Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CUI THE HAMBURG CENTRE FOR ULTRAFAST IMAGING

Integrative analysis of the dynamics of rhomboid protease GlpG from *Escherichia coli*

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften (*Dr. rer. nat.*) an der Fakultät für Mathematik, Informatik und Naturwissenschaften der Universität Hamburg, Fachbereich Chemie

vorgelegt von
Yasser Almeida-Hernández

Hamburg, September 2019

Die vorliegende Arbeit wurde im Zeitraum von Dezember 2014 bis September 2019 in der Arbeitsgruppe von Prof. Dr. H. Tidow am Institut für Biochemie und Molekularbiologie des Fachbereichs Chemie der Universität Hamburg durchgeführt.

Erstgutachter: Prof. Dr. Henning Tidow
Zweitgutachter: Prof. Dr. Johannes Kirchmair

Datum der Disputation: 08.11.2019

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.

Marie Skłodowska-Curie

Table of Contents

List of Figures

List of Tables

List of Publications

Publications associated with this work

- Yasser Almeida-Hernandez and Henning Tidow. (2019) ***Soluble Regions of GlpG Influence Protein-Lipid Interactions and Lipid Distribution***. Journal of Physical Chemistry B. 123, 37, 7852-7858. DOI: 10.1021/acs.jpcb.9b06943
- Yasser Almeida-Hernandez, Johann Klare, Heinz-Jürgen Steinhoff, and Henning Tidow. ***Dynamics of rhomboid protease GlpG in solution depicted by SAXS and DEER spectroscopy*** (manuscript in preparation)

Others

- Ivana G. Molina, Inokentijs Josts, Yasser Almeida-Hernandez, Sebastian Esperante, Mariano Salgueiro, Maria M. Garcia Alai, Gonzalo de Prat-Gay and Henning Tidow. (2018) ***Structure and stability of the Human respiratory syncytial virus M2–1 RNA-binding core domain reveals a compact and cooperative folding unit***. Acta Cryst**.** F74, 23-30. DOI: 10.1107/S2053230X17017381
- Katharina Veith, Maria Martinez Molledo, Yasser Almeida-Hernandez, Inokentijs Josts, Julius Nitsche, Christian Löw and Henning Tidow. (2017) ***Lipid-like Peptides can Stabilize Integral Membrane Proteins for Biophysical and Structural Studies.*** ChemBioChem 18 (17), 1735 –1742. DOI: 10.1002/cbic.201700235
- Inokentijs Josts, Yasser Almeida-Hernandez, Antonina Andreeva and Henning Tidow. (2016) ***Crystal Structure of a Group I Energy Coupling Factor Vitamin Transporter S Component in Complex with Its Cognate Substrate.*** Cell Chemical Biology. 23 (7), 827–836. DOI: 10.1016/j.chembiol.2016.06.008

Abbreviations

α	Regularization parameter
Å	Angstrom
Amp	Ampicillin
B_0	External static field
β	Beta
C	Celsius
CDL	Cardiolipin
CG	Coarse-grained
CMC	Critical micellar concentration
CW	Continuous-wave
CytoD	Cytoplasmatic domain
Da	Dalton
DDM	n-dodecyl-β-D-maltopyranoside
DEER	Double Electron-Electron Resonance
D_{max}	Maximum distance
DNA	Deoxyribonucleic acid
DPPE	di-C16:0-C18:0 PE
DPPG	di-C16:0-C18:0 PG
DOPE	di-C16:1-C18:1 PE
DOPG	di-C16:1-C18:1 PG
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
Δ	Modulation depth
EPR	Electron paramagnetic resonance
eV	Electronvolt
FC12	Fos-choline-12
fs	Femtosecond
GHz	Gigahertz
g	Gram
g	Gravitational force
h	Hour
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl -β-D-thiogalactopyranoside
K	Kelvin
k	Kilo
λ	Wavelength
l	Liter
LB	Lennox-Broth
Ln	Linker region
min	Minute
Ni-NTA	Nickel-Nitrilotriacetic acid
MD	Molecular dynamics
MHz	Megahertz
ml	Militer
mol	Mole

ms	Millisecond
M	Molar
MBP	Maltose binding protein
MTSL	(1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate
MW	Molecular weight
µ	Micro
N_A	Aggregation number
nm	Nanometer
ns	Nanosecond
ω_A, ω_B	Spin populations A and B
OD	Optical density
ps	Picosecond
PDB	Protein Data Bank
PELDOR	Pulsed Electron-Electron Double Resonance
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
POPE	C16:0/18:1 PE
POPG	C16:0/18:1 PG
R1	MTSL spin-label
r_{AB}	Interspin distance
RDF	Radial distribution function
R_g	Radius of gyration
rpm	Rounds per minute
s	Second
SAXS	Small-angle X-ray scattering
SDSL	Site-directed spin-labeling
SEC	Size exclusion chromatography
T	Tesla
TatA	Twin-arginine translocase component A
TMD	Transmembrane domain
Trx	Thioredoxin
UV	Ultraviolet
V	Volt
v/v	Volume per volume
wt	Wild type
W	Watt
w/v	Weight per volume
υ_{pump}	Pump frequency
υ_{obs}	Observer frequency

Abstract

Rhomboid proteases are membrane proteins wide present in all living kingdoms, which perform proteolytic reactions in the lipidic environment of the cell membrane. Rhomboid protease GlpG from *Escherichia coli* is the most studied molecule of this group, and a prototypical example of the intramembrane proteases. The protein is composed by a transmembrane domain (TMD) which contains the active site, a soluble and N-terminal cytoplasmic domain (CytoD), with unknown function at the moment, and a linker region (Ln) that connect both domains.

Most of the structural and functional knowledge of this protein have been obtained from its transmembrane domain since it can perform hydrolytic reactions without the rest of the protein. In this work, I addressed the study of the full-length protein *in vitro* and *in silico*, in order to gain a better understanding of the structure and dynamics of this molecule.

In the first part, I identified the detergent Fos-choline-12 (FC12) as a proper detergent to purify the molecule with high purity and stability, and suitable for biophysical studies in solution. In this direction, I carried out size-exclusion chromatography coupled with small-angle X-ray scattering (SEC-SAXS) experiments to depict the low-resolution shape of the molecule stabilized in FC12. Subsequently, I performed double electron-electron resonance (DEER) spectroscopy experiments, to support the SAXS data, showing that the protein exists in compacted and extended conformations, showing the high flexibility of GlpG in solution.

In the second part, I used coarse-grained molecular dynamics simulations (CGMD) to study full-length GlpG inserted in a native-like model of the *E. coli* membrane, and the influence of the soluble regions of the molecule on the protein/lipid interactions. I identified differences in the distribution and clustering of phosphoglycerol(PG)-based lipids around GlpG, depending on the presence or absence of the CytoD and Ln fragments. These data suggest a possible role of the cytoplasmic extensions of GlpG in the regulation of the lipid environment around GlpG, which may influence the activity of GlpG *in vivo*.

Taking together with these data and recent reports, I proposed a hypothetical mechanism for this protein, which takes into account the entire protein and its lipidic environment.

Zusammenfassung

Rhomboid-Proteasen sind in allen lebenden Reichen weit verbreitete Membranproteine, die im lipidischen Milieu der Zellmembran proteolytische Reaktionen durchführen. Die Rhomboid-Protease GlpG aus *Escherichia coli* ist das am meisten untersuchte Molekül dieser Gruppe und ein prototypisches Beispiel für die Intramembran-Proteasen. Das Protein besteht aus einer Transmembrandomäne (TMD), die das aktive Zentrum enthält, einer löslichen N-terminalen cytoplasmatischen Domäne (CytoD) mit derzeit unbekannter Funktion und einer Linkerregion (Ln), die beide Domänen verbindet.

Das meiste strukturelle und funktionelle Wissen über dieses Protein wurde aus seiner Transmembrandomäne gewonnen, da es ohne den Rest des Proteins hydrolytische Reaktionen durchführen kann. In dieser Arbeit habe ich mich mit der Untersuchung des vollständigen Proteins *in vitro* und *in silico* befasst, um ein besseres Verständnis der Struktur und Dynamik dieses Moleküls zu erlangen.

Im ersten Teil habe ich das Detergens Fos-Cholin-12 (FC12) als geeignetes Detergens zur Reinigung des Moleküls mit hoher Reinheit und Stabilität identifiziert, das für biophysikalische Untersuchungen in Lösung geeignet ist. In dieser Richtung führte ich Size-Exclusion Chromatography in Kombination mit Small-Angle X-ray Scattering (SEC-SAXS) Experimenten (SEC-SAXS) durch, um die Form des in FC12 stabilisierten Moleküls mit niedriger Auflösung darzustellen. Anschließend führte ich Double Electron-Electron Resonance (DEER)-Spektroskopie-Experimente durch, um die SAXS-Daten zu stützen. Dabei zeigte sich, dass das Protein in verdichteten und erweiterten Konformationen vorliegt, was die hohe Flexibilität von GlpG in Lösung zeigt.

Im zweiten Teil verwendete ich Coarse-Grained Molecular Dynamics (CGMD)-Simmulationen, um das in ein natives Modell der *E. coli*-Membran eingefügte GlpG in voller Länge und den Einfluss der löslichen Regionen des Moleküls auf das Protein / Lipid wechselwirkungen zu untersuchen. Ich identifizierte Unterschiede in der Verteilung und Clusterbildung von Phosphoglycol (PG)-basierten Lipiden um GlpG, abhängig von der Anwesenheit oder Abwesenheit der CytoD- und Ln-Fragmente. Diese Daten legen eine mögliche Rolle der cytoplasmatischen Verlängerungen von GlpG bei der Regulation der Lipidumgebung um GlpG nahe, die die Aktivität von GlpG *in vivo* beeinflussen kann.

Zusammen mit diesen Daten und letzte Berichten schlug ich einen hypothetischen Mechanismus für dieses Protein vor, der die Dynamik des gesamten Proteins und seiner Lipidumgebung berücksichtigt.

1 Introduction

1.1 Introduction to membrane proteins

Lipid membranes are a fundamental element in cell biology. They are the primary barrier separating the interior of cells and organelles from the environment, and constitute a platform where key processes take place, with paramount importance for cell physiology such as transport, signaling, sensing, and regulation, maintaining cell homeostasis (1), all of them mediated by proteins (Figure 1-1). Approximately 20 – 30% of sequenced genomes encode for membrane proteins (2, 3) and they are the target of around 50% of FDA approved drugs (4, 5).

Depending on how the molecule interacts with the lipidic bilayer, membrane proteins can be classified as peripheral or integral. Peripheral membrane proteins (PMP) lack a well-organized hydrophobic domain to intercalate in the membrane and are usually recruited to the membrane via electrostatic interactions, post-translational modifications or a combination thereof. These proteins can modulate membrane shape, lipid composition, and membrane fluidity. They play an important regulatory role in a variety of cellular processes including cytoskeletal interactions, vesicular trafficking, and signal transduction (6).

Integral membrane proteins (IMP) are defined as proteins permanently inserted into the biological membrane after mRNA translation and peptide chain folding. These proteins contain a region enriched with hydrophobic amino acids that are embedded in the lipidic bilayer and interact with the acyl chains of the lipids, while polar and charged residues are located in regions exposed to the solvent (1, 6).

IMPs can be categorized into three classes: single-pass, multi-pass, and β-barrel transmembrane proteins. The first two classes, also known as helical membrane proteins, carry out most of the functions in the inner membrane of Gram-negative bacteria and eukaryotes (1), while the β-barrel transmembrane proteins are largely found in the outer membrane in bacteria, as well as mitochondria and chloroplasts (7–9). Multi-pass transmembrane proteins represent the largest class of membrane-bound macromolecules and the number of α-helices span from 2 – 18 in *E. coli* and 2 – +30 in human (1, 10).

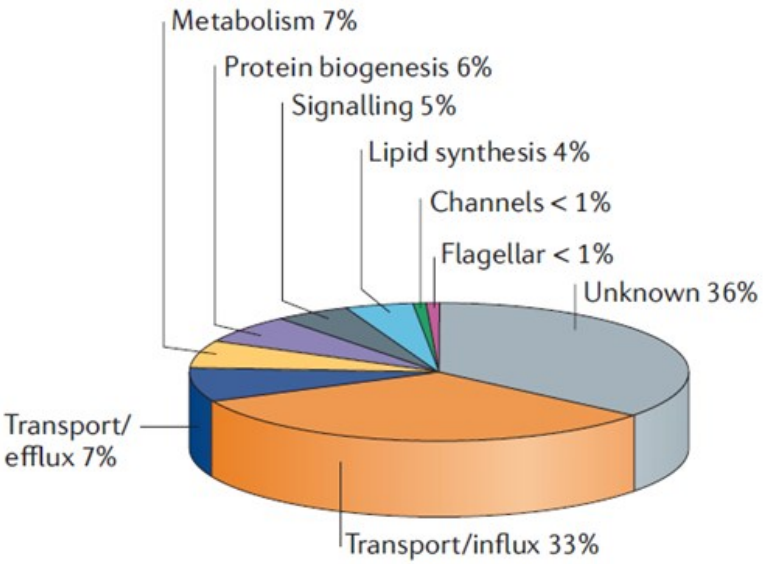

Figure 1-1: Functional distribution of transmembrane proteins in the proteome of *Escherichia coli* (taken and modified from ref. 5).

The folding of these proteins has been a conundrum for many years, trying to investigate how transmembrane proteins insert and adopt their three-dimensional structure in the lipid bilayer. The hydrophobic effect reflects a tendency of non-polar molecules to avoid contact with water, reducing the entropy of the system. This spontaneous interaction is a source of energetic stability because the unfolding or denaturation of a protein would leave the non-polar residues exposed to the unfavorable aqueous environment (11). This effect combined with ionic interactions and hydrogen bonds between residues located in different α-helices of the same protein seems to contribute to a similar extent. Ionic interactions have a dual effect because sometimes they can be relevant not only for achieving the folded state but also may be involved in specific changes related with the activity of the protein (12, 13).

1.2 Protein – lipid interactions

Due to its localization in the lipidic bilayer, membrane proteins are subjected to different thermodynamical, diffusional and kinetical restrictions, compared with soluble proteins. In this regard, lipids play a central role influencing the folding, stability and function of IMPs (14).

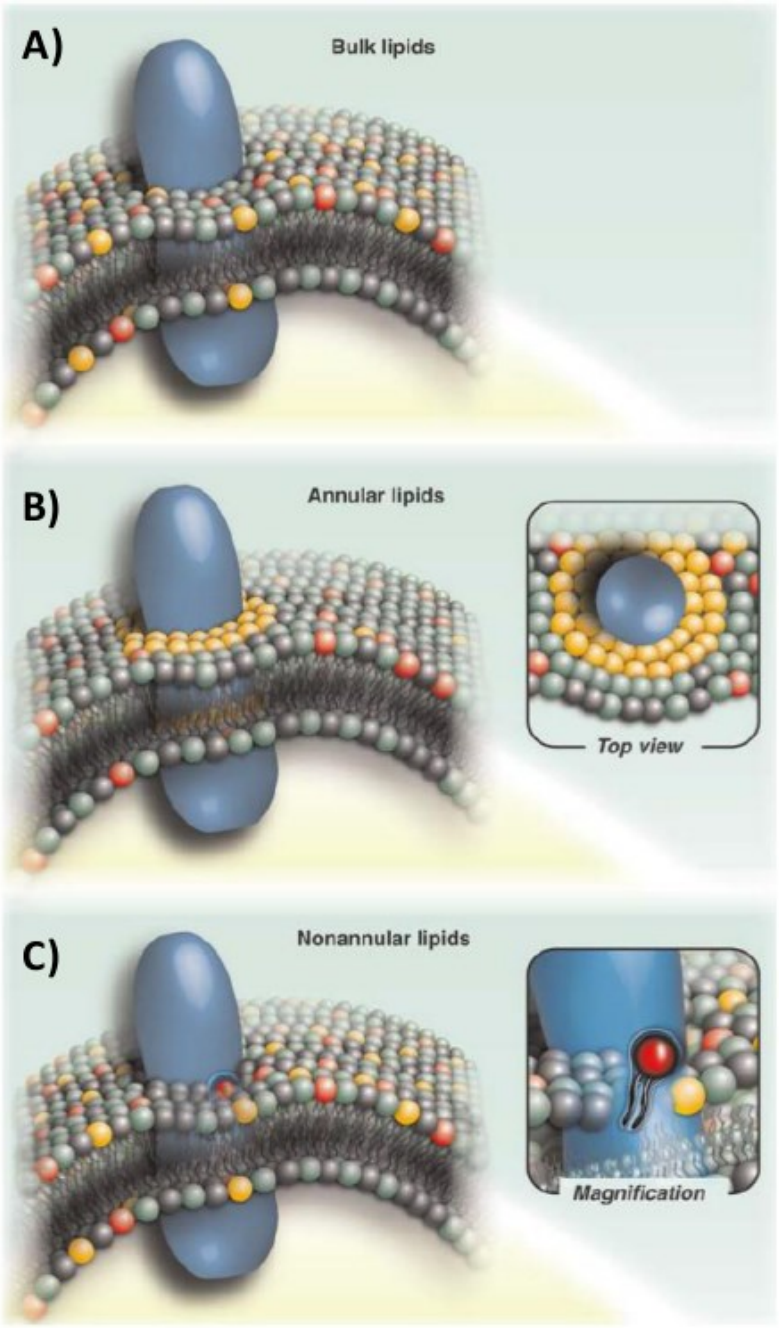

Figure 1-2: Intramembrane protein–lipid interactions within a cell membrane. A) Bulk lipids, B) annular lipids, C) non-annular lipids/lipid ligands (taken and modified from ref. 15).

Depending on how the lipids interact with the protein, these can be classified as bulk, annular or non-annular/ligand (14, 15). The "bulk" lipids are those within the membrane that diffuse rapidly in the bilayer plane and show a low residence time at the protein-lipid interface following random collisions (Figure 1-2.A). Typical lateral diffusion coefficients (D_L) for bulk lipids in bilayers are in the range of $D_L = 7 – 15\ \mu m^2/s$ (16, 17).

Annular lipids form a layer surrounding the protein, making favorable but transient contacts (Figure 1-2.B). Within this lipid shell, the diffusion rates, and hence the exchange rates with the bulk lipids, are found to be significantly reduced (18, 19). The first visualization of annular lipids was obtained with the structures of bacteriorhodopsin (bR) from *Halobacterium salinarum* (20, 21). 18 lipid chains were identified from the structure obtained by lipid cubic phase crystallization (21). In another study, employing mass spectrometry and quantitative lipidomics, annular lipids associated with the ABC transporter TmrAB were identified, with a clear preference to phosphatidylglycerol (PG). These lipids are also functionally important, with a 60-70% decrease in the ATPase activity after a complete delipidation (22).

Non-annular/ligand lipids have been identified by X-ray crystallography. These molecules come from native membranes and remain associated with the proteins despite protein delipidation. They are usually buried in the protein structure and clefts on the surface, and interact with the protein more specifically (Figure 1-2.C) (14, 15, 18). Hanson and coworkers solved the structure of the human β2-adrenergic receptor in complex with 2 cholesterol molecules, relevant for the packing and the stability of the molecule. They found a consensus motif, present in 44% of human class A receptors, responsible for specific cholesterol binding (23).

Another function of non-annular lipids is acting as "molecular glue", mediating the association of protein oligomers in the membrane. For example, 13 lipid molecules were resolved in the crystal structure of cytochrome oxidase, two cardiolipins (CL), one phosphatidylcholine (PC), three phosphatidylethanolamines (PE), four phosphatidylglycerols (PG) and three triglycerides. Four of the non-annular lipid molecules in cytochrome oxidase are important in homodimer formation (24).

1.3 Expression of membrane proteins

The study of the structure and dynamics of IMPs faces many challenges, due to the physical-chemical features of these molecules. Furthermore, the majority of IMPs with biomedical interest are naturally expressed at low abundance. Hence, the expression of sufficient quantities of the target protein becomes the first bottleneck to tackle for further structural studies.

In this regard, the first step is the selection of the host organism for recombinant expression. In the present days, many possibilities are available and several hosts can be used, from *E. coli*, yeast, insect or mammalian cells. Also, cell-free approaches are becoming more popular since

they may eliminate several obstacles known from conventional cell-based MP expression systems like problems with cell physiology, expression regulation and cell culture (25).

Escherichia coli is the most popular organism for recombinant expression because of the low costs, easy to handle, and the well-known molecular biology. Many strains are available in order to control the protein expression to maximize the quantity and the quality of the target protein. C41 and C43 strains are well suited for expressing membrane proteins in *E. coli* (26). These strains have an increased membrane production that allows increasing protein yields, and they constitute the base for obtaining new strains.

1.3.1 Solubilization and stabilization of IMPs

The next bottleneck for the study of IMPs is the isolation of sufficient amounts of the proteins with the required quality. This step includes the solubilization and stabilization of the target molecule, with a membrane-mimicking additive, that shields the hydrophobic regions of proteins or another component that keep the protein in its native or semi-native environment, maintaining the folding and the activity of the molecule in the selected buffer. Many additives have been developed for this task, and the selection of the proper one is key for further analysis.

Detergents

Detergents are surfactants (surface acting reagents) that decrease the interfacial tension between two immiscible liquids. The overall molecular structure of detergents consists of a hydrophilic polar head group and a hydrophobic non-polar tail group that renders them amphiphilic. The polar head group of detergents can be ionic, non-ionic or zwitterionic and usually has a strong attraction for aqueous solvent molecules whereas the detergent non-polar tail is generally an alkyl chain repelled from the aqueous solvent (27).

Detergent molecules persist as monomers in solution up to a particular concentration. As the detergent concentration increases, detergent molecules assemble into complex structures called micelles. The hydrophobic tails of the detergent molecules pack together, forming the core of the micelle and reducing their interaction with the water molecules. In contrast, the polar head groups orient themselves outwards from the micelle core, enabling interaction with the aqueous solvent. The minimal detergent concentration required for the formation of micelles in a defined concentration and temperature range is called the *critical micelle concentration* (CMC) and the number of detergent monomers required to form a micelle is called the *aggregation number* (N_A) (28, 29) (Table 1-1). The formation of micelles is the basis for membrane protein

solubilization. The hydrophobic parts of the membrane proteins are covered by the alkyl chains of the detergents, while polar groups are exposed to the solvent (27, 30).

During the solubilization and further purification steps, enough detergent needs to be available in the solution to accommodate all proteins in micelles, which typically occurs above the CMC; moreover, the membrane lipids form mixed micelles with the detergent used, changing the properties of the micelles. Some tightly packed membrane bilayers can be resistant to the detergent used, resulting in extracted proteins that are still associated with lipid molecules (31). In some cases, this can be beneficial when membrane lipids are relevant to maintain the structure and activity of the protein. In any case, an optimal detergent/protein ratio and detergent/lipid ratio is required for efficient solubilization of such membranes and for complete protein extraction (27, 30).

Table 1-1: Properties of common detergents used for membrane protein solubilization (31, 32)

Detergent (abbreviation)	MW (Da)	Micelle size (kDa)	CMC (mM/%)	N_A
n-dodecyl-β-D-maltopyranoside (DDM)	511	65–70	0.17/0.0087	80–150
n-decyl-β -D-maltopyranoside (DM)	483	40	1.8/0.087	69
n-nonyl-β -D-glucopyranoside (NG)	306	85	6.5/0.20	133
n-octyl-β -D-glucopyranoside (OG)	292	25	20/0.53	30–100
n-undecyl-β-D-Maltopyranoside (UDM)	496.6	50	0.59/0.029	71
n-dodecyl-N,N-dimethylamine-N-oxide (LDAO)	229.4	21.5	1-2/0.023	76
octaethylene-glycol-monododecylether (C12E8)	538.8	66	0.1/0.0048	90-120
n-dodecylphosphocholine (FC12)	351	38	1.5/0.047	54-108

Amphipols

Amphipols (APols) are short and flexible amphipathic polymers, designed to tightly bind to the transmembrane domain of membrane proteins by multiple hydrophobic contact points. Apols

remains associated with proteins, even at high dilutions and their dissociation rates are extremely slow. This would make them radically different from IMP/detergent complexes, in which the protein-bound detergent molecules are in rapid equilibrium with free monomers and micelles and dissociate upon dilution below the CMC (32, 33). APols are synthesized comprising a polyacrylic acid backbone onto which octylamine and isopropylamine side chains have been randomly grafted. A8-35 is the most popular amphipol, which have a molecular mass of ~4.3 kDa, composed of ~70 acrylate monomers. 35% of the carboxylic acid groups are ungrafted, 25% are derivatized with octyl chains, and the last 40% with isopropyl groups (32–34).

A major difference between APol-trapped and a detergent-solubilized membrane protein is their stability, which is usually much higher in Apols. For example, the incorporation of the G protein-coupled receptor (GPCR) BLT1 in A8-35 increases the thermal stability by ~11°C, when compared with the detergent Fos-choline-16 (35). However, depending on the protein, APols may affect the activity of the membrane protein they bind. For example, acetylcholine receptor from *Torpedo sp.* electric organ is not affected when it is trapped in A8-35 (36), while the sarcoplasmic calcium pump SERCA1a is reversibly inhibited after trapping (37). As a general rule, a positive correlation has been observed between the stabilization and the degree of functional inhibition (33).

Nanodiscs and lipodiscs

Nanodiscs are discoidal lipid bilayers of 8–16 nm in diameter, which are stabilized and rendered soluble in aqueous solutions by encircling amphipathic scaffolds, which can be either helical protein belts, termed membrane scaffold proteins, or organic polymers. The size of nanodiscs is determined by the length of the scaffold molecule and the stoichiometry of the lipids used in the self-assembly process. The resultant discoidal bilayers can be made homogeneous and monodisperse and can be obtained with high yield.

The first nanodisc was developed based on the engineering of apolipoprotein A-I (Apo A-I), a human protein that stabilized a transient form of high-density lipoproteins (HDL) particles in atherosclerosis. Modified variants of Apo A-I allowed its expression in *E. coli* and the modification of reconstitution procedures for the production of uniformly sized nanoparticles. The result of this genetic engineering exercise was a set of “membrane scaffold proteins” (MSPs) that were capable of self-assembly into discoidal phospholipid bilayers wrapped with an amphipathic helical belt surrounding the alkyl chains on the phospholipids (38).

Another approach developed, based on scaffold proteins, exploits the self-assembly of lipids and saposin into a saposin-lipoprotein nanoparticle (Salipro). The saposin protein family comprises four (saposin A–D) small proteins (~10 kDa) that derive from precursor proteins involved as cofactors in the catabolism of sphingolipids (39). Saposin A can be expressed and purified from *Escherichia coli*. Its general versatility lies in its ability to adapt to the size of the membrane protein, adjusting to transmembrane regions of varying size (40).

