-
 41 CTGGATGAATTCTGGCCTGGACACGACAAGGAAGTGATCAAGG-3', respectively. DNA fragments
 42 upstream and downstream of *plaF* gene were cloned into the pEMG vector by ligation of DNA
 43 fragments hydrolyzed with *EcoRI* restriction endonuclease. The white arrow indicates the neomycin
 44 phosphotransferase II gene (*npII*) necessary for plasmid selection in the presence of kanamycin
 45 antibiotic. DNA recognition sequence for I-SceI restriction endonuclease is indicated with the red
 46 arrows. b) Verification of *P. aeruginosa* Δ plaF mutant strain by PCR. Standard PCR was performed
 47 using Phusion DNA polymerase and the primers (5'- AAGGTCGCGCCAGGAATTTCCG-3'/5'-
 48 CCGCCTGCGTCCGACTACAAG-3') that bind to the up- and downstream regions of *plaF* gene. As
 49 PCR templates were used genomic DNA of *P. aeruginosa* PAO1 (WT), pEMG- Δ plaF plasmid and DNA
 50 obtained from the colony of *P. aeruginosa* Δ plaF scratched from the LB agar plate and suspended in
 51 water. The expected sizes of DNA fragments were 1494 bp, 552 bp, and 552 bp for WT, Δ plaF, and
 52 pEMG- Δ plaF, respectively. DNA products were analyzed by electrophoresis on agarose gel (1 % w/v)

53 in the presence of ethidium bromide and visualized by UV light exposure. The sizes of standard DNA
54 fragments are indicated on the left. c) Phospholipase A1 and esterase activity of *P. aeruginosa* Δ *plaF*
55 strain. *P. aeruginosa* PAO1 and Δ *plaF* strains were grown overnight in LB-medium at 37°C, and cells
56 were harvested by centrifugation. Cells were suspended in Tris-HCl buffer (100 mM, pH 8) to equal
57 cell count, and an enzyme activities of these samples were measured using standard esterase and
58 PLA1 assays with *p*-NPB and N-((6-(2,4-DNP)amino)hexanoyl)-1-(BODIPY®FL C5)-2-hexyl-sn-glycero-3-
59 phosphoethanolamine as the substrate, respectively. The results are mean \pm standard deviations of
60 three biological replicates each measured three times. d) Sequence alignment of DNA products of *P.*
61 *aeruginosa* Δ *plaF* and the wild-type (WT) strain obtained by PCR as described in Fig S1b. The start
62 and stop codons of *plaF* gene are indicated in red, identical nucleotides are indicated in grey, the
63 recognition site for *MluI* restriction endonuclease inserted on the chromosome using the
64 mutagenesis vector pEMG- Δ *plaF* in yellow. The company Eurofins Genomics (Ebersberg, Germany)
65 performed DNA sequencing.



66

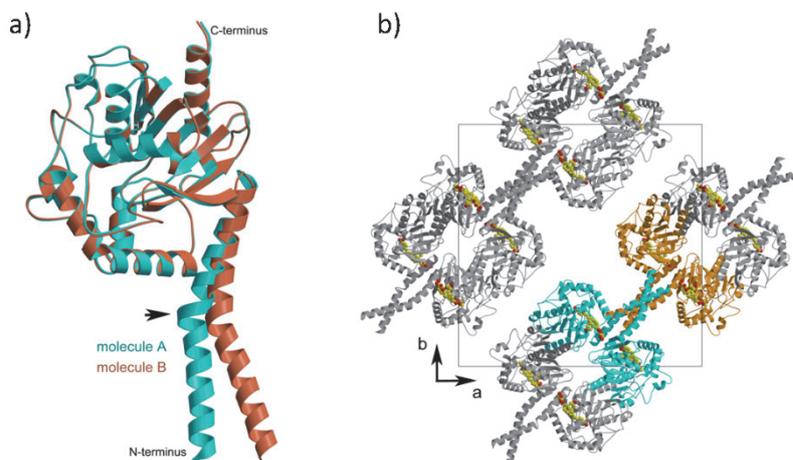
67 **Fig. S2:** The growth (a) twitching (b) and swarming (c) motilities of *P. aeruginosa* PAO1 and
68 $\Delta plaF$ do not differ. *P. aeruginosa* strains (n = 3) were grown in LB medium in Erlenmeyer
69 flasks at 37 °C. Twitching and swarming motilities were monitored after 24 h of incubation at
70 37 °C (n > 6).

71



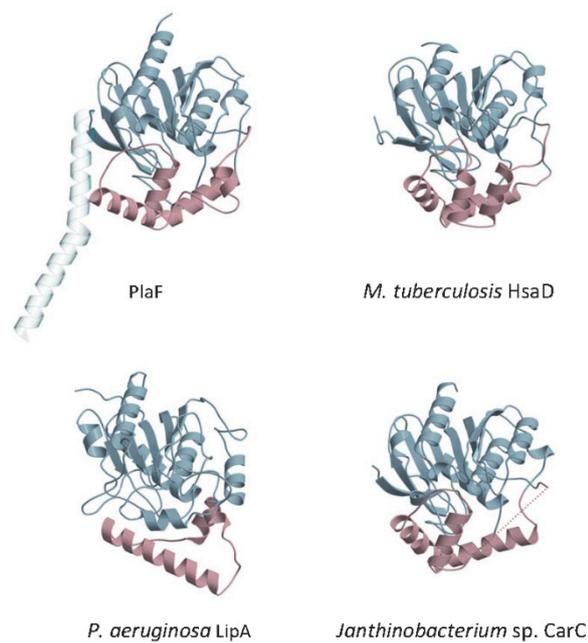
72

73 **Fig. S3: TM-JM helix of PlaF is not detected among PlaF structural homologs.** Pairwise structural
74 alignment of PlaF_B (colored orange) with 1cqz (a), 1orw (b), 1yr2 (c), 1z68 (d), 2bkl (e), 2ecf (f), 2bgc
75 (g), 2jbw (h), 2roq (i), 3mga (j), 3mun (k), 3o4j (l), 4hai (m), 4n8e (n), 5alj (o), 5l8s (p), 5t88 (q), 5yzm
76 (r), 6eop (s), 6hxa (t), 6igp (u). PlaF structural homologs (colored gray) containing conserved
77 sequence in TM-JM helix region of PlaF (yellow highlighted in table S6) and in catalytic domain as
78 identified by Dali server.

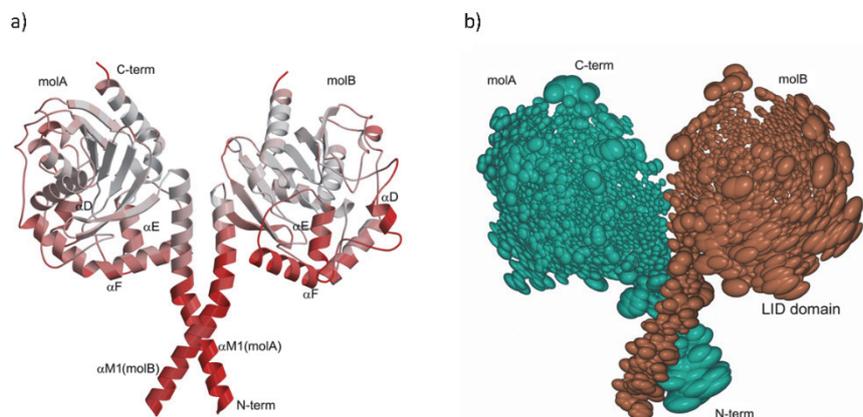


79
80

81 **Fig. S4:** a) Superposition of PlaF_A (cyan) and PlaF_B (orange). The arrow indicates the kink in
82 TM-JM helix. b) Crystal packing of PlaF showing a four-helix bundle formed by TM-JM
83 helices. For clarity, two dimers are shown in cyan and orange colors. The ligands present are
84 shown as filled circles and are colored by element (carbon yellow, oxygen red). The unit cell
85 (black square) and the a-b axes are labeled.



86
87 **Fig. S5:** The lid-like domains shown in brown vary considerably among PlaF and hydrolytic
88 enzymes (hydrolase HsaD from *Mycobacterium tuberculosis*, PDB ID 2VF2;² lipase LipA from
89 *P. aeruginosa*, PDB ID 1EX9;³ and hydrolase CarC from *Janthinobacterium* sp. strain J3, PDB
90 ID 1J11⁴) with structurally conserved α/β -hydrolase domains (light blue).

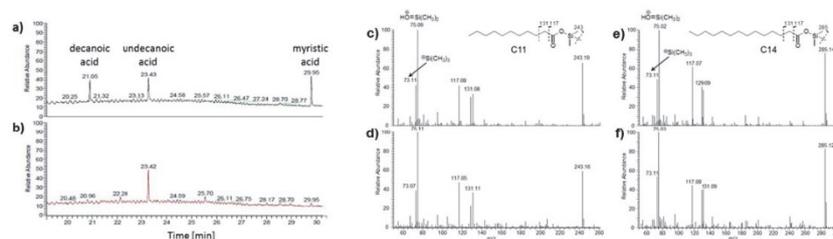


91

92

93 **Fig. S6:** a) Ribbon representation of the dimer structure colored according to B-factor. The
94 lid domain and the N-terminal helices show significantly higher B-factors (color spectrum
95 white - low B-factor, to red - high B-factor). The average B-factors of the α TM1 helix and the
96 lid domain in PlaF dimer are ~ 74 and ~ 55 \AA^2 , respectively. **b)** Thermal ellipsoid
97 representation of the dimer scaled by the B factors combined with TLS (translation-libration-
98 screw-rotation model) displays comparatively higher B-factors in the lid domain in molecule
99 B and in the α TM1 helix in molecule A. The figure was prepared using CCP4mg⁵.

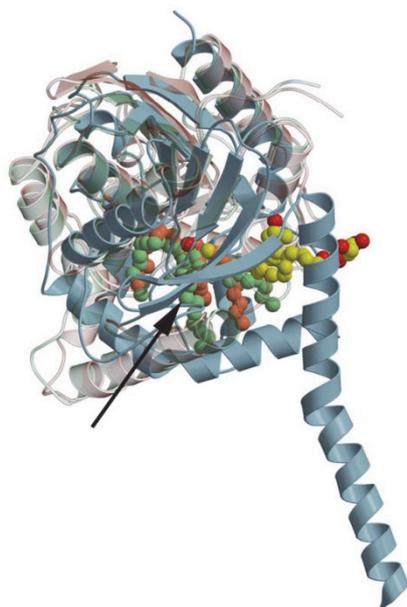
100



101

102

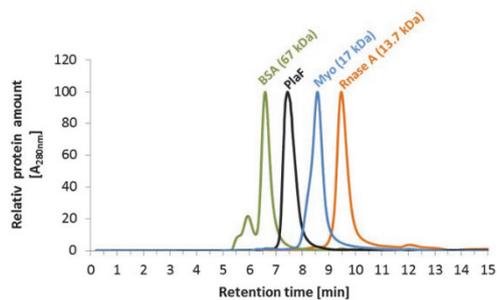
Fig. S7: Identification of fatty acid ligands co-purified with PlaF. Fatty acids were extracted
 103 from purified PlaF samples with organic solvent followed by silylation and gas
 104 chromatographic separation with mass spectrometric detection (GC-MS). GC
 105 chromatograms of pure C10, C11, and C14 fatty acids as standards (a) were compared to
 106 PlaF extracts (b). Mass spectrometric analysis of compounds with retention times
 107 23.43/23.42 min and 29.95 min for fatty acid standards (c and e) and PlaF extracts (d and f)
 108 revealed the presence of undecanoic and myristic acid trimethylsilylesters, respectively. The
 109 chemical structure of undecanoic and myristic acid trimethylsilylester and characteristic
 110 fragments⁶ (molecular weights in Da are indicated above) identified in mass spectra are
 111 shown.



112

113

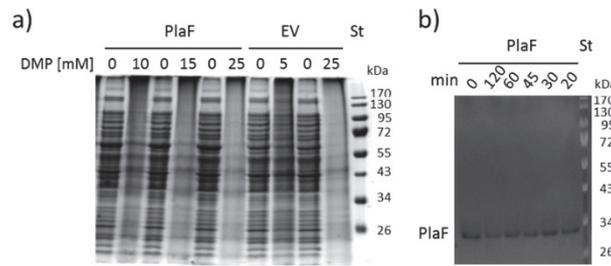
114 **Fig. S8:** The superposition of the α/β hydrolase domains. Ribbon representations of PlaF
115 (blue), *P. aeruginosa* lipase LipA (PDB ID 1EX9; green), *P. aeruginosa* lipase LipA (PDB ID 5LIP;
116 coral) were included in the analysis using the secondary structure matching program⁷ with
117 residues from one protein chain in each case. For clarity, the structures of 1EX9 and 5LIP are
118 made transparent. The ligands are shown as filled spheres and colored as follows: oxygen
119 red; carbon of OG and MYR in PlaF in yellow, carbon of OCP (R_C -(R_P , S_P)-1,2-
120 dioctylcarbamoylglycero-3-O-octylphosphonate) inhibitor in 5LIP in orange and carbon of R_C -
121 (R_P , S_P)-1,2-dioctylcarbamoyl-glycero-3-O-*p*-nitrophenyl octylphosphonate in 1EX9 in green.
122



123

124 **Fig. S9: Size exclusion chromatography of PlaF showed a monomer.** PlaF (10 μ l, 2 mg/ml)
125 purified with OG and standard proteins, bovine serum albumin (BSA), equine myoglobin
126 (Myo), and bovine ribonuclease A (RNase A) were separately analyzed using Biosep-SEC-
127 S2000 column. Proteins were detected by measuring absorbance at $\lambda = 280$ nm, and values
128 were normalized to the maximal $A_{280\text{nm}}$ of each protein.

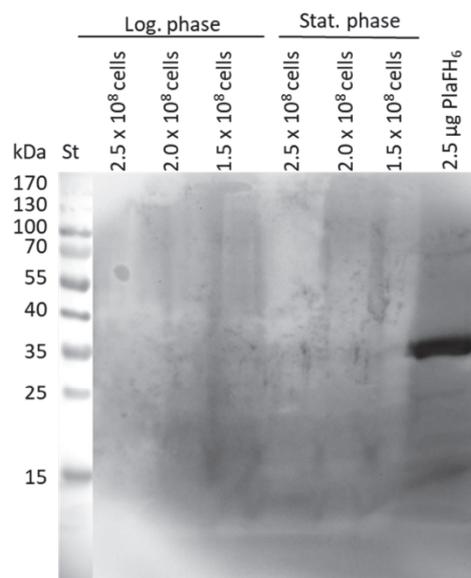
129



130

131 **Fig. S10: *In vivo* crosslinking and in-gel activity of PlaF.** a) The SDS-PAGE analysis of *in vivo*
132 cross-linking samples of *P. aeruginosa* strain carrying pBBR1mcs-3 (empty vector control, EV)
133 or *p-plaF* obtained in experiment shown in Fig. 5a. b) In-gel esterase activity of purified DMP
134 cross-linked PlaF samples obtained in the experiment shown in Fig. 5b. Esterase activity
135 monitored under UV light using the fluorescent substrate 4-methylumbelliferyl butyrate.
136 Molecular weights of protein standard (St) in kDa are indicated.

137

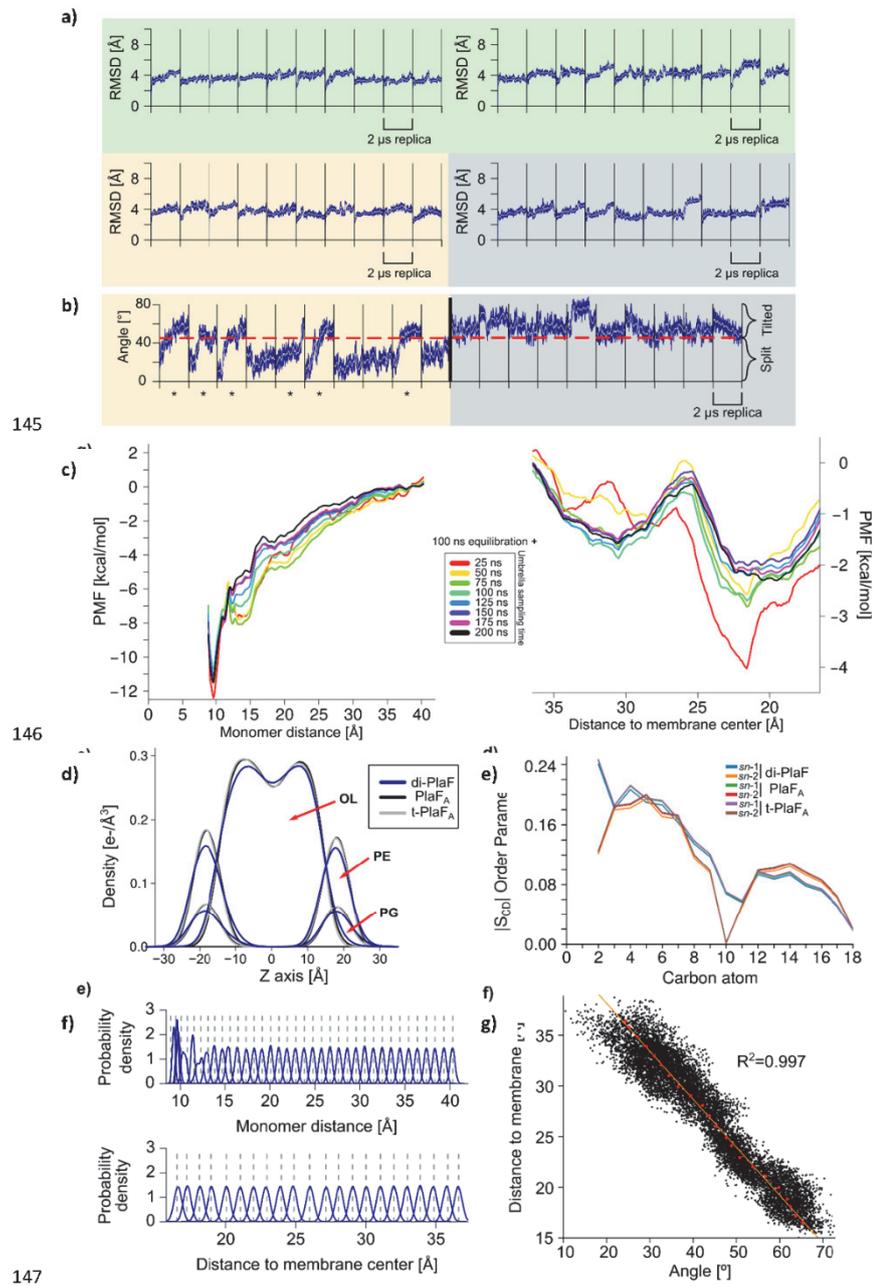


138

139 **Fig. S11:** Detection of PlaF by western blotting with anti-PlaF antiserum. *P. aeruginosa* cells
140 were cultivated until logarithmic (OD_{580nm} 1) and stationary phase (OD_{580nm} 3) cells were
141 harvested and lysed by SDS-PAGE sample buffer. Purified PlaF (2.5 μ g) was used as a positive
142 control.

143

144



145

146

147

148 **Figure S12: Structural variations, membrane parameters, and tests for PMF convergence. a) Root**
149 **mean square deviations during 10 independent, unbiased MD simulations of I)** the di-PlaF, where
150 each chain was measured independently (green, left and right), II) monomeric PlaF started in the
151 PlaF_A configuration (beige), and III) monomeric PlaF started in the tilted configuration (t-PlaF_A, gray),
152 computed with respect to the respective initial structure. Most simulations reached a plateau at ~
153 4 Å. **b) Time course of the orientation of monomeric PlaF starting from the PlaF_B (left) and t-PlaF_B**
154 **(right) configurations.** Six out of ten PlaF_B replicas show tilting of the monomer in less than 2 μs,
155 while all ten replicas that started in the t-PlaF_B configuration stayed tilted, similar to simulations
156 started from PlaF_A and t-PlaF_A. **c) Convergence of the PMFs for dimer separation (left) and**
157 **monomer tilting (right).** The plots show PMFs computed every 25 ns of umbrella simulations for
158 each window; the first 100 ns of umbrella simulations were considered equilibration phase and
159 removed. **d) Electron density profiles of membrane components averaged over 10 independent,**
160 **unbiased MD simulations of di-PlaF, PlaF_A or t-PlaF_A configurations** for the phospholipid head
161 groups (PE and PG) and oleic acid tails (OL). The obtained shapes correspond with those generally
162 found by experiment and MD simulations for biomembranes.⁸⁻¹⁰ **e) S_{CD} order parameter of the lipid**
163 **phase averaged over 10 independent, unbiased MD simulations of di-PlaF, PlaF_A or t-PlaF_A**
164 **configurations.** The carbon atoms are numbered according to their position in the phospholipid tail.
165 A clear dip is seen at the unsaturated position of the oleic acid tail, and the overall shape resembles a
166 structured membrane bilayer, as previously described^{8,11}. **f) Distribution of reaction coordinate**
167 **values obtained by umbrella sampling of dimer separation (top) and monomer tilting (bottom).**
168 The dashed lines represent the restrained distance used for each window. In both cases, a force
169 constant of 4 kcal mol⁻¹ Å⁻² was used, obtaining distributions with a median overlap of 8.2% and 8.6%,
170 respectively. For details, see the main text. **g) Distance to membrane center versus tilting angle.** The
171 scatter plot shows the distance of the COM of residues 33 to 37 to the membrane center *versus* the
172 tilting angle during the first microsecond of MD simulations of the tenth replica, starting from the
173 PlaF_A configuration ($R^2 = 0.997$, $p < 0.001$). The red dots represent the structures used as starting
174 conformations for calculating the PMF of the tilting process.

175

176 **Supplementary tables**

177

178 **Table S1:** All phospholipid species identified in *P. aeruginosa* PA01 and $\Delta plaF$ by Q-TOF

179 MS/MS.

180

181 [→ Excel file "Table S1 All phospholipid species"](#)

182 *Phospholipid nomenclature XX:Y. XX, the sum of carbon atoms in fatty acids bound to

183 phospholipid; Y, the number of double bonds in fatty acids bound to phospholipid.

184 **Table S2:** Phospholipid species significantly differentially abundant in *P. aeruginosa* wild-
185 type and Δ *plaF*.

Phospholipid*	<i>P. aeruginosa</i> PAO1	<i>P. aeruginosa</i> Δ <i>plaF</i>	Significance t-test	Shapiro-Wilk	
	nmol/mg(PL) \pm SD	nmol/mg(PL) \pm SD		WT	Δ <i>plaF</i>
accumulated in <i>P. aeruginosa</i> Δ<i>plaF</i>					
PE 35:1	116.458 \pm 23.489	178.208 \pm 8.242	0.05	0.05	0.75
PG 35:1	30.600 \pm 3.386	43.127 \pm 5.090	0.02	0.10	0.28
PE 36:2	40.806 \pm 5.577	53.303 \pm 2.930	0.02	0.26	0.42
PE 37:1	15.261 \pm 2.491	21.991 \pm 2.436	0.02	0.35	0.25
PE 38:2	4.056 \pm 2.041	9.715 \pm 0.923	0.01	0.18	0.26
PE 35:0	10.212 \pm 1.780	15.033 \pm 0.495	0.01	0.24	0.20
PE 38:8	4.263 \pm 1.129	7.386 \pm 0.716	0.01	0.06	0.30
PG 38:2	1.244 \pm 0.444	3.974 \pm 0.407	0.00	0.43	0.19
PG 38:3	1.690 \pm 0.370	3.694 \pm 0.669	0.01	0.39	0.93
PC 35:1	2.028 \pm 0.206	3.600 \pm 0.217	0.00	0.99	0.36
PG 37:1	1.069 \pm 0.311	1.927 \pm 0.317	0.02	0.95	0.49
PE 38:1	0.557 \pm 0.229	1.129 \pm 0.255	0.03	0.46	0.68
PE 37:2	0.414 \pm 0.078	0.939 \pm 0.202	0.01	0.15	0.48
PC 35:2	0.630 \pm 0.101	0.961 \pm 0.092	0.01	0.77	0.88
PG 38:1	0.401 \pm 0.109	0.602 \pm 0.065	0.04	0.71	0.39
PC 37:2	0.219 \pm 0.019	0.374 \pm 0.085	0.05	0.83	0.94
depleted in <i>P. aeruginosa</i> Δ<i>plaF</i>					
PE 38:3	0.090 \pm 0.027	0.175 \pm 0.019	0.01	0.31	0.31
PC 29:2	0.025 \pm 0.008	0.010 \pm 0.004	0.05	0.58	0.52
PE 21:1	0.037 \pm 0.007	0.016 \pm 0.008	0.02	0.69	0.20
PC 32:4	0.032 \pm 0.008	0.010 \pm 0.003	0.01	0.73	0.86
PC 35:3	0.036 \pm 0.009	0.013 \pm 0.004	0.02	0.23	0.73
PE 25:2	0.042 \pm 0.014	0.017 \pm 0.005	0.05	0.51	0.22
PG 23:0	0.070 \pm 0.019	0.028 \pm 0.013	0.03	0.57	0.26
PG 34:5	0.081 \pm 0.021	0.020 \pm 0.011	0.01	0.86	0.50
PA 30:1	0.107 \pm 0.028	0.031 \pm 0.001	0.02	0.06	-
PG 27:2	0.156 \pm 0.066	0.015 \pm 0.004	0.03	0.93	0.27
PG 30:6	0.289 \pm 0.062	0.123 \pm 0.033	0.01	0.13	0.16
PI 37:2	0.307 \pm 0.025	0.134 \pm 0.080	0.03	0.14	0.90
PA 28:1	0.291 \pm 0.105	0.095 \pm 0.008	0.05	0.53	0.63
PG 36:6	0.426 \pm 0.081	0.219 \pm 0.058	0.01	0.56	0.97
PE 30:1	0.801 \pm 0.134	0.434 \pm 0.086	0.01	0.27	0.87
PG 32:1	1.220 \pm 0.341	0.588 \pm 0.091	0.04	0.99	0.75
PE 36:4	4.591 \pm 0.689	2.244 \pm 0.384	0.00	0.79	0.42
PE 32:1	12.519 \pm 2.364	7.502 \pm 1.200	0.03	0.57	0.99
PE 34:2	51.967 \pm 4.764	38.784 \pm 6.733	0.04	0.25	0.15

186 *Phospholipid nomenclature XX:Y; XX, the sum of carbon atoms in fatty acids bound to

187 phospholipid; Y, the number of double bonds in fatty acids bound to phospholipid.

188

189 **Table S3:** Amount of GPL quantified by Q-TOF-MS/MS analysis.

	GPL amount [nmol/mg(GPL)]		
	Total*	Accumulated**	Depleted**
<i>P. aeruginosa</i> wild-type	959.7 ± 37.6	229.9	73.1
<i>P. aeruginosa</i> ΔplaF	966.1 ± 95.3	345.9	50.4

190 * Molecular species listed in table S2.

191 ** Molecular species listed in table S3.

192

193

194

195 **Table S4:** Properties of the cultures used for lipid extraction.

	<i>P. aeruginosa</i> PA01				<i>P. aeruginosa</i> ΔplaF				Significance t-test
	1	2	3	4	1	2	3	4	
Biological replicate									-
Optical density [OD_{580nm}]	4.66	5.11	4.48	4.58	5.08	4.93	5.02	5.22	0.08
mg(lipid)	3.04	2.53	2.75	2.94	2.99	3.28	3.09	3.10	0.07

196

197

198 **Table S5:** List of interactions involving the ligand molecules.

