

UNIVERSITÀ DEGLI STUDI DI TRIESTE
e
UNIVERSITÀ CA' FOSCARI DI VENEZIA

**XXXII CICLO DEL DOTTORATO DI RICERCA IN
_____CHIMICA_____**

**Characterization of two membrane proteins:
a P-IV ATPase from yeast and a bacterial
polymerase**

Settore scientifico-disciplinare: CHIM/03

**DOTTORANDA
CATERINA DEGANUTTI**

**SUPERVISORE DI TESI
PROF. RITA DE ZORZI**

ANNO ACCADEMICO 2018/2019



UNIVERSITÀ DEGLI STUDI DI TRIESTE
e
UNIVERSITÀ CA' FOSCARI DI VENEZIA

XXXII CICLO DEL DOTTORATO DI RICERCA IN
_____CHIMICA_____

**Characterization of two membrane proteins:
a P-IV ATPase from yeast and a bacterial
polymerase**

Settore scientifico-disciplinare: CHIM/03

DOTTORANDA
CATERINA DEGANUTTI

Caterina Degnutti

COORDINATORE
PROF. BARBARA MILANI

Barbara Milani

SUPERVISORE DI TESI
PROF. RITA DE ZORZI

Rita De Zorzi

ANNO ACCADEMICO 2018/2019

TABLE OF CONTENT

| | Page |
|---|----------|
| LIST OF ABBREVIATIONS | I |
| ABSTRACT | V |
| RIASSUNTO | VII |
| | |
| Chapter 1 | |
| The cellular membrane: a liquid barrier between outside and inside | 1 |
| 1.1 THE CELLULAR MEMBRANE | 1 |
| 1.1.1 The membrane characteristics | 1 |
| • Membrane fluidity | 1 |
| • Viscosity | 2 |
| • Continuity | 2 |
| • Cooperation | 2 |
| • Interleaflet coupling | 2 |
| • Asymmetry | 3 |
| • Self-assemblies | 3 |
| • Protein-islands | 3 |
| • Lateral pressure | 3 |
| 1.1.2 Bacterial cell envelope: a complex structure of membranes | 4 |
| • Gram-negative bacteria cell envelope | 4 |
| • Gram-positive bacteria cell envelope | 6 |
| • Corynebacterineae cell envelope | 6 |
| 1.1.3 Eukaryotic cell membranes | 7 |
| 1.2 LIPIDS: THE FATTY COMPONENTS OF THE CELLULAR MEMBRANE | 9 |
| 1.2.1 Classification of lipids present in cellular membranes | 9 |
| • Glycerophospholipids | 9 |
| • Sphingolipids | 10 |
| • Sterols | 10 |
| • Glycolipids | 11 |
| 1.2.2 Distribution of lipids presents in cellular membranes | 11 |

| | | |
|-------|--|----|
| 1.3 | MEMBRANE PROTEINS: THE FLESHY COMPONENTS OF CELLULAR MEMBRANES | 14 |
| 1.3.1 | Integral membrane proteins | 14 |
| | • β -barrel proteins: | 15 |
| | • α -helical proteins: | 15 |
| 1.3.2 | Peripheral membrane proteins | 16 |
| 1.3.3 | Challenges in membrane protein structural analysis | 17 |
| | • Protein expression | 17 |
| | • The detergent issue | 18 |
| | • Solving a membrane protein structure | 19 |
| 1.4 | PROTEIN AND LIPIDS WORK TOGETHER | 19 |
| | • Scramblases | 20 |
| | • P-IV atpases | 21 |
| | • ABC transporters: | 21 |
| | REFERENCES | 23 |

Chapter 2

Neo1: an integral membrane protein belonging to the family of P-IV

| | | |
|-------|---|----|
| | atpases | 30 |
| | INTRODUCTION | 31 |
| 2.1 | P-TYPE ATPASES | 31 |
| 2.1.1 | Structure of P-type ATPases | 32 |
| 2.1.2 | Catalytic cycle of P-type ATPases | 34 |
| 2.2 | P-IV ATPASES | 36 |
| 2.2.1 | Cryo-electron microscopy structures of P-IV ATPases | 40 |
| 2.2.2 | Human diseases connected to defects of p-iv atpases | 43 |
| | • Liver diseases | 44 |
| | • Cerebellar ataxia and mental retardation | 44 |
| | • Metabolic diseases | 44 |
| | • Cancer | 45 |
| | • Infertility | 45 |

| | |
|---|----|
| • Blood diseases | 45 |
| • Alzheimer's disease | 45 |
| 2.2.3 Yeast P-IV ATPases | 47 |
| • Drs2 | 48 |
| • Dnf1 | 48 |
| • Dnf2 | 49 |
| • Dnf3 | 49 |
| • Neo1 | 49 |
| RESULTS AND DISCUSSION | 51 |
| 2.3 NEO1 RECOMBINANT PROTEIN ANALYSIS | 51 |
| 2.3.1 Plasmid preparation for the overexpression of the protein Neo1 | 51 |
| 2.3.2 Overexpression conditions for the protein Neo1 | 52 |
| 2.3.3 Solubilization of the integral membrane protein Neo1 | 57 |
| 2.3.4 Purification protocol and analysis of protein stability for the P-IV ATPase Neo1 | 58 |
| 2.3.5 Circular Dichroism analysis | 62 |
| 2.3.6 Overexpression of mutants on conserved sequences involved in phosphorylation and dephosphorylation mechanisms | 64 |
| 2.3.7 Neo1 activity assay | 67 |
| 2.3.8 Neo1 crystallization trials | 70 |
| 2.3.9 Neo1 negative staining and TEM analysis | 71 |
| 2.3.10 Neo1 mutants | 73 |
| 2.3.11 Neo1 mutants on substrate selection capacity overexpression | 77 |
| 2.4 CONCLUSIONS AND FUTURE PERSPECTIVES | 79 |
| REFERENCES | 81 |
| Chapter 3 | |
| The prokaryotic integral membrane protein polymerase Wzy from <i>Pseudomonas aeruginosa</i> PAO1 | 88 |
| Introduction | 89 |
| 3.1 PSEUDOMONAS AERUGINOSA | 89 |

| | |
|---|-----|
| • Chronic lung infections in cystic fibrosis patients | 90 |
| • Bacteremia in severe burn patients | 91 |
| • Ulcerative keratitis associated to soft contact lenses use | 91 |
| 3.1.1 Pseudomonas aeruginosa antibiotic resistance | 92 |
| 3.2 THE BACTERIAL LIPOPOLYSACCHARIDE LPS IS FORMED OF THREE DIFFERENT UNITS: LIPID A, CORE AND O-ANTIGEN | 94 |
| 3.2.1 Lipid A | 95 |
| 3.2.2 Core oligosaccharide | 97 |
| 3.2.3 O antigen | 99 |
| 3.2.4 O antigen synthetic pathways | 102 |
| a) The Wzy-dependent pathway | 102 |
| b) The ABC transporter-dependent pathway | 104 |
| 3.2.5 The ligase WaaL | 105 |
| 3.2.6 The LPS Export Pathway Lpt | 105 |
| 3.3 THE INTEGRAL MEMBRANE PROTEIN WZY | 106 |
| 3.3.1 Structure of Wzy polymerase in <i>P. aeruginosa</i> PAO1 | 107 |
| 3.3.2 Wzy mechanism for specific O-antigen polymerization | 108 |
| RESULTS AND DISCUSSION | 110 |
| 3.4 WZY RECOMBINANT PROTEIN ANALYSIS | 110 |
| 3.4.1 Wzy _{pa} cloning | 110 |
| 3.4.2 Overexpression of Wzy | 111 |
| 3.4.3 Solubilization tests in different detergents | 114 |
| 3.4.4 Affinity chromatography purification | 115 |
| 3.4.5 Size exclusion chromatography purification | 120 |
| 3.4.6 Mass spectrometry analysis | 124 |
| 3.4.7 UV-visible spectrophotometric analysis | 125 |
| 3.4.8 Optimization of the affinity purification protocol | 127 |
| 3.4.9 Ion Exchange Chromatography test | 128 |
| 3.4.10 Comparative analysis of results from chromatographic and electrophoretic tests | 130 |
| 3.4.11 Circular Dichroism analysis on purified Wzy | 130 |
| 3.4.12 Fluorescence spectroscopy analysis of purified Wzy | 132 |

| | | |
|--------|--|-----|
| 3.4.13 | Raman spectroscopy | 134 |
| 3.4.14 | Wzy crystallization | 135 |
| 3.4.15 | Negative staining Electron Microscopy analysis | 140 |
| 3.5 | CONCLUSIONS AND FUTURE PERSPECTIVES | 142 |
| | REFERENCES | 145 |

Chapter 4

| | | |
|------------------------------|--|-----|
| Materials and methods | | 152 |
| 4.1 | CLONING | 152 |
| 4.1.1 | Vectors | 152 |
| | • P-IV atpase expression in <i>S. cerevisiae</i> | 152 |
| | • Wzy expression in <i>E. coli</i> | 153 |
| 4.1.2 | Polymerase chain reaction | 154 |
| | • Homologous recombination in yeast | 156 |
| | • Restriction-free cloning in <i>E. coli</i> | 156 |
| 4.1.3 | Plasmid transformation into <i>E. coli</i> or <i>S. cerevisiae</i> | 157 |
| 4.1.4 | Selection of positive colonies | 157 |
| 4.1.5 | Site-directed mutagenesis | 158 |
| 4.2 | MEMBRANE PROTEIN OVEREXPRESSION AND PURIFICATION | 159 |
| 4.2.1 | Small-scale expression | 159 |
| | • Yeast <i>S. cerevisiae</i> | 159 |
| | • Bacterium <i>E. coli</i> | 159 |
| 4.2.2 | Large-scale expression | 160 |
| 4.2.3 | Detergent solubilization | 160 |
| | Ionic detergents | 160 |
| | Non-ionic detergents | 161 |
| | Zwitterionic detergents | 161 |
| | Bile acid salts | 161 |
| 4.2.4 | Chromatographic methods for membrane protein purification | 162 |
| | Affinity chromatography | 162 |
| | Size exclusion chromatography (SEC) | 163 |
| | Ion exchange chromatography (IEX) | 163 |

| | | |
|-------|---|-----|
| 4.2.5 | Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis | |
| | SDS-PAGE | 164 |
| 4.2.6 | Use of green fluorescent protein for membrane proteins detection | 164 |
| 4.2.7 | Membrane protein quantification | 165 |
| | UV absorbance | 165 |
| | BCA assay | 165 |
| 4.3 | BIOPHYSICAL CHARACTERIZATION OF MEMBRANE PROTEINS | 166 |
| 4.3.1 | Circular dichroism on membrane proteins | 166 |
| 4.3.2 | Fluorescence spectroscopy | 167 |
| 4.3.3 | Raman spectroscopy for membrane proteins | 167 |
| 4.4 | STRUCTURAL STUDIES | 168 |
| 4.4.1 | Crystallization of membrane proteins | 169 |
| 4.4.2 | Negative staining electron microscopy for membrane proteins | 174 |
| 4.5 | COMPUTATIONL MODELS | 175 |
| | Protein alignment with Blosum62 matrix and computational model generation with swissmodel | 175 |
| 4.6 | BIOCHEMICAL ASSAYS | 176 |
| | ATP activity assay based on the Luciferin/Luciferase chemoluminescence | 176 |
| 4.7 | MATERIALS | 177 |
| 4.8 | Media and buffers | 178 |
| | Agarose gel for DNA separation | 180 |
| | SDS-PAGE for protein separation and identification | 180 |
| 4.7.1 | Neo1 protocols | 181 |
| | • Cloning protocol for P-IV ATPases: | 181 |
| | PCR for gene amplification | 181 |
| | Transformation in FGY217 and INVSc1 <i>S. cerevisiae</i> strains via homologous recombination | 182 |
| | Plasmid yield increase in <i>E. coli</i> and gene insertion control | 182 |
| | • Expression protocol for P-IV ATPase | 183 |
| | Small-scale overexpression protocol | 183 |

| | |
|---|-----|
| Large-scale expression protocol | 184 |
| • Neo1 mutants protocol | 184 |
| Mutants on Neo1 atpase activity | 184 |
| Mutants on the lipid substrate binding site | 185 |
| Mutagenesis protocol for Neo1 | 185 |
| • Membranes recovery | 186 |
| • Neo1 detergent test | 186 |
| • Protein purification | 187 |
| • Neo1 CD analysis | 188 |
| • Neo1 crystallization trials | 188 |
| • Neo1 negative staining and tem analysis | 188 |
| • Neo1 activity assay | 189 |
| • Mutants activity assays | 189 |
| 4.7.2 Wzy protocols | 190 |
| • RF cloning of Wzy from PAO1 | 190 |
| • <i>E. coli</i> competent cells | 191 |
| • Trasformation in <i>E. coli</i> Lemo21(DE3) or BL21(DE3)pLysE cells | 191 |
| • Wzy expression | 192 |
| Small-scale expression | 192 |
| Large-scale expression in Lemo21(DE3) cells | 192 |
| Membranes recovery | 193 |
| • Wzy detergent test | 193 |
| • Wzy affinity chromatography test | 193 |
| Wzy test in different affinity chromatography resins | 193 |
| Test on different detergents with Co ²⁺ resins | 194 |
| • Wzy purification | 194 |
| • Wzy CD analysis | 195 |
| • Wzy fluorometry analysis | 196 |
| • Wzy for Raman spectroscopy | 196 |
| • Wzy crystallization trials | 197 |
| • Negative staining and EM analysis of Wzy | 198 |

REFERENECEES

199

- AKNOWLEDGEMENTS

202

LIST OF ABBREVIATIONS

8eLeg 5,7-diamino-3,5,7,9-tetra-deoxy-L-glycero-D-galacto-non-2-ulosonic(8-epilegionaminic) acid

A Arginine

ABC ATP-Binding Cassette

Ac Acetyl

AMEs Aminoglycoside Modifying Enzymes

APS Ammonium Persulfate

BCA Bicinchoninic Acid Assay

BRIC1 Benign Recurrent Intrahepatic Cholestasis type 1

BSA Bovine Serum Albumine

CD Circular Dichroism

CF Cystic Fibrosis

CFTR Cystic Fibrosis Transmembrane Conductance Regulator

CL Cardiolipin

CL- Cytoplasmic Loop

CMC Critical Micelle Concentration

Cryo-EM Cryo Electron Microscopy

CUO Cytochrome O Ubiquinol Oxidase

CV Column Volume

D Aspartic Acid

DLOs lipid-linked oligosaccharides

dNTPs Deoxynucleotides triphosphates

E Glutamic Acid

EARS-Net European Antimicrobial Resistance Surveillance Network

EDTA Tris-acetate ethylenediaminetetraacetic acid

EM Electron Microscopy

ER Endoplasmic Reticulum

Fo Formyl

FucN 2-amino-2,6-dideoxy-galactose

GalN 2-amino-2-deoxy-D-galactose

GFP Green Fluorescent Protein

GL Glycosyltransferase

GLC D-glucose

GlcNA 2-amino-2-deoxy-glucuronic acid

GuINA 2-amino-2-deoxy-guluronic acid

HCAIs Health-Care Associated Infections

Hep L-glycero-manno-heptose

IATS Antigenic Typing System

IEX Ion Exchange Chromatography

IL1- β Interleukin 1- β

IM Inner Membrane

IMAC Immobilized Metal Affinity Chromatography

IP Immuno Precipitation

IPTG Isopropyl- β -D-thiogalactopyranoside

IR Infrared

K lysine

Kdo 3-deoxy-D-manno-oct-2-ulosonic acid

LPS lipopolysaccharide

LPT Lipopolysaccharide Export Pathway

L-Rha L-Rhamnose

MAD Multi-wavelength Anomalous Dispersion

ManN 2-amino-2-deoxy-mannose

ManNA 2-amino-2-deoxy-mannuronic acid

MIR Multiple Isomorphus Replacement

MDR Multidrug-Resistant

N- amino

N Asparagine

NAc acetamido

NMR Nuclear Magnetic Resonance spectroscopy

NS Negative Staining

OAc O-acetyl

OM Outer Membrane

PA Phosphatidic Acid

PC Phosphatidylcholine

PCR Polymerase Chain Reaction

PE Phosphatidylethanolamine

PEG polyethylene glycol

PFIC1 Progressive Familial Intrahepatic Coolestasis type1

PG Phosphatidylglycerol

PI Phosphatidylinositol

PI4P Phosphatidylinositol 4-phosphate

PL Periplasmic Loop

PM Plasma Membrane

PMSF Phenylmethylsulfonyl Fluoride

PS Phosphatidylserine

Pse 5,7-diamino-3,5,7,9-tetra-deoxy-L-glycero-L-manno-non-2-ulosonic (pseudoaminic) acid

QuiN 2-amino-2,6,dideoxy-glucose

R Arginine

RF Restriction Free

Rha Rhamnose

RHb (R)-3-hydroxybutanoyl

Rib Ribose

RT Room Temperature

S Serine

SDS Sodium Dodecyl Sulphate

SDS PAGE Sodium Dodecyl Sulphate Polyacrylamide Gele Electrophoresis

SCVs Small Colony Variants

SEC Size Exclusion Chromatography

SERCA Sarcoplasmatic Calcium Pump

SHb (S)-3-hydroxybutanoyl

SM Sphingomyelin

TBS Tris-Buffered Saline

TEM Transmission Electron Microscope

Temed Tetramethylethylenediamine

TEV Tobacco Etch Virus

TGN Trans-Golgi Network

TLR4 Toll-Like Receptor 4

TM Transmembrane

TNF- α Tumor Necrosis Factor α

UDP Undecaprenyl Phosphate

UK Ulcerative Keratitis

XDR Extensively-Drug Resistant

ABSTRACT

The cellular membrane is a physical barrier that divides the interior of the cell, the cytosol, from the extracellular environment. The membrane has the important role of protecting the cell against external attacks, but must be permeable for nutrients, molecules involved in signaling and other specific compounds. Despite a similar composition of lipids and proteins, cellular membranes of prokaryotic and eukaryotic organisms are structurally and functionally different. In this work I focused on the characterization of two membrane proteins involved in physiological processes that maintain the crucial features of cellular membranes.

My first target was the eukaryotic integral membrane protein Neo1 of the yeast *Saccharomyces cerevisiae*, belonging to the P-IV ATPase family and essential to maintain lipid asymmetry in the plasma membrane as well as in membranes delimiting internal organelles. To this aim, P-IV ATPases translocate phospholipids from the outer to the inner leaflets of the membranes. In humans, malfunction of these proteins are related to several pathologies, including liver diseases, Alzheimer's, obesity and type 2 diabetes. Considering the key role of these proteins, the clarification of their mechanism of action is of utmost importance. Exploiting the high homology between human and yeast P-IV ATPases, I carried on studies aimed at the structural and functional characterization of a specific *S. cerevisiae* P-IV ATPase, homologous of a subclass of human proteins. Neo1 is the only *S. cerevisiae* P-IV ATPase essential for cell survival, but its substrates are still a matter of debate. The target protein was expressed as a chimera with the green fluorescence protein (GFP) and a histidine-tag at its C-terminus. The presence of the GFP enabled the fast detection of the protein throughout the expression and purification processes, and allowed to optimize protein yield. For membrane proteins, the choice of the proper solubilization detergent is a crucial step for the characterization of their native conformation and activity. Among the various detergents tested for Neo1 solubilization, a pure and stable protein could only be obtained in LMNG. Circular dichroism analysis of the sample in this detergent indicated a melting temperature in the range between 40°C and 50°C. Activity assays were performed on the purified protein in presence of different lipids in order to identify Neo1 substrate(s). Mutants of the protein were expressed to test the role of specific residues in protein activity.

Negative staining EM images at high magnification showed a homogeneous sample without aggregation, representing an excellent starting point for further analysis.

A second project developed during this PhD, involved the study of the integral membrane polymerase Wzy from *Pseudomonas aeruginosa*. *P. aeruginosa* is a pathogen related to hospital-acquired infections and often found in the lungs of cystic fibrosis patients. In bacteria, a thick polysaccharide wall covers the cell membrane and is involved in crucial steps of the bacterial infection, such as antibiotic resistance mechanisms and bacterium-host interactions, including the stimulation of the host immune response. The glycosyltransferase Wzy is a fundamental part of the pathway for the production of the lipopolysaccharide (LPS) on the outer membrane of gram-negative bacteria and its role is to form new glycosidic bonds between short oligosaccharide units, thus contributing to the production of the exposed part of the LPS. As the LPS is highly specific for each bacterial species and strain, inhibitors of its biosynthetic route would be selective against a specific bacterium. Understanding Wzy structure would be a step forward in the rational design of novel antibiotics against *P. aeruginosa*.

In this study, Wzy was cloned and overexpressed in *E. coli* as a chimera with GFP and His tags. The protein was solubilized in different detergents and purified with chromatographic methods. The purification results allowed to identify the optimal detergents to obtain a pure and stable protein, suitable for crystallization or EM analysis. Stability of the protein was tested against a temperature gradient with Circular Dichroism, Raman spectroscopy, and fluorescence spectroscopy. In both CD and fluorescence spectroscopy, a melting point of about 58°C was determined, temperature at which half the protein sample is unfolded. Crystallization trials were set up using sparse matrices of conditions and obtaining different crystals. From diffraction experiments at the XRD2 beamline of the Elettra Synchrotron, crystallographic data up to 6 Å resolution were obtained, but structure solution revealed that the crystals were formed by a contaminant present in the sample, the Cytochrome O Ubiquinol Oxidase. Further purification experiments suggested a tight interaction between Wzy and the Cytochrome o Ubiquinol Oxidase, with the effect of stabilizing the former. Negative staining EM analysis was carried out on the protein-contaminant sample, confirming that the sample is monodisperse and homogeneous and is therefore suitable for further EM analysis.

RIASSUNTO

La membrana cellulare è una barriera fisica che divide l'interno della cellula, il citosol, dall'ambiente esterno. La membrana ha l'importante ruolo di proteggere la cellula da attacchi esterni, ma deve essere permeabile a nutrienti, molecole segnale e altri specifici composti. Nonostante una composizione simile di lipidi e proteine, le membrane cellulari di organismi eucarioti e procarioti sono strutturalmente e funzionalmente diverse. In questo lavoro mi sono concentrata sulla caratterizzazione di due proteine di membrana coinvolte in processi fisiologici per il mantenimento delle caratteristiche fondamentali delle membrane cellulari.

Il mio primo target è la proteina integrale di membrana Neo1 del lievito *Saccharomyces cerevisiae*, appartenente alla classe delle P-IV ATPasi, proteine essenziali nel mantenimento dell'asimmetria dei lipidi sulla membrana plasmatica così come in quello delle membrane che delimitano gli organelli interni della cellula. A questo scopo, le P-IV ATPasi trasferiscono fosfolipidi dalla parte esterna delle membrane a quella interna. Nell'uomo, malfunzionamenti di queste proteine sono connessi a diverse malattie, tra cui malattie del fegato, Alzheimer, obesità e diabete di tipo 2. Considerando il ruolo cruciale di queste proteine, la delucidazione del loro meccanismo di azione è di fondamentale importanza. Sfruttando l'alta omologia tra P-IV ATPasi umane e di lievito, ho portato avanti degli studi mirati alla caratterizzazione strutturale e funzionale di una specifica P-IV ATPasi di *S. cerevisiae*, omologa ad una sottoclasse di proteine umane. Neo1 è l'unica P-IV ATPasi di *S. cerevisiae* essenziale per la sopravvivenza cellulare, ma i suoi substrati sono tuttora sconosciuti. La proteina selezionata è stata espressa come chimera con la Green Fluorescent Protein (GFP) e una sequenza di istidine al suo C-terminale. La presenza della GFP ha permesso una veloce individuazione della proteina nei passaggi dall'espressione alla purificazione, consentendo l'ottimizzazione della resa della proteina. Per le proteine di membrana, la scelta di un detergente adatto alla loro solubilizzazione è un passaggio cruciale nella caratterizzazione della loro conformazione nativa e della loro attività. Tra i vari detergenti testati per la solubilizzazione di Neo1, solo in LMNG è stato possibile ottenere una proteina pura e stabile. L'analisi al dicroismo circolare del campione in questo detergente ha indicato una temperatura di fusione nell'intervallo tra 40°C e 50°C. Saggi di attività sono stati effettuati sulla proteina in presenza di diversi lipidi per identificare il

substrato di Neo1. Mutanti della proteina sono stati espressi per testare il ruolo di specifici residui coinvolti nella sua attività. Nelle immagini di microscopia elettronica di Neo1 con la tecnica di *negative staining* si vede un campione senza aggregazione. Questo risultato rappresenta pertanto un eccellente punto di partenza per analisi future.

Un secondo progetto sviluppato durante questo dottorato di ricerca, ha coinvolto lo studio della polimerasi integrale di membrana Wzy da *Pseudomonas aeruginosa*. *P. aeruginosa* è un patogeno connesso ad infezioni ospedaliere ed è spesso presente nei polmoni di pazienti affetti da fibrosi cistica. Nei batteri, un denso muro di polisaccaridi ricopre la membrana cellulare ed è coinvolto in processi cruciali durante l'infezione batterica, come meccanismi di resistenza agli antibiotici e interazioni tra il batterio e l'ospite, compresa la stimolazione della risposta del sistema immunitario. La glicosiltransferasi Wzy è una parte fondamentale del processo per la formazione del lipopolisaccaride (LPS) della membrana esterna nei batteri gram negativi e il suo ruolo è di formare nuovi legami glicosidici tra corte unità oligosaccaridiche, contribuendo così alla produzione della parte esposta dell'LPS. Poiché l'LPS è altamente specifico per ogni batterio e ceppo, inibitori della sua biosintesi sarebbero selettivi contro un batterio specifico. Pertanto la comprensione della struttura di Wzy rappresenterebbe un passo avanti per la creazione di nuovi antibiotici contro *P. aeruginosa*.

In questo studio, Wzy è stata clonata e overespressa in *E. coli* come chimera assieme alla GFP e ad una sequenza di istidine. La proteina è stata solubilizzata in diversi detergenti e purificata con metodi cromatografici. Dai risultati della purificazione è stato possibile identificare i detergenti migliori per ottenere una proteina pura e stabile per analisi cristallografiche e di microscopia elettronica. La stabilità della proteina è stata testata contro un gradiente di temperatura con dicroismo circolare (CD), spettroscopia Raman e spettroscopia di fluorescenza. Sia in CD che in spettroscopia di fluorescenza, è stato determinato un punto di fusione di circa 58°C, temperatura alla quale metà del campione proteico è sfoldato. Diversi cristalli sono stati ottenuti da prove di cristallizzazione usando matrici sparse di condizioni. Da esperimenti di diffrazione alla linea XRD2 del Sincrotrone Elettra, sono stati ottenuti dati cristallografici a circa 6 Å di risoluzione, ma la soluzione della struttura ha rivelato che questi cristalli sono formati da un contaminante presente nel campione, la Citocromo O Ubiquinolo Ossidasi. Esperimenti successivi hanno suggerito una stretta interazione tra Wzy e la Citocromo O Ubiquinolo Ossidasi, con effetto di stabilizzazione della prima. Sul campione di proteina e contaminante sono state effettuate

analisi di microscopia elettronica con la tecnica di *negative staining*, confermando che il campione è monodisperso e omogeneo, essendo pertanto promettente per future analisi.

Chapter 1

The cellular membrane: a liquid barrier between outside and inside

1.1 THE CELLULAR MEMBRANE

Cells are separated from the environment by a barrier of lipids that divide the inside of the cell, the cytoplasm, from the rest of the world, the extracellular fluid. These barriers are composed of asymmetric lipid bilayers containing many embedded proteins (Hedger and Sansom 2016). The lipids arrange to create a lamellar structure with a hydrophobic core and a polar region at both sides.

In cells, the membrane is composed of hundreds of different lipid species (de la Serna et al. 2016).

1.1.1 *The membrane characteristics*

During the years the description of the cellular membrane developed from a model of a fluid mosaic with no organization (Singer and Nicolson 1972), to a complex structure in which domains known as *rafts* are present with different composition of lipids and proteins (Simons and Ikonen 1997). The formation of these ordered structures have been discussed for long time, but recent experiments confirm their existence (Morozova, Guigas, and Weiss 2011) and the extraordinary characteristics that they confer to the membrane.

- *Membrane fluidity*: Fluidity is one of the most critical characteristics of the cellular membrane. Fluidity allows molecules to diffuse and rotate in order to adopt an optimal arrangement within membranes. Membrane fluidity is strongly dependent on acyl chain composition and its control is crucial. Short lipid chains with unsaturated fatty acids increase fluidity, while long chains with saturated fatty acids

and the presence of cholesterol confer rigidity to the membrane (Los and Murata 2004; S. T. Yang et al. 2016). The cell is able to modify its fluidity by changing lipid acyl chains composition of its membranes in response to environmental stimuli such as temperature (Buda et al. 1994).

- *Viscosity*: Membrane viscosity is another characteristic of similar importance. Different membrane structures show different viscosity according to their specific functions. The presence of long acid chains and cholesterol increase viscosity, whereas short fatty acids decrease it (Nipper et al. 2008). As in the case of fluidity, viscosity can be stimuli-adapted, too (Luby-Phelps et al. 1993).
- *Continuity*: A continuous lipid bilayer fully covers the cell surface. Continuity is needed not only to retain organelles and molecules inside the cell, but is also an important factor for intermolecular interactions and formation of multi-molecular assemblies (Janetopoulos and Firtel 2008).
- *Cooperation*: Molecules present in the cellular membrane cooperate with each other. In fact, every component of the membrane has a specific role and is in relationship with its surrounding. Lipid-lipid, lipid-protein and protein-protein interactions are present at the membrane level, each fulfilling a specific function (Singer and Nicolson 1972).
- *Interleaflet coupling*: Lipid acyl chains forming the two opposite leaflets interdigitate in the lipophilic membrane core. While their interactions could be seen only as dominated by hydrophobic repulsive forces making the two leaflets of the membrane behave independently from each other, the observation that molecules of the two leaflets can organize in coordination suggests that an interdependence is present.

The so-called *interleaflet coupling* induces the organization of molecules between the two leaflets (Nickels, Smith, and Cheng 2015; Williamson and Olmsted 2015).

- *Asymmetry*: Probably the most distinctive characteristic of natural membranes is their asymmetric distribution of phospholipids among the two leaflets. Membrane asymmetry involves not only lipids but also proteins and surface ion composition (Van Meer, Voelker, and Feigenson 2008). Moreover, each cellular membrane has a different degree of asymmetry and a different distribution of molecules between the leaflets, creating a remarkable diversity and complexity between cellular compartments. Asymmetry is pivotal for a number of cellular processes, including the maintenance of the cell shape (Marquardt, Geier, and Pabst 2015).
- *Self-assemblies*: Lipids self-assembly and segregation on the cellular membrane has been demonstrated and the importance of this phenomenon lies in the fact that it generates membrane sections enriched in one or more specific lipids (Ivankin, Kuzmenko, and Gidalevitz 2010).
- *Protein-islands*: On the cellular membrane, proteins are not uniformly distributed into the phospholipids bilayer, but protein-free and highly-dense protein areas are present simultaneously, creating so-called *protein islands* (Lillemeier et al. 2010).
- *Lateral pressure*: Steric and electrostatic effects generate different pressures within each monolayer of the phospholipid bilayer (J. M. Seddon 1990). A lateral pressure is generated in the external region from steric and electrostatic effects of the phospholipids headgroups (Anthony G. Lee 2004). In contrast, immediately below the lipid headgroups, in the region of the glycerol backbone, an attractive force is generated by the water exclusion and by hydrophobic interactions of the fatty acyl chains leading to tight lipid packing in this region. This produces an opposite lateral pressure that leads to a contraction of the bilayer. If these two pressures are

balanced the lipid layer stays flat. But if the lateral pressure generated by the acyl chain packing is greater than the one generated by the headgroups, the lipid layer tends to curl towards the aqueous ambient. Vice versa, if the pressure generated by the headgroups is greater than the one generated by the acyl chains, the monolayer curves towards the hydrophobic chain region. In order to maintain the integrity of the bilayer, when a monolayer curls in one direction, the second has to follow. For the bilayer to create an overall curve, the two monolayers need to have opposite unbalances of lateral pressure. This phenomenon generates a free elastic energy stored in the membrane bilayer (Anthony G. Lee 2004). Introduction of a protein in the membrane releases some of this energy, thus promoting this insertion.

1.1.2 *Bacterial cell envelope: a complex structure of membranes*

In order to survive in an often hostile environment, bacteria developed sophisticated envelopes for protection. These structures serve as protection barriers but allow selective passage of nutrients from the outside and waste products discharge from the inside (Silhavy, Kahne, and Walker 2010).

In 1884 Christian Gram classified bacteria in two groups according to their ability to retain a specific staining molecule (Gram 1884). Gram-positive bacteria are colored after Gram staining procedure while gram-negative are not*. This behavior is due to differences in their cellular envelope (Silhavy, Kahne, and Walker 2010) (**Table 1.1**).

** Throughout this dissertation, the capitalization of Gram name follows the the CDC guidelines: capitalized when referred to Gram stain, lowercase and hyphenated when referred to gram positive or gram negative bacteria. <https://wwwnc.cdc.gov/eid/page/preferred-usage>*

- *Gram-negative bacteria cell envelope*

The gram-negative bacteria envelope is composed of three layers: the outer membrane (OM) which is in contact with the external environment, the peptidoglycan cell wall that confers shape and rigidity to the cell and the inner membrane (IM) which is in contact with the cellular cytoplasm (Glauert and Thornley 1969) (**Figure 1.1a**).

- a) *Outer membrane*: The presence of the OM is the distinctive characteristic of gram-negative bacteria. This membrane is composed of lipids but with a high degree of asymmetry. Interestingly, phospholipids are confined to the internal layer of the OM, while the external layer is composed of glycolipids (Kamio and Nikaido 1976). The most abundant glycolipid in the external leaflet is the lipopolysaccharide (LPS). LPS is recognized by the immune system during bacterial infection and is therefore responsible for the endotoxic shock associated with septicemia (Raetz and Whitfield 2002). This polysaccharide creates a non-fluid barrier for hydrophilic molecules, making gram-negative bacteria highly resistant to antibiotics. While LPS functions as barrier, different proteins work as channels for the passage of selected molecules, such as nutrients. The OM contains two main classes of proteins: β -barrel proteins and lipoproteins (Silhavy, Kahne, and Walker 2010). β -barrel proteins are integral transmembrane proteins, while lipoproteins are bound to lipids embedded in the membrane (Sankaran and Wus 1994).
- b) *Peptidoglycan cell wall*: The peptidoglycan is a very large rigid polymer that determines the cellular shape (Vollmer, Blanot, and De Pedro 2008). Enzymes or antibiotics able to damage the peptidoglycan wall cause cell death through lysis.
- c) *Periplasm*: The periplasm is an aqueous compartment characteristic of gram-negative bacteria. A distinct set of proteins compared with the cytoplasm is present in the periplasm (Heppel 1967). Due to the high density of proteins, the periplasm is particularly viscous (Mullineaux et al. 2006). This compartment functions as insulation for potentially harmful molecules (Silhavy, Kahne, and Walker 2010).
- d) *Inner membrane*: The IM is a phospholipid bilayer that serves as locus for all membrane-associated functions such as energy production, protein production and secretion, lipids biosynthesis, nutrients transport and waste export (Silhavy, Kahne, and Walker 2010).

- *Gram-positive bacteria cell envelope*

Gram-positive bacteria are named after the positive coloration they assume when exposed to Gram staining. This effect is due to the presence of a large peptidoglycan layer, on their cellular surface, interacting with and holding the stain.

The main difference with gram-negative bacteria is the absence of the OM (**Figure 1.1b**). As protection from the external environment, gram-positive bacteria developed a thick layer of peptidoglycan around their cellular membrane. As for gram-negative bacteria, the peptidoglycan is constituted by linear strands of repeated disaccharide units, cross-linked with peptide segments to form a mesh-like framework around the cell. Cross-linking is unique for each bacterium (Vollmer 2008). While in gram-negative bacteria the peptidoglycan is only few nm thick, in gram-positive bacteria the thickness of the wall can reach 100 nm (Silhavy, Kahne, and Walker 2010) and is the site of action for many antibiotics (Chambers 2003). Teichoic acids and proteins are inserted in the peptidoglycan wall. Teichoic acids are long anionic polymers, termed as wall teichoic acids if covalently attached to peptidoglycan, or lipoteichoic acids if anchored on membrane lipids (Neuhaus and Baddiley 2003). The protein distribution is more complex, as it is the function of different proteins: some bind to the teichoic acids, some to the peptidoglycan, some possess transmembrane domains and others are bound to lipid anchors embedded in the membrane (Scott and Barnett 2006). Many of these proteins serve for motion and adhesion of the bacterium to the host tissues, others are involved in synthesis, turn-over and modifications of peptidoglycan (Navarre and Schneewind 1999).

- *Corynebacterineae cell envelope*

A distinct class not included in either gram-positive or gram-negative bacteria is the class of *Corynebacterineae*, such as *Mycobacterium tuberculosis*. Their cell envelope shares characteristics of both gram-positive and gram-negative bacteria. In fact, they possess a thick peptidoglycan wall similar to gram-positive bacteria, but possess also an OM like gram-negative bacteria (Fu and Fu-Liu 2002).

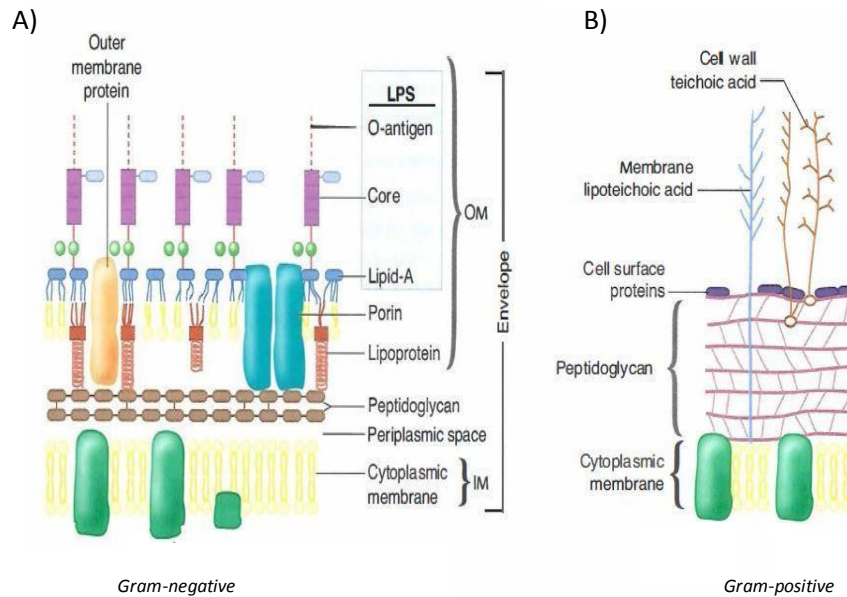


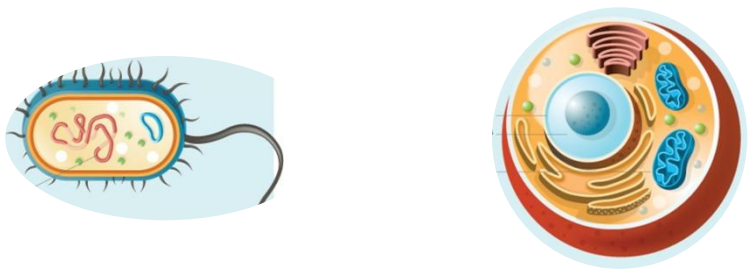
Figure 1.1: Differences between bacterial cell envelopes. **A)** Gram-negative bacteria have an inner membrane, a periplasmic space containing the peptidoglycan cell wall and an outer membrane, in which the LPS is inserted; **B)** Gram-positive bacteria have only a cytoplasmic membrane and a thick peptidoglycan cell wall, where teichoic acids are inserted. Images from <https://microbeonline.com/gram-staining-principle-procedure-results>.

1.1.3 Eukaryotic cell membranes

If for small organism such as bacteria the protection and transport functions are exerted by a complex set of different membrane types, in eukaryotic organisms the complexity of the membrane structure description is even increased by its heterogeneity and expanded roles (**Table 1.1**).

The most apparent difference between prokaryotes and eukaryotes is the internal organization of the latter, obtained through compartmentalization. Eukaryote membranes divide the cell in compartments defined organelles. More than constituting a boundary, membranes provide specific properties to each organelle. Consequentially, the composition of membranes of each organelle varies according to its functions (Jain and Holthuis 2017). Lipids and proteins are moved between organelles through vesicles, in vesicular trafficking. In the early stages of eukaryotic evolution, two distinct membrane systems were developed, centered on the endoplasmic reticulum (ER) and on the plasma membrane (PM), respectively (Jackson, Walch, and Verbavatz 2016). It is generally accepted that mitochondria derive from internalization of a bacterium into a eukaryotic cell (McLean and

Phillips 1981). An important proof of this theory is the absence of phosphatidylcholine (PC) in the mitochondrial membrane, a lipid present only in eukaryotic membranes (Vance and Tasseva 2013).



| | PROKARYOTES | | EUKARYOTES |
|--------------------------------|---|---|---|
| | Gram-positive | Gram-negative | |
| Membrane | Inner membrane | Inner membrane, periplasmic space, Outer membrane | Plasma membrane, organelles compartmentalization, vesicle Protection from the environment, cellular shape, |
| Properties and functionalities | Protection from the environment, cellular shape, incorporation of membrane proteins | | complex cellular architecture, vesiculation, cellular specialization depending on cell type |

Table 1.1: Differences between eukaryotic and prokaryotic membranes. The table is adapted from Simons K. and Sampaio J. (2011), *Cold Spring Harbor Perspectives in Biology* (Simons and Sampaio 2011). Images are taken from <https://www.ancestry.com/lp/where-is-dna-found/cells>.

Specific properties of each eukaryotic membrane are provided by a different distribution of sterols in the membranes of different organelles. Sterols are more abundant in the PM, where they confer to the membrane the physicochemical properties required for its barrier functions. In the ER, conversely, sterol levels are low in order to allow a looser membrane packing and favor insertion of newly synthesized membrane proteins (Jain and Holthuis 2017).

Membrane asymmetry is clearly visible in the PM, where lipids have the most divergent distribution between the leaflets, while it is absent in the ER (Van Meer, Voelker, and Feigenson 2008), where lipids have homogeneous distribution between sides.

1.2 LIPIDS: THE FATTY COMPONENTS OF THE CELLULAR MEMBRANE

The matrix of the cellular membrane is composed by amphipathic polar phospholipids, consisting in an hydrophilic and a lipophilic portion and crucial in several cellular functions (Van Meer, Voelker, and Feigenson 2008). Not only they form a barrier to delimit the cell, but they are also involved in intracellular messaging, growth, metabolism and protein activity (Dowhan, Mileykovskaya, and Bogdanov 2004). Due to their affinity to the membrane and their ability to cross it, in the last years phospholipids have gained interest as drug carriers.

1.2.1 *Classification of lipids present in cellular membranes*

- *Glycerophospholipids*: these lipids are composed of two fatty acids, one saturated and one containing double bonds, connected to a glycerol head. The last alcoholic group of the glycerol residue is attached to a phosphate moiety. In most of the glycerophospholipids, with the notable exception of the phosphatidic acid (PA), the phosphate group is linked to a hydrophilic small molecule (**Figure 1.2**). The phosphate group, its decoration and the glycerol backbone are the head of the glycerophospholipid, while the acyl chains are the hydrophobic part of the molecule, often termed tail (van Hoogevest and Wendel 2014). Fatty acids have different lengths, typically from 14 to 24, and degree of unsaturation (Burdge and Calder 2015; Poggi et al. 2015). Combining different fatty acids with different headgroups it is possible to obtain the huge number of phospholipids present in nature (Alagumuthu, Dahiya, and Singh Nigam 2019).

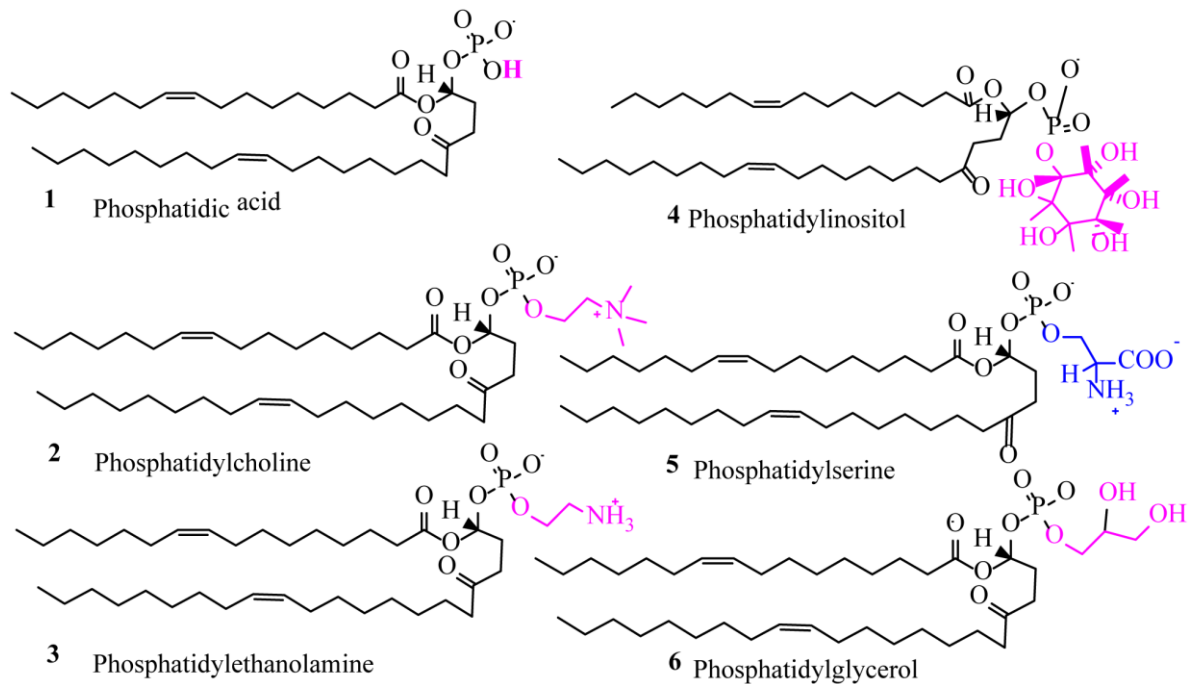


Figure 1.2: Glycerophospholipids differing for the head group. Phospholipid structure is divided in tail, composed of two acyl chains, and head, composed of a glycerol backbone, a phosphate group and small molecules. Exception is made for the head of phosphatidic acid that is composed only of glycerol backbone and phosphate group. Image is taken from Alagumuthu M. et al. (2019), AIMS Molecular Science (Alagumuthu, Dahiya, and Singh Nigam 2019).

- **Sphingolipids:** instead of a glycerol backbone these lipids are constituted by a serine backbone. They are critical components of the exofacial leaflet of eukaryotic PMs. The most abundant representative of this lipid class is sphingomyelin (SM), accounting for around 23% of the total phospholipids present in the PM (Y. Yang, Lee, and Fairn 2018).
- **Sterols:** unlike glycerophospholipids and sphingolipids, sterols are lipids with a large hydrophobic part and a small hydrophilic group. Cholesterol is the predominant sterol in mammalian cells and is important for modulation of membrane fluidity and permeability (Krause and Regen 2014).

- *Glycolipids*: the principal difference, and characteristic, of glycolipids is the lack of the phosphate headgroup, substituted with a sugar moiety, such as glucose. The backbone of glycolipids may be constituted either by glycerol or by sphingosine (Watson 2015).

1.2.2 Distribution of lipids present in cellular membranes

All organisms possess phosphatidylserine (PS) and phosphatidylethanolamine (PE), while PC is present only in eukaryotes (**Figure 1.2**). However, different membranes have different lipid composition and, within the bilayer, the distribution of lipids is asymmetric.

In eukaryotic cells, the abundances of choline-containing lipids such as PC or SM are higher in the exoplasmic leaflets of cellular membranes, while PE, PS and phosphatidylinositol (PI) are enriched in the cytosolic side (**Figure 1.3**) (Higgins and Evans 1978; Verkleij et al. 1973). Cholesterol is present in both sides with a preference for the cytoplasmic leaflet (Wood et al. 2011). Interestingly membrane rafts are enriched in SM and cholesterol (Pike 2006).

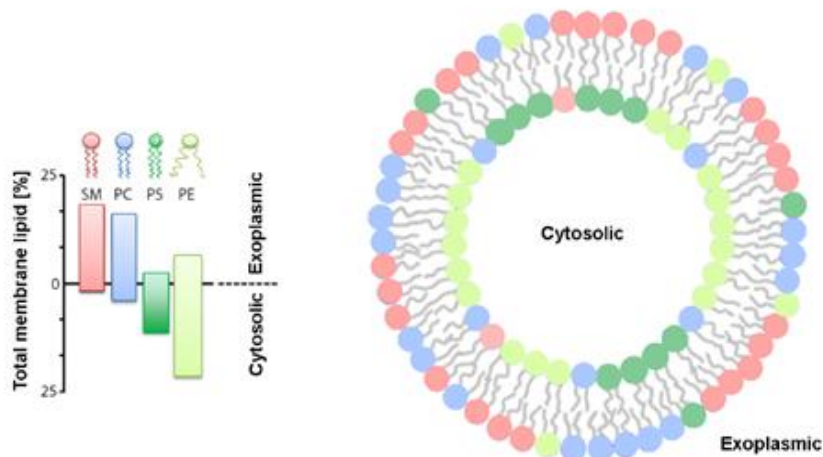


Figure 1.3: Schematic distribution of phospholipids in red blood cells. SM: Sphingomyelin; PC: Phosphatidylcholine; PS: Phosphatidylserine; PE: Phosphatidylethanolamine. SM and PC are enriched in the exoplasmic leaflet, while PE and PS are more present in the cytoplasmic side. Image from Marquardt D. et al. (2015), *Membranes* (Marquardt, Geier, and Pabst 2015).

In prokaryotes PI and PE are preferentially located in the inner membrane side, phosphatidylglycerol (PG) in the outer side. In gram-positive bacteria cardiolipin (CL) is distributed in both sides (Musatov and Sedlák 2017).

Due to its involvement in membrane potential, permeability, shape and signaling, loss of asymmetry has dire physiological consequences, such as cellular apoptosis or membrane permeability alterations in cancer cells, (Devaux 1991; Quinn 2002; Rivel, Ramseyer, and Yesylevskyy 2019).

PC is the most abundant phospholipid in the majority of eukaryotic cells representing from 41% to 57% of the total glycerophospholipids (Thewalt and Bloom 1992). PE is the second for abundance, being 17-38% of the total (Vance 2008). PS and PI are present in lower amounts, with percentages of 1-6% and 2-9%, respectively (Yeung et al. 2009). Other glycerophospholipids such as CL, PG or PA are rare (Y. Yang, Lee, and Fairn 2018). However, PG is highly abundant in mitochondria.

In eukaryotes, the ER is the biosynthetic organelle responsible for PC, PE, PI and PS production, whereas PG and CL are produced in mitochondria (Airola and Hannun 2013). A small amount of lipids is synthesized *de novo*, while the majority derives from recycled membranes (Watson 2015). After formation, phospholipids are symmetrically distributed in the ER membrane bilayer, but this symmetry is lost through the secretory pathway, to reach a completely asymmetric distribution in the PM (Van Meer, Voelker, and Feigenson 2008). Cholesterol is produced in the ER, but its abundance increases through membranes in the secretory pathway reaching the maximum in the PM. In fact, a high cholesterol content increases membrane thickness and rigidity, a quality useful for membranes deputed to protect the cell (Watson 2015; S. T. Yang et al. 2016).