The previous approaches reconstitute the membrane protein in a lipid mixture that can be different from the native one. Several efforts have been focused to develop methods for the solubilization of membrane protein preserving its native lipid environment. This has led to the development of the SMA- (styrene/maleic acid) and DIBMA (diisobutylene/maleic acid)-based lipodisc. These polymers intercalate the hydrophobic groups between the acyl chains of the bilayer, whereas the hydrophilic maleic acid groups face the solvent. The encapsulated bilayer retains many of the physical properties of the parent membrane, including the lipid mixture, structural organization and phase behavior (41, 42).

1.4 Intramembrane proteases

Proteases are proteins that perform hydrolytic cleavage of peptide bonds of other proteins. Depending on its physiological role, these proteins can be classified in two broad groups: A) degradative functions, where other proteins are broken into smaller peptides or single amino acids for nutritional purposes, or B) regulatory functions, where other proteins are cleaved in specific sites to activate or repress cellular processes. The very first protease (pepsin) was discovered in 1836 (43) and during a long time, the knowledge about the proteins was restricted to soluble proteases. Due to the unavailability of water in the cellular membrane, it was thought that those hydrolytic processes were not possible in the hydrophobic environment of the membrane. However, that idea was debunked in 1997 with the discovery of site-2 protease (S2P) as the first intramembrane protease (44, 45) (also known as intramembrane-cleaving proteases, I-CliPs). On the basis of catalytic mechanisms, the I-CLiPs can be divided into four major families: rhomboid serine proteases, Rce1-type glutamyl proteases, zinc-bound site-2 proteases (S2P), and aspartyl proteases (46–48).

In analogy to soluble proteases, intramembrane proteolysis requires substrate recognition, which is followed by substrate processing. Substrate processing requires exposure of the scissile bond to the catalytic residues, followed by the formation of a tetrahedral intermediate

structure leading to hydrolysis. Finally, the product is released. Compared to soluble proteases, intramembrane proteases tend to be rather slow enzymes (49–51). Although intramembrane proteases mainly fall into mechanistic classes that mimic the classical soluble proteases (metallo-, aspartyl-, and serine-), they are phylogenetically unrelated to the classical proteases, meaning that they have evolved convergently. They are also unrelated to each other except that they all share the common characteristics of having multiple transmembrane domains (TMDs), membrane-embedded active sites and are cleaving TMD substrates (52).

1.4.1 Rhomboid proteases family

Rhomboids are currently the most studied intramembrane protease family and it is the most extended and diverse in all living kingdoms (53). The first rhomboid protease was initially identified in *Drosophila melanogaster*, as a mutation that disrupts development. Since some genes are named after the phenotypical features of mutants, in the case rhomboid gene, it was named after the rhomboid-shaped head of the *D. melanogaster* mutant embryo. Subsequently, it was shown to be involved in generating the active ligand for EGF receptor signaling in *D. melanogaster* (54, 55). Finally, a combination of genetic and biochemical approaches identified this protein as the first in a new family of intramembrane serine proteases (56, 57).

Rhomboids also participate in diverse processes as well as in signaling. The rhomboid AarA of the bacterium *Providencia stuartii* cleaves and activates the type I membrane protein TatA, a central component of the twin-arginine translocase, which exports folded proteins across the inner membrane. One of these translocated proteins may be essential for quorum sensing, a mechanism for signaling cell density (58). In *Plasmodium falciparum* and the protozoan *Toxoplasma gondii* rhomboids regulate the shedding of parasite adhesins during the invasion of host cells (59). In mammals, four isoforms of rhomboid, named RHBDL1-4, were identified (60). Their putative substrates include thrombomodulin, ephrinB3, EGF, and EGFR, suggesting a conserved function in signaling pathways between fruit flies and mammals (61).

Sequence analysis has uncovered a widespread group of degenerate rhomboid-like proteins, termed iRhoms for ‘inactive rhomboids’, which usually lack one or more of the catalytic residues and have a GPX sequence in place of the typical GXS motif containing the catalytic serine. The iRhoms perhaps have chaperone or regulatory functions, as the peptide-binding capabilities of proteases may be useful even in the absence of catalytic activity (62, 63).

1.4.2 Rhomboid protease GlpG from *E. coli*

GlpG, the rhomboid homolog in *Escherichia coli*, represents a prototype for structural and mechanistic investigation of intramembrane proteases. The protein consists of a transmembrane domain (TMD) (Figure 1-3A), an N-terminal soluble domain (CytoD) (Figure 1-3B), both connected by a linker region (Ln). Several crystallographic structures of the TMD have been solved in different conformations and in complex with inhibitors that have been key in the elucidation of the catalytic mechanism (64, 65). On the other hand, the structure of CytoD has been solved by NMR and X-ray diffraction (66). Although plenty of structural information about GlpG exists, this is almost exclusively restricted to the TMD, since it contains the catalytic site, and many gaps remain still unsolved regarding the structure and dynamics of the full-length protein.

The crystal structures of the *E. coli* protease TMD and a related rhomboid from *Haemophilus influenzae* have been studied by several groups. With the exception of a surface loop (L5) and one of the TM helices (S5), the independently obtained structures, in detergent and lipid bicelles, are all very similar to each other (64, 65, 67–71). The catalytic TMD of GlpG is composed of six membrane-spanning segments (TM1–TM6), which harbor a number of highly conserved sequence motifs that are characteristic of the family (72). Crystallographic analyses revealed that the HxxxN motif in TM2 (H150 and N154 in GlpG), the GxSG motif near the N-terminus of TM4 (S201), and the (A/G)H motif in TM6 (H254) are all essential elements of the active site of the enzyme. S201 and H254 are hydrogen-bonded to each other and form a rudimentary catalytic dyad (Figure 1-3A). This is distinctive of GlpG and many other rhomboid proteases which, unlike soluble serine proteases, contain a catalytic S/H dyad, instead of the classical S/H/R triad.

As stated before, the structure of the isolated CytoD domain has been determined, but details about its function are now coming into light. In a recent and very elegant work, it was shown that CytoD increases in the diffusion rate of GlpG in the membrane, and influence the position of the protein embedded in the bilayer (73). The NMR structure of CytoD domain shows a compact globular architecture (66), while the crystal structure reveals a domain-swapped dimer (74).

This structure has been claimed to be physiologically relevant as it is dimeric not only in solution but also when in contact with the membrane domain, suggesting an allosteric regulatory role (74). The influence of CytoD on catalysis is controversial, and the claims depend on the substrate and the detergent used for the kinetics assays. Some reports showed that full-length protein cleaves faster that the TMD (56, 60, 64, 66), while others claim that the absence of the CytoD domain does not affect the activity of GlpG (75, 76).

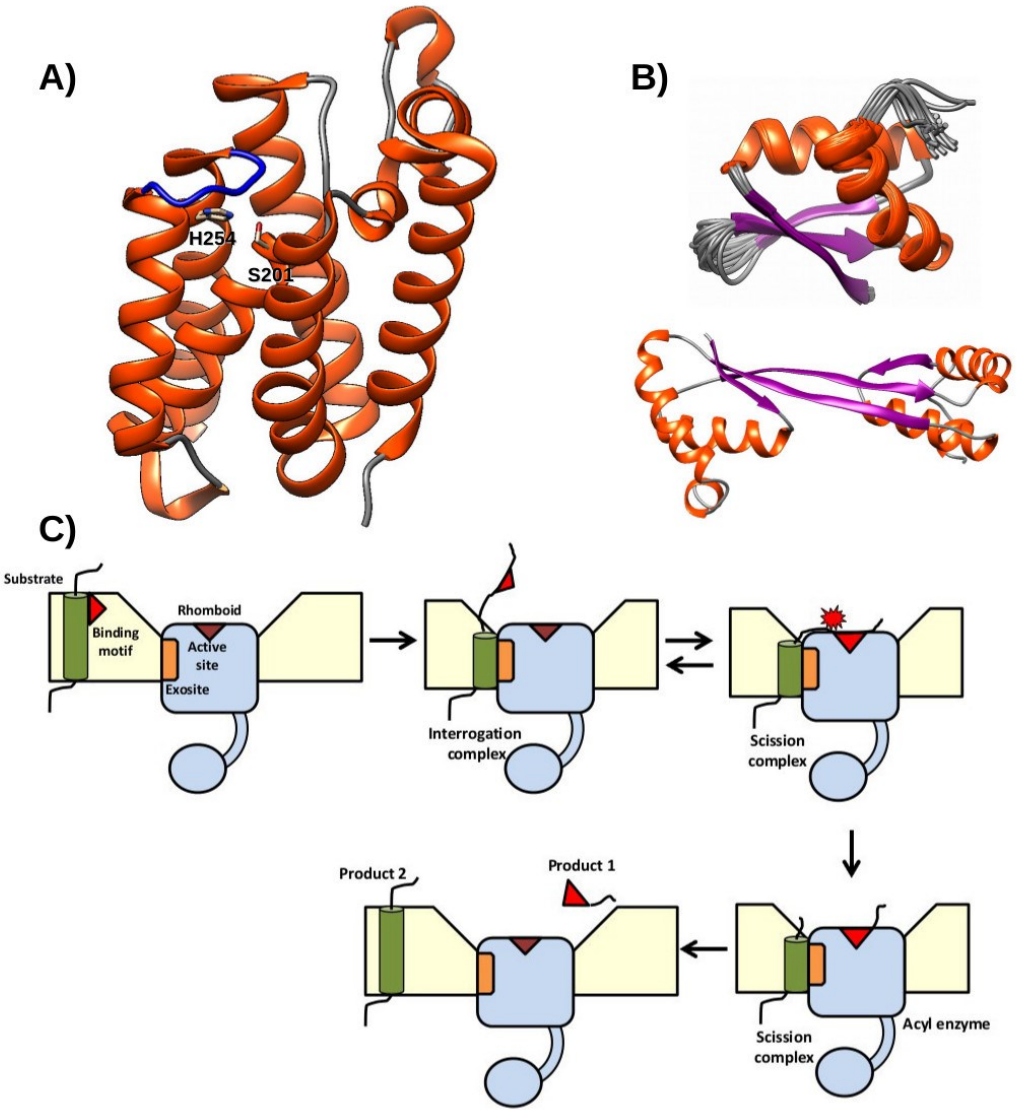

Figure 1-3: Structure and proposed mechanism of GlpG. A) Transmembrane domain of GlpG (PDB: 2XOV). Blue, L5 loop. B) CytoD domain of GlpG. Top, globular structure solved by NMR (PDB: 2LEP); bottom, domain-swapped crystallographic structure (PDB: 4HDD). C) "Interrogation-scission" mechanism of GlpG.

The membrane domain of prokaryotic rhomboid is also dimeric in detergent DDM, independent of its cytoplasmic domain (77). However, a recent report using single-molecule analysis demonstrated that rhomboid proteases exist as monomers in the cell membrane while dimers are non-physiological (78) (Figure 1-3B).

The Ln region has been shown to be important for maintaining maximum GlpG activity (66). Due to its position within the protein, it is likely to be in close contact with the membrane. In an elegant study by Reading and coworkers, they compared GlpG incorporated in nanodiscs or SMA-lipodisc, and used hydrogen-deuterium exchange mass spectrometry (HDX-MS), to study the dynamics of CytoD and Ln, by measuring the exposure to the solvent. It was shown that changes in the PE/PG ratio between C43 and BL21 *E. coli* strains did not seem to affect HDX significantly, whereas changes in the temperature-induced modifications in chain length and saturation of the membrane lipids, in C43 cells. This may be related to previous evidence from detergent micelle and bicelle systems that hydrophobic mismatch could exert an inhibitory effect on GlpG activity (76, 79).

1.4.3 Catalytic mechanism – From recognition to cleavage

Intramembrane protease substrates are transmembrane α-helices, which in principle need to be locally unwound before protease cleavage because most or all protease active sites bind substrates or inhibitors as extended β-strand (80). Helix-destabilizing residues in the substrate TM region may cause local unfolding, facilitating exposure of the membrane-embedded cleavage site via membrane thinning around the enzyme and/or its presentation to the proteolytic active site (81).

Rhomboid proteases show some level of sequence specificity: small amino acids are strongly preferred in the P1', negatively charged side chain are frequent in P1, and large hydrophobic residues in P4 and P2' positions of the substrate (82). Additionally, GlpG prefers positively charged residues in P3 and P2 positions (83, 84).

The full extent of substrate/protein interaction interface is unknown, but several studies suggest that the TMD of the substrate binds the enzyme first into a site that is spatially separated from the active site (exosite), and then the recognition motif interacts with the active site leading to the proteolytic reaction (49, 83, 85). This mechanism was supported by the observation that active site-directed peptidyl aldehydes inhibit the cleavage of a substrate of GlpG by a non-competitive mechanism, by binding to an enzyme-substrate complex. These data led to the

proposal of the “interrogation/scission” mechanism. The first encounter complex is termed “interrogation complex” complex, where the substrate binds to the exosite. If the binding is favorable, the second “scission complex” is formed, where the substrate’s recognition motif is fully aligned in the active site of the enzyme (83) (Figure 1-3C).

As stated before, the catalytic site of GlpG is formed by the residues S201 and H254, stabilized by an H-bond. The sequences surrounding the catalytic serine of rhomboids and chymotrypsin are similar, but the molecular architecture is different. In chymotrypsin, the backbone amide of the first glycine of the GxSG motif is pointed into the active site and, together with the amide of the serine, forms the oxyanion-binding site of the protease (86), while in rhomboid, the amide group of the glycine is pointed away from the active site and does not contribute to catalysis (87).

In GlpG, the catalytic serine acts as a nucleophile that attacks the scissile carbonyl carbon bond of the substrate. Most globular soluble serine proteases instead use a catalytic triad, typically Asp-Ser-His, to extract a proton from the OH group of the catalytic serine and thus poise it for nucleophilic attack of a substrate peptide bond (88). Molecular dynamics and quantum mechanics studies proposed that the lack of the third, proton-abstracting residue in rhomboid makes the catalytic H254 of GlpG more acidic than the OH group of S201 in the unliganded state of the enzyme (89), which means that S201 is protonated in the ground state of the enzyme. Gradual desolvation of the active site of GlpG by the substrate is proposed to induce an increase in the pKa of H254, leading to a concerted proton abstraction from OH of S201 and nucleophilic attack. This is different in classical serine proteases containing a catalytic triad, such as chymotrypsin, where the higher pKa of catalytic His facilitates deprotonation of catalytic serine and nucleophilic attack can be separated into distinct catalytic steps (90). This hypothesis could explain the markedly low catalytic efficiency of GlpG compared to classical serine proteases chymotrypsin or trypsin.

Serine protease catalysis involves two tetrahedral oxyanion intermediates whose negatively charged oxygen is stabilized by hydrogen-bonding in a structure termed “oxyanion pocket” (91), which in GlpG is formed by the side chains of amino acids H150 and N154 and the main-chain amide of S201. Rhomboid-catalyzed reaction occurs in the lipid membrane, and the reaction mechanism involves a covalent intermediate acyl-enzyme that must be hydrolyzed to complete the reaction cycle. The catalytic dyad of rhomboid is about 10Å below the membrane surface (64), the active site is open to bulk solvent, and the delivery of water molecules for the

deacylation reaction seems unhindered (67). However, molecular dynamics studies suggested that a cavity near the catalytic serine might act as a "water-retention site" that facilitates the delivery of water molecules from bulk solution into the membrane-immersed catalytic center and enhances the catalytic efficiency of GlpG (92).

All the steps described before occurs in connection with the lipid bilayer. It has been shown that the membrane environment can influence the activity and specificity of rhomboid protease (60, 93, 94). Proteins that are normally non-substrates become cleaved by rhomboid when reconstituted in membranes, hence inducing specificity by modifying the dynamics of the substrate (95). Additionally, a subclass of drugs known to modulate γ-secretase activity acted on the membrane directly and induced non-substrate cleavage by rhomboid proteases but left true substrate cleavage sites unaltered (94).

1.5 Aim of the work

Despite the fact that GlpG is the best characterized intramembrane protease, there are still many structural and functional questions open, since most of all the structural information has been obtained from the TMD, ignoring the rest of the protein. Therefore, the available information is mostly biased, overlooking a region of the protein that can be relevant for the function of the enzyme.

My objective in this work is to study the structure and dynamics of the full-length GlpG, by employing a set of experimental and computational tools. In the first part of this work, I studied the structure and dynamics of full-length GlpG in solution, solubilized in the detergent fos-choline-12, by using small-angle X-ray scattering (SAXS), double electron-electron resonance (DEER) spectroscopy and molecular dynamics simulations. In the second part, I employed coarse-grained molecular dynamics simulations to study the full-length GlpG, embedded in a native-like model lipidic bilayer, to model the dynamics of the enzyme in a membrane environment, its interactions with lipids and the role of the soluble regions in the protein/lipids interactions.

2 Materials and Methods

2.1 Experimental methods

2.1.1 Cloning

Full-length wild-type GlpG (wtGlpG) and mutants were amplified by polymerase chain reaction (PCR). For wtGlpG, the primers were modified to use *XhoI* (forward) and *BamHI* (reverse) restriction enzymes (Sigma-Aldrich, Germany) (Table II.1). The amplified products were cloned into a pnEA-vH plasmid with N-terminal His-tag and ampicillin antibiotic resistance. To obtain the mutants for site-directed spin labeling (SDSL), endogenous position C104 was mutated to alanine. Then, positions V17, A71 and V203 were mutated to cysteine, generating single- and a combination of double-mutants (Table 2-1).

Table 2-1: Primers employed

Primer	Protein
5'-TATTCTCGAGTGATGATTACCTCTTTTGC-3' (F)	wtGlpG
5'-TATTGGATCCTTATTTTCGTTTTCGCGCATTG-3' (R)	
5'-GTGATGATGATCGCCGCTGTGGTGGTGTTTATT-3' (F)	GlpG C104A
5'-AATAAACACCACCACAGCGGCGATCATCATCAC-3' (R)	
5'-GTGGCGCAGGCGTTTTGTGATTACATGGCGAC-3' (F)	GlpG V17C
5'-GTCGCCATGTAATCACAAAACGCCTGCGCCAC-3' (R)	
5'-GCGGCGAGCTGGCAGTGCGGCCATACCGGCAGT-3' (F)	GlpG A71C
5'-ACTGCCGGTATGGCCGCACTGCCAGCTCGCCGC-3' (R)	
5'-GGCGGGCTTTCTGGCTGTGTGTATGCGCTGATG-3' (F)	GlpG V203C
5'-CATCAGCGCATACACACAGCCAGAAAGCCCGCC-3' (R)	

Plasmid pKS508_SP encoding for a chimeric, non-native substrate of GlpG (C-terminal His-tag, ampicillin resistance), was kindly donated by Prof. Dr. Kvido Strisovsky (Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech). The substrate is composed of maltose-binding protein (MBP), 50 amino acids of protein translocase TatA from *Providencia stuartii*, and thioredoxin (Trx).

2.1.2 Protein expression and purification

wtGlpG, GlpG mutants, and chimeric substrate MBP-TatA-Trx were expressed in LB media (Carl Roth, Germany) in C43 *E. coli* cells. The cultures grew to an optical density of 0.6 – 0.8 (measured at a wavelength of 600 nm), and the temperature was reduced to 20°C. 0.1 mM isopropyl-thiogalactopyranoside (IPTG, Carl Roth, Germany) was added to induce the expression. Cells were harvested the next day and resuspended in lysis buffer (Appendix 4.1) and lysed using an Avestin EmulsiFlex-C3 high-pressure homogenizer (Avestin, Inc, Canada). Cell debris was pelleted at 20,000 x *g* for 30 min, and membrane fractions were isolated by centrifuging the supernatant at 200,000 x *g* for 1.5 hours. Isolated membranes were resuspended in solubilization buffer (Table 5-1) at a protein concentration of 1 – 5 mg/ml and solubilized with 1% n-dodecyl-β-D-maltopyranoside (DDM, Glycon Biochemicals GmbH, Germany) at room temperature, during 2 hours, or 0.5% DDM at 4°C overnight. The solubilized membranes were further centrifuged at 200,000 x g for 30 minutes, to eliminate non-solubilized fractions.

For immobilized metal affinity chromatography (IMAC) purification of wtGlpG, 5 ml of Protino® Ni-NTA Agarose beads (Macherey-Nagel, Germany) were equilibrated with 5x column volumes of Buffer A + 0.1% Fos-choline-12 (FC12, Anatrace, USA) (Appendix 4.1). Solubilized membranes were incubated with the equilibrated Ni-NTA beads at room temperature during 2 hours, and the beads washed with 5x columns volumes with Buffer A + 0.1% FC12. The protein was step-wise eluted with the IMAC elution buffer + 0.1% FC12, at 100 mM, 250 mM, and 1 M imidazole (Table 5-1).

2.1.3 Sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to estimate the size and purity of protein samples under denaturing conditions. Protein samples were analyzed by the Mini Protetra Cell electrophoresis system (Bio-Rad, USA). Depending on sample concentration the protein solution was mixed thoroughly with 5x sample buffer. Next, a 10% Bis-Tris Gel (1.0 mm x 15 wells) gel cassette was mounted in a gel tray for SDS-PAGE (Bio-Rad, USA), to which about 200 ml Laemmli SDS running buffer (1x) was added. 15 µl of the protein sample was loaded into individual wells. 5 µl of protein size standard PageRuler^{TM} unstained (Thermo Scientific, USA) was added to the first well. The gel was run at 200 V and a maximum of 80 mA for about 45 min.

The gel was stained with Coomassie Blue stain solution and destained with destaining solution (Table 5-1).

2.1.4 Functional cleavage assay

In order to check its functional integrity, 1.5 μM GlpG was incubated with 5 μM of the chimeric substrate MBP-TatA-Trx, in 20 μl of total assay buffer volume (gel filtration buffer), at 37°C up to 2h. The cleavage of the TatA sequence was checked using an SDS-PAGE gel, following the generation of the fragments correspondents to MBP (45 kDa) and Trx (20 kDa).

2.1.5 Thermal stability by differential scanning fluorimetry

Intrinsic protein fluorescence is associated mainly to tryptophan (Trp) residues, which are strongly sensitive to the polarity of the environment. Trp fluorescence is excited at 280 nm and emission occurs at 330 nm non-polar environment and at 350 nm for polar environments. Trp residues are usually hidden in the protein hydrophobic core or shielded by the detergent micelle. Upon protein unfolding, Trp is exposed to water resulting in a decrease of their fluorescence intensity and shifts the emission maximum to longer wavelengths. Thus, by measuring the changes in Trp fluorescence intensity, the melting temperature (Tm) of proteins can be determined in a dye-free approach (96, 97).

The stability of wtGlpG and mutants purified in FC12 was followed by using a nanoDSF differential scanning fluorimeter (Prometheus, NanoTemper Technologies, Munich). The intrinsic fluorescence at λ = 330 and 350 nm after excitation at λ = 280 nm was used to monitor the fluorescence change upon heat unfolding. 10 μl of the samples at a protein concentration of 0.5–1 mg/ml was loaded in a capillary, and the unfolding was then measured at a heating rate of 1°C min^{-1}. The first derivative of the unfolding curves was used to determine the transition midpoint and the Tm.

2.1.6 Site-directed spin labeling (SDSL)

SDSL has become a powerful tool to study the structure and dynamics of soluble and membrane proteins. In this technique, a spin-label side chain is introduced at a defined site of the protein, in order to extract information about the dynamics and/or the environment surrounding that site employing Electron Paramagnetic Resonance (EPR) spectroscopy. The most common method to introduce these labels is based on Cys substitution mutagenesis

followed by modification of the unique sulfhydryl group with a specific paramagnetic nitroxide reagent (98, 99). This method also requires the presence of Cys residues only in the sites of interest, and any other additional Cys residues must be replaced by mutation, usually by Ala or Ser.

Among the several spin labels available on the market, the most commonly used is (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSL) (100) due to its thiol specificity, and its molecular volume comparable to a Trp side chain. The spin-label binds to the protein by a disulfide bond with the respective cysteine and the resulting spin-labeled side chain is abbreviated as R1. This R1 is also flexible, minimizing disturbances of the native fold and the function of the protein it is attached to. In addition, the particular dynamic properties of R1 provide detailed structural and environmental information derived from the shape of its EPR spectrum (99).

The selection of the protein sites for SDSL was performed with the software MMM (101). We employed the "on column" SDSL approach, where the protein is labeled after binding to the IMAC resin, giving more control over the labeling and avoiding the formation of disulfide dimers. Membrane isolation and IMAC purification protocol of GlpG mutants were modified for SDSL by adding 2 mM DTT (Carl Roth, Germany) to maintain cysteines in a reduced state. After the protein binding in Buffer B, the resin (3 ml) was washed with 5 column volumes (CV) with Buffer A to eliminate the excess of DTT, and then incubated with 3x CV of SDSL buffer (Buffer A containing 1 mM MTSL (Toronto Research Chemicals)) overnight at 4°C. The protein was eluted from the resin as before, with the addition of 1 mM MTSL to the buffer in each elution step. Further gel filtration purification was carried out with a Superdex 200 Increase 10/300 GL column (GE Healthcare, USA), to eliminate the excess of MTSL and separate possible unlabeled protein dimers, and to obtain the labeled monomeric proteins.

2.1.7 Size exclusion chromatography coupled with small-angle X-ray scattering (SEC-SAXS)

Small-angle X-ray scattering (SAXS) is a powerful technique to study the structure and dynamics of biomacromolecules in solution, close to physiological conditions. The quality and interpretation of the data collected by SAXS are strongly dependent on the quality of the sample. In this regard, the monodispersity of the protein sample has to be ensured, as well as the use of a buffer solution that minimizes the background scattering (102, 103). These

requirements, however, cannot always be achieved. Some proteins exist in solution in different conformational states and may form complexes and assemblies with other components. In the case of membrane proteins, purified with detergents, this can be more challenging due to the presence of free micelles, which can have different sizes and scattering intensities, contributing to the polydispersity of the sample. Size-exclusion chromatography (SEC) is a powerful technique to ensure the purity and monodispersity of protein samples, and when coupled with SAXS, it can solve the quality requirements to obtain high-quality data (104).

SEC-SAXS data for GlpG/FC12 complexes were collected in the BioSAXS beamline B21, at Diamond Light Source (Harwell, UK). 45 μl of the samples (GlpG at 10 mg/ml with 0.1% FC12) was injected in an Agilent 1200 HPLC system (Agilent Technologies) with a Superdex 200 increase 3.2/300 column (GE Healthcare), equilibrated with gel filtration buffer + 0.1% FC12. Data were collected on an Eiger 4M detector (DECTRIS, Switzerland) with a fixed camera length of 4.014 m at 12.4 keV, with a q angular range of 0.0035 - 0.37 $Å^{-1}$.

Buffer subtraction and primary data reduction were done with CHROMIXS (105) and data processing and modeling were carried out with ATSAS (106).