Ligand Atom	Distance / Å	Interacting residue in PlaFA Atom	Residue	#	Ligand Atom	Distance / Å	Interacting residue in PlaFB Atom	Residue	#
MYR					11A				
O1	3.28	OD1	ASN	136	C3	3.62	OD1	ASN	77
O1	3.47	O	HOH	113	C3	3.34	O	HOH	90
O1	3.43	OGB	SER	137	C4	3.51	C8'	BOG	502
C1	3.64	O	PHE	71	C4	3.65	O	HOH	90
C3	3.20	OD1	ASN	77	C5	3.57	C5'	BOG	502
C3	3.89	CB	ALA	73	C5	3.74	C8'	BOG	502
C3	3.67	CG1	VAL	287	C6	3.04	O	HOH	90
C4	3.47	OD1	ASN	77	C7	3.54	O	HOH	90
C7	3.66	ND2	ASN	77	C10	3.67	CA	ARG	31
C8	3.83	C7'	BOG	502	C10	3.76	CG	ARG	31
C10	3.11	O	HOH	166	C10	3.64	CD1	LEU	79
O2	3.43	O	PHE	71	O1	3.12	ND2	ASN	77
OG					IPAS03				
O1	3.48	NH2	ARG	80	O1	2.99	O2	IPA	504
C5'	3.73	CD1	LEU	206	C	3.45	O2	IPA	504
C8'	3.85	CE2	PHE	200	O	3.39	O2	IPA	504
O6	2.58	OE2	GLU	34	C1	3.56	OD1	ASN	77
O6	3.56	CD	GLU	34	C2	3.62	OD1	ASN	77
C3'	3.73	NH2	ARG	80	IPAS03				
C3'	3.69	O	HOH	28	C1	3.16	OD1	ASN	136
C1'	3.63	CG1	VAL	30	C1	3.49	O	HOH	22
C7'	3.72	CE2	PHE	200	C1	3.25	OGB	SER	137
C7'	3.74	CZ	PHE	200	C2	3.28	O	PHE	71
C7'	3.83	C8	MYR	500	O2	3.45	C	11A	501
C8'	3.88	CD	GLN	203	O2	3.39	O	11A	501
C8'	3.56	CG	GLN	203	O2	2.99	O1	11A	501
C8'	3.62	NE2	GLN	203	C3	3.75	CD2	HIS	286
C8'	3.57	O	PRO	204	C3	3.69	NE2	HIS	286
O4	3.42	O	HOH	105	C3	3.32	O	HOH	171
C2	3.67	O	PRO	205	OG				
O2	2.96	O	PRO	205	O1	2.99	NH2	ARG	80
O2	3.23	CD	PRO	207	C5'	3.82	CD1	LEU	206
O2	3.51	C	PRO	205	C8'	3.87	CE2	PHE	200
O2	3.70	CG	PRO	207	O6	2.52	OE2	GLU	34
C6	3.67	O	HOH	49	O6	3.42	CD	GLU	34
O6	3.42	CG1	VAL	33	C3'	3.78	C3	IPA	503
					C2'	3.73	CD2	LEU	210
					C4'	3.62	NH1	ARG	80
					O1	3.36	O	HOH	147
					C5'	3.57	C5	11A	501
					C6'	3.81	CB	PRO	204
					C1	3.70	NH2	ARG	80
					C8'	3.51	C4	11A	501
					C8'	3.74	C5	11A	501
					O5	3.14	OE2	GLU	34
					C5	3.90	CG1	VAL	30
					C6	3.43	OE2	GLU	34
					O6	3.68	CG	GLU	34

van-der-waals
electrostatic
hydrogen
ionic

199

200 **Table S6:** Structure-based sequence alignment of PlaF with its structural homologs*
201 identified using DALI server.¹²
202 *Yellow are highlighted homologs containing the sequence in TM-JM helix region of PlaF.
203 → **Excel file:** Table S6 - PlaF sequence alignment
204
205
206 **Table S7:** PlaF-B structural homologs identified using DALI server.¹²
207 → **Excel file:** Table S7 – PlaF-B structural homologs
208
209
210 **Table S8:** PlaF-A structural homologs identified using DALI server.¹²
211 → **Excel file:** Table S8 – PlaF-A structural homologs

212 **Table S9:** List of interactions involving the dimer interface.

Source atoms			Target atoms			Distance (Å)
Leu	5A	CG	Val	9B	CG2	3.95
Leu	5A	CD1	Val	9B	CG2	4.49
Leu	8A	CG	Val	9B	CG1	3.87
Leu	8A	CD2	Val	9B	CG1	3.89
			Val	9B	CG2	4.43
Val	9A	CA	Ala	13B	CB	4.42
Val	9A	C	Ala	13B	CB	4.45
Val	9A	O	Ala	13B	CA	3.97
			Ala	13B	CB	3.66
Val	9A	CG1	Leu	12B	C	4.09
			Leu	12B	O	4.10
			Ala	13B	N	3.94
			Ala	13B	CA	3.80
			Ala	13B	CB	4.38
			Leu	12B	CB	4.41
			Ala	16B	CB	4.17
Val	9A	CG2	Val	9B	O	3.88
			Leu	12B	CB	4.33
			Val	9B	CA	4.45
			Val	9B	CG1	4.13
Leu	12A	C	Ala	13B	CB	4.41
Leu	12A	CB	Ala	13B	CB	3.76
Leu	12A	CD1	Ala	13B	CB	4.48
			Leu	10B	CD2	3.62
			Val	9B	CG1	3.84
Ala	13A	N	Ala	13B	CB	4.17
			Ala	13B	O	4.44 *
Ala	13A	CA	Gly	17B	CA	4.43
			Ala	13B	C	4.46
			Ala	13B	O	3.72
Ala	13A	C	Phe	21B	CE2	4.14
Ala	13A	O	Phe	21B	CD2	3.84
			Phe	21B	CE2	3.42
			Phe	21B	CZ	4.45
			Gly	17B	CA	4.24
Ala	13A	CB	Ala	16B	C	4.38
			Gly	17B	N	3.86
			Gly	17B	CA	4.02

			Ala	13B	C	4.38
			Ala	13B	O	3.51
			Ala	16B	CB	4.17
Val	14A	N	Phe	21B	CE2	4.27
Val	14A	CA	Phe	21B	CE2	3.66
			Phe	21B	CZ	4.06
Val	14A	C	Phe	21B	CE2	4.13
			Phe	21B	CZ	4.29
Val	14A	O	Phe	21B	CE2	3.78
			Phe	21B	CZ	3.63
Gly	17A	N	Phe	21B	CD2	4.27
			Phe	21B	CE2	4.07
Gly	17A	CA	Phe	21B	CG	3.77
			Phe	21B	CD2	3.54
			Phe	21B	CD1	3.97
			Phe	21B	CE2	3.52
			Phe	21B	CZ	3.75
			Phe	21B	CE1	3.96
Gly	17A	C	Phe	21B	CG	4.43
			Phe	21B	CD2	4.45
			Phe	21B	CD1	4.13
			Phe	21B	CE2	4.19
			Phe	21B	CZ	3.89
			Phe	21B	CE1	3.86
Gly	17A	O	Phe	21B	CD1	4.33
			Phe	21B	CE1	4.25
Val	18A	N	Phe	21B	CE2	4.29
			Phe	21B	CZ	3.78
			Phe	21B	CE1	4.00
Val	18A	CG2	Phe	21B	CZ	4.15
			Phe	21B	CE1	4.47
Phe	21A	CE1	Val	22B	CG2	4.38
Phe	21A	CZ	Val	22B	CA	4.23
			Val	22B	CB	4.42
			Thr	25B	OG1	3.95
			Val	22B	CG2	3.69
Phe	21A	CE2	Val	22B	CA	3.68
			Val	22B	CB	4.28
			Phe	21B	C	4.02
			Phe	21B	O	3.93
			Val	22B	N	3.93

			Thr	25B	CB	4.15
			Thr	25B	OG1	3.03
			Thr	25B	CG2	4.24
			Val	22B	CG2	3.75
Phe	21A	CD2	Phe	21B	C	4.23
			Phe	21B	O	4.28
			Val	22B	N	4.36
			Thr	25B	CB	4.44
			Thr	25B	OG1	3.69
			Thr	25B	CG2	4.08
			Val	22B	CG2	4.48
			Phe	21B	CB	3.90
Val	22A	CG2	Thr	25B	CG2	3.87
			Phe	21B	CD1	3.68
			Phe	21B	CE1	3.74
Thr	25A	CB	Thr	25B	CB	4.46
			Thr	25B	OG1	4.31
Thr	25A	OG1	Thr	25B	CB	3.88
			Thr	25B	OG1	4.16 *
			Thr	25B	CG2	4.43
Thr	25A	CG2	Thr	25B	CB	4.33
			Thr	25B	OG1	3.78
Ser	29A	O	Val	33B	CG2	3.97
Ser	29A	CA	Val	33B	CG2	4.26
Ser	29A	CB	Val	33B	CG2	4.27
			Thr	32B	CG2	3.75
			Ser	29B	CB	4.43
			Thr	32B	CB	4.13
Ser	29A	OG	Ser	29B	CA	3.64
			Ser	29B	CB	3.15
			Ser	29B	OG	3.55 *
			Ser	29B	C	4.29
			Ser	29B	O	4.09 *
			Thr	32B	CB	4.45
Ser	29A	CA	Val	33B	CG2	4.26
Ser	29A	CB	Val	33B	CG2	4.25
			Thr	32B	CG2	3.77
			Ser	29B	CB	4.40
			Thr	32B	CB	4.13
Ser	29A	OG	Thr	32B	CG2	3.52
			Thr	32B	CB	4.13

Thr	32A	CB	Val	33B	CG2	4.16
Thr	32A	CG2	Val	33B	CG1	4.05
			Val	33B	CG2	3.93
Val	33A	CB	Gly	36B	CA	4.45
Val	33A	CG1	Gly	36B	O	3.58
			Leu	37B	N	3.90
			Gly	36B	CA	3.66
			Gly	36B	C	3.46
Val	33A	CG2	Val	33B	C	4.18
			Val	33B	O	3.81
			Leu	37B	N	4.35
			Gly	36B	CA	3.95
			Gly	36B	C	4.32
			Thr	32B	C	4.47
			Val	33B	N	4.37
			Val	33B	CA	3.58
			Val	33B	CB	4.29
			Val	33B	CG1	4.38
			Val	33B	CG2	4.27
			Thr	32B	O	3.90
Val	33A	O	Leu	37B	CD2	4.16
Leu	37A	CG	Leu	37B	CD2	4.16
Leu	37A	CD1	Gly	36B	O	3.97
Leu	37A	CD2	Arg	83B	NH2	4.06
			Leu	37B	CD2	4.13
			Leu	37B	O	3.79
			Leu	37B	CA	4.26
			Leu	37B	C	4.44

213 #The cut off value is 4.5 Å (includes van der Waals interaction).

214 All contacts are part of the N-terminal TM-JM helix.

215 **Table S10:** List of interactions[#] involving the catalytic triad residues S137, D258 and H286.

<u>Source atoms</u>	<u>Target atoms</u>	<u>Distance (Å)</u>
<i>Monomer A</i>		
Ser 137A N	Asn 136A ND2	3.39
	Ile 160A O	3.20
	Asp 161A O	3.04
Ser 137A O	Ala 163A N	3.01
	Gly 139A N	3.17
	Gly 140A N	2.89
	His 141A N	2.97
<u>conformation A</u>		
Ser 137 A OG	His 286A NE2	3.04
	HOH 229S O	2.23
<u>conformation B</u>		
Ser 137 A OG	MYR 500A O1	3.43
	HOH 229S O	2.52
Asp 258A N	Gly 255A O	3.13
Asp 258A OD1	Leu 261A O	3.25
	Leu 261A N	2.93
	His 286A ND1	3.12
	Arg 259A N	3.00
	Val 260A N	2.75
Asp 258A OD2	Trp 254A NE1	3.49
	Leu 261A O	3.08
	Asp 161A OD2	2.63
	His 286A ND1	2.72
His 286A N	Val 199A O	3.08
His 286A ND1	Asp 258A OD1	3.12
	Asp 258A OD2	2.72
His 286A NE2	Ser 137A OG	3.04
His 286A O	Asn 136A ND2	3.25
<i>Monomer B</i>		
Ser 137B N	Asn 136B ND2	3.35
	Ile 160B O	3.16
	Asp 161B O	3.05
Ser 137B O	Ala 163B N	3.02

	Gly	140B	N	2.89
	His	141B	N	3.02
	Gly	139B	N	3.19
	<u>conformation A</u>			
Ser	137B	OG		
	His	286B	NE2	2.83
	HOH	171S	O	2.62
	<u>conformation B</u>			
Ser	137B	OG		
	IPA	504B	C1<<<	3.25
	HOH	171S	O	2.80
Asp	258B	N	Gly 255B O	3.08
Asp	258B	OD1	Arg 259B N	2.97
	Leu	261B	N	2.94
	Leu	261B	O	3.29
	Val	260B	N	2.77
	His	286B	ND1	3.12
Asp	258B	OD2	Trp 254B NE1	3.43
	Asp	161B	OD2	2.67
	Leu	261B	O	3.22
	His	286B	ND1	2.72
His	286B	N	Val 199B O	2.97
His	286B	ND1	Asp 258B OD2	2.72
	Asp	258B	OD1	3.12
His	286B	NE2	Ser 137B OG	2.83
His	286B	O	Asn 136B ND2	3.16

216 [#]The cut-off value for hydrogen bond selection used is ≤ 3.5 Å.

217 **Table S11:** Residues lining the active site cavity and their interactions with ligands.

Residue	Interacting ligand
L27	-
A28	-
V30	OG(A,B)
R31	11A
E34	OG(A,B)
G70	-
F71	MYR, IPA503
G72	-
A73	MYR
D74	-
D76	-
N77	MYR, 11A
W78	-
L79	MYR, 11A
R80	OG(A,B)
F81	-
N136	MYR, IPA503
S137*	MYR, IPA503
M138	-
H141	-
A163	-
F174	-
L173	-
L184	-
V185	-
V186	-
F192	-
L195	-
L196	-
V199	-
F200	OG(A,B)
N203	OG(A)
P204	OG(A,B)
L206	OG(A,B)
L210	OG(B)

L214	MYR
R217	-
A218	-
S222	-
N225	-
F229	-
L232	-
V260	-
L261	-
H286*	MYR, IPA503
M289	MYR
V287	-
V290	-

218

*active site residues

219 **Table S12:** Michaelis-Menten constants for inhibition of PlaF with decanoic acid (FA C10).

c(FA C10) [mM]	K_m [mM]*	v_{max} [U/mg]*
0.0	0.17±0.02	899.5±37.2
0.5	0.23±0.03	916.0±39.4
1.5	0.25±0.03	830.9±40.0
2.5	0.34±0.04	717.8±35.1
5.0	0.45±0.05	512.2±26.6
7.5	0.66±0.04	390.8±13.7

220 * Results are mean ± S.D. of three experiments each measured with three samples.

221

222

223

224

225

226 **Table S13:** Average 2D-RMSD_{all atom} of residues 25 to 315 of the structures sampled along MD

227 trajectories.^[a]

	di-PlaF _A ^[b]	di-PlaF _B ^[b]	PlaF _A ^[c]	t-PlaF _A ^[d]	PlaF _B ^[c]	t-PlaF _B ^[d]
di-PlaF _A ^[b]	3.42 ± 0.59	3.96 ± 0.69	3.63 ± 0.58	3.60 ± 0.61	3.94 ± 0.57	3.91 ± 0.60
di-PlaF _B ^[b]		4.01 ± 0.81	4.05 ± 0.68	4.05 ± 0.72	4.29 ± 0.70	4.23 ± 0.71
PlaF _A ^[c]			3.59 ± 0.60	3.71 ± 0.63	4.08 ± 0.58	4.02 ± 0.64
t-PlaF _A ^[d]				3.58 ± 0.72	4.05 ± 0.61	3.93 ± 0.65
PlaF _B					4.17 ± 0.76	4.21 ± 0.62
t-PlaF _B						3.99 ± 0.80

228 ^[a] RMSD values in Å, mean ± S.D., were computed in a pair-wise manner for respective
229 structures sampled every ns along the MD trajectories.

230 ^[b] PlaF molecules in dimeric form starting from the crystal structure.

231 ^[c] PlaF_A and PlaF_B obtained from the dimeric form by removal of the opposite chain.

232 ^[d] PlaF_A and PlaF_B in the tilted monomeric form.

233 **References**

- 234 1. Martinez-Garcia, E. & de Lorenzo, V. Engineering multiple genomic deletions in Gram-
235 negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida*
236 KT2440. *Environ Microbiol* **13**, 2702-16 (2011).
- 237 2. Lack, N.A. et al. Characterization of a carbon-carbon hydrolase from *Mycobacterium*
238 *tuberculosis* involved in cholesterol metabolism. *J. Biol. Chem.* **285**, 434-443 (2010).
- 239 3. Nardini, M., Lang, D.A., Liebeton, K., Jaeger, K.E. & Dijkstra, B.W. Crystal structure of
240 *Pseudomonas aeruginosa* lipase in the open conformation. The prototype for family I.1 of
241 bacterial lipases. *J. Biol. Chem.* **275**, 31219-31225 (2000).
- 242 4. Habe, H. et al. Crystal structure of a histidine-tagged serine hydrolase involved in the
243 carbazole degradation (CarC enzyme). *Biochem Biophys Res Commun* **303**, 631-9 (2003).
- 244 5. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D*
245 *Biol Crystallogr* **67**, 235-42 (2011).
- 246 6. Diekman, J., Thomson, J.B. & Djerassi, C. Mass spectrometry in structural and stereochemical
247 problems. CLXXIII. The electron impact induced fragmentations and rearrangements of
248 trimethylsilyl esters of *w*-phenoxyalkanoic acids. *The Journal of Organic Chemistry* **34**, 3147-
249 3161 (1969).
- 250 7. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein
251 structure alignment in three dimensions. *Acta Crystallographica Section D-Biological*
252 *Crystallography* **60**, 2256-2268 (2004).
- 253 8. Dickson, C.J. et al. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and*
254 *Computation* **10**, 865-879 (2014).
- 255 9. Liu, Y. & Nagle, J.F. Diffuse scattering provides material parameters and electron density
256 profiles of biomembranes. *Physical Review. E: Statistical, Nonlinear, and Soft Matter Physics*
257 **69**, 040901 (2004).
- 258 10. Tristram-Nagle, S., Petrache, H.J. & Nagle, J.F. Structure and Interactions of Fully Hydrated
259 Dioleoylphosphatidylcholine Bilayers. *Biophysical Journal* **75**, 917-925 (1998).
- 260 11. Baul, U. & Vemparala, S. Influence of lipid composition of model membranes on
261 methacrylate antimicrobial polymer-membrane interactions. *Soft Matter* **13**, 7665-7676
262 (2017).
- 263 12. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**,
264 W545-549 (2010).
- 265
- 266

Publication IV

Publication IV: Calcium-promoted interaction between the C2- domain protein EHB1 and metal transporter IRT1 inhibits *Arabidopsis* iron acquisition

Imran Kahn, Regina Gratz, Polina Denezhkin, Stephan Schott-Verdugo, Kalina Angrand, Lara Genders, Rubek Merina Basgaran, Claudia Fink-Straube, Tzvetina Brumbarova, Holger Gohlke, Petra Bauer and Rumen Ivanov

Plant Physiology, 2019, 180(3), 1564–1581

DOI: 10.1104/pp.19.00163



Calcium-Promoted Interaction between the C2-Domain Protein EHB1 and Metal Transporter IRT1 Inhibits Arabidopsis Iron Acquisition¹

Imran Khan,^a Regina Gratz,^{a,b} Polina Denezhkin,^b Stephan N. Schott-Verdugo,^{e,g} Kalina Angrand,^b Lara Genders,^a Rubek Merina Basgaran,^a Claudia Fink-Straube,^c Tzvetina Brumbarova,^{a,b} Holger Gohlke,^{e,f} Petra Bauer,^{a,b,d} and Rumen Ivanov^{a,b,2,3}

^aInstitute of Botany, Heinrich-Heine University, D-40225 Düsseldorf, Germany

^bFormer address: Department of Biosciences-Plant Biology, Saarland University, D-66123 Saarbrücken, Germany

^cLeibniz Institute for New Materials, D-66123 Saarbrücken, Germany

^dCluster of Excellence on Plant Sciences, Heinrich-Heine University, D-40225 Düsseldorf, Germany

^eInstitute for Pharmaceutical and Medicinal Chemistry, Heinrich-Heine University, D-40225 Düsseldorf, Germany

^fJohn von Neumann Institute for Computing, Jülich Supercomputing Centre & Institute of Complex Systems, ICS-6: Structural Biochemistry, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany

^gCentro de Bioinformática y Simulación Molecular, Facultad de Ingeniería, Universidad de Talca, CL-3460000 Talca, Chile

ORCID IDs: 0000-0001-9692-5961 (I.K.); 0000-0002-8820-7211 (R.G.); 0000-0001-5828-1608 (P.D.); 0000-0003-0735-1404 (S.N.S.-V.); 0000-0002-3104-6795 (T.B.); 0000-0001-7909-4123 (R.I.).

Iron is a key transition element in the biosphere and is crucial for living organisms, although its cellular excess can be deleterious. Maintaining the balance of optimal iron availability in the model plant *Arabidopsis thaliana* requires the precise operation of iron import through the principal iron transporter IRON-REGULATED TRANSPORTER1 (IRT1). Targeted inhibition of IRT1 can prevent oxidative stress, thus promoting plant survival. Here, we report the identification of an IRT1 inhibitor, namely the C2 domain-containing peripheral membrane protein ENHANCED BENDING1 (EHB1). EHB1 interacts with the cytoplasmically exposed variable region of IRT1, and we demonstrate that this interaction is greatly promoted by the presence of calcium. We found that EHB1 binds lipids characteristic of the plasma membrane, and the interaction between EHB1 and plant membranes is calcium-dependent. Molecular simulations showed that EHB1 membrane binding is a two-step process that precedes the interaction between EHB1 and IRT1. Genetic and physiological analyses indicated that EHB1 acts as a negative regulator of iron acquisition. The presence of EHB1 prevented the IRT1-mediated complementation of iron-deficient *fet3fet4* yeast (*Saccharomyces cerevisiae*). Our data suggest that EHB1 acts as a direct inhibitor of IRT1-mediated iron import into the cell. These findings represent a major step in understanding plant iron acquisition, a process that underlies the primary production of bioavailable iron for land ecosystems.

Deprived of the ability to escape unfavorable conditions, plants have developed powerful mechanisms for adapting to their environment (Haak et al., 2017). Nutrient availability in the soil can be a limiting factor for plant growth and plants respond to variations in nutrient concentrations by a variety of developmental and physiological responses (Giehl and von Wirén, 2014; Briat et al., 2015). Of particular importance is the acquisition of iron, an element abundant in the soil but poorly available to plants (Guerinot and Yi, 1994; Wedepohl, 1995). As a part of many key biological processes, iron is indispensable for life; however, in high doses it can cause deleterious effects and even death. Therefore, its controlled acquisition from the soil is essential for the plant. *Arabidopsis thaliana* employs a reduction-based iron acquisition strategy (Römheld and Marschner, 1983; Brumbarova et al., 2015), in which

actively-solubilized iron is first reduced by the FERRIC REDUCTASE-OXIDASE2 (FRO2; Robinson et al., 1999) and subsequently imported from the apoplastic space across the plasma membrane of the root epidermal cells by the bivalent metal transporter IRON-REGULATED TRANSPORTER1 (IRT1; Eide et al., 1996; Vert et al., 2002; Fourcroy et al., 2016).

The system is responsive to the amounts of available iron and is transcriptionally induced upon iron limitation (Brumbarova et al., 2015). Genes encoding IRT1 proteins are present throughout land plants and green algae, and show tight coregulation with the rest of the organism's iron acquisition and homeostasis machinery (Hanikenne et al., 2005; Ivanov et al., 2012a; Urzica et al., 2012; Ivanov and Bauer, 2017). In addition to this, the *Arabidopsis* IRT1 protein was shown to undergo posttranslational regulation involving covalent

modifications and regulated trafficking toward and away from the plasma membrane (Barberon et al., 2011, 2014; Shin et al., 2013; Ivanov et al., 2014; Dubeaux et al., 2018). Under iron deficiency, plasma membrane-localized IRT1 is modified by the E3 ubiquitin ligase IRT1-DEGRADATION FACTOR1 (IDF1), which leads to its clathrin-mediated endocytosis (Barberon et al., 2011, 2014; Shin et al., 2013). IRT1 was shown to sense the abundance of its noniron metal substrates zinc, manganese, and cobalt. When these are present in excess, IRT1 can bind them and recruit the CBL-INTERACTING PROTEIN KINASE 23 (CIPK23), which phosphorylates IRT1's variable region. Phosphorylated IRT1 is ubiquitinated by IDF1, internalized and degraded (Dubeaux et al., 2018). Internalized IRT1 can also be recycled and resorted back to the plasma membrane. Two proteins, SORTING NEXIN1 (SNX1) and FYVE DOMAIN PROTEIN REQUIRED FOR ENDOSOMAL SORTING 1 (FREE1/FYVE1), have been implicated in this process (Barberon et al., 2014; Ivanov et al., 2014). In addition, FREE1/FYVE1 was shown to participate in the maintenance of the polar localization of IRT1 to the plasma membrane facing the rhizosphere (Barberon et al., 2014). A tight multilevel control of IRT1 activity counteracts uncontrolled iron acquisition that may otherwise overload the cellular capacity to safely store iron and result in extensive oxidative damage (Reyt et al., 2015; Le et al., 2016). Therefore, mechanisms that prevent the excessive entry of iron into the cell are essential to the survival of plants.

Members of the C2-domain abscisic acid-related (CAR) protein family have recently emerged as key regulators of plant stress responses. These proteins share homology with the Arf GTPase activating protein family but lack the N-terminal Zn-finger motif, characteristic of the Arf GTPase activating protein family, and only contain a C2 calcium- and lipid-binding domain (Knauer et al., 2011). A characteristic feature of CAR proteins is a family-specific stretch of 48 amino acids inserted into the C2 domain that connects the two four-stranded beta sheets (Rodriguez et al., 2014). This

domain, referred to as the "CAR signature domain," was shown to be involved in the protein-protein interaction between Arabidopsis CAR4, also named "AtGAP1," and the abscisic acid (ABA) receptors PYRABACTIN RESISTANCE 1 (PYR1)-LIKE1 (PYL1) and PYL6 (Rodriguez et al., 2014). Members of the CAR family have been characterized in rice (*Oryza sativa*) and Arabidopsis, and are involved in the responses to wounding, salinity, blue light and ABA, as well as in the root gravitropic response (Cheung et al., 2008, 2010, 2013; Knauer et al., 2011; Rodriguez et al., 2014). The Arabidopsis CAR-family member ENHANCED BENDING1 (EHB1), also known as "CAR6," was identified in a yeast two-hybrid screen as an interactor of the blue light response regulator NONPHOTOTROPIC HYPOCOTYL3. Physiological analysis revealed EHB1 as a negative regulator of blue light responses and further showed its involvement in root gravitropic responses (Knauer et al., 2011; Dümmer et al., 2016).

Our aim was to investigate the interactome of IRT1 for the identification of processes and molecular players involved in the regulation of IRT1 function. Using a yeast two-hybrid screen with IRT1's cytoplasmically exposed variable region, we identified EHB1 as an IRT1 protein interactor. EHB1 was able to bind a subset of phosphatidylinositol lipids. The EHB1-IRT1 and EHB1-lipid interactions were significantly promoted by the presence of calcium. Molecular simulations suggested that the membrane binding of EHB1 precedes the interaction between EHB1 and IRT1. Physiological analysis showed that EHB1 presence negatively affects the import of apoplastic iron into the root and that EHB1 directly affects IRT1-dependent iron acquisition. The data suggest a role of EHB1 in mediating calcium signals for the inhibition of iron acquisition.

RESULTS

IRT1 Interacts with EHB1

To identify proteins involved in the regulation of iron import from the rhizosphere, we looked for interactors of IRT1. We used the cytoplasmically exposed variable region of IRT1 (referred to as "IRT1_{vr}," residues 145–192), as it constitutes the key IRT1 scaffold for regulatory and modification inputs. IRT1_{vr} was used as bait in a yeast two-hybrid screen against an expression library prepared from complementary DNA (cDNA) of iron-deficient Arabidopsis roots. Among the colonies growing on the selection medium, IRT1_{vr} together with the C2 domain-containing protein EHB1 were found in 34% of the cases. EHB1 is a part of a 10-member protein family; however, none of the other members were identified in the screen. UBIQUITIN 10, a previously reported IRT1 interactor (Barberon et al., 2011), was also identified in this screen. We could verify the interaction in a targeted yeast two-hybrid assay using a recloned full-length EHB1 (Fig. 1A). In this case, only

¹This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation Project no. 267205415-SFB 1208 and projects A03 to H.G. and B05 to P.B.); German Academic Exchange Service (DAAD grant no. 91541023 to I.K.); Saarland University (61cl/Anschub2012/ivanov to R.I.); and Heinrich Heine University (SFF-F 2014/730-15 Ivanov to R.I.).