A complex system of vesicular and non-vesicular trafficking is involved in the distribution of lipids in the different cell membranes (**Figure 1.4**). The vesicular trafficking consists in secretion and endocytosis of vesicles for inter- or intra-organellar transport, while soluble proteins able to carry lipids are responsible for non-vesicular trafficking (Sleight and Pagano 1983).

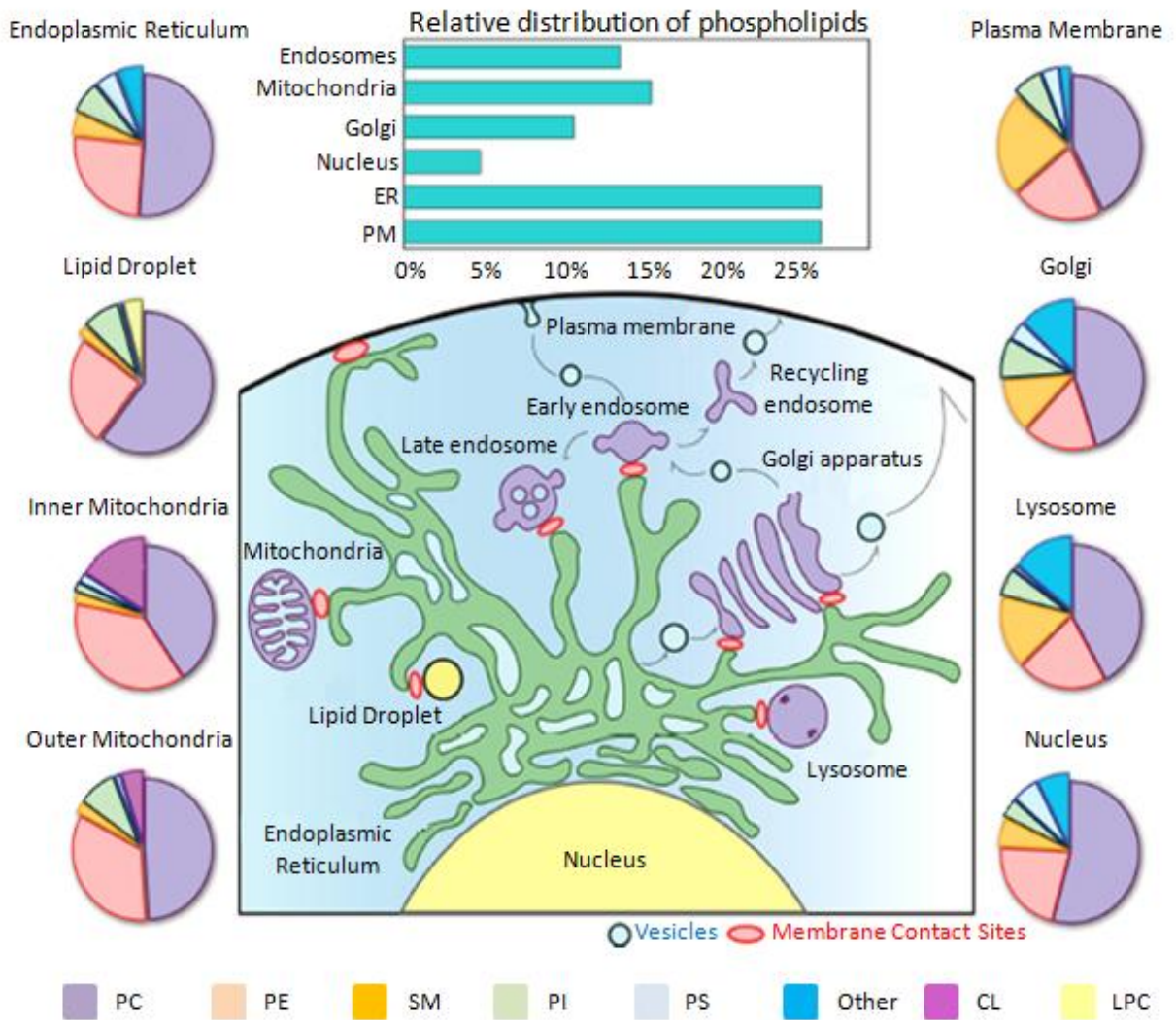


Figure 1.4: Lipid distribution in eukaryotic organelles. ER: Endoplasmic Reticulum; PM: Plasma Membrane; PC: Phosphatidylcholine; PE: Phosphatidylethanolamine; SM: Sphingomyelin; PI: Phosphatidylinositol; PS: Phosphatidylserine; Other: precursor and signal lipids; CL: Cardiolipin; LPC: Lyso-phosphatidylcholine. PC is the most abundant glycerophospholipid in the cell, PE is the second even if it is highly present in mitochondria inner membrane and less present in other membranes. Image from Yang Y. et al. (2018), *Journal of Biological Chemistry* (Y. Yang, Lee, and Fairn 2018).

1.3 MEMBRANE PROTEINS: THE FLESHY COMPONENTS OF CELLULAR MEMBRANES

Proteins are a fundamental part of the membrane. Membrane proteins are so important for cell activity and vitality that they represent nearly 20-30% of all proteins encoded in the genome of eukaryotic cells (Wallin & Heijne, 2008). Among them, integral membrane proteins are completely embedded into the lipid layer, while peripheral membrane proteins interact only transiently with one leaflet.

1.3.1 *Integral membrane proteins*

Integral membrane proteins cover different, fundamental cellular functions: transport of molecules and ions, signaling, intra- and inter-cellular transport, control of lipid composition and more (G. Von Heijne 2007).

Integral membrane proteins are inserted in the highly anisotropic environment of the lipid bilayer, hydrophilic on the surface and hydrophobic in its core (White et al. 2001). The typical membrane thickness is around 5 nm (W. Ding et al. 2015). Of these, lipid headgroups form a 15 Å-thick layer on both external sides, where complex interactions, such as electrostatic, hydrogen bond and van der Waals-type interactions, occur. Differently in the core the main interactions are hydrophobic. Proteins inserted in such environment, therefore, must have a structure adaptable to all these kinds of forces (G. Von Heijne 2007). Due to the numerous structural constraints imposed by the simultaneous presence of well-defined hydrophobic and hydrophilic regions, integral membrane proteins possess only two distinct architectures (**Figure 1.5**): an α -helical arrangement of transmembrane segments perpendicular to the membrane, and a β -barrel conformation, with the main axis of the barrel perpendicular to the membrane (Gunnar Von Heijne 1996). These two simple architectures are able to shield the peptide backbone from the lipophilic environment of the membrane.

While α -helical membrane proteins represent 20-30% of all genes encoding helical protein (Krogh et al. 2001), β -barrel proteins are rare (Wimley 2002).

Malfunctions of integral membrane proteins are related to a wide range of diseases, among which neurological and cardiac diseases caused by ion channels misbehavior (Abbott 2006; Gargus 2006), cystic fibrosis due to mutations of a chloride transporter (L. S. King, Kozono,

and Agre 2004), and some heritable disorders linked to peroxisomes and mutations of mitochondrial membrane proteins (Y. Suzuki et al. 2001).

- *β -barrel proteins:* Nearly all membrane proteins with a β -barrel folding are located in the OM of gram-negative bacteria, or in small organelles such as mitochondria and chloroplasts in eukaryotes. Their main role is to facilitate the passive diffusion of small molecules (G. Von Heijne 2007). Besides facilitating translocation of compounds, β -barrel proteins cover a wide range of functions, including various receptors and enzymes. The primary sequence of β -barrel proteins has a distinctive pattern containing a non-polar side chain every second residue, generating a structure with non-polar residues exposed towards the hydrophobic lipid environment and polar residues towards the interior of the protein, forming a hydrophilic channel (Vinothkumar and Henderson 2010).
- *α -helical proteins:* α -helical membrane proteins are present in most membranes. They cover a huge number of functions as enzymes, channels, receptors and transporters. Even as they exhibit a huge variety of sizes and shapes, they share some common features. The transmembrane helices are mainly formed of hydrophobic amino acids (alanine, leucine, isoleucine, phenylalanine and valine) in direct contact with the hydrophobic acyl chains of the surrounding lipids. At the extremities of the helices aromatic amino acid residue, as tyrosine and tryptophan, interact with the lipid headgroups (Gunnar Von Heijne 2000). Loops connecting transmembrane segments are usually polar, with a higher number of positively charged residues (arginine and lysine) in those exposed to the cytoplasm compared to the exoplasmic loops (Gunnar Von Heijne 1989). Phospholipids play key roles in integral membrane protein insertion. For example, the negative charge of PS, a lipid abundant in the inner leaflet, contributes to orient proteins in the membrane by forming electrostatic interactions with the positively charged cytoplasmic loops (Lin and London 2014).

Given their high hydrophobicity, integral membrane proteins are synthesized directly into the membrane. An N-terminal signal peptide directs the nascent polypeptide chain for fast and proper collocation of the integral membrane protein into its correct membrane (G. Von Heijne 2007).

Integral membrane proteins interact with their lipid surroundings in non-specific ways (annular lipids), acting as a sort of “solvent” (A. G. Lee 2003), or as specific binders, essential for protein activity. Among this second class, there are lipids acting as co-factors or even as substrates (Simmonds et al. 1982). Bulk lipids do not interact with membrane proteins but are still fundamental in defining the lipid bilayer thickness (Anthony G. Lee 2004).

The thickness of the membrane in which a protein is inserted plays a critical role. The main requirement to ensure the stability of the system is that the hydrophobic thickness of the lipid tails matches the hydrophobic surface of the protein (Anthony G. Lee 2004). Any discrepancy generates a distortion in the lipid bilayer or in the protein shape (Webb et al. 1998). In order to avoid mismatches, the membrane reorganizes in domains in which the length of the acyl chains is suitable for the specific protein hydrophobic thickness (Fattal and Ben-Shaul 1993).

1.3.2 Peripheral membrane proteins

Peripheral membrane proteins (PMPs) are soluble proteins that lack transmembrane domains, but bind to one leaflet of the phospholipids bilayer through a lipid-binding domain present on the protein surface (Fuglebakk and Reuter 2018) (**Figure 1.5**). PMPs are important for different cellular functions such as signaling (Kholodenko, Hancock, and Kolch 2010) or transport of intermediates of different pathways (Kirchhausen 2000). Through membrane-mediated interactions, PMSs form high-ordered structures that can extend across the lipid bilayer, a crucial signaling mechanism (Morozova, Guigas, and Weiss 2011).

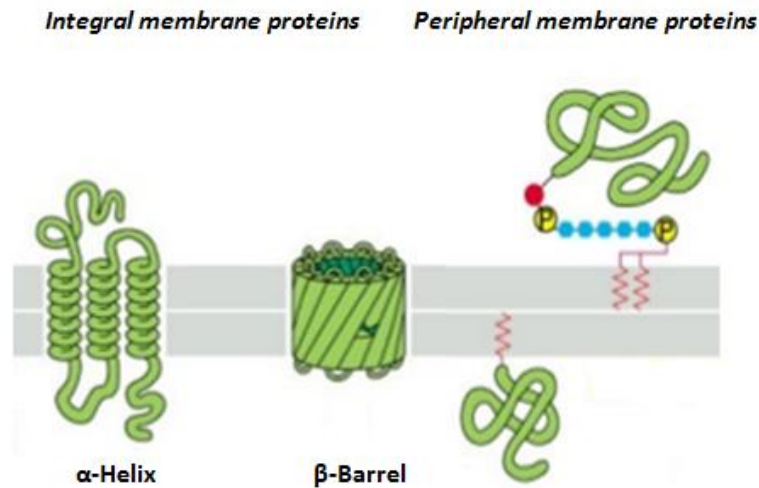


Figure 1.5: Integral and peripheral membrane proteins. Image adapted from slideshare.net/nedakhan3939/bio-480-cmdblec7.

1.3.3 Challenges in membrane protein structural analysis

While crystallography of soluble proteins dates back to the beginning of the 1950's, the first structure of a membrane protein, the photosynthetic reaction center of *Rhodospseudomonas viridis*, has been obtained via X-ray crystallography only in 1985 (Deisenhofer et al. 1985). Since then, the structures of many other membrane proteins have been solved. However, they still represent a challenging research topic due to the difficulties in their handling. From expression to crystallization, there are many crucial steps that require careful manipulation and that were greatly improved by recent methodological developments:

- **Protein expression:** Recombinant overexpression of the target membrane protein in a different host is the first step for its structural analysis. Widely used hosts are *E. coli* (D. Drew et al. 2005) and *S. cerevisiae* (Hays, Roe-Zurz, and Stroud 2010) for the easiness of their manipulation. However, the lipid composition of the host membrane is fundamental for the correct folding, membrane insertion and protein stability and it is a parameter that requires thorough preliminary analysis. In order to exploit the advantages of simpler systems such as bacteria, strains with specific lipid composition have been engineered specially for membrane protein production (K. Mathieu et al. 2019; Michou et al. 2019). The use of a Green Fluorescent Protein

(GFP) tag fused to the C-terminal of the selected membrane protein is an useful tool to monitor protein expression, as well as the following solubilization and purification steps (D. Drew et al. 2008).

- **The detergent issue:** Membrane proteins are stable in their highly hydrophobic lipid environment. In fact, structure and activity of integral membrane proteins are greatly affected by the presence of the surrounding phospholipids (Phillips et al. 2009). The use of amphipathic detergents is required to extract proteins from the membrane environment (A. M. Seddon, Curnow, and Booth 2004). The transition of the protein from the phospholipid environment to an aqueous solution through detergent solubilization is one of the most critical steps in protein purification. Detergents mimic the phospholipids bilayer generating a water soluble protein-detergent complex that reduces protein aggregation. Similarly to the phospholipids, detergents expose the polar headgroups to the aqueous environment, while the hydrophobic tails assemble in a lipophilic aggregate, forming micelles. The minimal detergent concentration required for their formation is a specific characteristic of each detergent, the Critical Micelle Concentration (CMC) (A. M. Seddon, Curnow, and Booth 2004).

Detergent can be classified as: (i) ionic, with a charged headgroup, further divided in cationic and anionic, (ii) non-ionic, with a polar but uncharged headgroup, or (iii) zwitterionic, having properties halfway between ionic and non-ionic detergents (Moraes et al. 2014). In protein purification, the most used detergents are non-ionic due to their milder effect on protein folding, since they are likely to disrupt lipid-lipid and protein-lipid interactions, rather than protein-protein.

Unfortunately, when aiming at solubilizing a membrane protein, there are no general rules for the detergent selection. Every membrane protein has its peculiar characteristics and its behavior in the presence of a detergent is difficult to predict. Different techniques can be used to assess protein stability in detergent, such as Thermal Shift Assays or spectroscopic techniques. Improvement of protein stability has been obtained through mutations, deletions and stabilizing terminal tags (Tate and Schertler 2009).

- *Solving a membrane protein structure*

If the solution of the structure of a protein is already a difficult matter, membrane proteins represent an even bigger challenge. In fact, their hydrophobicity and the flexibility of their domains make membrane proteins difficult to handle and to crystallize with the classical crystallization methodologies (Watson 2015). To partially overcome these issues, different commercial crystallization screens are available on the market. While X-ray crystallography remains a powerful tool to determine a membrane protein structure at atomic level (J. L. Parker and Newstead 2016), the slow pace at which membrane protein crystallization protocols were developed has prompted the use of other structural techniques, such as Nuclear Magnetic Resonance spectroscopy (NMR) and Cryo Electron Microscopy (Cryo-EM). However, the use of NMR techniques is limited to proteins below 40-50 kDa (Watson 2015; Howard 1998). On the contrary, Cryo-EM is suitable for proteins above 150-200 kDa, even if recent progress made possible to solve even smaller structures (Nwanochie and Uversky 2019; De Zorzi et al. 2016).

1.4 PROTEIN AND LIPIDS WORK TOGETHER

While lipids greatly affect membrane protein activity, sometimes even acting as co-factors, conversely proteins are required to maintain the lipid asymmetry of cellular membranes. The bilayer structure typical of the membrane forms spontaneously when phospholipids are dispersed in water, due to the amphipathic character of these molecules. In the bilayer, phospholipids undergo different intramolecular motions and exhibit rapid rotation and lateral diffusion, while reorientation of phospholipids across the membrane occurs very slowly (Kornberg and McConnell 1971; Mouritsen 2005) (**Figure 1.6**). Therefore, for an asymmetric membrane of phospholipids the spontaneous rearrangement towards symmetry is very slow. Considering the lipid distribution of natural membranes, it is obvious that the cell needs machineries for fast translocation of phospholipids across membranes in order to create and maintain the required asymmetry during growth, cellular stimuli, vesiculation, membrane-protein interactions and to face a range of environmental challenges (Mouritsen 2005).

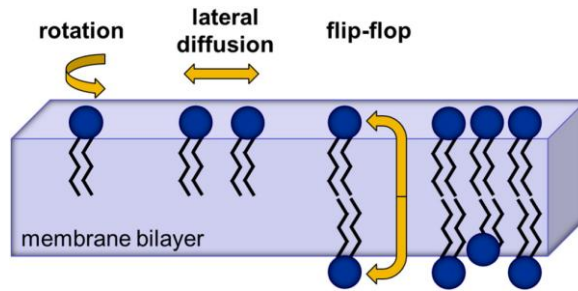


Figure 1.6: Movements of phospholipids within the membrane bilayer. Phospholipids can freely rotate or diffuse laterally, while the inversion movement occurs, either spontaneously or using an energy source, at a very slow pace and requires the action of proteins, such as scramblases, ABC transporter or P-IV ATPases, to be catalyzed and controlled. Image from Pomorki T. G. and Menon A. K. (2016), *Progress in lipid research* (T. G. Pomorski and Menon 2016).

Phospholipid translocation by specific proteins occurs through ATP-dependent or independent mechanisms, depending on the lipid gradient of the membrane (**Figure 1.7**). Lipid movements that reduce membrane asymmetry are spontaneous, but require catalysis due to their slow rate. ATP-independent transporters, proteins known as scramblases, allow a bidirectional translocation. Scramblases are either constitutively active or stimuli-regulated. On the contrary, to perform energetically unfavorable movements, i.e. against gradient, the cell expresses a number of different membrane proteins, generally classified according to directionality of lipid translocation. Flippases and floppases require ATP to translocate lipids and create or maintain the membrane asymmetry, the former with an outside-to-inside movement, the latter catalyzing the opposite translocation. Flippases are mainly P-IV ATPases, floppases are ABC transporter (Sanyal and Menon 2009).

- *Scramblases:* (i) *Constitutively active scramblases:* In eukaryotes, new lipids are synthesized in the cytoplasmic leaflet of the endoplasmic reticulum ER. Constitutively active scramblases are required to keep the symmetric distribution of the ER, but to date only few have been identified. In addition to their lipid translocation function, they play a key role in protein glycosylation in the ER, translocating the oligosaccharides involved in glycosylation from the cytosol to the lumen of the ER membrane (Schenk, Fernandez, and Waechter 2001). (ii) *Stimuli-regulated scramblases:* Stimuli-regulated scramblases activate only in presence of a chemical

signal and are able to quickly disrupt the membrane asymmetry when required by external conditions. An example of stimuli-regulated scramblases are the bacterial rhodopsin, a G-protein coupled receptor with a lipid translocation activity (Chauhan, Farine, and Bütikofer 2017). Some stimuli-regulated scramblases are involved in the exposure of PS on the membrane surface. In normal conditions, PS is sequestered in the cytoplasmic leaflet of the membrane. In particular conditions Ca^{2+} -dependent scramblases mediate loss of asymmetry, leading to an increased presence of PS on the cell surface (J. Suzuki et al. 2010). PS exposure is a cellular signal for phagocytosis, coagulation in red blood cells and apoptosis (J. Suzuki et al. 2013). In fact, Ca^{2+} -dependent scramblases are implicated in different bleeding disorders such as Scott syndrome (Bever and Williamson 2010).

- *P-IV ATPases*: These proteins, known also as flippases, catalyze the translocation of phospholipids from the outer to the inner part of the cellular membrane. They belong to the class of P-ATPases that couple ATP hydrolysis to the transport of different substrates against their concentration gradient. P-IV ATPases are present only in eukaryotes, where they are involved in flipping of PS, PC and PE (Alder-Baerens et al. 2006; X. Zhou and Graham 2009).
- *ABC transporters*: This superfamily of membrane proteins is characterized by the presence of two ATP-binding cassette (ABC) domains, able to hydrolyze ATP in order to transport different substrates across cellular membranes. Some ABC proteins are constituted by a single polypeptide chain with transmembrane domains and two ABC domains, while others are homo or hetero dimers with a single ABC domain on each polypeptide chain. The human genome codes for about 50 ABC transporters, classified in seven subfamilies and localized in different organelles. Only for few of these proteins (ABCB1, ABCA1, ABCC1, ABCA7) the lipid translocation activity has been assessed. The direction of their transport, from the inner to the outer membrane leaflet, supports their classification as floppases (Linton and Higgins 1998; Romsicki and Sharom 2001; Smit et al. 1993; S. L. Huang and MacDonald 2004). Among the ABC transporters, proteins with flippase activity have been also

recognized (Quazi and Molday 2013). An example of prokaryotic ABC transporter with floppase activity is the protein MsbA from *Escherichia coli*, whose function is to transport lipid A from the cytoplasm to the periplasm (K. Zhou et al. 1998).

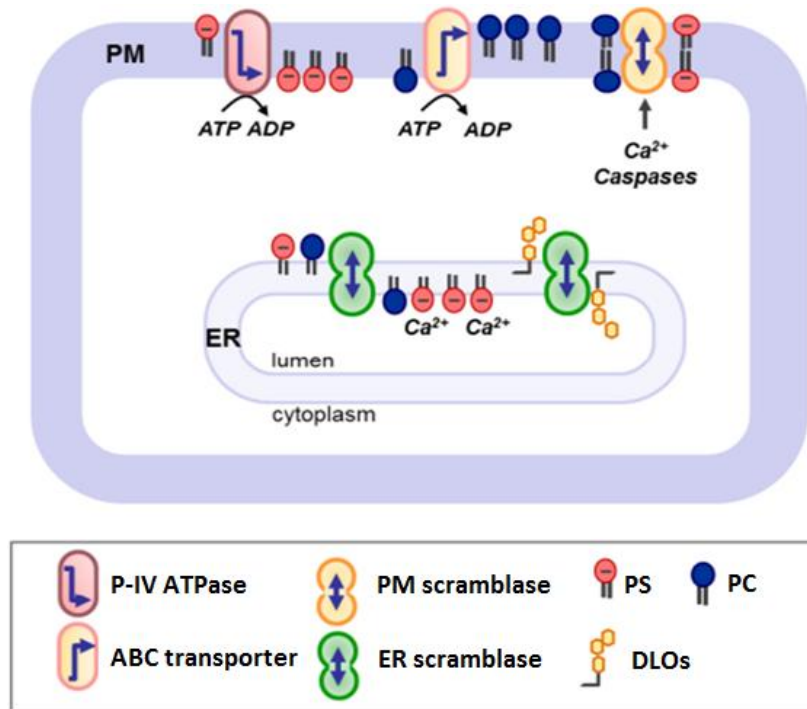


Figure 1.7: Proteins involved in lipid translocation: P-IV ATPases, ABC transporters and scramblases. P-IV ATPases catalyze the inward translocation of phospholipids while ABC transporters act in the opposite direction, both consuming ATP. Scramblases catalyze translocation according to gradient in both directions. PS: phosphatidylserine; PC: phosphatidylcholine; DLOs: lipid-linked oligosaccharides. Image from from Pomorki T. G. and Menon A. K. (2016), *Progress in lipid research* (T. G. Pomorski and Menon 2016).

REFERENCES

- 1) Abbott, Geoffrey. 2006. "Molecular Mechanisms of Cardiac Voltage-Gated Potassium Channelopathies." *Current Pharmaceutical Design* 12(28): 3631–44.
- 2) Airola, M. V., and Y. A. Hannun. 2013. "Sphingolipid Metabolism and Neutral Sphingomyelinases Michael." *Handbook of experimental pharmacology* 215: 57–76.
- 3) Alagumuthu, Manikandan, Divakar Dahiya, and Poonam Singh Nigam. 2019. "Phospholipid—the Dynamic Structure between Living and Non-Living World; a Much Obligatory Supramolecule for Present and Future." *AIMS Molecular Science* 6(1): 1–19.
- 4) Alder-Baerens, Nele et al. 2006. "Loss of P4 ATPases Drs2p and Dnf3p Disrupts Aminophospholipid Transport and Asymmetry in Yeast Post-Golgi Secretory Vesicles." *Molecular Biology of the Cell* 17: 1632–42.
- 5) Bevers, Edouard M., and Patrick L. Williamson. 2010. "Phospholipid Scramblase: An Update." *FEBS Letters* 584(13): 2724–30.
- 6) Buda, Csaba et al. 1994. "Structural Order of Membranes and Composition of Phospholipids in Fish Brain Cells during Thermal Acclimatization." *Proceedings of the National Academy of Sciences of the United States of America* 91(17): 8234–38.
- 7) Burdge, G. C., and P. C. Calder. 2015. "Introduction to Fatty Acids and Lipids." *World Rev Nutr Diet.* 112: 1–16.
- 8) Chambers, Henry F. 2003. "Solving Staphylococcal Resistance to β -Lactams." *Trends in Microbiology* 11(4): 145–48.
- 9) Chauhan, N, L Farine, and P Bütikofer. 2017. "Lipid Toponogenesis - 35 Years On." *Biochimica et Biophysica Acta* 1861(8): 757–66.
- 10) Deisenhofer, J. et al. 1985. "Structure of the Protein Subunits in the Photosynthetic Reaction Centre of Rhodospseudomonas Viridis at 3 Resolution." *Nature* 318(6047): 618–24.
- 11) Devaux, Philippe F. 1991. "Static and Dynamic Lipid Asymmetry in Cell Membranes." *Biochemistry* 30(5): 1163–73.
- 12) Ding, Wei, Michail Palaiokostas, Wen Wang, and Mario Orsi. 2015. "Effects of Lipid Composition on Bilayer Membranes Quantified by All-Atom Molecular Dynamics." *Journal of Physical Chemistry B* 119(49): 15263–74.
- 13) Dowhan, William, Eugenia Mileykovskaya, and Mikhail Bogdanov. 2004. "Diversity and Versatility of Lipid-Protein Interactions Revealed by Molecular Genetic Approaches." *Biochimica et Biophysica Acta - Biomembranes* 1666(1–2): 19–39.
- 14) Drew, David et al. 2005. "A Scalable, GFP-Based Pipeline for Membrane Protein Overexpression Screening and Purification." *Protein Science* 14(8): 2011–17.
- 15) Drew, David et al. 2008. "GFP-Based Optimization Scheme for the Overexpression and Purification of Eukaryotic Membrane Proteins in *Saccharomyces Cerevisiae*." *Nature Protocols* 3(5): 784–98.
- 16) Fattal, D. R., and A. Ben-Shaul. 1993. "A Molecular Model for Lipid-Protein Interaction in Membranes:

Chapter 1

- The Role of Hydrophobic Mismatch." *Biophysical Journal* 65(5): 1795–1809.
- 17) Fu, L. M., and C. S. Fu-Liu. 2002. "Is Mycobacterium Tuberculosis a Closer Relative to Gram-Positive or Gram-Negative Bacterial Pathogens?" *Tuberculosis* 82(2–3): 85–90.
 - 18) Fuglebakk, Edvin, and Nathalie Reuter. 2018. "A Model for Hydrophobic Protrusions on Peripheral Membrane Proteins." *PLoS Computational Biology* 14(7): 1–26.
 - 19) Gargus, J. Jay. 2006. "Ion Channel Functional Candidate Genes in Multigenic Neuropsychiatric Disease." *Biological Psychiatry* 60(2): 177–85.
 - 20) Glauert, A M, and M J Thornley. 1969. "The Topography of the Bacterial Cell Wall." *Annual Review of Microbiology* 23(1): 159–98.
 - 21) Gram, H. C. J. 1884. "Ueber Die Isolirte Färbung Der Schizomyceten in Schnitt- Und Trockenpräparaten." *Fortschritte der Medizin* 2: 185–89.
 - 22) Hays, Franklin A., Zygy Roe-Zurz, and Robert M. Stroud. 2010. "Overexpression and Purification of Integral Membrane Proteins in Yeast." *Methods in Enzymology* 470(C): 695–707.
 - 23) Hedger, George, and Mark S.P. Sansom. 2016. "Lipid Interaction Sites on Channels, Transporters and Receptors: Recent Insights from Molecular Dynamics Simulations." *Biochimica et Biophysica Acta - Biomembranes* 1858(10): 2390–2400.
 - 24) Von Heijne, G. 2007. "The Membrane Protein Universe: What's out There and Why Bother?" *Journal of Internal Medicine* 261(6): 543–57.
 - 25) Von Heijne, Gunnar. 1989. "Control of Topology and Mode of Assembly of a Polytopic Membrane Protein by Positively Charged Residues." *Nature* 341(6241): 456–58.
 - 26) Von Heijne, Gunnar. 1996. "Principles of Membrane Protein Assembly and Structure." *Progress in Biophysics and Molecular Biology* 66(2): 113–39.
 - 27) Von Heijne, Gunnar. 2000. "Recent Advances in the Understanding of Membrane Protein Assembly and Structure." *Quarterly Reviews of Biophysics* 32(4): 285–307.
 - 28) Heppel, Leon A. 1967. "Selective Release of Enzymes from Bacteria." *Science* 156(3781): 1451–55.
 - 29) Higgins, J. A., and W. H. Evans. 1978. "Transverse Organization of Phospholipids across the Bilayer of Plasma-Membrane Subfractions of Rat Hepatocytes." *Biochemical Journal* 174(2): 563–67.
 - 30) van Hoogevest, Peter, and Armin Wendel. 2014. "The Use of Natural and Synthetic Phospholipids as Pharmaceutical Excipients." *European Journal of Lipid Science and Technology* 116(9): 1088–1107.
 - 31) Howard, Mark J. 1998. "Protein NMR Spectroscopy." *Current Biology* 8(10): 331–33.
 - 32) Huang, Shao Ling, and Robert C. MacDonald. 2004. "Acoustically Active Liposomes for Drug Encapsulation and Ultrasound-Triggered Release." *Biochimica et Biophysica Acta - Biomembranes* 1665(1–2): 134–41.
 - 33) Ivankin, Andrey, Ivan Kuzmenko, and David Gidalevitz. 2010. "Cholesterol-Phospholipid Interactions: New Insights from Surface X-Ray Scattering Data." *Physical review letters* 104(10): 1–9.
 - 34) Jackson, Catherine L., Laurence Walch, and Jean Marc Verbavatz. 2016. "Lipids and Their Trafficking: An Integral Part of Cellular Organization." *Developmental Cell* 39(2): 139–53.
 - 35) Jain, Amrita, and Joost C.M. Holthuis. 2017. "Membrane Contact Sites, Ancient and Central Hubs of Cellular Lipid Logistics." *Biochimica et Biophysica Acta - Molecular Cell Research* 1864(9): 1450–58.

Chapter 1

- 36) Janetopoulos, Christopher, and Richard A. Firtel. 2008. "Directional Sensing During Chemotaxis." *FEBS Letters* 582(14): 2075–85.
- 37) Kamio, Yoshiyuki, and Hiroshi Nikaido. 1976. "Outer Membrane of Salmonella Typhimurium: Accessibility of Phospholipid Head Groups to Phospholipase C and Cyanogen Bromide Activated Dextran in the External Medium." *Biochemistry* 15(12): 2561–70.
- 38) Kholodenko, Boris N., John F. Hancock, and Walter Kolch. 2010. "Signalling Ballet in Space and Time." *Nature Reviews Molecular Cell Biology* 11(6): 414–26.
- 39) King, Landon S., David Kozono, and Peter Agre. 2004. "From Structure to Disease: The Evolving Tale of Aquaporin Biology." *Nature Reviews Molecular Cell Biology* 5(9): 687–98.
- 40) Kirchhausen, Tomas. 2000. "Three Ways to Make a Vesicle." *Nature Reviews Molecular Cell Biology* 1(3): 187–98.
- 41) Kornberg, Roger D., and Harden M. McConnell. 1971. "Inside-Outside Transitions of Phospholipids in Vesicle Membranes." *Biochemistry* 10(7): 1111–20.
- 42) Krause, Martin R., and Steven L. Regen. 2014. "The Structural Role of Cholesterol in Cell Membranes: From Condensed Bilayers to Lipid Rafts." *Accounts of Chemical Research* 47(12): 3512–21.
- 43) Krogh, Anders, Björn Larsson, Gunnar Von Heijne, and Erik L.L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305(3): 567–80.
- 44) de la Serna, Jorge Bernardino, Gerhard J. Schütz, Christian Eggeling, and Marek Cebecauer. 2016. "There Is No Simple Model of the Plasma Membrane Organization." *Frontiers in Cell and Developmental Biology* 4(article 106): 1–17.
- 45) Lee, A. G. 2003. 1612 *Biochimica et Biophysica Acta - Biomembranes Lipid-Protein Interactions in Biological Membranes: A Structural Perspective*.
- 46) Lee, Anthony G. 2004. "How Lipids Affect the Activities of Integral Membrane Proteins." *Biochimica et Biophysica Acta - Biomembranes* 1666(1–2): 62–87.
- 47) Lillemeier, Björn F et al. 2010. "TCR and Lat Are Expressed on Separate Protein Islands on T Cell Membranes and Concatenate during Activation." *Nature Immunology* 11(1): 90–96.
- 48) Lin, Qingqing, and Erwin London. 2014. "The Influence of Natural Lipid Asymmetry upon the Conformation of a Membrane-Inserted Protein (Perfringolysin O)." *Journal of Biological Chemistry* 289(9): 5467–78.
- 49) Linton, Kenneth J, and Christopher F Higgins. 1998. "MicroGenomics The Escherichia Coli ATP-Binding Cassette (ABC) Proteins." *molecular microbiology* 28(1): 5–13.
- 50) Los, Dmitry A., and Norio Murata. 2004. "Membrane Fluidity and Its Roles in the Perception of Environmental Signals." *Biochimica et Biophysica Acta - Biomembranes* 1666(1–2): 142–57.
- 51) Luby-Phelps, K. et al. 1993. "A Novel Fluorescence Ratiometric Method Confirms the Low Solvent Viscosity of the Cytoplasm." *Biophysical Journal* 65(1): 236–42.
- 52) Marquardt, Drew, Barbara Geier, and Georg Pabst. 2015. "Asymmetric Lipid Membranes: Towards More Realistic Model Systems." *Membranes* 5(2): 180–96.
- 53) Mathieu, Khadija et al. 2019. "Functionality of Membrane Proteins Overexpressed and Purified from E.

Chapter 1

- Coli Is Highly Dependent upon the Strain." *Scientific Reports* 9(1): 1–15.
- 54) McLean, L. R., and M. C. Phillips. 1981. "Mechanism of Cholesterol and Phosphatidylcholine Exchange or Transfer between Unilamellar Vesicles." *Biochemistry* 20(10): 2893–2900.
- 55) Van Meer, Gerrit, Dennis R. Voelker, and Gerald W. Feigenson. 2008. "Membrane Lipids: Where They Are and How They Behave." *Nature Reviews Molecular Cell Biology* 9(2): 112–24.
- 56) Michou, Myrsini, Charalampos Kapsalis, Christos Pliotas, and Georgios Skretas. 2019. "Optimization of Recombinant Membrane Protein Production in the Engineered Escherichia Coli Strains SuptoxD and SuptoxR ." *ACS Synthetic Biology* 8(7): 1631–41.
- 57) Moraes, Isabel et al. 2014. "Membrane Protein Structure Determination - The next Generation." *Biochimica et Biophysica Acta - Biomembranes* 1838(1 PARTA): 78–87.
- 58) Morozova, Diana, Gernot Guigas, and Matthias Weiss. 2011. "Dynamic Structure Formation of Peripheral Membrane Proteins." *PLoS Computational Biology* 7(6).
- 59) Mouritsen, Ole G . 2005. "Life – as a Matter of Fat. The Emerging Science of Lipidomics." *Springer*.
- 60) Mullineaux, Conrad W., Anja Nenninger, Nicola Ray, and Colin Robinson. 2006. "Diffusion of Green Fluorescent Protein in Three Cell Environments in Escherichia Coli." *Journal of Bacteriology* 188(10): 3442–48.
- 61) Musatov, Andrej, and Erik Sedláč. 2017. "Role of Cardiolipin in Stability of Integral Membrane Proteins." *Biochimie* 142: 102–11.
- 62) Navarre, William W., and Olaf Schneewind. 1999. "Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope." *Microbiology and molecular biology reviews* 63(1): 174–229.
- 63) Neuhaus, Francis C, and James Baddiley. 2003. "A Continuum of Anionic Charge: Structures and Functions Of." *Microbiology* 67(4): 686–723.
- 64) Nickels, Jonathan D., Jeremy C. Smith, and Xiaolin Cheng. 2015. "Lateral Organization, Bilayer Asymmetry, and Inter-Leaflet Coupling of Biological Membranes." *Chemistry and Physics of Lipids* 192: 87–99.
- 65) Nipper, Matthew E. et al. 2008. "Characterization of Changes in the Viscosity of Lipid Membranes with the Molecular Rotor FCVJ." *Biochimica et Biophysica Acta - Biomembranes* 1778(4): 1148–53.
- 66) Nwanochie, Emeka, and Vladimir N. Uversky. 2019. "Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) Is the Limit." *International Journal of Molecular Sciences* 20(17): 4186.
- 67) Parker, Joanne L, and Simon Newstead. 2016. "The Next Generation in Membrane Protein Structure Determination." *Advances in Experimental Medicine and Biology* 922: 61–72.
- 68) Phillips, Rob, Tristan Ursell, Paul Wiggins, and Pierre Sens. 2009. "Emerging Roles for Lipids in Shaping Membrane-Protein Function." *Nature* 459(7245): 379–85.
- 69) Pike, Linda J. 2006. "Rafts Defined: A Report on the Keystone Symposium on Lipid Rafts and Cell Function." *Journal of Lipid Research* 47(7): 1597–98.
- 70) Poggi, Paola et al. 2015. "Membrane Fatty Acid Heterogeneity of Leukocyte Classes Is Altered during in Vitro Cultivation but Can Be Restored with Ad-Hoc Lipid Supplementation." *Lipids in Health and*

Chapter 1

- Disease* 14(165): 1–13.
- 71) Pomorski, Thomas Günther, and K. Anant Menon. 2016. "Lipid Somersaults: Uncovering the Mechanism of Protein-Mediated Lipid Flipping." *Progress in Lipid Research* 64: 69–84.
 - 72) Quazi, Faraz, and Robert S. Molday. 2013. "Differential Phospholipid Substrates and Directional Transport by ATP-Binding Cassette Proteins ABCA1, ABCA7, and ABCA4 and Disease-Causing Mutants." *Journal of Biological Chemistry* 288(48): 34414–26.
 - 73) Quinn, Peter J. 2002. "Plasma membrane phospholipid asymmetry." *Sub-cellular Biochemistry* 36: 39–60.
 - 74) Raetz, Christian R. H., and Chris Whitfield. 2002. "Lipopolysaccharide Endotoxins." *Annual Review of Biochemistry* 71(1): 635–700.
 - 75) Rivel, Timothée, Ramseyer, Christophe and Yesylevskyy, Semen. 2019. "The asymmetry of plasma membranes and their cholesterol content influence the uptake of cisplatin." *Scientific Reports* 9: 1-14.
 - 76) Romsicki, Y., and F. J. Sharom. 2001. "Phospholipid Flippase Activity of the Reconstituted P-Glycoprotein Multidrug Transporter." *Biochemistry* 40(23): 6937–47.
 - 77) Sankaran, Krishnan, and Henry C Wus. 1994. "Lipid Modification of Bacterial Prolipoprotein." *The Journal of biological chemistry* 269(31): 19701–6.
 - 78) Sanyal, Sumana, and Anant K. Menon. 2009. "Flipping Lipids: Why an' What's the Reason For?" *ACS Chemical Biology* 4(11): 895–909.
 - 79) Schenk, Barbara, Fabiana Fernandez, and Charles J. Waechter. 2001. "The Ins(Ide) and Outs(Ide) of Dolichyl Phosphate Biosynthesis and Recycling in the Endoplasmic Reticulum." *Glycobiology* 11(5).
 - 80) Scott, June R., and Timothy C. Barnett. 2006. "Surface Proteins of Gram-Positive Bacteria and How They Get There." *Annual Review of Microbiology* 60(1): 397–423.
 - 81) Seddon, Annela M., Paul Curnow, and Paula J. Booth. 2004. "Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera." *Biochimica et Biophysica Acta - Biomembranes* 1666(1–2): 105–17.
 - 82) Seddon, John M. 1990. "Structure of the Inverted Hexagonal (HII) Phase, and Non-Lamellar Phase Transitions of Lipids." *BBA - Reviews on Biomembranes* 1031(1): 1–69.
 - 83) Silhavy, Thomas J, Daniel Kahne, and Suzanne Walker. 2010. "The Bacterial Cell Envelope." *Cold Spring Harbor Perspect Biol* 2: 1–16.
 - 84) Simmonds, A. C. et al. 1982. "Annular and Non-Annular Binding Sites on the (Ca²⁺ + Mg²⁺)-ATPase." *BBA - Biomembranes* 693(2): 398–406.
 - 85) Simons, Kai, and Elina Ikonen. 1997. "Functional Rafts in Cell Membranes." *Nature* 387(6633): 569–72.
 - 86) Simons, Kai, and Julio L. Sampaio. 2011. "Membrane Organization and Lipid Rafts." *Cold Spring Harbor Perspectives in Biology* 3(10): 1–17.
 - 87) Singer, S. J., and Garth L. Nicolson. 1972. "The Fluid Mosaic Model of the Structure of Cell Membranes." *Science* 175(4023): 720–31.
 - 88) Sleight, R. G., and R. E. Pagano. 1983. "Rapid Appearance of Newly Synthesized Phosphatidylethanolamine at the Plasma Membrane." *Journal of Biological Chemistry* 258(15): 9050–58.

Chapter 1

- 89) Smit, J. J. M. et al. 1993. "Homozygous Disruption of the Murine MDR2 P-Glycoprotein Gene Leads to a Complete Absence of Phospholipid from Bile and to Liver Disease." *Cell* 75(3): 451–62.
- 90) Suzuki, Jun et al. 2013. "Xk-Related Protein 8 and CED-8 Promote Phosphatidylserine Exposure in Apoptotic Cells." *Science* 341(6144): 403–6.
- 91) Suzuki, Jun, Masato Umeda, Peter J. Sims, and Shigekazu Nagata. 2010. "Calcium-Dependent Phospholipid Scrambling by TMEM16F." *Nature* 468(7325): 834–40.
- 92) Suzuki, Yasuyuki et al. 2001. "Genetic and Molecular Bases of Peroxisome Biogenesis Disorders." *Genetics in Medicine* 3(5): 372–76.
- 93) Tate, Christopher G., and Gebhard FX Schertler. 2009. "Engineering G Protein-Coupled Receptors to Facilitate Their Structure Determination." *Current Opinion in Structural Biology* 19(4): 386–95.
- 94) Thewalt, Jenifer L., and Myer Bloom. 1992. "Phosphatidylcholine: Cholesterol Phase Diagrams." *Biophysical Journal* 63(4): 1176–81.
- 95) Vance, Jean E. 2008. "Phosphatidylserine and Phosphatidylethanolamine in Mammalian Cells: Two Metabolically-Related Aminophospholipids." *Journal of Lipid Research* 49: 1377–87.
- 96) Vance, Jean E., and Guergana Tasseva. 2013. "Formation and Function of Phosphatidylserine and Phosphatidylethanolamine in Mammalian Cells." *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1831(3): 543–54.
- 97) Verkleij, A. J. et al. 1973. "The Asymmetric Distribution of Phospholipids in the Human Red Cell Membrane. A Combined Study Using Phospholipases and Freeze-Etch Electron Microscopy." *BBA - Biomembranes* 323(2): 178–93.
- 98) Vinothkumar, Kutti R., and Richard Henderson. 2010. 43 Quarterly Reviews of Biophysics *Structures of Membrane Proteins*.
- 99) Vollmer, Waldemar. 2008. "Structural Variation in the Glycan Strands of Bacterial Peptidoglycan." *FEMS Microbiology Reviews* 32(2): 287–306.
- 100) Vollmer, Waldemar, Didier Blanot, and Miguel A. De Pedro. 2008. "Peptidoglycan Structure and Architecture." *FEMS Microbiology Reviews* 32(2): 149–67.
- 101) Wallin, Erik, and Gunnar Von Heijne. 2008. "Genome-Wide Analysis of Integral Membrane Proteins from Eubacterial, Archaeal, and Eukaryotic Organisms." *Protein Science* 7(4): 1029–38.
- 102) Watson, Helen. 2015. "Biological Membranes." *Essays in Biochemistry* 59: 43–70.
- 103) Webb, Richard J., J. Malcolm East, Ram P. Sharma, and Anthony G. Lee. 1998. "Hydrophobic Mismatch and the Incorporation of Peptides into Lipid Bilayers: A Possible Mechanism for Retention in the Golgi." *Biochemistry* 37(2): 673–79.
- 104) White, Stephen H., Alexey S. Ladokhin, Sajith Jayasinghe, and Kalina Hristova. 2001. "How Membranes Shape Protein Structure." *Journal of Biological Chemistry* 276(35): 32395–98.
- 105) Williamson, J. J., and P. D. Olmsted. 2015. "Kinetics of Symmetry and Asymmetry in a Phase-Separating Bilayer Membrane." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 92(5): 1–17.
- 106) Wimley, William C. 2002. "Toward Genomic Identification of β -Barrel Membrane Proteins: Composition and Architecture of Known Structures." *Protein Science* 11(2): 301–12.
- 107) Wood, W. Gibson, Urule Igbavboa, Walter E. Müller, and Gunter P. Eckert. 2011. "Cholesterol

Chapter 1

- Asymmetry in Synaptic Plasma Membranes." *Journal of neurochemistry* 116(5): 684–89.
- 108) Yang, Sung Tae et al. 2016. "The Role of Cholesterol in Membrane Fusion." *Chemistry and Physics of Lipids* 199: 136–43.
- 109) Yang, Yanbo, Minhyoung Lee, and Gregory D. Fairn. 2018. "Phospholipid Subcellular Localization and Dynamics." *Journal of Biological Chemistry* 293(17): 6230–40.
- 110) Yeung, Tony et al. 2009. "Contribution of Phosphatidylserine to Membrane Surface Charge and Protein Targeting during Phagosome Maturation." *Journal of Cell Biology* 185(5): 917–28.
- 111) Zhou, Kemin et al. 1998. "Guanine Nucleotide Exchange Factors Regulate Specificity of Downstream Signaling from Rac and Cdc42." *Journal of Biological Chemistry* 273(27): 16782–86.
- 112) Zhou, Xiaoming, and Todd R. Graham. 2009. "Reconstitution of Phospholipid Translocase Activity with Purified Drs2p, a Type-IV P-Type ATPase from Budding Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 106(39): 16586–91.
- 113) De Zorzi, Rita, Wei Mi, Maofu Liao, and Thomas Walz. 2016. "Single-Particle Electron Microscopy in the Study of Membrane Protein Structure." *Microscopy (Oxford, England)* 65(1): 81–96.

Chapter 2

Neo1: an integral membrane protein belonging to the family of P-IV ATPases

Biological membranes are formed by a phospholipid bilayer where proteins, lipids and other components are inserted. Asymmetry of the lipid distribution between the two leaflets of the membrane is a pivotal characteristic of cellular membranes, fundamental for curvature and polarity of cell surface, vesiculation and intra-inter cellular signaling (Lenoir, Williamson, and Holthuis 2007). P-IV ATPases are integral membrane proteins present in eukaryotic cells and essential to maintain the asymmetric lipid distribution between different membrane leaflets. They act as ATP-powered translocases, moving specific lipids from the extracellular leaflet to the cytoplasmic one. These integral membrane proteins possess a large hydrophilic domain exposed towards the cytosol involved in ATP binding, whereas the hydrophobic regions are embedded in the cellular membrane and serve for substrate translocation (van der Mark, Oude Elferink, and Paulusma 2013). Substrates of P-IV ATPases are different phospholipids, mostly PC, PE or PS.

Since P-IV ATPases play a key role in establishing membrane asymmetry and curvature, they are involved also in vesicle-mediated protein transport in Golgi and endosomes (Sebastian et al. 2012). Mutations or expression defects of these proteins are linked to several human diseases, especially related to blood coagulation anomalies or hepatic cholestasis diseases (Andersen et al. 2016).

Considering their important role in cellular biology and the numerous issues caused by their malfunctioning, a deeper understanding of the structure and the flipping mechanism of P-IV ATPases is a long term goal of many research groups around the world. In this work, my study focused on a less-known member of the family of P-IV ATPases, the protein Neo1 from yeast *Saccharomices cerevisiae*. Neo1 stands out as an atypical P-IV ATPase, as it is the only yeast P-IV ATPase lacking a secondary subunit, required by other members of this family to correctly fold and to perform their activity (Takar, Wu, and Graham 2016). In addition, Neo1

specific substrates are unknown. The study of yeast proteins, simpler and easier to obtain, opens the way to understand mechanisms of action of human P-IV ATPases and sheds light on how their mutations influence the onset of different diseases.

In this study, I successfully obtained Neo1 from yeast in overexpression conditions, purified the protein in detergent, tested its stability and set up crystallization trials. In order to identify residues crucial for activity and substrate identification, I expressed and analyzed mutants of Neo1 in order to gain important information regarding its flipping mechanism.

INTRODUCTION

2.1 P-TYPE ATPASES

P-type ATPases are a large family of membrane proteins that transport substrates across membranes against their gradient. They share a common ATP-dependent catalytic mechanism that involves an intermediate where a conserved aspartic acid residue of the protein is phosphorylated (Palmgren and Nissen 2011; Pedersen, L. Peter and Carafoli 1987). Based on sequence similarity, P-ATPases are divided in 5 subfamilies, different for their function, substrate specificity and localization (Palmgren and Axelsen 1998):

- *P-I ATPases* are heavy metal transporters. Their role is to detoxify the cytoplasm and separate heavy metals in specific compartments.
- *P-II ATPases* are ion channels. The best known members are the Na^+/K^+ -ATPase, that maintains the electrochemical gradient across the plasma membrane, the H^+/K^+ -ATPase, responsible for stomach acidification, and the sarcoplasmic reticulum Ca^{2+} -ATPase, involved in muscle contraction.
- *P-III ATPases* are proton pumps. Localized on the plasma membrane, they extrude protons in order to generate a pH gradient.
- *P-IV ATPases* are responsible of an active unidirectional translocation of phospholipids from the extracellular to the cytoplasmic leaflet of membranes. They are present only in eukaryotic cells.

- *P-V ATPase* function is still unknown. Studies evidence the presence of this class, but to date little is known about their role (Schultheis et al. 2004; De La Hera et al. 2013).