Low-resolution *ab initio* model reconstruction from the low-q region (q < 0.12 $Å^{-1}$) of the experimental SAXS data, was conducted using the program DAMMIF (107), where a densely packed interconnected configuration of beads that best fits the experimental data are generated. Equation 1 was used to search for the macromolecular shape by minimizing the discrepancy:

$$\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{exp}(q_i) - cI_{calc}(q_i)}{\sigma(q_i)}\right]^2 \quad (1)$$

where N is the number of experimental points, c is a scaling factor and $I_{calc}(q_i)$ and $\sigma(q_i)$ are the calculated intensity and the experimental error at the momentum transfer q_j, respectively. For each data set, 10 independent reconstructions were generated to verify the stability of the solution and the program suite DAMAVER (108) used to calculate the average and representative models, with P1 symmetry.

2.1.8 Flexibility modeling

Initially, a model of the full-length GlpG/FC12 micelle complex was built. First, a model of GlpG was obtained from the server Phyre (109, 110), using the TMD structure pdb:2XOV and the CytoD structure pdb:2LEP. The program Packmol (111) was employed to assemble an

initial FC12 micelle around the TMD, containing 100 FC12 monomers. Coordinates and topologies for FC12 were obtained from ATB database (112, 113). The whole system was converted to coarse-grained resolution with *martinize* (114). The system was minimized with the steepest descent method up to a maximum of 10000 steps. Equilibration runs were performed using a Berendsen barostat (115) with a coupling time of 10 ps. 3 independent production runs were executed during 2 μs with different starting velocities with the GROMACS suite 5.1.2 (116), using the MARTINI 2.2 force field (117), using isotropic pressure coupling to a reference pressure of 1.0 bar with the Parrinello-Rahman barostat (118) and compressibility of 3.4 x 10^{-4} bar^{-1}. The temperature was controlled at 300 K using the velocity rescaling thermostat (119) with a time step of 10 fs. The final coordinates of the simulations were transformed back to atomistic representation with *backward* (120).

The coordinates of the assembled protein/micelle system were then used to model the flexibility of GlpG, with MultiFoXS (121). MultiFoXS accounts for multiple states contributing to a single observed SAXS profile. Multiple states can correspond to conformational heterogeneity (multiple conformations of the same protein or complex) and/or compositional heterogeneity (varying contents of protein and ligand molecules in the system). The scoring function and an enumeration procedure to compute multi-state models based on a SAXS profile is a modification of the Debye formula as following:

$$\chi^2 = \frac{1}{S}\sum_{i=1}^{S}\left[\frac{I_{exp}(q_i) - c\sum_n w_n I_n(q_i, c_1, c_2)}{\sigma(q_i)}\right]^2 \quad (2)$$

where $I_n(q, c_1, c_2)$ and w_n are the computed profile and the corresponding weight, respectively, for each of the N states in the model; this equation minimizes data overfitting by using a single set of c_1 (excluded volume) and c_2 (hydration layer density) values for all N states. The enumeration of multi-state models (subsets of conformations and their weights) of size N (typically, $1 < N < 5$), such that the corresponding sum of weighted SAXS profiles fits the experimental SAXS data, is performed iteratively using a branch-and-bound (122).

The method takes three inputs: a single atomic structure (or a model), a list of flexible residues and a SAXS curve, and runs in three steps. In the first step, it samples the input structure by exploring the space of the φ and ψ main chain dihedral angles of the defined flexible residues with a Rapidly Exploring Random Trees (RRTs) algorithm (123–126). In the second step, a theoretical SAXS profile and radius of gyration (*Rg*) are calculated for each sampled

conformation. In the final third step, an enumeration of the best-scoring multi-state models is performed using the multi-state scoring function (Equation 2) (127).

TMD/FC12 micelle and CytoD were defined as rigid bodies, while residues 60 – 95 were defined as the flexible linker Ln. During the modeling, TMD/FC12 was kept fixed, and Ln and CytoD were allowed to move.

2.1.9 Continuous-wave (CW) electron paramagnetic resonance (EPR) spectroscopy

CW-EPR spectroscopy reports about the nitroxide spin-label side-chain mobility, solvent accessibility, the polarity of its surrounding environment, and in some cases about the distance between the nitroxide and another paramagnetic center in the protein when the distance is less than 2 nm (128). Hence, a series of spin-labeled variants of a given protein allows the study of the secondary structure, solvent exposure, and the orientation of individual segments of the protein (98, 129).

CW-EPR spectra were recorded at room temperature (293 K) using a homemade X-band (9.3-9.4 GHz) EPR spectrometer fitted with a Bruker dielectric resonator MD5 (Bruker Biospin, Rheinstetten, Germany). Glass capillaries with 0.9 mm diameter were loaded with 20 µl sample containing 40-200 µM spin-labeled GlpG in size exclusion buffer with 0.1% FC12. All spectra were obtained at 0.5 mW incident microwave power and 0.15 mT B-field modulation amplitude.

Labeling efficiency was estimated by second integral analysis of the CW spectra and comparison with the spectrum of an MTSL standard at 100 µM in water (130).

2.1.10 Double Electron-Electron Resonance (DEER) spectroscopy

In DEER spectroscopy (also called Pulsed Electron-Electron Double Resonance (PELDOR) spectroscopy), molecular distance measurements, are based on the dependence of the electron dipole-dipole couplings under an external magnetic field, which scale as $1/r_{AB}^3$, where r_{AB} is the distance between spins centers. The commonly used 4-pulse DEER methodology determines the dipolar coupling between spins in the form of a modulation of the spin echo amplitude, with great sensitivity in the range of 20 to 80 Å (131, 132).

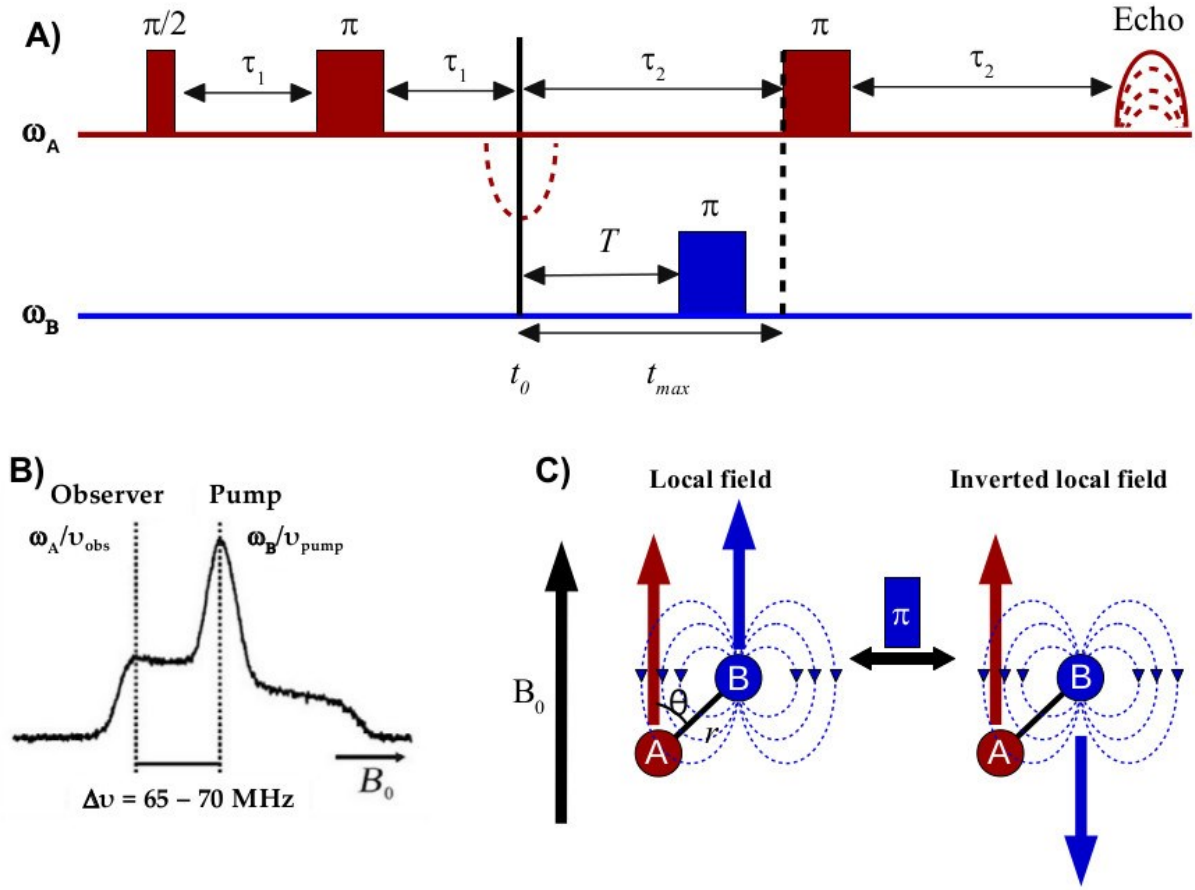

Figure 2-1: DEER spectroscopy. A) 4-pulse DEER sequence. ω_A corresponds to A spins (observer frequency, υ_{obs}) and ω_B corresponds to B spins (pump frequency, υ_{pump}). Pulses labeled with τ remain constant, T pulse is variable. B) Positions of the observer and pump frequencies on a typical EPR absorption spectrum (figure taken and modified from ref. 162). C) The π pump pulse at frequency ω_B inverts the state of spin B, inverting the local field imposed by spin B. The local field of B modulates the field of spin A.

In DEER experiments, microwave pulses are used to selectively excite two distinct spin populations (referred to as spins A and B). The B spins are flipped, which perturb and modulate the coupled A spins. The sequence begins with a two-pulse Hahn echo sequence on ω_A. After the appearance of the Hahn echo, a pump pulse is applied on ω_B with a varying time delay after the echo. At τ_2, the echo is refocused by an additional π-pulse on ω_A. Again, the echo intensity is recorded as a function of the time delay between the first echo and the pump pulse (Figure 2-1.A).

In four-pulse DEER, ω_A is the observed frequency (corresponding to A spins) and ω_B is the pump frequency (B spins). Both frequencies are chosen such that there is no overlap (or minimal overlap) between the excitation windows of the pulses and that the highest number of spins is excited (Figure 2-1.B).

The pump pulse flips the B spins at time *T*, which alters the effective magnetic field of the A spins that are coupled to the B spins. This change in the magnetic field changes the precession frequency of the coupled spins via electron-electron coupling (ω_{ee}), which results in the magnetization being out of phase by the angle $\Delta\varphi_{ee} = \omega_{ee}{}^{T}$. Thus, ω_{ee} can be determined by integrating the echo intensity as a function of *T* as:

$$\omega_{ee} = \frac{\mu_0 g_A g_B \beta_e^2}{4\pi h} \frac{1}{r_{AB}^3} (3cos^2\theta_{AB} - 1) \quad (3)$$

where r_{AB} is the distance between the spins, θ_{AB} is the angle between the static field B_0 and the vector between the spins, and ω_{dd} is the dipolar coupling between the electrons (Figure 2-1.C). Equation 3 is valid as long as the positions of the electron spins are relatively well defined in relation to the distance between them. This restriction is met for spins more than 15 Å apart, which is the lower limit for a DEER experiment.

Pulse EPR experiments were performed on a Bruker ELEXSYS E580 spectrometer (Bruker Biospin, Rheinstetten, Germany) at Q-band (34 GHz) using the ER 5106QT-2 resonator (Bruker). A sample volume of 50 µl containing 40-200 µM spin-labeled GlpG in detergent solution was loaded into quartz capillaries with 2.4 mm inner diameter. The system was equipped with an Oxford helium cryostat temperature regulation unit CF935 with a temperature controller ITC 503S (Oxford Instruments, Oxfordshire, UK). All measurements were performed at 50 K. The four-pulse DEER sequence (133, 134) was used as:

$$\pi/2\ (\upsilon_{obs}) - \tau_1 - \pi\ (\upsilon_{obs}) - t' - \pi\ (\upsilon_{pump}) - (\tau_1 + \tau_2 - t') - \pi\ (\upsilon_{obs}) - \tau_2 - \text{echo}$$

A two-step phase cycling (+‹x›, −‹x›) was performed on $\pi/2$ (υ_{obs}). Time t′ is varied, whereas τ_1 and τ_2 are kept constant. The dipolar evolution time is given by $t = t' - \tau_1$. Data were analyzed only for $t > 0$. The resonator was over-coupled.

Two different setups have been used in the DEER experiments. In both cases, the pump frequency υ_{pump} was centered at the maximum resonance whereas the observer frequency υ_{obs} was 50 MHz (Q-band) higher. The pump frequency υ_{pump} was centered either (i) in the resonator dip, or (ii) 30 MHz higher than the resonator dip center. For setup (i) the resonator was over-coupled to Q ~ 100, for setup (ii) a resonator Q of ~ 1300-1600 (as determined by the spectrometer) was adjusted. The observer pulse lengths were 32 ns for a π pulse and 16 ns for a $\pi/2$ pulse in both setups. The pump pulse length was either (i) 12 ns or (ii) 16 ns. In setup (i) proton modulation was averaged by adding traces at eight different τ_1 values, starting at $\tau_{1,0}$ =

320 ns and incrementing by $\Delta\tau_1$ = 16 ns. Setup (ii) has been used for the spin-diluted samples. For all other DEER experiments scheme (i) has been applied.

Analysis of the data was performed with *DeerAnalysis 2018* (135) and distance distributions were obtained by Tikhonov regularization after background correction. Validation of distance distributions was carried out with the validation tool included in the *DeerAnalysis* software. For the boundaries and number of trials tested with the respective datasets the suggestions from the software based on the result of the background correction performed have been taken. The parameters varied were the white noise, the background starting time and the background density.

2.2 Computational methods - Coarse-grained simulations (CGMD) of full-length GlpG and GlpG-TMD in a model lipidic bilayer

Computational modeling of biomolecular systems plays a fundamental role to get insights about how proteins function, the interpretation of experimental data, drug design, among others. Traditional atomic-level molecular modeling techniques can address most of these tasks, but its applications and performance are still limited by the size of the system under study, the algorithmic efficiency, and the available computing power (136).

Coarse-grained (CG) models are computationally more efficient compared with all-atoms models and allow simulations during longer time-scales and/or larger systems sizes. Moreover, coarse-grained models enable reasonable reconstruction of modeled structures back to all-atom resolution. These advantages give the possibility of multiscale modeling, based on a combination of the computational speed of coarse-grained models with the high accuracy of classical all-atom MD simulations (137–139).

CG molecular dynamics (CGMD) simulations have proven to be very successful in the modeling of lipid/solvent systems, in particular for biological membranes and membrane proteins, employing the MARTINI force field (140–142). This force field is designed following a “top-down” approach with an extensive calibration against experimental data, especially thermodynamic data such as oil/water partition coefficients (143). The resolution modification is based on a 4:1 mapping, where four heavy atoms are represented by a single interaction center (bead), except for ring-like molecules. For small ring-like fragments or molecules (e.g. benzene, cholesterol, and several of the amino acids), the general four-to-one mapping rule is insufficient and are therefore mapped with higher resolution, up to 2:1 (141, 143).

The Martini model has been extensively compared with other force fields, with respect to several experimental properties, showing that the model performs well semi-quantitatively for a broad range of systems. Among the properties accurately reproduced are liquid densities, area per lipid, accessible lipid conformations, the tilt angle of membrane-spanning helices, bilayer bending modulus, diffusion rates of lipids, peptides, and proteins (140, 142, 144–146).

2.2.1 Systems setup

The same previous structural model for full-length GlpG was employed for the simulations and was converted to coarse-grained representations using the program *martinize* (114). With the program *insane* (147), the proteins were embedded in a 12 x 12 x 12 nm solvated box, where the lipid bilayer is oriented in the XY plane, and the protein is oriented in the Z-axis, matching the hydrophobic region with the membrane lipid tails. The model membranes were constructed placing the following lipids in random positions across the bilayer planes: PE (DPPE: di-C16:0-C18:0 PE, DOPE: di-C16:1-C18:1 PE, POPE: C16:0/18:1 PE), PG (DPPG: di-C16:0-C18:0 PG, DOPG: di-C16:1-C18:1 PG, POPG: C16:0/18:1 PG)-based lipids, and cardiolipin, at a ratio of 70.5 % / 27 % / 2.5 % respectively (Table 2-2).

Table 2-2: Simulations setups

	GlpG		**GlpG-TMD**	
Lipid	**TOP**	**BOTTOM**	**TOP**	**BOTTOM**
DPPE	47	48	49	48
POPE	47	48	49	48
DOPE	47	48	49	48
DPPG	18	18	18	18
DOPG	18	18	18	18
POPG	18	18	18	18
CDL	3	3	3	3
Solvent				
Water	8966-9036		9224-9239	
Na^+	133		138	
Cl^-	66		68	

Lipids topologies and coordinates were obtained from the MARTINI force field web site (148). Na+ and Cl- were added at concentration of 0.15 M for neutralization (Table 2.2).

2.2.2 Simulations

The simulations were performed with the GROMACS suite 5.1.2 (116), using the MARTINI 2.2 force field (141). Elastic network *RubberBand* (149) implemented in *martinize*, was used to restrain the entire protein during the simulations, in order to avoid unfolding by keeping the secondary structure. The systems were minimized with the steepest descent method up to a maximum of 10000 steps. Next, equilibration runs were performed using a Berendsen barostat (115) with a coupling time of 10 ps. Independent production runs with different starting velocities were executed during 10 μs using semi-isotropic pressure coupling to a reference pressure of 1.0 bar with the Parrinello-Rahman barostat (118) and compressibility of 3.4×10^{-4} bar^{-1}. The temperature was controlled at 300 K using the velocity rescaling thermostat (119) with a time step of 10 fs. Five independently built trajectories were simulated for each system.

Calculations were performed at the Hummel computer cluster of the Regionales Rechenzentrum (RRZ) at the University of Hamburg and Dell Precision T5610 local workstations.

2.2.3 Analysis of the CGMD simulations

All analyses were performed on the last 1 μs of the simulations, in order to allow sufficient equilibration of the systems. Before the analysis, all trajectories were structurally aligned to the last frame of the GlpG in the first run.

Radial distribution functions (RDF) of the lipids in the plane of the membrane were calculated with the *xy_rdf* tool implemented in LOOS suite (v2.3.2) (150). The bead corresponding to the PO_4 group of the lipids was used for this calculation. The radial distribution function (RDF) *g(r)* of two particles A and B is a function of the distance *r* from a particle. The value of *g(r)* at a given distance *r* to a particle of type A is the density of particle type B at that distance, divided by the average density of the particle type B. Similarly, it is defined as the ratio between the probability of finding a particle in a given region and the analogous probability if the particles were uniformly distributed. The average probability of a group of particles homogeneously distributed is referred to as the bulk density ρ_{bulk}, and it is assumed to equal to

1. If there is another particle dissolved in the liquid, the density of the solute particles at a given distance is referred to as the local density ρ(r).

$$g(r) = \frac{\rho(r)}{\rho\text{bulk}} \quad (4)$$

A *g(r)* = 1 implies that the probability of finding any particular particle in the region of interest is equivalent to that which would be obtained if all the particles were uniformly distributed, whereas *g(r)* values greater/less than 1 imply an enhanced/reduced probability at certain distance *r*, relative to a uniform distribution (151). In our case *g(r)* is referred to the probability of finding a PO_4 bead of the lipid of interest at a certain distance from the surface of the protein along the *xy* plane of both membrane leaflets. The RDF curves were obtained from the trajectories with a time window of 1 ns.

2D density maps were obtained with the *densmap* tool from GROMACS. Each pixel of the maps corresponds to a spatial bin of 1 $Å^2$ and contains the number of events of the presence of a certain PO_4 bead in a XY coordinate bin, averaged among all simulations for each system.

Protein/PO_4 bead contacts were calculated with the tool *select* (implemented in GROMACS) and mapped onto the CG structure with *in house* scripts. Contacts were defined as protein backbone BB beads and lipid PO_4 beads at a cutoff distance of a smaller than 6 Å.

Structures and trajectories were visualized and analyzed with UCSF Chimera (152) and VMD (153).

3 Results and Discussion

3.1 Dynamics of GlpG in by SAXS and DEER spectroscopy experiments

3.1.1 Purification of GlpG/FC12 micelle complexes

As stated before, most of the structural information available to date comes from the transmembrane domain of GlpG, and only a few studies have accounted for the whole protein (73, 76, 78, 79). One of the reasons for this fact may come from the lack of a proper detergent to purify the protein at high yield and quality. GlpG was first purified in detergent DDM by Ni2+-affinity chromatography, obtaining high quantities of the protein with good purity (Figure 3-1A).

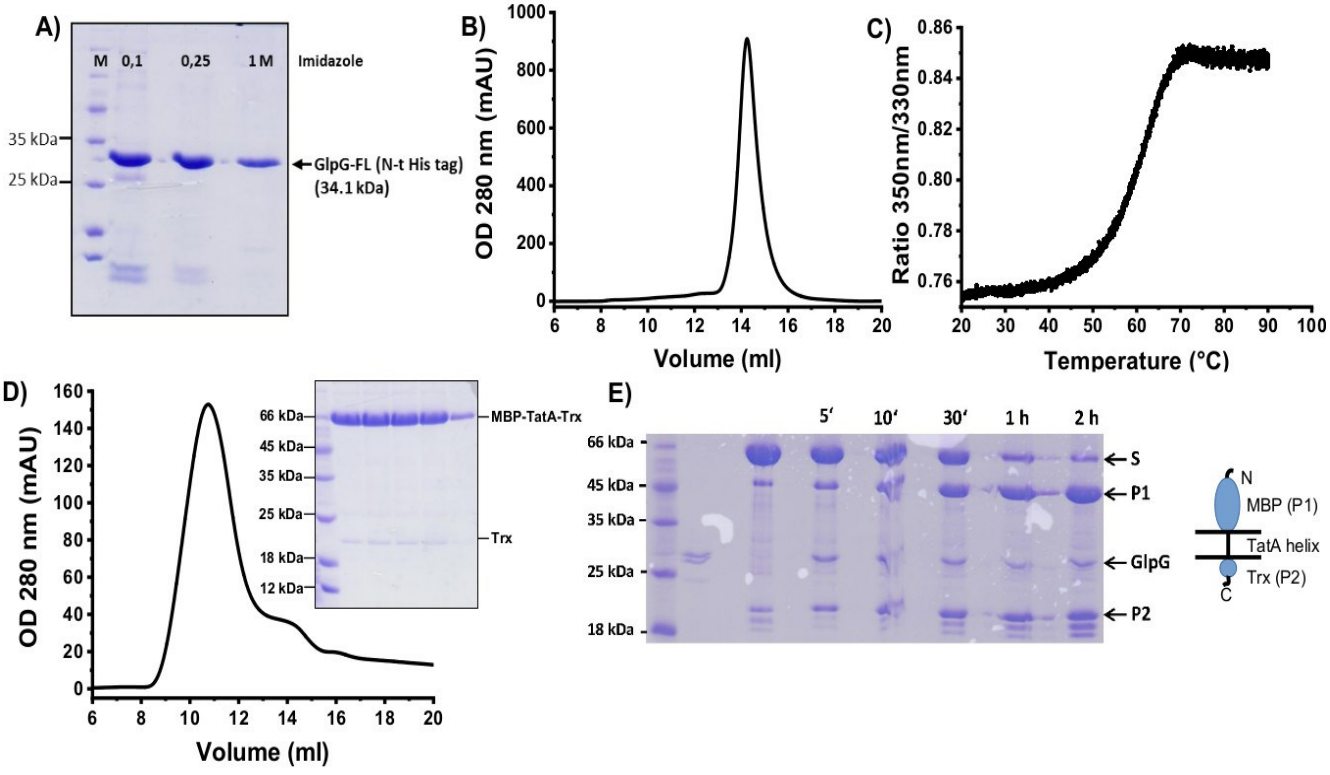

Figure 3-1: Purification of GlpG and chimeric substrate MBP-TatA-Trx. A) IMAC purification of GlpG. B) Size-exclusion purification of GlpG. C) Differential scanning fluorimetry of GlpG. Tm = 62.4°C. D) Size-exclusion purification of MBP-TatA-Trx. E) SDS-PAGE of a functional assay of GlpG. GlpG was incubated with MBP-TatA-Trx (S) up to 2 h at 37°C. The activity of GlpG was tested by the generation of fragments correspondents to MBP (P1, 45 kDa) and Trx (P2, 20 kDa).

In a subsequent size-exclusion chromatography purification of the molecule, the detergent was replaced by Fos-choline-12 (FC12), obtaining a nice profile, showing a pure, monodisperse, and stable protein (Figure 3-1.B). The stability was further checked by differential scanning fluorimetry, showing a typical sigmoidal behavior of the thermal unfolding of the protein, obtaining a transition melting temperature (Tm) of 62.4°C (Figure 3-1.C). Hence, we identified the detergent Fos-choline-12 (FC12) as a good detergent to purify GlpG, obtaining the protein at high yields, pure, stable.

We also purified the chimeric substrate MBP-TatA-Trx to test the functional integrity of GlpG (Figure 3-1.D). The protein was obtained with sufficient purity for the assay with only a short fragment corresponding to Trx, due to spontaneous cleavage of the substrate. In the functional assay, we incubated GlpG and MBP-TatA-Trx during 2 h, and we followed the cleavage of the TatA helix by the generation of the fragments correspondents to MBP and Trx (Figure 3-1.E). The product fragments are generated consistently along the time, demonstrating that GlpG is active in FC12.

With the successful purification of GlpG in FC12, we first tried to obtain crystals of the protein using vapor diffusion and lipidic cubic phase crystallization. After many trials, our efforts to crystallize GlpG with FC12 were unsuccessful.

3.1.2 Size-exclusion chromatography coupled to small-angle X-ray scattering (SEC-SAXS) experiments

We conducted SEC-SAXS measurements to determine the low-resolution structure of the GlpG/FC12 micelle in solution, and get some insights about the dynamics of the protein/micelle complex. The sample is initially separated with a size-exclusion column and then immediately exposed to X-rays (Figure 3-2.A). As before, the SEC purification yielded a nice, high and symmetrical peak corresponding to a well-separated GlpG/FC12 complex. The processing of the data with CHROMIXS (105) gave the trace of the I_0 SAXS intensities for every frame recorded, which matches perfectly with the UV recordings. During the processing, a region after the I_0 peak was selected for background subtraction. This step is very important when working with protein/micelle samples because it suppresses the scattering contribution of the free micelles. After this step, we analyzed the distribution of the radius of gyration (Rg) along the peak and we selected the region were Rg had less variation (Figure 3-2.B) and obtained an averaged SAXS profile (Figure 3-2.C).