²Author for contact: rumen.ivanov@uni-duesseldorf.de.

³Senior author.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Rumen Ivanov (rumen.ivanov@uni-duesseldorf.de).

R.I. and P.B. designed the research; R.I. designed experiments; H.G. and S.N.S.-V. designed computations; I.K., R.G., P.D., S.N.S.-V., K.A., L.G., R.M.B., C.F.-S., T.B., and R.I. performed research; T.B., R.G., P.B., H.G., S.N.S.-V., and R.I. analyzed data; R.I. wrote the article with contributions by S.N.S.-V. and H.G.; R.G., T.B., P.B., and R.I. commented and corrected the article; H.G., P.B., and R.I. acquired funding.

www.plantphysiol.org/cgi/doi/10.1104/pp.19.00163

Khan et al.

the IRT1vr-EHB1 combination but none of the negative control cotransformations produced colonies on the triple selection medium, suggesting a specific interaction between IRT1vr and EHB1. The known SNX1 homodimerization (Pourcher et al., 2010) was used as a positive control (Fig. 1A). To verify the interaction further, we used the two fragments, together with several deletion versions, for an interaction test in plant cells using bimolecular fluorescence complementation (BiFC; Fig. 1, B and L; Grefen and Blatt, 2012). A positive BiFC signal indicated the IRT1vr-EHB1 interaction in this system (Fig. 1, C–E). The CAR signature domain was found to be the protein interaction interface for the EHB1 homolog CAR4 interaction with PYR/PYL receptors (Rodriguez et al., 2014). EHB1 protein lacking the CAR signature domain was not able to interact with IRT1vr in transformed cells (Fig. 1, F–H). On the other hand, the signature domain alone was sufficient for an interaction with IRT1vr (Fig. 1, I–K), suggesting that this domain mediates the interaction on the side of EHB1. To delimit the protein interaction on the side of IRT1vr, we created four different deletion constructs of IRT1vr that contained or were devoid of the previously described regulatory residues for metal binding, phosphorylation, and ubiquitination, taking into account the predicted secondary structure (Fig. 1L; Ivanov and Bauer, 2017). EHB1 was able to bind three IRT1vr fragments (Fig. 1, M–R and V–X) but not IRT1vr Δ 3 (Fig. 1, S–U), which lacks the predicted helical regions in proximity to transmembrane domains III and IV (Ivanov and Bauer, 2017). These data show that the interaction requires the part of IRT1vr flanking the transmembrane regions and the CAR signature domain of EHB1. The involvement of the EHB1 CAR signature domain in protein–protein interaction is thus consistent with previous reports (Rodriguez et al., 2014).

IRT1 Interaction with EHB1 Is Calcium-Dependent

Because EHB1 was able to bind to the cytoplasmically exposed variable region of IRT1, we tested its ability to interact with the full-length IRT1 protein in plant cells. For this, we expressed EHB1-HA and IRT1-GFP fusions and performed coimmunoprecipitation. The presence of EHB1-HA in the IRT1-GFP-containing sample was visible after anti-GFP immunoprecipitation (Fig. 2A), suggesting that the full-length proteins can form a complex. EHB1-HA was present as a double band, one at the expected size of 25 kD and another of ~35 kD, presumably representing a modified form of the protein. As EHB1 is a calcium-dependent protein due to the presence of a C2 domain, we hypothesized that calcium concentration might influence EHB1's capacity to bind IRT1. When the coimmunoprecipitation procedure was performed in the presence of 100- μ M Ca²⁺, we observed a dramatic increase of recovered EHB1-HA (Fig. 2B). The EHB1 protein lacking the CAR signature domain was not able to interact with the full-length IRT1-GFP, irrespective of the calcium availability (Fig. 2, C and D). This confirms the

importance of the CAR signature domain for the formation of the IRT1-EHB1 complex and shows that the strength of the interaction is dependent on the local calcium concentration.

EHB1 and IRT1 Colocalize at the Plasma Membrane

C2 domain proteins are described as peripheral membrane proteins and previous studies have suggested that CAR-family proteins localize partially at the plasma membrane (Cheung et al., 2010; Demir et al., 2013; Rodriguez et al., 2014). We expected that as an IRT1 interactor, EHB1 should also localize to the membrane system of plant cells. An EHB1-GFP fusion protein expressed in *Nicotiana benthamiana* epidermis cells resulted in a broad localization pattern with signals visible also in the nucleus (Fig. 2, E–G). We performed a control immunoblot, which revealed two bands, one with the expected size of 57 kD and one at 67 kD (Fig. 2H), likely corresponding to a modified EHB1-GFP form, as observed with EHB1-HA. No signal was observed in the 27-kD region that would suggest the existence of free GFP. We first investigated the colocalization of free GFP and the plasma membrane marker AHA1-mRFP (Caesar et al., 2011). Distinct, spatially-separated signals were visible in cells coexpressing the two proteins, suggesting that they differ in their subcellular localization (Fig. 2, I–K). This was confirmed by intensity-based colocalization scatterplot (Fig. 2L), and by plasmolyzing the cells in the presence of mannitol. After plasmolysis, AHA1-mRFP was seen in Hechtian strands, by which the plasma membrane remains attached to the cell wall, whereas GFP was absent from these structures (Fig. 2, M–O). Next, we tested the EHB1-GFP plasma membrane localization. In the cell periphery, it displayed a good colocalization with the AHA1-mRFP marker. We could observe additional EHB1-GFP intracellular signals, not visible for AHA1-mRFP (Fig. 2, P–R). This demonstrates the partial presence of EHB1 at the plasma membrane in plant cells. Intensity-based colocalization scatterplot confirmed the observation (Fig. 2S). Upon mannitol-induced plasmolysis, we found both EHB1-GFP and the AHA1-mRFP marker in Hechtian strands (Fig. 2, T–V), confirming the presence of EHB1-GFP at the plasma membrane. We then performed a colocalization analysis between EHB1-GFP and IRT1-mCherry. Signals at the cell periphery showed a high degree of colocalization (Fig. 2, W–Z). Together, the data suggest that EHB1 is partially localized at the plasma membrane where it colocalizes with IRT1.

EHB1 Can Bind Phosphoinositides

To understand whether lipids are targeted by EHB1, we expressed and purified recombinant StrepII-tagged EHB1 in *Escherichia coli* (Supplemental Fig. S1) to use in a lipid overlay assay. Probing a set of immobilized

1566

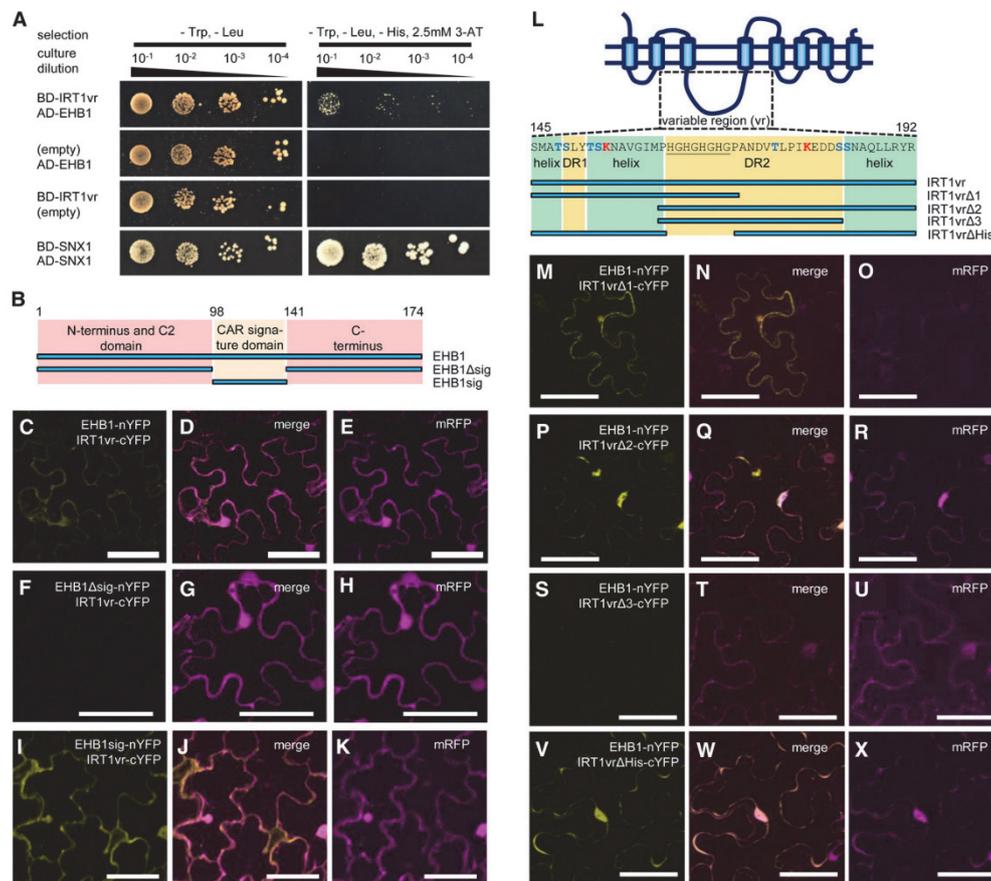


Figure 1. EHB1 interacts with the variable region of IRT1. **A**, Targeted yeast two-hybrid assay, showing the interaction between BD-IRT1vr (IRT1 variable region) and AD-EHB1. Growth on triple selection medium (right) indicates protein-protein interaction. **B**, Schematic representation of EHB1 protein and the two fragments used in the BiFC experiments testing interaction with IRT1vr. The position of the CAR signature domain, as defined in Rodriguez et al. (2014), together with the amino acid numbers, are indicated. **C** to **K**, Targeted BiFC experiment for verification of the interaction between EHB1 and its fragments (depicted in **B**) with the IRT1vr. Signal in the YFP channel indicates a reconstitution of a functional YFP protein as a consequence of an interaction. Signal in the RFP channel is used as a control, showing that the cell was transformed. Each of the shown protein combinations was tested a minimum of three times yielding comparable results. Bars = 50 μm. **L**, Schematic representation of the IRT1vr and the IRT1vr-derived fragments used in the BiFC experiments testing interaction with EHB1. The different predicted secondary structures within IRT1vr, as defined in Ivanov and Bauer (2017), are indicated. DR1 and DR2 indicate the predicted disordered regions. The two Lys residues, known as ubiquitination targets, are labeled in red. The known potential phosphorylation targets directing the interaction with IDFI E3 ubiquitin-ligase are shown in blue. The His-rich region, characteristic for the variable region of the ZIP-family transporters and involved in metal binding, is underlined. Numbers indicate the first and last amino acid of IRT1vr within IRT1 protein. **M** to **X**, Targeted BiFC experiment for verification of the interaction between the IRT1vr and its fragments (depicted in **L**) with EHB1. Signal in the YFP channel indicates a reconstitution of a functional YFP protein as a consequence of an interaction. Signal in the RFP channel is used as a control, showing that the cell was transformed. Each of the shown protein combinations was tested a minimum of three times yielding comparable results. Bars = 50 μm.

Khan et al.

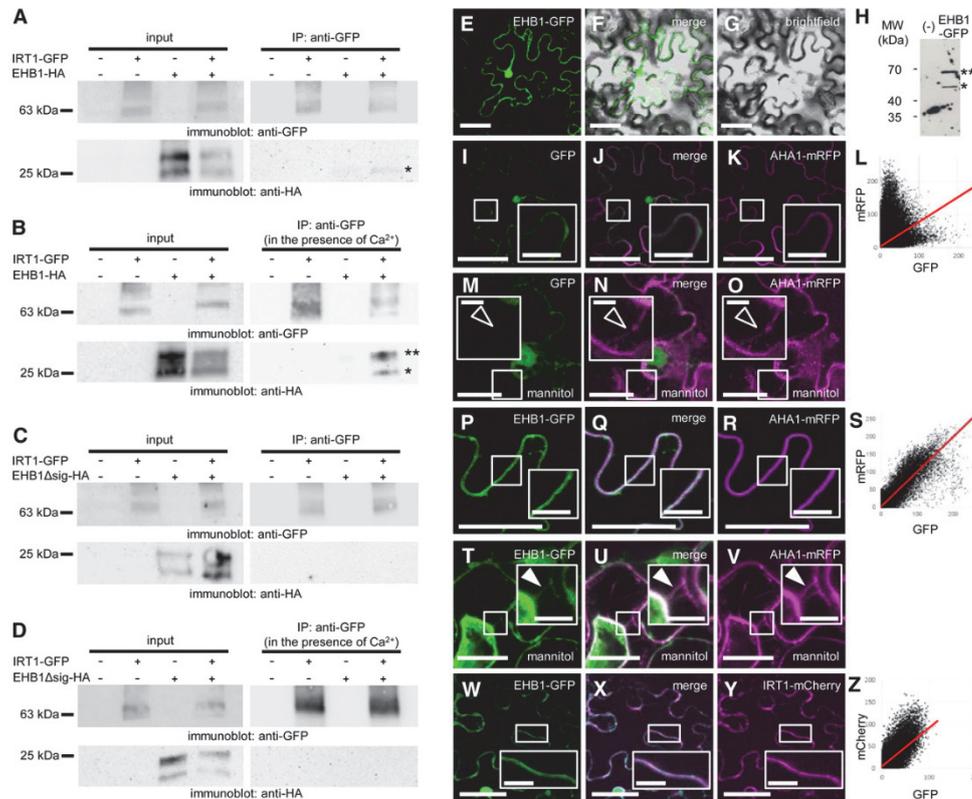


Figure 2. Interaction and subcellular localization of full-length IRT1 and EHB1. A, EHB1-HA coimmunoprecipitates with IRT1-GFP. IRT1-GFP, EHB1-HA or a combination of the two were expressed in *N. benthamiana* epidermis cells and used for anti-GFP immunoprecipitation. The samples before (left, input) and after (right, IP: anti-GFP) the procedure were tested by immunoblot analysis. Weak, yet specific, presence of EHB1-HA was found in the combined immunoprecipitated sample. Nontransformed samples were used as controls. The experiment was performed three times yielding comparable results. Asterisk indicates coimmunoprecipitated EHB1-HA. B, The same experiment as in (A) was performed in the presence of 100- μ M Ca²⁺. A marked increase of EHB1-HA signal in the combined sample could be observed, in comparison to (A). The experiment was performed three times yielding comparable results. Asterisks indicate coimmunoprecipitated EHB1-HA forms. C, EHB1 Δ sig-HA fails to coimmunoprecipitate with IRT1-GFP. IRT1-GFP, EHB1 Δ sig-HA, or a combination of the two was expressed in *N. benthamiana* epidermis cells and used for anti-GFP immunoprecipitation. The samples before (left, input) and after (right, IP: anti-GFP) the procedure were tested by immunoblot. No EHB1-HA signals could be detected in the combined immunoprecipitated sample. Nontransformed samples were used as controls. The experiment was performed three times yielding comparable results. D, The same experiment as in (C) was performed in the presence of 100- μ M Ca²⁺. No EHB1-HA signals could be detected in the combined immunoprecipitated sample. The experiment was performed three times yielding comparable results. E to G, Localization of EHB1-GFP in *N. benthamiana* epidermis cells. H, Anti-GFP immunoblot made on extracts either expressing or not EHB1-GFP. No obvious degradation products or free GFP can be seen. The single asterisk indicates the EHB1-GFP band at the predicted 57 kD and the double asterisk—an additional band at 67 kD. I to L, Colocalization between free GFP and ARABIDOPSIS H⁺-ATPASE 1 (AHA1)-monomeric Red Fluorescent Protein (mRFP). L, Scatterplot of the signals in the GFP and mRFP channels, showing the typical distribution of non-colocalizing signals. M–O, Colocalization between free GFP and AHA1-mRFP in mannitol-plasmolyzed cells. Open arrowheads point toward Hechtian strands indicating the presence of AHA1-mRFP at the plasma membrane. Free GFP could not be found in these structures. P–S, Colocalization between EHB1-GFP and AHA1-mRFP in the region of the plasma membrane. S, Represents a scatterplot of the signals in the GFP and mRFP channels, showing the typical distribution of colocalizing signals. T to V, Colocalization between EHB1-GFP and AHA1-mRFP in mannitol-plasmolyzed cells. Solid arrowheads point toward Hechtian strands indicating the presence of AHA1-mRFP at the plasma membrane. EHB1-GFP was also present in these structures, indicating its localization at the plasma membrane. W–Z, Colocalization between EHB1-GFP and

1568

Downloaded from on January 9, 2020 - Published by www.plantphysiol.org
Copyright © 2019 American Society of Plant Biologists. All rights reserved.

Plant Physiol. Vol. 180, 2019

lipids on the Membrane Lipid Strip (Echelon) showed a strong preference of the purified StrepII-EHB1 protein for phosphatidylinositol (PtdIns) and phosphatidylinositol 4-phosphate (PtdIns4P; Fig. 3A). The finding is particularly intriguing because PtdIns4P is known to accumulate in the plasma membrane of plant cells and has emerged as a key target for the attachment of regulatory proteins (Simon et al., 2016). To verify the finding in an actual membrane environment, we generated liposomes with phosphatidylcholine (PC) as the base lipid. After an incubation with StrepII-EHB1, we pelleted the liposomes by centrifugation. We then monitored the presence of StrepII-EHB1 in the pellet and the supernatant by protein immunoblot followed by band intensity quantification (Fig. 3, B and C). StrepII-EHB1 was copelleted with liposomes containing PC-PtdIns and PC-PtdIns4P mixture, indicating that the protein was able to bind the liposome surface. No signal could be detected in the pelleted fraction of liposomes containing exclusively PC, showing that StrepII-EHB1 binds specifically PtdIns and PtdIns4P lipids. Nor was there signal in the pellet fraction in the absence of liposomes, which demonstrates the absence of protein aggregation (Fig. 3B). Interestingly, when the experiment was performed in the presence of 100- μ M Ca^{2+} , a relative enrichment of the membrane-bound StrepII-EHB1 could be observed, especially in the case of PtdIns-containing liposomes (Fig. 3, B and C). The liposome binding activity of EHB1 was unchanged when the signature domain was deleted (Supplemental Fig. S2A). This shows that the signature domain does not affect the membrane-binding capacity of EHB1, and this finding is in agreement with the described structure of the CAR proteins (Rodriguez et al., 2014). Taken together, EHB1 can bind PtdIns and PtdIns4P-containing membranes and the binding is partially enhanced in the presence of calcium.

EHB1 Membrane Association Is Calcium-Dependent

To further understand the significance of calcium for EHB1 membrane association, we generated an Arabidopsis line expressing an HA-EHB1 fusion (Supplemental Figs. S3 and S4B). Using this, we performed ultracentrifugation-based separation of root membrane and soluble fractions (Fig. 3D; Supplemental Fig. S2B). Plants were grown in the two-week growth system under sufficient iron and iron-deficient conditions to account for potential changes in membrane composition under these two growth regimes. In the absence of added calcium, no significant presence of the HA-EHB1 protein could be observed in microsomal fractions. The presence of 100- μ M Ca^{2+} , however, resulted in a marked increase of membrane-bound HA-EHB1. The effect was

observed and was comparable in both standard-grown and iron-deficient plant samples (Fig. 3D; Supplemental Fig. S2B). This shows that the presence of calcium is an important prerequisite for EHB1 binding to plant membranes. Stimulus-driven relocalization of CAR family proteins to the plasma membrane has been observed in the case of the rice OsGAP1 (Cheung et al., 2010). Our data suggest that the process of EHB1 membrane association might be calcium-driven. At the same time the data show that iron starvation likely does not dramatically affect the availability of EHB1 binding sites in plant membranes.

Molecular Dynamics Simulations Suggest a Two-Step Binding Mode of EHB1 at the Membrane

To obtain a model of the binding of EHB1 at the plasma membrane at the atomistic level, a homology model of EHB1 was generated based on the structures of the homologous proteins CAR1 and CAR4. Molecular dynamics (MD) simulations of EHB1 binding were performed in an explicit solvent/membrane environment. The EHB1 model contained two calcium ions in the calcium binding site, as indicated by structures of CAR homologs (Diaz et al., 2016). The membrane bilayer was composed of a 4:4:1 ratio of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine:1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine:1,2-dioleoyl-*sn*-glycero-3-phosphoglycerol (DOPC/DOPE/DOPG), resembling the composition of a plant plasma membrane (Furt et al., 2011). To avoid any bias, in the starting configuration of the system, EHB1 was placed ~ 25 Å apart from the closest membrane surface. During the MD simulations, EHB1 was then allowed to diffuse freely. In all of the five independent replica simulations performed, after ~ 1 μ s of simulation time, EHB1 bound to the membrane via the calcium binding site (Fig. 3, E and F; Supplemental Movie). Furthermore, four out of five replicas showed a potassium ion in the close vicinity of the calcium binding site, similar to the homolog Protein Kinase C- α (PKC- α ; Fig. 3F; Guerrero-Valero et al., 2009). Finally, EHB1 showed a strong tendency to tilt and to engage in interactions with the membrane with its sides, as demonstrated by a tilt angle close to 90° with respect to the membrane normal. In one replica, the protein interacted that way during almost the complete simulation time (Fig. 3, G and H). Interestingly, the interaction does not occur via a site described as the polybasic patch in homologs (Li et al., 2006), but rather through residues 51–62, 140–142, and 164–168. In the tilted configuration, the CAR-signature domain and the polybasic patch remain close to the surface (Fig. 3H). We noted that the composition of amino acids within the polybasic patch, putatively interacting with phosphoinositides, differs between PKC- α and EHB1, as well as between EHB1

Figure 2. (Continued.)

IRT1-mCherry in the region of the plasma membrane. Z, Scatterplot of the signals in the GFP and mCherry channels showing the typical distribution of colocalizing signals. Bars = 20 μ m; bars in insets = 5 μ m.

Khan et al.

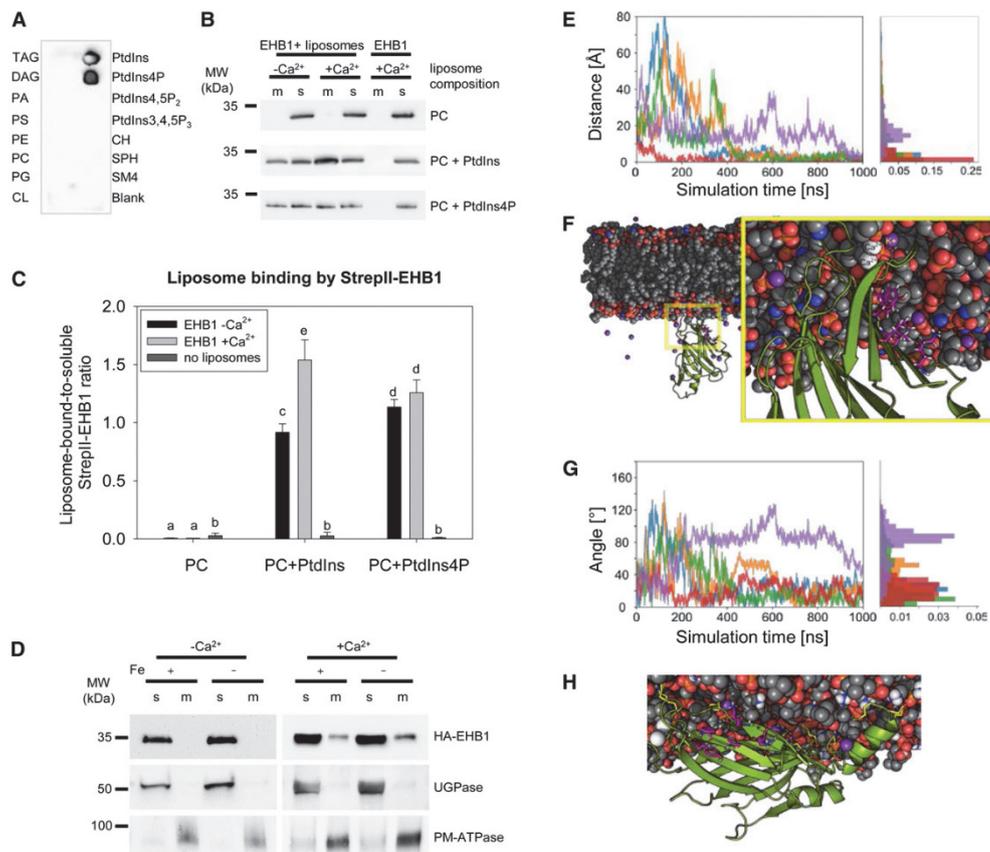


Figure 3. EHB1 binds membranes. A, Lipid overlay assay using StrepII-EHB1. Black signal shows the zones of the strip where StrepII-EHB1 was retained after washing. The experiment was performed three times yielding comparable results. TAG, triacylglycerol; DAG, diacylglycerol; PA, phosphatidic acid; PS, phosphatidyl-Ser; PE, phosphatidylethanolamine; PC, phosphatidylglycerol; CL, cardiolipin; CH, cholesterol; SPH, sphingomyelin; SM4, 3-*o*-sulfolactosylceramide. B, Binding of StrepII-EHB1 to membrane lipids. Liposomes containing PC were used as controls. After incubation with StrepII-EHB1, the liposomes were pelleted and the membrane (m), and soluble (s) fractions analyzed by immunoblot. Presence of StrepII-EHB1 in the membrane fraction indicates liposome binding. The experiment was performed three times yielding comparable results. MW, molecular weight. C, Quantification of data presented in (B). Error bars represent SD, $n = 3$. Letters above the bars indicate statistically significant difference ($P < 0.05$). D, Extracts from HA-EHB1-expressing Arabidopsis plants were fractionated into microsomal (m) and soluble (s) fraction in either the absence ($-Ca^{2+}$) or presence ($+Ca^{2+}$) of calcium. The composition of the fractions was analyzed using immunoblots. The soluble enzyme UDP-Glc pyrophosphorylase (UGPase) and the plasma membrane H^{+} -ATPase (PM-ATPase) were used as markers. The experiment was performed three times yielding comparable results. E–H, EHB1 membrane interactions through the calcium binding site based on five independent MD simulations of 1- μs length in the presence of a membrane bilayer with a 4:4:1 ratio of DOPC/DOPE/DOPG. E, Distance between the calcium ions bound to the protein and the center of mass of the phosphorous atoms of the phospholipids of the closest leaflet over the simulation time; in the starting configuration, EHB1 was placed ~ 25 Å away from the membrane surface. In all replicas, EHB1 binds to the membrane surface through the calcium binding site. F, Representative structure showing an EHB1 configuration (green) bound to the membrane (gray carbon atoms) through the calcium binding site (Ca^{2+} : white spheres). An additional potassium ion was found to bind to an extra calcium binding site (magenta). G, Tilt angle defined between the vector formed by the center of mass of EHB1 with the complexed calcium ions and the membrane normal over the simulation time. A repeated, frequent, and, in one case, persistent tilting of EHB1 was observed in the replica simulations. H, Representative structure of the “parallel” configuration of EHB1 at the membrane. Residues that show direct interactions with the membrane are colored yellow; residues that correspond to the described polybasic patch are colored magenta.

1570

Downloaded from on January 9, 2020 - Published by www.plantphysiol.org
Copyright © 2019 American Society of Plant Biologists. All rights reserved.

Plant Physiol. Vol. 180, 2019

and other CAR-family members (Supplemental Fig. S2, C–F). In PKC- α , these amino acids can interact with all P groups of PtdIns4,5P₂, whereas this is not possible for EHB1. Thus, the different amino acid composition in the case of EHB1 might contribute to the observed changes in lipid-binding specificity (Supplemental Fig. S2, C–F). To conclude, under the chosen simulation conditions, EHB1 shows a clear trend to bind to the membrane via the calcium binding site and then to tilt toward the membrane.