2.1.1 Structure of P-type ATPases

The first low-resolution structure of a P-type ATPase was obtained in 1983 for the sarcoplasmic Ca^{2+} pump (SERCA) from rabbit skeletal sarcoplasmic reticulum vesicles, analyzed by negative-staining electron microscopy (K. A. Taylor, Dux, and Martonosi 1986). However, a high-resolution structure for this P-II ATPase is available only since 2000, when a 2.6 Å-resolution crystal structure was obtained by Toyoshima and co-workers (Toyoshima et al. 2000) (**Figure 2.1**). The structure shows a Transmembrane Domain (TM) of 10 helices (TM1 to TM10) and three cytoplasmic domains: (1) the Nucleotide Binding Domain, or N domain, formed by the first segment of a large loop between TM4 and TM5; (2) the Phosphorylation Domain, or P domain, formed by the second segment of the loop between TM4 and TM5 and containing a conserved motif (DKTGT) that undergoes phosphorylation on the aspartic acid residue by an ATP molecule; and (3) the Actuator domain, or A domain, formed by the N-terminal segment of the protein and the loop between TM2 and TM3 and involved in the conformational change that triggers substrate translocation. In Toyoshima's structure, two Ca^{2+} ions are visible in the TM domain. This result, together with sequence alignments with other P-ATPases, suggested that the first six helices of the TM domain are involved in substrate recognition, showing the highest sequence variability (Bublitz, Morth, and Nissen 2011). Other, smaller domains are present on the N- and C-termini of some P-type ATPases, with regulatory or localization functions (Chalat et al. 2017; Malmström, Åkerlund, and Askerlund 2000).

In addition to the main polypeptide chain, also known as α -subunit, most P-type ATPases require a non-catalytic β -subunit, belonging to the CDC50/LEM3 family. To date it is not clear if this subunit is necessary for folding, localization or it is involved in the translocation mechanism. β -subunits have two TM domains linked by a highly glycosylated extracellular loop (**Figure 2.2**).

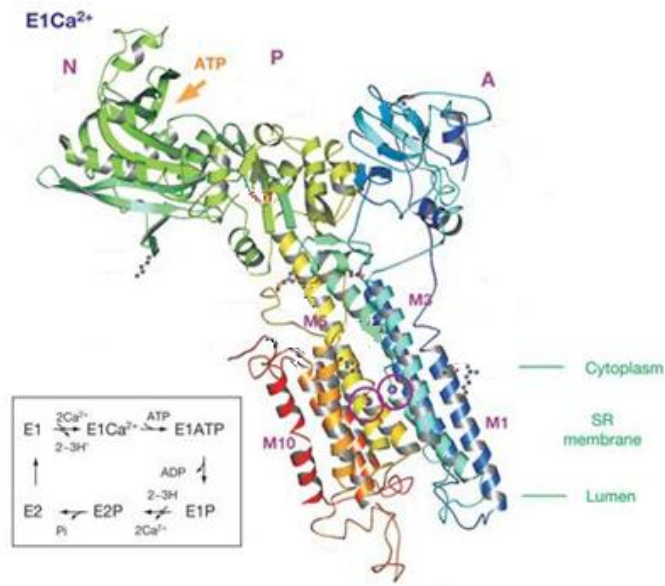


Figure 2.1: First crystal structure of a P-ATPase, the Ca²⁺ pump SERCA, at 2.6 Å resolution. Two Ca²⁺ ions bound to the TM domain are visible (circled). From Toyoshima et al. (2000), Nature (Toyoshima et al. 2000).

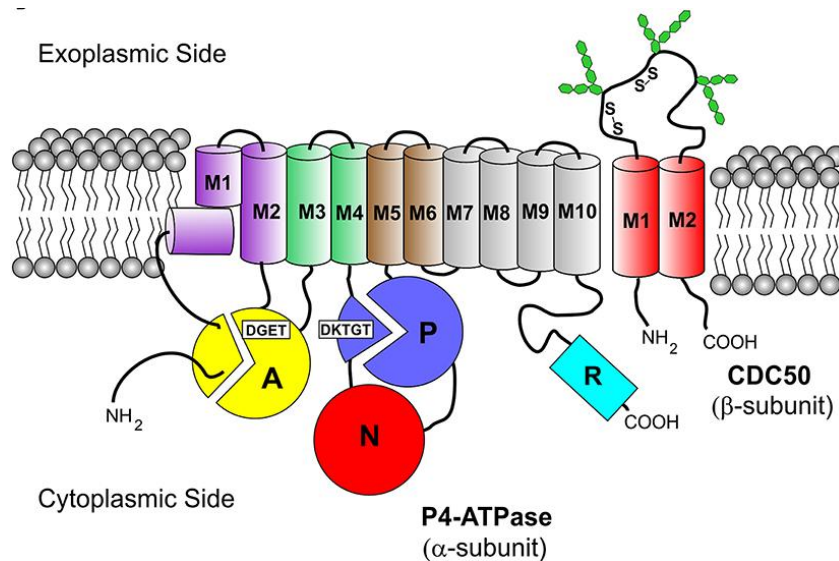


Figure 2.2: Domain distribution of P-ATPases. Actuator domain in yellow with the conserved DGET sequence for dephosphorylation (further described as LXGEX); Phosphorylation domain in blue with the conserved DKTGT sequence for phosphorylation; Nucleotide-binding domain in red; TM domain with helices in light purple (M1-M2), light green (M3-M4), brown (M5-M6) and grey (M7-M10); β subunit in light red, with glycosilation in green. From Andersen J. P. et al. (2016), Frontiers in Physiology (Andersen et al. 2016).

A second SERCA structure was obtained in 2002 by Cryo-EM (C. Xu et al. 2002), for the Ca^{2+} -free state. Since they represent different stages of the catalytic cycle, the two structures allowed to formulate the first hypothesis for the mechanism of action of P-type ATPases.

In the following years, structures of members of the other classes of P-type ATPases became available from both electron microscopy and crystallographic studies. Considering the essential roles of these proteins, new crystallographic studies are devoted to understand the mechanism and possibly improve inhibitors already widely used in medicine such as omeprazole, that targets the gastric H^+/K^+ pump, or the cardiac glycosides, that target the Na^+/K^+ pump (Yatime et al. 2009). Until 2019, however, no structures of P-IV ATPases were available, leaving open the crucial questions regarding their amphipathic substrate, considerably larger than ions transported by the other members of the family.

2.1.2 Catalytic cycle of P-type ATPases

All P-ATPases undergo a phosphorylation step and a dephosphorylation step that follow the Post-Albert scheme (**Figure 2.3**). In the *apo* state of the protein, the N domain binds an ATP molecule with high affinity (E1 state). This bond triggers a nucleophilic attack that results in the phosphorylation of the protein by ATP at a conserved aspartic acid residue in the P domain, leading to a high-energy intermediate (E1P). The classical Post-Albert E1-E2 scheme requires a strong binding of the E1 state with ATP and a less stable ADP-bound state (Jencks 1989), allowing the release of the nucleotide from the E1-P protein. A conformational change occurring after phosphorylation leads to the E2-P state, with lower affinity for the substrate bound to the TM domain. The substrate is therefore released on the opposite side of the membrane. If the protein is able to translocate different ions, as in the Na^+/K^+ pump, the second ion species binds to the E2-P state of the protein. Due to the conformational change occurring after substrate translocation, a water molecule bound to the A domain becomes available for the hydrolysis of the E2-P intermediate. When two substrates are involved, the dephosphorylated E2 state has a lower affinity for the second substrate, releasing it on the opposite side of the membrane. The rearrangement following dephosphorylation leads to the initial E1 conformation of the protein, ready for another cycle. Different P-ATPase proteins are able to translocate against their gradients one or two substrates, and a variable number of ions for each substrate, from 1 to 3 (Kühlbrandt 2004).

An important role in the catalytic cycle is played by a Mg^{2+} ion that functions as cofactor to form all the electrostatic interactions required for the stabilization of the phosphorylated intermediate (Bublitz et al. 2010). Interestingly, all steps are reversible and a P-ATPase could theoretically produce ATP using the membrane gradient of its specific substrate (Kühlbrandt 2004).

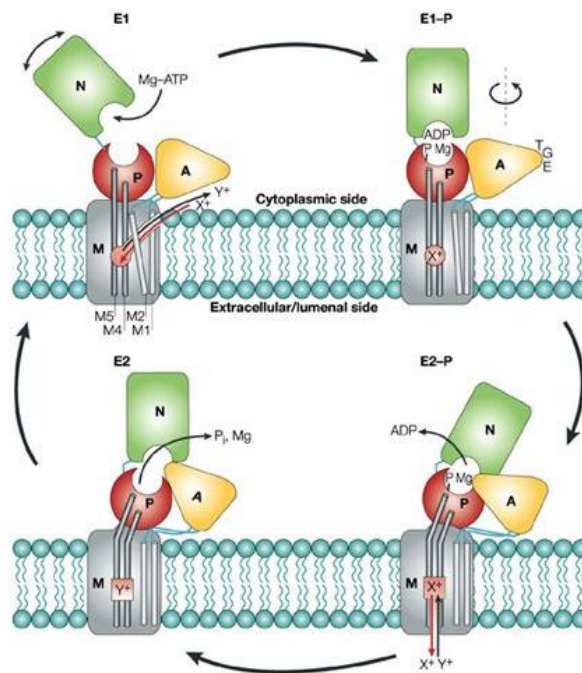


Figure 2.3: The Post-Albert cycle for P-ATPases. E1: initial state of the protein, apo with respect to ATP binding, but with a substrate molecule/ion interacting with the protein in the TM domain; E1-P: phosphorylated state of the protein in which the actuator domain undergoes a rearrangement; E2-P: ion release on the opposite side of the membrane; E2: dephosphorylated state of the protein with eventual binding of a second ion. From Kühlbrandt W. (2004), *Nature Reviews Molecular Cell Biology* (Kühlbrandt, 2004.)

In general, all P-type ATPases share a good homology, especially as regards the cytoplasmic domains, with some sequences particularly conserved in different species and classes. Among these, the DKTGT motif is present in all P-ATPases and it includes the aspartic acid residue phosphorylated during transport and the adjacent lysine residue that plays a coadjuvant role in this mechanism. The LXGEX motif of the A domain is also conserved among P-ATPases and its glutamic acid residue is involved in the dephosphorylation mechanism of the protein.

2.2 P-IV ATPASES

Translocation of aminophospholipids in presence of ATP was first discovered in bovine chromaffin granules, in 1988, and in red blood cells, in 1990 (Moriyama and Nelson 1988; Morrot, Zachowski, and Devaux 1990; Zimmerman and Daleke 1993). In 1994 one of the proteins responsible for this activity was reconstituted and studied in proteoliposomes (Auland et al. 1994). Later, this protein has been identified as a P-IV ATPase (now known as ATP8A1) and experiments proved its ability to translocate PS - and a lower activity for PE - from the exoplasm to the cytoplasm (J. Ding et al. 2000). Further studies followed the activity of these proteins using labeled phospholipids, analyzing the changes in membrane curvature or on detergent-solubilized and purified samples of P-IV ATPases (Roland and Graham 2016b).

Today, the role of P-IV ATPases in establishing and maintaining membrane asymmetry, together with the symmetric floppase activity of ABC transporters, is well-known. Due to their important function, P-IV ATPases are involved in vesicle-mediated protein transport in the Golgi compartment and in endosomes (Sebastian et al. 2012). Unfortunately, despite the efforts, many aspects of the structure and activity of this class of proteins are still unknown. As previously described, P-I, P-II and P-III ATPase are channels able to transport ions or metals across membranes, while the preferred substrates of P-IV ATPases are big amphipathic molecules. As expected, the hydrophilic channel present in the TM domain of other P-type ATPases is not conserved in P-IV ATPases, as it would not be suitable to host phospholipids. The large dimension of substrates of P-IV ATPases opens the so-called “giant substrate problem” (Roland and Graham 2016b). The class of P-IV ATPases is further divided in 7 subclasses (1a, 1b, 2, 3, 4, 5 and 6) according to sequence homology, function and localization (van der Mark, Oude Elferink, and Paulusma 2013).

In the proposed translocation mechanism for this class, known as “credit card” model, only the polar head of the phospholipid interacts with the TM domain of the protein, while the hydrophobic acyl chains remain in the lipophilic part of the membrane (**Figure 2.4**). The head group of the substrate binds to the TM domain of the P-IV ATPase, in the extracellular region, and is moved through the bilayer toward the cytoplasmic region. In this model, the acyl chains of the phospholipid follow the head group without direct interaction with a specific channel. As in other P-type ATPase, however, the phosphorylation and dephosphorylation triggers follow the same mechanistic route (**Figure 2.4a**). In the

conformational change occurring after phosphorylation, an affinity region for lipids, located in the TM domain, is exposed to the extracellular side favoring substrate binding. For this transition, a direct intervention of the β -subunit has been hypothesized to confer stability to the complex. However, this idea seems to be contradicted by the presence of P-IV ATPases that do not require any β -subunit (mainly subclass 2) (Takatsu et al. 2011). As in other ATPases, a water molecule is required for dephosphorylation and transition of the protein to a state with lower substrate affinity, leading to the release of the lipid in the cytoplasmic side of the membrane (van der Mark, Oude Elferink, and Paulusma 2013).

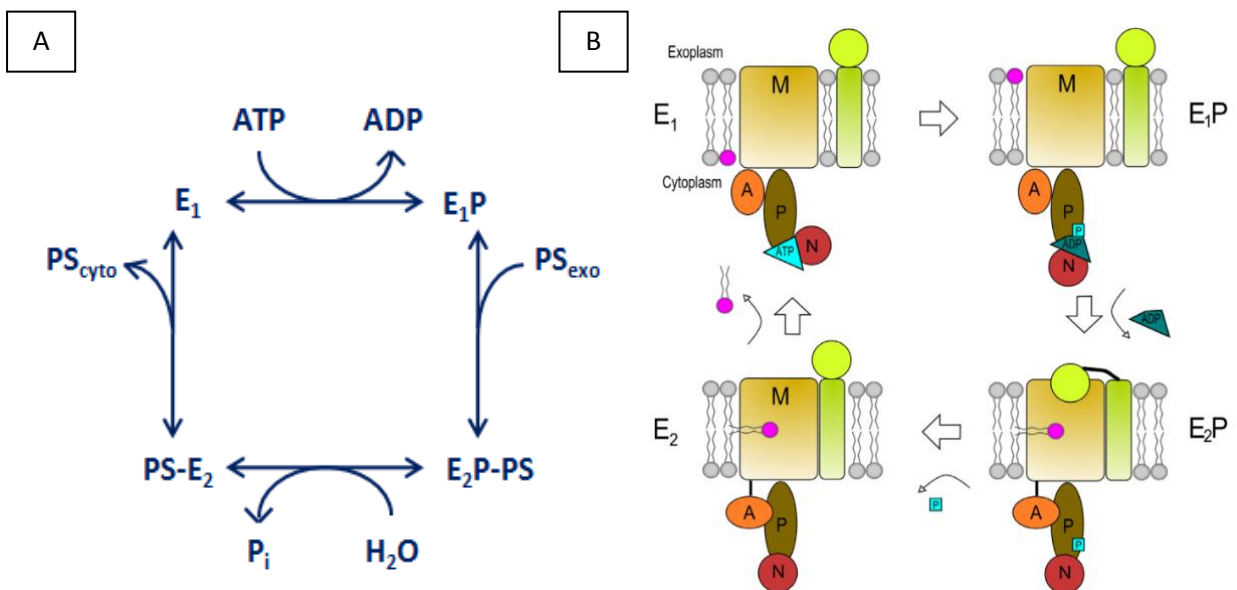


Figure 2.4: **A)** Schematic representation of the reaction cycle of a P-IV ATPase. **B)** The Post-Albert cycle for P-IV ATPases: the E_1 conformation of the protein has a phospholipid substrate bound to the TM domain; due to the high-affinity for ATP of the E_1 conformation, an ATP molecule binds to the N domain; phosphorylation of the P domain leads to the E_1P conformation; the release of ADP and a rearrangement of the TM domain allows phospholipid translocation, in the E_2P state; in the E_2 state, dephosphorylation of the P domain allows release of the phospholipid to the outer side of the membrane. From van der Mark V. A. et al. (2013), *International Journal of Molecular Sciences* (van der Mark, Oude Elferink, and Paulusma 2013).

As regards directionality of the translocation and substrate selectivity, two complementary models have been proposed (**Figure 2.5 a and b**):

- *Hydrophobic gate model*: in this model, hydrophobic residues in the middle of TM4 move in or out of the substrate channel as the conformation of the enzyme changes throughout the catalytic cycle, restricting or allowing the passage of the substrate.

During the conformational changes, the cytoplasmic side or the extracellular side of the channel become alternatively available for water penetration from opposite sides of the membrane. This model provides a good explanation for the unidirectional phospholipid transport but does not account for substrate specificity (Roland and Graham 2016b; Jensen et al. 2017).

- *Two-gate model*: according to this model, the substrate binds at the “entry gate”, located in the extracellular region of the TM domain, and is transferred to the “exit gate”, on the opposite side of the protein, along TM4, passing through a groove formed by TM1 and TM2 segments. This model highlights the presence of different residues at the entry and exit gates responsible for substrate selection, such as asparagine residues (N220 and N550) in the exit gate of yeast P-IV ATPase Dnf1. However, it lacks experimental support and does not explain the directionality of the phospholipid flipping (Roland et al. 2018; Roland and Graham 2016a).

A third model was proposed, the *central cavity model*, but received less attention due to the lack of experimental evidence (**Figure 2.5c**). The basic idea of this model was developed in analogy to H⁺-ATPases crystal structure and predicts the presence of a water-filled cavity between TM4, TM5 and TM6 that would contain the headgroup of the substrate during the phospholipid translocation (Jensen et al. 2017).

In addition to the canonical A, N, P and TM- domains, several ATP-ases possess a regulatory domain, or R domain, located in the cytoplasm. In some P-IV ATPases, this domain is formed by the C-terminal region of the protein (X. Zhou, Sebastian, and Graham 2013) and works as an autoinhibitory tail hampering binding of ATP molecules. When the protein is activated, the C-terminal tail changes position and exposes the N and P domains, allowing phosphorylation and subsequent phospholipid translocation.

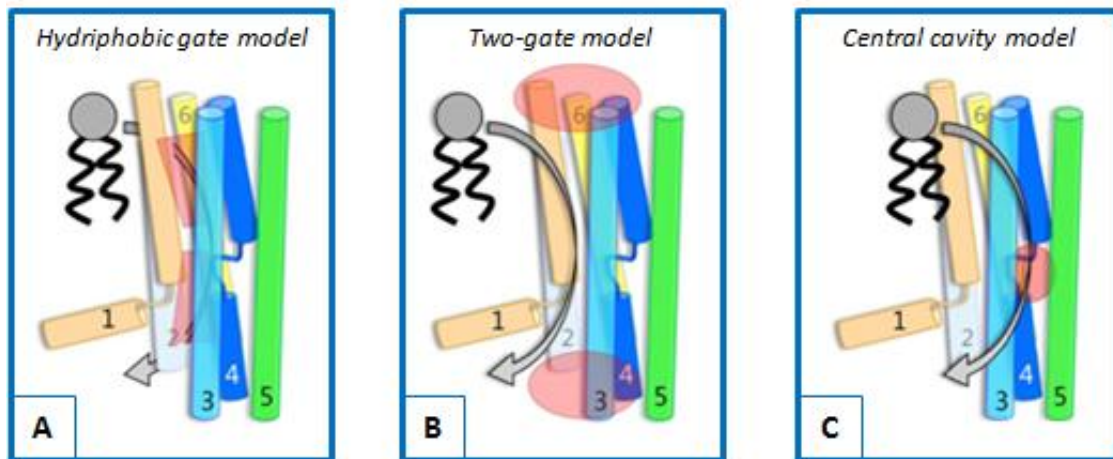


Figure 2.5: Proposed mechanisms for phospholipid translocation. A) Hydrophobic gate model: two water channels (red bars) are alternatively opened during phospholipid translocation; B) Two-gate model: critical residues (red ellipsoids) on extracellular and cytoplasmic sides of the TM domain select phospholipid headgroup, forming an entry- and an exit-gate; C) Central cavity model: a water-filled cavity (red circle) contains the substrate headgroup during translocation. From Jensen M. S. et al. (2017), *Scientific Reports* (Jensen et al. 2017).

Little is known about the regulation of these proteins, even though there is evidence for kinase-dependent phosphorylation (Roelants et al. 2010) and some sort of regulation is essential to manage the delicate cellular processes, such as vesiculation, in which these proteins are involved.

In vesiculation processes, for example, the role of P-IV ATPases is far from clear. It is possible that P-IV ATPases force the inward translocation of phospholipids, resulting in a numeric disequilibrium between leaflets that triggers an inward bending of the membrane and the subsequent vesicle formation. Another hypothesis is that P-IV ATPase activity confers to the membrane surface the characteristics that favor the recruitment of proteins involved in vesicle budding, such as clathrin or ADP-ribosylation factors. These mechanisms are likely to concur to the formation of vesicles (Lopez-Marques et al. 2014). In addition, experimental evidence on yeast mutants suggests that some P-IV ATPases may be involved in vesiculation pathways not only through their translocation function, but also as scaffolds for vesicle trafficking or cellular signaling (Van Der Velden, Van De Graaf, and Klomp 2010).

2.2.1 *Cryo-Electron Microscopy structures of P-IV ATPases*

Recently, a considerable step forward in understanding the mechanisms of action of P-IV ATPases has been made thanks to Cryo-EM structural studies that yielded the first structure of a P-IV ATPase, the protein Drs2 from *S. cerevisiae* in complex with its Cdc50 subunit (**Figure 2.6a**) (Timcenko et al. 2019). In this work, three different conformations were captured, consistent with an auto-inhibited, an intermediate phosphatidylinositol 4-phosphate (PI4P)-bound state and an outward-open fully active state, at 2.8 Å, 3.7 Å and 2.9 Å resolution, respectively. PI4P is a molecule found in many membranes with function of protein ligand or molecular precursor in intracellular signaling events (Gil de Rubio et al. 2018). Interestingly, this molecule is fundamental for Drs2 activity (Azouaoui et al. 2017).

The comparison of the three conformations highlights the rearrangement of transmembrane helices involved in lipid translocation. In the outward-open conformation, a movement of TM1 and TM2 (**in blue and pink in Figure 2.6a**) exposes the unwound segment of TM4 (**in light green in Figure 2.6a**) to the luminal membrane leaflet. The conserved PISL motif, located after TM4 in all P-IV ATPases, has a key role in phospholipid transport. The transmembrane rearrangement observed in the active state lines up helices TM1, TM2 and TM6 (**in brown in Figure 2.6a**), creating an “entry gate” for the lipid-transport path. In fact, this cleft contains residues already identified as important for P-IV substrate selection such as glutamine and isoleucine residues on TM1 and TM4, respectively.

Moreover, in these structures the interaction between α - and β -subunits is clearly visible. In the TM domain, the two helices of Cdc50 (**in magenta in Figure 2.6a**) interact with TM10 of the P-IV ATPase, while in the lumen the Cdc50 ectodomain interacts with the luminal loops of the α -subunit. The N-terminus of Cdc50 extends in the cytosol towards the final segment of TM4, in close proximity to the phosphorylation site. The auto-inhibited conformation clarifies the role of the C-terminus of Drs2: when interacting with the P-domain, it overlaps with the nucleotide binding site through its conserved GFAFS motif. The structure confirms that the terminal segment of Drs2 acts through an allosteric autoinhibitory mechanism that locks the conformation of the cytosolic domains. In the structure of the intermediate conformation, the interaction of PI4P with helices TM7, TM8 and TM10 induces a concerted movement of TM10 and the TM domains of Cdc50 that dissipates the auto-inhibitory interaction of Drs2 C-terminus with its P-domain.

Two months later, a new cryo-EM study was published, including six distinct structures of the human ATP8A1 P-IV ATPase in complex with its β -subunit CDC50a (**Figure 2.6b 1 and 2**), at 2.6-3.3 Å resolution (Hiraizumi et al. 2019). This ensemble of structures covers all the states in the phospholipid transport cycle. Their comparison highlights the greater flexibility of the cytosolic domains compared to the TM helices. In fact, only TM1 and TM2 (**in blue and purple in Figure 2.6b1**) display significant changes during the catalytic cycle and their movement, responsible for the formation of a cleft important for substrate translocation, occurs only after dephosphorylation of the P domain and rearrangement of the A domain. The polar head of PS, the preferred substrate of ATP8A1, interacts with this open cleft and is recognized by the PISL motif of TM4 (**in light green in Figure 2.6b1**), while the largest part of the acyl chains remains exposed to the membrane hydrophobic environment, with only a small portion fitting into a hydrophobic pocket formed by TM2 and TM4. Complementary activity assays demonstrate that the length of the acyl chain influences the ATPase activity, suggesting an interaction of the P-IV ATPase with the hydrophobic tail of the phospholipid. In fact, PS with short acyl chains induces a weaker Drs2 ATPase activity compared to PS with longer acyl chains. Despite the biochemical evidence, the position of the interaction site of the acyl chain on the ATPase remains an open question and suggests that the “credit card” model may not cover all aspects of P-IV ATPase cycle. Interestingly, PS binding site of ATP8A1 partially overlaps with the Ca^{2+} -binding site of SERCA.

The interaction between α - and β -subunits is similar in Drs2/Cdc50 and ATP8A1/CDC50a structures. In the extracellular part, the driving force for the complex formation is probably the electrostatic complementarity between the negative charges of CDC50a (**in magenta in Figure 2.6b 1 and 2**) and the positively charged extracellular loops of ATP8A1. In the TM domain, TM7 and TM10 (**in sea green and black in Figure 2.6b2**) helices of the α -subunit are in contact with the second helix of CDC50a. In the cytoplasm, the interaction involves the ATP8A1 loops between TM6 and TM7, TM8 and TM9 (**in orange and pink in Figure 2.6b2**) and the segment connecting TM4 to the P domain. Interestingly, in all structures CDC50 maintains a similar conformation. Taken together, these results suggest a great influence of CDC50a on the rigidity of the TM domain of ATP8A1.

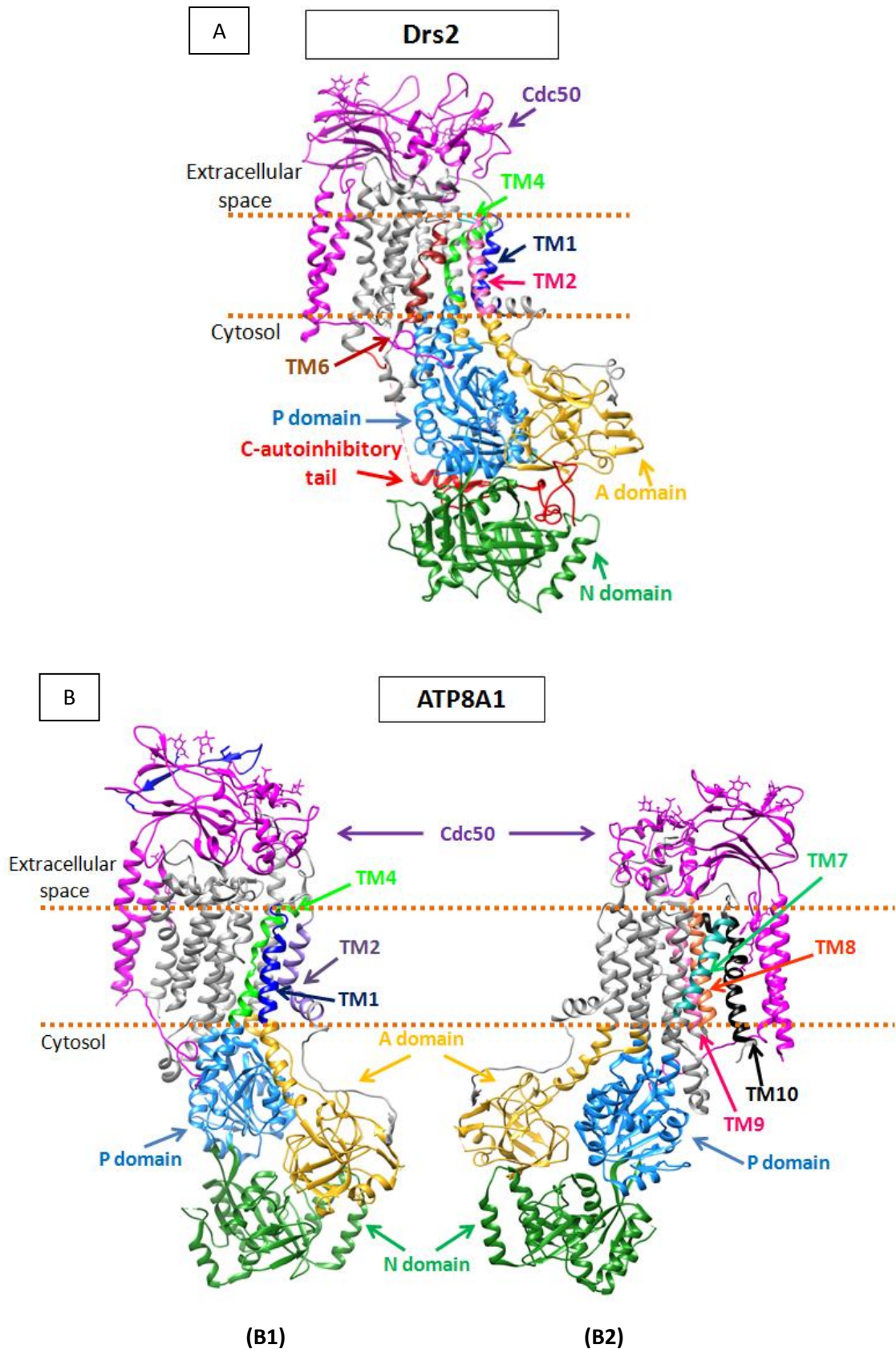


Figure 2.6: A) Cryo-EM structure of the complex Drs2-Cdc50 from *S. cerevisiae*. Drs2 in the autoinhibited state (PDB: 6ROH) (Timcenko et al. 2019). **B)** Structures of the human ATP8A1 with its subunit Cdc50a. In figure B1 are seen TM1-2-4, in figure B2 are seen TM7-8-9-10. ATP8A1 in the E1 state (PDB: 6K7G) (Hiraizumi et al. 2019).

2.2.2 Human diseases connected to defects of P-IV ATPases.

The human genome encodes for 14 different P-IV ATPases, while the yeast *S. cerevisiae* has only 5 paralogs and the plant *Arabidopsis thaliana* 12 (T. G. Pomorski and Menon 2016). All P-IV ATPases can be divided in three groups: proteins that flip preferentially PS and to a lower extent PE, proteins that flip preferentially PC and PE, and proteins whose substrate is still unknown. These enzymes are able to select their substrate mainly from the headgroup and the glycerol backbone, and only to a lesser extent from the acyl chain composition (Roland and Graham 2016b).

The involvement of P-IV ATPases in different human diseases is not surprising, considering their fundamental role in maintaining the cell membrane structure. Alterations of these proteins are associated to different diseases, including Alzheimer's disease, autism, obesity and liver disorders (**Figure 2.7 and Table 2.1**) (van der Mark, Oude Elferink, and Paulusma 2013). Moreover, it is known that an anomalous exposure of PS on the extracellular plasma membrane, possible consequence of an incorrect activity of a P-IV ATPase, is a signal for cellular apoptosis (Zwaal, Comfurius, and Bevers 2005).

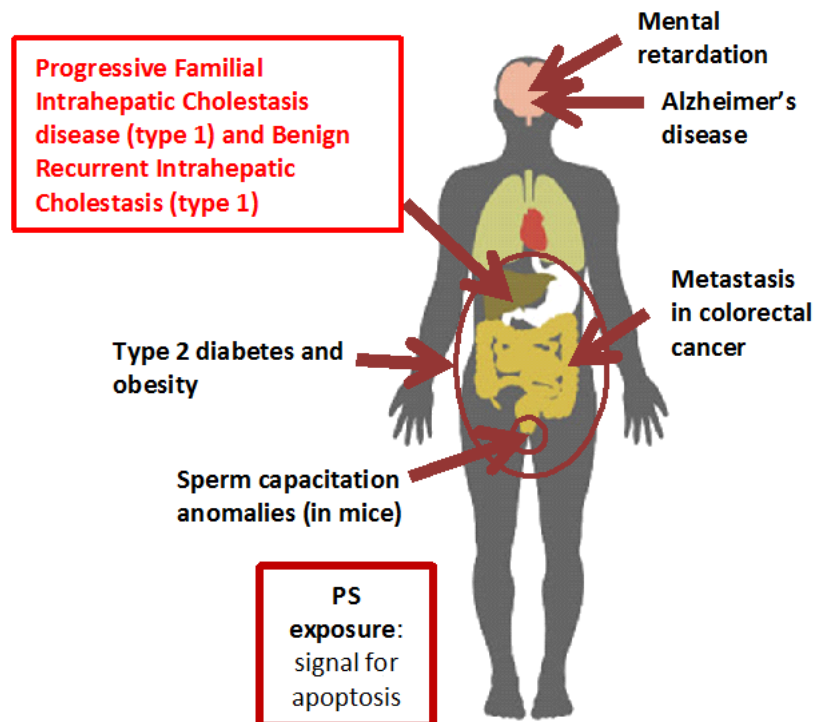


Figure 2.7: Schematic representation of the main diseases linked to malfunction of human P-IV ATPases.

- *Liver diseases:* The P-IV ATPase ATP8B1, belonging to class 1B, transports mainly PC and, while it is expressed ubiquitously in the body, reaches particularly high levels in the intestine and pancreas. Mutations on ATP8B1 are well characterized and have been identified as responsible for Familial Intrahepatic Cholestasis type 1 (PFIC1) and Benign Recurrent Intrahepatic Cholestasis type 1 (BRIC1), part of the more generic family of cholestasis diseases. PFIC1 often leads to liver failure before adulthood, instead BRIC1 is more benign and consists of episodic periods of disease that resolve spontaneously. In both cases patients suffer a reduced bile flow, resulting in difficulties in fat digestion, malabsorption of nutrients and even pancreatitis, diarrhea and hearing loss (Bull et al. 1998; Eppens et al. 2001; Folmer, Elferink, and Paulusma 2009). The I661T mutation of ATP8B1 is the most common in European patients. Some compounds used in the treatment of cystic fibrosis have recently shown to rescue the low activity of ATP8B1 due to the I661T mutation, opening the possibility for a therapeutic strategy against intrahepatic cholestasis (van der Woerd et al. 2016). Although the molecular mechanisms behind these diseases are not fully understood, it is possible that a lipid imbalance on the outer leaflet of the plasma membrane of cells of the bile canaliculi, thin tubes collecting bile secreted by hepatocytes, results in abnormal lipid packing, compromising their stability to bile salts present in the surrounding environment (Strautnieks et al. 1998).
- *Cerebellar ataxia and mental retardation:* Mutations on the P-IV ATPase ATP8A2 seem to be involved in a rare neurodegenerative disease called *cerebellar ataxia*, mental retardation, disequilibrium syndrome and hypotonia in humans (Cacciagli et al. 2010; Onat et al. 2013). This protein, belonging to subclass 1A is involved in the translocation of PS, and PE to a lesser degree, and is expressed in testis and in the membrane of the photoreceptor disc. Alterations of the paralog of ATP8A2 in mice induce visual and auditory functions loss (Coleman et al. 2014).
- *Metabolic diseases:* Studies have found that polymorphism on P-IV ATPase ATP10A is related to metabolic diseases, such as insulin resistance in African-American patients (single nucleotide polymorphism of ATP10A) (Dhar et al. 2004; 2006; Roland et al. 2018), type 2 diabetes and obesity (mutations of ATP10A). It is shown that mice with

a deficiency of ATP10A have insulin problems such as hyperinsulinemia and insulin-resistance. This ATPase of subclass 5 is responsible for PC translocation. Interestingly, ATP10A seems to be inherited only from the maternal chromosome (Sebastian et al. 2012). Obesity is also connected to mutations of the P-IV ATPase ATP10D, belonging to the same subclass 5.

- *Cancer*: The P-IV ATPase ATP11A (subclass 6), mainly involved in transport of PS, is frequently mutated in colorectal cancer patients and has been proposed as predictive marker for prognosis in this disease (Miyoshi et al. 2010). Mutations on another member of the same subclass, P-IV ATPase ATP11C, lead to hepatocellular carcinoma, hyperbilirubinemia, anemia and cholestasis in mice (Sebastian et al. 2012).
- *Infertility*: ATP8B3, a P-IV ATPase of subclass 1B, is present in testis and is implicated in fertility. Anomalies in ATP8B3 function lead to abnormalities in sperm capacitation in mice, resulting in the impossibility for spermatozoa to penetrate in the oocyte (Folmer, Elferink, and Paulusma 2009).
- *Blood diseases*: In healthy conditions, P-IV ATPases of subclasses 1A, 1B and 6 are responsible for asymmetry of PS distribution in the plasma membrane and are inactivated only during scramblase-induced PS exposure (Segawa, Kurata, and Nagata 2016). Anomalies on various P-IV ATPases of these subclasses are linked to abnormal PS exposure on the plasma membrane. For example, the ATPase ATP11C plays a critical role in generating PS asymmetry in erythrocytes during erythropoiesis (Yabas et al. 2014; Siggs et al. 2011). Mice with ATP11C deficiency show hyperbilirubinemia, hepatocellular carcinoma, B-cell development loss and anemia (Siggs et al. 2011; Yabas et al. 2014).
- *Alzheimer's disease*: The P-IV ATPase ATP8B4, seems to be implicated in Alzheimer disease in studies on different polymorphisms (Li et al. 2008).

Chapter 2

| Subclass | P-IV ATPase | Localization | Substrate | Linked diseases |
|-----------|---------------------|--|-----------|---|
| 1A | ATP8A1* | Mainly skeletal muscle and thyroid, but ubiquitous | PS>PE | Learning deficiencies in mice |
| | ATP8A2 [§] | Testis. Not present in liver, lung, kidney. Photoreceptor disc membrane. | PS>PE | Possibly tumorigenicity, mental retardation, hypotonia, <i>Cerebellar ataxia</i> , disequilibrium syndrome |
| 1B | ATP8B1 [€] | Mainly intestine and pancreas, but ubiquitous | PC | Mutations cause progressive familial intrahepatic cholestasis type 1 (PFIC1) and benign recurrent intrahepatic cholestasis type 1 (BRIC1) |
| | ATP8B2 [§] | Mainly brain, bladder, uterus, but ubiquitous. Not present in kidney and skeletal muscle | PC | Abnormal sperm-egg interactions |
| | ATP8B3* | Testis | Maybe PS | Sperm capacitation anomalies (in mice) |
| | ATP8B4 ^x | Ubiquitous | Unknown | Possibly Alzheimer disease |
| | ATP8B5* | Testis | Unknown | Sperm capacitation anomalies |
| 2 | ATP9A** | Ubiquitous. Not present in spleen. | Unknown | |
| | ATP9B** | Mainly testis, ubiquitous. Not present in spleen and muscle. | Unknown | |
| 5 | ATP10A ^ϕ | Mainly brain, kidney, lung, pancreas, but ubiquitous. Not present in small intestine. | PC | Obesity, insulin resistance, type 2 diabetes |
| | ATP10B [§] | Brain and testis. | Unknown | |
| | ATP10D [^] | Mainly liver, kidney, spleen, but ubiquitous. | Unknown | Obesity, myocardial infarction and atherosclerosis |
| 6 | ATP11A ^f | Ubiquitous. | PS>PE | Metastasis in colorectal cancer |
| | ATP11B ⁿ | Ubiquitous | PS>PE | |
| | ATP11C [*] | Mainly liver, pancreas, kidney, but ubiquitous. | PS>PE | Cholestasis, anemia, defects in B cells differentiation, hepatocarcinoma |

Table 2.1: Human P-IV ATPases and diseases connected to defects in their activity. *(Kato et al. 2013; S. Lee et al. 2015; Levano et al. 2012);[§](Vestergaard et al. 2014; Coleman, Kwok, and Molday 2009; van der Mark, Oude Elferink, and Paulusma 2013); [€](Bull et al. 1998; Eppens et al. 2001; Folmer, Elferink, and Paulusma 2009; Paulusma et al. 2006; Stapelbroek et al. 2009; Takatsu et al. 2014);^ϕ(Takatsu et al. 2014);[^](Wang, Beserra, and

Garbers 2004);*(Li et al. 2008; Van Der Velden, Van De Graaf, and Klomp 2010);*(P. Xu et al. 2009);**(Takatsu et al. 2011);*(Dhar et al. 2006; van der Mark, Oude Elferink, and Paulusma 2013); *(Folmer, Elferink, and Paulusma 2009); *(Coleman, Kwok, and Molday 2009; Flamant et al. 2003); *(Takatsu et al. 2014; van der Mark, Oude Elferink, and Paulusma 2013); *(Folmer, Elferink, and Paulusma 2009); *(Siggs et al. 2011; Takatsu et al. 2014).

2.2.3 Yeast P-IV ATPases

P-IV ATPases genes are well conserved and human and yeast genomes share a good sequence homology for these proteins. For this reason, yeast P-IV ATPases are regarded as good models of the human proteins and numerous studies, both functional and structural, have been conducted on yeast homologs, taking advantage of the easiness in manipulation of yeast *S. cerevisiae* compared to mammalian cell lines. *S. cerevisiae* expresses 5 P-IV ATPases, localized in different compartments of the cell (**Figure 2.8 and Table 2.2**). Dnf1 and Dnf2 are localized in the plasma membrane and their loss causes defects in the uptake of PE, PC and PS. Drs2, Dnf3 and Neo1 are localized in the Golgi system and in endosomes and are involved in formation and transport of vesicles (Noji et al. 2006; Dalton et al. 2017).

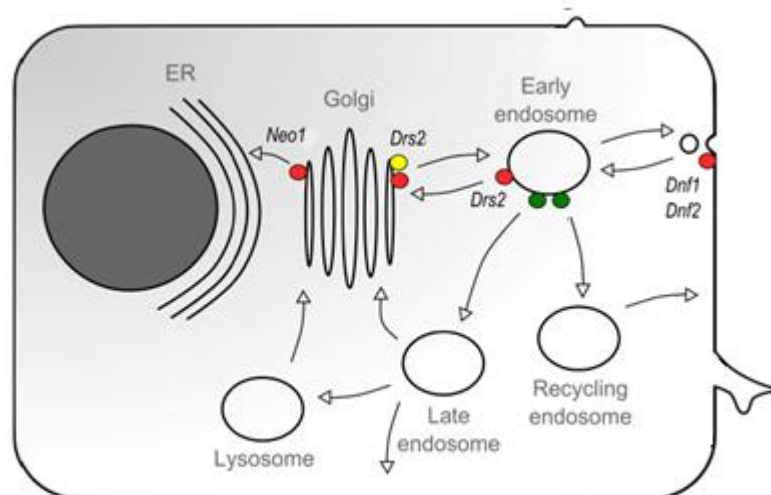


Figure 2.8: Localization of P-IV ATPases in yeast *S. cerevisiae*. From van der Mark V. A. et al. (2013), *International Journal of Molecular Sciences* (van der Mark, Oude Elferink, and Paulusma 2013).

Yeast P-IV ATPases share high sequence identity: Neo1, Dnf1, Dnf2 and Dnf3 share 29-31% of sequence identity and 45-48% similarity with Drs2. Among them Dnf1 and Dnf2 share the

highest sequence homology, with 69% of sequence identity and 83% similarity (T. Pomorski et al. 2003).

- The protein **Drs2** is formed by a single polypeptide chain of 1355 residues and works in association with its β -subunit, Cdc50. It localizes in the Trans-Golgi Network (TGN) and in early endosomes. Studies show that its preferred substrates are PS and PE (Natarajan et al. 2004). When reconstituted in liposomes, Drs2 shows high affinity for PS in presence Cdc50, suggesting that its β -subunit is necessary for phospholipid translocation (X. Zhou and Graham 2009).

PI4P is necessary for Drs2 flipping activity as it interacts with the RMKKQR motif present on the C-terminal of the protein, removing the autoinhibitory effect of this sequence (see above *Cryo-Electron Microscopy structures of P-IV ATPases*). Moreover, PI4P is implicated in the recruitment of proteins involved in vesicle formation in the TGN, suggesting a larger network of interactions. The C-terminal domain of Drs2 interacts also with the Arf GTPase Gea2, a regulator of protein trafficking and organelle structure, confirming Drs2 role in vesicle formation (Sakane, Yamamoto, and Tanaka 2006; Jacquot et al. 2012; Hanamatsu et al. 2014; Natarajan et al. 2009; Chantalat et al. 2004).

The N-terminal domain of Drs2 interacts with the F-box protein Rcy1 and the Arf-Gap protein Gcs1, both involved in recycling of vesicles between endosomes and TGN (Tsai et al. 2013; Sakane, Yamamoto, and Tanaka 2006).

- The protein **Dnf1**, whose name stands for Drs2/Neo1 family 1, is formed by a single polypeptide chain of 1571 residues, but works in close association with the protein Lem3 from the Cdc50 family, its β -subunit. It localizes in the plasma membrane, reaching higher concentrations in budding sites. Studies suggest that PC and PE, but also their precursors lyso-PE and lyso-PC are Dnf1 preferred substrates (T. Pomorski et al. 2003). Regulation mechanisms of Dnf1 seem to involve phosphorylation by Fpk1 and Fpk2 kinase proteins, whose action activates the flippase. Since Fpk1 and Fpk2 are stimulated in presence of sphingolipids, Dnf1 activity seems to be upregulated by the presence of sphingolipids in membrane (Tanaka, Fujimura-Kamada, and Yamamoto 2011; Nakano et al. 2008; Roelants et al. 2010).

- The protein ***Dnf2***, Drs2/Neo1 family 2, a single polypeptide chain of 1613 residues, has a similar localization, β -subunit, substrate affinity and activation as Dnf1. Studies demonstrate that these proteins are able to distinguish sphingolipids from glycerolipids, recognizing the backbone of the lipid and flipping only the glycerolipids (T. Pomorski et al. 2003). In the same study, cells deficient for the expression of Dnf1 and Dnf2 show sensitivity to the presence of metal ions in the surrounding environment and to exposure to low temperatures (20°C), suggesting a role for these proteins in metal homeostasis and in maintaining membrane fluidity.
- The protein ***Dnf3***, Drs2/Neo1 family 3, is formed by a polypeptide chain of 1656 residues. Its β -subunit is the protein Crf1 belonging to the Cdc50 family. This paralog is present in the Golgi system and in secretory vesicles, where it has been proposed to flip PS and PE (Alder-Baerens et al. 2006; Lopez-Marques et al. 2014).
- The protein ***Neo1***, named for the strain where it was first isolated, Neomycin Resistant 1 (Prezant, Chaltraw, and Fischel-Ghodsian 1996), is formed by a single polypeptide chain of 1151 residues, and is the only yeast P-IV ATPase that works without a β -subunit. Together with its human homologs ATP9A and ATP9B, Neo1 belongs to subclass 2. Substrates for these proteins have not yet been identified. These flippases are present in the endosomal system and in the Golgi network. Interestingly, the lethal effect caused by Neo1 depletion cannot be reduced by the overexpression of other members of this family. In fact, while knockout mutants for other P-IV ATPases are able to survive because other members of the family are able to compensate for the function of the deleted paralog, Neo1 is the only P-IV ATPase of yeast that is essential for cell survival.

Experimental evidence shows an in-vivo interaction between Neo1 and Ysl2, a protein required for endocytosis and vacuole formation, through immunoprecipitation (Wicky, Schwarz, and Singer-Kruger 2004; Hua, Fatheddin, and Graham 2002; Barbosa et al. 2010). Furthermore, the loss of asymmetry in the cis Golgi caused by the depletion of Neo1 alters the retrograde protein transport to the endoplasmic reticulum (Hua and Graham 2003).

Chapter 2

Various hypotheses have been developed regarding Neo1 substrate. Some experiments indicate that Neo1 has a key role in PS and PE membrane asymmetry, but it is not clear whether these lipids are substrates of Neo1 (Takar, Wu, and Graham 2016). The Neo1 analogue in the nematode worm *Caenorhabditis elegans*, TAT-5, is also involved in PE distribution, strengthening the hypothesis that PE is the preferred substrate of P-IV ATPases of subclass 2 (Wehman et al. 2011).

| Subclass | Protein | β -subunit | Localization | Substrate | Function | Interactor |
|----------|--|------------------|--------------------------------|---------------------------|---|--|
| 1 | Drs2* | Cdc50 | Golgi, secretory vesicle | PS, PE | Vesicle formation, cell polarity | PI4P, AP-1, Gcs1, Gea2, Rcy1, Sac1 |
| 2 | Neo1 ^ϕ | None | Golgi, Endosome | Unknown | Vesicular transport | Ysl2 |
| 3 | Dnf1 ^δ Dnf2 ^δ | Lem3 | Plasma membrane | PC, PE, (PS), LPC, LPE | Endocytosis, cell polarity, lysolipid uptake, protein sorting | Fpk1, Fpk2 |
| 4 | Dnf3 * | Crf1 | Golgi, secretory vesicle | PC, PE | Vesicular transport | |

Table 2.2 : Subclass division for *S. cerevisiae* P-IV ATPases. For each class β -subunit, cell localization, substrates and interactors are reported. *(Natarajan et al. 2004; Azouaoui et al. 2017; Sakane, Yamamoto, and Tanaka 2006; C. Chen et al. 1999; Lopez-Marques et al. 2014);^ϕ(Hua and Graham 2003; Wicky, Schwarz, and Singer-Kruger 2004); ^δ(T. Pomorski et al. 2003; Roelants et al. 2010); *(Alder-Baerens et al. 2006)

RESULTS AND DISCUSSION

Considering the large number of diseases involving P-IV ATPases (**Table 2.1**) and their fundamental role in biological processes, a better understanding of their structure and mechanism would be a major advancement in biochemical and medical research. In this study, I decided to work on a yeast analog of human P-IV ATPases, the protein Neo1, as the sequence of this protein shares a high homology with the corresponding human P-IV ATPases. In addition, the choice of a protein from a simpler eukaryotic organism compared to *Homo sapiens* is expected to allow higher expression levels in the first phase of this work. Neo1 is the only yeast P-IV ATPase essential for cell survival and that does not require the co-expression of a β -subunit. However, the structure and the preferred substrates of this protein are still unknown.

Neo1 was expressed as recombinant protein with a GFP tag and a His tag at its C-terminus. These two tags allowed for the detection and the purification of the protein. Recombinant Neo1 stability was then tested with circular dichroism, and X-ray crystallization experiment and negative staining EM tests were carried on in order to obtain structural data of this protein. To elucidate the role of specific residues in the translocation mechanism of this protein, several mutants were created for residues involved in ATP binding during Neo1 catalytic cycle and for residues possibly involved in substrate selection. These mutants were overexpressed as chimeric proteins with the same protocol as the Neo1-GFP-His recombinant wild-type protein. ATPase consumption tests in crude membranes and on purified protein samples allowed to detect changes in protein activity.

2.3 NEO1 RECOMBINANT PROTEIN ANALYSIS

2.3.1 *Plasmid preparation for the overexpression of the protein Neo1*

The gene coding for the ATPase protein Neo1 was amplified from *S. cerevisiae* genome with suitable extensions on the N- and C-termini. Using the homologous recombination of the yeast cell, the gene was inserted in the pDDGFP_LEU2D plasmid (D. Drew et al. 2008).