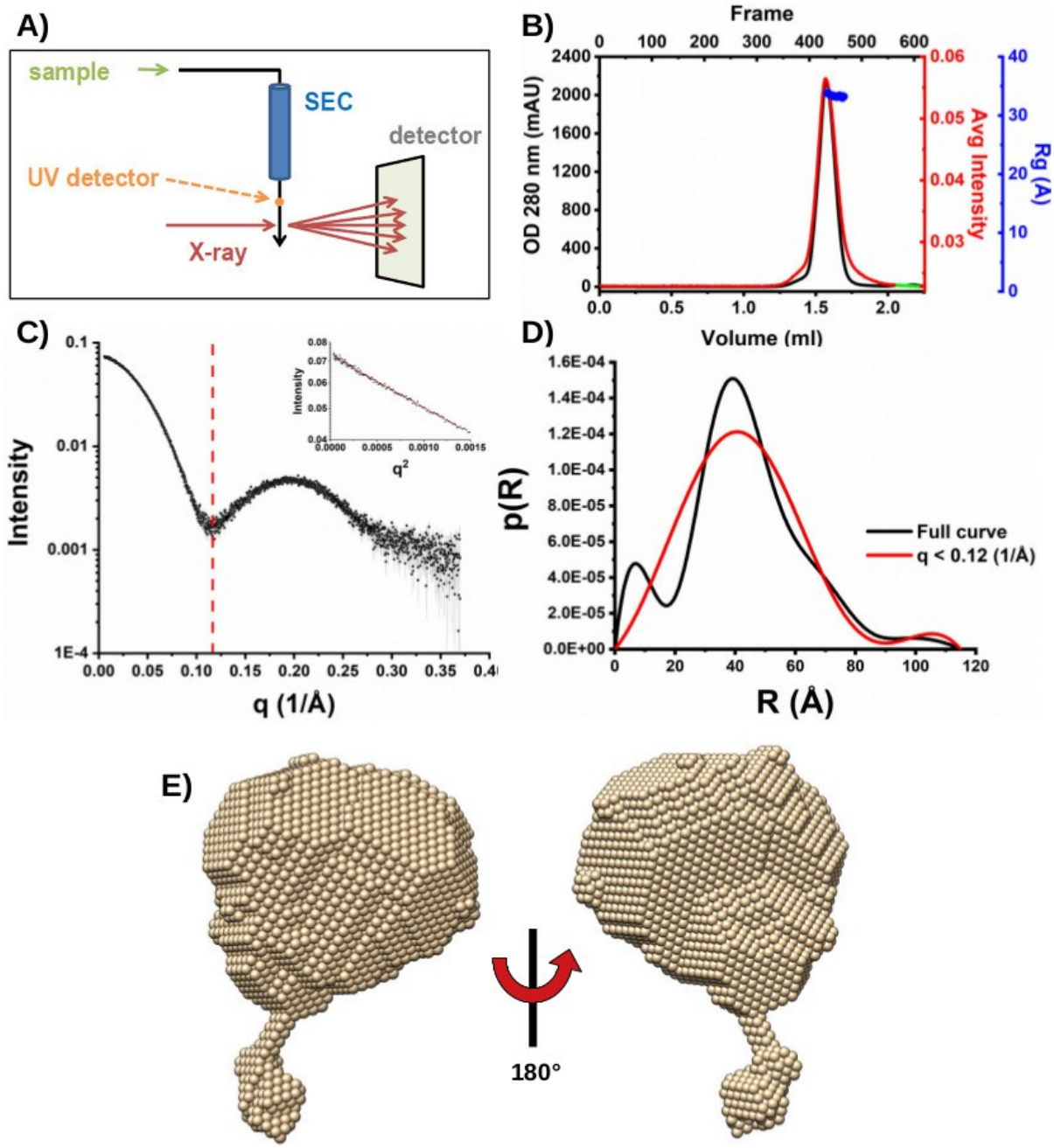

Figure 3-2: Size-exclusion chromatography coupled with SAXS (SEC-SAXS). A) Schematic setup of the SEC-SAXS run. B) SEC-SAXS profile. (black) UV trace of the eluting GlpG/FC12, (red) average scattering (I_0) intensity, (blue) radius of gyration (R_g) of the frames selected for analysis, (green) region selected for the buffer background subtraction. C) Averaged SAXS profile of the frames with the same Rg used for the analysis, and the Guinier region at very low angles (inset). The red dashed line represents the q region used for the modeling ($q < 0.12$ 1/Å). Distance distribution function (p(R) plot), calculated from the entire SAXS curve and from the region selected for modeling. E) Low-resolution ab initio envelope modeled with DAMMIF (108).

The profile suggests a shape similar to those found for protein/micelle and discoidal systems (154–157), with a minimum at 0.12 $Å^{-1}$ and a broad maximum at $0.15 < q < 0.25$ $Å^{-1}$. Model-free analysis of the SAXS data yielded an Rg = 33.4 Å from the Guinier plot (Figure 3.2B, inset), a Dmax of 115 Å, and a Porod volume of 71.3 nm^3.

The distance distribution plot of the whole profile shows some features typical of multi-contrast systems with two different maxima (Figure 3-2.D). This behavior originates from the differences in the electron density of the components of the sample. Alkyl chains of FC12 have a lower electron density of ~0.277 e/$Å^{-3}$ (158), while the buffer is ~0.334 e/$Å^{-3}$, contributing negatively to the scattering. Moreover, GlpG and FC12's head groups, both have a higher electron density of ~0.420-0.490 e/$Å^{-3}$ causing the maximum at 45 Å (154, 156, 158). The dumbbell p(R) function indicates that the scattering signal is dominated by the FC12 micelle, which interacts closely with the protein.

In order to get more insight into the overall shape of the GlpG/FC12 micelle particle, an *ab initio* low-resolution model was obtained with DAMMIF (107). For this, only the low-q region of the curve was used ($q < 0.12$ A^{-1}), to minimize the contribution from the inhomogeneous internal structure due to the low electron density and describing only the overall particle shape/dimensions. The model showed a globular body with a small tail (Figure 3-2.E). The first could correspond to the GlpG-TMD/micelle and the tail is thought to be the Ln and CytoD domains.

3.1.3 Dynamics of GlpG/FC12 complex

Ln region is thought to be very flexible, which suggest that the Ln-CytoD region may exist in different conformations in solution. However, the structure of this region is unknown, and in order to study the full-length structure of GlpG, it has to be modeled. We use the server Phyre2 (110) to obtain a model of the full-length GlpG. From the sequence, 240 residues (87%) were modelled at > 90% accuracy, by selecting PDB 2XOV (TMD) and PDB 2LEP (CytoD) as structural templates. We selected the structure 2XOV because it is the one with the best resolution among all the TMD's structures so far. The backbone Cα-Cα distances are extracted from these models while the regions not covered by templates (Ln region and C-terminal) are modeled *ab initio* using the Poing method (159).

Next, we need to generate a GlpG/FC12 micelle model to be used as a starting point for flexibility analysis. The size-exclusion purification profiles suggest that the GlpG/FC12 particles have a molecular weight (MW) of ~68 kDa. GlpG has a MW = 34 kDa and the FC12

micelle has a MW = 38 kDa, so the difference between the MW_{GlpG} and the estimated by SEC is filled by FC12, which has a monomeric MW of 351 Da. On the other hand, the aggregation number (N_A) of FC12 ranges between 54-108 depending on the method and conditions used for the determination (158, 160). With all these data from the literature, and considering that, membrane proteins are purified with an excess of detergent, at a concentration above the CMC, we assembled *in silico* 100 monomers of FC12 around GlpG-TMD, which gives a good compromise to our experimental conditions and the background information (Figure 3-3.A). With this then, $MW_{GlpG} + 100 \times MW_{FC12\ monomer} = MW_{GlpG/FC\ complex} = 69.1$ kDa.

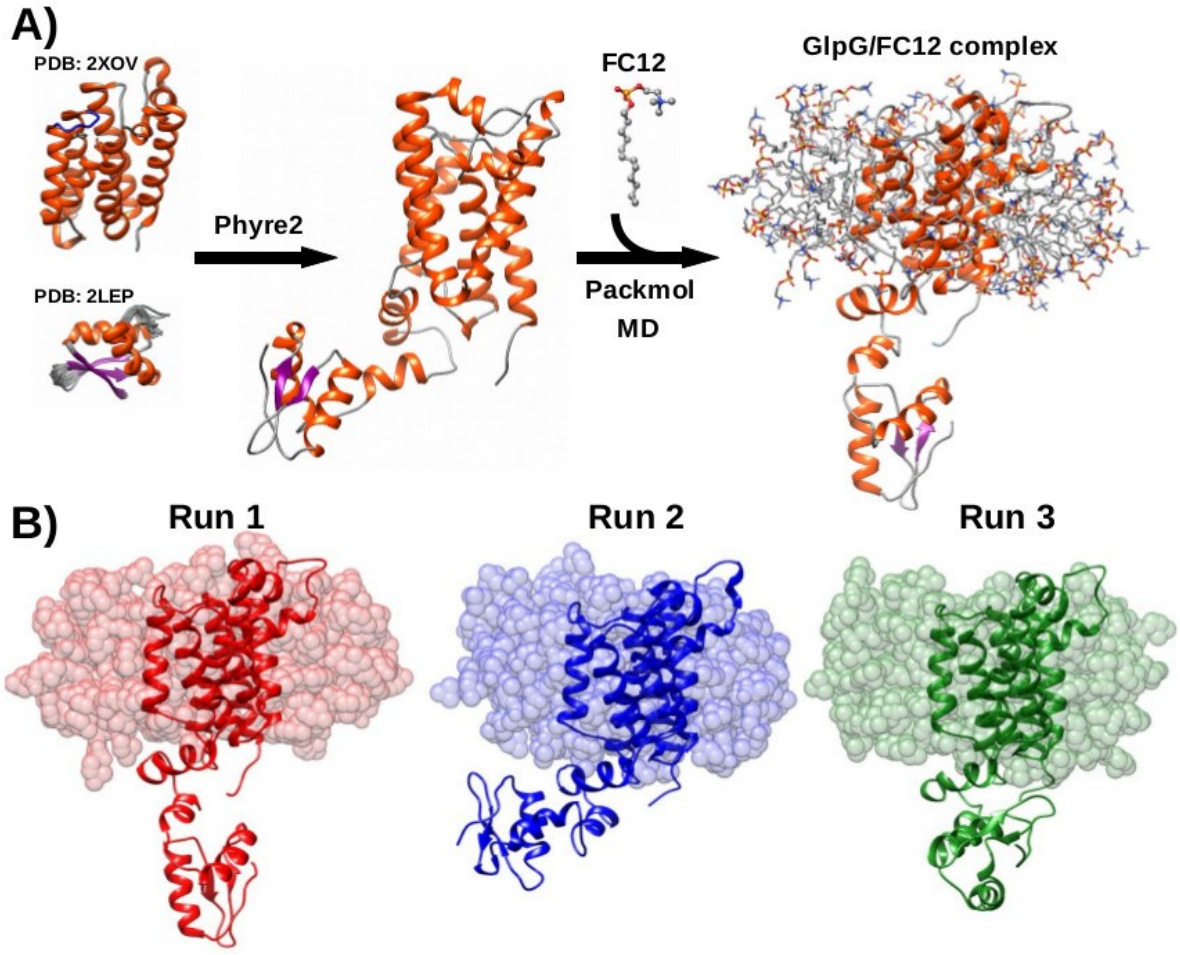

Figure 3-3: *in silico* GlpG/FC12 assembly. A) Full-length model of GlpG was obtained with the Phyre2 server (110), employing the structures PDB:2XOV (TMD) and PDB:2LEP (CytoD). Packmol software (111) was used to generate an initial model of the FC12 micelle around GlpG-TMD, with 100 monomers. Finally, an assembled GlpG/FC12 micelle model was obtained by MD simulation. B) Final snapshots from 3 different independent CGMD simulations. Models were back mapped to atomistic representation with backward code (120).

The entire system was converted to coarse-grained representation in order to save computational time and performed 3 independent runs during 2 μs (Figure 3-3.B). FC12 monomers wrap around GlpG-TMD with a discoidal shape, as suggested by the p(R) function.

I used MultiFoXS (121) to model the flexibility of the Ln region. The SAXS profile obtained is a description of the scattering properties of all species in the sample. In the case of monomeric flexible proteins, the SAXS profile describes the average of all conformations of the molecule. MultiFoXS addresses the problem under the same assumption that the SAXS profile is explained by an ensemble of N-states that coexist in solution. Our molecule is then modeled as two rigid bodies (TMD/FC12 micelle and CytoD) connected by a linker (Ln). 50000 conformations were generated with the RRTsample method, which explores the conformational space of the φ and ψ backbone dihedral angles of the defined flexible residues with the RRTs algorithm (124–126). This iterative strategy allows exploring a wide region of the conformational space in a significantly more efficient way compared to random sampling (127). Next, a theoretical SAXS profile is calculated for each conformation with FoXS (121) and multi-states models are enumerated and scored with MultiFoXS by using the Equation 2 (see Material and Methods, section 2.1.8) and the branch and bound combinatorial optimization approach (127) (Figure 3-4.A).

Given the input conformations generated and their computed SAXS profiles, multi-state models (subsets of conformations and their weights) of size N are searched, such that the corresponding sum of weighted SAXS profiles fits the experimental SAXS profile. Figure 3-4.B shows the distribution χ^2 (for $\chi^2 < 10$) of multi-states models for each N-state. For N = 1 there is a big variation in the χ^2 and the Rg distribution is very similar to the generated pool of conformations (Figure 3-4.D), showing that a single state model has a low probability to explain the SAXS curve. This variation was drastically reduced for $N \geq 2$, with values of $\chi^2 < 2$, showing that the sum of several conformations can explain the SAXS profile. These ensembles are composed of conformations that can be more "extended" or "compact" with respect to the distance between the TMD and CytoD (Figure 3-4.D).

These models were superimposed into the *ab initio* envelope, showing that the "extended" conformations overlap well while covering part of the big body and the tail of the envelope. On the other hand, the "compacted" conformations superpose well into the whole big blob of the envelope (Figure 3-4.E).

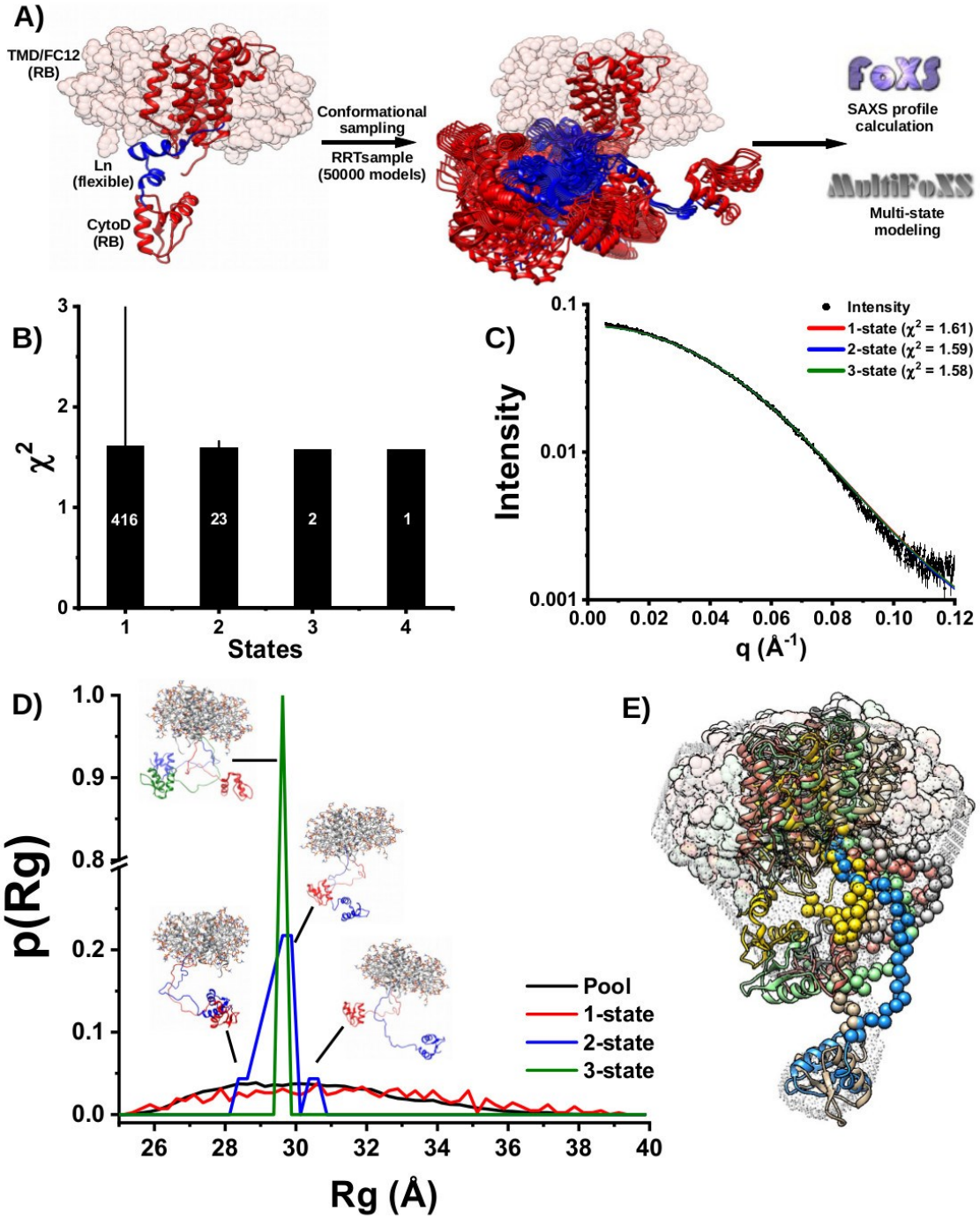

Figure 3-4: MultiFoXS modeling. A) For the flexibility modeling with MultiFoXS (15), the obtained GlpG/FC12 micelle model was used, defining the GlpG-TMD/micelle and the CytoD as rigid bodies (RB, red), while the Ln region (residues 60-95, blue) was defined as flexible. 50000 models were generated during the conformational sampling with the RRTsample method (16). A theoretical SAXS profile and radius of gyration is calculated for each sampled model with FoXS (16). Finally, multi-state modeling is performed with MultiFoXS. B) Distribution of χ^2 values for each state and N-state models with $\chi^2 < 10$. The only model for 4-state is the combination of the two models of 3-state. C) SAXS fitting of the best scoring N-state model. D) Distribution of the Rg for each N-state, and a representative model for each state. Error bars represent the standar deviation of the χ^2. E) Superposition of N-state models on the ab initio envelope obtained with DAMMIF (108).

3.1.4 Intramolecular distances measured by DEER spectroscopy

DEER spectroscopy has become a popular technique to probe the dynamics of macromolecules, by measuring intra- and intermolecular distance distributions. For this purpose, it is necessary to include spin labels containing an unpaired electron, whose absorption signal of electromagnetic radiation will be detectable under an external magnetic field (99). Most proteins do not contain unpaired electrons, so in order to employ this technique, the molecule has to be labeled in defined positions by site-directed spin labeling (SDSL). The most commonly used SDSL method takes advantage of the reactivity of the sulfhydryl group of Cys residues, introduced by site-directed mutagenesis (161), which react with MTSL spin-labels (Figure 3-5.A).

We used the software MMM (101) to select the positions to be mutated by Cys and labeled in the protein, located in the CytoD (residue 17), Ln region (residue 71), and TMD (residue 203). GlpG contains a constitutive Cys residue in position 104, which was replaced by Ala. In the following, all mutants contain the mutation C104A and will be considered as constitutive. We expressed and purified the single and the combination of double mutants. Figure 3-5.A (top) shows the size exclusion purification of the single mutants compared to wtGlpG. The SEC profiles show that V17C and V203C mutants are slightly bigger than wtGlpG but still monomeric, while A71C can form dimers, due to the high exposition and flexibility of this region that contains the introduced Cys residue, forming disulfide bonds (Figure 3-5.B). We tested the functionality of the single mutants by incubating them with the chimeric model substrate MBP-TatA-Trx, during 1h at 37°C (Figure 3-5.C). All mutants were active, checked by the generation of MBP and Trx fragments, by the cleavage of the TatA helix.

Mutant V203C was slightly less active and stable than the rest (Figure 3-5.D), most probably due to the position of the mutation in the core of the TMD, close to the active site. Once the mutants were labeled, dimer species disappear, and all have a higher weight, although still in a monomeric form (Figure 3-5.A (center) and Figure **3-5**.E). GlpG-17R1 shows a peak with the same size as the wtGlpG that can be attributed to non-labeled molecules or a different conformational state of the labeled protein. Double mutants also had a shift in the size, as 71R1-203R1 and 17R1-203R1 had a higher size, 17R1-71R1 had the same size of wtGlpG (Figure 3-5.A (bottom)), and all were monomeric (Figure 3-5.F).

We assessed the labeling by performing continuous-wave (CW) measurements at room temperature. MTSL side-chain dynamics strongly dependent on its surrounding environment, which is encoded in the shape of the CW-EPR spectrum, reflecting how restricted the motion of the label is (Figure 3-6.B) (19, 21). The CW experiments were performed in collaboration with Prof. Dr. Heinz-Jürgen Steinhoff's group at the University of Osnabrück.

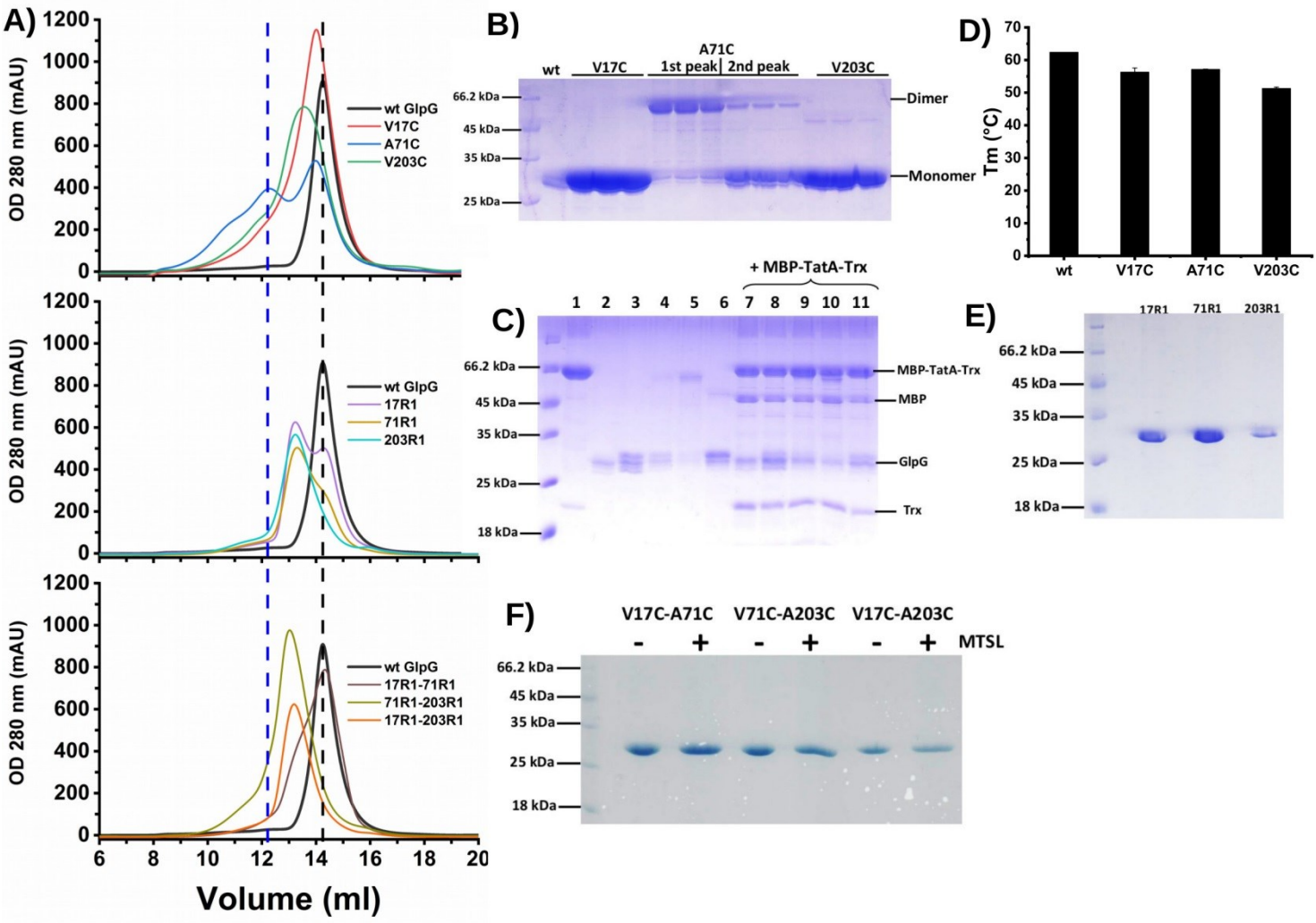

Figure 3-5: GlpG mutants and SDSL. A) Size-exclusion purification of Cys single-mutants of GlpG (top), spin-labeled single-mutants (center) and spin-labeled double-mutants (bottom). Black dashed line: monomer, blue dashed line: dimer. B) SDS-PAGE of the purified Cys single-mutants of GlpG. C) Functional assay of Cys single-mutants of GlpG. Lanes (1) chimeric substrate MBP-TatA-Trx, (2) wt GlpG, (3) V17C, (4) A71C monomer, (5) A71C dimer, (6) V203C, (7-11) correspond to the same order as before, incubated with the chimeric substrate at 37°C during 1h. D) Tm of wt and Cys single-mutants GlpG determined by DSF. Error bars represent the standard deviation of three measurements. E) SDS-PAGE of the size-exclusion purification of spin-labeled single-mutants GlpG. SDS-PAGE of the size-exclusion purified labeled and non-labeled double-mutants of GlpG. F) SDS-PAGE of the double mutants of GlpG, labeled and non-labeled with MTSL.