EHB1 Acts as a Negative Regulator of Iron Acquisition

CAR proteins that bind to signaling protein intermediates like the ABA receptor and NONPHOTOTROPIC HYPOCOTYL3 affect the physiological responses to ABA and blue light in a significant manner (Knauer et al., 2011; Rodriguez et al., 2014). Our next aim was to understand whether the interaction between IRT1 and EHB1 also affects downstream physiological reactions, in this case the plant capacity to acquire iron. To estimate the role of EHB1 in iron acquisition, we first analyzed two mutants carrying *EHB1* loss-of-function alleles, *ehb1-1* and *ehb1-2*, for which we confirmed the absence of full-length *EHB1* transcript, and two HA-EHB1 over-expressing lines (Supplemental Fig. S4). Control Col-0 wild-type, *ehb1*, and HA-EHB1 plants were grown in the 2-week growth system. Under iron-deficient conditions, wild-type as well as HA-EHB1-expressing plants developed the characteristic iron-deficiency chlorosis, whereas the effect was much less pronounced in the two *ehb1* lines (Fig. 4A). Leaf chlorosis is caused by a decrease

in the chlorophyll content. Chlorophyll content in all plant lines was lower under iron-deficient than that under iron-sufficient conditions. However, in agreement with the morphological observations, the chlorophyll content was lowest in the HA-EHB1 lines, whereas it was similar in *ehb1* mutant lines under iron deficiency compared with that in wild type grown with sufficient iron. This shows that, compared to wild-type plants, *ehb1* mutant plants are less sensitive whereas HA-EHB1 lines are more sensitive to iron deficiency (Fig. 4B). As a key readout for the efficiency of the reduction-based iron acquisition system, we measured the activity of the FRO2 protein, functioning as the principal root-surface ferric reductase, the step before uptake of ferrous iron by IRT1 (Fig. 4C). As expected, the activity was strongly increased under iron deficiency in all tested genotypes. The absence of EHB1, however, led to a significant increase of FRO2 activity under iron deficiency in comparison to that in the wild type. Conversely, HA-EHB1 lines showed a reduced activity under these conditions compared to that in both wild-type and *ehb1* plants (Fig. 4C). The result is consistent with the phenotypical observations and shows that, in the absence of EHB1, the iron acquisition system is more active. This suggests that EHB1 might function as a negative regulator of iron acquisition. We then examined the expression level of the *EHB1* gene in response to iron availability (Fig. 4D). In wild-type plants, we observed *EHB1* expression down-regulation under iron deficiency compared to that under iron-sufficient conditions. In *ehb1* mutant plants, the *EHB1* expression levels were very low, whereas in HA-EHB1 overexpressing plants, the expression levels were

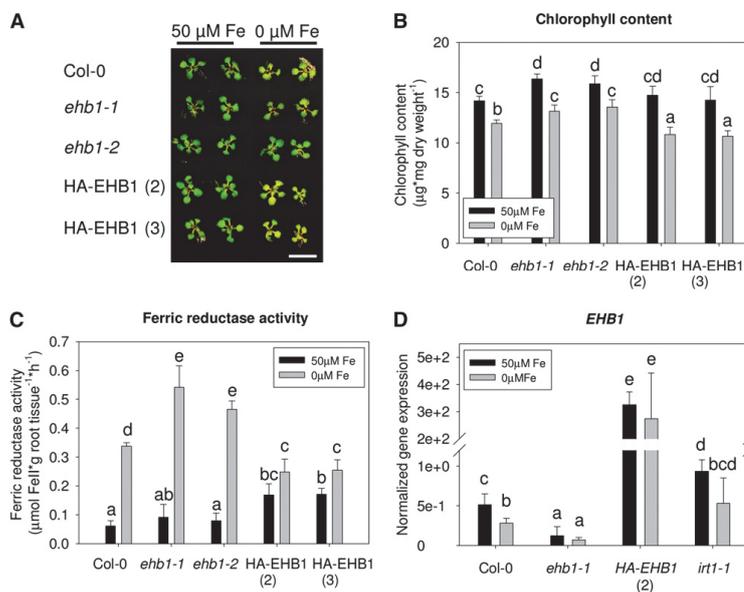


Figure 4. EHB1 inhibits Arabidopsis iron acquisition. A, Seedlings grown in the two-week growth system show iron deficiency chlorosis (leaf yellowing). In EHB1 loss-of-function plants, the effect is much less pronounced, compared to that in the wild-type (Col-0), whereas the effect is stronger in HA-EHB1-expressors. Bar = 1 cm. B, Total chlorophyll content of plants grown as in (A); $n = 5$. C, Ferric reductase activity of plants grown as in (A); $n = 5$. D, Expression level of *EHB1* under different iron regimes and in plants with modified iron acquisition capacity ($n = 3$). Error bars = SD. Letters above the bars indicate statistically significant difference ($P < 0.05$).

Khan et al.

high and not affected by the iron status of the plant (Fig. 4D; Supplemental Fig. S5). In *IRT1* loss-of-function plants, *EHB1* expression levels were higher than that in the wild type under the corresponding conditions (Fig. 4D). This suggests that *EHB1* gene expression is influenced not only by the iron status of the plant but it also responds to the presence of IRT1. The observed downregulation under iron deficiency is consistent with its proposed role as a negative regulator.

We further investigated the role of EHB1 in root responses to iron deficiency by evaluating the expression of key iron-deficiency marker genes (Fig. 5; Supplemental Fig. S5). The *FERRITIN1* (*FER1*) gene responds positively to the amounts of iron present in the plant (Reyt et al., 2015) and it showed downregulation under iron deficiency in wild-type plants. In *ehb1* loss-of-function plants, *FER1* expression under sufficient iron supply was higher than that in the wild type, suggesting that higher amounts of iron are present in the root in the absence of EHB1. Consistently, the opposite effect, reduced *FER1* expression compared to that in wild-type plants under sufficient iron, was observed in HA-EHB1 plants and *irt1* mutant plants, the latter lacking the iron transporter (Fig. 5A). The expression of the FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (*FIT*) gene was upregulated in response to iron deficiency compared to that under iron-sufficient conditions in wild-type, *ehb1* mutants, and HA-EHB1 plants (Fig. 5B; Supplemental Fig. S5). A marked upregulation of *FIT* under both iron conditions compared to that in the wild type could be observed in *IRT1* loss-of-function

plants (Fig. 5B). The *FRO2* gene showed a stronger upregulation under iron deficiency in *ehb1* mutant plants and tendency for reduced upregulation in HA-EHB1 compared to that in the wild type (Fig. 5C; Supplemental Fig. S5). This effect was consistent with the measured ferric reductase activity in these plants. The *FRO2* expression pattern in *irt1* mutants was similar to that in *ehb1-1* (Fig. 5C). Surprisingly, the induction of the *IRT1* gene was less pronounced under iron deficiency in *ehb1* mutant plants than that in the wild type (Fig. 5D; Supplemental Fig. S5). The effect was consistent, as HA-EHB1 plants showed increased *IRT1* expression in the absence of iron compared to that in the wild type (Fig. 5D). *FRO2* and *IRT1* genes are both FIT transcriptional targets, although their induction levels are not strictly coupled and may vary significantly (Lieberman et al., 2015; Wang et al., 2018). In addition, the documented broad stress-responsiveness of *IRT1* (Brumbarova et al., 2015), together with its expression and iron regulation in the absence of FIT (Barberon et al., 2011), suggests that additional transcriptional regulators may modulate its transcript levels. To test whether this effect somehow reflects the IRT1 protein abundance, we quantified the amount of IRT1 in total protein extracts of EHB1 loss-of-function plants. Compared to that in the wild type, roots of *ehb1* mutants were found to contain significantly more IRT1 protein (Fig. 5, E and F). This suggests that the absence of EHB1 leads to stabilization of IRT1 and the observed reduced gene expression in *ehb1* might be a compensatory mechanism for maintaining proper IRT1 levels.

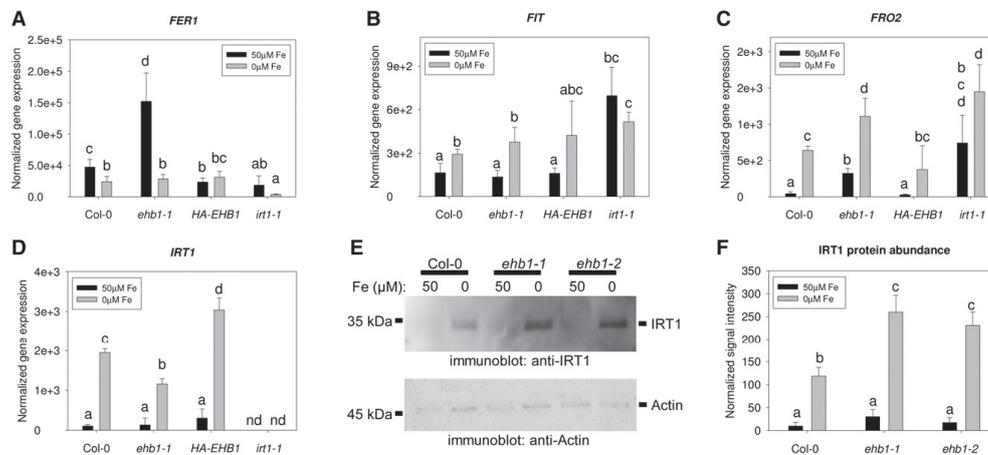


Figure 5. Regulation of iron homeostasis-related genes and proteins is affected by the absence of EHB1. A–D, Expression of genes related to iron storage (*FER1*, A) and acquisition (*FIT*, *FRO2*, and *IRT1*; B–D, respectively). Wild-type (Col-0) seedlings as well as one *ehb1* mutant and one HA-EHB1-expressing line were grown in the 2-week growth system. nd, not detected. Error bars = sd. Different letters above the bars indicate statistically significant difference ($P < 0.05$), $n = 3$. E, Anti-IRT1 immunoblot on extracts of plants grown in the 2-week growth system. Detection of actin was used as a loading control. The experiment was performed three times yielding comparable results. F, Quantification of the experiments presented in (E). Error bars = sd. Different letters above the bars indicate statistically significant difference ($P < 0.05$), $n = 3$.

1572

Presence of EHB1 Suppresses Iron Assimilation

Modulation of IRT1 activity was shown to affect the extent at which iron, found in the root apoplast, is taken up by the plant (Ivanov et al., 2014). Because EHB1 affects IRT1 protein, we suspected that iron uptake is modulated in *ehb1* mutants and that this can be visualized at the level of apoplastic iron. We performed Perls-3,3'-diaminobenzidine tetrahydrochloride staining to visualize the apoplastic iron. Roots of wild-type plants, grown for 8 d under standard iron supply, showed moderate staining at the base of the root (Fig. 6, A and B), which decreased in the early differentiation zone (Fig. 6, A and C). On the other hand, the roots of the *fit-3* mutant, which fails to express *IRT1*, showed intensive iron staining (Fig. 6, M–O), as previously described for *IRT1* loss-of-function plants (Ivanov et al., 2014). Roots of *ehb1* mutant plants showed strongly decreased iron staining in comparison to that in the wild-type (Fig. 6, D–I), whereas HA-EHB1 plants showed enhanced staining in the early differentiation zone, compared to that in the wild-type (Fig. 6, J–L). These results are consistent with the observations on the physiology of EHB1 loss- and gain-of-function plants, and show that EHB1 affects the import of apoplastic iron, a process mediated by IRT1.

We evaluated how the observed alterations in iron import affect the metal contents in the seeds of soil-grown EHB1 loss- and gain-of-function plants. The iron content in *ehb1* mutant seeds showed only slight but not statistically significant increase compared to that in the wild type; however, the seeds of the HA-EHB1 overexpressing line accumulated significantly less iron. This is consistent with the negative effect of

EHB1 on iron uptake. In the case of *ehb1*, probably other mechanisms interfered with the loading of the excess iron. To test whether this effect is specific to iron, we measured the content of manganese, which is imported by a different high-affinity transporter but is also a secondary substrate of IRT1 (Korshunova et al., 1999; Vert et al., 2002). The levels of manganese in *ehb1* mutant seeds were fully comparable to that in wild-type seeds, whereas HA-EHB1 plants showed slightly elevated manganese content. The two elements are known to be dependent on each other's transport in seeds (Chu et al., 2017), so the manganese accumulation in HA-EHB1 plants is potentially a secondary effect of compromised iron loading in these seeds. Thus, EHB1 preferentially affects the acquisition of iron at the import step, which in Arabidopsis is mediated by its interaction partner IRT1.

EHB1 Directly and Negatively Affects IRT1-Mediated Iron Import

To verify the direct connection between the negative effect of EHB1 on iron import and the function of IRT1, we reconstructed this binary system in a heterologous environment. We expressed combinations of the two proteins in yeast (*Saccharomyces cerevisiae*) cells. As a model, we employed the *fet3fet4* mutant yeast strain (Eide et al., 1996). Due to the absence of a functional multicopper oxidase FET3 and a bivalent iron transporter FET4, this strain has a severely reduced capacity of both bivalent and trivalent iron import, compared to that of wild-type yeast (Fig. 7A). Despite that, *fet3fet4*

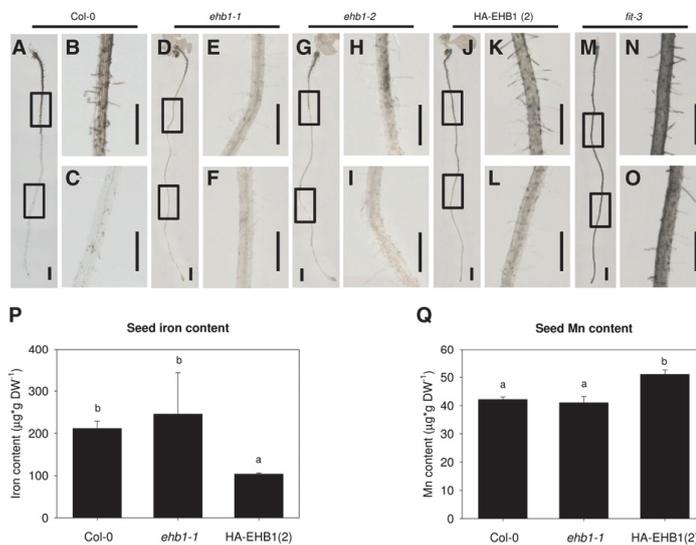


Figure 6. Iron import activity and storage in Arabidopsis depend on the presence of EHB1. A–O, Roots of wild-type (Col-0) seedlings, *ehb1* mutants, and an HA-EHB1-expressing line grown for 8 d under sufficient iron were stained for the presence of iron. The *fit-3* roots were used as a control, which cannot take up the iron available in the apoplast. Apoplastic iron is visible as dark precipitates. Rectangles in the overview images (A, D, G, J, and M) indicate the enlarged regions (B, C, E, F, H, I, K, L, N, and O). The experiment was performed three times yielding comparable results. Scale bars = 0.5 cm. P and Q, Measurement of the total iron (P) and manganese (Q) content in seeds of Col-0, *ehb1-1* and one HA-EHB1-expressing line grown on soil under standard growth conditions. DW, dry weight. Error bars = sd. Different letters above the bars indicate statistically significant difference ($P < 0.05$), $n = 3$.

Khan et al.

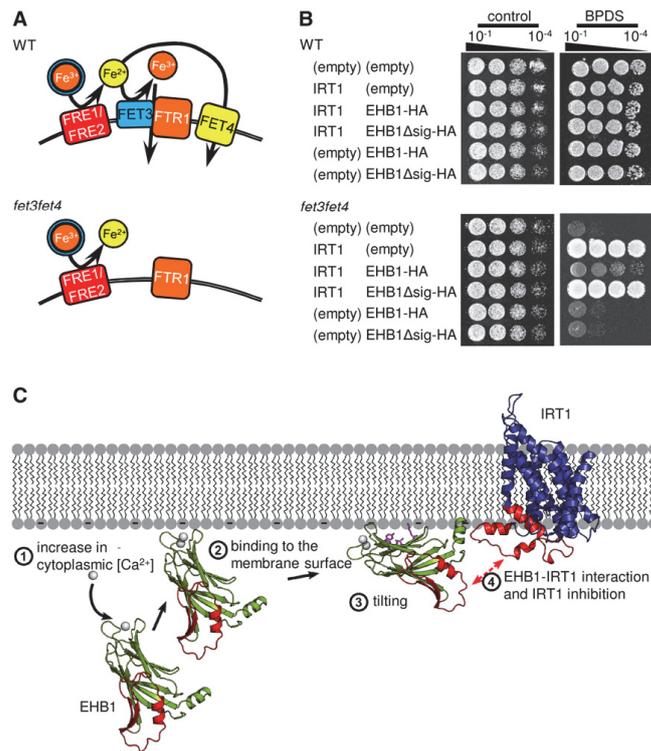
retains its ferric reductase activity and, due to the presence of low-affinity iron-chelate transporters, is capable of surviving under iron-replete conditions. We performed a complementation test using the full-length IRT1 in combination with HA-tagged EHB1 forms. The expression of none of the proteins in the wild type affected yeast growth on iron-depleted medium (Fig. 7B). Unlike the wild-type control, the *fet3fet4* strain was not able to grow in the absence of iron (chelated by bathophenanthrolinedisulfonic acid [BPDS]). However, the expression of IRT1 alone was sufficient to rescue this phenotype, as previously reported (Eide et al., 1996). The addition of EHB1-HA to the system reverted almost entirely this IRT1-mediated complementation, suggesting that EHB1 is sufficient to suppress IRT1 activity. To test this, we coexpressed IRT1 with the interaction-deficient form of EHB1 lacking the CAR-signature domain. This combination led to the complementation of the iron-deficient phenotype, suggesting full activity of IRT1 in the absence of IRT1-EHB1 interaction (Fig. 7B). Because the negative effect of EHB1-HA could not be observed in the wild-type strain, we can exclude the possibility that it may affect other components of the iron acquisition system, such as the ferric reductase, for example. Therefore we can

conclude that through its interaction with IRT1, which is mediated by its CAR-signature domain, EHB1 acts as a direct inhibitor of IRT1 function, leading to suppression of iron acquisition in Arabidopsis.

DISCUSSION

Although IRT1 is not the only transporter at the root surface capable of importing iron (Cailliatte et al., 2010), its controlled expression and subcellular localization in the root hair cells of the early differentiation zone (Blum et al., 2014; Marquès-Bueno et al., 2016; Dubeaux et al., 2018) make it the principal Strategy I iron importer and therefore a critical gateway for bioavailable iron. The capacity of iron to change oxidation states makes it a valuable component of many biological processes. However, its presence poses a challenge for the cell to prevent the deleterious effects of oxidative damage. In addition, IRT1 imports metals other than iron and the possibility to inhibit its function is an important step in preventing metal toxicity. Under iron deficiency, IRT1 substrates, such as zinc and manganese, might be imported and overaccumulate in the plant over time. Terminating this accumulation process might be the

Figure 7. EHB1 inhibits IRT1-mediated iron transport. **A**, Schematic representation of the iron acquisition system in wild-type (WT) yeast and the modified system in the *fet3fet4* strain. FRE1 and 2, Ferric Reductase 1 and 2; FET3 and 4, Ferrous Transport 3 and 4; FTR1, Fe Transporter 1. **B**, Growth of wild-type and *fet3fet4* yeast on media with sufficient iron (control) and with iron-depleted (BPDS) media. The *fet3fet4* strain is not capable of growing when iron is limiting and the introduction of the Arabidopsis IRT1 can rescue this phenotype. In the presence of EHB1, IRT1 loses the capacity to rescue *fet3fet4*, whereas the EHB1 Δ sig form, incapable of interacting with IRT1, has no effect on IRT1-mediated iron acquisition. **C**, Hypothetical mechanism of the EHB1-IRT1 interaction using information from other C2 domain proteins, homology models for EHB1 and IRT1, and results from our MD simulations. Structures highlighted in red relate to protein regions identified here as relevant for the EHB1/IRT1 interaction. (1) An increase in the cytosolic calcium concentration leads to the occupation of the EHB1 calcium binding sites. This increases EHB1's affinity for the plasma membrane surface (2), which contains negatively charged phospholipids. After binding in a "perpendicular" configuration, EHB1 would tilt to favor a "parallel" interaction (3). This would bring the CAR signature domain in close proximity to the IRT1_{vr} (4), allowing the two to interact, that way inhibiting iron uptake by a yet unknown mechanism.



1574

Downloaded from on January 9, 2020 - Published by www.plantphysiol.org
Copyright © 2019 American Society of Plant Biologists. All rights reserved.

Plant Physiol. Vol. 180, 2019

major reason for having an inhibitory protein, such as EHB1, act in these conditions.

We have identified the small peripheral membrane protein EHB1 as an interactor and negative regulator of IRT1. EHB1 belongs to the 10-member CAR protein family, several members of which have a well-documented involvement in the regulation of plant responses to environmental stimuli (Knauer et al., 2011; Cheung et al., 2013; Rodriguez et al., 2014).

We showed that EHB1 exists as both a soluble and membrane-associated protein. EHB1 was found in the nucleus, the cytoplasm, and at the plasma membrane. Such localization is consistent with its influence on the IRT1 transporter, and resembles the reported subcellular localization of other CAR-family proteins (Rodriguez et al., 2014). Of special interest was the discovery that EHB1 shows a specific affinity toward phosphoinositides, such as PtdIns4P, in both an immobilized and membrane-integrated form. PtdIns4P is a very low abundant lipid enriched at the plasma membrane where it generates negative electrostatic charge critical for the attachment of regulatory proteins, such as PINOID and BRI1 KINASE INHIBITOR1 (Simon et al., 2016). We show that calcium is a critical factor in EHB1 membrane association. In the absence of calcium, the protein was unable to associate with plant microsomes. This phenomenon cannot be explained by the EHB1 phosphoinositide binding alone. The calcium-dependence of PtdIns4P and mainly PtdIns binding by EHB1 observed in the liposome-binding experiments was quantitative, rather than a complete on-off situation as seen in the fractionation experiment. This suggests that EHB1 targets additional membrane lipids. Although the exact role of calcium in EHB1 membrane binding is not yet clear, the presence of canonical calcium-coordination sites in the EHB1 C2 domain suggests its involvement in the process. In contrast to CAR1 and CAR4 proteins (Rodriguez et al., 2014), EHB1 could neither bind PC, as observed in lipid strip experiments and liposome binding studies, nor phosphatidyl-Ser (lipid strip experiment). At this point, the data suggest a diversification of target lipids within the family and argues in favor of a functional specification of CAR family members.

Our experiments in yeast showed that EHB1 affected the capacity of IRT1 to complement the iron-deficient *fet3fet4* strain. The control experiment in the wild-type strain, where EHB1 did not affect the endogenous yeast iron acquisition components, demonstrates that the EHB1 effect was due to its direct and specific inhibition of IRT1. This conclusion is also supported by the fact that Basic Local Alignment Search Tool searches revealed no EHB1 homologs, as well as no CAR-family proteins in *Saccharomyces cerevisiae*. The experiment, together with the EHB1-IRT1 interaction, shows that the observed iron acquisition phenotypes in plants with modified EHB1 presence are due to the changes in the efficiency of IRT1 caused by the presence or absence of EHB1. We cannot exclude that EHB1 affects also other components of the Arabidopsis iron acquisition system.

We have observed consistent effects on the activity of the root surface ferric reductase. It may be possible that FRO2 and IRT1 form a protein complex for increased iron import efficiency, similarly to the Fet3pFtr1p complex in yeast. If so, the two proteins might influence each other's activity.

At present, the mechanism of EHB1-mediated IRT1 activity inhibition at the molecular level remains unclear. Because an experimental structure of the IRT1-EHB1 complex remains to be determined, we employed a comparative modeling combined with all-atom MD simulations to suggest an EHB1 membrane binding mode and propose an interaction mechanism between EHB1 and IRT1 (Fig. 7C). It has been previously proposed that the number of calcium ions bound to CAR proteins varies with respect to the surrounding calcium concentration (Diaz et al., 2016). An increase in calcium concentration would therefore favor interactions with negatively charged lipids in the plasma membrane, also causing increased protein concentration close to the membrane surface (Honigsmann et al., 2013; Diaz et al., 2016). Increased concentrations of cytoplasmic calcium have been shown in iron-deficient roots (Tian et al., 2016); however, calcium signatures may vary in intensity and duration (Steinhorst and Kudla, 2014). In a recent study, we were able to show that a calcium-decoding unit of CALCINEURIN B-LIKE PROTEIN1/9-CIPK11 is needed to perceive elevated calcium concentrations at the plasma membrane and translate it into a specific phosphorylation mark on the transcription factor FIT, which is thus activated and upregulates iron deficiency responses (Gratz et al., 2019). Another CIPK-family member, CIPK23, was shown to specifically interact with and regulate IRT1 (Dubeaux et al., 2018). Therefore, the EHB1 membrane association and EHB1-IRT1 interaction might occur in response to specific pulses of calcium waves that trigger the events and disappear shortly after. Thus, EHB1 might represent a rapid and dynamic response mechanism for switching off iron uptake under specific conditions. This may be particularly important under sufficient iron conditions, where the expression of IRT1 is transiently triggered (Hong et al., 2013) and the possibility should exist to quickly block iron import once sufficient iron amounts have been acquired. Such EHB1 function is well supported by the enhanced *EHB1* gene expression under iron-sufficient conditions, the reduced acquisition of apoplastic iron in roots of *ehb1* mutants, and the reduced iron content in seeds of HA-EHB1-expressing plants. Calcium is a second messenger for a wide variety of environmental cues (Steinhorst and Kudla, 2014). As iron acquisition is affected by the availability of other metals in the soil (Lešková et al., 2017), it is possible that EHB1-mediated IRT1 inhibition is not limited to iron-related signaling. It might occur under other stimuli as well, and thus, it might contribute to the global plant stress response modulation in a constantly changing environment.

Our EHB1 structural model was complexed with two calcium ions, reflecting a medium-to-high surrounding

Khan et al.

calcium concentration (Diaz et al., 2016). EHB1 bound to a membrane with a typical composition of a plant plasma membrane in all replicas of MD simulations; initially, the binding occurred with an orientation “perpendicular” (Honigsmann et al., 2013) to the membrane. Under our simulation conditions, the EHB1 concentration in the water phase is rather high (~1.5 mM), which may explain why we observe binding to a membrane that lacks phosphoinositides. At the same time, the membrane does contain negatively charged DOPG, and the binding is in line with the above suggestion that EHB1 targets additional membrane lipids. The MD simulations revealed in all but one replica a potassium ion bound in proximity to the calcium binding sites, which could represent how an additional calcium ion would bind under physiological conditions. In the course of all replicas, EHB1 tended to tilt and interact with the membrane through its positively charged surface, thereby leading to a binding “parallel” to the membrane. In one replica, the tilting persisted throughout the simulation time, mainly through interactions between residues 51–62, 140–142, and 164–168 with the membrane, whereas in the other replica, EHB1 frequently exchanged between “perpendicular” and “parallel” configurations. On the one hand, the lack of phosphoinositides in our model membrane may result in the less stable “parallel” binding mode. On the other hand, the “parallel” configuration may require further interactions to become stable. We note in this context that only in the “parallel” configuration the CAR signature domain, identified here to be responsible for the interaction with IRT1, can be expected to come close to the IRT1_{vr}, as inferred from a structural model containing both EHB1 and IRT1 at or in the membrane. For the homologous proteins PKC- α (Guerrero-Valero et al., 2009) and synaptotagmin-1 (Honigsmann et al., 2013), it was shown that the “parallel” configuration is mediated by interactions between PtdIns4,5P₂ in the membrane and a polybasic patch on the protein. The close plant homologs CAR1 and CAR4 have significant differences in the site of the patch that may be responsible for a lack of specific binding to PtdIns4,5P₂ (Diaz et al., 2016). Similarly, the lack of binding of EHB1 to PtdIns4,5P₂ can, at least in part, be explained by the lack of stabilizing interactions in the polybasic region, while supporting the binding with PtdIns4P.