This plasmid carries the selectable markers URA3 and LEU2, coding for the enzyme orotidine 5'-phosphate decarboxylase, required for the synthesis of pyrimidine ribonucleotides, and for the enzyme 3-isopropylmalate dehydrogenase, required for the synthesis of the amino acid L-leucine, respectively. Cells successfully transformed with this plasmid are able to survive in absence of both uracil and leucine in the growth media. The protein construct

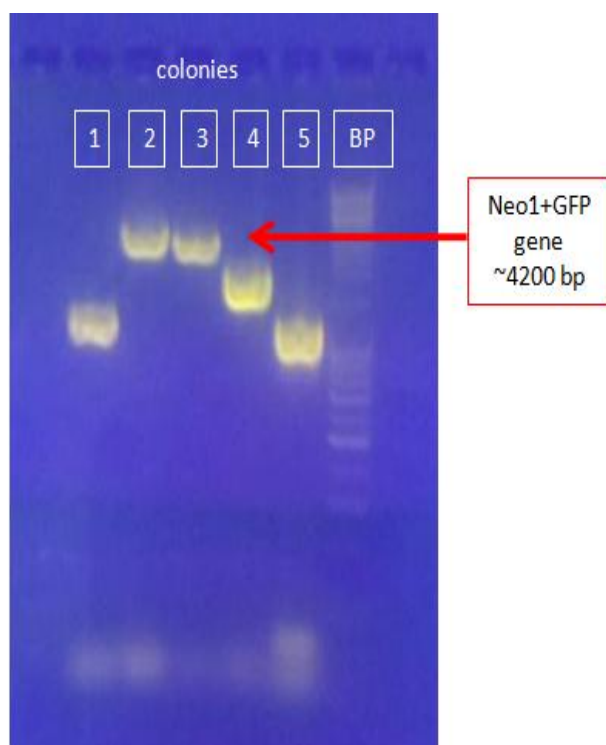


Figure 2.9: Electrophoretic separation of PCR-amplified products from *Neo1* transformed colonies. Lanes containing amplified products of 5 different colonies are numbered, lane BP contains markers for gene dimension. Colonies 2 and 3 (red arrow) show an amplification product with the correct gene length (~4200 bp).

inserted in the cloning site has a C-terminal tandem GFP/8X His tag. After transformation, different colonies were selected, the plasmid was extracted and the gene insertion was controlled by PCR amplification and agarose gel electrophoresis. **Figure 2.9** shows the result of electrophoretic separation of the amplified products from 5 different *S. cerevisiae* colonies, proving that in the second and third colonies the *Neo1* gene has been correctly inserted into the plasmid.

Correct gene insertion was further confirmed by sequencing of plasmids from positive colonies.

The plasmid from colony 2 was used to transform INVSc1 and FGY217 engineered yeast strains.

2.3.2 Overexpression conditions for the protein *Neo1*

The gene coding for the P-IV ATPase *Neo1* was inserted in the pDDGFP_LEU2D plasmid downstream the strong inducible GAL1 promoter gene. Thanks to this promoter, protein expression is triggered by the presence of galactose in the medium. For the present study, expression conditions included 2% w/v galactose in the medium, able to activate the GAL1

promoter. During protein expression, galactose is also used as carbon source for the yeast cells (J. L. Parker and Newstead 2014).

The overexpression of membrane proteins is often not trivial due to the low yields obtained for this kind of proteins and to the possible toxic effects for the host cell. Therefore, expression conditions have to be tightly controlled in order to maximize the amount of protein expressed by the host cell. An initial indication of the possible toxic effects on the host can be obtained by monitoring cell survival through optical cell density measurements, with visible light at 600 nm (OD_{600}). In addition, reporter tags are often fused at one end of the target protein, in order to quickly estimate the overexpression yield in different conditions. For the overexpression of the P-IV ATPase Neo1, a GFP tag at the C-terminus of the protein allowed monitoring overexpression yield through fluorescence measurements. The presence of the GFP at the C terminus of the recombinant protein can be also used as indication of its folding. In fact the GFP emits its fluorescence only if correctly folded. For membrane proteins an incorrect folding, or an incorrect insertion in the membrane, leads to precipitation or segregation into inclusion bodies. Therefore in presence of GFP fluorescence and absence of precipitate it can be assumed that the whole recombinant membrane protein is correctly folded and inserted into the cellular membrane (D. Drew et al. 2005; 2008).

Expression levels of the P-IV ATPase Neo1 were tested in small scale experiments in different conditions, using the INVSc1 strain of *S. cerevisiae* transformed with the pDDGFP_LEU2D plasmid, in a medium without uracil (URA-) (**Figure 2.10**). Considering that P-IV ATPase proteins seem to be involved in metal ion homeostasis (T. Pomorski et al. 2003), tests were conducted with media supplemented with different ions, such as Ca^{2+} 200 mM, Co^{2+} 100 μ M and Zn^{2+} 5 mM. In alternative, 2.5% w/v of DMSO was added to the growth media, since other studies proved that this additive facilitates membrane protein expression in yeast (D. Drew et al. 2008). Temperature is another parameter whose variation has often a great influence on expression levels. To determine the temperature that guarantees the highest yields, small scale expression tests were conducted at 20°C and the results compared to cell growth and protein expression at the more usual temperature for yeast growth, i.e. 30°C. Lowering the temperature may be beneficial for proteins that are difficult to express due to protease degradation in the host cell, since a lower temperature slows down cell metabolism and enzymatic reactions that could lead to host apoptotic processes.

Fluorescence of the yeast cells was measured for each expression test. Higher fluorescence signals correspond to higher expression yields. A negative control of cells, grown in the same conditions but with 2% w/v glucose addition instead of galactose, was analyzed together with the overexpression tests. Glucose is not able to activate the GAL1 promoter, but inhibits it. However, this sugar can be used by the yeast cell as carbon source, similarly to galactose. Expression of the *S. cerevisiae* P-IV ATPase Neo1 was compared with expression of the paralogs Drs2 and Dnf1, whose genes were cloned in the same construct as Neo1. As reference for the fluorescence measurement, INVSc1 cells were transformed with an empty pDDGFP_LEU2D vector and grown in the same conditions. Expression was monitored over 24 h after induction. **Figure 2.10** shows fluorescence measurements, indicative of expression levels, after 22 h from induction, for the conditions tested.

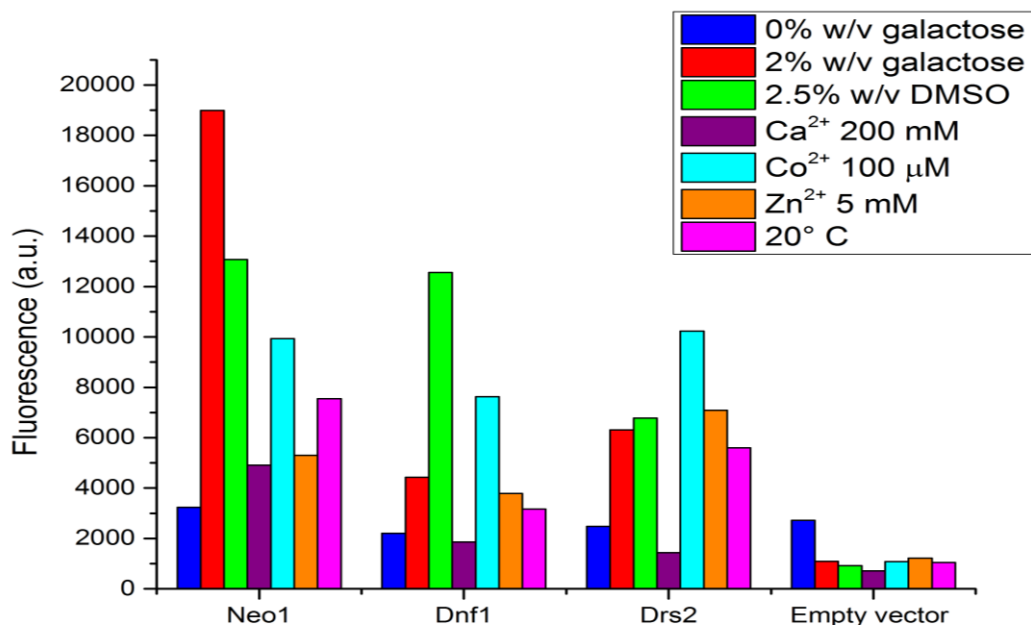


Figure 2.10: Overexpression of *Neo1*, *Dnf1*, and *Drs2* in INVSc1 cells, in different conditions, monitored by measuring GFP fluorescence. As negative controls, fluorescence measurements on cells transformed with an empty vector are reported for comparison. The blue bar represents inhibition of the GAL1 promoter with 2% w/v glucose, while in all other conditions expression was induced by addition of 2% w/v galactose. The green bars show expression tests with DMSO in the URA⁻ medium. Purple, cyan and orange bars represent conditions in which metal ions were added to the growth media. Magenta bars represents low temperature expression.

For the protein Neo1, the highest overexpression yield was obtained with an induction of 2% w/v galactose at 30°C (**Figure 2.10 red bar**). Addition of metal ions hampers the expression of the protein, as do lower temperature or addition of DMSO. As expected, low fluorescence was recorded when the GAL1 promoter was inhibited, possibly due to a residual basal expression of the construct. Fluorescence values below the instrument sensitivity were obtained for cells transformed with an empty vector.

In contrast to the optimal expression conditions for Neo1, higher expression levels for the proteins Dnf1 and Drs2 were obtained in different conditions, i.e. by adding DMSO or Co^{2+} ions to the medium, respectively. Differences in behavior could be linked to the different localization of the expressed protein in the yeast cell. In fact, it is possible that the expression of the protein Dnf1, localized on the plasma membrane, is more affected by the presence of an exogenous molecule as DMSO in the medium, while Drs2 and Neo1, localized in the Golgi network, are not in direct contact with additives present in the medium.

In parallel to the fluorescence measurements, OD_{600} values were monitored, indicative of the cell density in the liquid culture. Cell density was higher when protein expression was inhibited by glucose, confirming that the expression of the membrane protein imposes a strain to the host cell, leading to a slower cell replication.

Further, the overexpression of the P-IV ATPase Neo1 was tested in two different *S. cerevisiae* strains, namely FGY217 and INVSc1, by inducing protein expression with 2% w/v galactose and monitoring the results through fluorescence measurements over 24 h after induction. INVSc1 is a fast growing diploid auxotrophic *S. cerevisiae* strain, particularly suitable for protein expression (Tang et al. 2013). This strain needs addition of amino acids and nucleobases such as histidine, uracil, tryptophan and leucine to its culture medium, allowing for the selection of auxotrophic markers in exogenous plasmids, in cultures lacking one or more of these metabolites. The strain FGY217 is also frequently used for protein expression bearing a deletion for the vacuolar endopeptidase Pep4 gene and reduced expression levels for vacuolar hydrolases. Absence of peptidases and hydrolases is advisable during protein overexpression to partially prevent enzymatic digestion of the recombinant proteins (Newstead et al. 2007).

As the plasmid pDDGFP_LEU2D codes for two different selectable markers, LEU2 and URA3, both URA- and LEU- media, lacking uracil and leucine, respectively, were tested for protein production. In fact, different growth conditions can lead to different protein overexpression

levels (J. L. Parker and Newstead 2014). **Figure 2.11** shows fluorescence measurements, indicative of expression levels, after 22 h from induction for the different conditions tested.

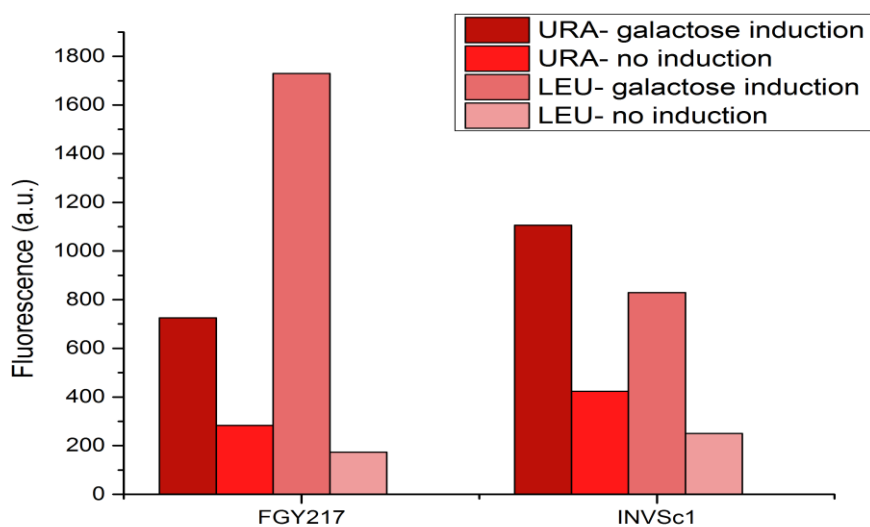


Figure 2.11: *Neo1* overexpression levels monitored by GFP fluorescence measurements in FGY217 and INVSc1 strains transformed with *Neo1 pDDGFP_LEU2D* plasmid, with different growth media. Protein expression was induced by addition of 2% w/v galactose. In negative controls, the *GAL1* promoter was inhibited by addition of 2% w/v glucose.

As expected, higher fluorescence was recorded for both *S. cerevisiae* strains when galactose was added. The highest expression yield for *Neo1* was obtained in the FGY217 strain with a LEU- medium. In parallel to the fluorescence measurements, OD_{600} values were monitored, indicative of the cell density in the liquid culture. As previously observed, the cell density was higher when protein expression was inhibited with glucose.

Overexpression yield was analyzed also in time, resulting in a maximum of protein production at 22 h after induction. Longer induction times resulted in a decrease of GFP fluorescence, indicating a possible protein degradation.

Overexpression experiments on a larger scale have been performed in the conditions that gave the best results from the previous small-scale tests. In this case, yeast cells were lysed at the end of the induction, cell debris removed by low-speed centrifugation, and membranes recovered by ultracentrifugation. Fluorescence measurements were performed at each step, confirming that membranes had notable fluorescence levels, while soluble fractions had negligible fluorescence. These results confirmed that the whole chimeric construct of *Neo1-GFP-His₈* can be recovered from membranes.

The different tests performed allowed to select the best conditions for large scale overexpression of the protein Neo1. A sufficient yield of expressed protein is crucial for the outcome of the following experiments of solubilization, purification and crystallization of the target protein. Considering the results of small scale overexpression tests, protein production was performed in FGY217 cells, with a galactose induction in LEU- medium for 22 h.

2.3.3 Solubilization of the integral membrane protein Neo1

Yeast membranes containing the P-IV ATPase Neo1 were solubilized in buffer (Tris 50 mM pH 7.5, NaCl 150 mM) with different detergents, to reach the same total volume. Concentrations of all detergents were kept at 10 times the CMC (10x CMC) of each specific compound. Following protocols reported in the literature, exceptions were made for LDAO, that was tested at a concentration of 10 times its CMC, but also at 22 times its CMC (22x CMC = 1% w/v), and C12E8 that was 200 times its CMC (200x CMC = 1% w/v) (Hirayama et al. 2013). Protein solubilization was monitored through the fluorescence value of the supernatant after ultracentrifugation, compared to the initial membranes. Solubilization experiments were carried on for 1 h or overnight, to assess kinetic aspects of the solubilization process (**Figure 2.12**). LDAO 1% w/v was tested only for 1 h in order to avoid protein degradation caused by a high detergent concentration.

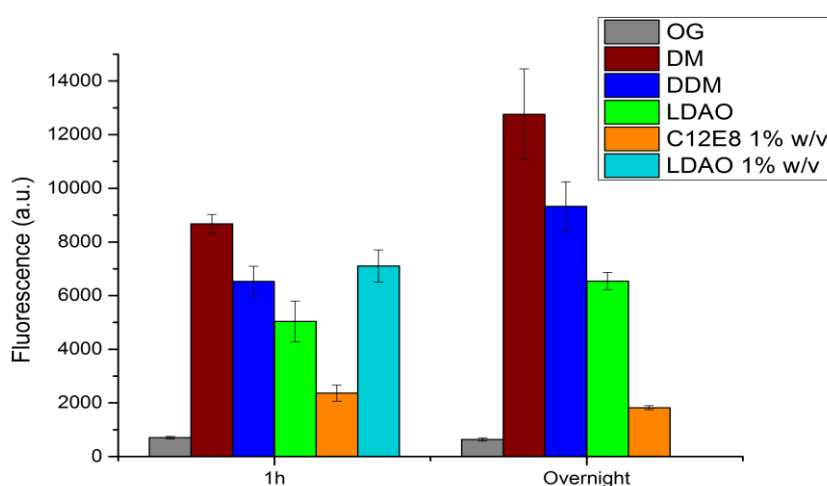


Figure 2.12: Solubilization tests of 200 μ l membranes containing the protein Neo1 with different detergents and solubilization lengths. All tests were conducted at 4° C under slow agitation. GFP fluorescence of the supernatant after ultracentrifugation is reported. Error bars correspond to the standard deviation of three experiments.

DM, DDM and LDAO detergents led to higher protein solubilization than OG. As expected, the mild C12E8 detergent led to low solubilization levels. For DM, DDM and LDAO an overnight solubilization resulted in higher protein recovery than the 1 h solubilization length. However, considering that the amount of solubilized protein does not improve significantly after the overnight incubation, all the following purification experiments were performed after a 1 h incubation period to avoid possible protein degradation from membrane bound proteases.

After the first solubilization tests, four detergents were selected for the following purification experiments: LDAO, DM, DDM and C12E8.

The publication in 2019 of the first P-IV ATPase structure (Timcenko et al. 2019), solubilized in LMNG, prompted us to test this detergent, not included in the initial experiments. LMNG was first used at 10x CMC (corresponding to 0.1 mg ml^{-1}) but this detergent concentration led to a low solubilization yield. However, increasing ten times the concentration (100x CMC = 1 mg ml^{-1}) significantly improved Neo1 solubilization. Therefore, a 100x CMC concentration was used for solubilization in the further purification tests.

2.3.4 *Purification protocol and analysis of protein stability for the P-IV ATPase Neo1*

A critical step for every structural study is the purification of the protein extracted from a natural source or overexpressed in a recombinant form. In this second case, the possibility to introduce purification tags facilitates the purification process, as a suitable affinity purification step can be planned for the tagged protein.

In the case of the P-IV ATPase Neo1 from *S. cerevisiae* membranes, the purification protocol included two steps. In the first step an immobilized metal affinity chromatography (IMAC) allowed to take advantage of the His-tag present at the C-terminus of the recombinant protein. In a second step, a size exclusion chromatography (SEC) was performed in order to separate aggregates and to control the protein oligomerization state. IMAC and SEC steps were performed in buffer (Tris 50 mM pH 7.5, NaCl 150 mM) with different detergents to analyze their effect on protein recovery after purification and on the aggregation state of the purified sample.

Table 2.3 shows the amount of protein recovered after the first purification step, monitored by measuring the GFP fluorescence. For purification tests conducted in DDM, the

solubilization was performed using DM that increases the yield of protein extraction from membranes but has a molecular structure very similar to DDM (see Materials and Methods). While a higher concentration of detergent is required to achieve high extraction yields from the membrane, it is not unusual to reduce the detergent concentration in the first purification step to decrease its negative influence on native protein folding. In the present study, such an expedient was applied to C12E8 and LMNG purification tests (C12E8 200x CMC for solubilization and 6x CMC for purification; LMNG 100x CMC for solubilization and 10x or 3x CMC for IMAC or SEC purification, respectively).

| Purification | Protein solubilization | Protein bound to the | Protein in the final |
|-------------------|------------------------|----------------------|----------------------|
| | % | resin % | elution % |
| LDAO (A) | 87 | 7 | 6 |
| DM/DDM (B) | 100 | 9 | 8 |
| C12E8 (C) | 86 | 15 | 13 |
| LMNG (D) | 96 | 53 | 46 |

Table 2.3: Results of the IMAC purification of the P-IV ATPase Neo1, monitored by measuring the GFP fluorescence after binding and after elution. Membranes, deriving from 1-2 l of culture, were solubilized in buffer with detergent and incubated with 1 ml Co^{2+} resin. Resin was washed and protein eluted in 0.5 ml buffer with 150 mM imidazole for 5 times. Elutions were collected and analyzed. A) LDAO concentration: 10x CMC (20 mM); B) Solubilization in DM 20x CMC (36 mM), purification in DDM 5x CMC (0.6 mM); C) C12E8 concentration: 200x CMC (1% w/v) during solubilization, reduced to 6x CMC (0,03% w/v) during purification; D) LMNG concentration: 100x CMC (1 mg mL^{-1}) during solubilization, reduced to 10x CMC (0.1 mg mL^{-1}) during purification.

The protein eluted from the first purification step underwent a quick buffer exchange or was immediately used for the second purification step, in order to avoid prolonged exposure to high imidazole concentrations. In fact, preliminary tests showed that the denaturation of the protein and consequent precipitation is a rapid process in buffers containing imidazole. The second purification step, the Size Exclusion Chromatography, was monitored following both the sample absorbance at 280 nm and the GFP fluorescence. **Figure 2.13** shows the results of SEC protein purification with different detergents.

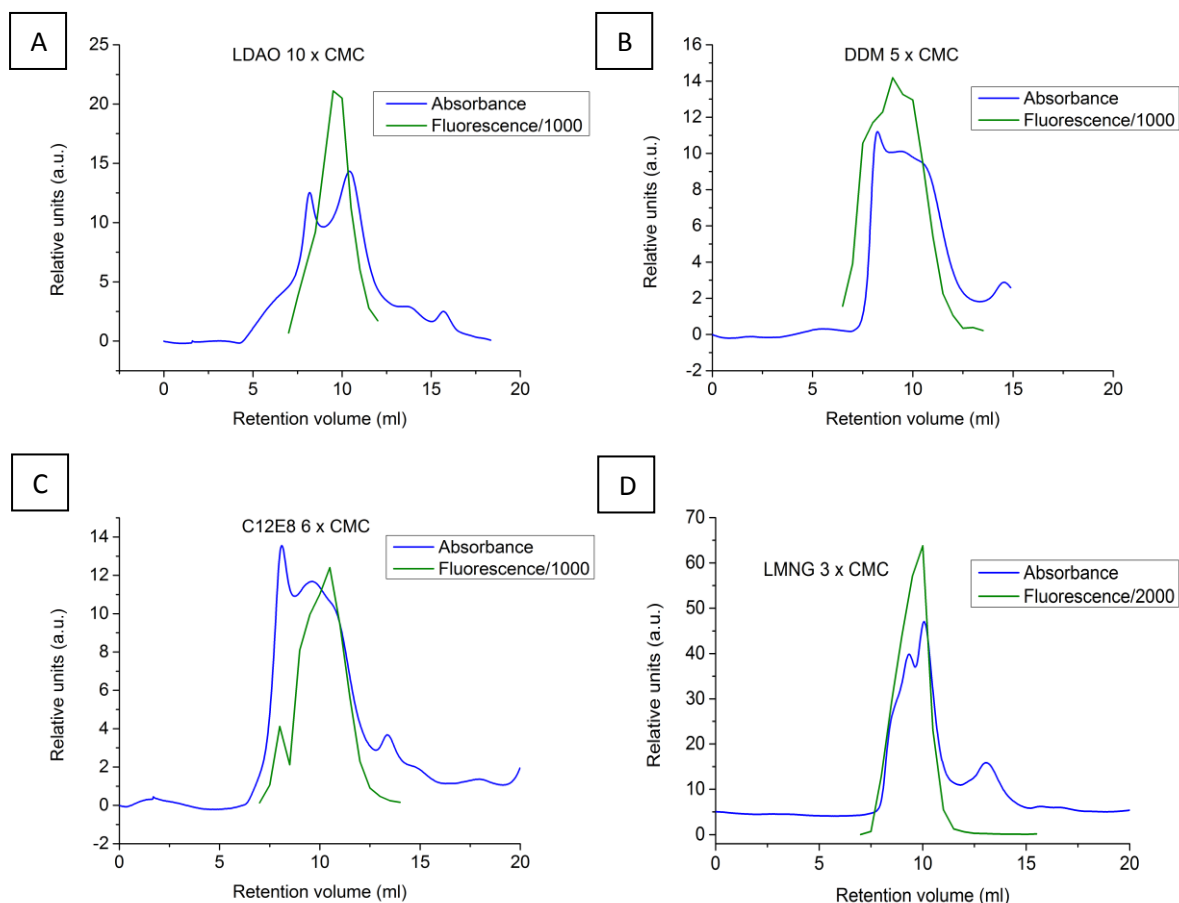


Figure 2.13: Results of the SEC purification of the P-IV ATPase Neo1, monitored by measuring the sample absorbance at 280 nm (blue) and the GFP fluorescence (green), on samples A-D from IMAC purification (Table 2.3). Detergent concentrations during SEC: A) LDAO:10x CMC (20 mM); B) DDM: 5x CDC (0.6 mM); C) C12E8: 6x CMC (0,03% w/v); D) LMNG: 3x CMC (0.03 mg mL⁻¹).

Results of the IMAC purification in LDAO show a weak binding to the resin, with only the 6% of the initial fluorescence recovered in the final elution, suggesting that the detergent is not able to maintain the correct protein folding and the protein undergoes aggregation/precipitation processes during the first purification step. The SEC result confirms this analysis (**Figure 2.13a**), as the chromatogram shows two peaks (**blue line**), while the maximum in absorbance signal of GFP fluorescence (**green line**) suggests that the protein is present in solution in multiple aggregated forms, confirming that the LDAO detergent is not suitable for Neo1 purification. For others detergents, i. e. DDM (**b**) and C12E8 (**c**), the behavior of the protein was slightly different than in LDAO, with high solubilization yields but still weak binding to the IMAC resin. The chromatograms of SEC purification tests performed with these detergents show a first absorbance peak immediately after the dead column volume followed by a hump of absorbance. The position

of the first peak is an indication of the presence of protein aggregates. Chromatographic tests suggested that DDM and C12E8 detergents are not suitable for Neo1 solubilization, as they lead to the formation of aggregates, a possible result of protein instability. The use of LMNG (**d**) allowed to obtain high solubilization yields and a significant amount of protein could be recovered from the IMAC purification, leading to a fluorescence in the final elution corresponding to 46% of the initial sample. For this detergent, the SEC chromatogram shows a first peak at a retention volume of 9 ml, followed by a second higher peak at 10 ml, while no aggregation peak after the dead column volume is visible. In this case the maximum of GFP fluorescence overlaps with the second absorbance peak, suggesting that Neo1 is predominantly present in this fraction.

Figure 2.14 shows SDS-PAGE separation of fractions obtained from IMAC and SEC purifications in LMNG, with both in-gel fluorescence detection and Coomassie staining. The presence of the GFP tag on the recombinant protein allows for protein identification directly from the SDS-PAGE experiment, since under UV light only the recombinant protein (or, eventually, the free GFP at lower molecular weight) is visible. In all cases fluorescent SEC fractions contained only Neo1, which is visible in two bands between 200 kDa and 116 kDa. The position of the higher molecular weight band on the SDS-PAGE is consistent with the presence of the recombinant protein, that has a molecular mass of 170 kDa, including the tags. The second, weaker band is probably the result of a partial proteolytic cleavage of the protein.

Coomassie staining and in-gel fluorescence confirm that Neo1 is present in the first and second absorbance fractions of the SEC purification, corresponding to retention volumes of 9 and 10 ml, respectively (**Figure 2.14 Frac 9 and Frac 10**). In addition, GFP fluorescence shows a higher protein concentration in the second fraction. These retention volumes suggest that the protein may be present as a dimer. However, the presence of the detergent micelle could impact on the sample total volume, pointing on the need of a different experimental technique to confirm Neo1 aggregation state.

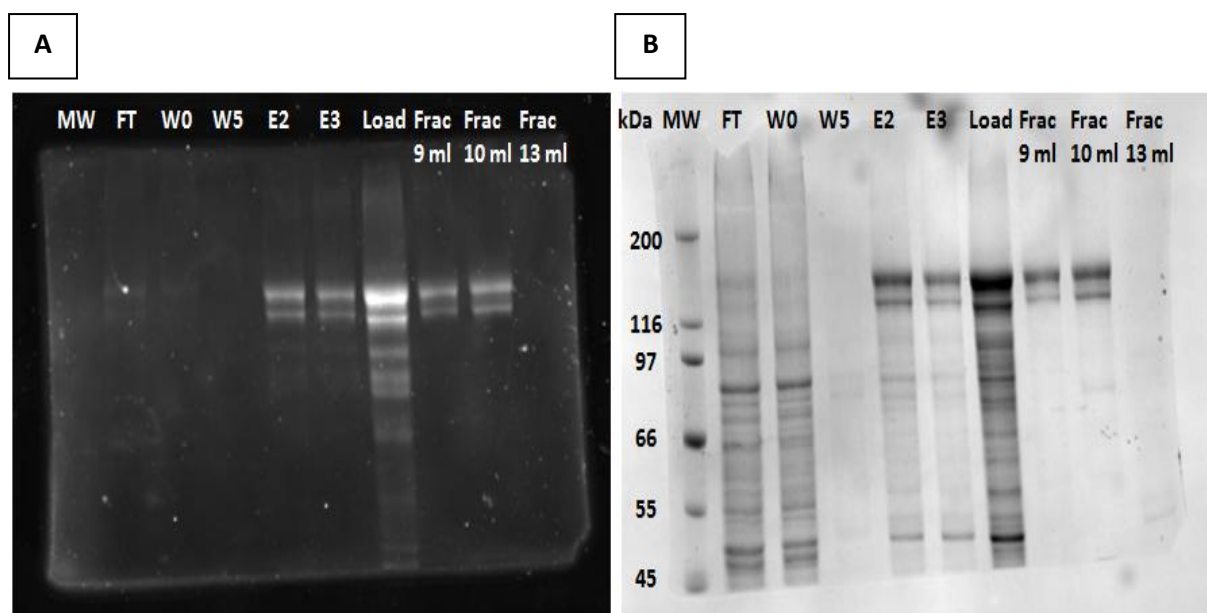


Figure 2.14: SDS-PAGE analysis of fractions of purification experiments on Neo1 in LMNG buffer. (A) In-gel fluorescence, (B) Coomassie staining. MW: Molecular Weight 200-45 kDa marker; FT: Flow through from IMAC resin; W0: first resin wash without imidazole; W5: second resin wash with 5 mM imidazole; E2 and E3: elution fractions with 150 mM imidazole; Load: sample loaded onto the SEC column; Frac 9 ml, Frac 10 ml and Frac 13 ml: SEC fractions with corresponding retention volumes.

2.3.5 Circular Dichroism analysis

Samples from LDAO, DDM and LMNG purification experiments were analyzed with Circular Dichroism (CD) spectroscopy in order to confirm the predicted secondary structure and to test Neo1 stability after purification in detergent (**Figure 2.15**). The CD techniques are based on the different absorption of light with opposite elliptical polarization, due to the chirality of the sample. For proteins, this analysis is particularly useful to detect secondary structure elements, i. e. α -helices, β -sheets, random coils, turns or other conformations. The predicted structure of the P-IV ATPase Neo1 has a high α -helical content, especially in the TM domain of the protein that is expected to be formed by 10 α -helices. Proteins with α -helical secondary structure yield a specific CD spectrum, with a maximum at 193 nm and two minima at 208 nm and 222 nm, respectively. Interestingly, in membrane proteins a few nm “red shift” of the absorbance frequency of the two minima has been often reported and it has been attributed to the presence of the detergent micelle surrounding the hydrophobic TM domain (Glaser and Singer 1971).

In addition to the secondary structure conformation, CD analysis can be informative also regarding the melting (or unfolding) temperature of the sample. In fact, while the protein undergoes unfolding processes, its secondary structure changes significantly and, therefore, yields a different CD signal. For this experiment, CD spectra are collected at different temperatures and the intensity of signals is compared.

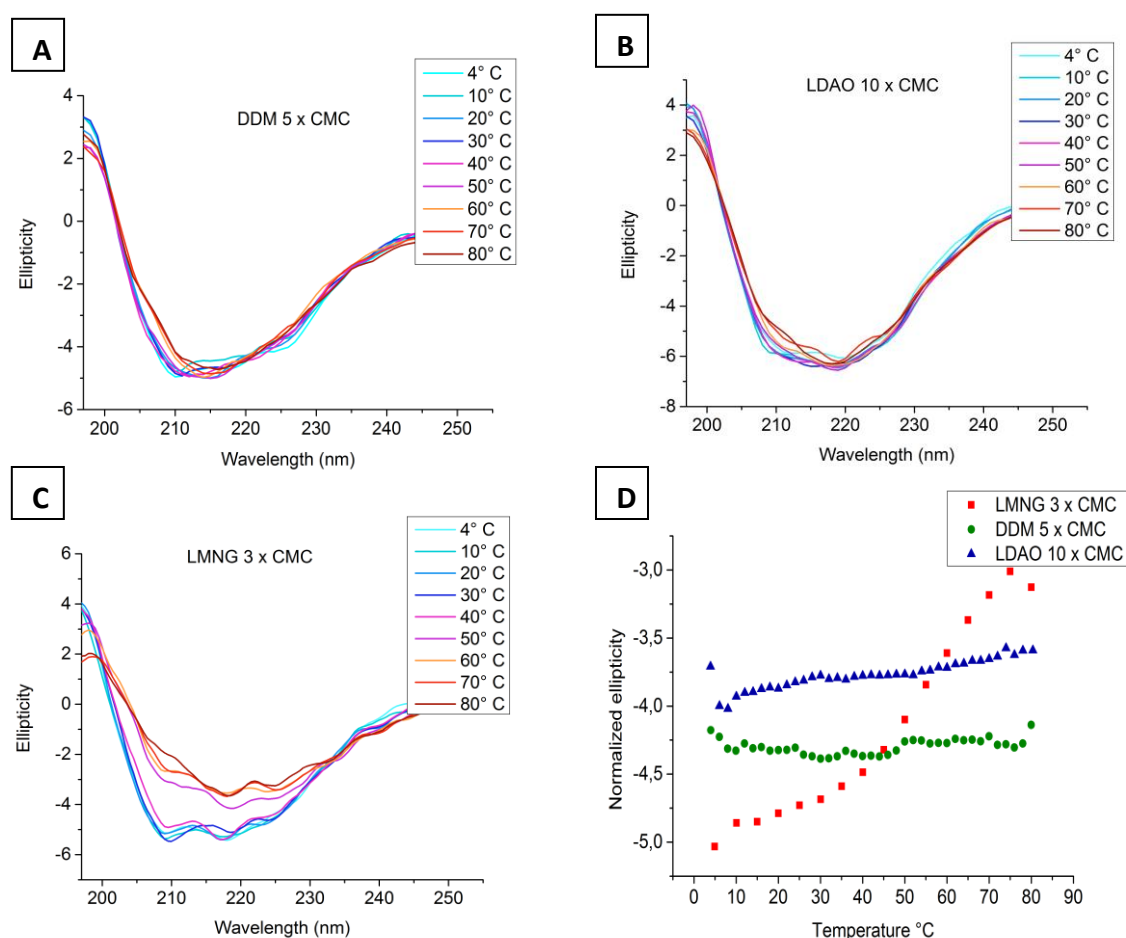


Figure 2.15: CD analysis in different detergents of samples of Neo1 obtained from SEC purification experiments. **A)** DDM 0.6 mM (5x CMC); **B)** LDAO 20 mM (10x CMC); **C)** LMNG 0.03 mg mL⁻¹ (3x CMC); **D)** Comparison of ellipticity values at 222 nm against temperature, for protein samples solubilized in DDM, LDAO and LMNG.

For samples in DDM (**Figure 2.15a**) and LDAO (**b**), the profiles of the CD spectra suggest that the protein is not in the expected α -helical conformation, but possibly in an unfolded or misfolded state, confirming results obtained from the analysis of SEC purification experiments. In fact, not only the minima corresponding to the α -helical secondary structure are not visible, but the ellipticity curve does not change during thermal treatment. On the contrary, the sample in LMNG (**c**) shows two minima, at 208 nm and 222 nm, corresponding

to α -helical elements and an additional minimum, possibly due to the presence of β -sheet elements in the cytoplasmic domains. In addition, the comparison of spectra measured at different temperatures shows a significant decrease in signal intensity, suggesting the presence of a transition between a folded and an unfolded state. **Figure 2.15d** shows the change in the CD spectrum at 222 nm, normalized for each experiment, against the temperature. For Neo1 in LMNG, a melting temperature between 40°C and 50°C can be inferred from CD data.

SEC and CD results confirm that LMNG is the ideal detergent for Neo1 solubilization and purification, as the protein in LMNG solution does not show aggregation or precipitation and the native folding is retained.

2.3.6 Overexpression of mutants on conserved sequences involved in phosphorylation and dephosphorylation mechanisms

In order to increase Neo1 overexpression to the levels required for structural studies, we decided to decrease its ATPase activity, and hence its toxicity for the host cell, by creating mutants of the protein in its P and A domains. This cytoplasmic domains of the protein are conserved in all P-ATPases and are responsible for the phosphorylation and dephosphorylation mechanisms, key steps of the catalytic cycle. Considering the alignment with the sequences of proteins belonging to the same family, we identified three residues involved in ATPase activity in two conserved sequences: residues D503 and K504 in the DKTGTLT motif, involved in protein phosphorylation, and residue E310 in the LDGET motif, involved in the dephosphorylation mechanism. Studies on other members of the P-ATPase family demonstrated that analogous mutations are able to decrease ATPase activity (Andersen et al. 2016).

In particular, 4 different mutations were selected:

1. The aspartic acid D503 residue was mutated to an asparagine residue (N), replacing the carboxyl group with a carboxamide. Substitution of the $-\text{COOH}$ with $-\text{CONH}_2$ inhibits phosphorylation in this site, thus preventing the start of the catalytic cycle.
2. For the lysine K504 residue, two different mutants were prepared. An asparagine was first substituted to the lysine, altering the electrostatic charge present in the site. Since the lysine residue is essential in orienting the ATP molecule in the catalytic site,

a substitution of this residue that removes the positive charge is expected to hamper protein phosphorylation.

3. Considering that a mutation that removes a positive charge may affect also protein folding, we decided to prepare another mutant of lysine K504, by substituting it with an arginine residue (R). In this case the electrostatic properties are retained, but the steric hindrance of the arginine alters the functional properties of the lysine.
4. The glutamic acid E310 was mutated to a serine (S). The change of the glutamic acid carboxyl group with an hydroxyl group locks the protein in the phosphorylated state. In fact, the glutamic acid is necessary for the dephosphorylation step of the catalytic cycle of P-ATPases.

Figure 2.16 shows the position of the selected residues on the Na^+/K^+ pump (Shinoda et al. 2009), which has an identical sequence to Neo1 in phosphorylation and dephosphorylation conserved motifs.

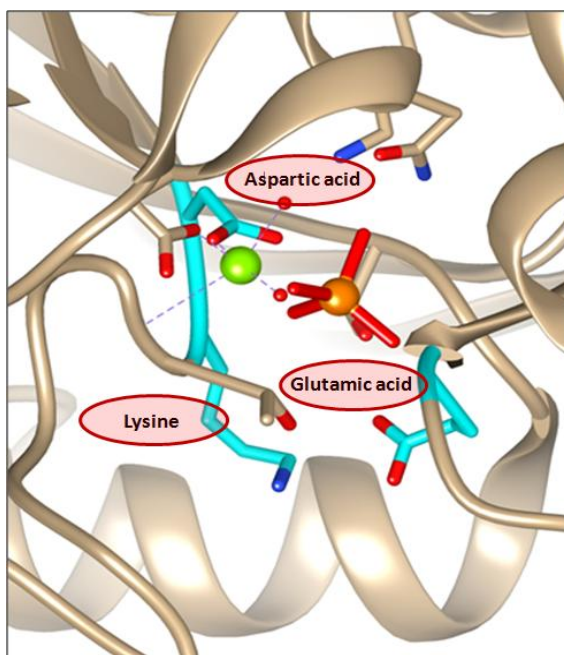


Figure 2.16: Residues involved in phosphorylation and dephosphorylation mechanisms on the Na^+/K^+ pump (PDB: 2ZXE) protein. The aspartic acid and the lysine are located in the P domain, while the glutamic acid is in the A domain of the protein.

Plasmids pDDGFP_LEU2D carrying E310S, D503N, K504N and K504R mutations were prepared with site-directed mutagenesis and verified through sequencing analysis. INVSc1 and FGY217 yeast strains were transformed with mutated plasmids. Small scale

overexpression tests were carried on in LEU- and URA- media, following the same protocol of the wild-type overexpression test (see above). **Figure 2.17** shows GFP fluorescence intensities measured on the cell cultures in different conditions, after 22 h from GAL1 promoter activation with 2% w/v galactose.

According to the fluorescence measurements, for all the mutants the highest yields were obtained from overexpression in INVSc1 yeast cells in a medium lacking leucine (**Figure 2.17, green bars**). In particular, the E310S mutant reaches higher overexpression levels compared to the others.

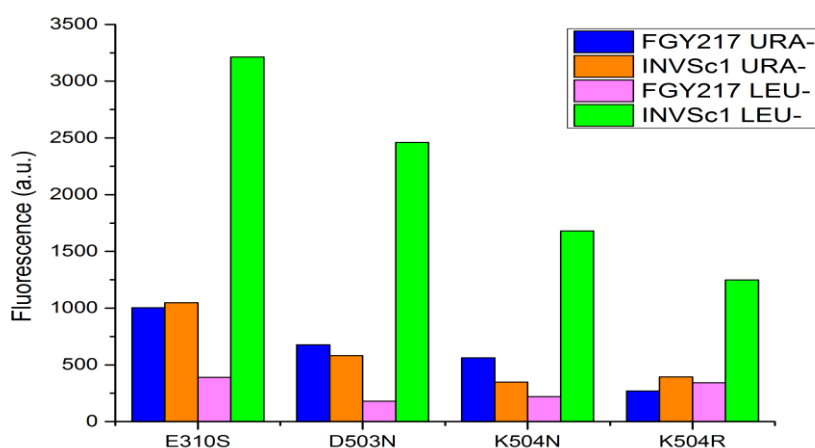


Figure 2.17: Overexpression tests on Neo1 mutants. GFP fluorescence of FGY217 cells grown in URA- medium (blue bars), INVSc1 cells grown in URA- medium (orange bars), FGY217 cells grown LEU- medium (magenta bars), and INVSc1 cells grown LEU- medium (green bars).

Scale-up experiments were set up for the E310S mutant in INVSc1 cells and LEU- medium. Membranes containing this mutant had high fluorescence levels, even compared to the wild type. Therefore, purification experiments were performed on these membranes using DM, DDM and LDAO detergent solutions. Unfortunately, the results obtained for this mutant E310S were comparable to those obtained for the wild type and protein aggregation was visible in the SEC chromatogram for all the detergents tested. However, this mutant is promising for further tests with the LMNG detergent.

2.3.7 *Neo1* activity assay

For the protein Neo1, as for all proteins of the P-ATPase family, ATP hydrolysis is strongly coupled with the translocation of the substrate across the membrane, in presence of magnesium ions that act as cofactors. Therefore, the analysis of the ATPase activity of the protein gives an indication on its capacity to transfer lipid substrates from the outer to the inner side of the membrane.

Activity tests were performed in presence of Mg^{2+} ions on wild-type and mutant proteins overexpressed in membranes, using commercial kits specific for the quantification of the ATP consumption. The magnesium ion is required as cofactor during protein catalytic cycle as previously described. **Figure 2.18** shows results of the ATP-consumption test of the mutants, compared to the wild-type protein and to membranes obtained from cells where overexpression was not induced. This test detects the amount of ATP still present in the reaction solution after an incubation period of 40 min in presence of ATP. Higher ATP levels are related to a lower protein activity. The wild-type (WT) protein was used in the test as reference for the highest ATPase activity, whereas membranes of *S. cerevisiae* where protein overexpression was not induced (WT NI) were used as reference for basal membrane ATP consumption, caused by different membrane ATPases. To further confirm ATPase activity, all tested conditions were repeated after incubation with Na_3VO_4 , a known inhibitor of the ATPase activity (Okumura and Kinoshita 2016).

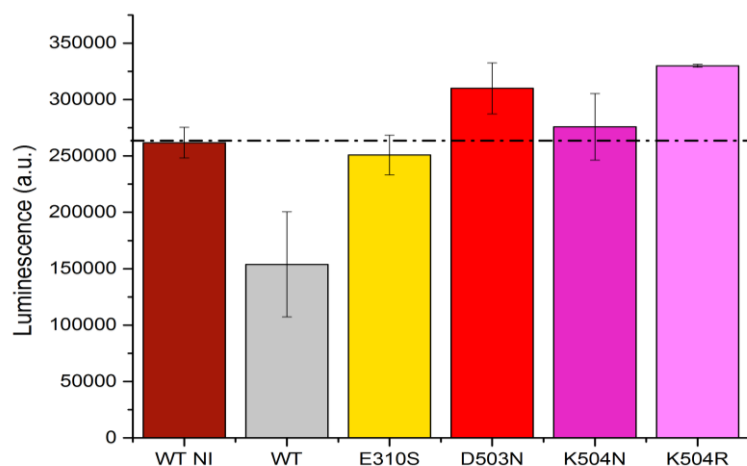


Figure 2.18: Residual ATPase levels of membranes containing Neo1 mutants compared to membranes containing the wild-type protein (WT) and membranes where overexpression was not triggered (WT NI). Membranes containing 25 μg of total proteins were incubated for 40 min at 37°C with 5 mM ATP in presence of 5 mM $MgCl_2$. Error bars: standard deviation on three experiments.

Membranes containing overexpressed mutants had a higher amount of unreacted ATP compared to the wild type, an indication of their reduced activity. As expected, membranes with no overexpression have a higher amount of residual ATP compared to the wild type. Mutations D503N and K504R had even higher residual ATP than membranes with no overexpression. The lower ATP activity of membranes of these mutants with respect to the non-induced membranes can be explained considering that in overexpression conditions membranes are overcrowded with mutants, reducing the space available for other ATP-consuming proteins. E310S and K504N mutations led to a similar ATP consumption as the negative control. These results confirm that residues selected for mutation are fundamental for protein activity, and their substitution with residues bearing even small differences in terms of charge or steric hindrance leads to high variation in ATPase activity.

ATPase activity assays were performed also on the purified protein samples in order to test their behavior in detergent, considering that activity is possible only in presence of the correct protein folding. In addition, to identify Neo1 substrates, the protein activity was analyzed after the addition of different lipids to the detergent solutions containing the protein. In detergent solution, the purified protein is surrounded by the detergent micelle and is expected to show no activity even when correctly folded, due to the absence of suitable substrate molecules. By adding to the solution the correct substrate, the ATPase activity should be restored.

Considering the low levels of activity that are expected for a membrane protein removed from its native environment, we decided to apply a more sensitive test that evaluates the amount of ADP produced, instead of the amount of residual ATP. The ADP produced corresponds to the ATP consumed.

Figure 2.19 shows the results of this activity test on samples of wild-type Neo1 purified in LMNG detergent. Different lipids (PC, PE or PS 0.12 mg ml^{-1}) were added to the sample and incubated for some minutes to allow lipid exchange with the detergent molecules. Considering that high DDM concentrations are required to solubilize the lipids, a final DDM concentration of 10x CMC was reached in all tests. After 30 min incubation with 0.5 mM ATP, ADP level was measured for each protein sample. Experiments were repeated with two different amounts of protein (25 or 50 ng).

As expected, purified Neo1 protein was inactive in a detergent solution with no lipids. Unexpectedly, addition of PC, PE and PS did not restore ATPase activity. Two possible explanations can be proposed: either the protein is inactive in the detergent solution, or the lipids added are not specific substrates for this P-IV ATPase. To verify the first hypothesis, crude *S. cerevisiae* membranes were added in the same amount as the lipids to the protein solution, as source of endogenous substrates: in this case protein activity was restored. To further verify the contribute of other ATP-consuming proteins present on the crude membranes, activity was measured also on a sample containing only *S. cerevisiae* membranes. In this case, a lower ATPase activity was measured compared to the activity in presence of the purified protein.

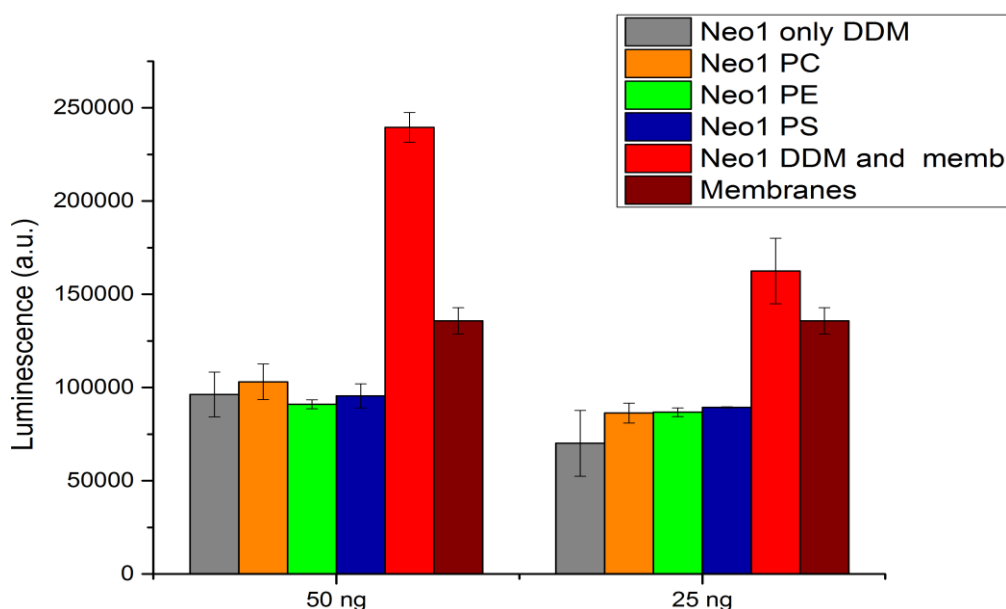


Figure 2.19: ATPase activity assay on samples of the purified protein in LMNG. Lipids PC, PE or PS (0.12 mg mL^{-1}) were added to the protein solutions in DDM 10x CMC buffer. As positive control, the same amount of *S. cerevisiae* membranes was added to the protein. Crude membranes in DDM buffer without purified protein were tested as well. Tests were repeated with two different amounts of protein. Error bars: standard deviation of three experiments.

Taken together these results confirm that Neo1 is active in LMNG solutions, a result that confirms the correct folding of the protein. In addition, the absence of activity with the tested lipids (PC, PE and PS) suggests that either the specific substrate of Neo1 is not among these, or the protein needs a co-factor (or an activator) to perform its activity. The substrate,

the activator or the co-factor may be present on the crude *S. cerevisiae* membranes, thus triggering Neo1 activation. Compared to other P-IV ATPases, Neo1 lacks the autoinhibitory C-terminal domain and it is not known to work with a β -subunit, suggesting that a different activation mechanism may be involved. On the other hand, since this protein is mainly found in Golgi and endosomes, it is possible that its substrate is an entirely different lipid, possibly with a critical role for these organelles. Further analyses are required to identify Neo1 specific substrates/activation mechanisms.

2.3.8 *Neo1 crystallization trials*

Among the main techniques used to structurally characterize biomolecules, X-ray crystallography occupies a prominent position, since in suitable conditions it can yield information at the atomic level. However, structure determination is only possible when crystals with high internal order are available and, therefore, the application of diffraction techniques is strictly dependent on the success of the crystallization process.

As crystallization conditions are highly diverse for different proteins, crystallization experiments should be set up with the largest possible matrices of conditions, testing different salts, precipitants and additives, varying the temperature or the parameters of the crystallization. However, protein purity and stability are among the main factors determining the success of a crystallization experiment. For membrane proteins, these requirements are challenging since these proteins are difficult to stabilize outside their hydrophobic membrane environment and their expression levels are usually low. Furthermore, the hydrophobic external surface of membrane proteins hampers the formation of protein-protein interactions required to obtain ordered crystals.