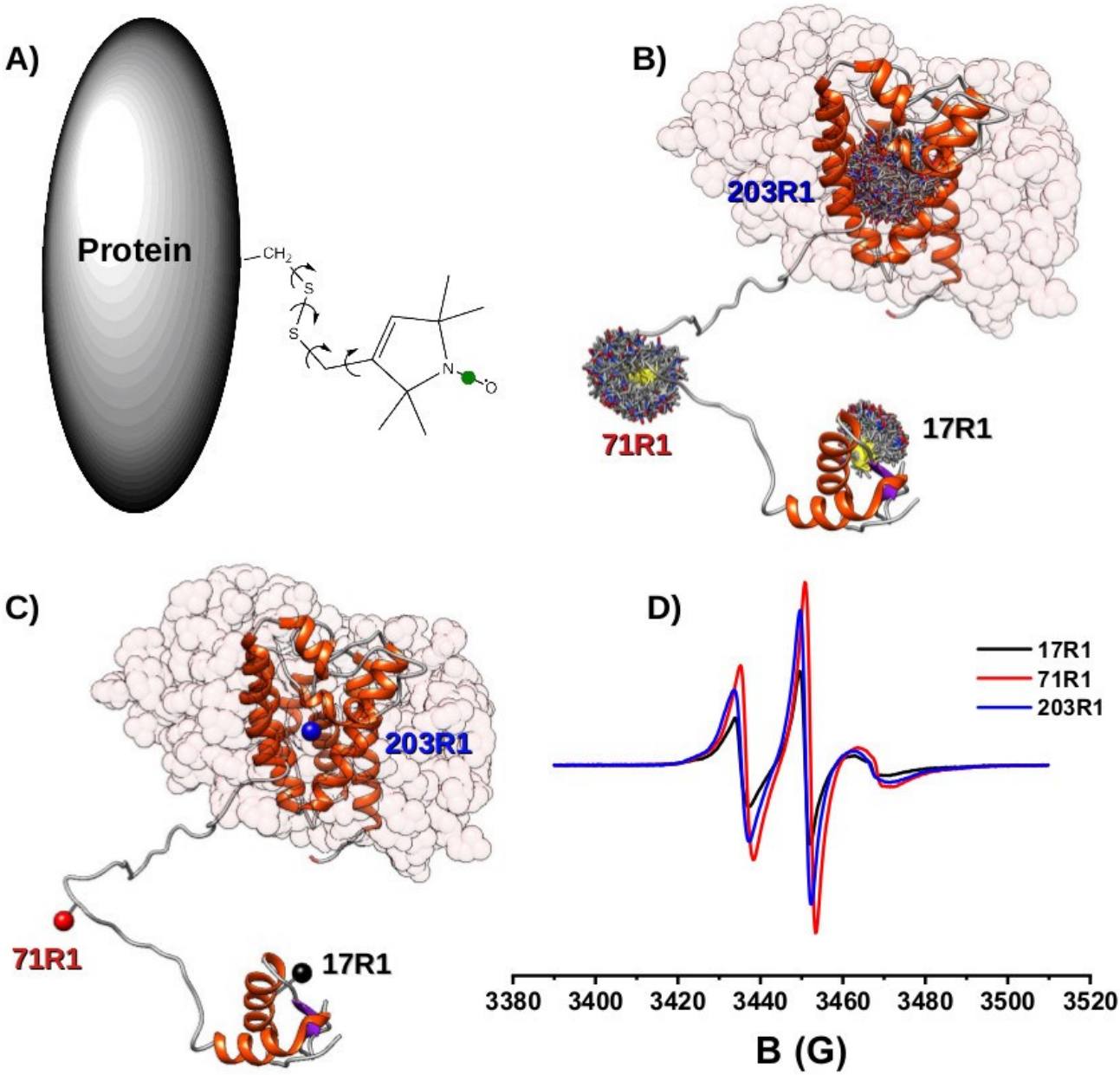

Figure 3-6: Site-directed spin-labeling. A) MTSL spin-label attached to a Cys residue of the protein. The unpaired electron of the nitroxide group is delocalized through the N-O bond (green circle). Curved arrows represent the rotatable bonds of MTSL. B) GlpG/FC12 micelle model with the MTSL label rotamers attached to positions 17 (CytoD), 71 (Ln) and 203 (TMD). C) The colored spheres represent the unpaired electron of the nitroxide. D) Continuous-wave (CW) spectra of single mutants.

CW-EPR spectra show that single- and double mutants of GlpG were successfully labeled (Figure 3-6.D). Interestingly, the spectrum of GlpG-203R1, for which the label is buried in the core of the TMD, has a similar shape as the others that are more exposed to the solvent. This means that the label is less restricted as expected to be in the core of the protein.

It is known that GlpG can cleave casein which is a soluble protein (75, 162), and a popular substrate to test the functionality of the protein in solution. At the moment it is unknown how GlpG can cleave soluble substrates in solution, but in order to do this, it is thought that the protein should undergo a major conformational change of the TMD, in particular, loop L5, to allow the access of the substrate and water for the cleavage, resulting in higher solvent access to the core of the protein, which would explain the shape of the GlpG-203R1 spectrum. This result is also in line with previous experimental evidence that showed increased dynamics of the transmembrane helix 5 and loop L5 in detergent solution (163).

The spectrum of the double-labeled 17R1-71R1 mutant (Figure 3-7.B) shows a shoulder in a region corresponding to an immobile component of the spin-label. This reflects the restriction in the dynamics of the spin-labels due to a strong coupling between the labels at short distances or the interaction with the protein environment, in the case of a compacted conformation of GlpG, as suggested by our SAXS data.

Next, we measured intramolecular distances of the double-labeled mutants of GlpG by dipolar electron-electron resonance (DEER) spectroscopy to characterize the dynamics of GlpG and validate the previous SAXS results. GlpG was labeled with pairs of spin labels in each domain/region of the molecule (Figure 3-7.A). After the microwave pulses, the spin-echo decay of spin-label A is modulated by intramolecular dipolar interaction with spin-label B on the same protein molecule and by intermolecular dipolar interaction with spins A or B on a distant molecule. The oscillating echo decay produced by intramolecular interaction directly reflects the average distance between the spin centers. On the other hand, the intermolecular dipolar interactions have an exponential decay that affects the oscillation and is referred to as the background (164). In order to minimize the background, all double-labeled GlpG mutants were mixed with wtGlpG, at 1:3 (mol/mol) ratio, to minimize the intermolecular dipolar interactions.

These experiments were performed at low temperatures, in the range of 40-60K, which allows maximizing τ_2 (spin-spin relaxation time) and minimize τ_1 (spin-lattice relaxation time) and as result, allowing to measure distances of at least 5 nm with high accuracy (131, 165). In addition, deuteration of the solvent, cryo-protectant or the protein itself, can further improve sensitivity by also increasing τ_{max} and thus obtaining even longer distances to 10 nm or more (132).

Figure 3-7.D shows the dipolar evolution functions of the double-labeled mutants. These functions are obtained after background correction, extracted from the raw dipolar evolution function (Figure 3-7.C) defined as $V(t) = F(t) \times B(t)$, where F(t) represents the form factors and

B(t) is the background. Assuming that the molecules in the sample are homogeneously distributed, the form factors are obtained after dividing the raw spin-echo signal by the background (Figure 3-7.D).

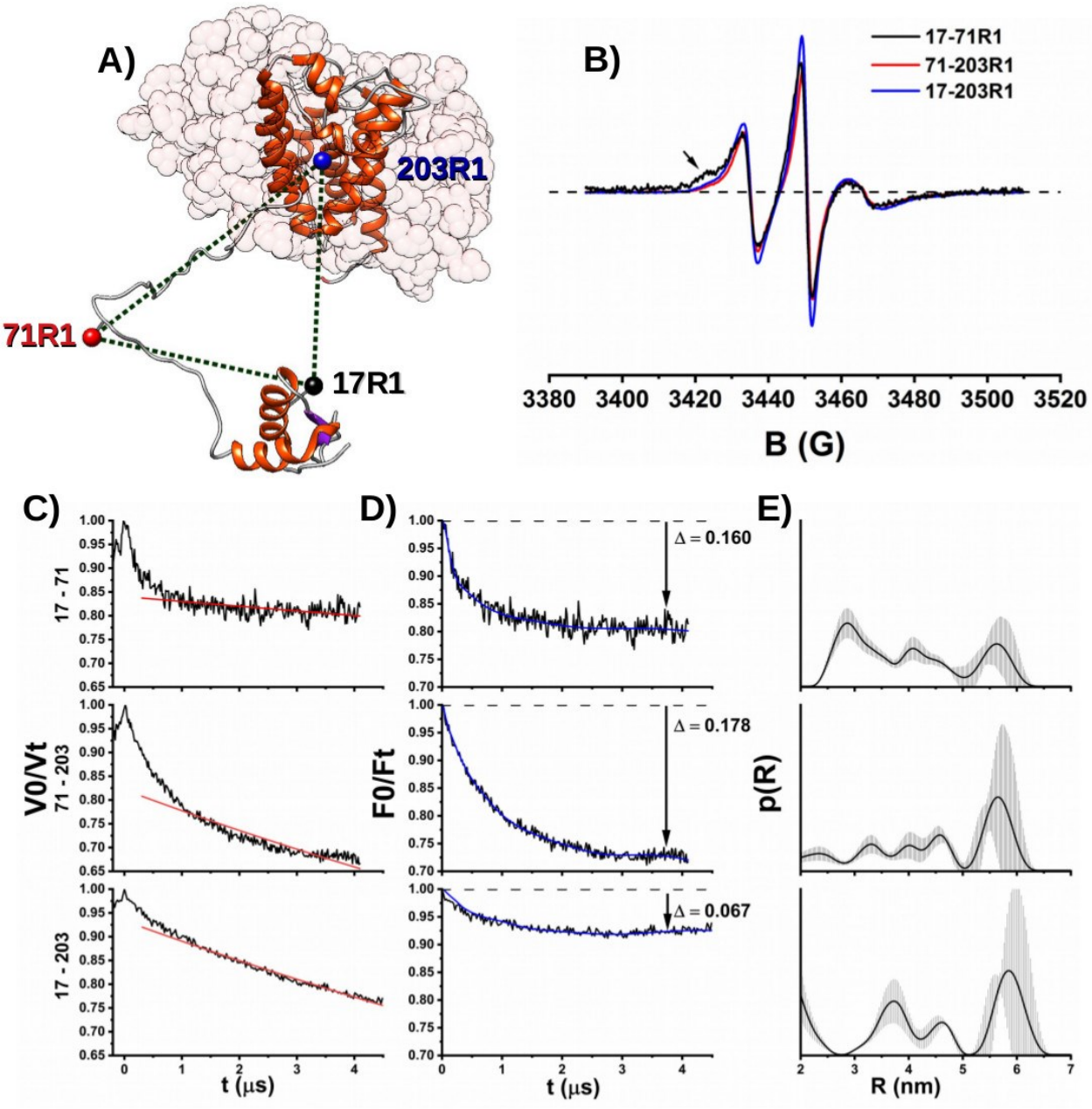

Figure 3-7: DEER spectroscopy of double-labeled GlpG constructs. A) GlpG/FC12 micelle model with the MTSL spin-labels. Dashed green lines show the distances between the unpaired electrons of MTSL labels. B) CW-EPR spectra of double-labeled GlpG mutants. The arrow indicates the immobile component of the 17R1-71R1 construct. Black dashed line indicates the baseline of the spectra. C) DEER dipolar evolution (V(t)) and background fits (red line). C) DEER form factors (F(t)) after background correction. Fits were obtained by Tikhonov regularization (blue line). Arrows show the interspin modulation depth (Δ). E) Interspin distance distributions and distance probability errors determined by the validation tool from DEERAnalysis software (28). $\alpha_{17\text{-}71}$, $\alpha_{71\text{-}203}$ = 200, and $\alpha_{17\text{-}203}$ = 100

With the background correction, values of the dipolar modulation (Δ) can be obtained for each spin pair. This value is influenced by the fraction of spins excited by the pump pulse and the number of interactions, which depends on the labeling efficiency. It was noticeable that the value of $\Delta_{17R1\text{-}203R1}$ was low compared to the rest. Labeling efficiency was determined after calibration with an MTSL sample with a known concentration (166, 167), obtaining labeling efficiencies of all mutants, close to the 100%, so this was not the cause of such value. The other reason for this difference refers to the presence of dipolar interactions coming from short and/or long distances that cannot be seen in the signal, and fall into the background, dampening the form factor and the Δ value of the 17R1-203R1 sample. Short distances are very unlikely to play a role because they are underrepresented in the DEER signal. However, due to the flexibility of the molecule, long distances are very likely to be present in the sample, but they cannot be measured accurately during the dipolar evolution time (τ_{max}) used, and hence they fall in the background, dampening the form factor and the $\Delta_{17R1\text{-}203R1}$ value.

The oscillation frequency of the dipolar evolution function, F(t), is inversely proportional to the third power of the mean distance between two electron spins, and the decay of the oscillations depends on the background and the width of the interspin distance distribution. The form factor F(t) can be then translated in terms of a distance distribution, p(R), by performing Tikhonov regularization (168) (Figure 3-7.E). This deconvolution approach seeks for the best compromise between the resolution of the distance distribution and the suppression of artifacts introduced by noise (smoothness), by screening the regularization parameter α. The determination of the optimum α value is performed using the L-curve criterion (135). This allows selecting the regularization parameter value that gives the distance distribution with maximum smoothness representing a good fit to the experimental data (169).

Figure 3-7.E shows the distance distributions obtained with the software DEERAnalysis (135) for each spin-label pair and the errors associated with the validation of these distances. The presence of several peaks for each pair illustrates again the flexibility of the structure. 17R1-71R1 shows peaks at short, medium, and long distances, 71R1-203R1 shows only a peak at long distances, while the rest are less defined, and 17R1-203R1 shows peaks at 3.5 nm < R < 6 nm. It is interesting that in all samples a peak around 5.7 nm appears which suggest the presence of extended conformations of the protein.

Figure 3-8 shows the accuracy of the distance distribution determination with the Tikhonov regularization validation, by plotting R (nm) vs. σ(p(R)). This plot shows that for pairs 17R1-71R1 and 71R1-203R1 the better accuracy is achieved for R < 5.5 nm, while for 17R1-203R1 it

is for R < 5.7 nm. Shorter distances with R < 1.8 nm cannot be seen by DEER spectroscopy (170, 171).

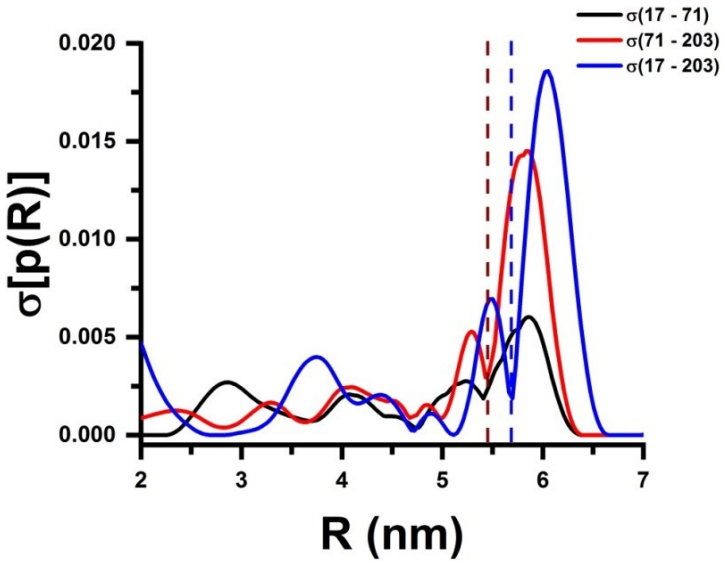

Figure 3-8: Accuracy of the distance distributions. The plot represents the accuracy (error) of the interspin distance determined by Tikhonov regularization. The dashed vertical lines represent the limit for the reliability of the distance determination for each spin pair. Dark red: 17R1-71R1 and 71R1-203R1, blue: 17R1-203R1.

This difference in the limits of the distances accuracy determination is due to the dipolar evolution time (τ_{max}) used for each experiment, where $\tau_{max,\ 17R1\text{-}71R1} = \tau_{max,\ 71R1\text{-}203R1} = 4$ µs, while $\tau_{max,\ 17R1\text{-}203R1} = 4.4$ µs, showing that the longer the evolution time used in the experiment, the better the accuracy of the determined distances, whose limit can be approximated by $R_{max,<R>} \approx 5\sqrt[3]{\frac{\tau_{max}}{2\ \mu s}}$ (132). Under these conditions, for our measurements, the upper limit would be $R_{max,\ 17R1\text{-}71R1} = R_{max,\ 71R1\text{-}203R1} = 6.3$ nm, and $R_{max,\ 17R1\text{-}203R1} = 6.5$ nm. At these distances, the accuracy achieved is not good and hence less reliable. However, despite the fact that the proposed presence of longer distances cannot be detected, the $\Delta_{17R1\text{-}203R1}$ value related with the presence of extended conformations in solution supports the SAXS data, that showed the tail in the *ab initio* model and extended conformations in the flexibility modeling with MultiFoXS.

Taking together, all these data show that GlpG stabilized with detergent FC12 in solution, is a very dynamic molecule due to the flexibility of the Ln region.

3.2 Coarse-grained molecular dynamics (CGMD) simulations of GlpG in a model *E. coli* membrane

As stated before, most of the currently available information about GlpG relies on the solved structure of GlpG-TMD, while the soluble regions have been widely ignored. Other members of the rhomboid family show that these soluble regions can have key functional roles in cell physiology (172, 173). On the other hand, the lipid environment around GlpG and how it influences the dynamics of the molecules was studied employing SMA lipodiscs (79). With this information, we built coarse-grained systems of GlpG and GlpG-TMD embedded in a model of *E. coli* inner membrane containing PE-, PG-based lipids, and cardiolipin, at the ratios determined experimentally (79). The investigated lipids differ with respect to their head group and to the number of unsaturations (double-bonds) in their acyl chains (Figure 3-9) (174).

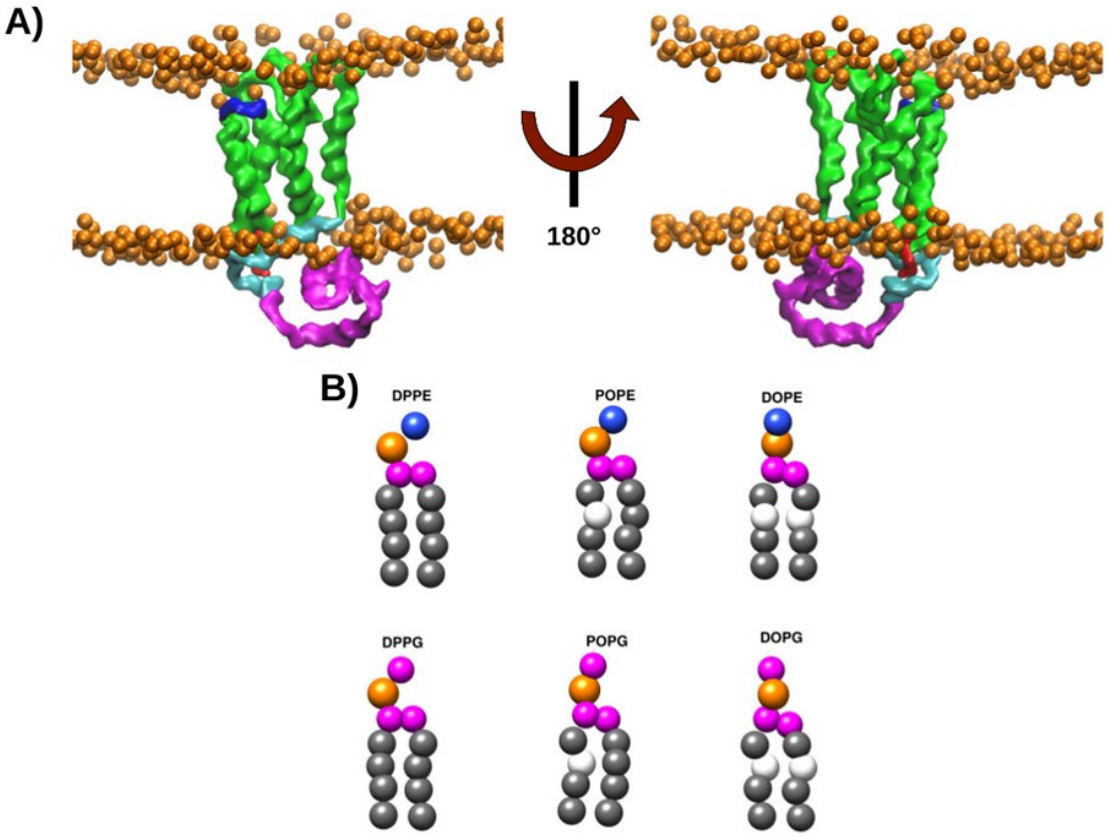

Figure 3-9: Full-length model of GlpG embedded in a phospholipid bilayer. A) Coarse-grained (CG) representation of GlpG embedded in the bilayer. Regions of the protein are shown: CytoD (magenta), Ln (cyan), TMD (green), L5 loop (blue) and C-terminal fragment (red). B) CG models of lipids used with the beads for every atoms group: ethanolamine (blue), PO4 (orange), glycerol carbons (magenta), acyl chain (gray). The white bead represents the unsaturation of the acyl chain.

We computed XY radial distribution functions (XY-RDF) for the phosphate group (coarse-grained PO4 bead) for each phospholipid, in the plane of both leaflets, for GlpG and GlpG-TMD (Figure 3-10). XY-RDF functions inform about the probability to find a certain particle at a given distance from the center of the system under study. In our case, the particles are the PO_4 beads of the lipids and the center of the system is the center of mass (COM) of the TMD. We choose the phosphate bead, because it gives a better estimate of the position of the whole lipid COM, because this does not include the acyl chains, which are very dynamic, and introduces variations in the determination of the COM, increasing the computation time.

RDFs show clear differences between lipids with respect to the interaction with GlpG, depending on the presence of the soluble fragments. DOPG dominates the interaction with the protein in both leaflets and systems, although this effect is more prominent in the bottom leaflet. POPG also shows a strong interaction with GlpG in the bottom leaflet, which is lost when CytoD is eliminated. The RDF plot for GlpG-TMD in the bottom leaflet shows that almost all lipids (in particular DOPG) are located closer to the center of mass of the transmembrane domain, filling the space that was occupied by the CytoD in GlpG.

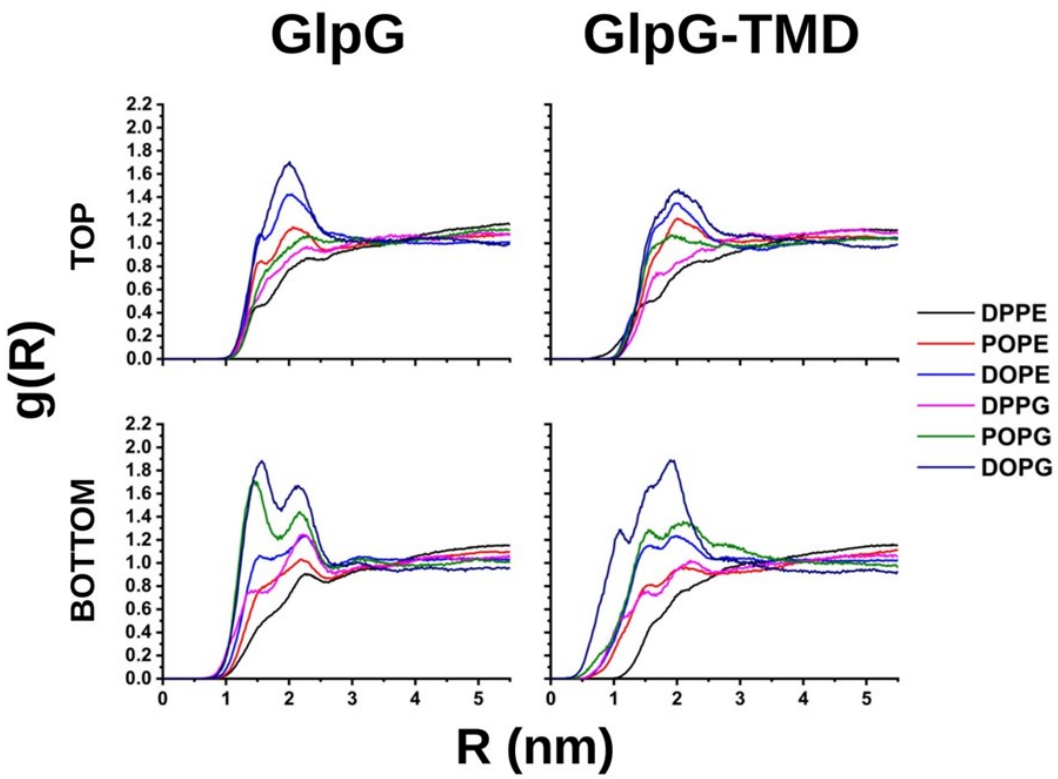

Figure 3-10: Lateral radial distribution function (XY-RDF) analysis, of the lipid phosphate beads in the top and bottom leaflets of the model bilayer, with embedded GlpG and GlpG-TMD. Curves for each lipid represent the average between five independent runs.

XY-RDF plots were also used to choose the last microsecond of the simulation for all the analysis. Since we work with a complex system enough time is needed for convergence, especially for the lipids to diffuse along the membrane (175–177). Lateral XY-RDF of the lipids gives a good estimate to assess how well the membrane is converged (Figure 3-11).

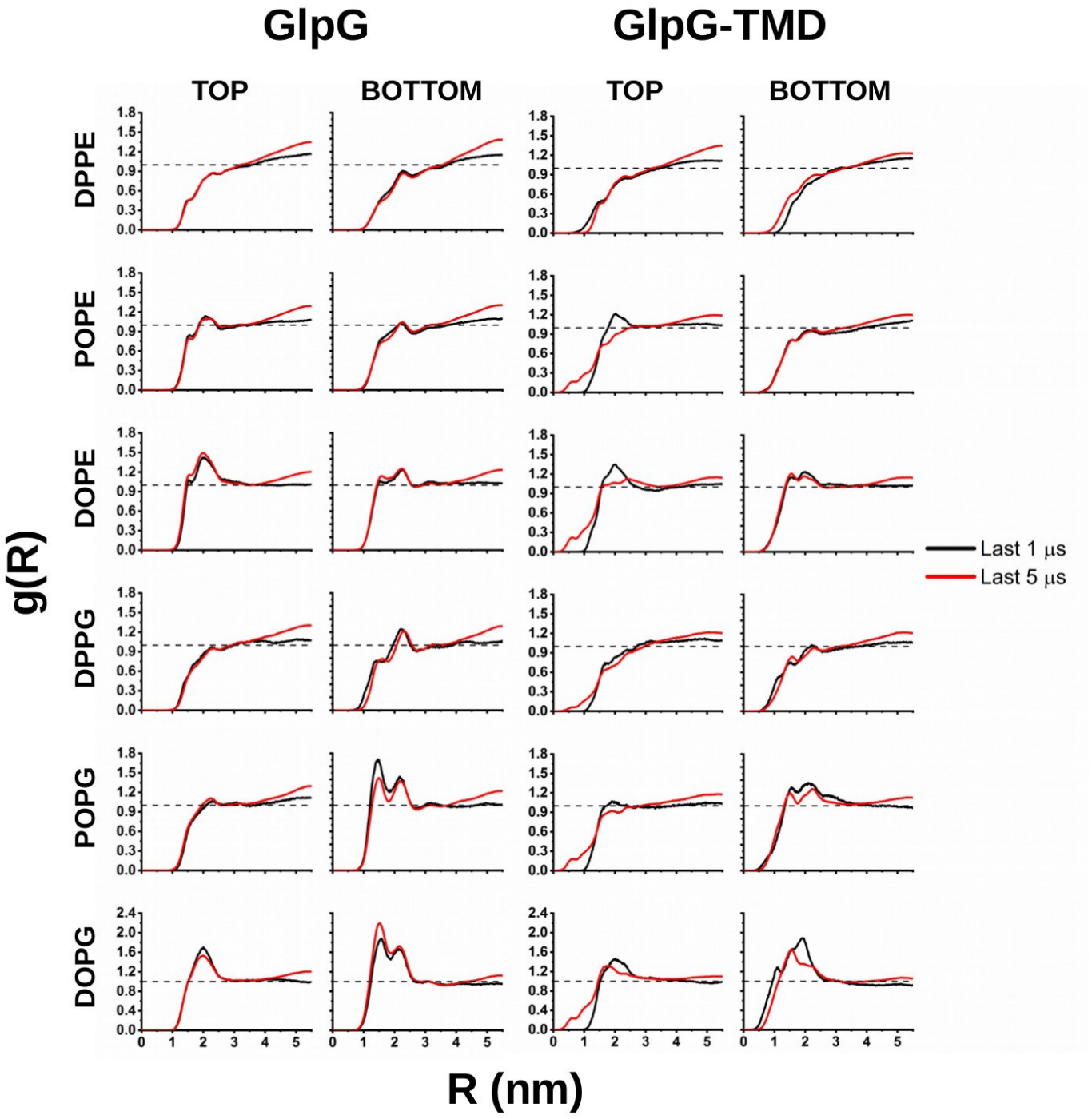

Figure 3-11: Average XY-RDF of the PO_4 bead of all phospholipids around GlpG and GlpG-TMD, computed during the last 1 (black) and 5 μs (red) of the simulations.