A plausible reason for the EHB1-mediated IRT1 inhibition may be the interaction with the IRT1 variable region. This cytoplasmically exposed stretch contains one of a total of three metal coordination sites in IRT1 (Eng et al., 1998; Rogers et al., 2000). The site may become hidden and therefore nonfunctional due to the proximity of EHB1. Additionally, two of the IRT1 transmembrane domains, IV and V, contain metal coordination sites. EHB1 interaction might force a change in their orientation, thus rendering them incapable of transferring the iron across the membrane. A second possibility for the inhibitory effect of EHB1 is based on the finding that CAR4 protein was shown to cause liposome membrane tubulation in the presence of

calcium in vitro (Diaz et al., 2016). It was thus proposed that its physiological role would be to react to a calcium signature and initiate membrane curvatures serving as signaling platforms for downstream processes (Diaz et al., 2016). In the case of EHB1, a calcium-dependent tubulation of IRT1-containing membrane might promote the endocytosis of the iron transporter. This would reduce the IRT1 protein available for iron transport and therefore the iron transport efficiency. The physiological data that we present in this article are generally in agreement with such a mechanism. This includes the observation of an increased IRT1 protein abundance in the lines lacking EHB1. The two *ehb1* lines show enhanced iron import, suggesting higher abundance and enhanced stability of active IRT1. These two possible mechanisms of EHB1 function are not mutually exclusive and might happen sequentially: first EHB1-mediated IRT1 inactivation, followed by endocytosis.

In summary, our data show that the peripheral membrane protein EHB1 can interact with the iron transporter IRT1. IRT1 is the primary importer of soil iron. To prevent iron overaccumulation in the cell, the otherwise soluble EHB1 protein is recruited to the plasma membrane in a calcium-dependent manner. In addition, calcium promotes the EHB1-IRT1 interaction, which results in the inhibition of transporter activity. This represents a novel mechanism for dynamic rebalancing of metal acquisition required for the prevention of metal toxicity.

MATERIALS AND METHODS

Yeast Transformation, Media, and Selection

The coding sequence of the IRT1 variable region (IRT1_{vr}) was amplified by PCR from cDNA obtained from iron-deficient wild-type (Col-0) roots using the primers IILB1 and IILB2 (Supplemental Table S1), subcloned into pDONR207 (Life Technologies) and transferred into the final destination vector pGBKT7-GW. This construct was then introduced into yeast (*Saccharomyces cerevisiae*) strain Y187 by the lithium acetate method (BD Biosciences). The resulting strain was mated with AH109 strain (BD Biosciences) harboring a cDNA library made from iron-deficient *Arabidopsis thaliana* roots in the pGADT7 vector (Lingam et al., 2011) and plated on a double-selective Synthetic Defined (SD) medium lacking Trp (selects for presence of pGBKT7-GW/IRT1_{vr}) and Leu (selects for presence of a library cDNA clone in pGADT7). Positive colonies were resuspended in distilled water and spread onto selection plates containing SD medium lacking Trp, Leu, and His, and supplemented with 4 mM of 3-amino-1,2,4-triazole (selection for protein-protein interaction). The plates were incubated for 10 d at 30°C and, starting from d 5, 43 newly-appeared colonies were picked, and insertions were amplified by PCR from the library-containing pGADT7 vector using primers Yeast seqF and Yeast seqR (Supplemental Table S1). For targeted yeast two-hybrid screens, EHB1 coding sequence was amplified using primers EHBInterB1 and EHBInterB2, subcloned into pDONR207, which was then used to create AD-EHB1 fusion in pACT2-GW. Yeast strain AH109 was cotransformed with AD- and BD fusion-expressing vector combinations. Cultures from the resulting transformation events were spotted in 10-fold dilutions onto SD medium-containing agar plates either lacking Trp and Leu or lacking Trp, Leu, and His, and supplemented with and 2.5 mM of 3-amino-1,2,4-triazole. Cotransformation of SNX1 AD and BD fusions, assaying for SNX1 homodimerization (Pourcher et al., 2010), was used as a positive control. Cotransformations with the respective non-recombined pACT2-GW and pGBKT7-GW were used as negative controls.

For complementation studies, the full-length IRT1 coding sequence was amplified from cDNA obtained from iron-deficient wild-type (Col-0) roots

1576

Plant Physiol. Vol. 180, 2019

using primers I1B1 and FLI1B2stop (Supplemental Table S1). The fragment was subcloned into pDONR207 and transferred to pAG426GPD-ccdB-eYFP vector (Susan Lindquist, Addgene plasmid # 14228). *EHB1* coding sequence was amplified from wild-type (Col-0) root cDNA using primers EHB1B1 and EHB1cterB2 (Supplemental Table S1) before subcloning into pDONR207 and transferred to pAG426GPD-ccdB-HA (Susan Lindquist, Addgene plasmid # 14250). Yeast strains INVSc1 (*MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52*, Thermo Fisher Scientific) and DEY1453 (*MATa/MATa ade2/+ can1/can1 his3/his3 leu2/leu2 trp1/trp1 ura3/ura3 fet3-2::HIS3/fet3-2::HIS3 fet4-1::LEU2/fet4-1::LEU2*; Eide et al., 1996) were used as hosts. To study the IRT1-mediated complementation of the DEY1453 iron-deficiency phenotype, 10-fold dilution of yeast cultures harboring combinations of IRT1 and EHB1, together with control transformants, were plated on agar plates containing yeast extract peptone dextrose medium either supplemented with 80 μ M of BPDS (iron-deficient conditions) or not (control conditions). Yeast growth was observed on d 3 after plating. Identical experiment in the INVSc1 strain was used to exclude any potential adverse effects to yeast growth from the expression of the constructs in the absence of the iron uptake-related mutations.

BiFC and Protein Localization

The 2in1 BiFC vector system was used to visualize protein-protein interactions in transiently transformed (*Nicotiana benthamiana*) epidermis cells. Full-length *EHB1* sequence was amplified from cDNA using primers EHB1B3 and EHB1cterB2. *EHB1 Δ sig* was created in a two-step reaction, first the *EHB1* coding sequence fragments neighboring the CAR signature domain were amplified in separate reactions using primers EHB1B3 and E1 Δ sig1, and E1 Δ sig2 and EHB1cterB2. In a second step, *EHB1 Δ sig* was amplified using primers EHB1B3 and EHB1cterB2, and the two fragments from the first step as a template. The *EHB1* and *EHB1 Δ sig* fragments were subcloned in pDONR221-B3B2 vector (Life Technologies). Primers for the amplification of the IRT1vr deletions were as follows: *IRT1 Δ vr*, IILatgB1 and IILnsB4; *IRT1 Δ vr Δ 1*, IILatgB1 and 2cternsB4; *IRT1 Δ vr Δ 2*, 3cterB1 and IILnsB4; *IRT1 Δ vr Δ 3*, 4cterB1 and 4cternsB4. The *IRT1 Δ vr Δ His* fragment was created in a two-step reaction as described for *EHB1 Δ sig*. The first-step amplifications were made using the primer pairs IILatgB1 and 5AM1, and 5BM2 and IILnsB4. The second reaction was performed with primers IILatgB1 and IILnsB4. All *IRT1 Δ vr* fragments were subcloned into pDONR221-P1P4 vector (Life Technologies). Combinations of *EHB1* and *IRT1 Δ vr* fragments were inserted into vector pBIFC2-2in1-CC vector (Grefen and Blatt, 2012) by recombination.

For investigating *EHB1* localization, the full-length *EHB1* sequence was amplified from cDNA using primers EHB1B1 and EHB1cterB2, subcloned into pDONR207 (Life Technologies) and introduced by recombination to pMDCS3 (Curtis and Grossniklaus, 2003). *GFP6* sequence was amplified from pMDCS3 using primers G6B1 and G6B2, subcloned into pDONR207 (Life Technologies) and introduced by recombination to pMDC7 (Curtis and Grossniklaus, 2003). IRT1-mCherry fusion was expressed from the pJNC1-IRT1 vector (Ivanov et al., 2014), and AHA1-mRFP was expressed from the pB7WG2:AHA1 vector (Caesar et al., 2011).

Final vectors were introduced into *N. benthamiana* through a *Rhizobium radiobacter*-mediated transformation (Hötzer et al., 2012) and fluorescence was investigated under a microscope. GFP expression was induced 24 h after infiltration by treating the leaf with 20 μ M of β -estradiol (Sigma-Aldrich) in dimethyl sulfoxide with 0.1% (w/v) TWEEN 20 (Sigma-Aldrich).

All mentioned primers are listed in Supplemental Table S1.

Fluorescence and Confocal Microscopy

For standard fluorescence imaging, AxioImager 2 microscope (Zeiss) equipped with the ApoTome.2 module was used. Images were taken with a Plan-Apochromat 40 \times /1.4 Oil objective and recorded by an AxioCam 503 monochromatic camera (Zeiss). Filter sets were as follows: GFP: Filter set 38 HE eGFP shift free (E) [EX BP 470/40, BS FT 495, EM BP 525/50]; YFP: Filter set 46 HE YFP shift free (E) [EX BP 500/25, BS FT 515, EM BP 535/30]; mCherry: Filter set 43 HE Cy 3 shift free (E) [EX BP 550/25, BS FT 570, EM BP 605/70].

Confocal microscopy was performed on an LSM 780 (Zeiss). For GFP visualization, excitation at 488 nm and detection between 505 and 545 nm were used. For mRFP/mCherry, excitation at 561 nm and detection from 575 to 615 nm were used. Pinholes for both channels were set to 1 Airy Unit resulting in optical slices of 0.8 mm. Images were recorded in a 1,024-pixel format. Colocalization analysis was performed on 8-bit grayscale image pairs, representing the GFP

and mRFP/mCherry channels. Images were loaded in the software ImageJ (<http://rsb.info.nih.gov/ij/>) and analyzed using the JACoP v2.0 plug-in (Bolte and Cordelières, 2006). Threshold values were automatically adjusted by the software.

Protein Expression and Purification

EHB1 coding sequence was amplified from cDNA using primers E1SpeI Δ and E1XhoI Δ containing *SpeI* and *XhoI* restriction sites, respectively (Supplemental Table S1). The fragment was digested by the two nucleases, purified and inserted into the *SpeI* and *XhoI* sites of the pET-StrepII vector (Novagen). The resulting pET-StrepII-EHB1 vector was introduced into BL21(DE3) cells (New England Biolabs). For StrepII-EHB1 fusion protein expression, a culture with *OD*₆₀₀ 0.6 was induced by 0.1 mM of isopropyl β -D-1-thiogalactopyranoside. Cells were collected 3 h after the induction and lysed in a buffer containing 100 mM of Tris-HCl at pH 8.0, 150 mM of NaCl, 5 mM of 1,4-dithiothreitol, 0.35 mg/mL of lysozyme, 1 mM of Brij-35, and 1 \times CIP protease inhibitor (Roche). After 30 min, 0.1% (w/v) of Triton X-100 was added and the suspension was sonicated. StrepII-EHB1 was solubilized in the presence of 100 mM of Tris-HCl at pH 8.0, 150 mM of NaCl, 1 mM of Brij-35, and 8 M of Urea. The unspecific biotin-binding sites were blocked by the addition of avidin to a concentration of 100 μ g/mL. The solution was then diluted 10-fold in 100 mM of Tris-HCl at pH 8.0 and 1 mM of Brij-35 before the addition of Strep-Tactin Macro Prep suspension (IBA Life Sciences). The beads were washed in 100 mM of Tris-HCl at pH 8.0, 150 mM of NaCl, and 1 mM of Brij-35 and then eluted in the same buffer in the presence of 2.5 mM of α -Desthiobiotin.

Protein Coimmunoprecipitation

N. benthamiana epidermis was transformed as above with combinations of IRT1-GFP and EHB1-HA or EHB1 Δ sig-HA expressing vectors. To create the *EHB1-HA* expression cassette, the *EHB1* coding sequence was amplified from cDNA using the EHB1B1 and EHB1cterB2 primers (Supplemental Table S1) introduced to pDONR207 and then transferred to pAUL1 vector (Lyska et al., 2013). The same cloning strategy was used for the generation of the *EHB1 Δ sig-HA* cassette. One gram of plant material was ground under liquid nitrogen and resuspended in immunoprecipitation (IP) buffer (50 mM of Tris HCl at pH 8.0, 150 mM of NaCl, 1 mM of EDTA, 1% [w/v] Triton-x-100, and 1 \times CIP protease inhibitor; Roche). Where indicated, 100 mM of CaCl₂ was added to the IP buffer. The cell debris was removed by centrifugation for 10 min (16,800g, 4°C). GFP-Trap_A beads (ChromoTek) were washed twice in IP buffer and resuspended to a final volume of 100 μ L. The supernatant was incubated with 25 μ L of the beads suspension at a revolving wheel for 3 h at 4°C. Beads were collected by centrifugation and washed three times with 1-mL IP buffer each. Elution was done using 50- μ L SDG buffer (62 mM of Tris-HCl at pH 8.6, 2.5% [w/v] SDS, 2% [w/v] 1,4-dithiothreitol, and 10% [v/v] glycerol) at room temperature. Samples were taken from the original cleared extract (designated "input") and after elution (designated "IP"). The experiments were repeated twice, yielding comparable results.

SDS-PAGE and Immunoblot

Protein electrophoresis and protein immunoblots were performed as described in Ivanov et al. (2014). For electrophoresis, 4% to 20% 10-well Mini-Protein stain-free gradient gels (Bio-Rad) were used. The antibody dilutions were as follows: mouse anti-GFP (11814460001; Roche) 1:1,000, mouse anti-actin (MabGPa; Sigma-Aldrich) 1:1,000, rabbit anti-IRT1 (AS11 1780; Agrisera) 1:5,000, rabbit anti-PM(H⁺)-ATPase (AS07 260; Agrisera) 1:1,000, rabbit anti-UGPase (AS05 086; Agrisera) 1:1,000, rat monoclonal anti-HA horseradish peroxidase conjugated (3F10; Roche) 1:5,000, goat anti-mouse IgG horseradish peroxidase (W4011; Promega) 1:5,000, and goat anti-rabbit IgG horseradish peroxidase (AS09 602; Agrisera) 1:5,000. Strep-tag was detected using Strep-Tactin horseradish peroxidase conjugate (IBA Life Sciences) at a dilution of 1:1,000.

Lipid Overlay and Liposome Binding Assays

The Membrane Lipid Strip (Echelon) membrane was blocked for 3 h with 3% (w/v) fatty acid-free bovine serum albumin (Sigma-Aldrich) in Tris-buffered saline with TWEEN 20 (20 mM of Tris-HCl at pH 7.4, 180 mM of NaCl, 0.1% [w/v] Tween 20) buffer and then incubated overnight with the purified StrepII-EHB1

Khan et al.

protein (2 µg/mL). The presence of immobilized StrepII-EHB1 was detected using Strep-Tactin horseradish peroxidase conjugate (IBA Life Sciences). Where indicated, the Tris-buffered saline with TWEEN 20 buffer contained an additional 100 µM of CaCl₂.

For the liposome-binding assay, PC or a 75:25 (w/w) mixture of PC:PtdIns, or PC:PtdIns4P (Sigma-Aldrich) was prepared. A 2% (w/v) lipid suspension was prepared in a buffer containing 50-mM HEPES and 100 mM of NaCl at pH 6.8 by sonication and frozen. Upon defreezing, the suspension was additionally sonicated and the resulting liposomes were extruded through a 0.2-µm pore size polycarbonate membrane using a Mini Extruder (Avanti Polar Lipids). The liposomes were incubated with 5 µg of purified StrepII-EHB1 for 30 min at 30°C in the above HEPES-NaCl buffer. The suspension was then centrifuged at 100,000g for 30 min in a SW40Ti Swinging Rotor (Beckman). Supernatant and resuspended pellet were collected and analyzed for the presence of StrepII-EHB1. Where indicated, the buffer contained an additional 100 µM of CaCl₂. The experiments were performed three times with new lipid membranes or freshly prepared liposomes, yielding comparable results.

Isolation of Membrane Fractions

Roots of HA-EHB1-expressing Arabidopsis plants, grown in the 2-week growth system, were used in the experiment. The separation of microsomal and soluble fractions was performed as described in Alexandersson et al. (2008), except that where indicated the buffer contained additional 100-µM CaCl₂. Equal amounts of soluble and microsomal fraction were loaded for analysis. The experiment was performed three independent times and yielded comparable results.

Immunoblot Band Intensity Quantification

Densitometry analysis of immunoblot images was performed using the software ImageJ (National Institutes of Health) as described in Ivanov et al. (2012b).

Homology Modeling and MD Simulations

EHB1 and IRT1 (UniProt accession IDs Q9S764 and Q38856, respectively) were modeled with the software TopModel (Mulhaes and Gohlke, 2018) as described in Milić et al. (2018). EHB1 was modeled based on structures with PDB IDs 4V29 and 5A52 (46.9% and 52.2% of sequence identity, respectively; in both cases, chain "A" was used). For IRT1, only PDB ID 5TSA was identified as a template (15.5% sequence identity; chain "A" was used), and the first 50 residues were omitted as no coverage for this portion was available.

MD simulations were prepared to evaluate the binding mode of EHB1 with respect to a membrane. The system was prepared with the tool Packmol-Memgen (Case et al., 2018), placing the protein at ~25 Å from the membrane surface. The membrane composition was 4:4:1 DOPC/DOPE/DOPG, resembling main components of a plant plasma membrane (Furt et al., 2011). Two calcium sites of EHB1 were occupied by using the coordinates found in the crystal structure of PDB ID 4V29. The system charges were neutralized by adding 0.15 M of KCl in the solvation box. The GPU implementation of the program AMBER18 pmemd (Darden et al., 1993; Le Grand et al., 2013) with the ff14SB (Maier et al., 2015) and the software Lipid17 (Dickson et al., 2014; Skjevik et al., 2014; Case et al., 2018) parameters for the protein and the membrane lipids, respectively, were used. Water molecules were added using the TIP3P model (Jorgensen et al., 1983). After thermalization to 300 K and density adaptation, five independent MD simulations of 1-µs length were performed in the NPT ensemble. Covalent bonds to hydrogens were constrained with the SHAKE algorithm (Ryckaert et al., 1977) in all simulations, allowing use of a time step of 2 fs.

To obtain a representative depiction of IRT1 embedded in the membrane, the homology model of IRT1 was treated with the same MD protocol as EHB1, simulating five replicas for 500 ns. The structure considered is a representative conformation of the biggest cluster identified by using the Density-based spatial clustering of applications with noise (DBSCAN) algorithm with an *e*-parameter of 1 Å across the last 200 ns of all replicas. All analyses were performed with the software CPPTRAJ (Roe and Cheatham, 2013).

Plant Material

The Arabidopsis *ehb1-1*, *ehb1-2* (Knauer et al., 2011), and *irt1-1* (Vert et al., 2002) mutants were previously described. Absence of a full-length transcript in

ehb1-1 and *ehb1-2* was verified by PCR on cDNA from the two mutants, as well as on wild-type and HA-EHB1 cDNA. PCR was performed using primers EHB1f and EHB1r (Supplemental Table S1). For the generation of the HA-EHB1 lines, full-length EHB1 fragment was amplified from cDNA using primers EHB1interB1 and EHB1B2 (Supplemental Table S1), subcloned into pDONR207 vector (Life Technologies) and transferred to pAlligator-2 vector by recombination. The resulting vector was introduced into *R. radiobacter* C58Cl (pCV2260) strain. Transformation of Arabidopsis wild-type Col-0 ecotype was performed using the floral dip method (Clough and Bent, 1998). Homozygous T4 plants were used in the study.

Plant Growth Conditions

Arabidopsis plants were surface-sterilized and grown upright on agar plates containing Hoagland medium as in Brumbarova and Ivanov (2016). The medium either contained 50-µM FeNaEDTA (50-µM Fe, sufficient iron condition) or FeNaEDTA was omitted (0-µM Fe, iron-deficient conditions). Plants were either grown directly on 50-µM Fe or 0-µM Fe plates for 8 d (referred to as "8-d growth system"), or were grown on 50-µM Fe plates for 14 d and then separated to either 50-µM Fe or 0-µM Fe plates for three additional days (referred to as "2-week growth system").

For localization and interaction analysis in leaf epidermis, *N. benthamiana* plants were germinated on soil and grown for 3 weeks in a greenhouse. When needed, artificial light was used to ensure a 16-h day and 8-h night cycle.

Histochemical Detection of Iron

The visualization of iron in roots of plants, grown in the 8-d growth system, was performed using the Perls-3,3'-diaminobenzidine tetrahydrochloride method, as described in Brumbarova and Ivanov (2014). The stained roots were imaged on AxioImager.M2 microscope (Zeiss) at 20× magnification using the "Tiles" module and the "stitching" function of the software ZEN 2 (Zeiss) to assemble the collected images. The experiment was performed on three independently grown batches of plants, yielding comparable results.

Chlorophyll Content Measurement

Leaves were collected from plants grown in the 2-week growth system. Chlorophyll content was analyzed as described in Ivanov et al. (2014). The experiment was performed on five independently grown sets of plants.

Measurement of Ferric Reductase Activity

Plants were grown in the 2-week growth system. The activity of the root surface ferric reductase was measured spectrophotometrically as described in Le et al. (2016). The experiment was performed on five independently grown sets of plants.

Measurement of Root Length

Root length was measured from frontal images of plant roots using the JMicroVision software (<http://www.jmicrovision.com>) as described in Ivanov et al. (2014). A minimum of 25 plants were measured per genotype per condition. The whole experiment was performed on five independently grown sets of plants.

Gene Expression Analysis by Reverse-Transcription Quantitative PCR

Total RNA was prepared from Arabidopsis plants grown in the 2-week growth system using the RNeasy Plant Mini Kit (Qiagen). Oligo dT primer and RevertAid first-strand cDNA synthesis kit (Thermo Fisher Scientific) were used for cDNA preparation. Reactions were prepared with DyNAmo Color-Flash SYBR Green qPCR Kit (Thermo Fisher Scientific) and were carried out in a C100 Touch PCR Cycler equipped with the CFX96 Real-Time System (Bio-Rad). Data were analyzed with the software CFX Manager (Bio-Rad). Reactions using mass standard dilution series for each gene were used to generate standard curves, based on which the quantification of the samples was done. Samples were normalized to the expression of *EF1Bα*. Amplification of the unsplined form of *EF1Bα* was used to control for genomic DNA contamination. The

1578

Plant Physiol. Vol. 180, 2019

experiment was performed on three independently grown sets of plants with each cDNA sample analyzed in two replicates to control for technical variations. Primer sequences for all analyzed genes are available in Supplemental Table S1.

Fe and Mn Content Measurement

Seed samples of soil-grown plants were ground and melted at high temperature in a solution of 65% (v/v) HNO₃:30% (v/v) H₂O₂ – 5:2. The cooled solutions were measured with inductively coupled plasma optical emission spectrometry (conical atomizer; 0.61 l/min argon flow; pressure: 2.41 bar, Ultima 2; HORIBA Jobin Yvon). The wavelengths used were as follows: Fe – 259.940 nm; Mn – 257.610 nm. Three independent batches of seeds were measured.

Statistical Analysis

For statistical analysis, data were analyzed using one-way ANOVA followed by Fisher's least significant difference post hoc test in the software SPSS Statistics (IBM). The obtained *P* values were used for identifying statistically significant differences between groups (*P* < 0.05).

Accession Numbers

CAR1, A15g37740; CAR4, A13g17980; EF1Bα, A15g19510; EHB1, A11g70800; FER, A15g01600; FT, A12g28160; FRO2, A11g01580; IRT1, A14g19690; SNX1, A15g06140; PKC-α, GCL1P066302.

Data Availability

Data supporting the findings of this study are available within the article and its Supplemental Data, and from the corresponding author on request.

Supplemental Data

The following supplemental information is available.

Supplemental Figure S1. Heterologous expression and purification of StrepII-EHB1 and StrepII-EHB1Δsig in *E. coli*.

Supplemental Figure S2. Binding of StrepII-EHB1Δsig to membrane lipids.

Supplemental Figure S3. Generation of HA-EHB1-expressing Arabidopsis plants.

Supplemental Figure S4. The *ehb1* T-DNA insertion mutants used in this study.

Supplemental Figure S5. Gene expression analysis in the *ehb1-2* mutant.

Supplemental Table S1. Primers used in this study.

Supplemental Movie. MD simulation of EHB1 membrane binding.

ACKNOWLEDGMENTS

We thank Dr. Christoph Forreiter for the *ehb1* mutants, Dr. Christopher Grefen for the pBIFC2-in1 vectors, Prof. Klaus Harter for the pB7WG2-AHA1 vector, Dr. Yves Jacob for the pACT2-GW and pGBKT7-GW vectors, Dr. Nicole Linka for help in the liposome preparation, Dr. Francois Parcy for the pAlligator-2 vector, and Prof. Peter Westhoff for the pAUL1 vector. We thank Sarah Knitter, Karolin Montag, and Daniela Lichtblau for help in cloning. We are grateful to the Jülich Supercomputing Centre at the Forschungszentrum Jülich for computing time on the supercomputers JURECA and JUWELS (NIC project ID: HKF7) and to the Zentrum für Informations und Medientechnologie at the Heinrich Heine University Düsseldorf for computational support. Received February 13, 2019; accepted April 12, 2019; published April 30, 2019.

Plant Physiol. Vol. 180, 2019

LITERATURE CITED

- Alexander E, Gustavsson N, Bernfur K, Karlsson A, Kjellbom P, Larsson C (2008) Purification and proteomic analysis of plant plasma membranes. *Methods Mol Biol* 432: 161–173
- Barberon M, Zelazny E, Robert S, Conéjéro G, Curie C, Friml J, Vert G (2011) Monoubiquitin-dependent endocytosis of the iron-regulated transporter 1 (IRT1) transporter controls iron uptake in plants. *Proc Natl Acad Sci USA* 108: E450–E458
- Barberon M, Dubeaux G, Kolb C, Isono E, Zelazny E, Vert G (2014) Polarization of IRON-REGULATED TRANSPORTER1 (IRT1) to the plant-soil interface plays crucial role in metal homeostasis. *Proc Natl Acad Sci USA* 111: 8293–8298
- Blum A, Brumbarova T, Bauer P, Ivanov R (2014) Hormone influence on the spatial regulation of *IRT1* expression in iron-deficient *Arabidopsis thaliana* roots. *Plant Signal Behav* 9: e28787
- Bolte S, Cordelières FP (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* 224: 213–232
- Briat JF, Dubos C, Gaymard F (2015) Iron nutrition, biomass production, and plant product quality. *Trends Plant Sci* 20: 33–40
- Brumbarova T, Ivanov R (2014) Perls staining for histochemical detection of iron in plant samples. *Bio Protoc* 4: e1245
- Brumbarova T, Ivanov R (2016) Differential gene expression and protein phosphorylation as factors regulating the state of the Arabidopsis SNX1 protein complexes in response to environmental stimuli. *Front Plant Sci* 7: 1456
- Brumbarova T, Bauer P, Ivanov R (2015) Molecular mechanisms governing Arabidopsis iron uptake. *Trends Plant Sci* 20: 124–133
- Caesar K, Elgass K, Chen Z, Huppenberger P, Witthöft J, Schleifenbaum F, Blatt MR, Oecking C, Harter K (2011) A fast brassinolide-regulated response pathway in the plasma membrane of *Arabidopsis thaliana*. *Plant J* 66: 528–540
- Cailliatte R, Schikora A, Briat JF, Mari S, Curie C (2010) High-affinity manganese uptake by the metal transporter NRAMP1 is essential for Arabidopsis growth in low manganese conditions. *Plant Cell* 22: 904–917
- Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham III TE, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi DMK, Gilson G, et al (2018) AMBER 2018. University of California, San Francisco, CA
- Cheung MY, Zeng NY, Tong SW, Li WY, Xue Y, Zhao KJ, Wang C, Zhang Q, Fu Y, Sun Z, et al (2008) Constitutive expression of a rice GTPase-activating protein induces defense responses. *New Phytol* 179: 530–545
- Cheung MY, Xue Y, Zhou L, Li MW, Sun SS, Lam HM (2010) An ancient P-loop GTPase in rice is regulated by a higher plant-specific regulatory protein. *J Biol Chem* 285: 37359–37369
- Cheung MY, Li MW, Yung YL, Wen CQ, Lam HM (2013) The unconventional P-loop NTPase OsYchF1 and its regulator OsCAP1 play opposite roles in salinity stress tolerance. *Plant Cell Environ* 36: 2008–2020
- Chu HH, Car S, Socha AL, Hindt MN, Punshon T, Gueriot ML (2017) The Arabidopsis MTP8 transporter determines the localization of manganese and iron in seeds. *Sci Rep* 7: 11024
- Crough SJ, Bent AF (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735–743
- Curtis MD, Grossniklaus U (2003) A Gateway cloning vector set for high-throughput functional analysis of genes in plants. *Plant Physiol* 133: 462–469
- Darden TA, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089–10092
- Demir F, Horntrich C, Blachutzik JO, Scherzer S, Reinders Y, Kierszniowska S, Schulze WX, Harms GS, Hedrich R, Geiger D, et al (2013) Arabidopsis nanodomain-delimited ABA signaling pathway regulates the anion channel SLAH3. *Proc Natl Acad Sci USA* 110: 8296–8301
- Diaz M, Sanchez-Barrena MJ, Gonzalez-Rubio JM, Rodriguez L, Fernandez D, Antoni R, Yunta C, Belda-Palazon B, Gonzalez-Guzman M, Peirats-Llobet M, et al (2016) Calcium-dependent oligomerization of CAR proteins at cell membrane modulates ABA signaling. *Proc Natl Acad Sci USA* 113: E396–E405
- Dickson CJ, Madej BD, Skjervek AA, Betz RM, Teigen K, Gould IR, Walker RC (2014) Lipid14: The AMBER lipid force field. *J Chem Theory Comput* 10: 865–879

1579

Khan et al.