In the case of the recombinant protein Neo1, the presence of the GFP tag could be an advantage, as it increases the polar surfaces of the protein, allowing for the formation of polar contacts within the crystal. Protein samples eluted in the first and second peaks of the SEC chromatography in LMNG were concentrated and used for crystallization experiments, without further post-purification treatments. Crystallization trials were set up using commercial screens of sparse matrices of conditions (MemGold1, MemGold2, MemSyst and MemStart kits of Molecular Dimensions). To date, no crystal formation was observed in the

screen plates. However, as crystallization may require longer incubation times, experiments are still in progress.

2.3.9 *Neo1 negative staining and TEM analysis*

In recent years, an alternative technique for the structural characterization of biomolecules at high resolution became very popular, due to the amazing technical and computational advancements that it experienced: the cryo-Electron Microscopy (cryo-EM). This technique allows the visualization of proteins in aqueous solution, at lower concentrations compared to X-ray crystallography, and without the need for ordered proteins lattices. However, requirements for protein purity and stability are similar as for X-ray crystallography.

The first step for a successful cryo-EM experiment is the determination of the protein aggregation state, purity and stability. To this aim, a negative staining EM analysis is advisable before any attempt at preparing samples for cryo-EM. The negative staining technique takes advantage of the higher diffraction power of heavy atoms, such as tungsten or uranium, compared to the light atoms present in proteins (mainly carbon, hydrogen, nitrogen, oxygen and sulfur). The sample is dispersed on an electron microscopy grid, dried and covered with the heavy atom solution. The grid is analyzed at the microscope and positions occupied by proteins appear as white particles, while the heavy atom staining forms a dark background.

Figure 2.20 shows negative staining EM images of the purified Neo1 sample in LMNG, at 50'000x, 80'000x and 100'000x magnification, diluted in a range of 40-10 $\mu\text{g ml}^{-1}$. For this experiment, the use of Formvar carbon grid was crucial to favor adhesion of the hydrophobic protein on the carbon layer.

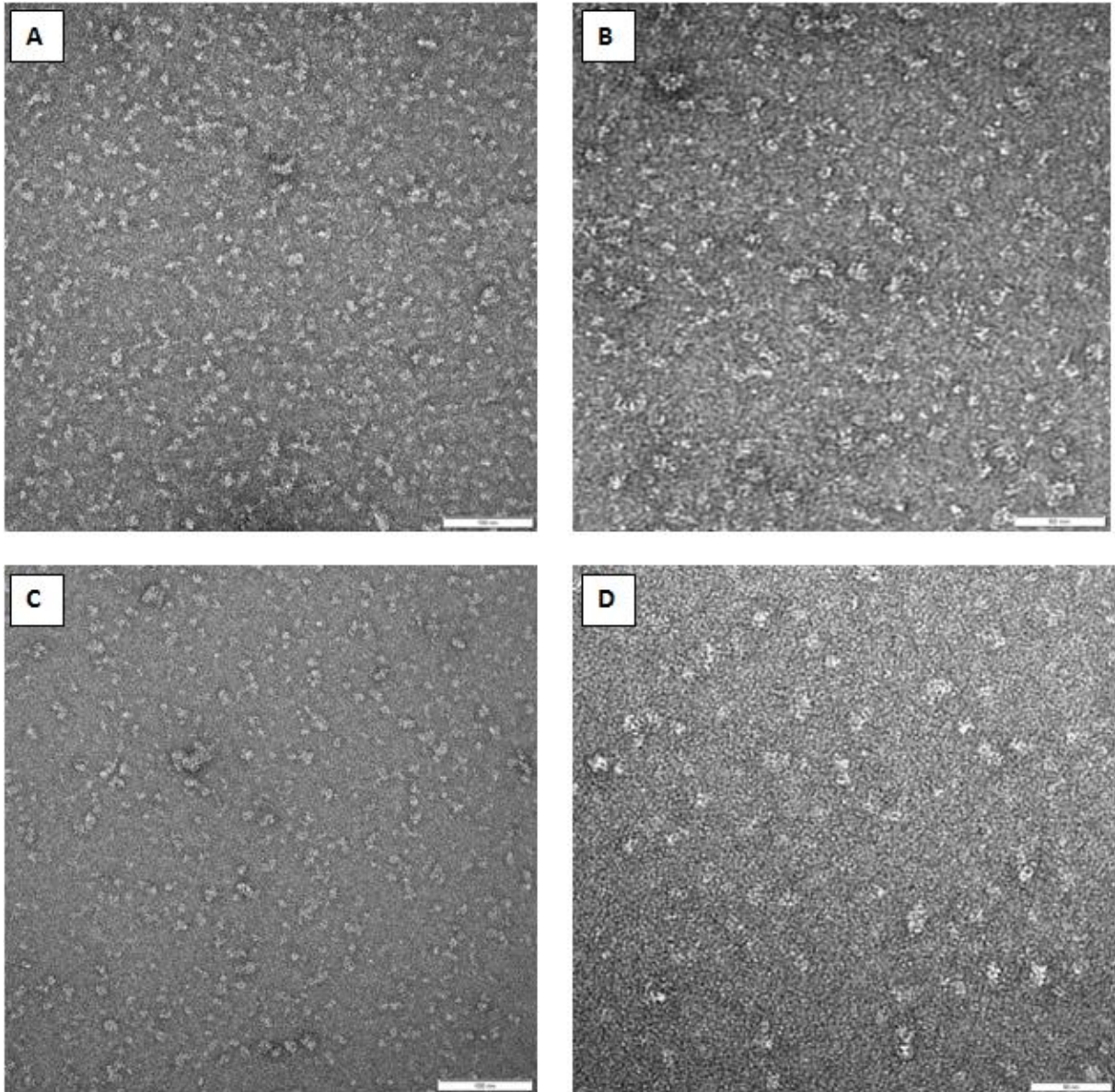


Figure 2.20: Negative staining EM images of the protein Neo1 in LMNG detergent. **A)** Protein concentration $40 \mu\text{g ml}^{-1}$ at 50'000x magnification, white bar corresponding to 100 nm; **B)** Protein concentration $40 \mu\text{g ml}^{-1}$ at 80'000x magnification, white bar 50 nm; **C)** Protein concentration $20 \mu\text{g ml}^{-1}$ at 50'000x magnification, white bar 100 nm; **D)** Protein concentration $20 \mu\text{g ml}^{-1}$ at 100'000x magnification, white bar 50 nm. Images were recorded in defocus mode from a TEM Philips EM208 at 100 kV, with a QUEMESA camera, using RADIUS software.

The images show the presence of a homogeneous sample, with very little protein aggregation. Particles with a similar shape can be easily observed in all the grids. In particular, a jellyfish shape can be distinguished, as shown in **Figure 2.21**. This shape could be due to the tightly packed TM domain, forming the head of the jellyfish, and the more

flexible cytosolic domains of Neo1, forming the tentacles. However, class averaging of a higher number of data is required to confirm the preliminary observations.

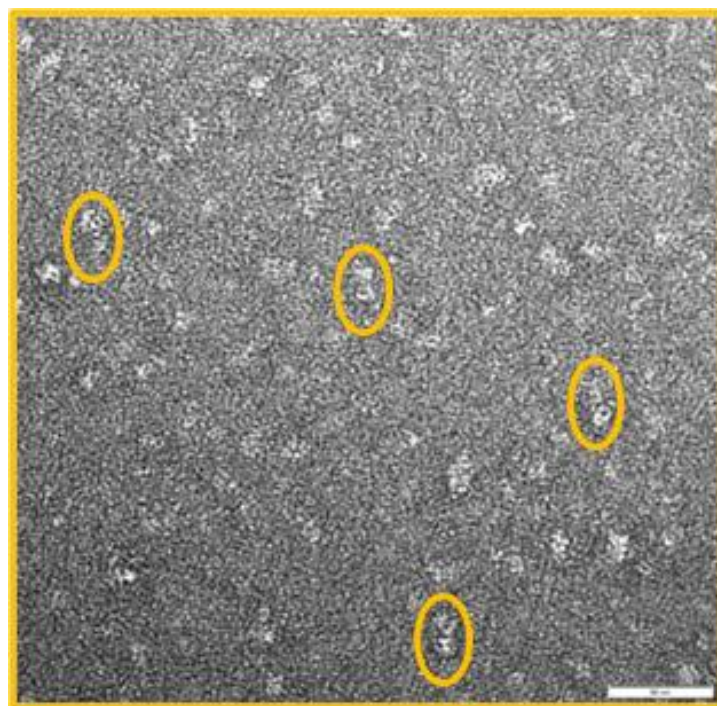


Figure 2.21: Recurrent jellyfish shape in negative staining EM images at 100'000x magnification. White bar: 50 nm.

Considering the quality of the images, grids can be used for further EM studies in negative staining. Moreover, the sample of the protein Neo1 purified in LMNG detergent is suitable for Cryo-EM structural studies.

2.3.10 *Neo1* mutants

The mutagenesis of specific residues is a useful tool to elucidate protein mechanisms at the molecular level. The analysis of changes in activity of the mutants compared to the wild-type protein allows to identify the role of each residue in key aspects of the protein mechanism, such as catalytic activity, structural stability, or substrate selection. However, a preliminary knowledge of the protein structure and the analysis of its sequence is essential for the selection of crucial residues for the mutagenesis study.

For the P-IV ATPase Neo1, the selection of residues of the polypeptide sequence possibly involved in essential aspects of the lipid translocation was the result of a careful analysis of

the protein alignment with other P-ATPases, mainly from the human and yeast genomes. In order to better understand Neo1 mechanism, my study focused on residues involved in substrate selection or translocation. The effect of the mutation of these residues was monitored through the analysis of changes in the ATPase activity of the protein.

Alignments between Neo1 and other *S. cerevisiae* and human P-IV ATPases allowed to identify four residues (R446, R460, R946 and E1065), possibly involved in interactions with the substrate, and located in protein segments that are expected to form the transmembrane helices TM4, TM5 and TM8, according to predictions based on the protein sequence. The secondary structure assignment for transmembrane helices of membrane proteins is based on an algorithm that considers the pattern of hydrophobic residues and the “positive-inside/negative-outside” empirical rule (Baker et al. 2017). In addition, the prediction is confirmed by sequence alignments to known P-IV ATPases.

The first two arginine residues selected, R446 and R460, are the first at the exoplasmic terminus of the TM4 helix and the second, at the opposite, at the cytosolic terminus of the same segment. Considering their location, these residues are expected to be involved in substrate engagement and disengagement. The second of these arginine residues, R460, in particular, precedes the conserved PISL motif, required for phospholipids identification (Y. Z. Chen et al. 2019). On the contrary, residues R956 and E1065 are located in the core of the transmembrane domain, along helices TM5 and TM8, and are typical of P-IV ATPases of subclass 2, a possible indication of their involvement in binding of the yet unknown specific substrate. In fact, the presence of an arginine residue in the center of TM5 helix is particularly interesting, since this is a specific feature of all P-IV ATPases of subclass 2, while other P-IV ATPases have a lysine residue in this position. In the P-IV ATPase ATP8A2, belonging to subclass 1A, this residue is involved in PS binding (Coleman et al. 2012). This particular location in the middle of TM5 has been recognized as an important “hot spot” involved in substrate binding also in other members of the larger family of P-ATPases, such as the Ca^{2+} -ATPase SERCA and in the Na^+/K^+ pump.

Alignment results are summarized in **Table 2.4**.

| CLASS | NAME | LOCATION | SPECIFICITY | Substrate selection | | | |
|-------|---------------|---------------------------|------------------------------|---------------------|------|-------|-------|
| | | | | TM4 | TM4 | TM5 | TM8 |
| 2 | Neo1 | Endosomes, late Golgi | Possibly PE | R446 | R460 | R946 | E1065 |
| | ATP9A | Endosomes, TGN | Unknown | R334 | R348 | R849 | E965 |
| | ATP9B | Endosomes, TGN | Unknown | R411 | R425 | R938 | E1054 |
| 4 | DNF3 | Golgi, vesicles | PC (PE) | S503 | Y517 | K1293 | I1416 |
| 5 | ATP10A | Plasma (ER) | PC (PE) | T364 | Y378 | K1095 | F1215 |
| | ATP10B | Vesicles | Unknown | T370 | Y384 | K1119 | I1239 |
| | ATP10D | Plasma | Unknown | T375 | Y389 | K1120 | V1240 |
| 3 | DNF1 | Plasma | PC, PE (PS), LysoPC, Lyso PE | V604 | Y618 | K1194 | C1323 |
| | DNF2 | Plasma | PC, PE (PS), LysoPC, Lyso PE | V650 | Y664 | K1238 | C1367 |
| 6 | ATP11B | Plasma, endosomes | PS>PE | A344 | Y358 | K885 | V1013 |
| | ATP11A | Plasma, endosomes | PS>PE | A351 | Y365 | K889 | V1018 |
| | ATP11C | Plasma | PS>PE | S349 | Y363 | K883 | V1012 |
| 1B | ATP8B3 | Plasma | Possibly PS | S432 | F446 | K1003 | I1128 |
| | ATP8B1 | Plasma, apical membrane | Possibly PC | G391 | Y405 | K957 | V1084 |
| | ATP8B2 | Plasma | PC | S348 | Y362 | K897 | V1024 |
| | ATP8B4 | Plasma | Unknown | S329 | Y343 | K879 | V1006 |
| 1A | DRS2 | Golgi, vesicles | PS, PE | / | F511 | K1018 | V1146 |
| | ATP8A1 | Golgi, vesicles | PS>PE | / | L360 | K865 | V992 |
| | ATP8A2 | Golgi, photoreceptor disc | PS>PE | / | L379 | K885 | V1012 |

Table 2.4: Results of the yeast and human P-IV ATPases alignments as regards residues selected for mutagenesis studies in this work. Blue and light-blue background: positively charged arginine and lysine residues, respectively; red background: negatively charged residues; yellow background: polar residues; grey background: non-polar residues.

Considering the structural information available on the Protein Data Bank and the sequence alignments, a first computational model of the protein Neo1 was generated with the online SwissModel tool based on the structure of the P-ATPase SERCA (PDB: 5MPM) (Sitsel et al. 2019). In this software, model quality is assessed through two parameters, the Global Model Quality Estimation, GMQE, and the QMEAN, a more complex value that considers both local and global contributions to the quality estimation. GMQE has values between 0 and 1, increasing with the quality of the structure, while for QMEAN values below -4 are considered

a strong indication of a poor model quality. For the model based on the SERCA structure, GMQE and QMEAN values of 0.47 and -5.33, respectively, were obtained, a strong indication that the sequence identity between Neo1 and SERCA (18.16%) is not sufficient to calculate a reliable model.

When structures of the P-IV ATPase Drs2 became available (PDB: 6ROH, 6ROJ and 6ROI) (Timcenko et al. 2019), new models of Neo1 were obtained using these paralog, with sequence identity of 31.48%. The better GMQE and QMEAN scores were obtained using the structure with PDB code 6ROJ, with values of 0.61 and -3.68, respectively.

The two computational models obtained for Neo1 were compared with the software Chimera, as shown in **Figure 2.22**, and did not reveal significant differences as regards the overall protein arrangement, nor the local environment of residues selected for mutagenesis in the TM domain. Surprisingly, the main differences between the predicted models involve the cytosolic domains and may be related to the different stages of the catalytic cycle that are represented by the experimental structures chosen as starting models.

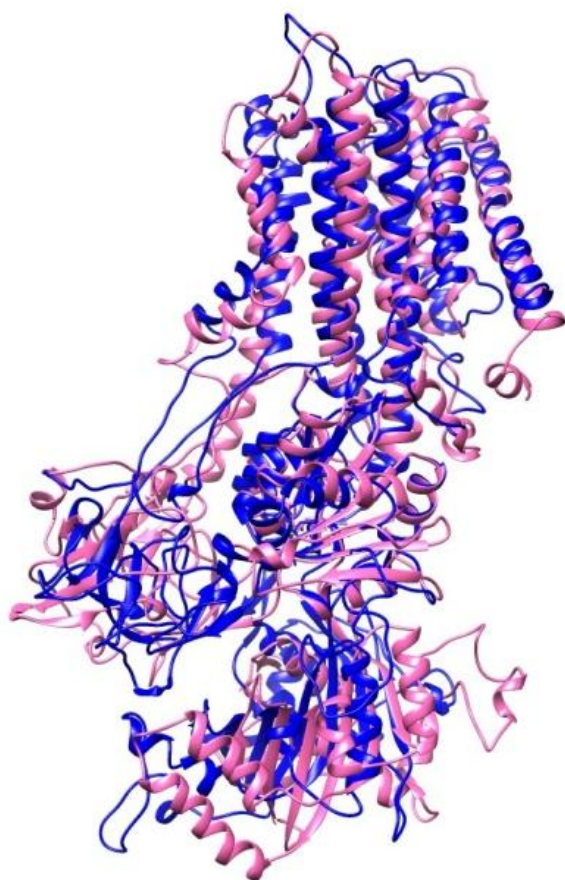


Figure 2.22: Homology models obtained for the P-IV ATPase Neo1 using the SwissModel software. Blue: model obtained from the crystallographic structure of SERCA (5MPM); Magenta: model obtained from the cryo-EM structure of Drs2 (6ROJ).

The analysis of the model structures shows that the residues selected for mutagenesis, R446, R460, R946 and E1065, form a sort of charged channel (**Figure 2.23a**). The residue R446 may

be involved in the identification of the polar headgroup of the phospholipid, at the entry of the channel. Residues R946 and E1065, located in the middle of the channel, could shift the lipid towards the end of the channel. Finally, residue R460 may have a role in the release mechanism at the channel exit. The positions of residues R946 and E1065 are particularly interesting, as they seem to be involved in a salt-bridge interaction in the center of the TM domain (**Figure 2.23b**).

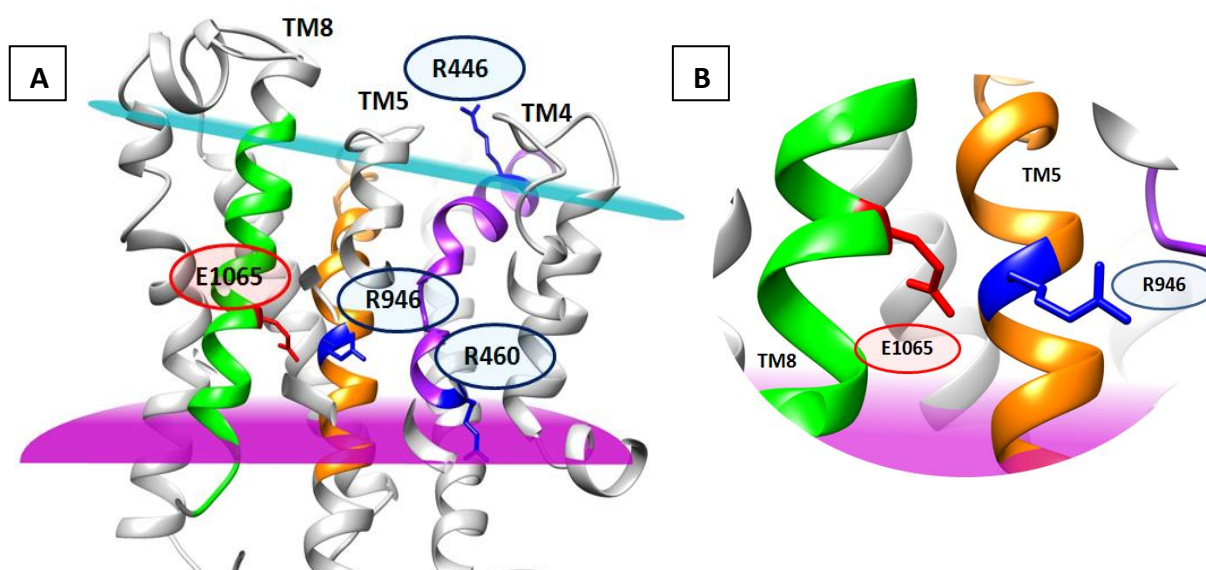


Figure 2.23: Residues possibly involved in substrate binding in the computational model of Neo1, forming a charged channel for lipid sliding. **A)** Overview of the TM domain. **B)** Close view of the R946-E1065 salt bridge forming at the center of the charged channel.

2.3.11 Neo1 mutants on substrate selection capacity overexpression

To verify the hypothesis drawn from the sequence alignment and the computational modeling, we performed further mutagenesis experiments on residues R446, R460, R946 and E1065. The positive charges of the arginine residues were replaced with: i) an alanine residue (A) whose methyl group completely removes the positive charge; ii) an aspartic acid residue (D), with the opposite charge of the carboxylic group and a significantly shorter carbon chain; iii) a glutamic acid residue (E), with an opposite charge compared to the original arginine, but a longer carbon chain than the aspartic acid; and iv) a lysine residue (K), that maintains the positive charge, but removes the guanidinium group. In mutants of the glutamic acid residue E1065, the negative charge was replaced with: i) an alanine residue, completely removing charges; ii) an aspartic acid residue, that maintains the negative charge

but decrease the carbon chain length; and iii) an arginine residue, with an opposite charge. Furthermore, a double mutation on R946 and E1065 was created by inverting both charges on these residues, in the R946E/E1065R mutant, in order to maintain the salt-bridge interaction in the middle of the TM channel, but changing the positions of the positive and negative poles.

pDDGFP-LEU2D plasmids carrying the mutations on R446, R460, R946 or E1065 amino acids have been obtained with site-directed mutagenesis and verified by sequencing analysis. The FGY217 yeast strains were transformed with 15 plasmids encoding for the mutants R446A, R446D, R446E, R446K, R460D, R460E, R460K, R946A, R946D, R946K, R946K, E1065A, E1065D, E1065R, and R946E/E1065R. **Figure 2.24** shows the GFP fluorescence signals indicative of protein overexpression for small scale cultures in URA⁻ or LEU⁻ media, after 22 h from induction with 2% w/v galactose .

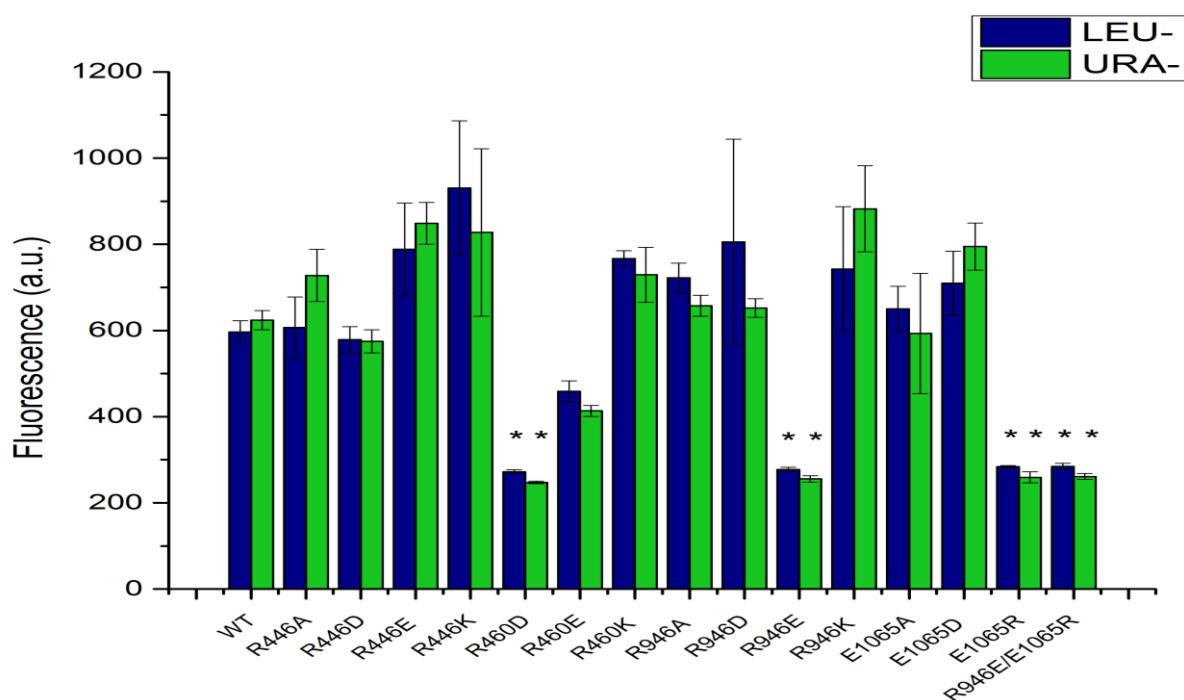


Figure 2.24: Overexpression tests on the Neo1 mutants in LEU⁻ and URA⁻ media, compared to the WT, followed by GFP fluorescence signal. A negative control with no induction (not shown) was added to the test. Blue bars: induction in LEU⁻ medium; green bars: induction in URA⁻ medium. Asterisks (*) show conditions where no protein expression was detected. Error bars: standard deviation of three fluorescence measurements.

The medium without leucine seems to yield a slightly better protein overexpression than the medium without uracil. Interestingly, mutants R460D, E1065R, R946E and R946E/E1065R could not be expressed in this system, suggesting an important role for these residues in

protein folding or toxicity. Mutants E1065R and R946E invert the charge in the middle of the TM domain, probably affecting the salt-bridge formation and the stability of the overall TM domain. However, the double mutant R946E/E1065R, that could potentially restore the electrostatic interaction, was not expressed in the tested conditions.

2.4 CONCLUSIONS AND FUTURE PERSPECTIVES

P-IV ATPases are membrane proteins whose role is to translocate phospholipids from the external to the internal side of the cellular membrane. Their role is so crucial for the cell life, that malfunctions of these proteins in humans lead to severe diseases like hepatic cholestasis, diabetes, cancer and Alzheimer's disease. Given their relevance, the understanding of their structure and mechanism of action is fundamental for their medical implications.

This study focused on Neo1, an atypical P-IV ATPase of *S. cerevisiae* whose substrates are still unknown. I was able to set up and optimize a protein production protocol from cloning to expression and purification that allowed to obtain high amounts of pure Neo1, necessary for further characterization tests. The protein was cloned into the pDDGFP_LEU2D plasmid, that allows for the expression of a recombinant protein with a C-terminal GFP/8His tandem tag. The His tag was required for purification of Neo1, while the GFP was useful for protein detection and identification in all the overexpression and purification processes. Many overexpression conditions were tested, leading to the selection of the best conditions for expression of Neo1 in high yields. Several detergents were tested to purify this protein, whose instability outside the hydrophobic membrane environment is well known. Only using LMNG I successfully purified a stable form of the protein, while all other detergents tested resulted in aggregation. After the optimization of the purification conditions in LMNG, I obtained about 0.5 mg of pure protein per liter of culture. Mutants on the phosphorylation and dephosphorylation conserved motifs were created to decrease the ATPase activity of Neo1, reducing its toxicity for the host cell during overexpression. I was able to identify different mutants that, in particular growth conditions, led to higher protein overexpression.

Analyzing the protein activity of a wild-type Neo1 purified sample in LMNG, I concluded that the protein is properly folded and still active even in the detergent buffer, as it is able to translocate lipids from crude *S. cerevisiae* membranes. Addition of PC, PE or PS lipids to the purified sample, however, did not result in ATP consumption, indicating that, while the protein is still active, PC, PE and PS lipids are not translocated in these conditions. This result leaves an open question regarding Neo1 substrates and activation mode, as the lack of activity may be related to the absence of an activator molecule, a role covered by PI4P for other P-ATPases, or to the unsuitable substrates added to the protein solution. To this date, Neo1 activation mechanism is still unknown. Further experiments with different lipids as substrates will be performed on purified Neo1 in order to answer this question.

A high number of crystallization screens were tested on the purified Neo1 recombinant protein, but experiments have not yet yielded positive results and crystallization trials are still ongoing. As crystals of membrane proteins are difficult to obtain, I decided to investigate Neo1 structure through a different technique, using EM analysis. Negative staining EM analysis on this sample gave excellent results and good quality images at 50-100k magnification were obtained. In these images, an homogeneous specimen is clearly visible and the overall shape of the protein can be distinguished. These results are a valuable starting point for further Cryo-EM analysis of Neo1, that will be executed in collaboration with Prof. Venien-Bryan in Paris

After accurate alignment with other P-IV ATPases and calculation of a computational model that predicts Neo1 structure, four residues were identified for their potential involvement in the substrate translocation mechanism. These residues were mutated to change their charge or steric hindrance and cloned in the same pDDGFP-LEU2D vector, obtaining a final number of 11 rationally-designed mutants expressed in yeast. However, some of the identified mutations did not produce viable protein, suggesting the crucial role of these residues in protein folding. For the continuation of this study, activity tests will be performed on the overexpressed mutants.

REFERENCES

- 1) Alder-Baerens, Nele et al. 2006. "Loss of P4 ATPases Drs2p and Dnf3p Disrupts Aminophospholipid Transport and Asymmetry in Yeast Post-Golgi Secretory Vesicles." *Molecular Biology of the Cell* 17: 1632–42.
- 2) Andersen, Jens P. et al. 2016. "P4-ATPases as Phospholipid Flippases-Structure, Function, and Enigmas." *Frontiers in Physiology* 7(JUL): 1–23.
- 3) Auland, Merran E., Basil D. Roufogalis, Philippe F. Devaux, and Alain Zachowski. 1994. "Reconstitution of ATP-Dependent Aminophospholipid Translocation in Proteoliposomes." *Proceedings of the National Academy of Sciences of the United States of America* 91(23): 10938–42.
- 4) Azouaoui, Hassina et al. 2017. "High Phosphatidylinositol 4-Phosphate (PI4P)-Dependent ATPase Activity for the Drs2p-Cdc50p Flippase after Removal of Its N- and C-Terminal Extensions." *Journal of Biological Chemistry* 292(19): 7954–70.
- 5) Baker, James Alexander et al. 2017. "Charged Residues next to Transmembrane Regions Revisited: 'Positive-inside Rule' Is Complemented by the 'Negative inside Depletion/Outside Enrichment Rule.'" *BMC Biology* 15(66): 1–29.
- 6) Barbosa, Sónia et al. 2010. "Oligomeric Dop1p Is Part of the Endosomal Neo1p-Ysl2p-Arl1p Membrane Remodeling Complex." *Traffic* 11(8): 1092–1106.
- 7) Bublitz, Maike, J. Preben Morth, and Poul Nissen. 2011. "P-Type ATPases at a Glance." *Journal of Cell Science* 124(22): 2515–19.
- 8) Bublitz, Maike, Hanne Poulsen, J. Preben Morth, and Poul Nissen. 2010. "In and out of the Cation Pumps: P-Type ATPase Structure Revisited." *Current Opinion in Structural Biology* 20(4): 431–39.
- 9) Bull, Laura N. et al. 1998. "A Gene Encoding a P-Type ATPase Mutated in Two Forms of Hereditary Cholestasis." *Nature Genetics* 18(3): 219–24.
- 10) Cacciagli, Pierre et al. 2010. "Disruption of the ATP8A2 Gene in a Patient with a t(10;13) de Novo Balanced Translocation and a Severe Neurological Phenotype." *European Journal of Human Genetics* 18(12): 1360–63.
- 11) Chalal, Madhavan, Kody Moleschi, Robert S. Molday, and Thomas F.J. Martin. 2017. "C-Terminus of the P4-ATPase ATP8A2 Functions in Protein Folding and Regulation of Phospholipid Flippase Activity." *Molecular Biology of the Cell* 28(3): 452–62.
- 12) Chantalat, Sophie et al. 2004. "The Arf Activator Gea2p and the P-Type ATPase Drs2p Interact at the Golgi in *Saccharomyces Cerevisiae*." *Journal of Cell Science* 117(5): 711–22.
- 13) Chen, Chih-ying, Michael F Ingram, Peter H Rosal, and Todd R Graham. 1999. "Role for Drs2p, a P-Type ATPase and Potential Aminophospholipid Translocase, in Yeast Late Golgi Function." *the journal of cell biology* 147(6): 1223–36.
- 14) Chen, Yu Zen et al. 2019. "Structure and Function Analysis of the C. Elegans Aminophospholipid Translocase TAT-1." *Journal of Cell Science* 132(5): 1–7.
- 15) Coleman, Jonathan A. et al. 2012. "Critical Role of a Transmembrane Lysine in Aminophospholipid

Chapter 2

- Transport by Mammalian Photoreceptor P 4-ATPase ATP8A2." *Proceedings of the National Academy of Sciences of the United States of America* 109(5): 1449–54.
- 16) Coleman, Jonathan A. et al. 2014. "Phospholipid Flippase ATP8A2 Is Required for Normal Visual and Auditory Function and Photoreceptor and Spiral Ganglion Cell Survival." *Journal of Cell Science* 127(5): 1138–49.
 - 17) Coleman, Jonathan A., Michael C.M. Kwok, and Robert S. Molday. 2009. "Localization, Purification, and Functional Reconstitution of the P4-ATPase Atp8a2, a Phosphatidylserine Flippase in Photoreceptor Disc Membranes." *Journal of Biological Chemistry* 284(47): 32670–79.
 - 18) Dalton, Lauren E., Björn D.M. Bean, Michael Davey, and Elizabeth Conibear. 2017. "Quantitative High-Content Imaging Identifies Novel Regulators of Neol1 Trafficking at Endosomes." *Molecular Biology of the Cell* 28(11): 1539–50.
 - 19) Dhar, Madhu S. et al. 2004. "Mice Heterozygous for Atp10c, a Putative Amphipath, Represent a Novel Model of Obesity and Type 2 Diabetes." *The Journal of Nutrition* 134(4): 799–805.
 - 20) Dhar, Madhu S., Joshua S. Yuan, Sarah B. Elliott, and Carla Sommardahl. 2006. "A Type IV P-Type ATPase Affects Insulin-Mediated Glucose Uptake in Adipose Tissue and Skeletal Muscle in Mice." *Journal of Nutritional Biochemistry* 17(12): 811–20.
 - 21) Ding, Jiantao et al. 2000. "Identification and Functional Expression of Four Isoforms of ATPase II, the Putative Aminophospholipid Translocase: Effect of Isoform Variation on the ATPase Activity and Phospholipid Specificity." *Journal of Biological Chemistry* 275(30): 23378–86.
 - 22) Drew, David et al. 2005. "A Scalable, GFP-Based Pipeline for Membrane Protein Overexpression Screening and Purification." *Protein Science* 14(8): 2011–17.
 - 23) Drew, David et al. 2008. "GFP-Based Optimization Scheme for the Overexpression and Purification of Eukaryotic Membrane Proteins in *Saccharomyces Cerevisiae*." *Nature Protocols* 3(5): 784–98.
 - 24) Eppens, Elaine F. et al. 2001. "FIC1, the Protein Affected in Two Forms of Hereditary Cholestasis, Is Localized in the Cholangiocyte and the Canalicular Membrane of the Hepatocyte." *Journal of Hepatology* 35(4): 436–43.
 - 25) Flamant, Stéphane et al. 2003. "Characterization of a Putative Type IV Aminophospholipid Transporter P-Type ATPase." *Mammalian Genome* 14(1): 21–30.
 - 26) Folmer, Dineke E., Ronald P.J. Oude Elferink, and Coen C. Paulusma. 2009. "P4 ATPases - Lipid Flippases and Their Role in Disease." *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1791(7): 628–35.
 - 27) Gil de Rubio, Rafael et al. 2018. "Phosphatidylinositol 4-Phosphate Is a Major Source of GPCR-Stimulated Phosphoinositide Production." *Science Signaling* 11(547).
 - 28) Glaser, Michael, and S. J. Singer. 1971. "Circular Dichroism and the Conformations of Membrane Proteins. Studies with Red Blood Cell Membranes." *Biochemistry* 10(10): 1780–87.
 - 29) Hanamatsu, Hisatoshi et al. 2014. "Interaction of the Phospholipid Flippase Drs2p with the F-Box Protein Rcy1p Plays an Important Role in Early Endosome to Trans-Golgi Network Vesicle Transport in Yeast." *Journal of Biochemistry* 155(1): 51–62.
 - 30) Hiraizumi, Masahiro, Keitaro Yamashita, Tomohiro Nishizawa, and Osamu Nureki. 2019. "Cryo-EM

Chapter 2

- Structures Capture the Transport Cycle of the P4-ATPase Flippase." *Science* 365(6458): 1149–55.
- 31) Hirayama, Hiroshi et al. 2013. "ATPase Activity of Human ABCG1 Is Stimulated by Cholesterol and Sphingomyelin." *Journal of Lipid Research* 54(2): 496–502.
 - 32) Hua, Zhaolin, Parvin Fatheddin, and Todd R. Graham. 2002. "An Essential Subfamily of Drs2p-Related P-Type ATPases Is Required for Protein Trafficking between Golgi Complex and Endosomal/Vacuolar System." *Molecular biology of the cell* 13: 3162–77.
 - 33) Hua, Zhaolin, and Todd R. Graham. 2003. "Requirement for Neo1p in Retrograde Transport from the Golgi Complex to the Endoplasmic Reticulum." *Molecular Biology of the Cell* 14: 4971–83.
 - 34) Jacquot, Aurore et al. 2012. "Phosphatidylserine Stimulation of Drs2p-Cdc50p Lipid Translocase Dephosphorylation Is Controlled by Phosphatidylinositol-4-Phosphate." *Journal of Biological Chemistry* 287(16): 13249–61.
 - 35) Jencks, William P. 1989. "Utilization of Binding Energy and Coupling Rules for Active Transport and Other Coupled Vectorial Processes." *Methods in Enzymology* 171(C): 145–64.
 - 36) Jensen, M. S. et al. 2017. "Phospholipid Flipping Involves a Central Cavity in P4 ATPases." *Scientific Reports* 7(1): 1–13.
 - 37) Kato, Utako et al. 2013. "Role for Phospholipid Flippase Complex of ATP8A1 and CDC50A Proteins in Cell Migration." *Journal of Biological Chemistry* 288(7): 4922–34.
 - 38) Kühlbrandt, Werner. 2004. "Biology, Structure and Mechanism of P-Type ATPases." *Nature Reviews Molecular Cell Biology* 5(4): 282–95.
 - 39) De La Hera, Diego P., Gerardo R. Corradi, Hugo P. Adamo, and Felicitas De Tezanos Pinto. 2013. "Parkinson's Disease-Associated Human P5B-ATPase ATP13A2 Increases Spermidine Uptake." *Biochemical Journal* 450(1): 47–53.
 - 40) Lee, Shoken et al. 2015. "Transport through Recycling Endosomes Requires EHD 1 Recruitment by a Phosphatidylserine Translocase." *The EMBO Journal* 34(5): 669–88.
 - 41) Lenoir, Guillaume, Patrick Williamson, and Joost C. Holthuis. 2007. "On the Origin of Lipid Asymmetry: The Flip Side of Ion Transport." *Current Opinion in Chemical Biology* 11(6): 654–61.
 - 42) Levano, Kelly, Vineet Punia, Michael Raghunath, and Probal Banerjee. 2012. "Atp8a1 Deficiency Is Associated with Phosphatidylserine Externalization in Hippocampus and Delayed Hippocampus-Dependent Learning." *Journal of neurochemistry* 120(2): 302–13.
 - 43) Li, Hao et al. 2008. "Candidate Single-Nucleotide Polymorphisms from a Genomewide Association Study of Alzheimer Disease." *Archives of Neurology* 65(1): 45–53.
 - 44) Lopez-Marques, Rosa L., Lisa Theorin, Michael G. Palmgren, and Thomas Günther Pomorski. 2014. "P4-ATPases: Lipid Flippases in Cell Membranes." *Pflugers Archiv European Journal of Physiology* 466(7): 1227–40.
 - 45) Malmström, Susanna, Hans Erik Åkerlund, and Per Askerlund. 2000. "Regulatory Role of the N Terminus of the Vacuolar Calcium-ATPase in Cauliflower." *Plant Physiology* 122(2): 517–26.
 - 46) van der Mark, Vincent A., Ronald P.J. Oude Elferink, and Coen C. Paulusma. 2013. "P4 ATPases: Flippases in Health and Disease." *International Journal of Molecular Sciences* 14(4): 7897–7922.
 - 47) Miyoshi, Norikatsu et al. 2010. "ATP11A Is a Novel Predictive Marker for Metachronous Metastasis of

Chapter 2

- Colorectal Cancer." *Oncology reports* 23: 505–10.
- 48) Moriyama, Y., and N. Nelson. 1988. "Purification and Properties of a Vanadate- and N-Ethylmaleimide-Sensitive ATPase from Chromaffin Granule Membranes." *Journal of Biological Chemistry* 263(17): 8521–27.
- 49) Morrot, Gil, Alain Zachowski, and Philippe F. Devaux. 1990. "Partial Purification and Characterization of the Human Erythrocyte Mg²⁺-ATPase A Candidate Aminophospholipid Translocase." *FEBS Letters* 266(1–2): 29–32.
- 50) Nakano, Kenzi et al. 2008. "Protein Kinases Fpk1p and Fpk2p Are Novel Regulators of Phospholipid Asymmetry." *Molecular biology of the cell* 19: 1783–97.
- 51) Natarajan, Paramasivam et al. 2009. "Regulation of a Golgi Flippase by Phosphoinositides and an ArfGEF." *Nature Cell Biology* 11(12): 1421–26.
- 52) Natarajan, Paramasivam, Jiyi Wang, Zhaolin Hua, and Todd R. Graham. 2004. "Drs2p-Coupled Aminophospholipid Translocase Activity in Yeast Golgi Membranes and Relationship to in Vivo Function." *Proceedings of the National Academy of Sciences of the United States of America* 101(29): 10614–19.
- 53) Newstead, Simon et al. 2007. "High-Throughput Fluorescent-Based Optimization of Eukaryotic Membrane Protein Overexpression and Purification in *Saccharomyces Cerevisiae*." *Proceedings of the National Academy of Sciences of the United States of America* 104(35): 13936–41.
- 54) Noji, Takehiro et al. 2006. "Mutational Analysis of the Lem3p-Dnf1p Putative Phospholipid-Translocating P-Type ATPase Reveals Novel Regulatory Roles for Lem3p and a Carboxyl-Terminal Region of Dnf1p Independent of the Phospholipid-Translocating Activity of Dnf1p in Yeast." *Biochemical and Biophysical Research Communications* 344(1): 323–31.
- 55) Okumura, Masaki, and Toshinori Kinoshita. 2016. "Measurement of ATP Hydrolytic Activity of Plasma Membrane H⁺-ATPase from *Arabidopsis Thaliana* Leaves." *Bio-Protocol* 6(23): e2044.
- 56) Onat, Onur Emre et al. 2013. "Missense Mutation in the ATPase, Aminophospholipid Transporter Protein ATP8A2 Is Associated with Cerebellar Atrophy and Quadrupedal Locomotion." *European Journal of Human Genetics* 21(3): 281–85.
- 57) Palmgren, Michael G, and Kristian B Axelsen. 1998. "Evolution of P-Type ATPases." 1365: 37–45.
- 58) Palmgren, Michael G, and Poul Nissen. 2011. "P-Type ATPases Phosphorylated Intermediate-Type ATPase (P-Type ATPase): A Family of Membrane-Embedded Biological Pumps with a Single Catalytic Subunit That Form a Phosphorylated Intermediate during Each Catalytic Cycle." *Annu. Rev. Biophys* 40: 243–66.
- 59) Parker, Joanne L., and Simon Newstead. 2014. "Method to Increase the Yield of Eukaryotic Membrane Protein Expression in *Saccharomyces Cerevisiae* for Structural and Functional Studies." *Protein Science* 23(9): 1309–14.
- 60) Paulusma, Coen C. et al. 2006. "Atp8b1 Deficiency in Mice Reduces Resistance of the Canalicular Membrane to Hydrophobic Bile Salts and Impairs Bile Salt Transport." *Hepatology* 44(1): 195–204.
- 61) Pedersen, L. Peter, and Ernesto Carafoli. 1987. "Ion Motive ATPases. I. Ubiquity, Properties, and Significance to Cell Function." *Trends in Biochemical Science* 12: 146–50.

Chapter 2

- 62) Pomorski, Thomas et al. 2003. "Drs2p-Related P-Type ATPases Dnf1p and Dnf2p Are Required for Phospholipid Translocation across the Yeast Plasma Membrane and Serve a Role in Endocytosis." *Molecular Biology of the Cell* 14: 1240–54.
- 63) Pomorski, Thomas Günther, and K. Anant Menon. 2016. "Lipid Somersaults: Uncovering the Mechanism of Protein-Mediated Lipid Flipping." *Progress in Lipid Research* 64: 69–84.
- 64) Prezant, Toni R., William E. Chaltraw, and Nathan Fischel-Ghodsian. 1996. "Identification of an Overexpressed Yeast Gene Which Prevents Aminoglycoside Toxicity." *Microbiology* 142(12): 3407–14.
- 65) Roelants, Françoise M. et al. 2010. "A Protein Kinase Network Regulates the Function of Aminophospholipid Flippases." *Proceedings of the National Academy of Sciences of the United States of America* 107(1): 34–39.
- 66) Roland, Bartholomew P. et al. 2018. "Yeast and Human P4-ATPases Transport Glycosphingolipids Using Conserved Structural Motifs." *Journal of Biological Chemistry* 294(6): 1794–1806.
- 67) Roland, Bartholomew P., and Todd R. Graham. 2016a. "Directed Evolution of a Sphingomyelin Flippase Reveals Mechanism of Substrate Backbone Discrimination by a P4-ATPase." *Proceedings of the National Academy of Sciences of the United States of America* 113(31): E4460–66.
- 68) Roland, Bartholomew P., and Todd R. Graham. 2016b. "Decoding P4-ATPase Substrate Interactions." *Critical Reviews in Biochemistry and Molecular Biology* 51(6): 513–27.
- 69) Sakane, Hiroshi, Takaharu Yamamoto, and Kazuma Tanaka. 2006. "The Functional Relationship between the Cdc50p-Drs2p Putative Aminophospholipid Translocase and the Arf GAP Gcs1p in Vesicle Formation in the Retrieval Pathway from Yeast Early Endosomes to the TGN." *Cell Structure and Function* 31(2): 87–108.
- 70) Schultheis, Patrick J. et al. 2004. "Characterization of the P 5 Subfamily of P-Type Transport ATPases in Mice." *Biochemical and Biophysical Research Communications* 323(3): 731–38.
- 71) Sebastian, Tessy T., Ryan D. Baldrige, Peng Xu, and Todd R. Graham. 2012. "Phospholipid Flippases: Building Asymmetric Membranes and Transport Vesicles." *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1821(8): 1068–77.
- 72) Segawa, Katsumori, Sachiko Kurata, and Shigekazu Nagata. 2016. "Human Type IV P-Type ATPases That Work as Plasma Membrane Phospholipid Flippases and Their Regulation by Caspase and Calcium." *Journal of Biological Chemistry* 291(2): 762–72.
- 73) Shinoda, Takehiro, Haruo Ogawa, Flemming Cornelius, and Chikashi Toyoshima. 2009. "Crystal Structure of the Sodium-Potassium Pump at 2.4 Resolution." *Nature* 459(7245): 446–50.
- 74) Siggs, M. Owen et al. 2011. "The P4-Type ATPase ATP11C Is Essential for B Lymphopoiesis in Adult Bone Marrow." *Nature Immunology* 12(5): 434–40.
- 75) Sitsel, Aljona et al. 2019. "Structures of the Heart Specific SERCA 2a Ca²⁺ - ATP Ase ." *The EMBO Journal* 38(5): 1–17.
- 76) Stapelbroek, Janneke M. et al. 2009. "ATP8B1 Is Essential for Maintaining Normal Hearing." *Proceedings of the National Academy of Sciences of the United States of America* 106(24): 9709–14.
- 77) Strautnieks, Sandra S et al. 1998. "A Gene Encoding a Liver-Specific ABC Transporter Is Mutated in Progressive Familial Intrahepatic Cholestasis." *Nature Genetics* 20(3): 233–38.

Chapter 2

- 78) Takar, Mehmet, Yuntai Wu, and Todd R. Graham. 2016. "The Essential Neo1 Protein from Budding Yeast Plays a Role in Establishing Aminophospholipid Asymmetry of the Plasma Membrane." *Journal of Biological Chemistry* 291(30): 15727–39.
- 79) Takatsu, Hiroyuki et al. 2011. "ATP9B, a P4-ATPase (a Putative Aminophospholipid Translocase), Localizes to the Trans-Golgi Network in a CDC50 Protein-Independent Manner." *Journal of Biological Chemistry* 286(44): 38159–67.
- 80) Takatsu, Hiroyuki et al. 2014. "Phospholipid Flippase Activities and Substrate Specificities of Human Type IV P-Type ATPases Localized to the Plasma Membrane." *Journal of Biological Chemistry* 289(48): 33543–56.
- 81) Tanaka, Kazuma, Konomi Fujimura-Kamada, and Takaharu Yamamoto. 2011. "Functions of Phospholipid Flippases." *Journal of Biochemistry* 149(2): 131–43.
- 82) Tang, Yuqian et al. 2013. "Secretory Expression and Characterization of a Novel Peroxiredoxin for Zearalenone Detoxification in *Saccharomyces Cerevisiae*." *Microbiological Research* 168(1): 6–11.
- 83) Taylor, Kenneth A., Laszlo Dux, and Anthony Martonosi. 1986. "Three-Dimensional Reconstruction of Negatively Stained Crystals of the Ca²⁺-ATPase from Muscle Sarcoplasmic Reticulum." *Journal of Molecular Biology* 187(3): 417–27.
- 84) Timcenko, Milena et al. 2019. "Structure and Autoregulation of a P4-ATPase Lipid Flippase." *Nature* 571(7765):366-370.
- 85) Toyoshima, Chikashi, Masayoshi Nakasako, Hiromi Nomura, and Haruo Ogawa. 2000. "Crystal Structure of the Calcium Pump of Sarcoplasmic Reticulum at 2.6 Å Resolution." *Nature* 405(6787): 647–55.
- 86) Tsai, Pei Chin et al. 2013. "Arl1p Regulates Spatial Membrane Organization at the Trans-Golgi Network through Interaction with Arf-GEF Gea2p and Flippase Drs2p." *Proceedings of the National Academy of Sciences of the United States of America* 110(8):E668-77.
- 87) Van Der Velden, Lieke M., Stan F J Van De Graaf, and Leo W J Klomp. 2010. "Biochemical and Cellular Functions of P4 ATPases." *Biochemical Journal* 431(1): 1–11.
- 88) Vestergaard, Anna L. et al. 2014. "Critical Roles of Isoleucine-364 and Adjacent Residues in a Hydrophobic Gate Control of Phospholipid Transport by the Mammalian P4-ATPase ATP8A2." *Proceedings of the National Academy of Sciences of the United States of America* 111(28): E1334–43.
- 89) Wang, Lei, Crystal Beserra, and David L. Garbers. 2004. "A Novel Aminophospholipid Transporter Exclusively Expressed in Spermatozoa Is Required for Membrane Lipid Asymmetry and Normal Fertilization." *Developmental Biology* 267(1): 203–15.
- 90) Wehman, Ann M et al. 2011. "The P4-ATPase TAT-5 Inhibits the Budding of Extracellular Vesicles in *C. Elegans* Embryos." *Current Biology* 21(23): 1951–59.
- 91) Wicky, S., H. Schwarz, and B. Singer-Kruger. 2004. "Molecular Interactions of Yeast Neo1p, an Essential Member of the Drs2 Family of Aminophospholipid Translocases, and Its Role in Membrane Trafficking within the Endomembrane System." *Molecular and Cellular Biology* 24(17): 7402–18.
- 92) van der Woerd, W. et al. 2016. "Defective Plasma Membrane Targeting of P.I661T-ATP8B1, Associated with Familial Intrahepatic Cholestasis, Can Be Rescued in Vitro by CFTR Correctors." *Journal of*

Chapter 2

- Hepatology* 64(2): S304.
- 93) Xu, Chen, William J. Rice, Wanzhong He, and David L. Stokes. 2002. "A Structural Model for the Catalytic Cycle of Ca²⁺-ATPase." *Journal of Molecular Biology* 316(1): 201–11.
- 94) Xu, Peng et al. 2009. "Identification of a Novel Mouse P4-ATPase Family Member Highly Expressed during Spermatogenesis." *Journal of Cell Science* 122(16): 2866–76.
- 95) Yabas, Mehmet et al. 2014. "Mice Deficient in the Putative Phospholipid Flippase Atp11c Exhibit Altered Erythrocyte Shape, Anemia, and Reduced Erythrocyte Life Span." *Journal of Biological Chemistry* 289(28): 19531–37.
- 96) Yatime, Laure et al. 2009. "P-Type ATPases as Drug Targets: Tools for Medicine and Science." *Biochimica et Biophysica Acta - Bioenergetics* 1787(4): 207–20.
- 97) Zhou, Xiaoming, and Todd R. Graham. 2009. "Reconstitution of Phospholipid Translocase Activity with Purified Drs2p, a Type-IV P-Type ATPase from Budding Yeast." *Proceedings of the National Academy of Sciences of the United States of America* 106(39): 16586–91.
- 98) Zhou, Xiaoming, Tessy T. Sebastian, and Todd R. Graham. 2013. "Auto-Inhibition of Drs2p, a Yeast Phospholipid Flippase, by Its Carboxyl-Terminal Tail." *Journal of Biological Chemistry* 288(44): 31807–15.
- 99) Zimmerman, Michael L., and David L. Daleke. 1993. "Regulation of a Candidate Aminophospholipid-Transporting ATPase by Lipid." *Biochemistry* 32(45): 12257–63.
- 100) Zwaal, R. F.A., P. Comfurius, and E. M. Bevers. 2005. "Surface Exposure of Phosphatidylserine in Pathological Cells." *Cellular and Molecular Life Sciences* 62(9): 971–88.