Figure 3-11 shows the comparison of the XY-RDFs of all phospholipids PO4 beads during the last 5 and 1 μs. A well converged system will have an XY-RDF tending to 1 at high distances. This is only achieved by averaging during the last 1 μs. If done with the last 5 μs, we are introducing artifacts, produced by a not properly equilibrated system. This can be seen in the higher values of the red curves at high distances.

In order to study the spatial distribution of the lipids in the XY plane of the membrane surrounding GlpG and GlpG-TMD, we computed 2D density maps, averaged during the last 1 μs of the simulations for PE- and PG-based lipids in both leaflets. Each pixel of the maps corresponds to a spatial bin of 1 $Å^2$ and contains the number of events of the presence of a certain PO4 bead in an XY coordinate bin, averaged among all simulations for each system (Figure 3-12 and Figure **3-13**).

The inspection of the maps revealed interesting features about the interaction of the lipids with the protein in each leaflet, in particular for POPG and DOPG (Figure 3-12.B). In the bottom leaflet, the absence of CytoD affects the protein/lipid interactions, in particular for POPG. We observed a cluster of POPG to two regions of full-length GlpG. The first region corresponds to the residues G170, S171, G172 and A272 (TMD), and the second to residues R81, R82 (Ln) and R227 (TMD). These two clusters were affected when the soluble region was eliminated. In particular, the second cluster is formed by the interaction of Arg residues from the TMD and the Ln region, which provides a positively charged patch, responsible for the clustering of the lipid in this area. However, when the soluble region is removed, the clustering and number of contacts between POPG and R227 were significantly reduced (Figure 3-12.B). For DOPG, there was also a strong clustering in the bottom leaflet in GlpG, in a region corresponding to the Ln segment, as well as an accumulation around the protein forming an annulus. This clustering was diminished in the area corresponding to the Ln segment. This indicated that, although residues located in the TMD seem to be sufficient for the lipid-interaction, the presence of certain residues in the cytoplasmic domain and the linker region provides further stabilization to retain the protein/lipid interaction. In this respect, the amount of positively charged residues located in Ln and the region of TMD facing the bottom leaflet may play a role.

The differences between POPG and DOPG in the bottom leaflet of GlpG-TMD can be attributed to unsaturations and the hydrophobic mismatch effect. It has been reported that unsaturated lipids show stronger interactions with several membrane proteins through this effect, which can subsequently affect their activity (178–180). Compared to DPPG, which is

completely saturated, POPG and DOPG show increased interactions with the protein. In addition, the two double bonds of DOPG make the length of the acyl chains shorter compared to POPG (contains one double bond), leading to membrane thinning.

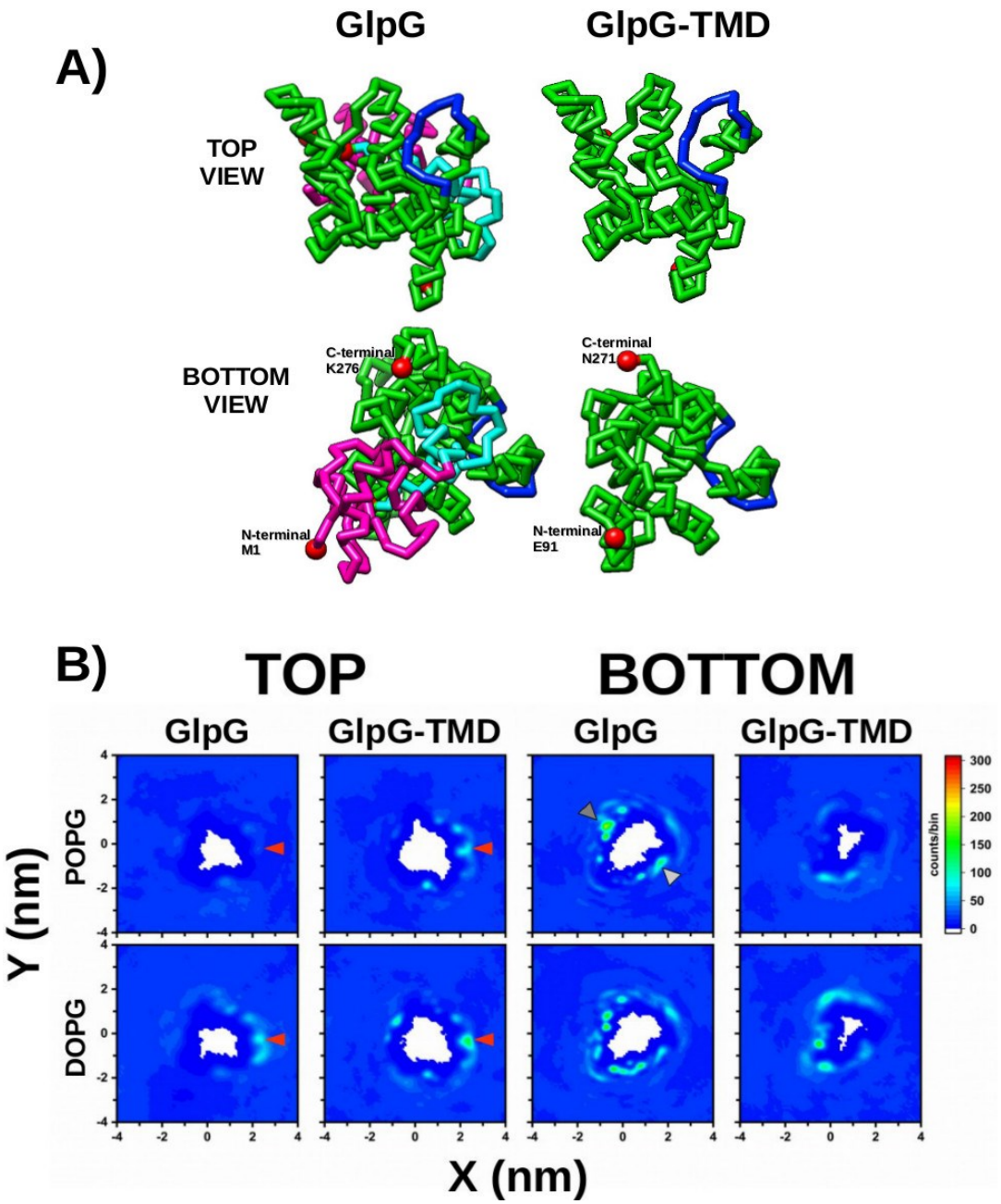

Figure 3-12: Lipid density maps. A) Top and bottom views of superposed GlpG and GlpG-TMD. Magenta: CytoD, Cyan: Ln region, Green: TMD, Blue: L5 loop. B) 2D density maps of the phosphate beads of POPG and DOPG, with embedded GlpG and GlpG-TMD. Each pixel of the maps represents a bin of 1 $Å^2$ and contains the count of PO_4 beads in each grid point along the last microsecond of the simulations averaged between five independent runs. The red arrowheads (top leaflet) indicate the lipid accumulation close to the L5 loop and the substrate-gating site. Dark grey arrowhead (bottom leaflet, POPG) shows the lipid interaction cluster with G170, S171, G172, and A272 (TMD). Light arrow indicates an interaction cluster with R81, R82 (Ln) and R227 (TMD).

This behavior of DOPG could make it more prone to interact with GlpG by matching the hydrophobic surface area of GlpG forming the before mentioned annulus, which is thinner than the average thickness of the membrane (hydrophobic mismatch) (181). POPG, however, with only one unsaturation can be more extended and interacts with the protein in defined residues.

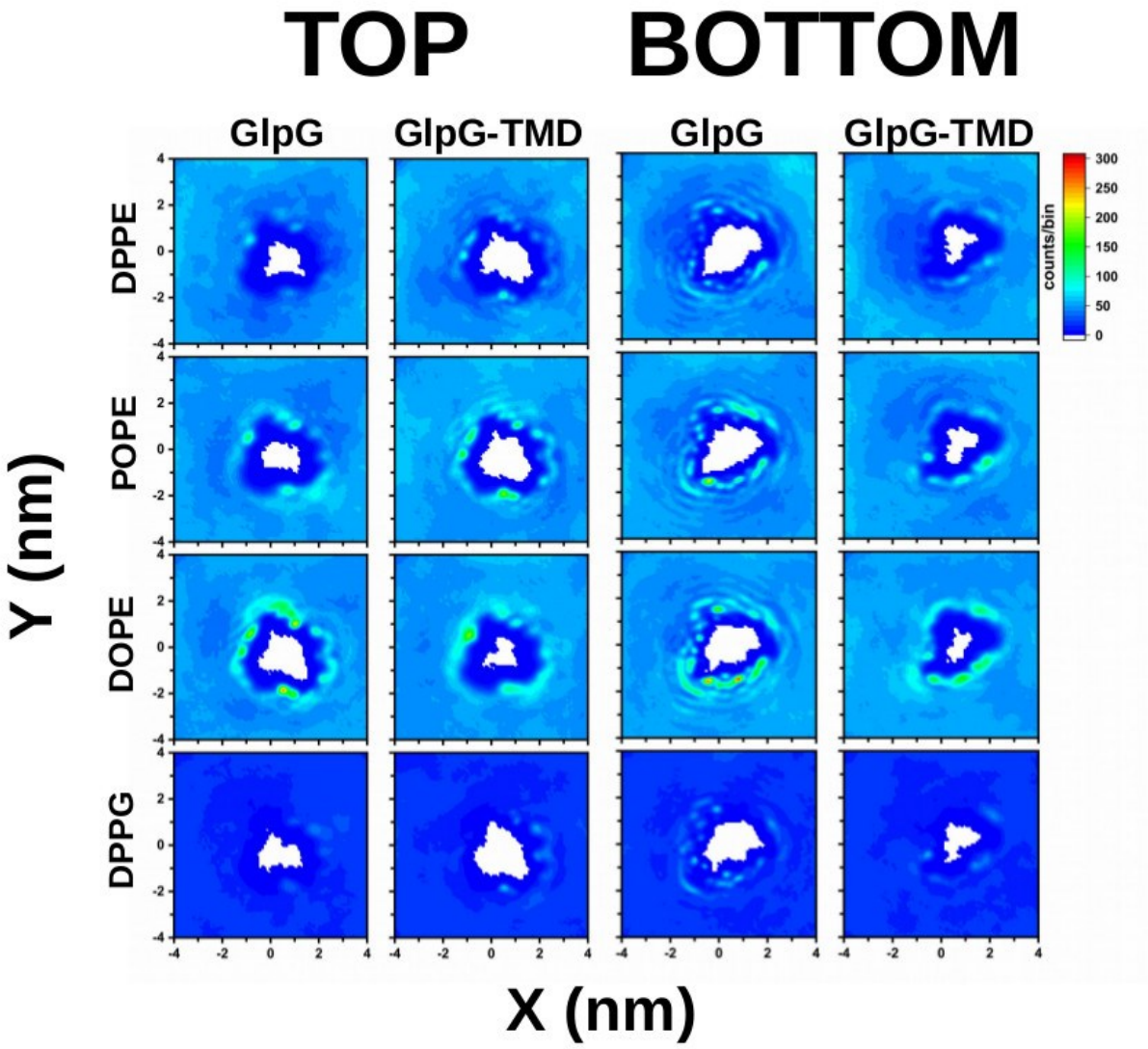

Figure 3-13: Averaged density maps of PE-based lipids and DPPG in the top and bottom leaflet for GlpG and GlpG-TMD.

The hydrophobic mismatch effect on GlpG and its impact on the protein diffusion in the membrane were recently also experimentally demonstrated by Kreutzberger and co-workers (73). They determined the diffusion coefficients of membrane proteins in bilayers with different. Most of the membrane proteins diffused faster in thinner membranes, while rhomboid proteins diffusion slowed in thin membranes that matched its hydrophobic belt.

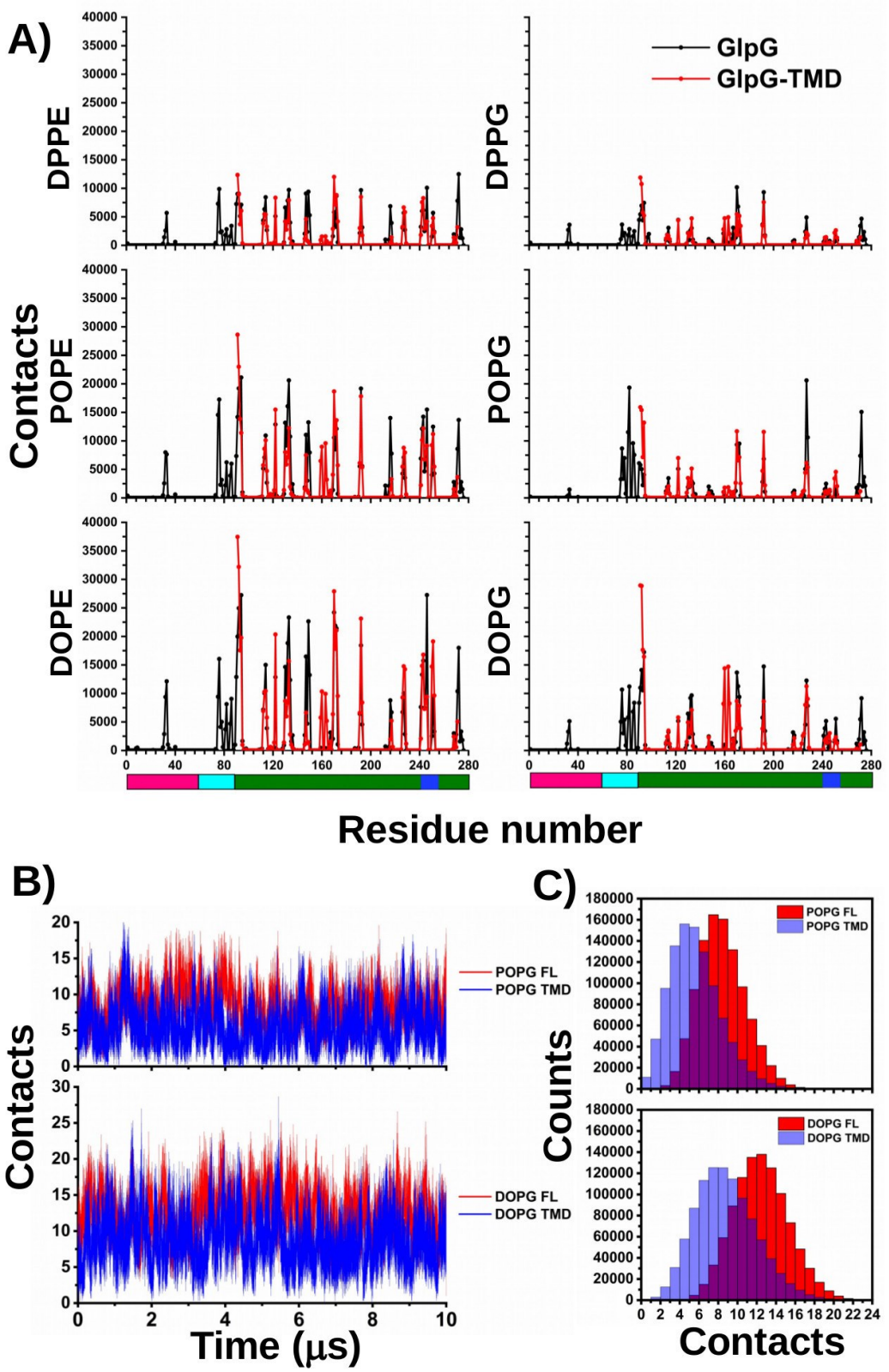

Figure 3-14: Residue-based protein-lipid contacts. A) Number of contacts between backbone BB beads of the protein and PO_4 beads of each lipid. Contacts were defined as the PO_4 at a cutoff distance of smaller than 6 Å. The colored bars represent the regions in the protein. Magenta: CytoD, Cyan: Ln region, Green: TMD, Blue: L5 loop. B) Averaged total number of contacts (both leaflets) of POPG and DOPG with GlpG and GlpG-TMD, along entire simulations. C) Distribution of the contacts along entire simulations.

Conversely, in thicker membranes, rhomboid diffusion was faster. EPR spectroscopy analysis revealed that the soluble regions influence the position of the TMD in the membrane, resulting in an acceleration of the lateral diffusion. In living cells, this acceleration resulted in increased proteolysis. Interestingly, we also detected some changes in POPG and DOPG clustering in the top leaflet (Figure 3-12), in a region corresponding to an interaction with the L5 loop, which is responsible for the protein gating to allow substrates to enter in the catalytic site.

This effect was also evident in the number of protein/PO_4 bead contacts. The contact was defined as the PO4 bead located in a cutoff distance of 6 Å from the protein backbone (176, 182). Differences in the number of contacts were located in the L5 loop and the N- and C-terminals of the TMD (Figure 3-14.A and Figure 3-15). In the case of POPG and DOPG, these had more contacts with GlpG compared to GlpG-TMD, even along the entire simulation (Figure 3-14.B and 3-14.C)

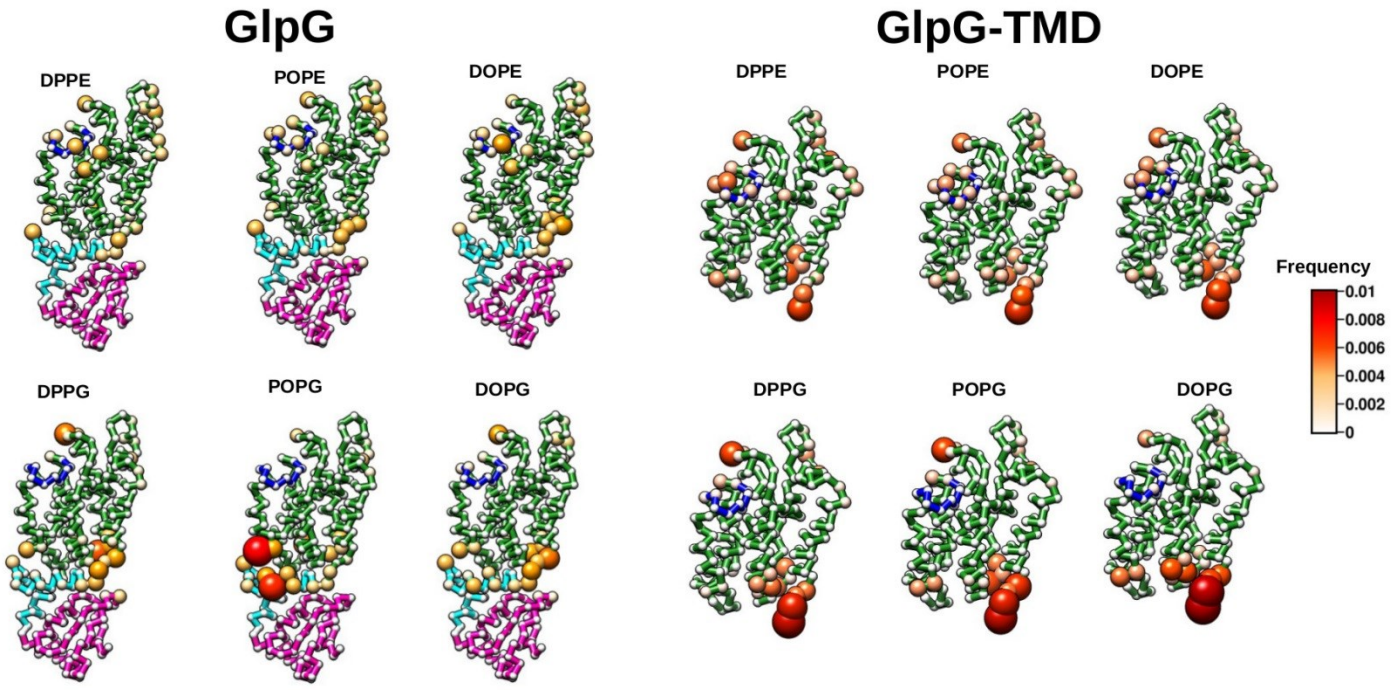

Figure 3-15: 3D mapping of the protein-lipid contacts on the structure. The color scale and the size of the BB beads represent the contact frequency of each lipid PO_4 bead with the protein residues, normalized against the total number of contacts of each lipid. The color of the bonds represents each region of the protein. Magenta: CytoD, Cyan: Ln region, Green: TMD, Blue: L5 loop.

Previous studies have shown the importance of loop L5 *in vitro* and *in vivo* for the gating during substrate binding, and mutations in this region completely abolished the activity (183–185). These variations in the lipid clustering and contacts around the L5 loop may suggest that

these could regulate the access of substrates to the active site and hence the activity of the protein. Similar behavior has recently been described in atomistic simulations, where lipid head groups could compete with the substrate in the membrane by transient contacts in the active site of the protein (186). Furthermore, crystal structures of GlpG-TMD show a detergent molecule in the active site as well as lipid molecules interacting with residues in the cytoplasmatic side of the molecule (70). We also found that the positively charged C-terminal fragment interacts with all lipids. This fragment together with CytoD and Ln is usually lost during the standard proteolytic purification of TMD (64) and thus not present in the available crystal structures.

In summary, our simulations revealed that full-length GlpG preferentially interacts with unsaturated and PG-based lipids. These lipids preferentially accumulate in the annular belt of GlpG, which may lead to membrane thinning due to hydrophobic mismatch (181). In addition, residues in the linker and bottom region of GlpG are involved in the regulation of the lipid distribution in the bottom leaflet of the membrane.

Although our simulations did not show large conformational changes for the linker region, it may still be flexible and adopt extended conformations, as indicated by our SAXS and DEER data, in which the cytoplasmic domain and the linker region are more distant from the membrane, resulting in an arrangement resembling the GlpG-TMD simulated in this study. In that case, the interactions with the lipids in the bottom leaflet would change, affecting their distribution in both leaflets.

We propose this as a putative a mechanism to regulate the activity of GlpG, where the soluble regions of GlpG act as a relay by regulating the lipid distribution around the protein. A similar mechanism has been demonstrated before in yeast, where the cytosolic domain of Rbd2 rhomboid protease influences the organization of the lipids, which in turn regulates actin assembly during clathrin-mediated endocytosis (173). Our results indicate that the cytoplasmatic extensions of GlpG affect the lipidic environment around its transmembrane domain.

4 Final remarks

In this study, we employed experimental and computational techniques to explore the dynamics of the full-length rhomboid protease GlpG from *E. coli*. The soluble regions of GlpG are widely ignored in the literature, although a few recent reports shed new lights about their role (73, 79). Uncovering the molecular functions of these regions will improve our understanding of intramembrane proteolysis and the influence of the lipidic environment in this family of proteins.

Although the structure and dynamics of the protein may differ depending on whether is in a detergent solution or in a native membrane, the fact that the protein remains active in complex with FC12 detergents shows that the protein is properly folded to perform the cleavage of the α-helical substrate. Furthermore, the structure of GlpG-TMD has been solved in both lipidic and detergent environment, with minor changes between them, and only changes in the flexible L5 loop, responsible for the substrate gating to the cleavage site (70). However, despite the fact that the function of the protein is retained in both lipidic and detergent environment, this has an influence on the substrate recognition and specificity. In a lipidic environment, the gating dynamics, concentrated in the transmembrane helix H5 and the L5 loop are restricted as compared to the protein stabilized in detergents, conferring more specificity in the recognition of substrates in the membrane (163).

Our SAXS and DEER data show that the Ln region of GlpG is responsible for the high flexibility of this protein, in agreement with the results obtained by Reading and coworkers about the solvent accessibility of GlpG in native membranes, obtained by hydrogen-deuterium exchange mass spectrometry (HDX-MS) (79). SAXS and DEER data show that GlpG can adopt compacted and extended conformations of the Ln region. As demonstrated before, Ln showed differences in the HDX profile influenced by the lipid composition of the nanodiscs, suggesting that this region is lipid sensitive.

In order to integrate SAXS and DEER data, I tried unsuccessfully to perform hybrid modeling of both data. I think that this could be caused by two reasons. 1) We do not have enough DEER restraints that reproduce the SAXS data, even when the first recapitulates the flexibility of the Ln region. The lack of structure of this part of the protein makes it difficult to select a position for labeling, so we selected the middle point in that section. Also, we have to take into

consideration that the SAXS profile represents an envelope that shows the average all conformations of the protein in solution. On the other hand, DEER data represent a discretization of this population of conformations, since it shows the distribution of conformational subpopulations (as the sum of several gaussians) defined by the distance between discrete positions of the protein. Hence, in order to be able to reproduce a highly flexible system by DEER data, an extensive orthogonalization of many positions in the protein is required. 2) The flexibility of the system imposes a problem for the conformational sampling in order to explore those configurations that satisfy both SAXS and DEER restraints, which arises as a major computational problem. Current software can deal with big conformational ensembles, but they usually fail to find the proper weights to balance the fitting of SAXS, DEER and other data types, particularly when there are several states that describe the whole ensemble (101, 187, 188).

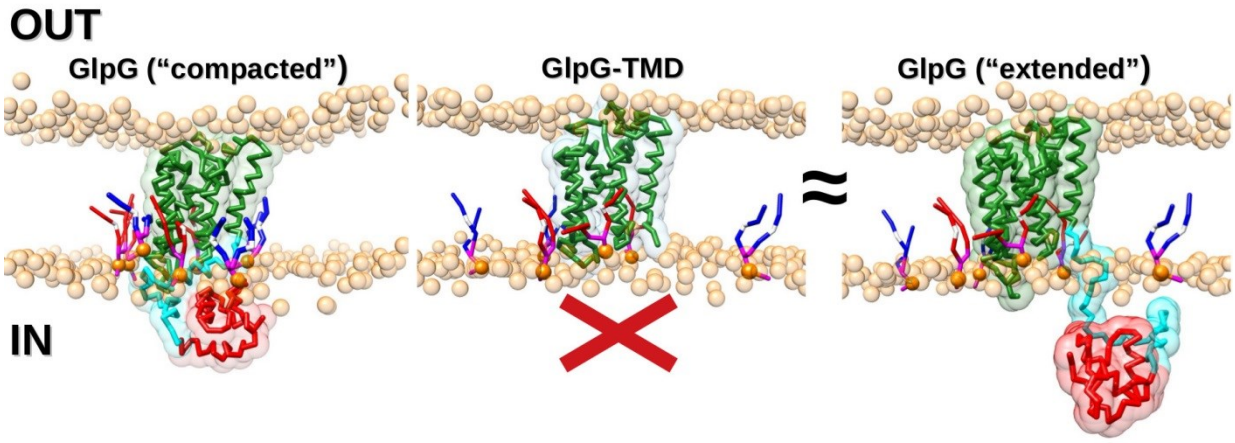

Figure 4-1: Dynamics of soluble regions of GlpG regulate protein/lipid interactions, which could affect the activity of the protein *in vivo*.