- Dubeaux G, Neveu J, Zelazny E, Vert G (2018) Metal sensing by the IRT1 transporter-receptor orchestrates its own degradation and plant metal nutrition. *Mol Cell* 69: 953–964
- Dümmer M, Michalski C, Essen LO, Rath M, Galland P, Forreiter C (2016) EHB1 and AGD12, two calcium-dependent proteins affect gravitropism antagonistically in *Arabidopsis thaliana*. *J Plant Physiol* 206: 114–124
- Eide D, Broderius M, Fett J, Guerinot ML (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc Natl Acad Sci USA* 93: 5624–5628
- Eng BH, Guerinot ML, Eide D, Saier MH, Jr. (1998) Sequence analyses and phylogenetic characterization of the ZIP family of metal ion transport proteins. *J Membr Biol* 166: 1–7
- Fourcroy P, Tissot N, Gaymard F, Briat JF, Dubos C (2016) Facilitated Fe nutrition by phenolic compounds excreted by the Arabidopsis ABCG37/PDR9 transporter requires the IRT1/FRO2 high-affinity root Fe²⁺ transport system. *Mol Plant* 9: 485–488
- Furt F, Simon-Plas F, Mongrand S (2011) Lipids of the plant plasma membrane. In AS Murphy, B Schulz, W Peer, eds, *The Plant Plasma Membrane*. Springer, Heidelberg, pp 3–30
- Giehl RF, von Wirén N (2014) Root nutrient foraging. *Plant Physiol* 166: 509–517
- Gratz R, Manishankar P, Ivanov R, Koster P, Mohr I, Trofimov K, Steinhilber L, Meiser J, Mai HJ, Drerup M, et al (2019) CIPK11-dependent phosphorylation modulates FIT activity to promote Arabidopsis iron acquisition in response to calcium signaling. *Dev Cell* 48: 726–740
- Grefen C, Blatt MR (2012) A 2in1 cloning system enables ratiometric bi-molecular fluorescence complementation (rBiFC). *Biotechniques* 53: 311–314
- Guerinot ML, Yi Y (1994) Iron: Nutritious, noxious, and not readily available. *Plant Physiol* 104: 815–820
- Guerrero-Valero M, Ferrer-Orta C, Querol-Audí J, Marín-Vicente C, Fita I, Gómez-Fernández JC, Verdager N, Corbalán-García S (2009) Structural and mechanistic insights into the association of PKCa-C2 domain to PtdIns(4,5)P₂. *Proc Natl Acad Sci USA* 106: 6603–6607
- Haak DC, Fukao T, Grene K, Hua Z, Ivanov R, Perrella G, Li S (2017) Multilevel regulation of abiotic stress responses in plants. *Front Plant Sci* 8: 1564
- Hanikenne M, Krämer U, Demoulin V, Baurain D (2005) A comparative inventory of metal transporters in the green alga *Chlamydomonas reinhardtii* and the red alga *Cyanidioschyzon merolae*. *Plant Physiol* 137: 428–446
- Hong S, Kim SA, Guerinot ML, McClung CR (2013) Reciprocal interaction of the circadian clock with the iron homeostasis network in Arabidopsis. *Plant Physiol* 161: 893–903
- Honigsmann A, van den Bogaart G, Iraheta E, Risselada HJ, Milovanovic D, Mueller V, Müller S, Diederichsen U, Fasshauer D, Grubmüller H, et al (2013) Phosphatidylinositol 4,5-bisphosphate clusters act as molecular beacons for vesicle recruitment. *Nat Struct Mol Biol* 20: 679–686
- Hötzer B, Ivanov R, Brumbarova T, Bauer P, Jung G (2012) Visualization of Cu²⁺ uptake and release in plant cells by fluorescence lifetime imaging microscopy. *FEBS J* 279: 410–419
- Ivanov R, Bauer P (2017) Sequence and coexpression analysis of iron-regulated ZIP transporter genes reveals crossing points between iron acquisition strategies in green algae and land plants. *Plant Soil* 418: 61–73
- Ivanov R, Brumbarova T, Bauer P (2012a) Fitting into the harsh reality: Regulation of iron-deficiency responses in dicotyledonous plants. *Mol Plant* 5: 27–42
- Ivanov R, Tiedemann J, Czihal A, Baumlein H (2012b) Transcriptional regulator AtET2 is required for the induction of dormancy during late seed development. *J Plant Physiol* 169: 501–508
- Ivanov R, Brumbarova T, Blum A, Janitke AM, Fink-Straube C, Bauer P (2014) SORTING NEXIN1 is required for modulating the trafficking and stability of the Arabidopsis IRON-REGULATED TRANSPORTER1. *Plant Cell* 26: 1294–1307
- Jørgensen WL, Chandrasekhar J, Madura JD, Impey RD, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935
- Knauer T, Dümmer M, Landgraf F, Forreiter C (2011) A negative effector of blue light-induced and gravitropic bending in Arabidopsis. *Plant Physiol* 156: 439–447
- Korshunova YO, Eide D, Clark WG, Guerinot ML, Pakrasi HB (1999) The IRT1 protein from *Arabidopsis thaliana* is a metal transporter with a broad substrate range. *Plant Mol Biol* 40: 37–44
- Le CT, Brumbarova T, Ivanov R, Stoof C, Weber E, Mohrbacher J, Fink-Straube C, Bauer P (2016) ZINC FINGER OF ARABIDOPSIS THALIANA12 (ZAT12) interacts with FER-LIKE IRON DEFICIENCY-INDUCED TRANSCRIPTION FACTOR (FIT) linking iron deficiency and oxidative stress responses. *Plant Physiol* 170: 540–557
- Le Grand S, Götz AW, Walker RC (2013) SPPF: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun* 184: 374–380
- Lešková A, Giehl RFH, Hartmann A, Fargašová A, von Wirén N (2017) Heavy metals induce iron deficiency responses at different hierarchical and regulatory levels. *Plant Physiol* 174: 1648–1668
- Li L, Shin OH, Rhee JS, Araç D, Rah JC, Rizo J, Südhof T, Rosenmund C (2006) Phosphatidylinositol phosphates as co-activators of Ca²⁺ binding to C2 domains of synaptotagmin 1. *J Biol Chem* 281: 15845–15852
- Liberman LM, Sparks EE, Moreno-Risueno MA, Petricka JJ, Benfey PN (2015) MYB36 regulates the transition from proliferation to differentiation in the Arabidopsis root. *Proc Natl Acad Sci USA* 112: 12099–12104
- Lingam S, Mohrbacher J, Brumbarova T, Potuschak T, Fink-Straube C, Blondet E, Genschik P, Bauer P (2011) Interaction between the bHLH transcription factor FIT and ETHYLENE INSENSITIVE3/ETHYLENE INSENSITIVE3-LIKE1 reveals molecular linkage between the regulation of iron acquisition and ethylene signaling in Arabidopsis. *Plant Cell* 23: 1815–1829
- Lyska D, Engelmann K, Meierhoff K, Westhoff P (2013) pAUL: A gateway-based vector system for adaptive expression and flexible tagging of proteins in Arabidopsis. *PLoS One* 8: e53757
- Maier JA, Martínez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11: 3696–3713
- Marqués-Bueno MDM, Morao AK, Cayrel A, Platre MP, Barberon M, Caillieux E, Colot V, Jaillais Y, Roudier F, Vert G (2016) A versatile multisite Gateway-compatible promoter and transgenic line collection for cell type-specific functional genomics in Arabidopsis. *Plant J* 85: 320–333
- Milić D, Dick M, Mulnaes D, Pflieger C, Kinnen A, Gohlke H, Groth G (2018) Recognition motif and mechanism of ripening inhibitory peptides in plant hormone receptor ETR1. *Sci Rep* 8: 3890
- Mulnaes D, Gohlke H (2018) TopScore: Using deep neural networks and large diverse data sets for accurate protein model quality assessment. *J Chem Theory Comput* 14: 6117–6126
- Pourcher M, Santambrogio M, Thazar N, Thierry AM, Fobis-Loisy I, Miège C, Jaillais Y, Gaude T (2010) Analyses of sorting nexins reveal distinct retromer-subcomplex functions in development and protein sorting in *Arabidopsis thaliana*. *Plant Cell* 22: 3950–3991
- Reyt G, Boudouf S, Boucherez J, Gaymard F, Briat JF (2015) Iron- and ferritin-dependent reactive oxygen species distribution: Impact on Arabidopsis root system architecture. *Mol Plant* 8: 439–453
- Robinson NJ, Procter CM, Connolly EL, Guerinot ML (1999) A ferric-chelate reductase for iron uptake from soils. *Nature* 397: 694–697
- Rodríguez L, Gonzalez-Guzman M, Diaz M, Rodrigues A, Izquierdo-García AC, Peirats-Llobet M, Fernandez MA, Antoni R, Fernandez D, Marquez JA, et al (2014) C2-domain abscisic acid-related proteins mediate the interaction of PYR/PYL/RCAR abscisic acid receptors with the plasma membrane and regulate abscisic acid sensitivity in Arabidopsis. *Plant Cell* 26: 4802–4820
- Roe DR, Cheatham III TE (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9: 3084–3095
- Rogers EE, Eide DJ, Guerinot ML (2000) Altered selectivity in an Arabidopsis metal transporter. *Proc Natl Acad Sci USA* 97: 12356–12360
- Römheld V, Marschner H (1983) Mechanism of iron uptake by peanut plants: I. Fe reduction, chelate splitting, and release of phenolics. *Plant Physiol* 71: 949–954
- Ryckaert J-P, Cicotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys* 23: 327–341
- Shin LJ, Lo JC, Chen GH, Callis J, Fu H, Yeh KC (2013) IRT1 degradation factor1, a ring E3 ubiquitin ligase, regulates the degradation of iron-regulated transporter1 in Arabidopsis. *Plant Cell* 25: 3039–3051

1580

Plant Physiol. Vol. 180, 2019

EHB1 Inhibits IRT1-Based Fe Import

- Simon ML, Platre MP, Marquès-Bueno MM, Armengot L, Stanislas T, Bayle V, Caillaud MC, Jaillais Y (2016) A PtdIns(4)P-driven electrostatic field controls cell membrane identity and signalling in plants. *Nat Plants* 2: 16089
- Skjerveik AA, Mileni M, Baumann A, Halskau O, Teigen K, Stevens RC, Martinez A (2014) The N-terminal sequence of tyrosine hydroxylase is a conformationally versatile motif that binds 14-3-3 proteins and membranes. *J Mol Biol* 426: 150–168
- Steinhorst L, Kudla J (2014) Signaling in cells and organisms—calcium holds the line. *Curr Opin Plant Biol* 22: 14–21
- Tian Q, Zhang X, Yang A, Wang T, Zhang WH (2016) CIPK23 is involved in iron acquisition of Arabidopsis by affecting ferric chelate reductase activity. *Plant Sci* 246: 70–79
- Urzica EI, Casero D, Yamasaki H, Hsieh SI, Adler LN, Karpowicz SJ, Blaby-Haas CE, Clarke SG, Loo JA, Pellegrini M, et al (2012) Systems and trans-system level analysis identifies conserved iron deficiency responses in the plant lineage. *Plant Cell* 24: 3921–3945
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat JF, Curie C (2002) IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14: 1223–1233
- Wang H, Tang J, Liu J, Hu J, Liu J, Chen Y, Cai Z, Wang X (2018) Abscisic acid signaling inhibits brassinosteroid signaling through dampening the dephosphorylation of BIN2 by ABI1 and ABI2. *Mol Plant* 11: 315–325
- Wedepohl KH (1995) The composition of the continental-crust. *Geochim Cosmochim Acta* 59: 1217–1232

1 **Supplemental data**

2 **Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter**
3 **IRT1 inhibits Arabidopsis iron acquisition**

4
5 Imran Khan,^a Regina Gratz,^{a,b} Polina Denezhkin,^b Stephan N. Schott-Verdugo,^{c,g} Kalina Angrand,^b
6 Lara Genders,^a Rubek Merina Basgaran,^a Claudia Fink-Straube,^d Tzvetina Brumbarova,^{a,b} Holger
7 Gohlke,^{e,f} Petra Bauer^{a,b,d} and Rumen Ivanov,^{a,b,*}

8
9 ^a Institute of Botany, Heinrich-Heine University, Universitätsstrasse 1, D-40225 Düsseldorf,
10 Germany.

11 ^b Former address: Department of Biosciences-Plant Biology, Saarland University, Campus
12 A2.4, D-66123 Saarbrücken, Germany.

13 ^c Leibniz Institute for New Materials, D-66123 Saarbrücken, Germany.

14 ^d Cluster of Excellence on Plant Sciences (CEPLAS), Heinrich-Heine University,
15 Universitätsstrasse 1, D-40225 Düsseldorf, Germany.

16 ^e Institute for Pharmaceutical and Medicinal Chemistry, Heinrich Heine University Düsseldorf,
17 D-40225 Düsseldorf, Germany

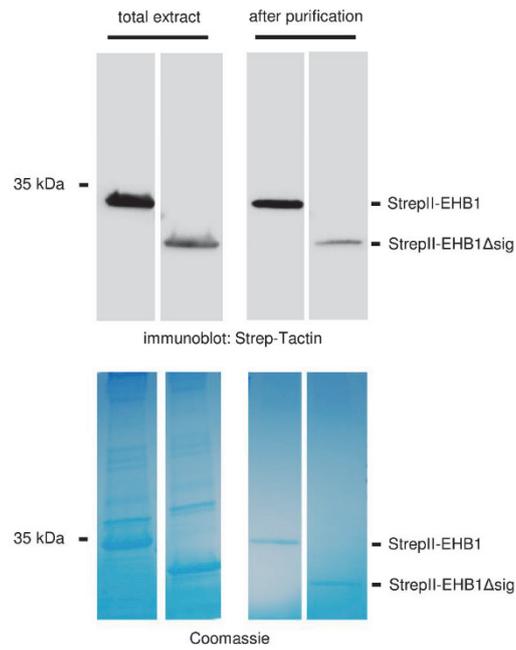
18 ^f John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC) &
19 Institute of Complex Systems, ICS-6: Structural Biochemistry, Forschungszentrum Jülich
20 GmbH, D-52425 Jülich, Germany

21 ^g Centro de Bioinformática y Simulación Molecular (CBSM), Facultad de Ingeniería,
22 Universidad de Talca, 2 Norte 685, CL-3460000 Talca, Chile

23
24

25 **Supplemental Figures**

26
27
28
29
30



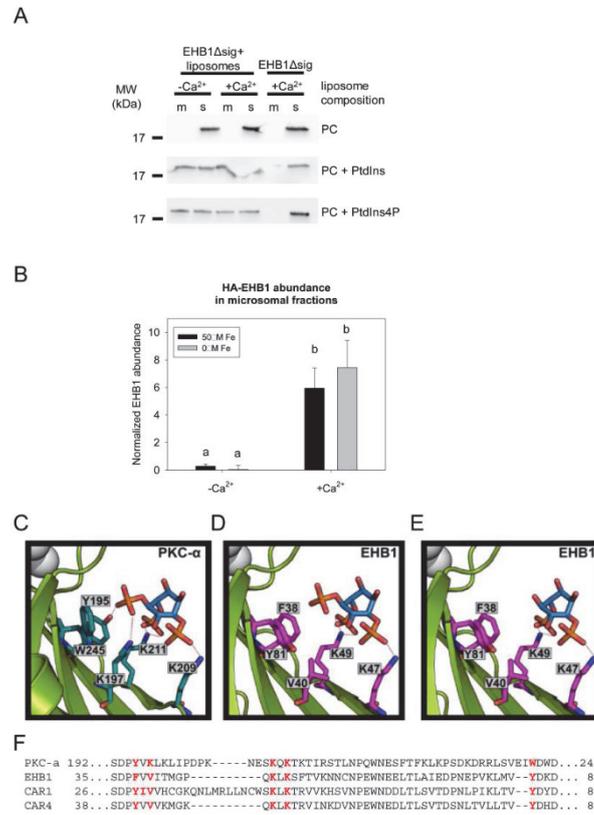
31
32
33
34
35
36
37
38
39
40

Supplemental Figure S1. Heterologous expression and purification of StrepII-EHB1 and StrepII-EHB1Δsig in *E. coli*.

Total bacterial extracts and affinity purified StrepII-EHB1, and StrepII-EHB1Δsig were analyzed by immunoblot with Strep-Tactin. Coomassie-stained gels are shown as controls. The positions of StrepII-EHB1 and StrepII-EHB1Δsig are indicated.

41

42



43

44 **Supplemental Figure S2. Binding of StrepII-EHB1 Δ sig to membrane lipids.**

45 (A) Liposomes containing PC were used as controls. Following incubation with StrepII-
 46 EHB1 Δ sig, the liposomes were pelleted and the membrane (m), and soluble (s) fractions analyzed
 47 by immunoblot. Presence of StrepII-EHB1 Δ sig in the membrane fraction indicates membrane
 48 binding. The experiment was performed three times yielding comparable results.

49 (B) Quantification of HA-EHB1 presence in microsomal fractions prepared from Arabidopsis
 50 roots. Percentages of signal in the microsome fraction were calculated for HA-EHB1 and UGPase.
 51 The HA-EHB1 signal was then normalized to UGPase. The data represents three independent

52 experiments. Error bars indicate standard deviation. Letters indicate statistically significant
53 differences ($P < 0.05$).

54 **(C)** to **(F)** Mutations in the polybasic patch can explain the lack of binding of EHB1 to PtdIns4,5P₂
55 **(C)** C2 domain protein PKC- α (PDB ID 3GPE) preferably binds to PtdIns4,5P₂ (blue carbon
56 atoms) through its polybasic binding site (marine carbon atoms), which contains Tyr and Lys to
57 interact with the phosphate at O5.

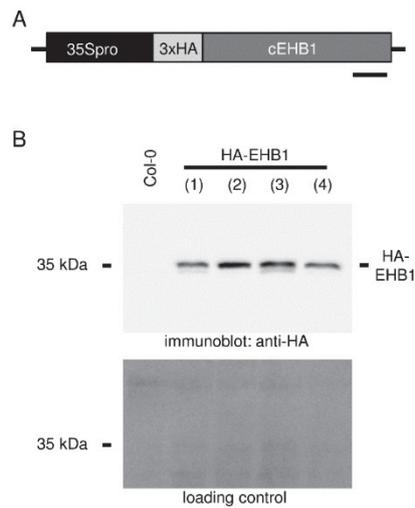
58 **(D)** In contrast, EHB1 has a Phe and Val at this site (magenta carbon atoms), which may explain
59 the lack of EHB1 binding to PtdIns4,5P₂ in the strip analysis (Figure 3A).

60 **(E)** This site could bind PtdIns4P, as for this lipid no stabilization of the negative charge of the
61 phosphate at O5 is required. The PtdIns4,5P₂- and PtdIns4P-bound structures were modeled based
62 on PDB ID 3GPE.

63 **(F)** Alignment of the polybasic region of PKC- α , CAR1, CAR4 and EHB1. The amino acids
64 described in **(C)**-**(E)** are highlighted in red.

65

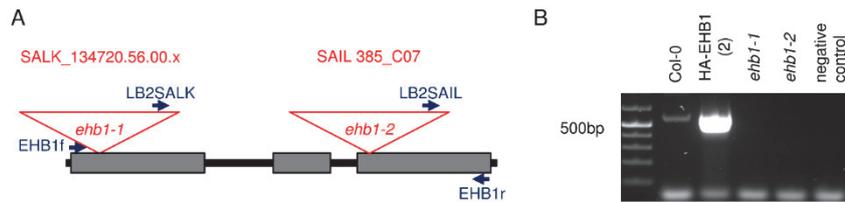
66
67
68
69
70



71
72
73
74
75
76
77
78
79
80
81
82

Supplemental Figure S3. Generation of HA-EHB1-expressing Arabidopsis plants.
(A) Schematic representation of the expression cassette used for generating the HA-EHB1 lines.
The size bar represents 100 base pairs.
(B) Presence of the HA-EHB1 protein was verified by immunoblot. HA-positive band at an estimated size of 35 kDa was present in protein extracts from transgenic plants but absent in wild type extracts.

83
84
85
86
87
88



89
90
91
92
93
94
95
96

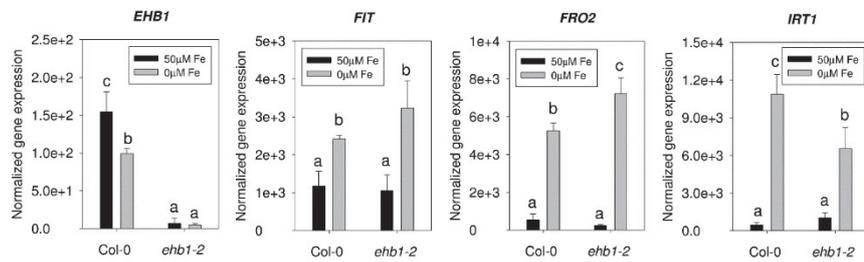
97 **Supplemental Figure S4.** The *ehb1* T-DNA insertion mutants used in this study.

98 **(A)** Schematic representation of the *EHB1* gene and the two alleles used in the study. The primers
99 used for genotyping the two alleles are indicated.

100 **(B)** Detection of full-length *EHB1* transcript. cDNA from wild-type, *EHB1*-overexpressing and
101 two independent homozygous T-DNA insertion plants was tested using primers EHB1f and
102 EHB1r, to amplify the full-length coding *EHB1* sequence. The expected product size was 525 base
103 pairs. As expected, the HA-EHB1 line contained large amounts of EHB1 transcript, while
104 amplification was not possible in the two *ehb1* mutant alleles.

105 Primer sequences are provided in Supplemental Table S1.

106
107
108
109
110
111
112
113
114



115
116
117
118
119
120
121
122
123
124
125
126
127
128
129

Supplemental Figure S5. Gene expression analysis in the *ehb1-2* mutant.

The expression of *EHB1*, *FIT*, *FRO2* and *IRT1* was analyzed in Col-0 and *ehb1-2* plants grown under sufficient (50µM Fe) and iron-deficient (0µM Fe) in the two-week growth system. Error bars indicate standard deviation. Letters indicate statistically significant differences (P<0.05).

130 **Supplemental Table S1.** Primers used in this study.