Chapter 3

The prokaryotic integral membrane protein polymerase Wzy from Pseudomonas aeruginosa PAO1

The fast rise of antibiotic-resistant bacteria leads to an increasing need for novel drugs against these pathogens. The polysaccharide capsule present in these organisms allows them to overcome most of the defense mechanisms of the host and to cause invasive diseases. The capsule constitutes an anti-phagocytic barrier for the bacterium, but at the same time is also important as serotype-determining factor and as recognition target for the host immune system. Moreover, there is a correlation between thickness of the capsule and resistance to neutrophils. The fundamental role of capsular polysaccharides in cell survival and the high variability of their chemical composition suggests that proteins involved in their biosynthesis may be ideal targets for the development of new, very specific antibacterial drugs.

During the years, three different pathways for capsule biosynthesis have been identified: the Wzy-dependent pathway, the synthase-dependent pathway and the ABC transporter-dependent pathway. Wzy, an integral membrane protein present on the inner membrane of both gram-positive and gram-negative bacteria, acts as polymerase, or more specifically as glycosyltransferase, adding sugar repeated subunits to a growing O-antigen chain. The low sequence conservation of this class of proteins among bacterial species, and even strains, makes them a potential target for rational drug design.

This thesis focuses on the protein Wzy with the final aim of exploiting it for the development of new anti-bacterial drugs. To gain a deeper understanding of the mechanism of action of Wzy, I set the goal of elucidating the structural characteristics of this polymerase. Our working plan includes the overexpression of a recombinant form of Wzy, its purification and the characterization of its folding and stability. As structural knowledge of this protein would lay the ground for the future development of novel drugs specifically targeted against the

capsular polysaccharide biosynthetic pathway, crystallization experiments will be conducted to obtain crystals suitable for X-ray diffraction experiments.

INTRODUCTION

3.1 PSEUDOMONAS AERUGINOSA

Pseudomonas aeruginosa is a non-fermentative gram-negative rod-shaped bacterium able to colonize a huge variety of ecological niches. Isolated for the first time in the 1890s from injured patients with a characteristic blue-green pus in their wounds, this opportunistic pathogen was initially called *Bacillus pyocyaneus* by the French doctor Carl Gessard (Villavicencio 1998). Since the beginning of the 20th century *P. aeruginosa* has been associated with hospital-acquired infections, and it is still one of the major causes of nosocomial epidemics, particularly in patients in intensive care units or with reduced immune response as AIDS patients, cancer patients, cystic fibrosis (CF) patients or patients with burns (Sadikot et al. 2005). Intra-abdominal and urinary tract infections (Ruiz-Garbajosa and Cantón 2017), ocular infections as acute ulcerative keratitis associated to soft contact lenses use (Lyczak, Cannon, and Pier 2000; Subedi, Vijay, and Willcox 2017), bacteremia in severe burn patients and lung infections in CF patients (Stefani et al. 2017) are among the most common diseases with a significant *P. aeruginosa* involvement. The ability of this bacterium to grow on different surfaces such as implanted biomaterials, hospital surfaces and water supplies, represents an additional risk for contamination (Mulcahy, Isabella, and Lewis 2014). Acquisition of this pathogen is related to environmental contamination as well as patient-to-patient contagion.

Concerns regarding the resistance of an increasing number of strains to antimicrobial agents are rising in both the medical and the scientific communities. *P. aeruginosa* is among the species of bacteria known for their ability to survive to strong antibiotic treatments. In particular, multidrug-resistant (MDR) or even extensively-drug-resistant (XDR) strains can easily develop in hospital facilities. In the years 2011-2012, the European Centre for Disease Prevention and Control found that 9% of the health-care associated infections (HCAIs) were

caused by *P. aeruginosa*, a similar percentage of 7% was recorded in the United States in 2011 by the Center for Disease Control and Prevention (Magill et al. 2014).

In 2015, the European Antimicrobial Resistance Surveillance Network (EARS-Net) reported that, among all *P. aeruginosa* invasive isolates, 20% was simultaneously resistant to piperacillin/tazobactam, carbapenems and fluoroquinolones, and 13% was ceftazidime and aminoglycosides resistant. Higher resistance rates, with peaks of 30% of MDR *P. aeruginosa*, were recorded in Southern and Eastern European countries compared to strains prevalent in the North (Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net) 2015) .

Multidrug efflux pumps, enzymes able to inactivate different antibiotics, and a strong polysaccharide capsule render this bacterium particularly difficult to eradicate, once the host has been infected (Mulcahy, Isabella, and Lewis 2014).

- *Chronic lung infections in cystic fibrosis patients*

CF patients possess mutations in the cyclic AMP-regulated chloride ion channel, also known as Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), that cause an electrolytic imbalance which in turn leads to dehydration and production of viscous mucus in lungs. This affects the mucociliary clearance mechanism, allowing pathogens to colonize and persist in this environment (H. L. Rocchetta, Burrows, and Lam 1999). Moreover, the CFTR protein has a role as epithelial cell receptor for *P. aeruginosa* LPS promoting bacterium internalization (Pier, Grout, and Zaidi 1997).

The major cause of morbidity and mortality for patients with CF is represented by lung infections and *P. aeruginosa* is one of the most common pathogens found in CF lung infections. The survival and recovery is strictly dependent on the ability of antibiotic therapy to treat pulmonary exacerbations and to eradicate infections, including chronic cases. In recent years, the increasingly frequent appearance of MDR strains in *P. aeruginosa* isolates from patients is becoming a serious issue, as routine antibiotic treatments, including those based on fluoroquinolones, aminoglycosides and β -lactams, are ineffective in these kinds of infections. The Cystic Fibrosis Foundation reported that in 2014 18.1% of CF patients had MDR strains of *P. aeruginosa*, a proportion that is even higher for older patients due to

cumulative exposure to antibiotics that leads to survival and selection of MDR bacteria (Stefani et al. 2017).

P. aeruginosa lung infection has a complex development. The bacterium first enters by patient inhalation in the lower airways, where the host immune system can defeat it. However, if the defense mechanisms fail, the bacterium colonizes all the airways, up to the lung. The long-term persistence of this bacterium is related to many different adaptation mechanisms, among them: macrophage-phagocytosis resistance, antibiotic-resistance, development of a mucoid phenotype, formation of highly adherent small colony variants (SCVs) that are able to grow slowly and have superior adherence due to reduced motility, or production of biofilm. In particular, the biofilm allows *P. aeruginosa* to acquire a higher tolerance to inflammatory defense mechanisms and antibiotic therapies, becoming up to a thousand fold more resistant than normal bacteria (Mulcahy, Isabella, and Lewis 2014).

Lately, pharmaceutical research efforts were aimed at the development of vaccines against *P. aeruginosa*, but so far tests show no efficacy. To date, the most common therapies for CF patients with *P. aeruginosa* infections are based on tobramycin, a third generation aminoglycoside, and on the β -lactam aztreonam lysine (Stefani et al. 2017).

- *Bacteremia in severe burn patients*

In burn injuries, the loss of the skin barrier represents an obvious debilitating factor against infections. Since *P. aeruginosa* is commonly present in the environment of hospital facilities where burn patients are treated, there is a high risk of infection. In addition, MDR strains of *P. aeruginosa* are particularly frequent on hospital floors, bed rails, other patients and even the skin of hospital staff (Chitkara and Feierabend 1981). *P. aeruginosa* pili and flagella play an important role in this kind of infections, as strains deficient in these appendages lose the ability to persist in wounds (Sato, Okinaga, and Saito 1988).

- *Ulcerative keratitis associated to soft contact lenses use*

The ulcerative keratitis (UK) is a corneal inflammatory response to bacterial infections that leads to the destruction of the cornea (Galentine et al. 1984). The human body protects the eye from infections through different mechanisms, among which tear production and eye

blinking. Body fluids, including tears, are rich in protective enzymes such as secretory immunoglobulin A, amylase and lysozyme, able to kill bacteria before they start the infection (Masinick et al. 1997). However, the repeated use of contact lenses decreases the levels of protective enzymes in tears, promoting *P. aeruginosa* adherence to eye tissues (Fleiszig et al. 1994). In the eye, the LPS present on the bacterial envelope interacts with the CTFR protein in the same way as for CF patients, promoting the infection, albeit through different mechanisms (Zaidi et al. 1999). While in the lungs of CF patients the epithelium expressing CTFR is a one cell-thick layer that undergoes desquamation and allows internalization of the bacterium, in the eye CTFR is mainly expressed by the basal epithelial cells (**Figure 3.1**). These cells are covered by layers of different cell types and bacterial internalization is harder. However, once the bacterium is internalized in the basal epithelial cells, the interaction with CTFR allows its replication within the cornea, generating the infection and the subsequent tissue damage (Fleiszig, Zaidi, and Pier 1995). *P. aeruginosa* flagella are fundamental in eye adhesion and infection since they play a critical role in cell invasion and cytotoxicity (Comolli et al. 1999).

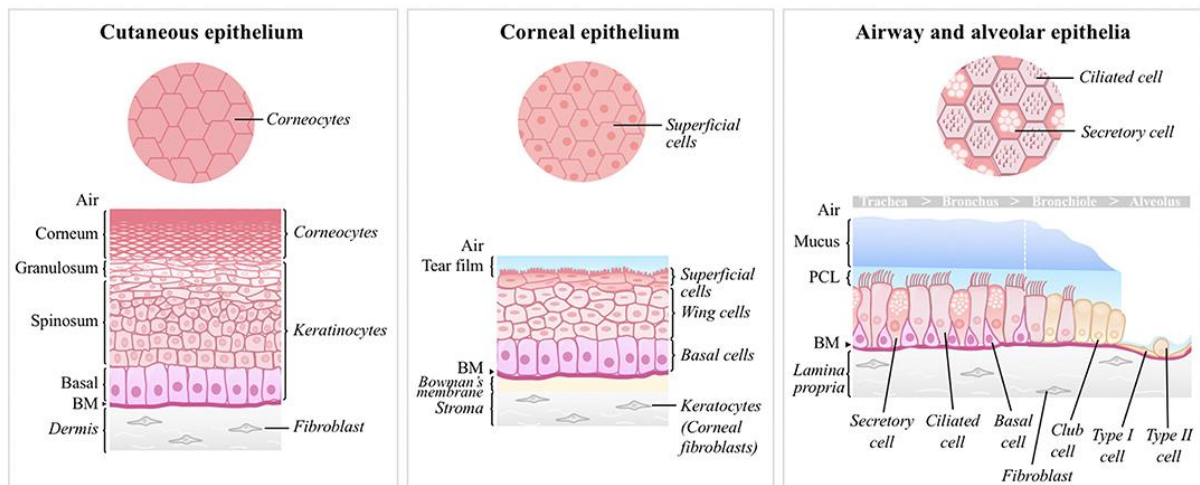


Figure 3.1: Epithelium structures in cute, cornea and airways. From M. Ruffin and E. Brochiero (2019), *Frontiers in Cellular and Infection Microbiology* (Ruffin and Brochiero 2019).

3.1.1 *Pseudomonas aeruginosa* antibiotic resistance

Antibiotic treatments are the primary therapy in cases of *P. aeruginosa* infections. However, the ability of the bacterium to mutate to an MDR strain requires the development of new therapeutic strategies and dictates some caution in the extensive use of antibiotics.

Survival mechanisms of *P. aeruginosa* develop through mutations and follow different strategies: its modest outer membrane permeability, increased by repression or inactivation of carbapenem porins and the presence of a polysaccharide barrier; the up-regulation of multi-drug efflux pumps; the production of inducible enzymes against antibiotics (Ruiz-Garbajosa and Cantón 2017); and the development of biofilms. MDR isolates often display broad resistance to pharmacological therapy due to the combined activation of more than one survival strategy.

Reduced outer membrane permeability is attributable to the absence of large diffusion porins in *P. aeruginosa* (Chevalier et al. 2017), together with the presence of a thick LPS that creates a physical barrier against antibiotic penetration. In fact, LPS molecules interact to form a shield that allows penetration of nutrients but impedes diffusion of dangerous compounds, such as antibiotics or detergents, promoting survival of these organisms even in harsh environmental conditions (Zhanga, Meredith, and Kahne 2013).

Resistance to β -lactams, among the most diffuse classes of antibiotics, develops through mutations in the regulation of β -lactamases. Isolates able to produce AmpC β -lactamase or AmpC cephalosporinase are resistant to high concentrations of β -lactam drugs. In addition, strains able to express metallo- β -lactamases are insensitive to therapies that combine β -lactam antibiotics with β -lactamase inhibitors, as the metallo- β -lactamases are often unaffected by these compounds (Stefani et al. 2017).

The acquisition of plasmids and integrons carrying genes encoding for aminoglycoside-modifying enzymes (AMEs), such as rRNA methylases, grants the bacteria resistance to aminoglycosides. This mechanism can be coupled with de-repression of efflux systems already present in the cell: in these cases, efflux pumps (MexXY) are overexpressed leading to elimination of aminoglycosides through the bacterial membrane (Stefani et al. 2017).

Similarly, fluoroquinolones resistance, particularly in the case of ciprofloxacin, is the result of mutations on expression regulators of efflux pumps, leading to multidrug efflux pump overexpression (Stefani et al. 2017).

Polymyxins such as colistin are another class of antibiotics commonly used to treat gram-negative bacterial infections. This class of compounds are effective through an interaction with the LPS present on the surface of different bacteria. The exposure of *P. aeruginosa* to polymyxins can trigger the activation of a complex mechanism leading to the overexpression

of LPS-modifying enzymes. Once the LPS on the cell surface changes its overall structure, this class of antibiotics becomes ineffective having lost the primary target (Stefani et al. 2017).

P. aeruginosa, like other bacteria, is able to form an extracellular matrix mainly composed by polysaccharides known as biofilm. A reduced susceptibility of *P. aeruginosa* to antibiotic treatment results from biofilm formation and is linked to the limited penetration of these molecules through the complex biofilm matrix, further hampered by adsorption in the polysaccharide or deactivation of the compounds. In addition, the biofilm represents a protection barrier against the host immune system (Mulcahy, Isabella, and Lewis 2014).

3.2 THE BACTERIAL LIPOPOLYSACCHARIDE LPS IS FORMED BY THREE DIFFERENT UNITS: LIPID A, CORE AND O-ANTIGEN

The LPS is the main component of the outer membrane of gram-negative bacteria (Rhee 2014). The presence of this polymer is pivotal for bacterial pathogenicity since it is involved in different mechanisms that allow bacterial survival, among which the formation of a permeability barrier against antibiotics (Zhanga, Meredith, and Kahne 2013), of a physical barrier against host defenses and the direct interaction with host receptors (J. D. King et al. 2009).

When the bacterium enters into contact with the host tissue, its LPS stimulates the host immune response. First, LPS is recognized by the LPS-binding protein, that flags it for the subsequent binding of the CD14 receptor of macrophages. These events cause the secretion of cytokines, primary markers of inflammatory response. Excessive LPS stimulation of the immune system, as in systemic infections, can result in septic shock and death of the host. *P. aeruginosa* is one of the gram-negative bacteria with the highest occurrence in sepsis, since its LPS is particularly effective in overstimulation of the immune system (H. L. Rocchetta, Burrows, and Lam 1999).

LPS is constituted by three different components: the lipid A, known also as endotoxin, the core oligosaccharide and the O-polysaccharide (Woodward et al. 2010) (**Figure 3.2**).

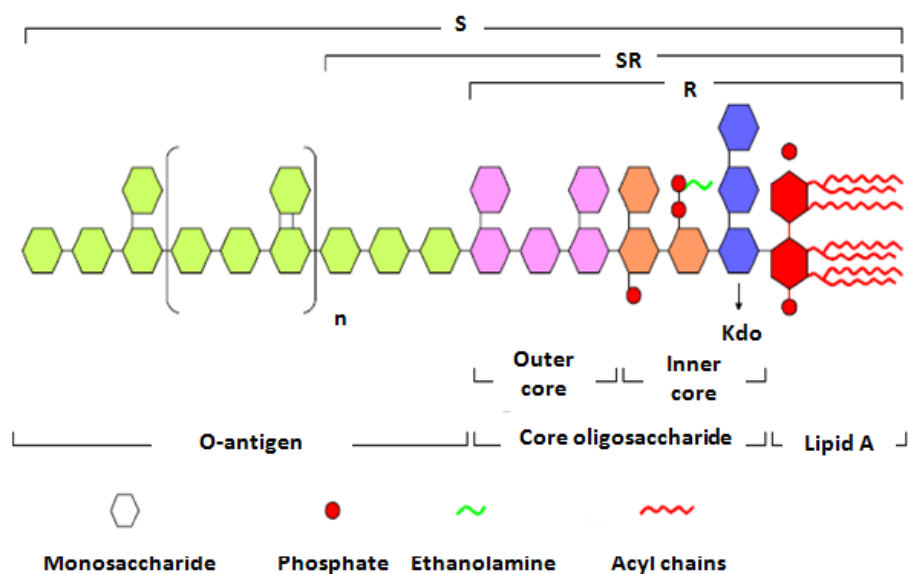


Figure 3.2: Structure of the LPS. Red: lipid A with its fatty acyl chains; Blue and Orange: inner core; Purple: outer core; Green: O-antigen with a variable number of O units; S: smooth LPS; SR: semi-rough LPS; R: rough LPS. From E. Pupo and E. Hardy (2009), *Biotecnologia Aplicada* (Pupo and Hardy 2009).

While the lipid A and the core oligosaccharides are conserved in different bacterial strains, the high variability of the O antigen is responsible for the huge number of different serotypes even within the same species (Merino, Gonzalez, and Tomás 2016). The structure of the O antigen varies in monosaccharide composition, in their arrangement and in their linkage (Zhao et al. 2015). In addition, some *P. aeruginosa* strains synthesize different kinds of LPS (**Figure 3.2**): the smooth LPS presents a complete O antigen; the semi-rough LPS possesses only one short oligosaccharide sequence, the O unit, whose polymerization forms the polymeric O antigen in the smooth type; the rough LPS lacks completely the O antigen and is only formed by lipid A and core oligosaccharides (J. D. King et al. 2009).

3.2.1 Lipid A

Lipid A is a hydrophobic glycolipid necessary for the anchorage of LPS into the external lipid layer of the bacterial outer membrane. It is formed by a disaccharide backbone, composed of two molecules of glucosamine, phosphorylated on opposite sides (carbon 1 of the first sugar and carbon 4 of the second), and holding five or six acyl chains bound to the nitrogen and oxygen atoms of the disaccharide (**Figure 3.3**) (J. D. King et al. 2009; Kulshin et al. 1991). The length of the acyl chains differs between bacterial species: for example, *E. coli* has only

saturated 14-carbon fatty acids, with the exception for one 12-carbon chain, while *P. aeruginosa* has saturated 10-carbon fatty acids bound to the oxygen atoms and 12-carbon chains bound to the nitrogen atoms (**Figure 3.3**). This difference is determined by the capacity of two enzymes, an O-acyl transferase and a N-acyl transferase, to discriminate between fatty chains of different lengths (Wyckoff et al. 1998).

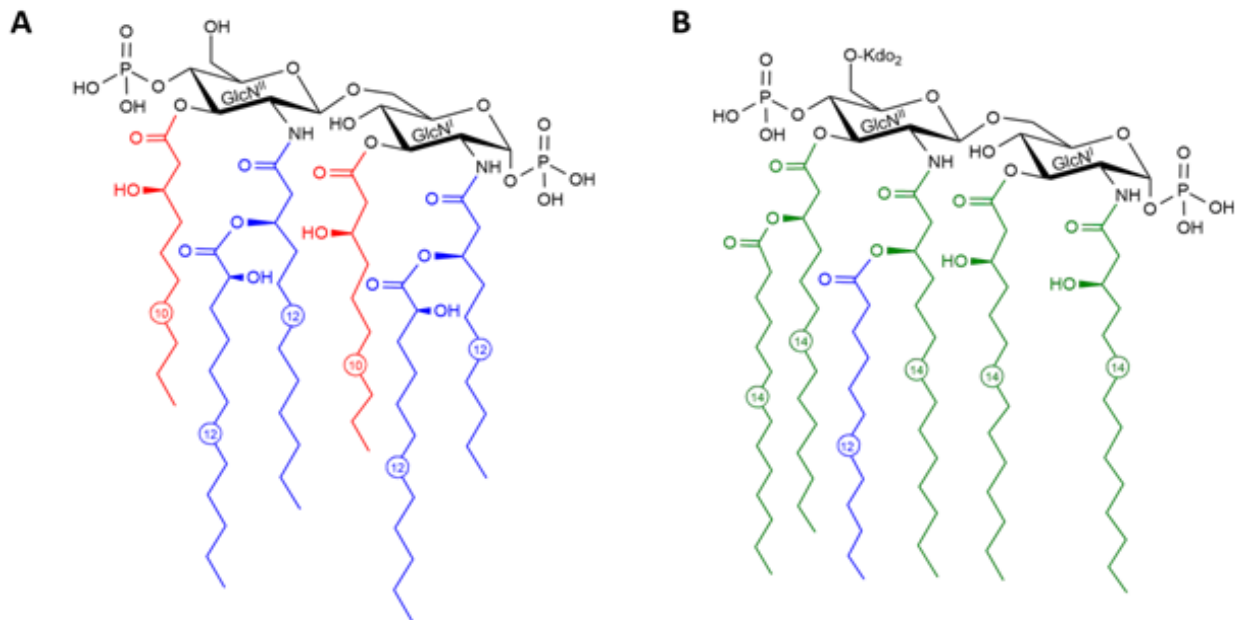


Figure 3.3: Lipid A from *P. aeruginosa* (A) and *E. coli* (B) laboratory strains. The structure of lipid A from *P. aeruginosa* and *E. coli* is composed of two phosphorylated glucosamine residues, differing for acyl chain length (10 or 12 for *P. aeruginosa* and 12 or 14 for *E. coli*) and linkages. Adapted from King J. D. et al. (2009), *Innate Immunity* (J. D. King et al. 2009).

In *P. aeruginosa*, the lipid A structure can be modified in response to environmental stimuli, in order to modulate bacterium-host interactions or resist to antimicrobial and bactericidal agents (Ernst et al. 1999). For example, lipid A from *P. aeruginosa* isolates of CF patients possesses six or seven acyl chains bound to the diglucosamine, and a 4-amino-4-deoxy-L-arabinose bound to one or both the phosphate groups (Bhat et al. 1990). Another classical modification in lipid A structure is the deacetylation induced by low magnesium concentration in the environment and low temperature (Trent et al. 2001; Ernst et al. 2006). In addition, in low magnesium conditions, a secondary saturated 16-carbon fatty acid chain may be transferred on the sugar backbone (Ernst et al. 1999).

For most bacterial species, lipid A is the LPS domain mainly responsible for the activation of inflammatory response-induced endotoxicity (Galanos et al. 1985), gaining its endotoxin

name. The fatty chains of lipid A interact with the Toll-like receptor 4 (TLR4) (Alexander and Rietschel 2001) and its activation triggers the production of many inflammatory mediators such as the tumor necrosis factor (TNF- α) or the interleukin 1- β (IL1- β) (Beutler and Cerami 1988; Dinarello 1991). Since this interaction depends on the length of fatty acid chains, in bacteria with longer carbon chains, like *E. coli*, the lipid A causes a huge inflammatory response, while in *P. aeruginosa*, where the chains are shorter, this response is weaker (Bäckhed et al. 2003). However, an exceptionally severe inflammatory response to *P. aeruginosa* infection is due to the hyperacylated forms of lipid A present in clinical isolates from CF patients (Ernst et al. 2003).

3.2.2 Core oligosaccharide

In LPS, lipid A is connected to the O antigen through an oligosaccharide, known as core oligosaccharide, composed of nine or ten sugars, usually heptoses and keto-deoxyoctulosonic acid derivatives. The oligosaccharide composition depends on bacterial species and strains.

In the *P. aeruginosa* strain PAO1, one of the most common laboratory strains, the core oligosaccharide has two distinct components, the inner and the outer core. The inner core oligosaccharide is composed of two residues of 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) and two residues of L-glycero-manno-heptose (Hep), sites of further phosphorylation (J. D. King et al. 2009) (**Figure 3.4**). Hyperphosphorylation of LPS seems to be related to *P. aeruginosa* antibiotic resistance (Walsh et al. 2000).

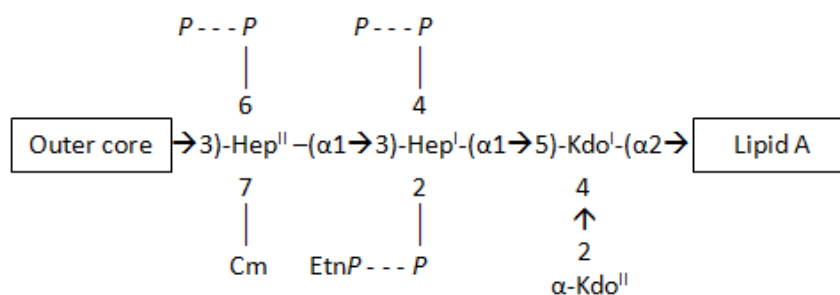


Figure 3.4: Structure of the inner core of *P. aeruginosa*.

The outer core oligosaccharide of *P. aeruginosa* is composed by three different sugars: D-glucose (Glc), L-rhamnose (L-Rha) and 2-amino-2-deoxy-D-galactose (GalN), the last further modified by an alanine residue (J. D. King et al. 2009). Two different outer core glycoforms are composed of a similar GalN-Gcl-Gcl sequence, but differ in the position of L-Rha residue (**Figure 3.5**): L-Rha is bound to the first Gcl residue in glycoform 1, characteristic of rough LPS, and to the second Glc residue in glycoform 2 (Sadovskaya et al. 1998). These two glycoforms are present on the bacterial envelope in comparable amount. In addition, a less common glycoform 1b presents an additional glucose residue (Bystrova et al. 2006). In the smooth LPS-isoform, the O-antigen is bound to the oxygen atom in position 3 of the L-Rha of glycoform 2 (Knirel et al. 2006).

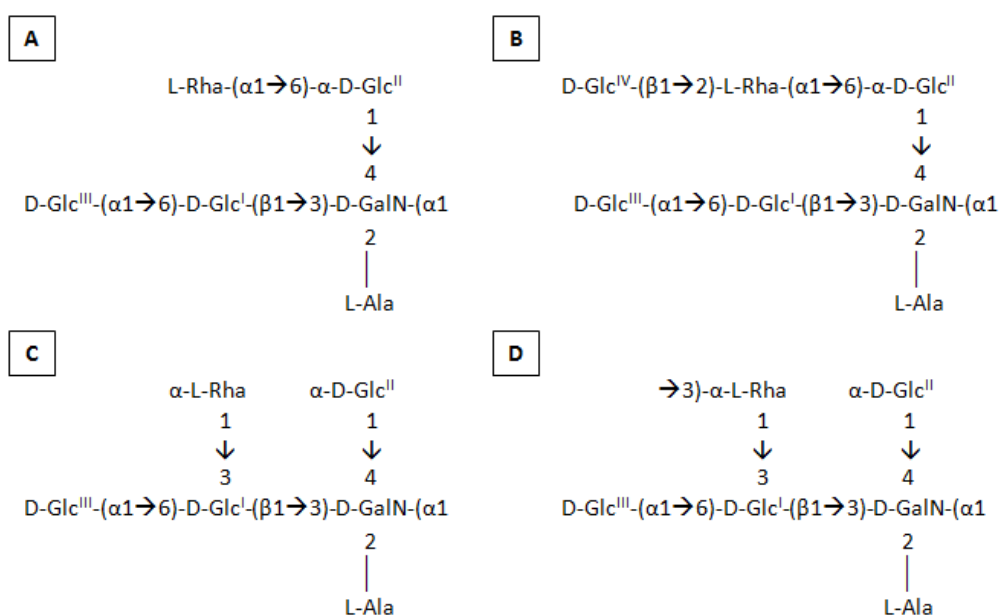


Figure 3.5: Structures of outer core forms in *P. aeruginosa*: **A)** glycoform 1a, **B)** glycoform 1b, and glycoform 2 **C)** in LPS rough strains, and **D)** in LPS smooth strains, with an arrow showing the binding site for the O antigen.

The core oligosaccharide is the binding site of the CTFR protein receptor, that mediates epithelial-cell internalization in CF and has a key role in keratitis associated with contact lenses use (Fleiszig, Zaidi, and Pier 1995)

Once the lipid A-core complex is formed on the cytosolic side of the inner bacterial membrane, the protein MsbA, an essential ABC transporter, is responsible for its translocation to the periplasmic side, where this unit is linked to the O antigen and is further transferred to the external side of the outer membrane (Doerrler et al. 2004). The intrinsic ATPase activity of MsbA is stimulated by the presence of the lipid A-core complex (Ghanei, Abeyrathne, and Lam 2007). In addition, the 4.5 Å resolution crystal structure of *E. coli* MsbA

(Chang and Roth 2001) seems to confirm MsbA function as floppase/channel for lipid A transport.

3.2.3 O antigen

The last component of the LPS is the O polysaccharide, known also as O antigen, constituted of repeated units of homo- or hetero-polysaccharides (Merino, Gonzalez, and Tomás 2016) and covalently bound to the core oligosaccharide (J. D. King et al. 2009).

P. aeruginosa strains produce two O-antigen types differing for their structure and named A-band O antigen, or common polysaccharide antigen, and B-band O antigen, or O-specific antigen (Stanislavsky and Lam 1997). A-band and B-band O antigens are synthesized via different mechanisms, sharing only the initial glycosyltransferase, the protein WbpL. B-band O-antigen biosynthesis occurs via the Wzy-dependent pathway, while the A-band synthesis follows the ABC transporter-dependent pathway (H. L. Rocchetta, Burrows, and Lam 1999).

The A-band O antigen is a homopolymer of approximately 70 residues of D-rhamnose and is able to induce only a weak antibody response (Arsenault et al. 1991; Yokota et al. 1987). Despite the fact that the monomers forming the A-band polymer are the same, their bonds are different and the whole common antigen can be viewed as a repetition of a trisaccharide sequence (**Figure 3.6**) (Arsenault et al. 1991). Its biosynthesis initiates in the inner membrane with the addition of a sugar residue with a phosphorylation in position 1, to an undecaprenyl phosphate, catalyzed by the WbpL protein. The action of 3 different glycosyltransferase, WbpX, WbpY and WbpZ, allows the formation of the repeated units of the common O polysaccharide (Heather L. Rocchetta et al. 1998). Once formed, the common O polysaccharide is translocated across the inner membrane by the ABC O-polysaccharide exporter to the periplasmic side of the membrane.

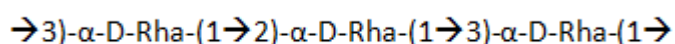


Figure 3.6: Scheme of a common polysaccharide antigen repeated unit.

The B-band O antigen is a highly immunogenic heteropolymer, with a very specific composition for each *P. aeruginosa* strain (J. D. King et al. 2009). In the O-specific antigen,

units of di- to penta-oligosaccharides form long chains of more than 50 repetitions (J. S. Lam et al. 1992).

The variability of the O-specific antigen is remarkable and gives rise to 20 serotypes in *P. aeruginosa* alone, each of them with a different sequence of sugars in its O unit (**Table 3.1**).

The International Antigenic Typing System (IATS) classifies each serotype according to its serological reactivity against specific antibodies (Liu and Wang 1990).

| Serotype (IATS) | O Antigen repeated unit |
|-----------------|--|
| O1 | →4)-D-GalNAc-(1→4)-β-D-GlcNAc3NAcA-D-(1→3)-D-FucNAc-(1→3)-D-QuiNAc-(1→ |
| O2 | →4)-β-D-ManNAc3NAmA-(1→4)-L-GulNAc3NAcA-(1→3)-β-D-FucNAc-(1→ |
| O3 | →2)-L-Rha3OAc-(1→6)-D-GlcNAc-(1→4)-L-GalNAc4OAcA-(1→3)-β-D-QuiNAc4NSHb-(1→ |
| O4 | →2)-L-Rha-(1→3)-L-FucNAc-(1→3)-L-FucNAc-(1→3)-D-QuiNAc-(1→ |
| O5 | →4)-β-D-ManNAc3NAmA-(1→4)-β-D-ManNAc3NAcA-(1→3)-β-D-FucNAc-(1→ |
| O6 | →3)-L-Rha-(1→4)-D-GalNAc3OAcAN-(1→4)-D-GalNfoA-(1→3)-D-QuiNAc-(1→ |
| O7 | →4)Pse4OAc5NRHb7Nfo-(2→4)-β-D-Xyl-(1→3)-D-FucNAc-(1→ |
| O8 | →4)Pse4OAc5Nac7Nfo-(2→4)-β-D-Xyl-(1→3)-D-FucNAc-(1→ |
| O9 | →3)Pse4OAc5Nac7NRHb-(2→4)-D-FucNAc-(1→3)-β-D-QuiNAc-(1→* |
| O10 | →3)L-Rha2OAc-(1→4)-L-GalNAcA-(1→3)-D-QuiNAc-(1→ |
| O11 | →2)-β-D-Glc-(1→3)-L-FucNAc-(1→3)-D-FucNAc-(1→ |
| O12 | →8)-8eLeg5Nac7Nac-(2→3)-L-FucNAc-(1→3)-D-QuiNAc-(1→ |
| O13 | →2)-L-Rha-(1→3)-L-Rha-(1→4)-D-GalNAc3OAcA-(1→3)-β-D-QuiNAc-(1→ |
| O14 | →2)-[D-Glc-(1→3)]-L-Rha-(1→3)-L-Rha-(1→4)-D-GalNAc3A-(1→3)-β-D-QuiNAc-(1→ |
| O15 | →2)-β-D-Ribf-(1→4)-D-GalNAc-(1→ |
| O16 | →4)-β-D-ManNAc3NAmA-(1→4)-β-D-ManNAc3NAcA-(1→3)-β-D-FucNAc-(1→ |
| O17 | →4)-β-L-Rha-(1→3)-D-ManNAc-(1→ |
| O18 | →4)-L-GulNAc3NAmA-(1→4)-β-D-ManNAc3NAcA-(1→3)-β-D-FucNAc-(1→ |
| O19 | →3)L-Rha-(1→4)-L-GalNAcA-(1→3)-D-QuiNAc-(1→ |
| O20 | →4)-L-GulNAc3NAmA-(1→4)-β-D-ManNAc3NAcA-(1→3)-β-D-FucNAc4OAc-(1→ |

Table 3.1: *P. aeruginosa* O-antigen serotypes. Ac: acetyl; 8eLeg: 5,7-diamino-3,5,7,9-tetradecyloxy-L-glycero-D-galacto-non-2-ulonic(8-epilegionaminic) acid; Fo: formyl; FucN: 2-amino-2,6-dideoxy-galactose; GalN: 2-amino-2-deoxy-galactose; GlcNA: 2-amino-2-deoxy-glucuronic acid; GulNA: 2-amino-2-deoxy-guluronic acid; ManN: 2-amino-2-deoxy-mannose; ManNA: 2-amino-2-deoxy-mannuronic acid; N: amino; NAc: acetamido; NAm: acetamidino; OAc: O-acetyl; Pse: 5,7-diamino-3,5,7,9-tetradecyloxy-L-glycero-L-manno-non-2-ulonic (pseudoaminic) acid; QuiN: 2-amino-2,6,dideoxy-glucose; Rha: rhamnose; RHb. (R)-3-hydroxybutanoyl; Rib: ribose; SHb: (S)-3-hydroxybutanoyl. Anomeric conformations are α, unless otherwise marked. Sugars are in the pyranose form, except for the ribose residue of serotype O15.

Among the 20 O serotypes of *P. aeruginosa* only 14 are able to produce A-band LPS (J. D. King et al. 2009). The most common strain is the serotype O5, that includes the laboratory strain PAO1, and is able to produce both A- and B- band LPS. PAO1 specific O antigen is formed by repeated trisaccharide units of two N-acetylmanosaminuronic acid and a N-acetyl-6-deoxygalactose (H. L. Rocchetta, Burrows, and Lam 1999).

Since O-specific antigens differ among strains, each of them possesses specific proteins for its biosynthesis, with the exception of the protein WbpM, a nucleotide sugar epimerase/dehydratase conserved in all serotypes. A series of serotype-specific glycosyl transferase are responsible for the formation of the repeated O-specific units. Once synthesized, the O-repeated units are:

- (1) translocated to the periplasmic layer of the inner membrane by the protein Wzx;
- (2) polymerized to the correct length by proteins Wzy and Wzz, that form the Wzy-dependent pathway, together with Wzx;
- (3) attached to the lipid A-core complex by the ligase Waal, for the formation of mature LPS;
- (4) and, finally, exported to the surface of the outer membrane by proteins of the Lpt-pathway (J. D. King et al. 2009).

Different parameters influence the production of the LPS and its length. While cellular mechanisms of response are highly complex, the LPS structure produced in different environmental conditions can be easily defined. High temperature decreases B-band O-antigen length, as do hypertonic or low pH conditions. In parallel, an elevated temperature and variation in nutrients, such as high MgCl₂ and NaCl levels or low phosphate media, increase significantly A-band O-antigen length (Kropinski, Lewis, and Berry 1987; McGroarty and Rivera 1990). Under particular conditions, *P. aeruginosa* strains may lose the ability to produce O antigen, developing in rough or semi-rough strains (J. D. King et al. 2009).

The LPS O antigen, forming the surface of the outer membrane, is involved in all bacterium-environment interactions. In fact, the main role of this component is to protect the bacterium from phagocytosis and complement-mediated killing (Dasgupta et al. 1994), but it intervenes also in protection against the host response to oxidative stress (Berry et al. 2009). In chronic infections, such as lung infections of CF patients, a selective pressure leads to the development of strains that lack the O-specific antigen and are able to evade the huge host

immunogenic response activated by antibody recognition of the LPS (M. Y. C. Lam et al. 1989; Hancock et al. 1983). Moreover, repeated exposure to antibiotics reduces O-antigen production, since some antibiotics such as gentamicin bind to the cellular surface (Kadurugamuwa, Lam, and Beveridge 1993).

3.2.4 *O-antigen synthetic pathways*

Analyzing in more detail the synthetic pathways responsible for the formation of the O antigen, it is interesting to notice that the energy required for lipid A-core/O-antigen ligation is stored in the UDP-O-antigen intermediate both for the Wzy-dependent pathway and for the ABC transporter-dependent pathway (Raetz and Whitfield 2002) (**Figure 3.7**). Another mechanism for O-antigen biosynthesis is present in some bacterial species, the synthase-dependent pathway (Woodward et al. 2010).

In *P. aeruginosa*, for both the Wzy-dependent and the ABC transporter-dependent mechanisms, the first step is the formation of the bond between the 1-phosphate derivative of the first sugar residue of the O unit and a molecule of undecaprenyl phosphate (UDP), catalyzed by the protein WbpL. However, after the formation of this precursor, the two mechanisms follow different routes.

a) *The Wzy-dependent pathway*

Starting from the undecaprenyl phosphate precursor, a series of specific transferases catalyze the ordered and specific subsequent addition of sugar monomers forming the O-antigen repeated unit. The UDP-O Unit, positioned in the cytoplasmic leaflet of the inner membrane, is transferred to the periplasm where it undergoes polymerization, leading to the production of the complete O antigen. The O-specific antigen follows the Wzy-dependent pathway for its translocation and polymerization (J. D. King et al. 2009). The integral membrane protein Wzx is responsible for the translocation of the UDP-O Unit from the cytoplasmic to the periplasmic layer of the inner membrane while functioning also as cation antiport (Islam, Eckford, et al. 2013). In the periplasmic leaflet of the inner membrane, the UDP-O Unit is the substrate of another integral membrane protein, Wzy, composed by 10-14 transmembrane helices (Woodward et al. 2010). Wzy starts the

polymerization of the polysaccharidic O-antigen chain transferring a nascent UDP-O Polymer to the non-reducing end of a UDP-O Unit, creating α -glycosidic bonds or β -glycosidic bonds (Bray and Robbins 1967; V. L. Taylor et al. 2016). The integral membrane protein Wzz acts as a ruler regulating the O-antigen chain length (Batchelor et al. 1991). During the years different models for the mechanism of polymer length control of Wzz have been proposed, but the most likely are the Ruler model from the Goldberg group (Kintz and Goldberg 2011), and the Chain Feedback model from Kalynych (Kalynych, Valvano, and Cygler 2012). In the first model there is a direct interaction between the surface of Wzz and the polysaccharide: in this mechanism the length of the O antigen is connected to the diameter of the specific Wzz protein. In the second model, the growing O antigen generates higher-order structures that could interfere with the active site of Wzy during polymerization. The presence of Wzz determines the length of the O antigen chain by binding with the polymeric chain and preventing the formation of ordered structures. With no inhibition on the Wzy polymerase, the elongation of the O-antigen chain continues, until the glycosidic polymer is too long to bind to Wzz. Probably the real mechanism is a hybrid, with Wzy and Wzz interacting, while the polymerization proceeds, and the interaction of the growing sugar chain with Wzz contributing to maintain an orientation that allows the addition of other O-antigen repeated units. While the chain grows, it creates more rigid higher-order structures that destabilize the interaction with Wzz resulting in the polymer disengaging from Wzy (Islam and Lam 2014).

Once the polymerization is complete, the protein WaaL catalyzes the binding of the O antigen to the lipid A-core complex, forming the LPS molecule. Finally, the LPS is transported to the exoplasmic side of the outer membrane via the Lpt pathway (Islam and Lam 2014; Merino, Gonzalez, and Tomás 2016). Evidence of the formation of stable complexes between Wzy and Wzz has been obtained in *S. Flexneri* with cross-linking studies (Nath and Morona 2015).

While to date no experimental Wzy structure is available, the structure of Wzz has been determined in different conditions. It is known that Wzz forms multimers, but their dimension spans from trimers to dodecamers, or even larger complexes, in different conditions suggesting a correlation between oligomer size and O-antigen length. Other studies propose that surface exposure of specific residues in peculiar Wzz regions may be the determining factor for O-antigen length regulation, instead of the Wzz multimer

dimension (Collins et al. 2017), thus leaving the debate open on the real mechanism of O-antigen length control by Wzz.

b) *The ABC transporter-dependent pathway*

In *P. aeruginosa*, the synthesis of the common O antigen follows the ABC transporter-dependent pathway (H. L. Rocchetta, Burrows, and Lam 1999). In this biosynthetic route, the extension of the A-band O-antigen chain is catalyzed by the progressive addition of sugar residues to the reducing terminus of the nascent UDP-polymer on the cytosolic part of the inner membrane. No specific polymerases are required and, once the polymer reaches a defined length, it is exported to the periplasmic side of the inner membrane for ligation with the lipid A-core complex. This translocation is performed by an ABC transporter driven by ATP hydrolysis. The ABC transporter is formed by two proteins, Wzm and Wzt (Leslie Cuthbertson, Powers, and Whitfield 2005). The Wzm/Wzt interaction and the conformational change induced by ATP binding to Wzt drive the export of the common polysaccharide chain to the periplasmic side of the inner membrane, where it undergoes a further ligation to lipid A-core complex (Raetz and Whitfield 2002).

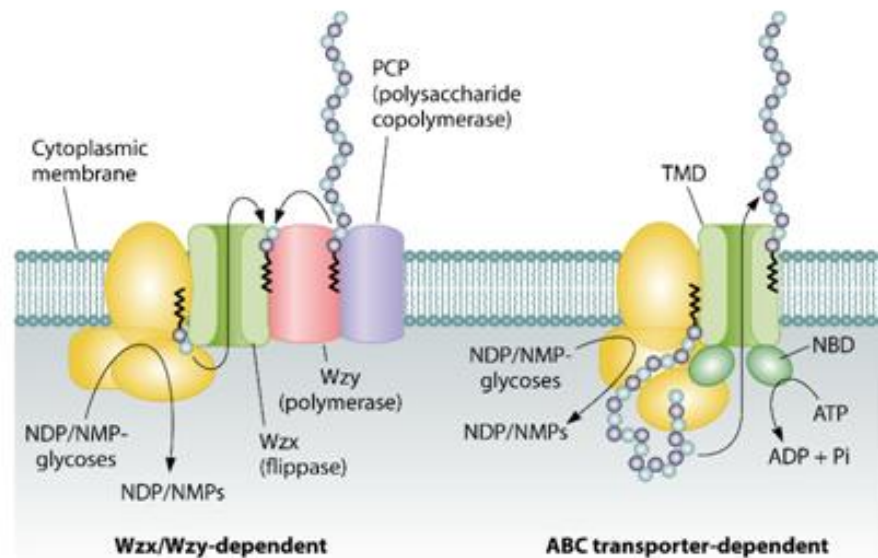


Figure 3.7: Different pathways for O-antigen synthesis. **A)** Wzy-dependent pathway; **B)** ABC transporter-dependent pathway. From Cuthbertson L. et al. (2010), *Microbiology and Molecular Biology Reviews* (L. Cuthbertson, Kos, and Whitfield 2010).

3.2.5 *The ligase WaaL*

When both O antigen and lipid A-core complex are present on the periplasmic side of the inner membrane, the O-antigen ligase WaaL catalyzes the assembly of these two parts in order to form the complete LPS (Abeyrathne et al. 2005). The protein WaaL is a membrane enzyme that catalyzes the formation of a glycosidic bond between a sugar on the proximal end of the UDP-O Antigen and the terminal sugar of the lipid A-core complex (Ruan et al. 2012).

3.2.6 *The LPS Export Pathway Lpt*

Once fully formed on the periplasmic side of the inner membrane, the Lipopolysaccharide Export Pathway (Lpt), a route common to many bacterial species, transports the complete LPS to the outer membrane, to be exposed on the cell surface (Polissi and Sperandio 2014) (**Figure 3.8**). Lpt pathway is composed by seven proteins, named with subsequent letters from LptA to LptG and localized throughout the periplasmic space. The 7 polypeptide chains form a single large complex, working together as a single machinery (Ruiz, Kahne, and Silhavy 2009). This large complex can be divided in three subassemblies differing for their position within the periplasm: LptBFGC located in the IM, LptA located in the periplasm and LptDE located on the OM (Sperandio et al. 2007).

The first subassembly, LptBFGC, works as an ABC transporter, provided with an atypical bitopic subunit (LptC), an ATP-binding domain (LptB), and two TM domains (LptF and LptG) (Narita and Tokuda 2009). The function of LptA is to connect the inner membrane and the outer membrane subassemblies (Sperandio et al. 2007) and its structure was determined through X-ray crystallography at 2.75 Å resolution (Bollati et al. 2015). On the outer membrane the LptDE translocon, composed of the β -barrel protein LptD and the lipoprotein LptE, is responsible for the final translocation of the LPS on the external leaflet of the outer membrane (Chng et al. 2010).

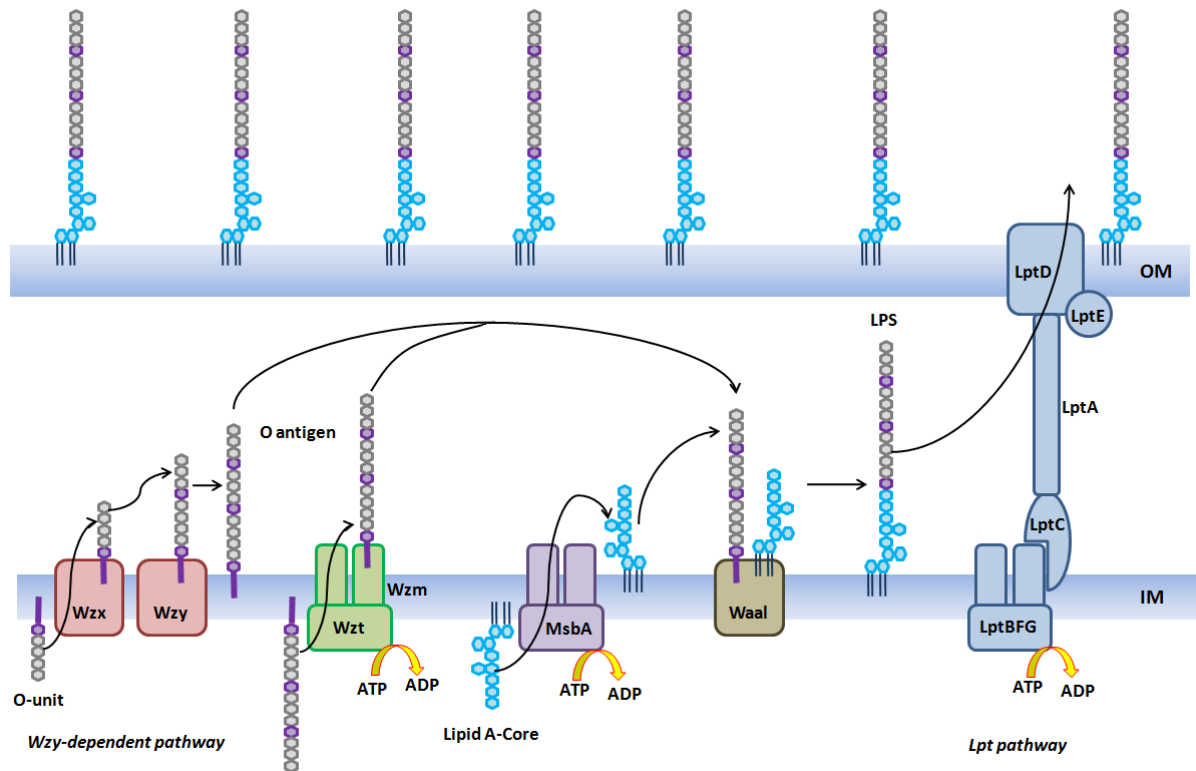


Figure 3.8: Synthetic pathways involved in LPS formation. Wzy- and ABC transporter-dependent pathways are responsible for formation of the O antigen on the periplasmic side of the inner membrane; the protein MsbA translocates the complex Lipid A-Core oligosaccharide; the protein WaaL assembles O antigen and Lipid A-Core oligosaccharide for LPS formation; the Lpt pathway translocates the LPS to the outer side of the outer membrane.