Kreutzberger and coworkers demonstrated that the soluble regions of GlpG increase the lateral diffusion in the membrane with functional implications, by the distortion of the membrane by increasing the mismatch effect between the protein and the bilayer (73, 76). Our simulations suggest that this effect may be caused also by the redistribution of lipids around the protein induced by the soluble regions of GlpG. Due to the dynamics of the Ln region, whose fold is still unknown, this effect may be transient depending on the conformation (depicted by SAXS and DEER) and the interaction of this region with the membrane, pointing to a potential regulatory "relay" mechanism (Figure 4-1). In this way, in a case where Ln has an "extended" conformation, where its residues have low or no contact with the membrane, and the CytoD is

away from it, GlpG would behave like GlpG-TMD, which would have less interaction with unsaturated PG-based lipids, diffusing slowly on the membrane, and hence less active. On the other hand, in the case of a "compacted" conformation, where both CytoD and Ln are in close contact with the membrane, these will recruit unsaturated PG-based lipids close to the protein, increasing the hydrophobic mismatch, the lateral diffusion, and the activity. Our "relay" model, although speculative, integrates the current knowledge of full-length GlpG interacting with biological membranes, and suggests further experimental approaches to confirm it.

5 Appendices

5.1 Buffers, instrumentation, and chemicals

Table 5-1: Buffers and solutions

Cell lysis and protein purification	
Lysis buffer	TBS buffer 1x, Glycerol 20%, DNAse 0.01 mg/ml, EDTA 1 mM, PMSF 1 mM, $MgCl_2$ 5 mM, Lysozyme 1 mg/ml, pH 7.5
Solubilization buffer	Tris-HCl 50 mM, NaCl 300 mM, Glycerol 20%, Imidazol 10 mM, pH 8, DDM 1%
Buffer A	Tris-HCl 50 mM, NaCl 300 mM, Glycerol 5%, Imidazol 10 mM, pH 8
IMAC elution buffer	Tris-HCl 50 mM, NaCl 300 mM, Glycerol 5%, Imidazol 100 mM – 1 M, pH 8
Gel filtration buffer	Tris-HCl 25 mM, NaCl 150 mM, Glycerol 5%, pH 7.2
“on column” SDSL	
Buffer B	Buffer A + 2 mM DTT
SDSL buffer	Buffer A + 1 mM MTSL
SDSL elution buffer	IMAC elution buffer + 1 mM MTSL
SDS-PAGE	
1x Electrode buffer	190 mM Glycine, 24.8 mM Tris-HCl, 0.1% (w/v) SDS
Stacking Buffer	0.5 M Tris-HCl pH 6.8, 0.4% (w/v) SDS
Separating Buffer	1.5 M Tris-HCl pH 8.8, 0.4% (w/v) SDS
5x Sample Buffer	50 mM Tris-HCl pH 6.8, 10% (w/v) SDS, 50% (v/v) Glycerine, 125 mM DTT, 0.1% (w/v) Bromophenol blue
Coomassie Staining Solution	2 tablets PhastGel Blue R in 95 % EtOH (40% (v/v)), 25% (v/v) 2-Propanol, 10% (v/v) Acetic Acid
Destaining Solution	10% (v/v) Acetic Acid, 25% (v/v) 2-Propanol

Table 5-2: Instrumentation (listed alphabetically)

Instrument	Manufacturer
Agilent 1200 HPLC system	Agilent Technologies, USA
Äkta*pure*	GE Healthcare, USA
Analogue tube rollers: SRT6	Stuart Equipment, United Kingdom
Analytical balance: ABJ	Kern, Germany
Autoclave: VX-120	Systec, Germany
BioSAXS beamline B21	Diamond Light Source (Harwell, UK)
Chirascan CD spectrometer	Applied Photophysics, UK
Centrifuge: 5415 R	Eppendorf, Germany
Centrifuge: 5810 R	Eppendorf, Germany
Centrifuge: Avanti J-26SXP	Beckman & Coulter, USA
Centrifuge rotor: 70 Ti UZ	Beckman & Coulter, USA
Centrifuge rotor: JA 25.50	Beckman & Coulter, USA
Centrifuge rotor: JLA-8.1	Beckman & Coulter, USA
Eiger 4M detector	DECTRIS, Switzerland
Electrophoresis system: Mini Protetra Cell	Bio-Rad, USA
Electroporation system: MicroPulser	Bio-Rad, USA
CW EPR spectrometer (X-Band, Flexline MD-5)	Bruker Biospin, Rheinstetten, Germany
Pulsed EPR spectrometer (Q-Band, ELEXSYS E580, ER 5106QT-2 resonator)	Bruker Biospin, Rheinstetten, Germany
Gel documentation: Gel Doc 200 & ChemiDoc MP	Bio-Rad, USA
Gel tray for SDS-PAGE	Bio-Rad, USA
Heat block: MKR13	HLC, Germany
High-pressure homogeniser: Emulsiflex-C3	Avestin, Canada

Incubator: Ecotron	Infors HT, Germany
Incubator: Multitron Standard	Infors HT, Germany
Laminar airflow cabinet: BDK	Weiss Technik, Germany
Magnetic stirrer: Hei-Mix L	Heidolph Instruments, Germany
Mass spectrometer MALDI-TOF	Bruker UltrafleXtreme, Germany
Micropipette puller: P-1000 equipped with a squared box filament (2.5 x 2.5 mm)	Sutter Instruments, USA
Microwave	Bosch (via Carl Roth), Germany
PCR instrument: Thermocycler peqSTAR2x Gradient	Peqlab (via VWR International), USA
pH-Meter: peqMeter 1,14	Peqlab (via VWR International), USA
Pipetting aid: accu-jet pro	Brand, Germany
Platform shaker: Polymax 1040	Heidolph Instruments, Germany
Power supply unit: peqPower	Peqlab (via VWR International), USA
Precision balance: EG	Kern & Sohn, Germany
nanoDSF Prometheus NT.48	NanoTemper, Germany
Scanner: LiDE 110	Canon, Japan
Spectrophotometer: Nanodrop 2000	Thermo Fisher Scientific, USA
Spectrophotometer: SmartSpec Plus	Bio-Rad, USA
Superdex 200 Increase 3.2/300 column	GE Healthcare, USA
Superdex 200 Increase 10/300 column	GE Healthcare, USA
Ultracentrifuge: Optima XE-90	Beckman & Coulter, USA
Ultrasonic cell disruptor: Sonopuls	Bandelin, Germany
Vacuum pump: BVC control	Vacuubrand, Germany
Vortexer: Vortex-Genie 2	Scientific Industries, USA

Table 5-3: Chemicals (listed alphabetically)

Compound	CAS-No.	Supplier	GHS hazard	Hazard Statements	Precautionary Statements
2-Nitrobenzylbromid	3958-60-9	Th. Geyer	GHS05, GHS07	H314, H335	P280, P304+P340 P305+P351+P338 P310
2-Mercaptoethanol	60-24-2	Sigma-Aldrich	GHS06 GHS09	H302, H411, H315, H335, H311, H319	P280, P312, P302+350, P261, P273, P301+312, P305+351+338
4-Benzoyl-L-phenylalanine	104504-45-2	Bachem	-	-	-
5,5′-Dithiobis (2-nitrobenzoic acid)	69-78-3	Sigma-Aldrich	GHS07	H315, H319, H335	P261, P305+P351+P338
Acetic acid, 96 %	64-19-7	Carl Roth	GHS02, GHS05	H226, H314	P280, P305+351+338, P310
Acetone	67-64-1	Carl Roth	GHS02, GHS07	H225, H319, H336	P210, P280, P304+P340+P312, P305+P351+P338, P337+P313, P403+P235
Active carbon	7440-44-0	Carl Roth	-	-	-
Acrylamide 37%	79-06-1	Carl Roth	GHS06, GHS08	H301, H312, H315, H317, H319, H332, H340, H350, H361f, H372	P201, P280, P301+310, P305+351+338, P308+313
Agar-Agar	9002-18-0	Sigma-Aldrich	-	-	-
Agarose	9012-36-6	Sigma-Aldrich	-	-	-

Albumin Fraktion V BSA	90604-29-8	Carl Roth	-	-	-
Ammonium formate	540-69-2	Sigma Aldrich	GH202	H315, H319, H335	P261, P305+351+338
Ammonium acetate	631-61-8	Carl Roth	-	-	-
Ampicillin	69-52-3	Carl Roth	GHS 08	H334, H317	P280, P261, P302+352, P342+311
APS	7727-54-0	Carl Roth	GHS03 GHS07 GHS08	H272, H302, H315, H317, H319, H334, H335	P280, P305+351+338, P302+352, P304+341, P342+311
Arabinose	5328-37-0	Carl Roth	-	-	-
β-Mercaptoethanol	60-24-2	Sigma-Aldrich	GHS05, GHS06, GHS08, GHS09	H301 + H331-H310-H315-H317-H318-H373-H410	P260, P262, P273, P280, P301+P310+P330, P302+P352+P310, P305+P351+P338+ P310, P391, P403+P233
BamHI	81295-09-2	Sigma-Aldrich	-	-	-
Brilliant blue G 250	6104-58-1	Carl Roth	-	-	-
Bromphenol blue	115-39-9	Sigma-Aldrich	-	-	-
$CaCl_2$	10043-52-4	Sigma-Aldrich	GHS07	H319	P305+P351+P338
Chymotrypsin	9004-07-3	Carl Roth			
Chloramphenicol	56-75-7	Sigma-Aldrich	GHS08	H351	P280
Citric acid	77-92-9	Carl Roth	GHS05	H318	P305+P351+P338, P311
Coomassie Brilliant Blue R250	6104-59-2	Sigma-Aldrich	-	-	-
Copper (II) sulphate	7758-98-7	Sigma-Aldrich	GHS05, GHS09	H302-H315-H319-H410	P264, P273, P280, P337+P313, P391, P501
D_2O	7789-20-0	Sigma-Aldrich	-	-	-

Di-Potassium hydrogen phosphate	16788-57-1	Carl Roth	-	-	-
Dimethylformamid	68-12-2	Carl Roth	GHS02, GHS07, GHS08	H226-H312+H332-H319-H360D P	P201-P210-P261-P280-P308+P313-P370+P378
Dnase I	-	Sigma-Aldrich	-	-	-
DTT	578517	Carl Roth	GHS07	H302, H315, H319, H335	P302+352, P305+351+338
EDTA	60-00-4	Carl Roth	GHS07	H319	P305+351+338
Ethanol	64-17-5	Carl Roth	GHS02	H225	P210
Formaldehyde	50-00-0	Carl Roth	GHS02, GHS05, GHS06, GHS08	H226-H301+H311+H331-H314-H317-H335-H341-H350-H370	P201-P210-P260-P280-P301+P310+P330-P303+P361+P353-P304+P340+P310-P305+P351+P338+P310-P308+P311-P370+P378-P403+P233
Fos-Choline-10	70504-28-8	Anatrace	-	-	-
Fos-Choline-12	29557-51-5	Anatrace	-	-	-
Glycerol	56-81-5	Carl Roth	-	-	-
Hydrochloric acid, 6 N	7647-01-0	Carl Roth	GHS05 GHS07	H290 H314, H335	P260-P280-P303+P361 +P353-P304+P340+P310-P305+P351+P338
HEPES	7365-45-9	Carl Roth	-	-	-
Protino® Ni-NTA Agarose	64-17-5d	Macherey-Nagel	-	H226, H317	P261, P272, P280, P302+P352, P333+P313,
Imidazole	288-32-4	Carl Roth	GHS05 GHS06 GHS08	H301, H314, H361	P260, P281, P303+361+353, P301+330+331, P305+351+338, P308+313
IPTG	367-93-1	Carl Roth	-	-	-
Isopropanol	67-63-0	Carl Roth	GHS02 GHS07	H225, H319, H336	P210, P233, P305+351+338

Kanamycin sulfate	25389-94-0	Carl Roth	GHS08	H360	P201-P308+P313
LB-medium Lennox	-	Carl Roth	-	-	-
Lysozyme (*Gallus gallus)*	12650-88-3	Sigma-Aldrich	-	-	-
MES	4432-3-9	Carl Roth	-	-	-
Methanol	67-56-1	Sigma-Aldrich	GHS02, GHS06, GHS08	H225-H301+H311+H331-H370	P210-P280-P302+P352+P312-P304+P340+P312-P370+P378-P403+P235
$MgCl_2$	7786-30-3	Carl Roth	-	-	-
Milk powder	68514-61-4	Carl Roth	-	-	-
n-decyl-β-D-maltopyranoside (DM)	82494-09-5	Anatrace	-	-	-
n-Dodecyl-N,N-Dimethylamine-N-Oxide (LDAO)	1643-20-5	Anatrace		H315, H319	P280, P302+P352, P305+P351+P338
n-dodecyl-β-D-maltopyranoside (DDM)	69227-93-6	Glycon Biochemicals	-	H303, H313, H333	P261, P280, P302+P352
n-nonyl-β-D-glucopyranoside (NG)	106402-05-5	Anatrace	-	-	-
NaOAc	127-09-3	Sigma-Aldrich	GHS07	H319	P305+P351+P338
NaCl	7647-14-5	Carl Roth	-	-	-
NaOH	1310-73-2	Carl Roth	GHS05	H314	P280, P310, P305+351+338
Ni-NTA-Agarose		Sigma-Aldrich	GHS02 GHS07 GHS08	H226-H317-H350i-H360D-H373-H412	P201-P273-P280-P308+P313-P333+P313-P370+P378
$Ni(II)SO_4$	10101-97-0	Carl Roth	GHS07 GHS08 GHS09	H302+332, H315, H317, H334, H341, H350i, H360d,	P201, P261, P273, P280, P284, P304+340+312

				H372 H410	
Potassium Carbonate	584-08-7	Sigma-Aldrich	GHS07	H315, H319, H335	P302+P352, P305+P351+P338
Penicillin-Streptomycin	-	Gibco	GHS07 GHS07	H315, H317, H334, H335	P280, P261, P264, P284, P271, P302+P352, P333+P313, P304+P340, P342+P311, P312, P403+P233, P501
Potassium cyanide	151-50-8	Sigma-Aldrich	GHS05 GHS06 GHS08 GHS09	H290-H300+H310 +H330-H370-H371-H372-H410	P260, P262, P264, P273, P280, P284, P301+P310,P301+ P310 +P330, P302+P352+P310, P304+P340+P310, P391, P403+P233
Potassium di-hydrogen phosphate	7778-77-0	Carl Roth	-	-	-
Roti-Gel Stain	-	Carl Roth	-	-	-
S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate	81213-52-7	Toronto Research Chemicals	H315, H319, H335	-	P261, P280, P305, P351, P338
SDS	151-21-3	Carl Roth	GHS02 GHS06	H228, H302, H311, H315, H319, H335	P210, P261, P280, P312, P305+351+338
Sodium citrate	1545832	Sigma-Aldrich	-	-	-
TEMED	1185-53-1	Carl Roth	GHS07	H315, H319, H335	P261, P305+351+338
Tris hydrochloride	1185-53-1	Carl Roth	-	-	-
Triton X 100	9002-93-1	Carl Roth	GHS05 GHS07 GHS09	H302-H315-H318-H410	P280, P301+P312+P330, P305+P351+P338+ P310
Tryptone/Peptone	8952.4	Carl Roth	-	-	-
Trypsin	9002-07-7	Carl Roth	GHS07	H315-	P261, P280, P284,

			GHS08	H319- H334-H335	P304+P340, P337+P313, P342+P311
XhoI	81295-43-4	Sigma-Aldrich	-	-	-
Yeast Extract	8013-01-2	Carl Roth	-	-	-

5.3 GHS and risks symbols

Figure Appendix 4.4: GHS pictograms (from https://www.ohsa.gov)

5.4 GHS hazards statements

H225 Highly flammable liquid and vapour
H226 Flammable liquid and vapour
H228 Flammable solid
H272 May intensify fire; oxidizer
H290 May be corrosive to metals
H301 Toxic if swallowed
H302 Harmful if swallowed
H303 May be harmful if swallowed
H311 Toxic in contact with skin
H312 Harmful in contact with skin
H313 May be harmful in contact with skin
H314 Causes severe skin burns and eye damage
H315 Causes skin irritation
H316 Causes mild skin irritation
H317 May cause an allergic skin reaction
H318 Causes serious eye damage
H319 Causes serious eye irritation
H330 Fatal if inhaled
H331 Toxic if inhaled
H332 Harmful if inhaled
H333 May be harmful if inhaled
H334 May cause allergy or asthma symptoms or breathing difficulties if inhaled
H335 May cause respiratory irritation
H336 May cause drowsiness or dizziness
H340 May cause genetic defects
H341 Suspected of causing genetic defects

H350	May cause cancer
H350i	May cause cancer by inhalation
H360	May damage fertility or the unborn child
H360D	May damage the unborn child
H360F	May damage fertility
H360Fd	May damage fertility. Suspected of damaging the unborn child
H360FD	May damage fertility. May damage the unborn child
H361	Suspected of damaging fertility or the unborn child
H361d	Suspected of damaging the unborn child.
H361f	Suspected of damaging fertility
H370	Cause damage to organs
H372	Causes damage to organs through prolonged or repeated exposure
H373	May cause damage to organs through prolonged or repeated exposure.
H400	Very toxic to aquatic life with long-lasting effects
H410	Very toxic to aquatic life with long lasting effects
H411	Toxic to aquatic life with long lasting effects
H412	Harmful to aquatic life with long lasting effects.
EUH032	Contact with acids liberates very toxic gas
EUH066	Repeated exposure may cause skin dryness or cracking

5.5 GHS precautionary statements

P101	If medical advice is needed, have product container or label at hand
P201	Obtain special instructions before use
P210	Keep away from heat/sparks/open flames/hot surfaces – No smoking
P233	Keep container tightly closed
P260	Do not breathe dust/fume/gas/mist/vapours/spray
P261	Avoid breathing dust/fume/gas/mist/vapours/spray
P264	Wash thoroughly after handling
P270	Do not eat, drink or smoke when using this product
P273	Avoid release to the environment
P281	Use personal protective equipment as required
P280	Wear protective gloves/protective clothing/eye protection/face protection
P284	Wear respiratory protection
P309	IF exposed or you feel unwell
P310	Immediately call a POISON CENTER or doctor/physician
P311	Call a POISON CENTER or doctor/physician
P312	Call a POISON CENTER or doctor/physician if you feel unwell
P321	Specific treatment (see respective MSDS)
P330	Rinse mouth
P362	Take off contaminated clothing and wash before reuse
P391	Collect spillage
P405	Store locked up
P501	Dispose of contents/container in accordance with local/regional/national/international regulations
P301+P310	IF SWALLOWED: Immediately call a POISON CENTER or doctor/physician
P301+P312	IF SWALLOWED: Call a POISON CENTER or doctor/physician if you feel unwell
P301+P330+P331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting
P302+P352	IF ON SKIN: Wash with soap and water
P303+P361+P353	IF ON SKIN (or hair): Remove/Take off immediately all contaminated clothing. Rinse skin with water/shower
P304+P341	IF INHALED: lf breathing is difficult, remove victim to fresh air and keep at rest in a position comfortable for breathing

P305+P351+P338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do - continue rinsing
P308+P313	IF exposed or concerned: Get medical advice/attention
P309+P311	IF exposed or you feel unwell: Call a POISON CENTER or doctor/physician
P332+P313	lf skin irritation occurs: Get medical advice/attention
P333+P313	If skin irritation or rash occurs: Get medical advice/attention
P337+P313	If eye irritation persists: Get medical advice/attention
P342+P311	Call a POISON CENTER or doctor/physician
P370+P378	In case of fire: Use for extinction: Alcohol resistant foam
P370+P378	In case of fire: Use for extinction: Fire-extinguishing powder
P370+P378	In case of fire: Use for extinction: Carbon dioxide
P403+P233	Store in a well-ventilated place. Keep container tightly closed
P403+P235	Store in a well-ventilated place. Keep cool

Disposal

All chemicals used during the experimental phase were handled and disposed according to their H- and P-Sentences. Solvents and contaminated waste was stored in the specific boxes and disposed according to the safety instructions. Genetically modified organisms and related waste were autoclaved according to the "Gentechnikgesetz" before disposal for 20 min at 121°C and 5 bar.

6 References

1. Y. Cao, *Advances in Membrane Proteins*, Y. Cao, Ed. (Springer Singapore, 2018).
2. J. Liu, B. Rost, Comparing function and structure between entire proteomes. *Protein Sci.* **10**, 1970–1979 (2001).
3. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
4. K. Liszewski, Dissecting the Structure of Membrane Proteins. *Genet. Eng. Biotechnol. News* **35**, 1, 14, 16–17 (2015).
5. J. P. Overington, B. Al-Lazikani, A. L. Hopkins, How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
6. R. Wardhan, P. Mudgal, *Textbook of Membrane Biology* (Springer Singapore, 2017).
7. J. W. Fairman, N. Noinaj, S. K. Buchanan, The structural biology of β-barrel membrane proteins: a summary of recent reports. *Curr. Opin. Struct. Biol.* **21**, 523–531 (2011).
8. A. I. C. Höhr, *et al.*, Membrane protein insertion through a mitochondrial β-barrel gate. *Science.* **359**, eaah6834 (2018).
9. W. C. Wimley, The versatile β-barrel membrane protein. *Curr. Opin. Struct. Biol.* **13**, 404–411 (2003).
10. G. Von Heijne, Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.* **7**, 909–918 (2006).
11. H. Hong, Toward understanding driving forces in membrane protein folding. *Arch. Biochem. Biophys.* **564**, 297–313 (2014).
12. J. U. Bowie, Solving the membrane protein folding problem. *Nature* **438**, 581–589 (2005).
13. K. G. Fleming, Energetics of Membrane Protein Folding. *Annu. Rev. Biophys.* **43**, 233–255 (2014).
14. O. Gursky, *Lipids in Protein Misfolding*, O. Gursky, Ed. (Springer International Publishing, 2015).
15. F. X. Contreras, A. M. Ernst, F. Wieland, B. Brügger, Specificity of intramembrane protein-lipid interactions. *Cold Spring Harb. Perspect. Biol.* **3**, 1–18 (2011).
16. R. Macháň, M. Hof, Lipid diffusion in planar membranes investigated by fluorescence

correlation spectroscopy. *Biochim. Biophys. Acta - Biomembr.* **1798**, 1377–1391 (2010).

17. G. Lindblom, "Lipid Lateral Diffusion" in *Encyclopedia of Biophysics*, (Springer Berlin Heidelberg, 2013). 1264–1268.
18. A. G. Lee, Biological membranes: The importance of molecular detail. *Trends Biochem. Sci.* **36**, 493–500 (2011).
19. D. Marsh, Electron spin resonance in membrane research: Protein-lipid interactions from challenging beginnings to state of the art. *Eur. Biophys. J.* **39**, 513–525 (2010).
20. K. Takeda, *et al.*, A novel three-dimensional crystal of bacteriorhodopsin obtained by successive fusion of the vesicular assemblies. *J. Mol. Biol.* **283**, 463–474 (1998).
21. H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J. K. Lanyi, Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
22. C. Bechara, *et al.*, A subset of annular lipids is linked to the flippase activity of an ABC transporter. *Nat. Chem.* **7**, 255–262 (2015).
23. M. A. Hanson, *et al.*, A Specific Cholesterol Binding Site Is Established by the 2.8 Å Structure of the Human β2-Adrenergic Receptor. *Structure* **16**, 897–905 (2008).
24. K. Shinzawa-Itoh, *et al.*, Structures and physiological roles of 13 integral lipids of bovine heart cytochrome c oxidase. *EMBO J.* **26**, 1713–1725 (2007).
25. I. Mus-Veteau, *Membrane Proteins Production for Structural Analysis*, I. Mus-Veteau, Ed. (Springer, 2014).
26. B. Miroux, J. E. Walker, Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **260**, 289–298 (1996).
27. I. Moraes, G. Evans, J. Sanchez-Weatherby, S. Newstead, P. D. Stewart, *The Next Generation in Membrane Protein Structure Determination*, I. Moraes, Ed. (Springer International Publishing, 2016).
28. A. Helenius, D. R. McCaslin, E. Fries, C. Tanford, "Properties of detergents" in *Methods in Enzymology*, (1979). 734–749.
29. J. M. Neugebauer, "Detergents: An overview" in *Methods in Enzymology*, (1990). 239–253.
30. D. Linke, "Detergents. An Overview" in *Methods in Enzymology*, 1st Ed., (Elsevier Inc., 2009). 603–617.
31. H. Ilgü, *et al.*, Variation of the Detergent-Binding Capacity and Phospholipid Content of Membrane Proteins When Purified in Different Detergents. *Biophys. J.* **106**, 1660–1670

(2014).

32. M. Zoonens, F. Zito, K. L. Martinez, J.-L. Popot, “Amphipols: A General Introduction and Some Protocols” in *Membrane Proteins Production for Structural Analysis*, (Springer New York, 2014). 173–203.
33. J.-L. Popot, *et al.*, Amphipols From A to Z. *Annu. Rev. Biophys.* **40**, 379–408 (2011).
34. Y. Gohon, *et al.*, Partial specific volume and solvent interactions of amphipol A8-35. *Anal. Biochem.* **334**, 318–334 (2004).
35. T. Dahmane, M. Damian, S. Mary, J.-L. Popot, J.-L. Banères, Amphipol-Assisted *in vitro* Folding of G Protein-Coupled Receptors. *Biochemistry* **48**, 6516–6521 (2009).
36. K. L. Martinez, *et al.*, Allosteric transitions of *Torpedo* acetylcholine receptor in lipids, detergent and amphipols: molecular interactions vs. physical constraints. *FEBS Lett.* **528**, 251–256 (2002).
37. M. Picard, *et al.*, Protective and Inhibitory Effects of Various Types of Amphipols on the Ca^{2+}-ATPase from Sarcoplasmic Reticulum: A Comparative Study. *Biochemistry* **45**, 1861–1869 (2006).
38. M. A. Schuler, I. G. Denisov, S. G. Sligar, Nanodiscs as a new tool to examine lipid-protein interactions. *Methods Mol. Biol.* **974**, 415–433 (2013).
39. J. A. Lyons, A. Bøggild, P. Nissen, J. Frauenfeld, “Saposin-Lipoprotein Scaffolds for Structure Determination of Membrane Transporters” in *Methods in Enzymology*, 1st Ed., (Elsevier Inc., 2017). 85–99.
40. J. Frauenfeld, *et al.*, A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* **13**, 345–351 (2016).
41. A. O. Oluwole, *et al.*, Formation of Lipid-Bilayer Nanodiscs by Diisobutylene/Maleic Acid (DIBMA) Copolymer. *Langmuir* **33**, 14378–14388 (2017).
42. M. Jamshad, *et al.*, Structural analysis of a nanoparticle containing a lipid bilayer used for detergent-free extraction of membrane proteins. *Nano Res.* **8**, 774–789 (2015).
43. T. Schwann, Ueber das wesen des Verdauungsprocesses. *Ann Pharm Chem* **20**, 28–34 (1836).
44. M. S. Brown, J. Ye, R. B. Rawson, J. L. Goldstein, Regulated Intramembrane Proteolysis. *Cell* **100**, 391–398 (2000).
45. R. B. Rawson, *et al.*, Complementation Cloning of S2P, a Gene Encoding a Putative Metalloprotease Required for Intramembrane Cleavage of SREBPs. *Mol. Cell* **1**, 47–57 (1997).