131 An alphabetically-ordered list of primers appearing in the Materials and Methods section.

Primer name	Primer sequence	Purpose	Origin
2cternsB4	GGGGACAACCTTTGTATAGAAAAGTTGGGT GGGGGCCGTGACCATGACCATG	GATEWAY™-compatible cloning of <i>IRT1vrΔ1</i> (for pBiFC-2in1-CC)	this study
3cterB1	GGGGACAAGTTTGTACAAAAAGCAGGCT ATGCCCCATGGTCATGGTCATGGT	GATEWAY™-compatible cloning of <i>IRT1vrΔ2</i> (for pBiFC-2in1-CC)	this study
3cternsB4	GGGGACAACCTTTGTATAGAAAAGTTGGGT GTCGGTATCGCAAGAGCTGTG	GATEWAY™-compatible cloning of <i>IRT1vrΔ2</i> (for pBiFC-2in1-CC)	this study
4cterB1	GGGGACAAGTTTGTACAAAAAGCAGGCT ATGCCCCATGGTCATGGTCATGGT	GATEWAY™-compatible cloning of <i>IRT1vrΔ3</i> (for pBiFC-2in1-CC)	this study
4cternsB4	GGGGACAACCTTTGTATAGAAAAGTTGGGT GCGAATCATCTTCTTTATTGG	GATEWAY™-compatible cloning of <i>IRT1vrΔ3</i> (for pBiFC-2in1-CC)	this study
5AM1	AAGGTAACATCATTGCGGGGGGCAT GATACCAACTGCGT	Generation of the <i>IRT1vrΔHis</i> fragment	this study
5BM2	ACGCAGTTGGTATCATGCCCCCGCAA ATGATGTTACCTT	Generation of the <i>IRT1vrΔHis</i> fragment	this study
BHLH39rt1	GACGGTTTCTCGAAGCTTG	BHLH039 RT-PCR	Wang et al., 2007
BHLH39rt2	GGTGGCTGCTTAACGTAACAT	BHLH039 RT-PCR	Wang et al., 2007
E1deltaSig1	AGATGGGAGATGCGCAGATAGGCAAGATC GTTCAGGACAT	generation of the <i>EHB1Δsig</i> clone	this study
E1deltaSig2	ATGTCCTGAACGATCTTGCTATCTGCGCA TCTCCCATCT	generation of the <i>EHB1Δsig</i> clone	this study
E1SpelFw	TTTTACTAGTATGGAGAAAACAGAGG AAGAG	Cloning of <i>EHB1</i> for the generation of the pET-StrepII <i>EHB1</i> vector	this study
E1XhoRev	TTTTCTCAGTCAGAGTCCACTACCAC CTG	Cloning of <i>EHB1</i> for the generation of the pET-StrepII <i>EHB1</i> vector	this study
EFc1	ACTTGTACCAGTTGGTTATGGG	EF1Ba RT-PCR	Wang et al., 2007
EFc2	CTGGATGTA CTGTTGTTAGGC	EF1Ba RT-PCR	Wang et al., 2007
EFg1	TCCGAACAATACCAGAACTACG	EF1Balpha (genomic) RT-PCR	Wang et al., 2007
EFg2	CCGGGACATATGGAGGTAAG	EF1Balpha (genomic) RT-PCR	Wang et al., 2007
EHB1B1	GGGGACAAGTTTGTACAAAAAGCAGGCT ATGGAGAAAACAGAGGAAGAG	GATEWAY™-compatible cloning of <i>EHB1</i>	this study
EHB1B2	GGGGACCACTTTGTACAAGAAAGCTGGGT TCAGAGTCCACTACCACCTGGAT	GATEWAY™-compatible cloning of <i>EHB1</i>	this study
EHB1B3	GGGGACAACCTTTGTATAATAAAGTTGATGG AGAAAACAGAGGAAGAG	GATEWAY™-compatible cloning of <i>EHB1</i> and <i>EHB1Δsig</i> (for pBiFC-2in1-CC)	this study
EHB1cterB2	GGGGACCACTTTGTACAAGAAAGCTGGGT TGAGTCCACTACCACCTGGAT	GATEWAY™-compatible cloning of <i>EHB1</i> (for C-terminal fusions)	this study

EHB1f	ATGGAGAAAACAGAGGAAGA	genotyping	this study
EHB1nterB1	GGGGACAAGTTTGTACAAAAAAGCAGGCT TTATGGAGAAAACAGAGGAAGAG	GATEWAY™-compatible cloning of <i>EHB1</i> (for N-terminal fusions)	this study
EHB1r	TCAGAGTCCACTACCACCTG	genotyping	this study
EHB1rt1	GCTTGTCTGAAGATAGCATA	EHB1 RT-PCR	this study
EHB1rt2	CGCATTGACATTCTTCAACAG	EHB1 RT-PCR	this study
FER1_3'	GAAAGGCTGGAACACGACTC	FER1 RT-PCR	Klatte et al., 2009
FER1_5'	ACGCACTCTCGTCTTTCACC	FER1 RT-PCR	Klatte et al., 2009
FITrt1	GGAGAAGGTGTTGCTCCATC	FIT RT-PCR	Wang et al., 2007
FITrt2	TCCGGAGAAGGAGAGCTTAG	FIT RT-PCR	Wang et al., 2007
FL1B2Stop	GGGGACCACTTTGTACAAGAAAGCTG GGTTTAAGCCCATTTGGCGATAATCG	Amplification of <i>IRT</i> coding sequence. Reverse primer. Includes the native stop codon	this study
FRO2rt1	CTTGGTCATCTCCGTGAGC	FRO2 RT-PCR	Wang et al., 2007
FRO2rt2	AAGATGTTGGAGATGGACGG	FRO2 RT-PCR	Wang et al., 2007
G6B1	GGGGACAAGTTTGTACAAAAAAGCAG GCTATGAGTAAAGGAGAAGAAGT	GATEWAY™-compatible cloning of <i>GFP6</i>	this study
G6B2	GGGGACCACTTTGTACAAGAAAGCTG GGTTTAGTGGTGGTGGTGGT	GATEWAY™-compatible cloning of <i>GFP6</i>	this study
I1B1	GGGGACAAGTTTGTACAAAAAAGCAG GCTTTATGGCTTCAAATTCAGCACTT	Amplification of <i>IRT</i> coding sequence. Forward primer.	Ivanov et al., 2014
I1LatgB1	GGGGACAAGTTTGTACAAAAAAGCAGGCT ATGTCCATGGCCACGAGCCTATA	GATEWAY™-compatible cloning of <i>IRT1vr</i> , <i>IRT1vrΔ1</i> and <i>IRT1vrΔHis</i> (for pBiFC-2in1-CC)	this study
I1LB1	GGGGACAAGTTTGTACAAAAAAGCAGGCT TTTCCATGGCCACGAGCCTATA	GATEWAY™-compatible cloning of <i>IRT1vr</i> (for N-terminal fusions)	this study
I1LB2	GGGGACCACTTTGTACAAGAAAGCTGGGT TTATCGGTATCGCAAGAGCTGTG	GATEWAY™-compatible cloning of <i>IRT1vr</i> (with an artificial stop codon)	this study
I1LnsB4	GGGGACAACCTTTGTATAGAAAAGTTGGGT GTCGGTATCGCAAGAGCTGTG	GATEWAY™-compatible cloning of <i>IRT1vr</i> , <i>IRT1vrΔ2</i> and <i>IRT1vrΔHis</i> (for pBiFC-2in1-CC)	this study
IRT1rt1	AAGCTTTGATCACGGTTGG	IRT1 RT-PCR	Wang et al., 2007
IRT1rt2	TTAGGTCCCATGAACTCCG	IRT1 RT-PCR	Wang et al., 2007
LB2SAIL	GCTTCCTATTATATCTTCCCAAATTACCAAT ACA	genotyping of the <i>ehb1-2</i> allele	http://signal.salk.edu/dnaprimers.2.html
LB2SALK	ACCGAGCTCGAATTTCCCG	genotyping of the <i>ehb1-1</i> allele	Jaillais et al., 2006
Yeast_seqF	AATACCACTACAATGGAT	Amplification and sequencing of fragments from the pGADT7 vector	this study
Yeast_seqR	GAGATGGTGCACGATGCAC	Amplification and sequencing of fragments from the pGADT7 vector	this study

132

133

134 **Supplemental References**

135

136 Ivanov R., Brumbarova T., Blum A., Jantke A-M., Fink-Straube C. and Bauer P. (2014) SORTING
137 NEXIN1 is required for modulating the trafficking and stability of the Arabidopsis IRON-
138 REGULATED TRANSPORTER1. *Plant Cell* 26, 1294-1307.

139

140 Jaillais, Y., Fobis-Loisy, I., Miege, C., Rollin, C., and Gaude, T. (2006). AtSNX1 defines an
141 endosome for auxin-carrier trafficking in Arabidopsis. *Nature* 443, 106-109.

142

143 Klatte, M., Schuler, M, Wirtz, M., Fink-Straube, C, Hell, R. and Bauer P. (2008) The Analysis of
144 Arabidopsis Nicotianamine Synthase Mutants Reveals Functions for Nicotianamine in Seed
145 Iron Loading and Iron Deficiency Responses. *Plant Physiol* 150, 257-271.

146

147 Wang, H.Y., Klatte, M., Jakoby, M., Baumlein, H., Weisshaar, B., and Bauer, P. (2007). Iron
148 deficiency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis*
149 *thaliana*. *Planta* 226, 897-908.

150

151

152

153 **Supplemental Movie Legend**

154

155 **Supplemental Movie S1.** Molecular dynamics simulation of EHB1 membrane binding.

156 EHB1 protein starts in an unbound state and freely diffuses to the membrane, attracted by the
157 positive charges in the calcium binding site (white spheres) and the positively charged residues on
158 the protein surface. Once bound, the protein stays either in a perpendicular orientation by the
159 interacting calcium ions, or rests on its side, interacting through the polybasic patch (magenta
160 sticks). It is expected that the latter orientation is stabilized by interactions with PtdIns4P present
161 in the plant plasma membrane. The movie shows the yellow trajectory depicted in Figures 3E and
162 3G. The simulation time is 1 μ s.

163

7 Curriculum vitae

Personal information

Name: Stephan Nicolás Schott Verdugo
 Place of Birth: Valdivia, 5090000, Chile
 Date of Birth: 7th November 1989

Education

HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF, DÜSSELDORF, GERMANY

PhD Student, Computational biology 2016-Present

Institut für Pharmazeutische und Medizinische Chemie

UNIVERSIDAD DE TALCA, TALCA, CHILE

PhD Student, Program in Applied Sciences 2015-Present

UNIVERSIDAD AUSTRAL DE CHILE, VALDIVIA, CHILE

Biochemist 2014

Thesis: Citrato interacciona con fructosa 1,6 bisfosfatasa en la interfaz C1-C4 con efectos controvertidos sobre la actividad catalítica (“Citrate interacts with fructose 1,6-bisphosphatase in the C1-C4 interface with controverted effects on the catalytic activity”)

Honors: High distinction

Licentiate in Biochemistry 2012

INSTITUTO ALEMÁN CARLOS ANWANDTER

Primary and Secondary Education 2007

Distinctions

Best enrolled Biochemistry student 2008

Best annual score in the Biochemistry undergraduate program 2009, 2010, 2011

Teaching experience

Assistant in General Chemistry practical lectures (UACH) 2009-2013 (6 Semesters)

- Assistance to students during the development of the practical lecture

Assistant in Biotechnology practical lectures (UACH) 2013 (1 Semester)

- Preparation of materials and solutions for practical lectures

Assistant in Molecular Modeling course for Pharmacy and MSc. in Industrial Pharmacy students (HHU) 2016-Present

- Theory seminars
- Assistance to students during the development of practical lectures

Related experience

Research Assistant in Molecular enzymology laboratory, Instituto de Bioquímica y Microbiología, Universidad Austral de Chile April 2014 - March 2015

- Help in development of research project, particularly in protein expression and enzymatic assays. Assistance in maintenance and installation of laboratory equipment

Skills

Computational skills: Molecular modeling and related software (PyMOL, VMD, Rosetta)

Molecular dynamics (AMBER, NAMD, GROMACS)

UNIX system administration and maintenance

Working knowledge in Python programming

Generation of publication-quality vector graphics in Inkscape

Biochemistry skills: Protein expression and purification

FPLC

Enzymatic assays

Molecular biology techniques

Language skills

Spanish – Mother tongue

English – Proficient user. B1 certification (Preliminary English Test, 2006)

German –Independent user. B2 certification (Deutsches Sprachdiplom Stufe II, 2007).

Publications

Articles

Schott-Verdugo, S., Müller, L., Classen, E., Gohlke, H., & Groth, G. (2019). Structural Model of the ETR1 Ethylene Receptor Transmembrane Sensor Domain. *Scientific Reports*, 9(1), 1–14.

Khan, I., Gratz, R., Denezhkin, P., Schott-Verdugo, S. N., Angrand, K., Genders, L., Basgaran, R.M., Fink-Straube, C., Brumbarova, T. Gohlke, H., Bauer, P. & Ivanov, R. (2019). Calcium-Promoted Interaction between the C2-Domain Protein EHB1 and Metal Transporter IRT1 Inhibits Arabidopsis Iron Acquisition. *Plant Physiology*, 180(3), 1564–1581.

Schott-Verdugo, S., & Gohlke, H. (2019). PACKMOL-Memgen: A Simple-To-Use, Generalized Workflow for Membrane-Protein-Lipid-Bilayer System Building. *Journal of Chemical Information and Modeling*, 59, 2522–2528.

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., Sucurro, A., Zuccaro, A., Gould, S.B., Bauer, P., Schmitt, L. & Gohlke, H. (2019). Nutrient exchange in arbuscular mycorrhizal symbiosis from a thermodynamic point of view. *New Phytologist*, 222(2), 1043–1053.

Schott, S., Valdebenito, B., Bustos, D., Gomez-Porras, J. L., Sharma, T., & Dreyer, I. (2016). Cooperation through Competition—Dynamics and Microeconomics of a Minimal Nutrient Trade System in Arbuscular Mycorrhizal Symbiosis. *Frontiers in Plant Science*, 7, 912.

Zarate, X., Schott-Verdugo, S., Rodriguez-Serrano, A., & Schott, E. (2016). The nature of the donor motif in acceptor-bridge-donor dyes as an influence in the electron photo-injection mechanism in DSSCs. *The Journal of Physical Chemistry A*, 120(9), 1613-1624.

Arrue, L., Zarate, X., Schott-Verdugo, S., & Schott, E. (2015). Substituted phenylhydrazono derivatives of curcumin as new ligands, a theoretical study. *Chemical Physics Letters*, 623, 42-45.

Oral presentations

“Dimerization Free Energy Profile and Equilibrium of the Phospholipase PlaF from *Pseudomonas aeruginosa*” Workshop on Computer Simulation and Theory of Macromolecules, Hünfeld, Germany (2018)

“PACKMOL MEMGEN: PACKMOL MEMbrane GENerator” AMBER Developers Meeting, Tampa, Florida, U.S.A. (2019)

Poster presentations

“Structural model of the transmembrane sensor domain of the ETR1 ethylene receptor.” 2nd CRC 1208 Conference, Düsseldorf, Germany (2019)

“TopContact: Protein contact meta-predictions through machine learning.” CASP13
Meeting, Cancún, México (2018)

“Free energy profile and the dimer-to-monomer equilibrium of the phospholipase PlaF from
Pseudomonas aeruginosa.” NIC Symposium, Jülich, Germany (2018)

“Structural model of the ethylene receptor (ETR1) transmembrane domain and energetics of
phospholipase B (PlbF) dimerization and loading.” CRC 1208 Conference, Düsseldorf,
Germany (2017)

“Evaluation of the Interaction of Citrate with fructose 1,6-bisphosphatase.” XXXVII
Meeting of the Society for Biochemistry and Molecular Biology of Chile, Valdivia, Chile
(2014)

Bibliography

1. Robertson, J.L., *The lipid bilayer membrane and its protein constituents*. J. Gen. Physiol., 2018. **150**(11): p. 1472-1483.
2. Collawn, J.F. and Z. Bebök, *Structure and Functions of Biomembranes*, in *Free Radical Effects on Membranes*, S. Matalon, Editor. 2008, Elsevier. p. 1-21.
3. Harayama, T. and H. Riezman, *Understanding the diversity of membrane lipid composition*. Nat. Rev. Mol. Cell Biol., 2018. **19**(5): p. 281-296.
4. Tan, S., H.T. Tan, and M.C. Chung, *Membrane proteins and membrane proteomics*. Proteomics, 2008. **8**(19): p. 3924-32.
5. Guidotti, G., *Membrane proteins*. Annu. Rev. Biochem., 1972. **41**: p. 731-52.
6. Segrest, J.P., et al., *A molecular theory of lipid-protein interactions in the plasma lipoproteins*. FEBS Lett., 1974. **38**(3): p. 247-253.
7. Eisenberg, D., et al., *Hydrophobic moments and protein structure*. Faraday Symposia of the Chemical Society, 1982. **17**: p. 109.
8. Singer, S.J. and G.L. Nicolson, *The fluid mosaic model of the structure of cell membranes*. Science, 1972. **175**(4023): p. 720-31.
9. Nicolson, G.L., *The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years*. Biochim. Biophys. Acta, 2014. **1838**(6): p. 1451-66.
10. Fujiwara, T., et al., *Phospholipids undergo hop diffusion in compartmentalized cell membrane*. J. Cell Biol., 2002. **157**(6): p. 1071-81.
11. Marrink, S.J., et al., *Computational Modeling of Realistic Cell Membranes*. Chem. Rev., 2019. **119**(9): p. 6184-6226.
12. Marianayagam, N.J., M. Sunde, and J.M. Matthews, *The power of two: protein dimerization in biology*. Trends Biochem. Sci., 2004. **29**(11): p. 618-25.
13. Mulgrew-Nesbitt, A., et al., *The role of electrostatics in protein-membrane interactions*. Biochim. Biophys. Acta, 2006. **1761**(8): p. 812-26.
14. Lindahl, E. and M.S. Sansom, *Membrane proteins: molecular dynamics simulations*. Curr. Opin. Struct. Biol., 2008. **18**(4): p. 425-31.
15. D.A. Case, I.Y.B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, *AMBER 2018*. 2018, University of California: San Francisco.
16. Van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: Where they are and how they behave*. Nat. Rev. Mol. Cell Biol., 2008. **9**: p. 112-124.
17. Sud, M., et al., *LMSD: LIPID MAPS structure database*. Nucleic Acids Res., 2007. **35**(Database issue): p. D527-32.
18. *LIPID MAPS Lipidomics Gateway*. [cited 2019 Nov 4th]; Available from: <http://www.lipidmaps.org/resources/databases/index.php?tab=lmsd>.
19. Rowlett, V.W., et al., *Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation*. J. Bacteriol., 2017. **199**(13).
20. Clejan, S., et al., *Membrane lipid composition of obligately and facultatively alkalophilic strains of Bacillus spp*. J. Bacteriol., 1986. **168**(1): p. 334-40.
21. Mileykovskaya, E. and W. Dowhan, *Role of membrane lipids in bacterial division-site selection*. Curr. Opin. Microbiol., 2005. **8**(2): p. 135-42.

22. Leonard, A.N., et al., *Developing and Testing of Lipid Force Fields with Applications to Modeling Cellular Membranes*. Chem. Rev., 2019. **119**(9): p. 6227-6269.
23. Mileykovskaya, E. and W. Dowhan, *Cardiolipin membrane domains in prokaryotes and eukaryotes*. Biochim. Biophys. Acta, 2009. **1788**(10): p. 2084-91.
24. Falkenburger, B.H., et al., *Phosphoinositides: lipid regulators of membrane proteins*. The Journal of Physiology, 2010. **588**(Pt 17): p. 3179-85.
25. Marques, J.T., H.S. Marinho, and R.F.M. de Almeida, *Sphingolipid hydroxylation in mammals, yeast and plants - An integrated view*. Prog. Lipid Res., 2018. **71**: p. 18-42.
26. Lingwood, D. and K. Simons, *Lipid rafts as a membrane-organizing principle*. Science, 2010. **327**(5961): p. 46-50.
27. Zhang, G., T.C. Meredith, and D. Kahne, *On the essentiality of lipopolysaccharide to Gram-negative bacteria*. Curr. Opin. Microbiol., 2013. **16**(6): p. 779-85.
28. Raetz, C.R. and C. Whitfield, *Lipopolysaccharide endotoxins*. Annu. Rev. Biochem., 2002. **71**: p. 635-700.
29. Saenz, J.P., et al., *Functional convergence of hopanoids and sterols in membrane ordering*. Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(35): p. 14236-40.
30. Saenz, J.P., et al., *Hopanoids as functional analogues of cholesterol in bacterial membranes*. Proc. Natl. Acad. Sci. U. S. A., 2015. **112**(38): p. 11971-6.
31. Poger, D. and A.E. Mark, *The relative effect of sterols and hopanoids on lipid bilayers: when comparable is not identical*. J. Phys. Chem. B, 2013. **117**(50): p. 16129-40.
32. Holthuis, J.C. and A.K. Menon, *Lipid landscapes and pipelines in membrane homeostasis*. Nature, 2014. **510**(7503): p. 48-57.
33. Zhang, Y.M. and C.O. Rock, *Membrane lipid homeostasis in bacteria*. Nat. Rev. Microbiol., 2008. **6**(3): p. 222-233.
34. Vinothkumar, K.R. and R. Henderson, *Structures of membrane proteins*. Q. Rev. Biophys., 2010. **43**(1): p. 65-158.
35. Rojas, E. and M. Luxoro, *Micro-Injection of Trypsin into Axons of Squid*. Nature, 1963. **199**: p. 78-9.
36. Tusnady, G.E., Z. Dosztanyi, and I. Simon, *Transmembrane proteins in the Protein Data Bank: identification and classification*. Bioinformatics, 2004. **20**(17): p. 2964-72.
37. Chaturvedi, D. and R. Mahalakshmi, *Transmembrane beta-barrels: Evolution, folding and energetics*. Biochim. Biophys. Acta, 2017. **1859**(12): p. 2467-2482.
38. White, S.H., et al., *How membranes shape protein structure*. J. Biol. Chem., 2001. **276**(35): p. 32395-8.
39. Lee, A.G., *How lipids affect the activities of integral membrane proteins*. Biochim. Biophys. Acta, 2004. **1666**(1-2): p. 62-87.
40. Almeida, J.G., et al., *Membrane proteins structures: A review on computational modeling tools*. Biochim. Biophys. Acta, 2017. **1859**: p. 2021-2039.
41. Fagerberg, L., et al., *Prediction of the human membrane proteome*. Proteomics, 2010. **10**(6): p. 1141-9.
42. Rask-Andersen, M., M.S. Almen, and H.B. Schioth, *Trends in the exploitation of novel drug targets*. Nat. Rev. Drug Discov., 2011. **10**(8): p. 579-90.
43. Deisenhofer, J., et al., *Structure of the protein subunits in the photosynthetic reaction centre of Rhodospseudomonas viridis at 3A resolution*. Nature, 1985. **318**(6047): p. 618-24.
44. Jo, S., et al., *CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes*. Biophys. J., 2009. **97**: p. 50-58.
45. Pastor, R.W. and A.D. Mackerell, Jr., *Development of the CHARMM Force Field for Lipids*. J. Phys. Chem. Lett., 2011. **2**(13): p. 1526-1532.

46. Jambeck, J.P. and A.P. Lyubartsev, *Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids*. J. Phys. Chem. B, 2012. **116**(10): p. 3164-79.
47. Ermilova, I. and A.P. Lyubartsev, *Extension of the Slipids Force Field to Polyunsaturated Lipids*. J. Phys. Chem. B, 2016. **120**(50): p. 12826-12842.
48. Schmid, N., et al., *Definition and testing of the GROMOS force-field versions 54A7 and 54B7*. Eur. Biophys. J., 2011. **40**: p. 843-856.
49. Marzuoli, I., C. Margreitter, and F. Fraternali, *Lipid Head Group Parameterization for GROMOS 54A8: A Consistent Approach with Protein Force Field Description*. J. Chem. Theory Comput., 2019. **15**(10): p. 5175-5193.
50. Skjevik, A.A., et al., *LIPID11: a modular framework for lipid simulations using amber*. J. Phys. Chem. B, 2012. **116**(36): p. 11124-36.
51. Dickson, C.J., et al., *Lipid14: The Amber Lipid Force Field*. J. Chem. Theory Comput., 2014. **10**: p. 865-879.
52. Lyubartsev, A.P. and A.L. Rabinovich, *Force Field Development for Lipid Membrane Simulations*. Biochim. Biophys. Acta, 2016. **1858**(10): p. 2483-2497.
53. Wang, J., et al., *Development and testing of a general amber force field*. J. Comput. Chem., 2004. **25**(9): p. 1157-74.
54. Rosso, L. and I.R. Gould, *Structure and dynamics of phospholipid bilayers using recently developed general all-atom force fields*. J. Comput. Chem., 2008. **29**(1): p. 24-37.
55. Dickson, C.J., et al., *GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid*. Soft Matter, 2012. **8**(37): p. 9617.
56. Pluhackova, K., et al., *A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers*. J. Phys. Chem. B, 2016. **120**(16): p. 3888-903.
57. Sandoval-Perez, A., K. Pluhackova, and R.A. Bockmann, *Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface*. J. Chem. Theory Comput., 2017. **13**(5): p. 2310-2321.
58. Jo, S., et al., *CHARMM-GUI 10 years for biomolecular modeling and simulation*. J. Comput. Chem., 2017. **38**(15): p. 1114-1124.
59. Shell, M.S., A. Panagiotopoulos, and A. Pohorille, *Methods Based on Probability Distributions and Histograms*, in *Free Energy Calculations*, C. Chipot and A. Pohorille, Editors. 2007. p. 77-118.
60. Trzesniak, D., A.P. Kunz, and W.F. van Gunsteren, *A comparison of methods to compute the potential of mean force*. Chemphyschem, 2007. **8**(1): p. 162-9.
61. Kästner, J., *Umbrella sampling*. Wiley Interdiscip. Rev. Comput. Mol. Sci, 2011. **1**: p. 932-942.
62. Provasi, D., J.M. Johnston, and M. Filizola, *Lessons from free energy simulations of δ -opioid receptor homodimers involving the fourth transmembrane helix*. Biochemistry, 2010. **49**: p. 6771-6776.
63. Johnston, J.M., et al., *Assessing the Relative Stability of Dimer Interfaces in G Protein-Coupled Receptors*. PLOS Comp. Biol., 2012. **8**.
64. Baker, D. and A. Sali, *Protein structure prediction and structural genomics*. Science, 2001. **294**(5540): p. 93-6.
65. Lam, S.D., et al., *An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences*. Acta Crystallogr. D, 2017. **73**(Pt 8): p. 628-640.
66. Forrest, L.R., C.L. Tang, and B. Honig, *On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins*. Biophys. J., 2006. **91**(2): p. 508-17.

67. Nguyen, H., et al., *Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent*. J. Am. Chem. Soc., 2014. **136**: p. 13959-13962.
68. Lee, J., P.L. Freddolino, and Y. Zhang, *Ab Initio Protein Structure Prediction*, in *From Protein Structure to Function with Bioinformatics*, D.J. Rigden, Editor. 2017, Springer. p. 3-35.
69. Bu, L., W. Im, and C.L. Brooks, *Membrane assembly of simple helix homo-oligomers studied via molecular dynamics simulations*. Biophys. J., 2007. **92**: p. 854-63.
70. Im, W. and C.L. Brooks, 3rd, *De novo folding of membrane proteins: an exploration of the structure and NMR properties of the fd coat protein*. J. Mol. Biol., 2004. **337**(3): p. 513-9.
71. Jones, D.T., *Successful ab initio prediction of the tertiary structure of NK-lysin using multiple sequences and recognized supersecondary structural motifs*. Proteins, 1997. **29**: p. 185-191.
72. Simons, K.T., et al., *Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions*. J. Mol. Biol., 1997. **268**(1): p. 209-25.
73. Kaufmann, K.W., et al., *Practically useful: What the Rosetta protein modeling suite can do for you*. Biochemistry, 2010. **49**: p. 2987-2998.
74. Rubenstein, A.B., et al., *Systematic Comparison of Amber and Rosetta Energy Functions for Protein Structure Evaluation*. J. Chem. Theory Comput., 2018. **14**(11): p. 6015-6025.
75. Rohl, C.A., et al., *Protein Structure Prediction Using Rosetta*. Methods Enzymol., 2004. **383**: p. 66-93.
76. Yarov-Yarovoy, V., J. Schonbrun, and D. Baker, *Multipass membrane protein structure prediction using Rosetta*. Proteins, 2006. **62**: p. 1010-1025.
77. Pellegrini-Calace, M., A. Carotti, and D.T. Jones, *Folding in lipid membranes (FILM): A novel method for the prediction of small membrane protein 3D structures*. Proteins, 2003. **50**: p. 537-545.
78. Vendruscolo, M., E. Kussell, and E. Domany, *Recovery of protein structure from contact maps*. Fold Des., 1997. **2**(5): p. 295-306.
79. Bowers, P.M., C.E. Strauss, and D. Baker, *De novo protein structure determination using sparse NMR data*. J. Biomol. NMR, 2000. **18**(4): p. 311-8.
80. Sathyapriya, R., et al., *Defining an essence of structure determining residue contacts in proteins*. PLOS Comp. Biol., 2009. **5**(12): p. e1000584.
81. Kim, D.E., et al., *One contact for every twelve residues allows robust and accurate topology-level protein structure modeling*. Proteins, 2014. **82 Suppl 2**: p. 208-18.
82. Neher, E., *How frequent are correlated changes in families of protein sequences?* Proc. Natl. Acad. Sci. U. S. A., 1994. **91**(1): p. 98-102.
83. Gobel, U., et al., *Correlated mutations and residue contacts in proteins*. Proteins, 1994. **18**(4): p. 309-17.
84. Marks, D.S., T.a. Hopf, and C. Sander, *Protein structure prediction from sequence variation*. Nat. Biotechnol., 2012. **30**: p. 1072-80.
85. Weigt, M., et al., *Identification of direct residue contacts in protein-protein interaction by message passing*. Proc. Natl. Acad. Sci. U. S. A., 2008. **106**(1): p. 67-72.
86. Morcos, F., et al., *Direct-coupling analysis of residue coevolution captures native contacts across many protein families*. Proc. Natl. Acad. Sci. U. S. A., 2011. **108**(49): p. E1293-301.
87. Jones, D.T., et al., *PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments*. Bioinformatics, 2012. **28**: p. 184-190.