3.3 THE INTEGRAL MEMBRANE PROTEIN WZY

The integral membrane protein Wzy is localized in the IM of bacteria, where it has a fundamental role in the Wzy-dependent pathway. As expected considering the high diversity observed in bacterial O-specific antigens, Wzy has low sequence conservation among species and even bacterial strains. The catalytic role of Wzy involves the formation of new glycosidic bonds between O-repeated units to form a longer polysaccharide chain. For this reason, this protein is either referred to as a polymerase or a glycosyltransferase (GT) enzyme. Despite the significant differences in their sequence, the class of Wzy proteins is divided in α -polymerases (Wzy_{α}) or β -polymerases (Wzy_{β}) depending on the stereochemistry of the glycosidic bond that they catalyze (V. L. Taylor et al. 2016). Studies demonstrate that Wzy proteins are able to discriminate between identical O units that differ only in anomeric bonds, indicating a high substrate selectivity (Hong et al. 2015). Furthermore, Wzy activity is

strongly dependent on the first sugar of the repeated glycosidic unit, reinforcing its ability to discriminate different O units (Merino, Gonzalez, and Tomás 2016). Some evidence suggests that Wzy works as a dimer, but no conclusive proof of its functional assembly is available, nor is its 3-dimensional structure (Islam and Lam 2014).

3.3.1 Structure of Wzy polymerase in *P. aeruginosa* PAO1

Most of the available studies regarding this protein class have been conducted on the common laboratory strain of *P. aeruginosa* PAO1 (serotype O5). The Wzy protein expressed by this strain (Wzy_{pa}) is a 438 residue protein classified in the subclass of Wzy_α proteins (V. L. Taylor et al. 2016). According to sequence analysis, it is formed by 14 TM helices, two large periplasmic loops of comparable size and two smaller cytoplasmic loops (**Figure 3.9**). Despite the good sequence homology shown by the two periplasmic loops (PL3 and PL5), they have opposite ionic charge at physiological pH (Islam et al. 2011). The presence of similar RX₁₀G motifs ensures the binding of the same sugar substrate but with different affinity (Islam et al. 2010)

The loop PL3 is located between the transmembrane segments 5 and 6. Since its isoelectric point is 8.59, the PL3 loop has a positive charge at physiological pH. Arginine residues 175, 176 and 180 (R175, R176 and R180), belonging to the RX₁₀G motif, play a crucial role in the polymerization of the O antigen. In fact, substitution of arginine residues with alanine significantly reduces the O-antigen production. The functional importance of these residues, however, does not depend only on their charge, as substitution of R176 with lysine (R176K) results in the suppression of polymerization activity, indicating a crucial role for the guanidinium group of this specific residue. Analogous mutations on the other two arginine residues, R175K and R180K, do not affect the protein activity (Islam et al. 2011).

The loop PL5 is located between transmembrane segments 9 and 10. Its 5.49 isoelectric point results in a completely opposite charge compared to PL3, as at physiological pH PL5 is negatively charged. As for PL3, arginine residues 290 and 291 (R290 and R291), located in the first positions of the RX₁₀G motif, are fundamental for O-antigen formation. For PL5, however, single mutations on both residues with lysine (R290K and R291K) significantly reduce protein activity, emphasizing the role of the guanidinium group (Islam et al. 2011).

The cytoplasmic loops CL2 and CL6 connect transmembrane segments 4 and 5, and 12 and 13, respectively. Mutations in CL6 affects the O-antigen phenotype, particularly as regards the length of the polysaccharide chain, suggesting a key role in length control regulation. Considering the role of the protein Wzz in controlling the dimension of the O antigen, it has been proposed that CL6 is involved in the Wzy-Wzz interaction (Islam, Huszczyński, et al. 2013). In *P. aeruginosa*, Wzz is present in two isoforms, Wzz₁ and Wzz₂, involved in the production of long and very-long O antigens, respectively (Daniels et al. 2002). Mutations in CL6 lead to a reduced formation of long O antigen, while in some cases increasing the amount of very-long O antigen (Islam, Huszczyński, et al. 2013). This evidence suggests that the correct sequence and folding of CL6 is required for Wzy_{Pa} and Wzz₁ binding, but not for the interaction with Wzz₂.

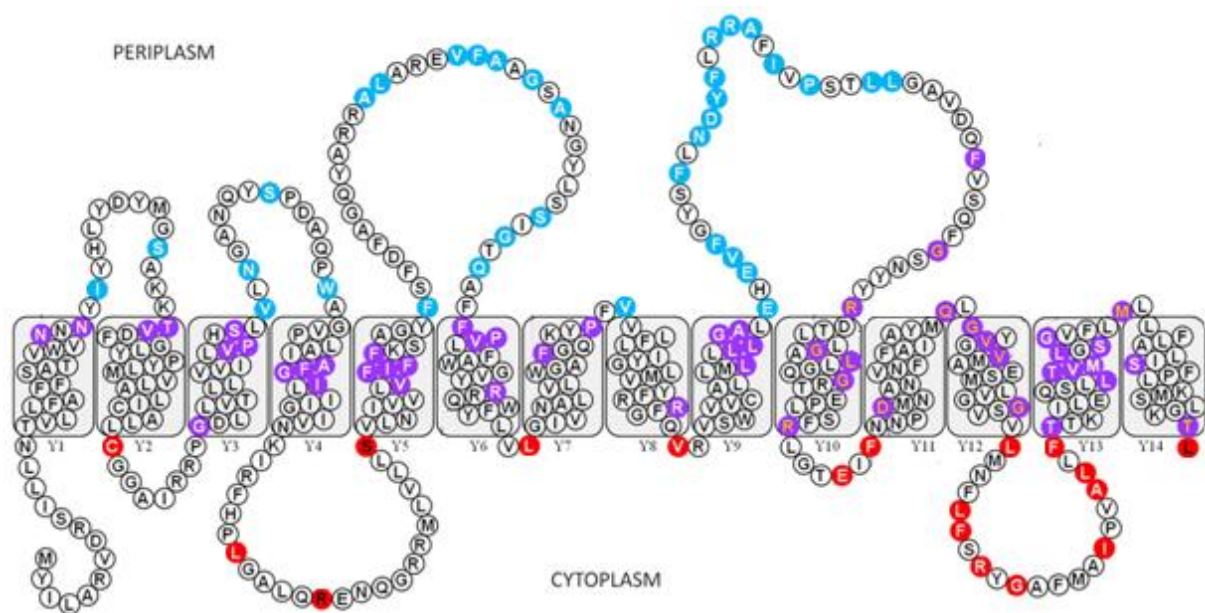


Figure 3.9: Topological map of the protein Wzy from *P. aeruginosa* strain PAO1. From Islam S. T. et al. (2010), *mBio* (Islam et al. 2010).

3.3.2 Wzy mechanism for specific O-antigen polymerization

To perform its function, the glycosyltransferase Wzy_{Pa} needs to recruit a single UDP-linked O unit, while at the same time retaining the growing polymer chain. The mechanism proposed

for Wzy activity, known as “catch-and-release” mechanism, involves the positively charged PL3 and the negatively charged PL5 loops with opposite functions (Sachdeva et al. 2017). The role of PL3 in this mechanism is to “catch” the negatively charged sugars of the single repeated O-antigen unit, and recruit it for further polymerization. PL5 plays the role of a “retention arm”, with its negative charge involved in a relatively transient interaction with O-antigen subunits, due to unfavorable electrostatic interactions. Its lower affinity for the sugar units allows for a fast exchange, suggesting that this is the catalytic site of the protein where the glycosidic bond is formed between the non-reducing end of the repeated unit and the reducing end of the UDP-linked growing polymer (**Figure 3.10**). In this mechanism, despite the lower affinity for sugars, PL5 is able to retain the growing polymer chain. After the formation of the glycosidic bond, a new O-antigen repeated units starts the cycle, allowing O-antigen elongation. The interaction of the polymerase with the chain-length regulator Wzz, probably at CL6, limits the number of repeated units that are added to the growing chain, allowing the formation of an O antigen of the correct length, that is released from Wzy_{Pa} (Islam et al. 2011).

Studies conducted on the homologous Wzy protein from *Salmonella enterica* suggest that Wzy is able to discriminate between O units with identical sugar composition but different internal glycosidic bonds. Wzy selectivity towards specific O units has important implications for pathogenesis and bacterial colonization. During infections, mutations of Wzy are selected to bind to sugar oligomers that produce a LPS that induces a reduced inflammatory response in the host, improving bacterial survival (Hong et al. 2015).

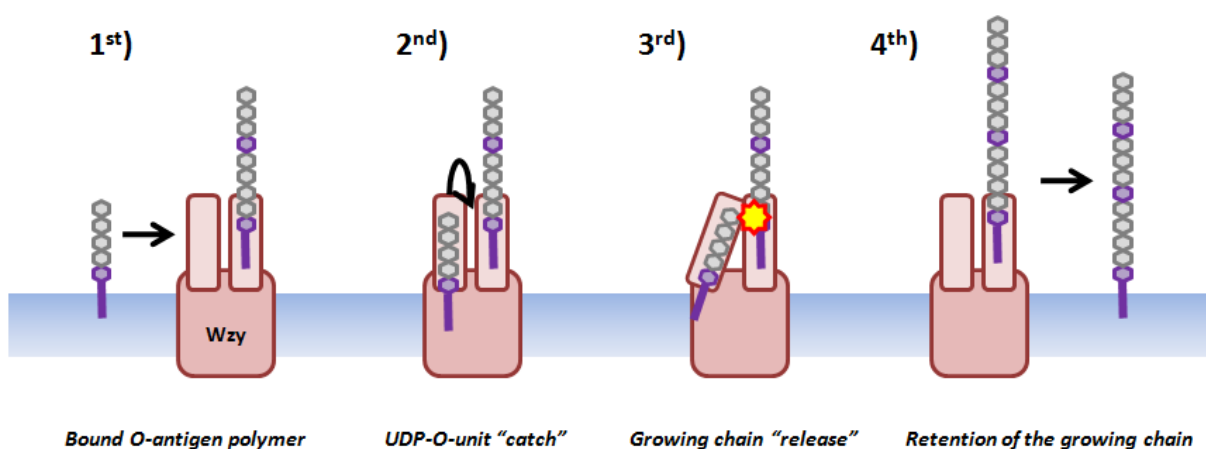


Figure 3.10: Catch and release mechanism for Wzy. Image adapted from Islam S. T. et al (2011), *Journal of Biological Chemistry* (Islam et al. 2011).

RESULTS AND DISCUSSION

The integral membrane protein Wzy from *P. aeruginosa* was chosen among the proteins involved in LPS production for its fundamental role and its species and strain specificity. To date, no experimental structure is available for this protein, but structural information on Wzy would be an important starting point for rational drug design targeted at this protein. Inhibitors of Wzy would reduce O-specific antigen production, thus impairing bacterial capsule formation and, consequently, making *P. aeruginosa* more susceptible to antibiotic treatment in invasive infections.

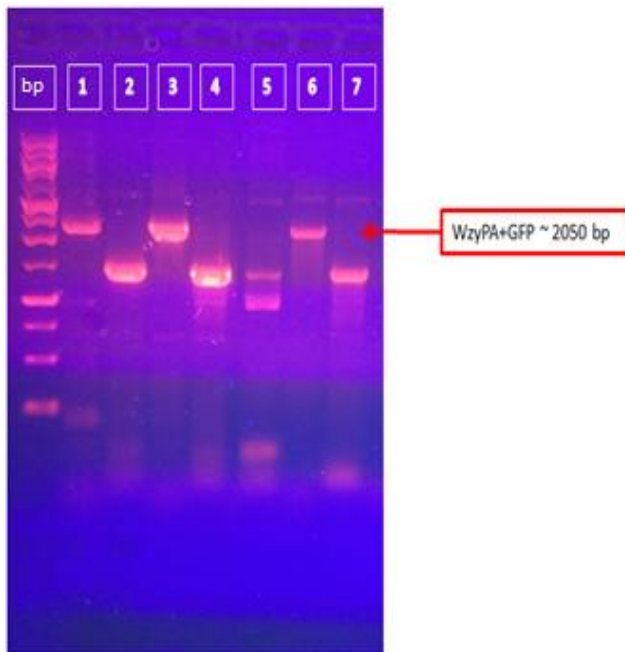
The strategy followed in this work involves the overexpression of the Wzy protein from *P. aeruginosa* in a recombinant form in a non-pathogenic host such as *E. coli*. A GFP tag and a His tag will be added at the C-terminus of the protein in order to detect its presence during all steps, from overexpression to purification and sample characterization.

3.4 WZY RECOMBINANT PROTEIN ANALYSIS

3.4.1 *Wzy*_{PA} cloning

The *P. aeruginosa* PAO1 gene encoding for the polymerase Wzy was cloned from the bacterial genome and inserted into pWALDO_GFP plasmid with the Restriction Free (RF) cloning method (see Materials and Methods). The vector has been developed for membrane protein overexpression and is designed to add a tandem GFP/His tag at the C-terminus of the target gene. The protein transcription is under control of a T7 RNA polymerase promoter and can be used in inducible host systems. The T7 RNA polymerase is more active than bacterial host polymerases, synthesizing RNA at high rates and terminating synthesis less frequently (Tabor 2001). In addition, for colony selection, the plasmid contains a gene coding for Kanamycin resistance. *E. coli* XL1-Blue cells were transformed with the plasmid and plated on a Kanamycin-containing medium.

Transformed colonies were screened to control correct gene insertion by PCR and agarose gel electrophoresis. In **Figure 3.11**, only colonies 1, 3 and 6 show the correct plasmid insertion. Furthermore, insertion and gene sequence were confirmed by sequencing positive colonies. Plasmids with correct insertion were used for protein overexpression.



In the construct carried by the pWaldo_GFP plasmid, the sequence of Wzy_{PA} is followed by a GFP/10x His tag at its C-terminus (from this point on defined only Wzy).

In the construct carried by the pWaldo_GFP plasmid, the sequence of Wzy_{PA} is followed by a GFP/10x His tag at its C-terminus (from this point on defined only Wzy).

Figure 3.11: Control PCR of Wzy_{PA} with T7/T7-term primers on different *E. coli* colonies (1-7). The first lane contains markers for the number of base pairs. Red arrow: DNA bands of correct base pair size (~2050 bp).

3.4.2 Overexpression of Wzy

Two different *E. coli* strains were selected to test Wzy overexpression: BL21(DE3)pLysE and Lemo21(DE3). Both these strains contain the DE3 prophage that codes for the T7 RNA polymerase under control of a *lac* promoter. In this system, protein production is triggered only in presence of isopropyl- β -D-thiogalactopyranoside (IPTG). IPTG mimics the molecular structure of a lactose metabolite, the allolactose, able to bind to the *lac* operon, thus stimulating the transcription of the T7 RNA polymerase, that in turn initiates the transcription of the target gene. Moreover, both strains contain additional plasmids, pLysE and pLemo, respectively, that reduce to a minimum the levels of basal expression of the target protein, by expressing T7 lysozyme, an inhibitor of the T7 polymerase.

In BL21(DE3) pLysE, low levels of T7 lysozyme are constitutively expressed by the bacterial cell. Additional repression of the basal expression level in this strain was obtained by adding glucose to the culture medium. Glucose works as inhibitor for the cyclic AMP receptor, that bind upstream the *lac* promoter stimulating RNA polymerase activity (Novy and Morris 2001).

In the Lemo21(DE3) *E. coli* strain, the pLemo plasmid contains a T7 lysozyme gene under the control of a rhamnose promoter, therefore allowing a tight modulation of the basal expression inhibition. This strain is specifically engineered for overexpression of toxic or difficult proteins, including membrane proteins (Schlegel et al. 2012).

Both pLysE and pLemo contain a gene for Chloramphenicol resistance.

BL21(DE3)pLysE and Lemo21(DE3) *E. coli* cells were transformed with pWALDO_GFP Wzy plasmid. Overexpression was tested in a small scale in Luria-Bertani (LB) medium with Kanamycin and Chloramphenicol, varying different parameters in the two expression systems. In BL21(DE3)pLysE cells, the amount of IPTG (0.2-0.5-1 mM) for induction and glucose (1-2% w/v) for repression of the basal expression were changed. In Lemo21(DE3) cells, the amount of IPTG was kept at a constant value of 0.4 mM, but the rhamnose concentration was changed (0-0.25-0.5-0.75-1-2 mM) in order to modulate repression of the basal expression level. In addition, for both systems, expression levels were tested after 5 h or overnight from induction at two different temperatures (37°C or 20°C respectively).

For all tests performed, the amount of protein expressed was evaluated by measuring the GFP fluorescence (**Figure 3.12**), while cell replication rate was monitored by measuring the optical density OD₆₀₀ of the culture. Results were compared with not induced cells in the same media and growth conditions.

Protein overexpression was low for BL21(DE3)pLysE cells in all tested conditions, since GFP fluorescence values reach a maximum of 1500 a.u. for the tested conditions. On the contrary, for Lemo21(DE3) the condition of overnight growth at 20°C with 0 mM rhamnose (**Figure 3.12 B, first red column**) led to higher overexpression levels, reaching 2800 a.u. of fluorescence. As expected, the OD₆₀₀ values measured in all conditions were significantly lower for the induced sample compared to the negative control, indicating the expression of a potentially toxic protein. As expected, lowering the temperature of the culture had a positive effect on protein production.

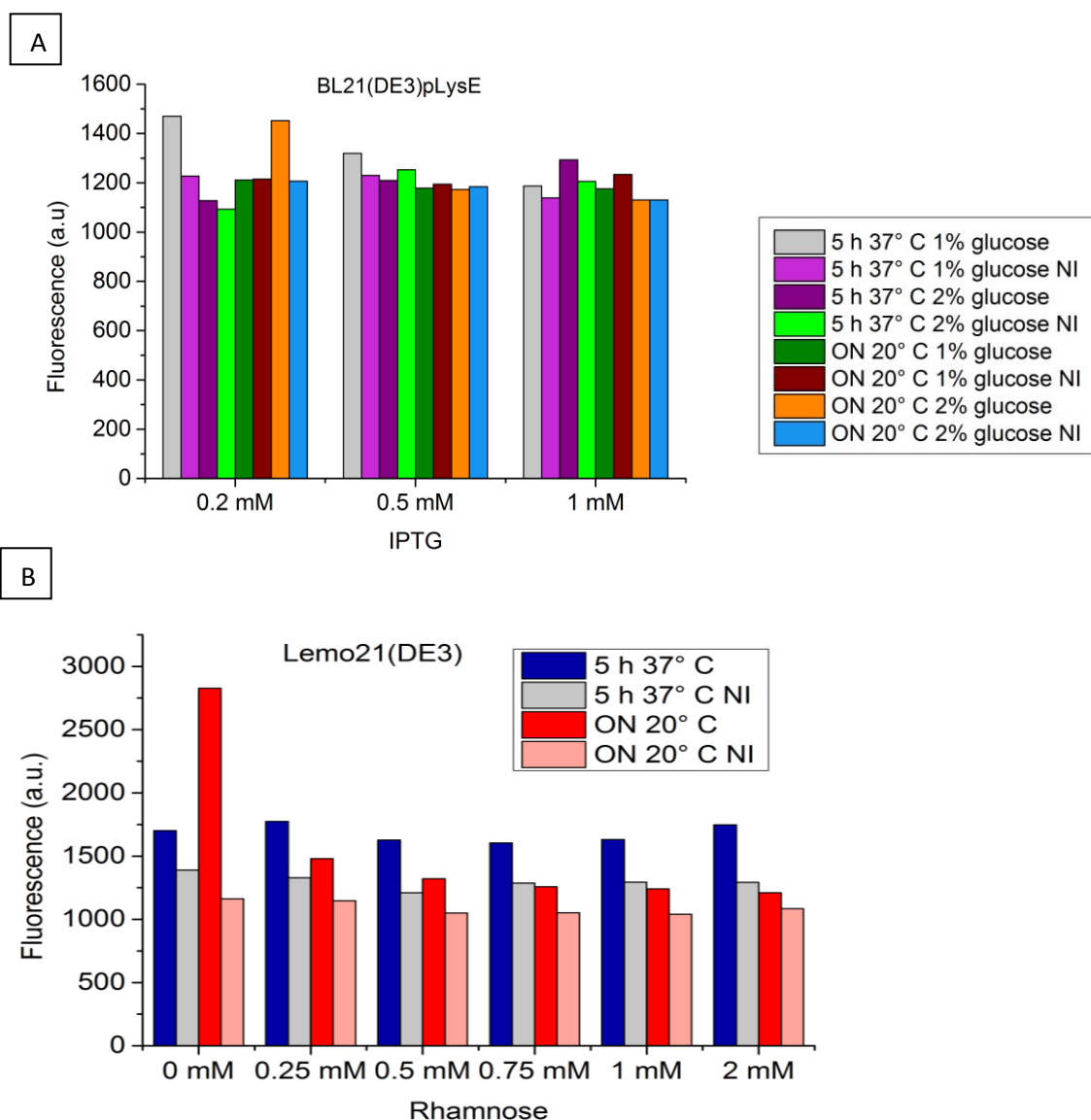


Figure 3.12: Wzy overexpression in different *E. coli* strains. **A)** Expression in BL21(DE3)pLysE. **B)** Expression in Lemo21(DE3). GFP fluorescence was recorded during growth. ON: overnight; NI: no induction.

Considering results obtained from preliminary tests, large scale overexpression of Wzy was performed in Lemo21(DE3) cells grown in LB medium without rhamnose, inducing expression with 0.4 mM IPTG at 20°C, and collecting the cells after overnight expression. However, during scale-up the addition of 0.25 mM rhamnose showed better results for protein expression compared to the condition with no rhamnose. Apparently, inhibition of T7 RNA polymerase achieved through expression of T7 lysozyme becomes necessary for protein expression during scale-up.

After expression, cells were collected and lysed. From the cell lysate, membranes were recovered after ultracentrifugation. Membranes had high fluorescence levels, confirming the membrane localization of the Wzy chimeric protein.

3.4.3 Solubilization tests in different detergents

To test detergent capacity to solubilize the polymerase protein Wzy, OG, DM, DDM, LDAO and C12E8 detergents were chosen. In order to maintain and stabilize the native structure of the protein, only mild detergents such as non-ionic or zwitterionic molecules were selected. Equal amounts of membranes, deriving from 40 ml of culture for each test, with overexpressed Wzy protein were diluted in buffer (Tris 20 mM pH 8.2, NaCl 150 mM) with detergent to same final volume. Detergents were used at 10x CMC, concentration that allows extraction from membranes, except for LDAO, that was tested both at 10x CMC and at 22x CMC (1% w/v), and C12E8, that was used at a concentration of 200x CMC (1% w/v), according to protocols present in the literature (Hirayama et al. 2013). Solubilization of the protein was tested after 1 h and overnight incubation, except for LDAO 22x CMC that was incubated only for 1 h to prevent protein aggregation with the harsher zwitterionic detergent. **Figure 3.13** shows fluorescence levels measured on the supernatant, after centrifugation to remove non solubilized components. Among the five detergents tested, only OG led to poor protein solubilization, while the higher levels were obtained with DM and LDAO solutions, as expected considering the stronger detergent properties of LDAO and DM compared to the other detergents used in this test. For DDM and C12E8, a lower but acceptable protein solubilization was obtained, probably due to their milder characteristics.

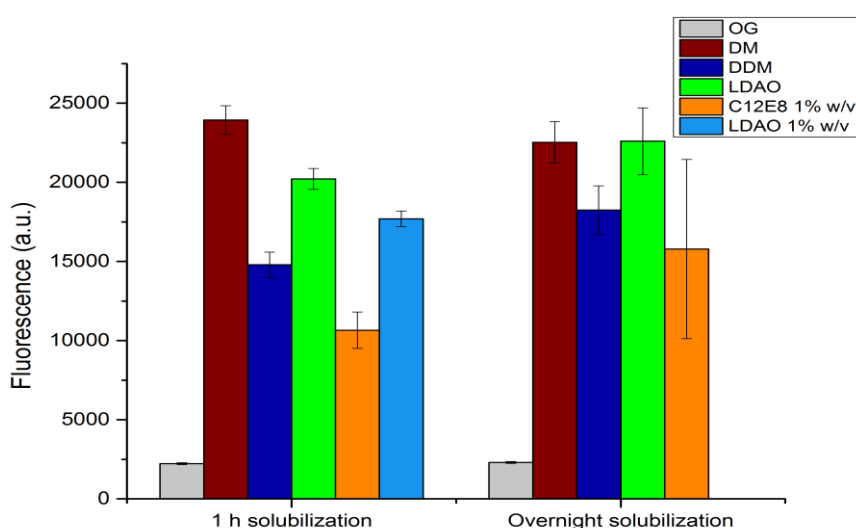


Figure 3.13: GFP fluorescence obtained in the solubilization test on Wzy with different detergents for 1 h or overnight. Detergent solutions were at 10x CMC (OG 180 mM, DM 18 mM, DDM 1.2 mM, LDAO 20 mM), except C12E8 (200x CMC, 1% w/v) and LDAO (22x CMC, 1% w/v). Error bars: standard deviation of three fluorescence measurements.

Comparing the results from the short (1 h) and the overnight incubation, only small differences in the amount of solubilized protein were observed between the two protocols. Therefore, the shorter incubation was preferred, in order to reduce potential protein denaturation occurring over longer incubation times. However, since high solubilization rates are not always indicative of good protein stability in the detergent, the protein samples in different detergents were purified using different chromatographic techniques, to better assess Wzy behavior in detergent.

3.4.4 *Affinity chromatography purification*

Three different affinity techniques were tested for the purification of Wzy taking advantage of the presence of a GFP tag and a 10x His-tag at the C-terminus of the recombinant protein. The first technique involved the interaction of the sample with a resin specifically developed for immunoprecipitation of GFP-tagged proteins and containing an antibody that specifically and selectively binds the GFP. Unfortunately, this technique can only be used for explorative analysis, due to the lack of a suitable interactor with an affinity to the resin higher than the GFP itself. The resin can be recycled using an acidic pH, but the protein is damaged during elution.

The other two affinity chromatography techniques take advantage of the affinity of histidine residues for divalent cations, specifically Ni^{2+} and Co^{2+} ions, and are suitable for the purification of His-tagged proteins. The protein is eluted from the resin when imidazole is added due to the higher affinity of this molecule for the metal ions. Aspecific interactions may occur between the resin and negatively charged protein or histidine-rich proteins. However, in most cases a thorough wash of the resin with solutions with a low imidazole concentration can reduce the aspecific binding.

Similar amounts of membranes, deriving from 0.2 l of culture for each test, were solubilized in buffer (Tris 20 mM pH 8.2, NaCl 150 mM) with LDAO 10x CMC and incubated with the three affinity resins for 1 h at low temperature (4°C). Resins were separated from the supernatant solution (flowthrough).

In the case of the Ni^{2+} and Co^{2+} resins, elution was performed with buffers containing 250 mM imidazole and 150 mM imidazole, respectively. The different concentration used for

elution is related to the different affinity of the metal ions for the histidine moiety. In the case of the resin containing anti-GFP antibody, bound protein was removed using a glycine 0.2 M solution at pH 2.5. Elution fractions were collected until complete removal of Wzy from the resins. **Figure 3.14** shows SDS-PAGE analysis of the eluted fractions, assessed using Coomassie staining and in-gel fluorescence of chimeric-GFP protein. Flow through, resin washes and elution were analyzed. In all gels Wzy was detected in eluates indicating that the protein binds to the resin with good affinity.

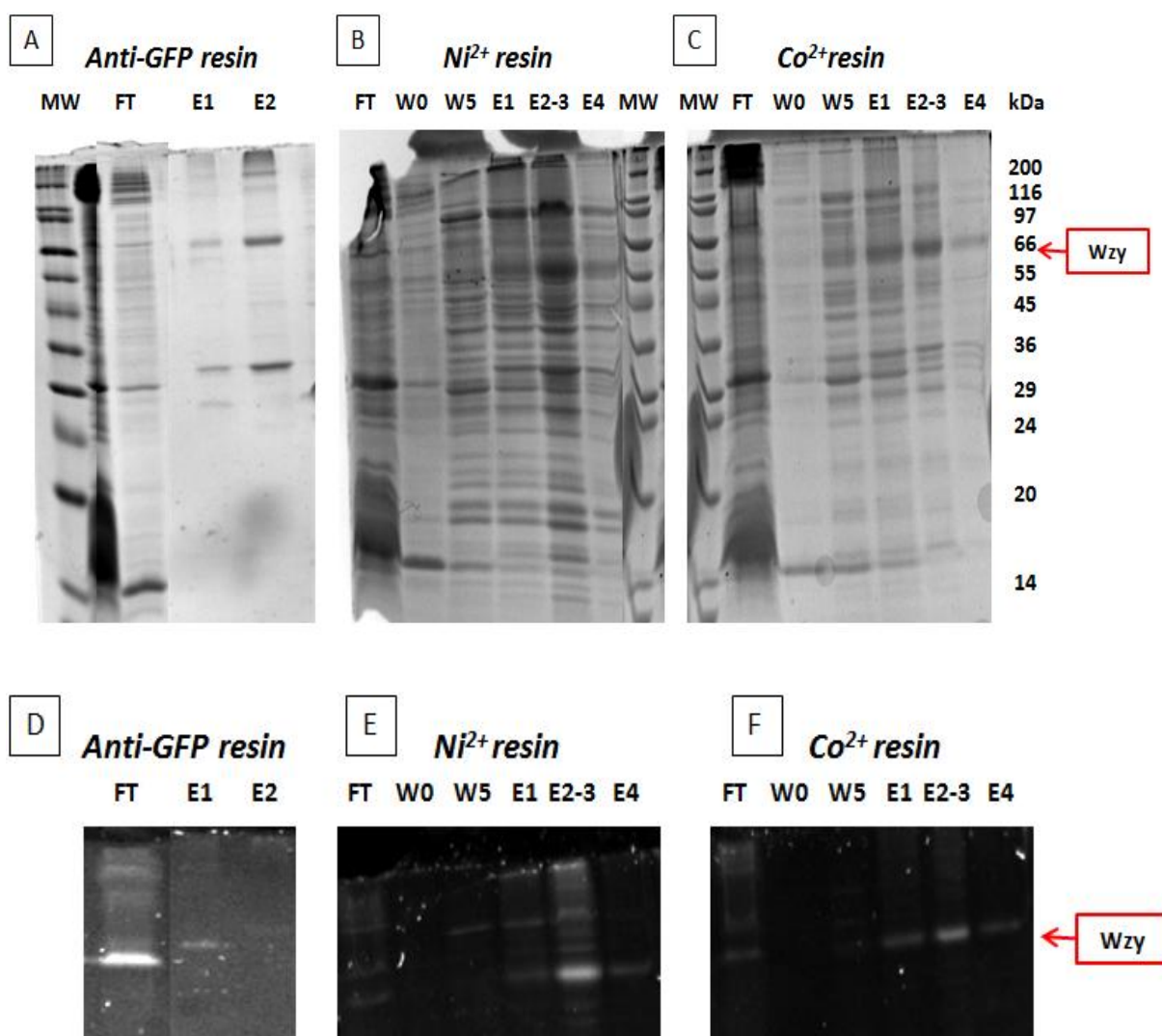


Figure 3.14: SDS-PAGE analysis of Wzy affinity chromatography purification. **a-c)** Coomassie staining; **d-f)** In-gel fluorescence detection. FT: resin flow through; W0 (only for Ni-NTA and Talon Superflow resins): first wash with Wzy buffer with LDAO 20 mM; W5 (only for Ni-NTA and Talon Superflow resins): second wash with Wzy buffer with LDAO 20 mM and 5 mM imidazole. E: protein elution from the resin according to type of resin; MW: molecular weights. Red label: Wzy.

The higher purity obtained for the test using the anti-GFP resin is visible comparing the Coomassie stained gels (**Figure 3.14a** compared to **Figure 3.14b and c**), confirming a very specific binding and a high affinity of the antibody. Unfortunately, while useful to confirm the presence and the identity of the protein, this method cannot be used for preparative purification due to the lack of an elution method that preserves the protein native conformation. In all affinity purification tests, including the separation with the very specific anti-GFP resin, the eluted fractions contain also a second protein at a molecular weight of 36 kDa.

Wzy apparent molecular weight in SDS-PAGE analysis is between 55-66 kDa, even if the total molecular weight of the recombinant protein is 78 kDa. This behavior is not unusual for membrane proteins. In fact, due to their high hydrophobicity, their migration in the electrical field differs significantly from the soluble proteins used as molecular weight standards. However, in-gel fluorescence allowed to unequivocally assign the bands to Wzy.

The feeble Wzy fluorescence observed in fractions E1-2 in **Figure 3.14d** is probably due to protein denaturation in the harsh elution conditions (acidic pH); the protein is, however, clearly visible in the Coomassie stained gel (**Figure 3.14a**).

The fluorescence measurements performed directly on the resin allowed to detect a significant amount of Wzy bound to Ni⁺² resin even after an imidazole elution. This is probably the result of the denaturation induced by the larger amount of imidazole required for elution from Ni⁺² resin (250 mM) compared to the Co⁺² resin: the denatured protein precipitates within the resin and cannot be eluted. In fact, imidazole is known to cause denaturation and should be quickly removed from the protein solution.

Due to the denaturing conditions required for protein elution, anti-GFP and Ni⁺² resins cannot be used for preparative purification of Wzy. On the contrary, purification using Co⁺² resin showed the presence of the protein in all eluted fractions and the protein was fully recovered from the resin. However, the presence of other proteins in the eluted fractions from the Co⁺² resin highlights the need for further purification steps.

Further experiments were performed to test protein purification using different detergents, with the aim of reducing contaminant proteins in the eluted fractions. Co²⁺ affinity purifications were performed using protein solutions in the detergents that gave positive results in the solubilization test, i.e. DM, DDM, C12E8 and LDAO. DM and LDAO (**Figure 3.13**)

concentrations were kept at 10x CMC for both solubilization and purification. DDM was used at a concentration of 20x CMC in the whole process, considering the reduced solubilization capacity of this mild detergent inferred from the previous experiments. C12E8 was used at a concentration of 200x CMC (1% w/v) during solubilization and 6x CMC (0.03% w/v) for purification, as in literature protocols (Hirayama et al. 2013). Considering the good solubilization obtained in LDAO, the purification test was performed also using the zwitterionic detergent CHAPS.

For all tested detergents, similar amounts of membranes, deriving from 0.2 l of culture for each test, containing overexpressed Wzy were diluted to the same volume with buffer (Tris 20 mM pH 8.2, NaCl 150 mM) and detergent. Affinity chromatography experiments were performed with the same protocol used in the previous experiment and the results analyzed with SDS-PAGE, both by staining the gel with Coomassie staining and by measuring the in-gel fluorescence (**Figure 3.15**). Wzy band was visible in all the detergents tested (DM, DDM, CHAPS, C12E8, and LDAO).

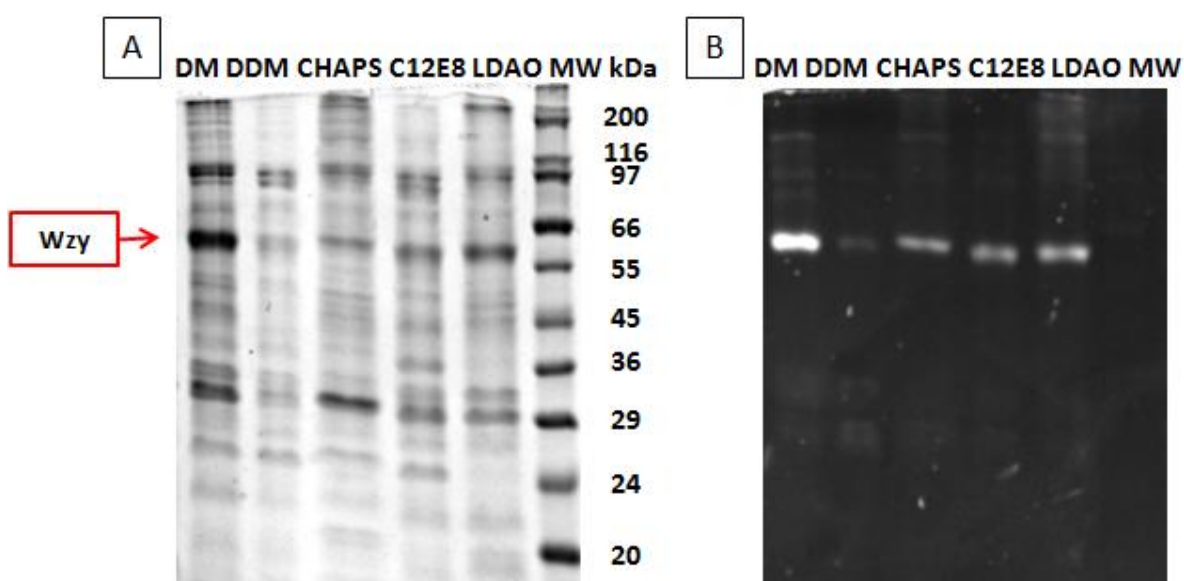


Figure 3.15: SDS-PAGE analysis of Co^{2+} -affinity chromatography purification of Wzy in different detergents. For each detergent, eluted fractions were loaded on polyacrylamide gel. **A)** Coomassie staining and **B)** in-gel fluorescence.

To confirm the results obtained from gel electrophoresis analysis, GFP fluorescence of the eluted fractions for each affinity purification test was measured (**Figure 3.16**). The highest recovery rates were obtained for proteins solubilized in DM and LDAO. The use of DDM in all the steps of the procedure yielded unsatisfactory results considering that, according to the

GFP fluorescence measured in the eluted fractions, only 9% of the initial protein was recovered. The comparison of the value of the final yield with the satisfactory results of the solubilization test suggests that the DDM reduces the affinity of the protein tag for the metal ions of the resin. This result is surprising considering the similar structures of DM, a harsher detergent, and DDM, a milder one. It is possible that the larger size of the DDM micelle compared to that of DM has the effect of shielding the tag.

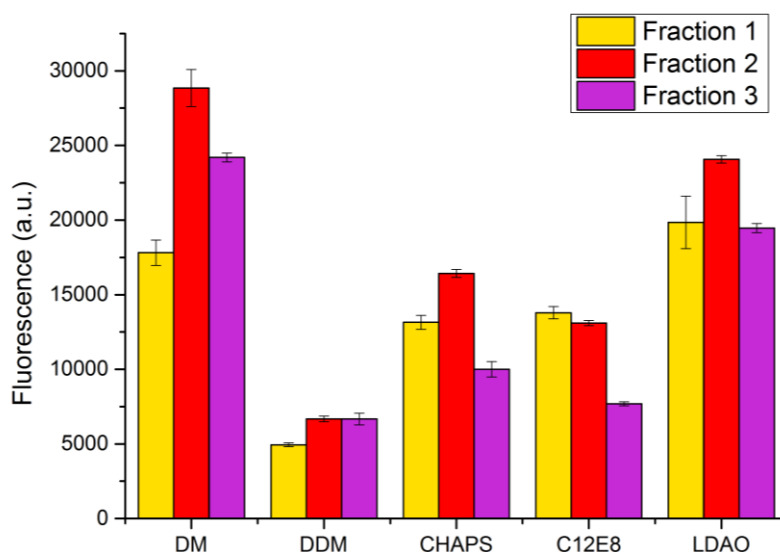


Figure 3.16: GFP fluorescence measurements on elution fractions from affinity purification of Wzy with Co^{2+} resin in different detergents. Error bars: standard deviation of three fluorescence measurements.

Considering the results of solubilization (**Figure 3.13**) and purification tests (**Figure 3.15**) in different detergents, DM and LDAO were chosen as the ideal detergents for Wzy, as they not only demonstrate the highest solubilization yields, but also have been proved not to interfere with the His-tag binding to the metal ions of the resin. However, in both detergents a considerable amount of contaminant proteins is still present in the eluted fractions, pointing to the need of a further purification step.

DM, DDM and LDAO were chosen for the following purification experiments using Size Exclusion Chromatography (SEC). DDM was added despite the poor results from IMAC purification, as this detergent might have a less denaturing effect on the protein conformation compared to DM.

3.4.5 Size Exclusion Chromatography purification

After IMAC purification with Co^{2+} resin, samples solubilized in different detergents underwent a further purification step using a Superdex 200 10/300 GL column (from GE Healthcare Life Sciences).

For this experiments, DM and LDAO detergents were used at 10x CMC both for solubilization and purification. The SEC purification experiment in DDM, however, followed a different route. In order to overcome the lower solubilization power of DDM and the reduced binding affinity to the IMAC induced by the presence of this detergent, the protein sample was first solubilized in LDAO 10x CMC and loaded on the IMAC resin. The following elution was carried out in DDM at a concentration of 5x CMC. Protein elution was detected by measuring the 280 nm absorbance of the solution (**Figure 3.17**).

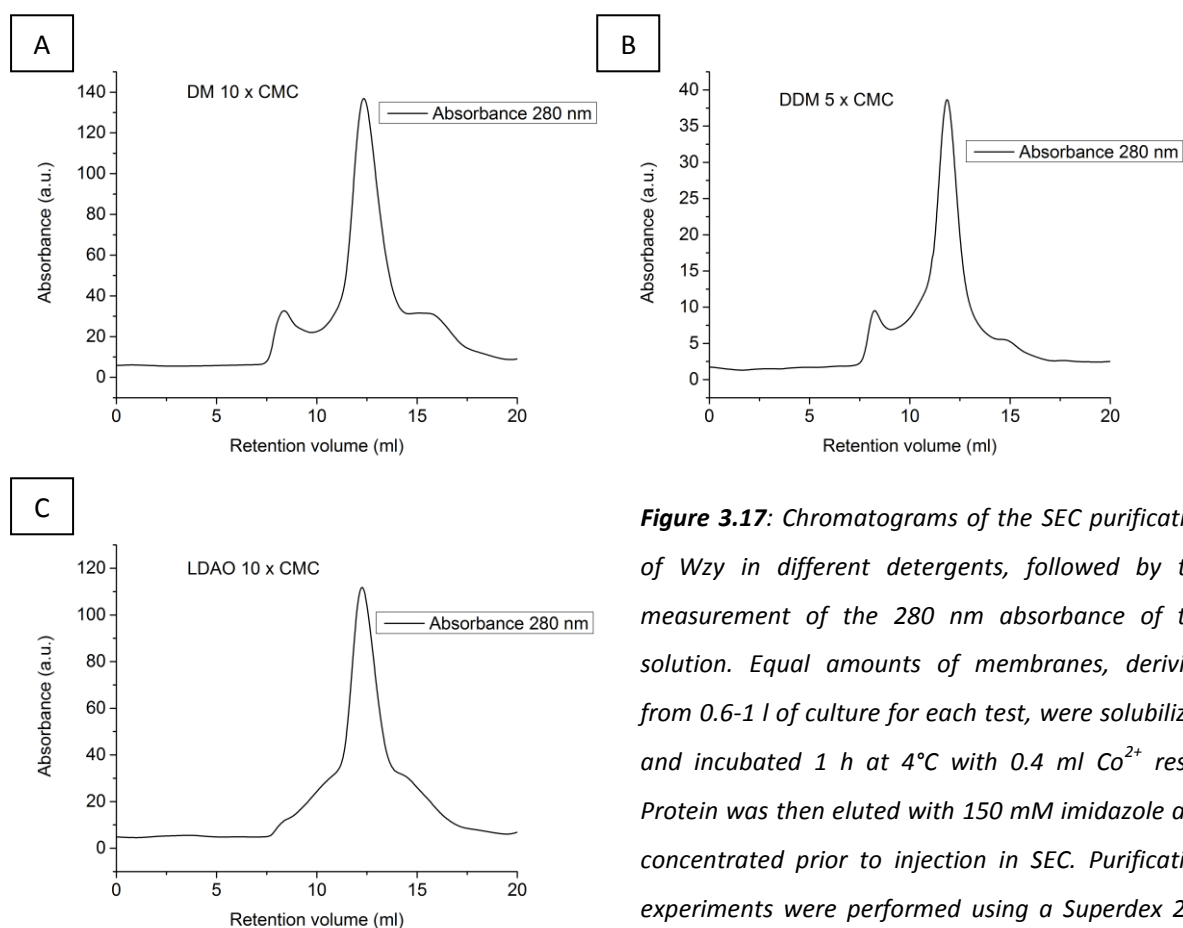


Figure 3.17: Chromatograms of the SEC purification of Wzy in different detergents, followed by the measurement of the 280 nm absorbance of the solution. Equal amounts of membranes, deriving from 0.6-1 l of culture for each test, were solubilized and incubated 1 h at 4°C with 0.4 ml Co^{2+} resin. Protein was then eluted with 150 mM imidazole and concentrated prior to injection in SEC. Purification experiments were performed using a Superdex 200 10/300 GL column.

The chromatograms of the SEC purification in DM, DDM and LDAO show a similar trend: an initial shoulder, followed by a narrow peak around 12-12.4 ml retention volume (**Figure 3.17a,b,c**). In DM the absorbance reached the highest values, closely followed by the LDAO experiment, indicating high protein amount in the eluted solution. Compared to other detergents, the chromatogram of LDAO purification (**Figure 3.17c**) showed a lower signal for the shoulder preceding the main peak, a possible indication of a reduced protein aggregation. For the DDM experiments, values were lower due to poor IMAC resin binding. GFP fluorescence measurements confirmed the presence of Wzy in the sharp peaks of all three purifications. SDS-PAGE analysis showed that the main peak of all three purification experiments contained the target protein Wzy.

To assess the stability of the protein over time, after a week from the first purification the SEC experiment in LDAO was run again. The fractions from the main peak were stored at 4°C for a week, then pooled together, concentrated and re-loaded in the SEC column. The chromatogram of the second purification showed little difference compared to the first, indicating a good protein stability.

Considering the unsatisfactory separation of the shoulder from the main peak of the SEC experiments, a new, more performing column was used to repeat the purification. The Superdex 200 Increase 10/300 GL column (from GE Healthcare Life Sciences) has a higher nominal resolution compared to the Superdex 200 10/300 GL column from the same manufacturer. When the SEC experiment in LDAO was repeated with the new column, the chromatogram displayed a better peak resolution, as shown in **Figure 3.18** (blue line for the 280 nm absorbance measurement). To understand this result, GFP fluorescence was detected during the whole SEC elution (**Figure 3.18 green line**).

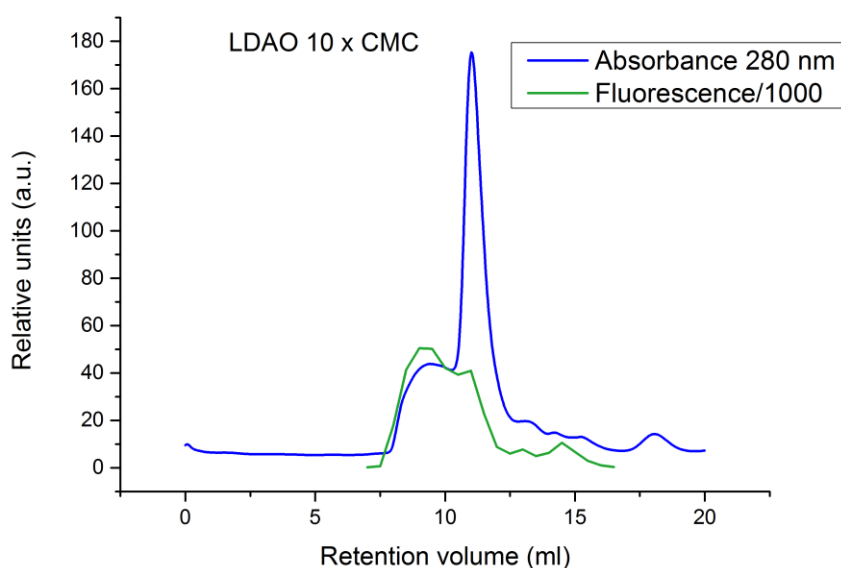


Figure 3.18: Chromatogram of the SEC purification of Wzy in LDAO using a Superdex 200 Increase column, followed by the measurements of the protein absorbance at 280 nm (blue line) and GFP fluorescence (green line).

The high-resolution column was able to better separate the first broad peak from the sharp one. A small shift in the retention volume of the main peak (from 12 ml to 10.5-11 ml) can be attributed to the slightly different value of the column volume in the two different systems. Surprisingly, GFP fluorescence is higher in the broader peak at lower retention volume than in the sharper peak, suggesting that two different protein forms are present and can be separated using a high-resolution column. Considering the results of column standardization with globular proteins of known molecular weight, the peak at elution volume of 10.5-11 ml corresponds to a recombinant protein dimer in LDAO micelle, while in the band at 9 ml the protein is probably eluted in a multimeric form, not suitable for structure analysis. SDS-PAGE analysis was performed on fractions from the SEC purification in LDAO using the Superdex 200 Increase column (**Figure 3.19**).

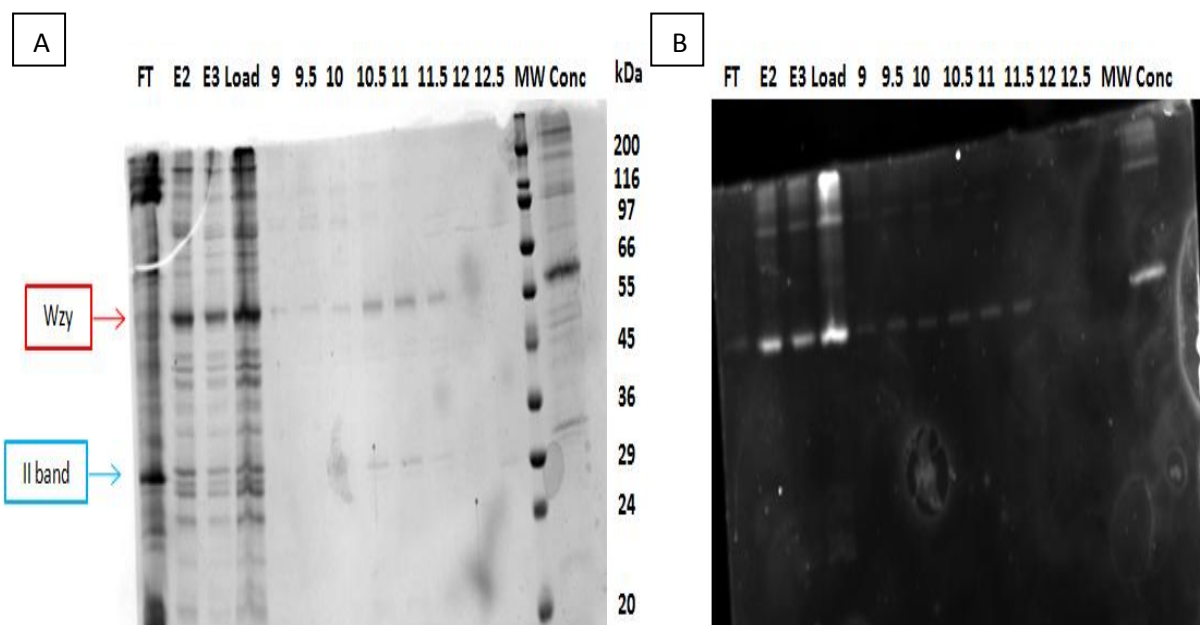


Figure 3.19: SDS-PAGE analysis of fractions from the SEC purification of Wzy in LDAO, performed using the Superdex 200 Increase column. **(A)** Coomassie staining and **(B)** in-gel fluorescence. FT: IMAC flow through; E2-3: IMAC elution; Load: IMAC eluates pooled and concentrated before SEC purification; 9-12.5: fractions of SEC purification, marked with the corresponding retention volume; MW: molecular weight reference; Conc: fractions between 10.5 and 11.5 ml of retention volume, pooled and concentrated to 1.2 mg ml^{-1} of protein.