46. E. Erez, D. Fass, E. Bibi, How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**, 371–378 (2009).
47. D. Langosch, C. Scharnagl, H. Steiner, M. K. Lemberg, Understanding intramembrane proteolysis: from protein dynamics to reaction kinetics. *Trends Biochem. Sci.* **40**, 318–327 (2015).
48. S. Urban, Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr. Opin. Genet. Dev.* **12**, 512–518 (2002).
49. S. W. Dickey, R. P. Baker, S. Cho, S. S. Urban, Proteolysis inside the membrane is a rate-governed reaction not driven by Substrate Affinity. *Cell* **155**, 1270–1281 (2013).
50. F. Kamp, *et al.*, Intramembrane Proteolysis of β-Amyloid Precursor Protein by γ-Secretase Is an Unusually Slow Process. *Biophys. J.* **108**, 1229–1237 (2015).
51. E. Arutyunova, *et al.*, Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* **33**, 1–13 (2014).
52. M. Freeman, The Rhomboid-Like Superfamily: Molecular Mechanisms and Biological Roles. *Annu. Rev. Cell Dev. Biol.* **30**, 235–254 (2014).
53. S. Urban, S. W. Dickey, The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol.* **12**, 231 (2011).
54. A. Guichard, *et al.*, rhomboid and Star interact synergistically to promote EGFR/MAPK signaling during Drosophila wing vein development. *Development* **126**, 2663–76 (1999).
55. J. D. Wasserman, S. Urban, M. Freeman, A family of rhomboid-like genes: *Drosophila* rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling. *Genes Dev.* **14**, 1651–63 (2000).
56. J. R. Lee, S. Urban, C. F. Garvey, M. Freeman, Regulated Intracellular Ligand Transport and Proteolysis Control EGF Signal Activation in *Drosophila*. *Cell* **107**, 161–171 (2001).
57. S. Urban, J. R. Lee, M. Freeman, *Drosophila* Rhomboid-1 Defines a Family of Putative Intramembrane Serine Proteases. *Cell* **107**, 173–182 (2001).
58. L. G. Stevenson, *et al.*, Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1003–8 (2007).
59. T. J. Dowse, K. Koussis, M. J. Blackman, D. Soldati-Favre, Roles of proteases during invasion and egress by *Plasmodium* and *Toxoplasma*. *Subcell. Biochem.* **47**, 121–39 (2008).

60. O. Lohi, S. Urban, M. Freeman, Diverse Substrate Recognition Mechanisms for Rhomboids: Thrombomodulin Is Cleaved by Mammalian Rhomboids. *Curr. Biol.* **14**, 236–241 (2004).
61. L. Fleig, *et al.*, Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins. *Mol. Cell* **47**, 558–569 (2012).
62. M. Hulko, A. N. Lupas, J. Martin, Inherent chaperone-like activity of aspartic proteases reveals a distant evolutionary relation to double-psi barrel domains of AAA-ATPases. *Protein Sci.* **16**, 644–53 (2007).
63. C. Spiess, A. Beil, M. Ehrmann, A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339–47 (1999).
64. Y. Wang, Y. Zhang, Y. Ha, Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179–183 (2006).
65. Y. Wang, Y. Ha, Open-cap conformation of intramembrane protease GlpG. *Proc. Natl. Acad. Sci.* **104**, 2098–2102 (2007).
66. A. R. Sherratt, D. R. Blais, H. Ghasriani, J. P. Pezacki, N. K. Goto, Activity-Based Protein Profiling of the *Escherichia coli* GlpG Rhomboid Protein Delineates the Catalytic Core. *Biochemistry* **51**, 7794–7803 (2012).
67. Z. Wu, *et al.*, Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **13**, 1084–1091 (2006).
68. A. Ben-Shem, D. Fass, E. Bibi, Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl. Acad. Sci.* **104**, 462–466 (2007).
69. M. J. Lemieux, S. J. Fischer, M. M. Cherney, K. S. Bateman, M. N. G. James, The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc. Natl. Acad. Sci.* **104**, 750–754 (2007).
70. K. R. Vinothkumar, Structure of Rhomboid Protease in a Lipid Environment. *J. Mol. Biol.* **407**, 232–247 (2011).
71. K. R. R. Vinothkumar, O. A. a. Pierrat, J. M. M. Large, M. Freeman, Structure of Rhomboid Protease in Complex with β-Lactam Inhibitors Defines the S2′ Cavity. *Structure* **21**, 1051–1058 (2013).
72. E. V Koonin, *et al.*, The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* **4**, R19 (2003).
73. A. J. B. Kreutzberger, M. Ji, J. Aaron, L. Mihaljević, S. Urban, Rhomboid distorts lipids

to break the viscosity-imposed speed limit of membrane diffusion. *Science.* **363**, eaao0076 (2019).

74. E. Arutyunova, *et al.*, Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* **33**, 1869–1881 (2014).
75. C. Lazareno-Saez, E. Arutyunova, N. Coquelle, M. J. Lemieux, Domain Swapping in the Cytoplasmic Domain of the *Escherichia coli* Rhomboid Protease. *J. Mol. Biol.* **425**, 1127–1142 (2013).
76. A. C. Y. Foo, B. G. R. Harvey, J. J. Metz, N. K. Goto, Influence of hydrophobic mismatch on the catalytic activity of *Escherichia coli* GlpG rhomboid protease. *Protein Sci.* **24**, 464–473 (2015).
77. P. Sampathkumar, *et al.*, Oligomeric state study of prokaryotic rhomboid proteases. *Biochim. Biophys. Acta - Biomembr.* **1818**, 3090–3097 (2012).
78. A. J. B. Kreutzberger, S. Urban, Single-Molecule Analyses Reveal Rhomboid Proteins Are Strict and Functional Monomers in the Membrane. *Biophys. J.* **115**, 1755–1761 (2018).
79. E. Reading, *et al.*, Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angew. Chemie Int. Ed.* **56**, 15654–15657 (2017).
80. P. K. Madala, J. D. A. Tyndall, T. Nall, D. P. Fairlie, Update 1 of: Proteases Universally Recognize Beta Strands In Their Active Sites. *Chem. Rev.* **110**, PR1–PR31 (2010).
81. Y. Wang, S. Maegawa, Y. Akiyama, Y. Ha, The Role of L1 Loop in the Mechanism of Rhomboid Intramembrane Protease GlpG. *J. Mol. Biol.* **374**, 1104–1113 (2007).
82. Y. Akiyama, S. Maegawa, Sequence features of substrates required for cleavage by GlpG, an *Escherichia coli* rhomboid protease. *Mol. Microbiol.* **64**, 1028–1037 (2007).
83. S. Cho, S. W. Dickey, S. Urban, Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis. *Mol. Cell* **61**, 329–340 (2016).
84. S. Zoll, *et al.*, Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate – peptide complex structures. **33**, 2408–2421 (2014).
85. K. Strisovsky, H. J. Sharpe, M. Freeman, Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Mol. Cell* **36**, 1048–1059 (2009).
86. A. Fersht, *Structure and Mechanism in Protein Science* (WORLD SCIENTIFIC, 2017).
87. Y. Ha, Y. Akiyama, Y. Xue, Structure and Mechanism of Rhomboid Protease. *J. Biol.*

Chem. **288**, 15430–15436 (2013).

88. L. Hedstrom, Serine Protease Mechanism and Specificity. *Chem. Rev.* **102**, 4501–4524 (2002).
89. N. Uritsky, M. Shokhen, A. Albeck, The Catalytic Machinery of Rhomboid Proteases: Combined MD and QM Simulations. *J. Chem. Theory Comput.* **8**, 4663–4671 (2012).
90. N. Uritsky, M. Shokhen, A. Albeck, Stepwise Versus Concerted Mechanisms in General-Base Catalysis by Serine Proteases. *Angew. Chemie - Int. Ed.* **55**, 1680–1684 (2016).
91. R. Henderson, Structure of crystalline α-chymotrypsin. IV. The structure of indoleacryloyl-α-chymotrypsin and its relevance to the hydrolytic mechanism of the enzyme. *J. Mol. Biol.* **54**, 341–354 (1970).
92. Y. Zhou, S. M. Moin, S. Urban, Y. Zhang, An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* **20**, 1255–1263 (2012).
93. C. L. Brooks, M. J. Lemieux, Untangling structure–function relationships in the rhomboid family of intramembrane proteases. *Biochim. Biophys. Acta - Biomembr.* **1828**, 2862–2872 (2013).
94. S. Urban, S. M. Moin, A Subset of Membrane-Altering Agents and γ-Secretase Modulators Provoke Nonsubstrate Cleavage by Rhomboid Proteases. *Cell Rep.* **8**, 1241–1247 (2014).
95. S. M. Moin, S. Urban, Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife* **1**, e00173 (2012).
96. A. B. T. Ghisaidoobe, S. J. Chung, Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on förster resonance energy transfer techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).
97. J. R. Lakowicz, Ed., *Principles of Fluorescence Spectroscopy* (Springer US, 2006).
98. E. Bordignon, H.-J. Steinhoff, "Membrane Protein Structure and Dynamics Studied by Site-Directed Spin-Labeling ESR" in *ESR Spectroscopy in Membrane Biophysics*, (Springer US, 2009), pp. 129–164.
99. J. P. Klare, Site-directed spin labeling EPR spectroscopy in protein research. *Biol. Chem.* **394**, 1281–1300 (2013).
100. L. J. Berliner, J. Grunwald, H. O. Hankovszky, K. Hideg, A novel reversible thiol-specific spin label: Papain active site labeling and inhibition. *Anal. Biochem.* **119**, 450–

455 (1982).

101. G. Jeschke, MMM: A toolbox for integrative structure modeling. *Protein Sci.* **27**, 76–85 (2018).
102. C. M. Jeffries, *et al.*, Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **11**, 2122–2153 (2016).
103. J. Trewhella, Small-angle scattering and 3D structure interpretation. *Curr. Opin. Struct. Biol.* **40**, 1–7 (2016).
104. T. M. Ryan, *et al.*, An optimized SEC-SAXS system enabling high X-ray dose for rapid SAXS assessment with correlated UV measurements for biomolecular structure analysis: *J. Appl. Crystallogr.* **51**, 97–111 (2018).
105. A. Panjkovich, D. I. Svergun, CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
106. D. Franke, *et al.*, *ATSAS 2.8*: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1151–1158 (2017).
107. D. Franke, D. I. Svergun, *DAMMIF*, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
108. V. V. Volkov, D. I. Svergun, Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
109. L. a Kelley, M. J. E. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
110. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
111. L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).
112. A. K. Malde, *et al.*, An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **7**, 4026–4037 (2011).
113. K. B. Koziara, M. Stroet, A. K. Malde, A. E. Mark, Testing and validation of the Automated Topology Builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput. Aided. Mol. Des.* **28**, 221–233 (2014).
114. D. H. de Jong, *et al.*, Improved Parameters for the Martini Coarse-Grained Protein Force

Field. *J. Chem. Theory Comput.* **9**, 687–697 (2013).

115. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, a DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
116. M. J. Abraham, *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).
117. K. M. Ravikumar, W. Huang, S. Yang, Coarse-grained simulations of protein-protein association: An energy landscape perspective. *Biophys. J.* **103**, 837–845 (2012).
118. M. Parrinello, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).
119. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 1–7 (2007).
120. T. a. Wassenaar, K. Pluhackova, R. a. Böckmann, S. J. Marrink, D. P. Tieleman, Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **10**, 676–690 (2014).
121. D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).
122. E. L. Lawler, D. E. Wood, Branch-and-Bound Methods: A Survey. *Oper. Res.* **14**, 699–719 (1966).
123. B. Raveh, N. London, O. Schueler-Furman, Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins Struct. Funct. Bioinforma.* **78**, 2029–2040 (2010).
124. J. Cortes, *et al.*, A path planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics* **21**, i116–i125 (2005).
125. N. M. Amato, G. Song, Using Motion Planning to Study Protein Folding Pathways. *J. Comput. Biol.* **9**, 149–168 (2002).
126. S. M. LaValle, J. J. Kuffner, Rapidly-Exploring Random Trees: Progress and Prospects in *Algorithmic and Computational Robotics: New Directions*, (2000). 293–308.
127. D. Schneidman-Duhovny, M. Hammel, "Modeling Structure and Dynamics of Protein Complexes with SAXS Profiles" in *Protein Complex Assembly: Methods and Protocols*, (2018). 449–473.
128. I. M. S. de Vera, M. E. Blackburn, L. Galiano, G. E. Fanucci, "Pulsed EPR Distance

Measurements in Soluble Proteins by Site-Directed Spin Labeling (SDSL)" in *Current Protocols in Protein Science*, (John Wiley & Sons, Inc., 2013). 17.17.1-17.17.29.

129. I. D. Sahu, *et al.*, DEER EPR measurements for membrane protein structures via bifunctional spin labels and lipodisq nanoparticles. *Biochemistry* **52**, 6627–6632 (2013).
130. E. Bordignon, "Site-Directed Spin Labeling of Membrane Proteins" in *Topics in Current Chemistry*, (2012). 121–158.
131. G. Jeschke, The contribution of modern EPR to structural biology. *Emerg. Top. Life Sci.* **2**, 9–18 (2018).
132. G. Jeschke, DEER Distance Measurements on Proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446 (2012).
133. R. E. Martin, *et al.*, Determination of End-to-End Distances in a Series of TEMPO Diradicals of up to 2.8 nm Length with a New Four-Pulse Double Electron Electron Resonance Experiment. *Angew. Chemie Int. Ed.* **37**, 2833–2837 (1998).
134. M. Pannier, S. Veit, A. Godt, G. Jeschke, H. . Spiess, Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins. *J. Magn. Reson.* **142**, 331–340 (2000).
135. G. Jeschke, *et al.*, DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30**, 473–498 (2006).
136. M. Vendruscolo, C. M. Dobson, Protein Dynamics: Moore's Law in Molecular Biology. *Curr. Biol.* **21**, R68–R70 (2011).
137. S. C. L. Kamerlin, S. Vicatos, A. Dryga, A. Warshel, Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annu. Rev. Phys. Chem.* **62**, 41–64 (2011).
138. H. I. Ingólfsson, *et al.*, The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 225–248 (2014).
139. S. Kmiecik, *et al.*, Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936 (2016).
140. S. J. Marrink, A. H. de Vries, A. E. Mark, Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **108**, 750–760 (2004).
141. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
142. L. Monticelli, *et al.*, The MARTINI Coarse-Grained Force Field: Extension to Proteins.

J. Chem. Theory Comput. **4**, 819–834 (2008).

143. X. Periole, S.-J. Marrink, "The Martini Coarse-Grained Force Field" in *Essentials of Micro- and Nanofluidics*, (2013), pp. 533–565.
144. R. Baron, *et al.*, Comparison of Thermodynamic Properties of Coarse-Grained and Atomic-Level Simulation Models. *ChemPhysChem* **8**, 452–461 (2007).
145. S. Ramadurai, *et al.*, Influence of Hydrophobic Mismatch and Amino Acid Composition on the Lateral Diffusion of Transmembrane Peptides. *Biophys. J.* **99**, 1447–1454 (2010).
146. X. Periole, T. Huber, S.-J. Marrink, T. P. Sakmar, G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **129**, 10126–10132 (2007).
147. T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, S. J. Marrink, Computational Lipidomics with *insane*: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).
148. S.-J. Marrink, MARTINI: Coarse Grain Forcefield for Biomolecules. http:// cgmartini.nl/index.php/force-field-parameters/lipids/ (accessed August 8, 2017)
149. X. Periole, M. Cavalli, S.-J. J. Marrink, M. a. Ceruso, Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).
150. T. D. Romo, N. Leioatts, A. Grossfield, Lightweight object oriented structure analysis: tools for building tools to analyze molecular dynamics simulations. *J. Comput. Chem.* **35**, 2305–2318 (2014).
151. K. Younge, *et al.*, The use of radial distribution and pair-correlation functions to analyze and describe biological aggregations. *Limnol. Oceanogr. Methods* **4**, 382–391 (2006).
152. E. F. Pettersen, *et al.*, UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
153. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
154. A. Flayhan, *et al.*, Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **26**, 345-355.e5 (2018).
155. D. I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
156. N. Skar-Gislinge, L. Arleth, Small-angle scattering from phospholipid nanodiscs:

derivation and refinement of a molecular constrained analytical model form factor. *Phys. Chem. Chem. Phys.* **13**, 3161–3170 (2011).

157. A. Y. Shih, I. G. Denisov, J. C. Phillips, S. G. Sligar, K. Schulten, Molecular Dynamics Simulations of Discoidal Bilayers Assembled from Truncated Human Lipoproteins. *Biophys. J.* **88**, 548–556 (2005).
158. R. C. Oliver, *et al.*, Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS One* **8** (2013).
159. B. R. Jefferys, L. A. Kelley, M. J. E. Sternberg, Protein Folding Requires Crowd Control in a Simulated Cell. *J. Mol. Biol.* **397**, 1329–1338 (2010).
160. J. Lipfert, L. Columbus, V. B. Chu, S. A. Lesley, S. Doniach, Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. *J. Phys. Chem. B* **111**, 12427–12438 (2007).
161. J. P. Klare, H. J. Steinhoff, Spin Labeling Studies of Transmembrane Signaling and Transport: Applications to Phototaxis, ABC Transporters and Symporters. *Methods Enzymol.* **564**, 315–347 (2015).
162. Y. Xue, Y. Ha, Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.* **287**, 3099–3107 (2012).
163. S. M. Moin, S. Urban, Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife* **1**, e00173 (2012).
164. H. S. McHaourab, P. R. Steed, K. Kazmier, Toward the fourth dimension of membrane protein structure: Insight into dynamics from spin-labeling EPR spectroscopy. *Structure* **19**, 1549–1561 (2011).
165. I. Hänelt, The Synergetic Effects of Combining Structural Biology and EPR Spectroscopy on Membrane Proteins. *Crystals* **7**, 117 (2017).
166. B. E. Bode, *et al.*, Counting the Monomers in Nanometer-Sized Oligomers by Pulsed Electron−Electron Double Resonance. *J. Am. Chem. Soc.* **129**, 6736–6745 (2007).
167. Y. Polyhach, *et al.*, High sensitivity and versatility of the DEER experiment on nitroxide radical pairs at Q-band frequencies. *Phys. Chem. Chem. Phys.* **14**, 10762 (2012).
168. Y. W. Chiang, P. P. Borbat, J. H. Freed, The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *J. Magn. Reson.* **172**, 279–295 (2005).
169. J. P. Klare, H.-J. Steinhoff, "Site-directed Spin Labeling and Pulse Dipolar Electron Paramagnetic Resonance" in *Encyclopedia of Analytical Chemistry*, (John Wiley &

Sons, Ltd, 2010).

170. J. E. Banham, *et al.*, Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides. *J. Magn. Reson.* **191**, 202–218 (2008).

171. A. D. Milov, B. D. Naumov, Y. D. Tsvetkov, The effect of microwave pulse duration on the distance distribution function between spin labels obtained by PELDOR data analysis. *Appl. Magn. Reson.* **26**, 587–599 (2004).

172. R. P. Baker, S. Urban, Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating. *Nature* **523**, 101–105 (2015).

173. C. L. Cortesio, E. B. Lewellyn, D. G. Drubin, Control of lipid organization and actin assembly during clathrin-mediated endocytosis by the cytoplasmic tail of the rhomboid protein Rbd2. *Mol. Biol. Cell* **26**, 1509–22 (2015).

174. Y. Almeida-Hernandez, H. Tidow, Soluble Regions of GlpG Influence Protein-Lipid Interactions and Lipid Distribution. *J. Phys. Chem. B* **123**, 7852-7858 (2019).

175. V. Corradi, *et al.*, Emerging Diversity in Lipid–Protein Interactions. *Chem. Rev.* **119**, 5775–5848 (2019).

176. H. Koldsø, D. Shorthouse, J. Hélie, M. S. P. Sansom, Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers. *PLoS Comput. Biol.* **10** (2014).

177. K. A. Marino, D. Prada-Gracia, D. Provasi, M. Filizola, Impact of Lipid Composition and Receptor Conformation on the Spatio-temporal Organization of μ-Opioid Receptors in a Multi-component Plasma Membrane Model. *PLOS Comput. Biol.* **12**, e1005240 (2016).

178. A. Grossfield, S. E. Feller, M. C. Pitman, A role for direct interactions in the modulation of rhodopsin by omega-3 polyunsaturated lipids. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4888–93 (2006).

179. S.-L. Niu, D. C. Mitchell, B. J. Litman, Optimization of Receptor-G Protein Coupling by Bilayer Lipid Composition II. *J. Biol. Chem.* **276**, 42807–42811 (2001).

180. D. Oliver, *et al.*, Functional conversion between A-type and delayed rectifier K+ channels by membrane lipids. *Science* **304**, 265–70 (2004).

181. K. Strisovsky, Structural and mechanistic principles of intramembrane proteolysis–lessons from rhomboids. *FEBS J.* **280**, 1579–603 (2013).

182. S. Lee, *et al.*, How Do Short Chain Nonionic Detergents Destabilize G-Protein-Coupled

Receptors? *J. Am. Chem. Soc.* **138**, 15425–15433 (2016).

183. S. Urban, R. P. Baker, *In vivo* analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol. Chem.* **389**, 1107–1115 (2008).
184. R. P. Baker, K. Young, L. Feng, Y. Shi, S. Urban, Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl. Acad. Sci.* **104**, 8257–8262 (2007).
185. Y. Xue, Y. Ha, Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J. Biol. Chem.* **288**, 16645–16654 (2013).
186. A.-N. Bondar, Mechanisms by Which Lipids Influence Conformational Dynamics of the GlpG Intramembrane Protease. *J. Phys. Chem. B* **123**, 4159–4172 (2019).
187. E. C. Ihms, M. P. Foster, *MESMER*: minimal ensemble solutions to multiple experimental restraints. *Bioinformatics* **31**, 1951–1958 (2015).
188. G. Hagelueken, D. Abdullin, O. Schiemann, "*mtsslSuite*" in *Methods in Enzymology*, (2015), pp. 595–622.

Acknowledgments

I want to thank many people who have supported me along this journey. First I want to thank my supervisor Prof. Dr. Henning Tidow, to accept me in his lab and give me the opportunity to carry on my research with big support and independence. Also, many thanks to my lab colleagues, Dr. Inokentijs Josts, Dr. Katharina Jungnickel, Katharina Veith, Julius Nitsche, Stephanie Kesgin-Schäfer, Tina Bohstedt, Anne Creon, Ellen Gattkowski, Zeinab Eskandarian, Simon Sander and Dominique-Maurice Kehlenbeck for the help, the fruitful discussions, beamtimes, and the occasional beers.

Moreover, I want to thanks Prof. Dr. Kvido Strisovsky for the supply of the plasmid to express the substrate of my protein. Also, many thanks to Prof. Dr. Heinz-Jürgen Steinhoff and especially to Dr. Johann Klare, for all the help and great discussions about the DEER experiments. Also, to Prof. Dr. Johannes Kirchmair, Prof. Dr. Andrew Torda, and Prof. Dr. Thomas Hackl, for agreed to evaluate my thesis and disputation.

Special thanks to the Center of Ultrafast Imaging (CUI), the CUI Cluster Office, and the CUI Graduate School, for the support in every aspect of my doctoral training and for the great environment for research and collaborations.

This work is also the result of the dedication of my previous teachers, professors, and supervisors, from whom I have learned so much. To all of them, thanks.

Many thanks to all my friends, scattered around the world, who have become my family and have walked by my side in the best and the worst moments. Many thanks to Bárbara Ibarrola, wherever you are, without whose help I wouldn't have been able to get here. RIP.

The biggest thanks go to my family, especially to my mum and my sister, who have sacrificed and always supported unconditionally from the other side of the world.

To each and every one of you, Gracias, Thanks, Danke, Teşekkürler, سد پا سد گزارم !

Curriculum vitae

Education / Professional experience

Dec. 2014 – Nov. 2019 **Research Assistant / PhD student**
Hamburg Centre for Ultrafast Imaging (CUI) & Institute of Biochemistry and Molecular Biology. University of Hamburg, Germany
Project: "Integrative analysis of the structure and dynamics of rhomboid protease GlpG from *Escherichia coli*"
Supervisor: Prof. Dr. Henning Tidow

Apr. 2014 – Sep 2014 **Guest Research Fellow / Internship**
Institute of Physical Chemistry "Rocasolano" & Spanish Center of Biotechnology (CNB). Madrid, Spain
Project: "Crystallization of Fab fragment of monoclonal antibody hR3 (Nimotuzumab), in complex with domain III of Human Epidermal Growth Factor Receptor (EGFR)"
Supervisors: Dr. Armando Albert and Dr. José F. Rodríguez

Oct. 2011 – Jan 2012 **Guest Research Fellow / Internship**
Department of Pathology. Oslo University Hospital. Norway
Project: "Characterization of the endocytosis of anti-EGFR mono- and polyclonal antibodies in tumoral cells lines by confocal microscopy"
Supervisor: Prof. Dr. Inger Helene Madshus

Sep. 2009 – Oct. 2014 **Junior Researcher Fellow / MSc. student**
Center of Molecular Immunology (CIM). University of Havana, Cuba
Thesis: "Statistical and energetic patterns in molecular recognition of protein-ligand complexes"
Supervisor: Prof. Dr. Ernesto Moreno-Frías

Sep. 2004 – Jul. 2009 **Study of Biochemistry / Licentiate student (BSc.+2)**
Center of Molecular Immunology (CIM). University of Havana, Cuba
Thesis: "Development and optimization of a machine-learning method for prediction of class II T-cell epitopes"
Supervisor: Dr. Amaury Pupo-Meriño

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

Hamburg, 24.09.2019