88. Marks, D.S., et al., *Protein 3D structure computed from evolutionary sequence variation*. PLOS ONE, 2011. **6**.
89. Kaján, L., et al., *FreeContact: fast and free software for protein contact prediction from residue co-evolution*. BMC Bioinform., 2014. **15**: p. 85.
90. Seemayer, S., M. Gruber, and J. Soding, *CCMpred--fast and precise prediction of protein residue-residue contacts from correlated mutations*. Bioinformatics, 2014. **30**(21): p. 3128-30.
91. Hopf, T.A., et al., *Sequence co-evolution gives 3D contacts and structures of protein complexes*. eLife, 2014. **3**: p. 1-45.
92. Schaarschmidt, J., et al., *Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age*. Proteins, 2018. **86 Suppl 1**: p. 51-66.
93. Shrestha, R., et al., *Assessing the accuracy of contact predictions in CASP13*. Proteins, 2019. **87**(12): p. 1058-1068.
94. Jones, D.T. and S.M. Kandathil, *High precision in protein contact prediction using fully convolutional neural networks and minimal sequence features*. Bioinformatics, 2018. **34**(19): p. 3308-3315.
95. Jones, D.T., et al., *MetaPSICOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins*. Bioinformatics, 2015. **31**: p. 999-1006.
96. Wang, S., et al., *Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model*. PLOS Comp. Biol., 2017. **13**(1): p. e1005324.
97. Michel, M., D. Menendez Hurtado, and A. Elofsson, *PconsC4: fast, accurate and hassle-free contact predictions*. Bioinformatics, 2019. **35**(15): p. 2677-2679.
98. Adhikari, B., J. Hou, and J. Cheng, *DNCON2: improved protein contact prediction using two-level deep convolutional neural networks*. Bioinformatics, 2018. **34**(9): p. 1466-1472.
99. Liu, Y., et al., *Enhancing Evolutionary Couplings with Deep Convolutional Neural Networks*. Cell Systems, 2018. **6**(1): p. 65-74 e3.
100. Khelashvili, G., et al., *Computational modeling of the N-terminus of the human dopamine transporter and its interaction with PIP2 -containing membranes*. Proteins, 2015. **83**(5): p. 952-69.
101. Watschinger, K., et al., *Catalytic residues and a predicted structure of tetrahydrobiopterin-dependent alkylglycerol mono-oxygenase*. Biochem. J., 2012. **443**(1): p. 279-86.
102. Antala, S., et al., *Computation and functional studies provide a model for the structure of the zinc transporter hZIP4*. J. Biol. Chem., 2015. **290**: p. 17796-17805.
103. Zhang, T., et al., *Crystal structures of a ZIP zinc transporter reveal a binuclear metal center in the transport pathway*. Sci. Adv., 2017. **3**: p. e1700344.
104. Neljubow, D., *Ueber die horizontale Nutation der Stengel von Pisum sativum und einiger anderen Pflanzen*. Beihefte zum Botanischen Centralblatt, 1901. **10**(3): p. 128-139.
105. Crocker, W. and L.I. Knight, *Effect of Illuminating Gas and Ethylene upon Flowering Carnations*. Bot. Gaz., 1908. **46**(4): p. 259-276.
106. Knight, L.I., *Toxicity of Smoke*. Bot. Gaz., 1913. **55**(5): p. 337-371.
107. Harvey, E.M. and R.C. Rose, *The Effects of Illuminating Gas on Root Systems*. Bot. Gaz., 1915. **60**(1): p. 27-44.
108. Doubt, S.L., *The Response of Plants to Illuminating Gas*. Bot. Gaz., 1917. **63**(3): p. 209-224.
109. Knight, L.I., R.C. Rose, and W. Crocker, *Effect of Variows Gases and Vapors upon Etiolated Seedlings of the Sweet Pea*. Science, 1910. **31**(799): p. 635-36.

110. Bakshi, A., et al., *History of Research on the Plant Hormone Ethylene*. J. Plant Growth Regul., 2015. **34**(4): p. 809-827.
111. Moore, T.C., *Ethylene*, in *Biochemistry and Physiology of Plant Hormones*. 1979, Springer. p. 208-229.
112. Gane, R., *Production of Ethylene by Some Ripening Fruits*. Nature, 1934. **134**(3400): p. 1008-1008.
113. Gane, R., *The Formation of Ethylene by Plant Tissues, and its Significance in the Ripening of Fruits*. Journal of Pomology and Horticultural Science, 1935. **13**(4): p. 351-358.
114. Laties, G.G., *Franklin Kidd, Charles West and F.F. Blackman: The start of modern postharvest physiology*. Postharvest Biol. Technol., 1995. **5**(1-2): p. 1-10.
115. Wang, K.L., H. Li, and J.R. Ecker, *Ethylene biosynthesis and signaling networks*. Plant Cell, 2002. **14 Suppl**: p. S131-51.
116. Argueso, C.T., M. Hansen, and J.J. Kieber, *Regulation of Ethylene Biosynthesis*. J. Plant Growth Regul., 2007. **26**(2): p. 92-105.
117. Pech, J.-C., et al., *Ethylene and Fruit Ripening*, in *The Plant Hormone Ethylene*, M.T. McManus, Editor. 2018, John Wiley & Sons. p. 275-304.
118. Puri, M., *How access to energy can influence food losses. A brief overview*, in *Environment and Natural Resources Management*. 2016, Climate and Environment division, FAO: Rome.
119. *The Surprising Science Behind the World's Most Popular Fruit*. National Geographic 24/10/2017 [cited 2019 04/12]; Available from: <https://www.nationalgeographic.com/environment/urban-expeditions/food/food-journeys-graphic/>.
120. Wills, R.B. and J.B. Golding, *Reduction of energy usage in postharvest horticulture through management of ethylene*. J. Sci. Food Agric., 2015. **95**(7): p. 1379-84.
121. Monforti-Ferrario, F., et al., *Energy use in the EU food sector: State of play and opportunities for improvement*, in *Publications Office of the European Union*, F. Monforti-Ferrario and I. Pinedo Pascua, Editors. 2015, Institute for Energy and Transport and Institute for Environment and Sustainability: Luxembourg.
122. Field, R.J., *The Effect of Low Temperature on Ethylene Production by Leaf Tissue of Phaseolus vulgaris L.* Ann. Bot., 1981. **47**(2): p. 215-223.
123. Tong, C.B.S. and S.F. Yang, *Chilling-induced ethylene production by beans and peas*. J. Plant Growth Regul., 1987. **6**(4): p. 201-208.
124. Wang, C.Y. and D.O. Adams, *Chilling-Induced Ethylene Production in Cucumbers (Cucumis sativus L.)*. Plant Physiol., 1982. **69**(2): p. 424-7.
125. Wills, R.B.H., et al., *Reduction of energy usage during storage and transport of bananas by management of exogenous ethylene levels*. Postharvest Biol. Technol., 2014. **89**: p. 7-10.
126. Kader, A.A., *Prevention of ripening in fruits by use of controlled atmospheres*. Food Technol., 1980. **34**(3): p. 51-54.
127. Serek, M., E.C. Sisler, and M.S. Reid, *Novel Gaseous Ethylene Binding Inhibitor Prevents Ethylene Effects in Potted Flowering Plants*. J. Am. Soc. Hort. Sci., 1994. **119**(6): p. 1230-1233.
128. Park, Y.-M., *1-MCP application for horticultural commodities in Korea: Practical potential and future task*. Hortic. Environ. Biotechnol., 2013. **53**(6): p. 441-446.
129. Payasi, A. and G.G. Sanwal, *Ripening of Climacteric Fruits and Their Control*. J. Food Biochem., 2010. **34**(4): p. 679-710.
130. Bleecker, A.B., et al., *Insensitivity to Ethylene Conferred by a Dominant Mutation in Arabidopsis thaliana*. Science, 1988. **241**(4869): p. 1086-9.

131. Hall, A.E., et al., *The relationship between ethylene binding and dominant insensitivity conferred by mutant forms of the ETR1 ethylene receptor*. Plant Physiol., 1999. **121**: p. 291-300.
132. Schaller, G.E., et al., *The ethylene response mediator ETR1 from Arabidopsis forms a disulfide-linked dimer*. J. Biol. Chem., 1995. **270**(21): p. 12526-30.
133. Aravind, L. and C.P. Ponting, *The GAF domain: an evolutionary link between diverse phototransducing proteins*. Trends Biochem. Sci., 1997. **22**(12): p. 458-459.
134. Chang, C., et al., *Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators*. Science, 1993. **262**(5133): p. 539-44.
135. O'Malley, R.C., et al., *Ethylene-binding activity, gene expression levels, and receptor system output for ethylene receptor family members from Arabidopsis and tomato*. The Plant Journal, 2005. **41**(5): p. 651-9.
136. Wang, W., et al., *Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from Arabidopsis is not required for signal transmission*. Proc. Natl. Acad. Sci. U. S. A., 2003. **100**(1): p. 352-7.
137. Wang, W., et al., *Identification of important regions for ethylene binding and signaling in the transmembrane domain of the ETR1 ethylene receptor of Arabidopsis*. Plant Cell, 2006. **18**: p. 3429-3442.
138. Wuriyangan, H., et al., *The Ethylene Receptor ETR2 Delays Floral Transition and Affects Starch Accumulation in Rice*. Plant Cell, 2009. **21**: p. 1473-1494.
139. Shakeel, S.N., et al., *Mechanisms of signal transduction by ethylene: Overlapping and non-overlapping signalling roles in a receptor family*. AoB PLANTS, 2013. **5**: p. 1-16.
140. Bisson, M.M.A. and G. Groth, *New insight in ethylene signaling: Autokinase activity of ETR1 modulates the interaction of receptors and EIN2*. Molecular Plant, 2010. **3**: p. 882-889.
141. Gallie, D.R., *Ethylene receptors in plants - why so much complexity?* F1000Prime Reports, 2015. **7**: p. 39.
142. Chen, Y.F., et al., *Localization of the ethylene receptor ETR1 to the endoplasmic reticulum of Arabidopsis*. J. Biol. Chem., 2002. **277**: p. 19861-19866.
143. Ju, C., et al., *CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(47): p. 19486-91.
144. Qiao, H., et al., *Interplay between ethylene, ETP1/ETP2 F-box proteins, and degradation of EIN2 triggers ethylene responses in Arabidopsis*. Genes Dev., 2009. **23**(4): p. 512-21.
145. Binder, B.M., et al., *The Arabidopsis EIN3 binding F-Box proteins EBF1 and EBF2 have distinct but overlapping roles in ethylene signaling*. Plant Cell, 2007. **19**(2): p. 509-23.
146. Li, W., et al., *EIN2-directed translational regulation of ethylene signaling in Arabidopsis*. Cell, 2015. **163**(3): p. 670-83.
147. Zhang, F., et al., *EIN2 mediates direct regulation of histone acetylation in the ethylene response*. Proc. Natl. Acad. Sci. U. S. A., 2017. **114**(38): p. 10274-10279.
148. Müller-Dieckmann, H.-J., A.A. Grantz, and S.-H. Kim, *The structure of the signal receiver domain of the Arabidopsis thaliana ethylene receptor ETR1*. Structure, 1999. **7**(12): p. 1547-1556.
149. Mayerhofer, H., et al., *Structural model of the cytosolic domain of the plant ethylene receptor 1 (ETR1)*. J. Biol. Chem., 2015. **290**: p. 2644-2658.
150. Milic, D., et al., *Recognition motif and mechanism of ripening inhibitory peptides in plant hormone receptor ETR1*. Sci. Rep., 2018. **8**(1): p. 3890.
151. Rodriguez, F.I., et al., *A copper cofactor for the ethylene receptor ETR1 from Arabidopsis*. Science, 1999. **283**: p. 996-998.

152. Thompson, J.S., R.L. Harlow, and J.F. Whitney, *Copper(I)-olefin complexes. Support for the proposed role of copper in the ethylene effect in plants*. J. Am. Chem. Soc., 1983. **105**(11): p. 3522-3527.
153. Burg, S.P. and E.A. Burg, *Molecular requirements for the biological activity of ethylene*. Plant Physiol., 1967. **42**(1): p. 144-52.
154. Light, K.M., et al., *Perception of the plant hormone ethylene: known-knowns and known-unknowns*. J. Biol. Inorg. Chem., 2016. **21**: p. 715-728.
155. Bleecker, A.B., et al., *The ethylene-receptor family from Arabidopsis: structure and function*. Philosophical Transactions of the Royal Society B, 1998. **353**(1374): p. 1405-12.
156. McDaniel, B.K. and B.M. Binder, *Ethylene receptor 1 (ETR1) is sufficient and has the predominant role in mediating inhibition of ethylene responses by silver in Arabidopsis thaliana*. J. Biol. Chem., 2012. **287**: p. 26094-26103.
157. Hoppen, C., et al., *Soluble and membrane-bound protein carrier mediate direct copper transport to the ethylene receptor family*. Sci. Rep., 2019. **9**(1): p. 10715.
158. Gellatly, S.L. and R.E.W. Hancock, *Pseudomonas aeruginosa : new insights into pathogenesis and host defenses*. Pathogens and Disease, 2013. **67**: p. 159-173.
159. Hauser, A.R., *Pseudomonas aeruginosa: So many virulence factors, so little time**. Crit. Care Med., 2011. **39**: p. 2193-2194.
160. Klockgether, J. and B. Tummler, *Recent advances in understanding Pseudomonas aeruginosa as a pathogen*. F1000Research, 2017. **6**: p. 1261.
161. Weiner-Lastinger, L.M., et al., *Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network, 2015-2017*. Infect. Control Hosp. Epidemiol., 2019: p. 1-18.
162. Morales, E., et al., *Hospital costs of nosocomial multi-drug resistant Pseudomonas aeruginosa acquisition*. BMC Health Serv. Res., 2012. **12**: p. 122.
163. Tacconelli, E. and N. Magrini, *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics*, in *Essential Medicines and Health Products*. 2017, World Health Organization: Geneva. p. 7.
164. Tacconelli, E., 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): p. 318-327.
165. Bender, J. and A. Flieger, *Lipases as Pathogenicity Factors of Bacterial Pathogens of Humans*, in *Handbook of Hydrocarbon and Lipid Microbiology*, K.N. Timmis, Editor. 2010, Springer. p. 3241-3258.
166. Istivan, T.S. and P.J. Coloe, *Phospholipase A in Gram-negative bacteria and its role in pathogenesis*. Microbiology, 2006. **152**(Pt 5): p. 1263-74.
167. Jaeger, K.-E. and T. Eggert, *Lipases for biotechnology*. Curr. Opin. Biotechnol., 2002. **13**(4): p. 390-397.
168. Richmond, G.S. and T.K. Smith, *Phospholipases A(1)*. Int. J. Mol. Sci., 2011. **12**(1): p. 588-612.
169. Dennis, E.A., *Introduction to Thematic Review Series: Phospholipases: Central Role in Lipid Signaling and Disease*. J. Lipid Res., 2015. **56**: p. 1245-1247.
170. Benamara, H., et al., *Characterization of membrane lipidome changes in Pseudomonas aeruginosa during biofilm growth on glass wool*. PLOS ONE, 2014. **9**(9): p. e108478.
171. Ryan, R.P. and J.M. Dow, *Communication with a growing family: Diffusible signal factor (DSF) signaling in bacteria*. Trends Microbiol., 2011. **19**: p. 145-152.
172. Jimenez, P.N., et al., *The multiple signaling systems regulating virulence in Pseudomonas aeruginosa*. Microbiol. Mol. Biol. Rev., 2012. **76**(1): p. 46-65.

173. Kovacic, F., *Novel phospholipases A of Pseudomonas aeruginosa: biochemical characterisation and cellular localisation*, in *Mathematisch-Naturwissenschaftlichen Fakultät*. 2010, Heinrich Heine University Düsseldorf.
174. Kovacic, F., et al., *Structural and functional characterisation of TesA - a novel lysophospholipase A from Pseudomonas aeruginosa*. PLOS ONE, 2013. **8**(7): p. e69125.
175. Salacha, R., et al., *The Pseudomonas aeruginosa patatin-like protein PlpD is the archetype of a novel Type V secretion system*. Environ. Microbiol., 2010. **12**(6): p. 1498-512.
176. Kovacic, F., et al., *A membrane-bound esterase PA2949 from Pseudomonas aeruginosa is expressed and purified from Escherichia coli*. FEBS Open Bio, 2016. **6**(5): p. 484-93.
177. Bleffert, F., et al., *Pseudomonas aeruginosa esterase PA2949, a bacterial homolog of the human membrane esterase ABHD6: expression, purification and crystallization*. Acta Crystallogr. F, 2019. **75**(4): p. 270-277.
178. Hutchings, M.J. and H. de Kroon, *Foraging in Plants: the Role of Morphological Plasticity in Resource Acquisition*. Adv. Ecol. Res., 1994. **25**(C): p. 159-238.
179. Hermans, C., et al., *How do plants respond to nutrient shortage by biomass allocation?* Trends Plant Sci., 2006. **11**(12): p. 610-7.
180. Shemesh, H., et al., *The Effects of Nutrient Dynamics on Root Patch Choice*. PLOS ONE, 2010. **5**(5).
181. Passioura, J.B., *'Soil conditions and plant growth'*. Plant, Cell and Environment, 2002. **25**(2): p. 311-318.
182. Cosme, M., et al., *Non-Mycorrhizal Plants: The Exceptions that Prove the Rule*. Trends Plant Sci., 2018. **23**(7): p. 577-587.
183. van der Heijden, M.G.A., et al., *Mycorrhizal ecology and evolution: the past, the present, and the future*. New Phytol., 2015. **205**: p. 1406-1423.
184. Schott, S., et al., *Cooperation through Competition—Dynamics and Microeconomics of a Minimal Nutrient Trade System in Arbuscular Mycorrhizal Symbiosis*. Front. Plant Sci., 2016. **7**: p. 1-13.
185. Dreyer, I., et al., *Nutrient exchange in arbuscular mycorrhizal symbiosis from a thermodynamic point of view*. New Phytol., 2019. **222**(2): p. 1043-1053.
186. Welch, R.M. and L. Shuman, *Micronutrient Nutrition of Plants*. Crit. Rev. Plant Sci., 2011. **14**(1): p. 49-82.
187. Briat, J.F. and S. Lobréaux, *Iron transport and storage in plants*. Trends Plant Sci., 1997. **2**(5): p. 166-166.
188. Colombo, C., et al., *Review on iron availability in soil: interaction of Fe minerals, plants, and microbes*. J. Soils Sed., 2013. **14**(3): p. 538-548.
189. Rajniak, J., et al., *Biosynthesis of redox-active metabolites in response to iron deficiency in plants*. Nat. Chem. Biol., 2018. **14**(5): p. 442-450.
190. Harada, E., et al., *Structural element responsible for the Fe(III)-phytosiderophore specific transport by HvYSL transporter in barley*. FEBS Lett., 2007. **581**(22): p. 4298-302.
191. Nozoye, T., et al., *Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants*. J. Biol. Chem., 2011. **286**(7): p. 5446-54.
192. Connorton, J.M., J. Balk, and J. Rodriguez-Celma, *Iron homeostasis in plants - a brief overview*. Metallomics, 2017. **9**(7): p. 813-823.
193. Vert, G., et al., *IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth*. Plant Cell, 2002. **14**(6): p. 1223-33.
194. Guerinot, M.L., *The ZIP family of metal transporters*. Biochim. Biophys. Acta, 2000. **1465**(1-2): p. 190-198.

195. Kerkeb, L., et al., *Iron-induced turnover of the Arabidopsis IRON-REGULATED TRANSPORTER1 metal transporter requires lysine residues*. *Plant Physiol.*, 2008. **146**(4): p. 1964-73.
196. Eide, D., et al., *A novel iron-regulated metal transporter from plants identified by functional expression in yeast*. *Proc. Natl. Acad. Sci. U. S. A.*, 1996. **93**(11): p. 5624-8.
197. Dubeaux, G., et al., *Metal Sensing by the IRT1 Transporter-Receptor Orchestrates Its Own Degradation and Plant Metal Nutrition*. *Mol. Cell*, 2018. **69**(6): p. 953-964 e5.
198. Shin, L.J., et al., *IRT1 degradation factor1, a ring E3 ubiquitin ligase, regulates the degradation of iron-regulated transporter1 in Arabidopsis*. *Plant Cell*, 2013. **25**(8): p. 3039-51.
199. Barberon, M., et al., *Monoubiquitin-dependent endocytosis of the iron-regulated transporter 1 (IRT1) transporter controls iron uptake in plants*. *Proc. Natl. Acad. Sci. U. S. A.*, 2011. **108**(32): p. E450-8.
200. Barberon, M., et al., *Polarization of IRON-REGULATED TRANSPORTER 1 (IRT1) to the plant-soil interface plays crucial role in metal homeostasis*. *Proc. Natl. Acad. Sci. U. S. A.*, 2014. **111**(22): p. 8293-8.
201. Vishwakarma, K., et al., *Abscisic Acid Signaling and Abiotic Stress Tolerance in Plants: A Review on Current Knowledge and Future Prospects*. *Front. Plant Sci.*, 2017. **8**: p. 161.
202. Ma, Y., et al., *Regulators of PP2C phosphatase activity function as abscisic acid sensors*. *Science*, 2009. **324**(5930): p. 1064-8.
203. Park, S.Y., et al., *Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins*. *Science*, 2009. **324**(5930): p. 1068-71.
204. Edel, K.H. and J. Kudla, *Integration of calcium and ABA signaling*. *Curr. Opin. Plant Biol.*, 2016. **33**: p. 83-91.
205. Rodriguez, L., et al., *C2-Domain Abscisic Acid-Related Proteins Mediate the Interaction of PYR/PYL/RCAR Abscisic Acid Receptors with the Plasma Membrane and Regulate Abscisic Acid Sensitivity in Arabidopsis*. *Plant Cell*, 2014. **26**: p. 4802-4820.
206. Stahelin, R.V. and W. Cho, *Roles of calcium ions in the membrane binding of C2 domains*. *Biochem. J.*, 2001. **359**(3): p. 679-685.
207. Corbalan-Garcia, S., et al., *The C2 domains of classical/conventional PKCs are specific PtdIns(4,5)P(2)-sensing domains*. *Biochem. Soc. Trans.*, 2007. **35**(Pt 5): p. 1046-8.
208. Guerrero-Valero, M., et al., *Structural and mechanistic insights into the association of PKC α -C2 domain to PtdIns(4,5)P2*. *Proc. Natl. Acad. Sci. U. S. A.*, 2009. **106**(16): p. 6603-7.
209. Diaz, M., et al., *Calcium-dependent oligomerization of CAR proteins at cell membrane modulates ABA signaling*. *Proc. Natl. Acad. Sci. U. S. A.*, 2016. **113**: p. E396-E405.
210. Corbalan-Garcia, S. and J.C. Gómez-Fernández, *Signaling through C2 domains: More than one lipid target*. *Biochim. Biophys. Acta*, 2014. **1838**(6): p. 1536-1547.
211. Guillen, J., et al., *Structural insights into the Ca²⁺ and PI(4,5)P2 binding modes of the C2 domains of rabphilin 3A and synaptotagmin 1*. *Proc. Natl. Acad. Sci. U. S. A.*, 2013. **110**: p. 20503-20508.
212. Knauer, T., et al., *A negative effector of blue light-induced and gravitropic bending in Arabidopsis*. *Plant Physiol.*, 2011. **156**(1): p. 439-47.
213. Michalski, C., et al., *Impact of EHB1 and AGD12 on Root and Hypocotyl Phototropism in Arabidopsis thaliana*. *J. Plant Growth Regul.*, 2017. **36**(3): p. 660-668.
214. Sommer, B., *Membrane Packing Problems: A short Review on computational Membrane Modeling Methods and Tools*. *Comput Struct Biotechnol J*, 2013. **5**: p. e201302014.
215. Javanainen, M. and H. Martinez-Seara, *Efficient preparation and analysis of membrane and membrane protein systems*. *Biochim. Biophys. Acta*, 2016. **1858**(10): p. 2468-2482.

216. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. J. Mol. Graphics, 1996. **14**(1): p. 33-38.
217. *Maestro*. 2019, Schrödinger, LLC: New York, NY.
218. Martínez, L., et al., *PACKMOL: a package for building initial configurations for molecular dynamics simulations*. J. Comput. Chem., 2009. **30**(13): p. 2157-64.
219. Martínez, J.M. and L. Martínez, *Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking*. J. Comput. Chem., 2003. **24**: p. 819-825.
220. Word, J.M., et al., *Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation*. J. Mol. Biol., 1999. **285**(4): p. 1735-47.
221. Nugent, T. and D.T. Jones, *Membrane protein orientation and refinement using a knowledge-based statistical potential*. BMC Bioinform., 2013. **14**: p. 276.
222. Ho, B. *pdbremix Git-Hub repository*. Available from: <https://github.com/boscoh/pdbremix>.
223. Koenig, B.W. and K. Gawrisch, *Specific volumes of unsaturated phosphatidylcholines in the liquid crystalline lamellar phase*. Biochim. Biophys. Acta, 2005. **1715**: p. 65-70.
224. Kutzner, C., et al., *Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail*. Biophys. J., 2011. **101**(4): p. 809-17.
225. Tyagi, C., et al., *Accelerated Molecular Dynamics Applied to the Peptaibol Folding Problem*. Int. J. Mol. Sci., 2019. **20**(17).
226. Stavrou, A., R.J. Evans, and R. Schmid, *Identification of a distinct desensitisation gate in the ATP-gated P2X2 receptor*. Biochem. Biophys. Res. Commun., 2019.
227. Chen, Y.F., et al., *Ethylene receptors function as components of high-molecular-mass protein complexes in Arabidopsis*. PLOS ONE, 2010. **5**(1): p. e8640.
228. Suenaga, Y., et al., *Structure and 1H NMR study of copper(I) complex with ethylene and tetramethylethylenediamine*. Polyhedron, 1997. **16**: p. 67-70.
229. Hirsch, J., et al., *Raman and extended X-ray absorption fine structure characterization of a sulfur-ligated Cu(I) ethylene complex: modeling the proposed ethylene binding site of Arabidopsis thaliana ETR1*. Inorg. Chem., 2001. **40**(10): p. 2439-41.
230. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices*. J. Mol. Biol., 1999. **292**(2): p. 195-202.
231. Dobson, L., I. Reményi, and G.E. Tusnády, *CCTOP: A Consensus Constrained TOPology prediction web server*. Nucleic Acids Res., 2015. **43**: p. W408-W412.
232. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score*. Nucleic Acids Res., 2005. **33**(7): p. 2302-9.
233. Barth, P., B. Wallner, and D. Baker, *Prediction of membrane protein structures with complex topologies using limited constraints*. Proc. Natl. Acad. Sci. U. S. A., 2009. **106**: p. 1409-1414.
234. Dutta, A. and I. Bahar, *Metal-binding sites are designed to achieve optimal mechanical and signaling properties*. Structure, 2010. **18**: p. 1140-1148.
235. Dominguez, C., R. Boelens, and A.M.J.J. Bonvin, *HADDOCK: A protein-protein docking approach based on biochemical or biophysical information*. J. Am. Chem. Soc., 2003. **125**: p. 1731-1737.
236. Im, W., M.S. Lee, and C.L. Brooks, *Generalized Born Model with a Simple Smoothing Function*. J. Comput. Chem., 2003. **24**: p. 1691-1702.
237. Parsons, J.B. and C.O. Rock, *Bacterial lipids: metabolism and membrane homeostasis*. Prog. Lipid Res., 2013. **52**(3): p. 249-276.
238. Sinensky, M., *Homeoviscous adaptation--a homeostatic process that regulates the viscosity of membrane lipids in Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A., 1974. **71**(2): p. 522-525.

239. Cossins, A.R. and Society for Experimental Biology (Great Britain), *Temperature adaptation of biological membranes : proceedings of the meeting held in Cambridge under the auspices of the Society for Experimental Biology in conjunction with its US/Canadian counterparts*. Portland Press proceedings. 1994, London ; Chapel Hill, NC: Portland Press. xiv, 227 p.
240. Sohlenkamp, C., *Membrane Homeostasis in Bacteria upon pH Challenge*, in *Biogenesis of Fatty Acids, Lipids and Membranes*, O. Geiger, Editor. 2017. p. 1-13.
241. Fozo, E.M. and E.A. Rucks, *The making and taking of lipids: the role of bacterial lipid synthesis and the harnessing of host lipids in bacterial pathogenesis*, in *Adv. Microb. Physiol.*, K.P. Robert, Editor. 2016, Academic Press. p. 51-155.
242. Lomize, M.A., et al., *OPM database and PPM web server: resources for positioning of proteins in membranes*. *Nucleic Acids Res.*, 2012. **40**(Database issue): p. D370-6.
243. Grossfield, A. *WHAM: the weighted histogram analysis method*. 21/11/2016]; Available from: <http://membrane.urmc.rochester.edu/content/wham>.
244. Doudou, S., N.A. Burton, and R.H. Henchman, *Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force*. *J. Chem. Theory Comput.*, 2009. **5**: p. 909-918.
245. Mulnaes, D., et al., *TopModel: Template-based protein structure prediction at low sequence identity using top-down consensus and deep neural networks*. *J. Chem. Theory Comput.*, 2020.
246. Furt, F., F. Simon-Plas, and S. Mongrand, *Lipids of the Plant Plasma Membrane*, in *The Plant Plasma Membrane*, A.S. Murphy, W. Peer, and B. Schulz, Editors. 2011, Springer. p. 3-30.
247. Simon, M.L., et al., *A PtdIns(4)P-driven electrostatic field controls cell membrane identity and signalling in plants*. *Nat. Plants*, 2016. **2**: p. 16089.