The comparison between the SDS-PAGE analysis, particularly as regards the in-gel fluorescence, and the fluorescence measurements on the eluted fractions shows a significantly different result: while in the fluorescence measurements performed directly from the elution fractions the highest intensity was detected in the first peak, centered at about 9 ml, the SDS-PAGE in-gel fluorescence shows a higher protein content in the second peak, at 10.5-11.5 ml retention volume (**Figure 3.19b**).

Coomassie staining revealed the presence of a second band at lower molecular weight, around 36 kDa, whose appearance coincides with the second peak (**Figure 3.19 blue label**).

In addition, at higher molecular weight, starting roughly at 200 kDa, aggregates of Wzy are clearly visible in the in-gel fluorescence image. The first of these bands is well defined and has a molecular weight that suggests the presence of trimeric or higher aggregates of Wzy, confirming the presence of aggregation phenomena inferred from the elution volume of the SEC experiment. The intensity of the band is higher in the first fractions of the purification,

consistent with their higher molecular weight. However, after concentration, the band is visible also in the fraction of the second peak.

3.4.6 Mass spectrometry analysis

In order to identify the protein co-purified with Wzy in the previous experiments, mass spectrometry analysis was performed on both the bands obtained from SDS-PAGE. Mass spectrometry analysis is particularly difficult for membrane proteins, due to their hydrophobicity that hampers their charging in the ionization phase, and therefore the separation in the magnetic field. As an alternative, enzymatic protein digestion can be performed before injecting the sample in the spectrometer. In this case, digested peptides are detected in the mass spectrometry analysis, allowing to identify the protein from the fragments by comparison with sequence databases. However, for membrane proteins the tight membrane domain packing and the reduced number of enzymatic cleavage sites make it more difficult to obtain a high yield in the digestion step.

SDS-PAGE bands at 55-60 kDa and at 36 kDa in the Wzy purified sample were analyzed by mass spectrometry after digestion with trypsin. Comparison with sequence databases allowed to assign the peptides identified to the corresponding proteins. Results are reported in **Table 3.2**. In the 55-60 kDa band, peptides belonging to GFP and Wzy_{PA} were identified, confirming the presence of the recombinant Wzy-GFP protein. In addition, peptides of subunit 1 of the Cytochrome O Ubiquinol Oxidase (CUO) from *E. coli* were detected. The second band was identified as the subunit 2 of the same *E. coli* protein. All these data together point to the contamination of the sample with a host protein.

| | Identification | Protein |
|-----------------------|--|--|
| Band 55-60 kDA | Chain A of Green Fluorescent Protein From <i>Aequorea Victoria</i> | GFP |
| | Hypothetical protein M004_29040 [<i>Pseudomonas aeruginosa</i> M10] | Wzy from PAO1 |
| | Cytochrome bo(3) ubiquinol oxidase subunit 1 OS= <i>Escherichia coli</i> (strain K12) | Cytochrome O Ubiquinol Oxidase (sub 1) <i>E. coli</i> |
| Band 36 kDA | Cytochrome o ubiquinol oxidase subunit II [<i>Escherichia coli</i>] | Cytochrome O Ubiquinol Oxidase (sub 2) <i>E. coli</i> |

Table 3.2 Fragment assignment after mass spectrometry analysis of the digested samples of the first and second bands obtained from the SDS-PAGE analysis.

Cytochrome O Ubiquinol Oxidase is a component of the electron transport respiratory chain in *E. coli*. This protein is composed of 4 subunits with molecular weights of 66 kDa, 32 kDa, 20 kDa, and 10 kDa, respectively, with the larger subunit containing two heme groups and a copper ion. CUO is expressed in high concentrations on the bacterial membrane and it was probably co-eluted during Wzy purification. Unfortunately, the presence of the subunit at higher molecular weight cannot be inferred from SDS-PAGE analysis, as its band overlaps with Wzy, while the two lower molecular weight bands are not visible in the gel. Interestingly, contamination was present only in the second peak of SEC purification and not in the first, as demonstrated by the absence of the second band in the SDS-PAGE analysis of fractions corresponding to the first peak (**Figure 3.19**).

3.4.7 UV-visible spectrophotometric analysis

Taking advantage of the known absorbance in the visible spectrum of heme groups, the presence of the contaminant in the SEC fractions was confirmed by UV-visible spectrophotometric analysis (**Figure 3.20**). In fact, heme groups show a strong absorption around 400-450 nm corresponding to electronic transitions in their delocalized aromatic systems and commonly referred as Soret band (Anderson and Robertson 1995). Fractions corresponding to the second peak of the SEC purification in LDAO exhibited an absorbance band around 400-410 nm, as expected in presence of heme-containing proteins, in addition to the typical 280 nm absorption of protein samples containing aromatic residues. However,

the sample from the first chromatographic peak of the same purification experiment shows only the absorbance band centered at 280 nm, confirming the presence of contaminant only in the second peak.

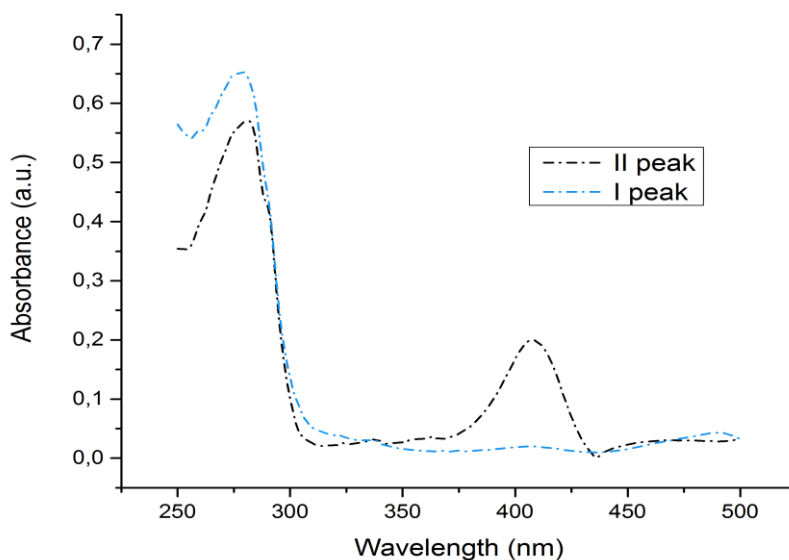


Figure 3.20: UV analysis of purified fractions ($0.5\text{--}0.7\text{ mg ml}^{-1}$) from the SEC experiment.

The UV analysis allowed to calculate the ratio between the target protein Wzy and the contaminant Cytochrome O Ubiquinol Oxidase.

Considering the molar extinction coefficient of the heme group at 404 nm ($1.7 \cdot 10^6\text{ M}^{-1}\text{ cm}^{-1}$) (Karnaukhova et al. 2014), heme concentration could be calculated according to Beer-Lambert law. Since each molecule of Cytochrome O Ubiquinol Oxidase binds 2 heme groups contaminant concentration was estimated. This value was used to calculate the absorbance of the complex at 280 nm according to the molar extinction coefficient calculated from the protein sequence ($3.0 \cdot 10^6\text{ M}^{-1}\text{ cm}^{-1}$). The difference between the experimental value of absorbance of the sample at 280 nm and the calculated absorbance of the contaminant was used to estimate the concentration of the Wzy protein, whose molar extinction coefficient was calculated from the protein sequence ($9.2 \cdot 10^5\text{ M}^{-1}\text{ cm}^{-1}$). According to this estimation, the relative abundance of Wzy: Cytochrome O Ubiquinol Oxidase appears to be 3:1.

3.4.8 Optimization of the affinity purification protocol

Considering the results of the previous SEC, SDS-PAGE, mass spectrometry and UV-visible analysis, further IMAC experiments were devoted to obtain a better separation between the Wzy protein, containing a His-tag, and the contaminant, that probably binds to the Co^{2+} column through aspecific interactions.

In particular, detergent and buffer solutions with incremental amounts of imidazole were applied to the Co^{2+} resin and the eluted fractions analyzed by SDS-PAGE. This procedure allowed the quantification of the exact imidazole amounts required to (1) separate Wzy from Cytochrome O Ubiquinol Oxidase and (2) elute Wzy from the resin. **Figure 3.21** shows electrophoresis gel with Coomassie staining and in gel-fluorescence detection of IMAC fractions with increasing imidazole gradient.

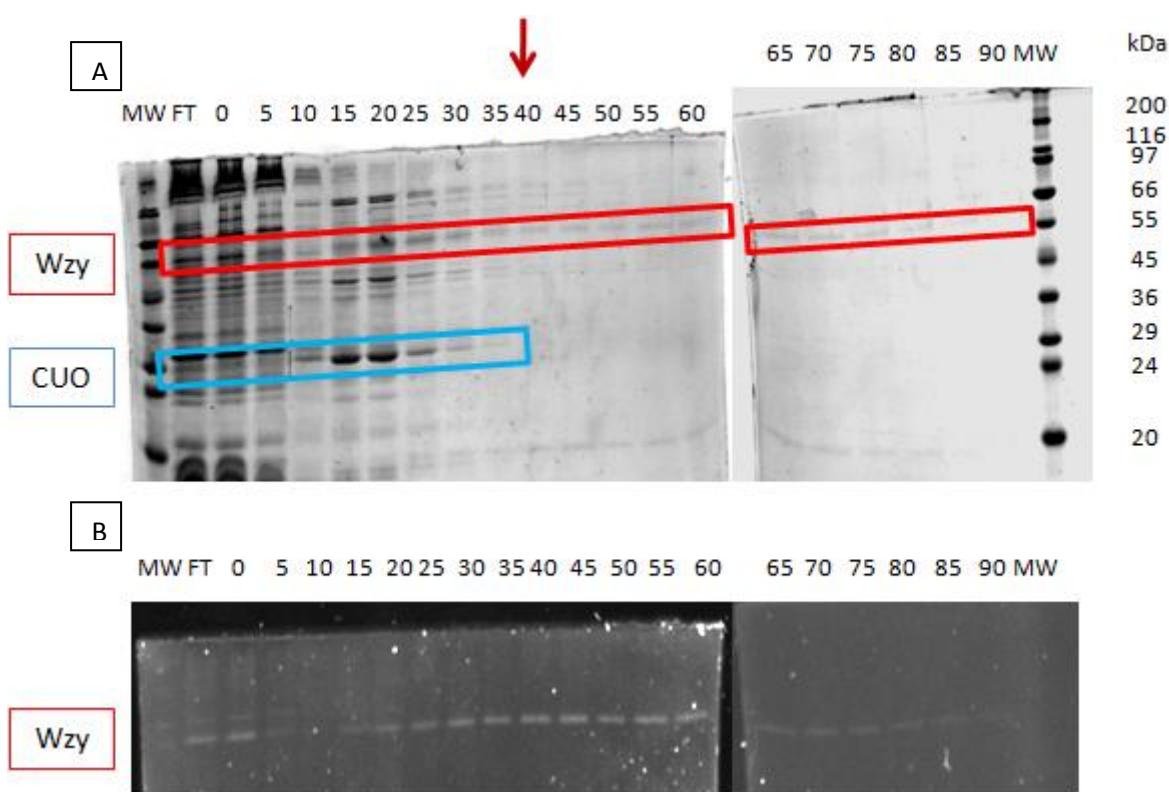


Figure 3.21: SDS-PAGE analysis of the IMAC separation of Wzy with increasing imidazole concentration (from 0 to 90 mM). **A)** Coomassie staining; **B)** In-gel fluorescence of Wzy.

The presence of the contaminant could only be inferred from the band corresponding to the 32 kDa subunit (**Figure 3.21, blue mark**), since the band of the 66 kDa subunit overlaps with Wzy. The band relative to the second subunit of the contaminant, detected at a molecular

weight of 36 kDa, was found in fractions with imidazole concentrations up to 40 mM (**red arrow**). The Wzy protein was present in all fractions up to an imidazole concentration of 85 mM. However, this was the only protein present in fractions above 40 mM of imidazole. Unfortunately, as proved by the presence of amounts of Wzy in all eluted fractions up to a 85 mM concentration of imidazole, the protocol that allows removal of the contaminant reduces significantly the yield of the Wzy purification. To reduce the protein loss, wash steps were reduced to two, at concentrations of 0 mM and 40 mM imidazole, respectively, followed by a final 80 mM elution. This protocol behaved well allowing for complete removal of the CUO. However, SEC purification of this sample revealed a severe protein degradation, as shown in **Figure 3.22**.

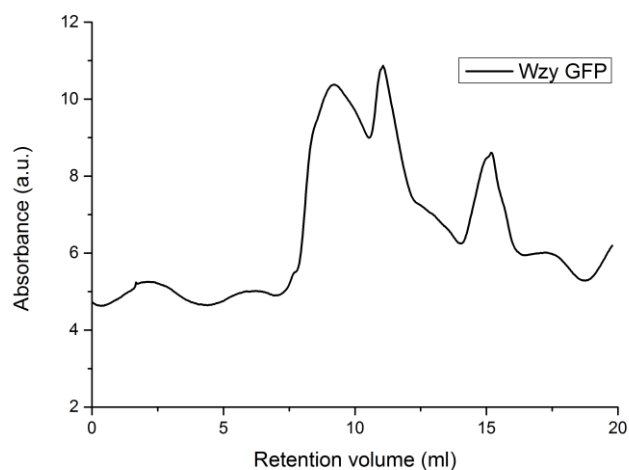


Figure 3.22: Chromatogram of the SEC purification of Wzy in LDAO after an IMAC purification with the optimized protocol.

3.4.9 Ion Exchange Chromatography test

Further efforts were made with anion exchange chromatography (IEX), taking advantage of the different isoelectric points of the two proteins (8.3 for Wzy and 6.7 for the contaminant). The experiment was performed at a pH of 8, in buffer (Tris 20 mM pH 8, NaCl from 150 mM to 1 mM) with LDAO 10x CMC detergent. However, separation of the proteins could not be obtained, probably due to the presence of the LDAO micelle whose effect dominates over protein charges. The chromatogram of this purification is shown in **Figure 3.23**.

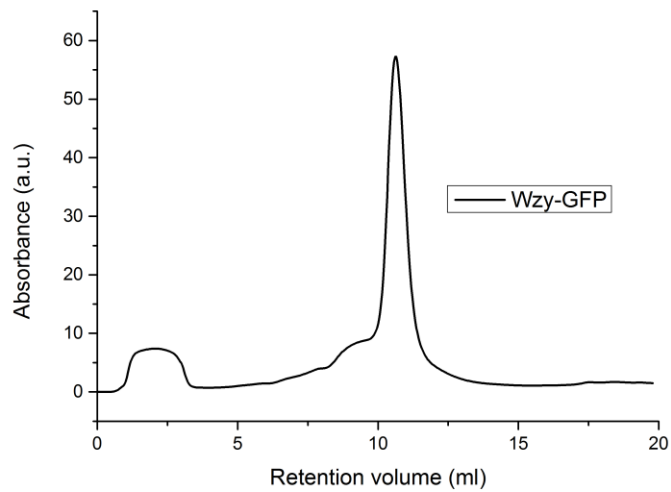


Figure 3.23: Chromatogram of the IEX purification of Wzy in LDAO 10x CMC, followed by measuring the 280 nm absorbance of the solution.

The chromatogram of the IEX purification shows a single absorbance peak at 450 mM NaCl concentration. Fractions of the IEX purification were analyzed with SDS-PAGE (**Figure 3.24**).

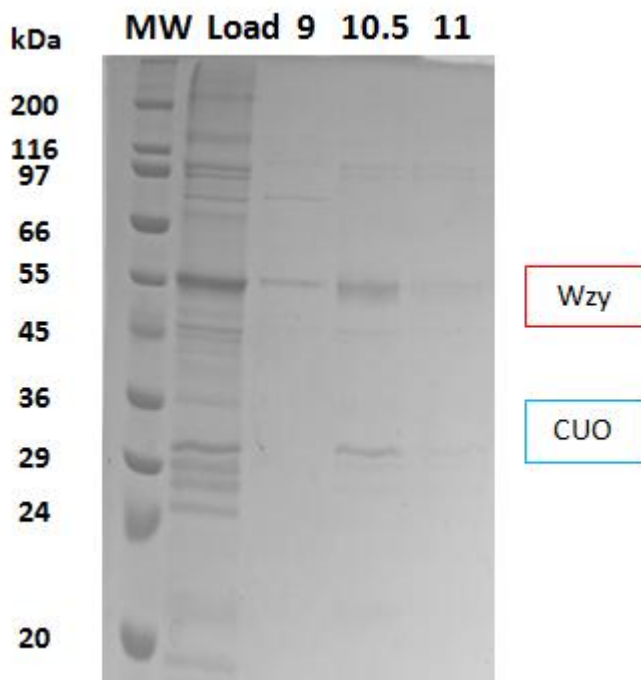


Figure 3.24: SDS-PAGE analysis of the fractions of the IEX purification. MW: molecular weight reference; Load: sample injected in the column; 9-10.5-11 ml: fractions of the IEX chromatogram.

The Coomassie Blue staining of peak fractions, revealed the presence of Wzy in the central peak of the chromatogram, together with the contaminant.

3.4.10 *Comparative analysis of results from chromatographic and electrophoretic tests*

As proved by the chromatographic analysis, removal of the Cytochrome O Ubiquinol Oxidase contaminant could not be achieved with chromatography techniques. In particular, even with specific and very specific affinity purifications, i.e. using the Co^{+2} and the anti-GFP resin, respectively, the contaminant co-elutes with the target protein. These results suggest an interaction of the two proteins within the same micelle. Considering the possibility that target protein and contaminant are co-solubilized in the same micelle, we decided to perform further separation tests by changing the solubilization detergent. However, detergents that allowed protein solubilization and stability were not able to disrupt the interaction between Wzy and the Cytochrome O Ubiquinol Oxidase.

Moreover, when a harsher separation protocol was applied and a sample of Wzy without contamination could be obtained, the polymerase was not viable due to aggregation effects. All these data together suggested a role for Cytochrome O Ubiquinol Oxidase in improving Wzy stability. For this reason, we decided to proceed without contaminant removal, with the aim of characterizing the whole Wzy-contaminant complex formed in the membrane. A role for this interaction is not expected in physiological conditions as the two proteins belong to different pathways and different organisms. However, the analogous protein from *P. aeruginosa* has a very similar sequence (68% identity and 90% similarity) and this protein is conserved among bacterial species.

Proper folding and protein stability were tested prior to structural characterization analysis. Circular Dichroism, fluorescence and Raman spectroscopy techniques were chosen to assess protein secondary structure and unfolding temperature.

3.4.11 *Circular Dichroism analysis on purified Wzy*

Wzy sample purified in LDAO without removing the co-purified Cytochrome O Ubiquinol Oxidase was analyzed by Circular Dichroism at UV wavelengths. To test protein stability, spectra were measured while increasing sample temperature.

The predicted structure of Wzy has a TM domain composed of 14 α -helices, and similarly the experimental structure of Cytochrome O Ubiquinol Oxidase, determined at 3.5 Å resolution by X-ray crystallography (PDB: 1FFT) (Abramson et al. 2000), shows a folding dominated by

α -helices. Therefore, the typical shape of α -helical secondary structures is expected to be present in the CD spectra in the range between 190 and 250 nm. α -helical proteins have spectra characterized by the presence of two minima (at 208 nm and 222 nm, respectively) and a maximum at 191-193 nm (Wei, Thyparambil, and Latour 2014).

When heated above their unfolding temperature, proteins denature and lose the secondary structures. This phenomenon causes a significant change in the CD spectrum that can be detected by collecting CD signals over a range of temperature, allowing to determine the melting point. In the case of a protein with a mainly α -helical secondary structure, signals at 208 nm and 222 nm are expected to increase the absolute value of ellipticity, while the signal at 191-193 nm decreases.

Figure 3.25 shows CD spectra of samples obtained from purification of Wzy in LDAO 10x CMC, without Cytochrome O Ubiquinol Oxidase removal. Total protein concentration was kept at 2-3.7 mg ml⁻¹ and CD analysis was performed over a range from 4°C to 80°C. Values below 200 nm are strongly affected by the absorption of the buffer solution containing chloride ions, and due to the high signal-to-noise ratio were removed from the spectra. Attempts at buffer exchange to a similar buffer containing NaF failed due to protein precipitation.

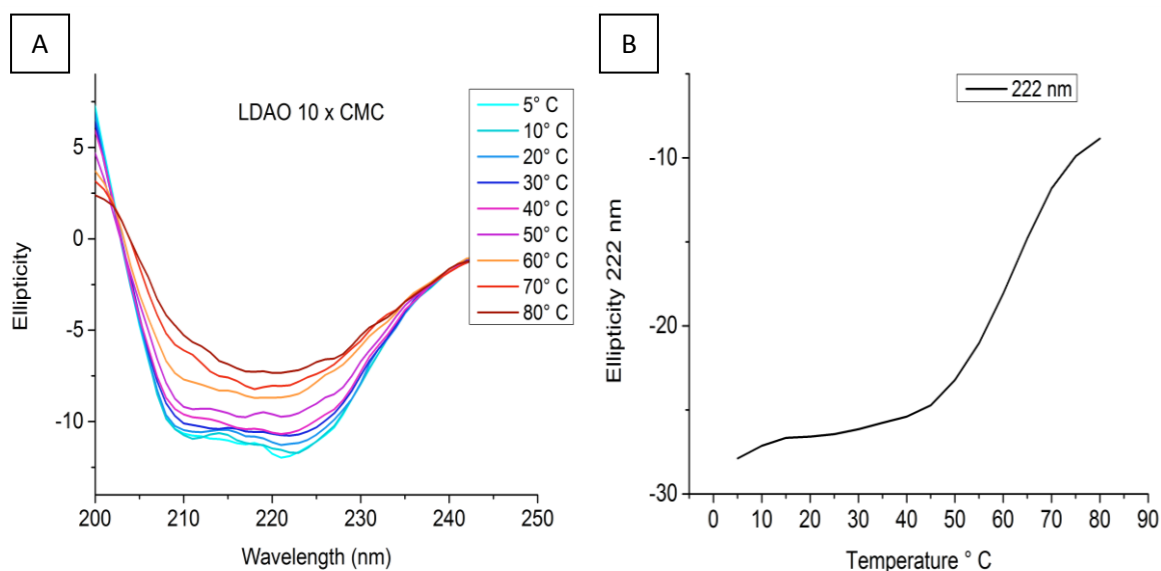


Figure 3.25: CD analysis of samples of Wzy-Cytochrome O Ubiquinol Oxidase. **A)** CD spectra of Wzy-CUO; **B)** Plot of the ellipticity value at 222 nm (α -helix minimum) against temperature. For this analysis a cuvette with a double path length compared to the previous spectra was used in order to increase the signal-to-noise ratio.

Analysis of the low temperature CD spectrum showed the typical α -helix features in the LDAO-solubilized sample (**Figure 3.25a**). During heating, the positive and, particularly, the negative peaks in the spectrum decrease in intensity, indicating helices unfolding and protein denaturation, as expected. This effect could be measured with more precision using a longer path length and a higher concentration of protein. In order to precisely identify Wzy denaturing temperature, absorbance at 222 nm was measured in a cuvette with a double path length (0.2 mm) compared to the previous (0.1 mm) (**Figure 3.25b**). In this case, measurements below 215 nm are affected by buffer absorption, but the ellipticity value at 222 nm has a higher signal-to-noise ratio and can be plotted against the temperature. The CD signal at 222 nm is stable for temperatures between 4 and 45°C, but after 45°C it steadily increases. The temperature range of this variation is quite large, possibly due to kinetic effects that require a longer time than the CD measurement. However, from the inflection point a melting temperature around 55-60°C was inferred.

In order to validate the unfolding temperature resulting from CD analysis, samples were further analyzed by fluorescence spectroscopy.

3.4.12 *Fluorescence spectroscopy analysis of purified Wzy*

Exploiting the 7 tryptophan residues present in Wzy-GFP recombinant protein, 5 of which are expected to be in TM α -helices (**Figure 3.9**), and the 44 tryptophan residues in the Cytochrome O Ubiquinol Oxidase, 28 of which are found in the TM helices, fluorescence was recorded in the tryptophan emission range (300–350 nm) while heating the sample from 8°C to 75°C. The fluorescence signal of the aromatic residues is sensitive to the chemical environment surrounding the fluorophore. In particular, hydrophobicity plays a key role in determining both the position and the intensity of the tryptophan fluorescence. The presence of tryptophan residues located in the TM domain, hence surrounded by a hydrophobic environment, allows to observe changes in the fluorescence signal as the protein loses its native folding. In fact, protein unfolding causes an exposure of the fluorophores to the polar aqueous environment and a consequent change in the fluorescence spectrum. Fluorescence spectra were recorded for specimen obtained with DM and LDAO purification protocols, in a concentration range of 0.75-0.9 mg ml⁻¹ (**Figure 3.26 a and b**).

The intensity of the tryptophan fluorescence emission at the maximum was plotted against the temperature to identify the unfolding point of the proteins (**Figure 3.26 c and d**). For the sample solubilized and purified in DM, the intensity decreased constantly. This behavior can be attributed to slow unfolding processes that occur in the sample at every temperature and is an indication of an intrinsic instability of the protein in this detergent. On the contrary, a sudden decrease in fluorescence intensity occurred for the sample in LDAO between 55°C and 65°C. The shape of the plot of fluorescence intensity against temperature is typical for a protein unfolding that exposes tryptophan residues previously embedded in the TM domain to the hydrophilic environment. This plot indicates an unfolding temperature in the range of 55-65°C, consistent with the value obtained by CD analysis in LDAO detergent solution.

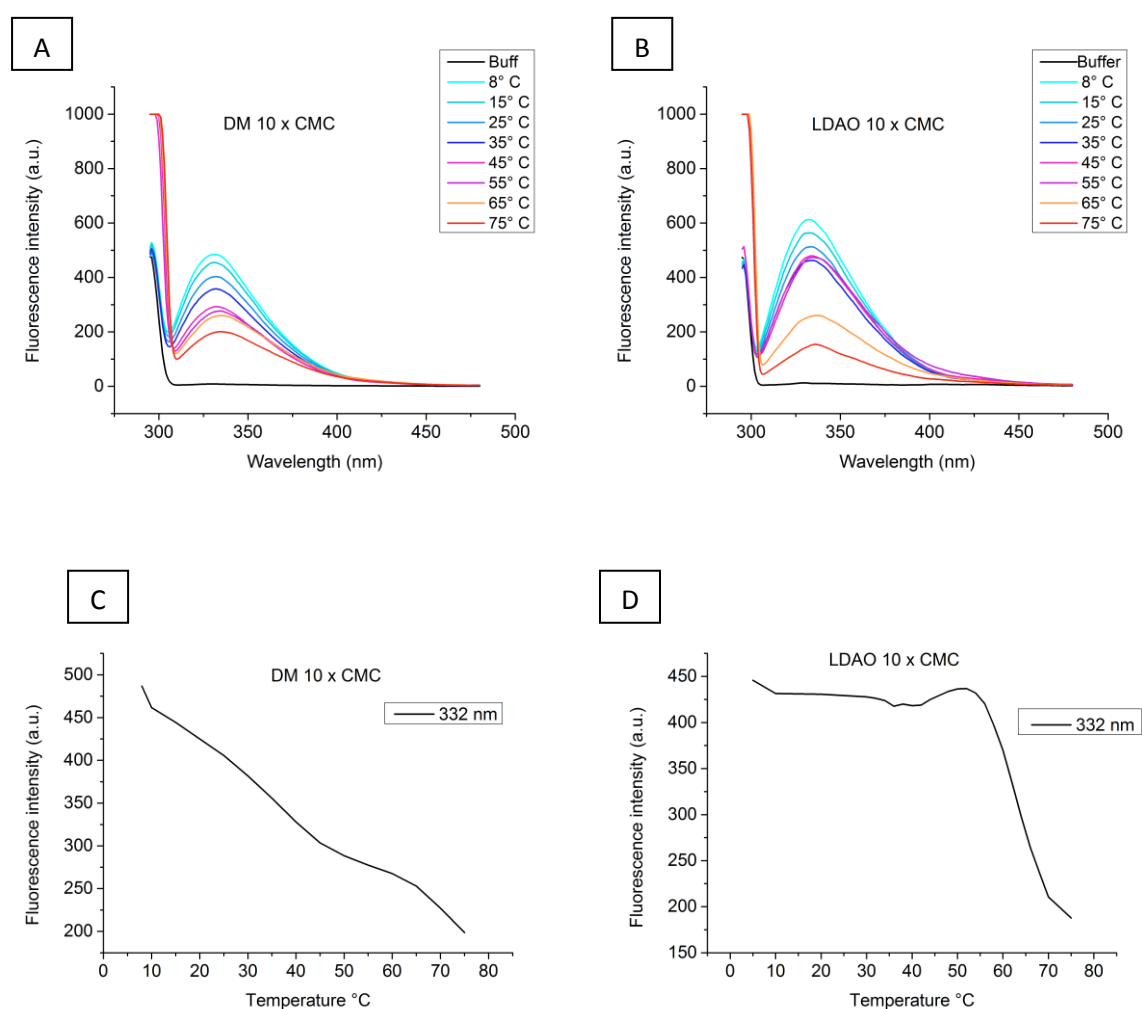


Figure 3.26: Fluorescence spectrometry of the Wzy-Cytochrome O Ubiquinol Oxidase samples in (**a and c**) DM and (**b and d**) LDAO buffer. (**a-b**) Spectra measured at different temperatures in the range of 8-75°C. (**c-d**) Variation of fluorescence intensity at the maximum with temperature.

3.4.13 Raman spectroscopy

Raman spectroscopy is a technique based on inelastic scattering of UV photons and it gives information on the vibrational state of a molecule. For protein samples, the technique is mainly used at wavelengths compatible with the absorbance of the aromatic residues of the protein, i.e. tryptophan, tyrosine and phenylalanine residues. Considering the low probability of the Raman transition, lasers are usually employed to increase the intensity of the signal.

In presence of a single tryptophan residue, the Raman technique can give indications on the precise surroundings of the aromatic moiety, by the analysis of the vibrational transitions of the spectrum. However, when multiple aromatic residues are present, it becomes more challenging to assign a vibrational transition to a single residue. Nevertheless, as for fluorescence spectroscopy, the drastic change in the surroundings of a residue during the transition from a folded to an unfolded state has a significant influence on the Raman spectrum of the specimen (Oladepo et al. 2012). Therefore, recording spectra at different temperatures can give indications regarding the thermostability of the system.

In order to validate the hypothesis of an increased stability of the Wzy protein when in complex with the Cytochrome O Ubiquinol Oxidase, both chromatographic peaks obtained in the SEC purification in LDAO with the Superdex 200 Increase column were analyzed at a concentration of 1 mg ml⁻¹. UV Resonant Raman measurements were carried out at the BL10.2-IUVS beamline of the Elettra Synchrotron and spectra were collected with an excitation wavelength of 244 nm in backscattering geometry with a spectral resolution of about 8 cm⁻¹. Samples were measured at RT (around 19°C) and after heating to 70°C. Raman peaks involving amide III (1238 cm⁻¹), II (1559 cm⁻¹) and I (1666 cm⁻¹) transitions of the peptide backbone, as well as tryptophan (1362 cm⁻¹ and 1623 cm⁻¹) and tyrosine (1623 cm⁻¹) residues could be identified in the spectra. **Figure 3.27** shows Raman spectra of the first and second peak from the SEC purification, at 19°C and 70°C. At low temperature amide I and tryptophan signals are particularly intense, whereas after thermal denaturation these signals were almost undetectable for the sample of the first peak and very weak for the sample of the second peak, i.e. the solution containing Cytochrome O Ubiquinol Oxidase.

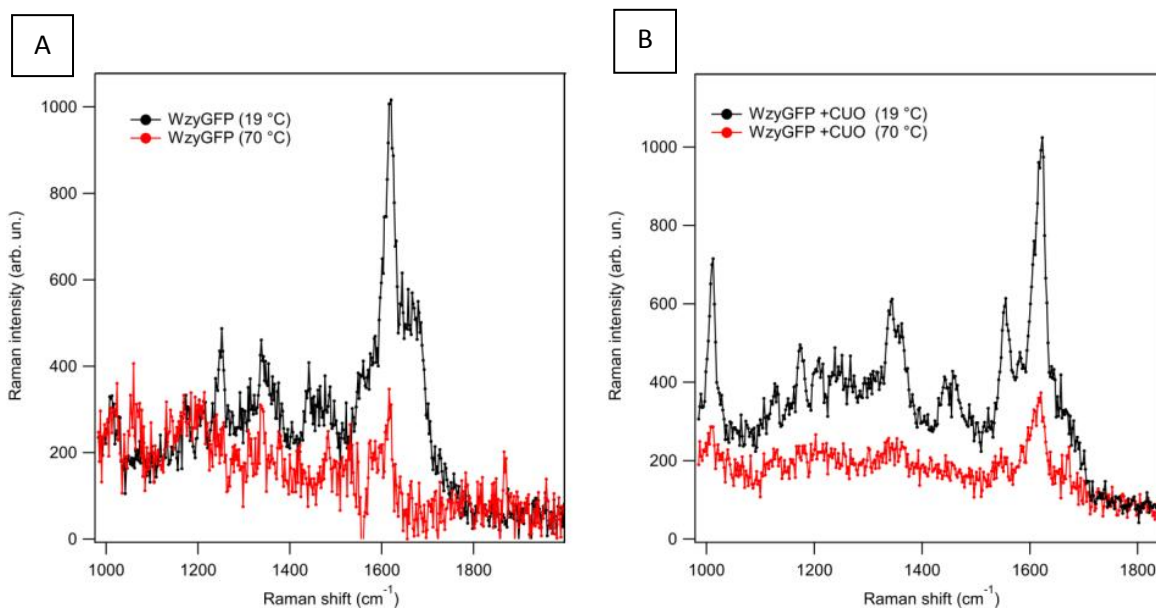


Figure 3.27: Raman spectra of samples from the first (A) and second chromatographic peak (B). Black lines represent spectra obtained at RT, red lines show spectra at 70°C.

Raman intensities were affected by scattering phenomena involving detergent micelles, with a reduction in both the intensity of incident beam and back-scattered Raman radiation. Despite this phenomenon, spectra were interpreted. Data show that Wzy in the first chromatographic peak is completely unfolded at 70°C (Figure 3.27). In fact, amide I signal at 1666 cm⁻¹ was completely absent after heating (Figure 3.27a, red line), suggesting secondary structure loss. On the contrary, for the protein of the second chromatographic peak, both amide I and tryptophan signals were still visible at high temperature (3.27b red line). Possibly, these results point to a higher stability of Wzy in presence of the Cytochrome O Ubiquinol Oxidase. These data are consistent with chromatographic results.

3.4.14 Wzy crystallization

Considering the results of the purification experiments and the stability of the protein assessed with different methods, crystallization tests were set up without further purification. In addition, the GFP/His tag was not removed, as it can help in the formation of the protein-protein interactions required for crystal packing, especially in the case of membrane proteins. In fact, this kind of proteins can hardly create electrostatic interactions

due to their prevalence of hydrophobic residues, while the formation of hydrophobic interactions is hampered by the presence of the detergent micelle.

Crystallization experiments were performed on the samples obtained from the main peak of the SEC purification in DM and from the second peak of the purification with the Superdex 200 Increase column in LDAO, in a protein concentration range between 2.7 and 3.7 mg ml⁻¹. For the initial tests, sparse matrices of conditions were used with the commercial kits MemGold1, MemGold2, MemSyst and MemStart from Molecular Dimensions. Tests were performed at 4°C and 18°C.

From the initial screening, three conditions at 4°C gave positive results with the formation of small crystals. Interestingly, the two hits obtained from tests with the DM-solubilized protein have high pH values (9.0-9.5), while all the positive results included the presence of PEG.

Figure 3.28 shows the crystals obtained in the initial crystallization tests. The crystal habit was different for crystals obtained from DM and LDAO solutions. In the first case, larger needle shaped crystals were obtained, with a pink/red color, while for the second detergent crystals had a cubic habit, were colorless and smaller.

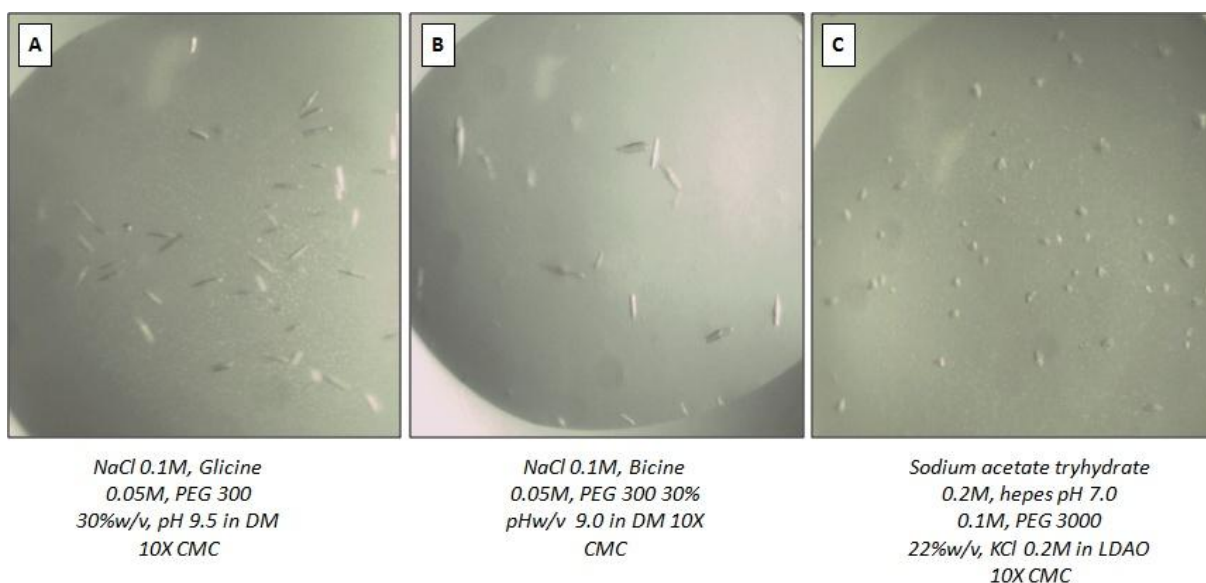


Figure 3.28: Crystals obtained from MemGold1 and MemGold2 screens at 4°C after several weeks. **A-B)** Crystals obtained from the DM solution. **C)** Crystals obtained from LDAO-solubilized protein.

Optimization screens were set up around the conditions that yielded positive results, in order to increase the size of the crystals. In the following experiments, pH, protein amount, precipitant or salt concentrations were varied compared to the initial conditions. Both sitting

and hanging drop vapor diffusion experiments were performed. In addition, dimension of the drops was varied.

In the optimization screens, some of the tested conditions yielded crystals suitable for X-ray diffraction analysis. Data collection was performed at the XRD2 beamline at the Elettra Synchrotron, using an incident radiation of 1 Å wavelength. For 4 crystals complete datasets of diffraction data could be recorded. **Table 3.3** summarizes crystallization conditions, cell parameters and resolution of crystallographic data obtained for these crystals. Interestingly, data obtained from DM-solubilized protein had similar cell parameters. Crystals obtained from LDAO solution had similar cells, too, but they differed significantly from the former. In some cases, data collection could not be completed due to the radiation sensitivity of the sample.

| Crystal | Wzy conc | Condition 4° C | Cell (a,b,c=Å; α,β,γ=°) | Symmetry | Resolution |
|-------------------------|--|--|---|---|---------------------------------|
| A DM | 3 mg ml ⁻¹ drop 0.5 μl | NaCl 0.1M, glicine 0.05 M, pH 9.5, PEG 300 27% w/v | a=68.959, b=186.398, c=186.398 α=β=γ=90.000 | P422 | >7 Å |
| B DM | 3 mg ml ⁻¹ drop 0.5 μl | NaCl 0.1 M, Bicine 0.05 M, pH 9.0. PEG 300 30% w/v | a=180.775, b=182.613, c=67.727 α=β=γ=90.000 orthorhombic | P2 ₁ 2 ₁ 2 | 6.32 Å Completeness 98.8% |
| C LDAO | 3.7 mg ml ⁻¹ drop 0.7 μl | NaOAc*3H ₂ O 0.2 M, hepes 0.1 M, pH 7.0, PEG 3000 22% w/v, KCl 0.2 M | a=127.337, b=147.176, c=196.283 α=89.738, β=89.942, γ=89.941 | P1 | >7 Å Completeness 78% |
| D LDAO | 3 mg ml ⁻¹ drop 0.5 μl | NaOAc*3H ₂ O 0.2 M, hepes 0.1 M, pH 7.0, PEG 3000 22% w/v, KCl 0.2 M | a=125.818, b=146.783, c=191.632 α=β=γ=90.000 | P2 ₁ 2 ₁ 2 ₁ | >7 Å Completeness 70% |

Table 3.3: Results obtained in the crystallization tests and diffraction experiments.

For crystal B, the higher resolution of the data allowed to solve the phase problem using molecular replacement with the structure of the Cytochrome O Ubiquinol Oxidase as a starting model (PDB: 1FFT) (Abramson et al. 2000). Despite the low resolution, the positions

of the metal ions were clearly visible in the electron density. **Table 3.4** reports results from low-resolution refinement of the structure.

| <i>Structure refinement</i> | | | | | |
|-----------------------------|--|---------------------------------------|----------------------------------|-------------------------------|-------------------|
| <i>Resolution</i> | <i>Cell parameters</i> | <i>Space group</i> | <i>Completeness</i> | <i>R factor</i> | <i>R free</i> |
| 6.321 Å | a=180.775, b=182.613, c=67.727 $\alpha=\beta=\gamma=90.000$ | P2 ₁ 2 ₁ 2 | 98.8% | 0.4032 | 0.4326 |
| <i>N° Used reflections</i> | <i>Percentage observed</i> | <i>Percentage of free reflections</i> | <i>Overall weighted R factor</i> | <i>Free weighted R factor</i> | <i>Wavelength</i> |
| 4875 | 99.00 | 5.08 | 0.4132 | 0.4132 | 1.00 Å |

Table 3.4: Structure refinement values.

Unfortunately, the unit cell contains only molecules of Cytochrome O Ubiquinol Oxidase and no further electron density was observed in the independent unit, nor the empty volumes of the structure would allow the introduction of a large molecule such as the Wzy protein.

Interestingly, the unit cell parameters obtained for this structure of the Cytochrome O Ubiquinol Oxidase differ from those reported in PDB for this protein (a=92.1 Å, b=372.5 Å, c=232.7 Å, $\alpha=\beta=\gamma=90^\circ$, space group C222₁), probably due to the different crystallization conditions (Abramson et al. 2000).

Figure 3.29 shows the fitting of the 3D structure in the electron density map and its crystal packing.

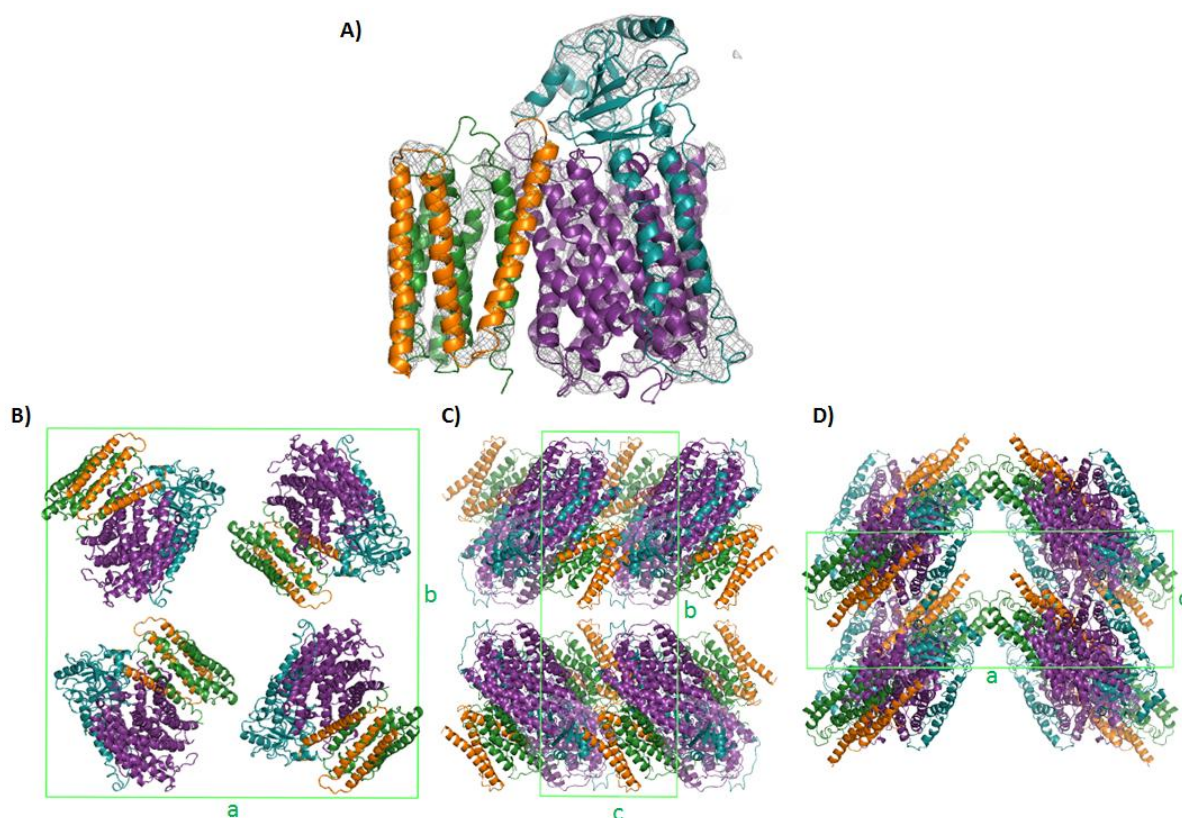


Figure 3.29: Cytochrome O Ubiquinol Oxidase crystal structure at 6 Å resolution. **A)** Fitting of the model in the electron density map. **B-D)** Cell along *a*, *b* and *c* axes. Purple: Subunit 1; Blue: Subunit 2; Green: Subunit 3; Orange: Subunit 4.

In the electron density, the 4 subunits of the Cytochrome O Ubiquinol Oxidase could be recognized. Subunit 1 (**in purple in Figure 3.29**) is the largest (66 kDa), formed by 15 TM helices and short cytoplasmic or periplasmic loops, and is probably responsible for proton translocation activity. This subunit contains a copper ion, stabilized by the interaction with three histidine residues (His284, His333 and His334), and two iron ions chelated by the heme o and heme b prosthetic groups present in the protein. Subunit 2 (**in blue in Figure 3.29**) has a large periplasmic domain interacting with subunits 1, 3 and 4. Subunits 3 and 4 (**in green and orange, respectively**) are smaller and their TM domains interact only with subunit 1.

Particular attention was paid to the analysis of subunit 1, shown in **Figure 3.30**, as this protein has a molecular weight similar to Wzy and, analogously to the polymerase its overall structure is constituted by a bundle of α -helices. In order to confirm the identity of this subunit in the low-resolution structure obtained from the data collection on crystal B, the presence of the metal ions was confirmed by the analysis of the difference electron density

maps obtained after refinement cycles conducted by omitting the metal ions. Strong difference density peaks were observed in the maps, confirming the presence of all expected atoms for a Cytochrome O Ubiquinol Oxidase protein.

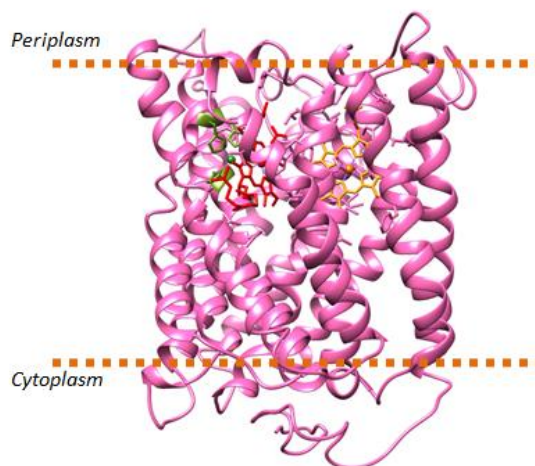


Figure 3.30: The subunit 1 of Cytochrome O Ubiquinol Oxidase.

The presence of the iron ions is also confirmed by the pink/red color of the crystals grown from DM solutions.

Unit cell parameters of crystals grown from LDAO-solubilized protein samples show a significant difference from those obtained for the Cytochrome O Ubiquinol Oxidase structure. In addition, their colorless and cubic crystal habit suggested a different composition, possibly including the Wzy protein. Unfortunately, their smaller crystal size did not allow to collect diffraction data at a resolution sufficient for structure phasing. The presence of the Wzy in these crystals is also consistent with results from CD and fluorescence spectrometry analyses, that indicate that the LDAO-solubilized protein is more stable than that in DM detergent solution. It is possible that DM-solubilized Wzy denatures during the long incubation time of the crystallization experiments, facilitating the crystallization of the Cytochrome O Ubiquinol Oxidase present as a contaminant in the sample. Further crystallization experiments of the LDAO-solubilized protein are ongoing.

3.4.15 Negative staining Electron Microscopy analysis

Negative staining of the sample combined with Electron Microscopy (EM) analysis, allows for the visualization of the proteins as white particles on a black background. The heavy atoms

of the stain (in these experiments, U atoms of the uranyl acetate solution) have a stronger scattering power for electrons compared to the light atoms present in the protein sample (C, H, N, O and, in lower amount, S). When the protein sample deposited on the microscopy grid is covered with the stain solution, the presence of the protein particles reduce the amount of stain deposited and, thus, the sample scattering in the specific grid positions. This technique is suitable for large proteins, as small proteins can be hardly distinguished against the background. The small size of the Wzy recombinant protein (78 kDa) could be an obstacle for the visualization of the protein in negative staining experiments. However, the presence of the larger Cytochrome O Ubiquinol Oxidase contaminant in complex with Wzy may increase the dimension of the particles and allow their recognition. Therefore, negative staining EM analysis was performed in order to confirm the interaction between the two proteins in the solubilized and purified sample. In addition, negative staining EM analysis provides an assessment of the aggregation state of the protein.

Samples obtained after LDAO 10x CMC solubilization, IMAC purification and SEC purification with the Superdex 200 Increase column (second peak) were applied on discharged carbon grids in order to facilitate protein attachment. Uranyl acetate stain was added twice and then completely removed. Dried grids were analyzed with the EM microscope present at the *Institut de minéralogie, de physique des matériaux et de cosmochimie* of the Sorbonne University in Paris, in collaboration with Prof. Catherine Venien-Bryan. Images were taken with a Jeol JEM2100 microscope with a LaB₆ source, equipped with a Gatan 2k×2k camera, at 32500x magnification. **Figure 3.31** shows representative images of the protein sample. Proteins were clearly visible in all grids. Some detergent micelles are present in the sample as larger, whiter and more circular particles. Micelles could be removed by reducing the detergent concentration prior to grid preparation, to a value above the CMC to avoid protein aggregation.

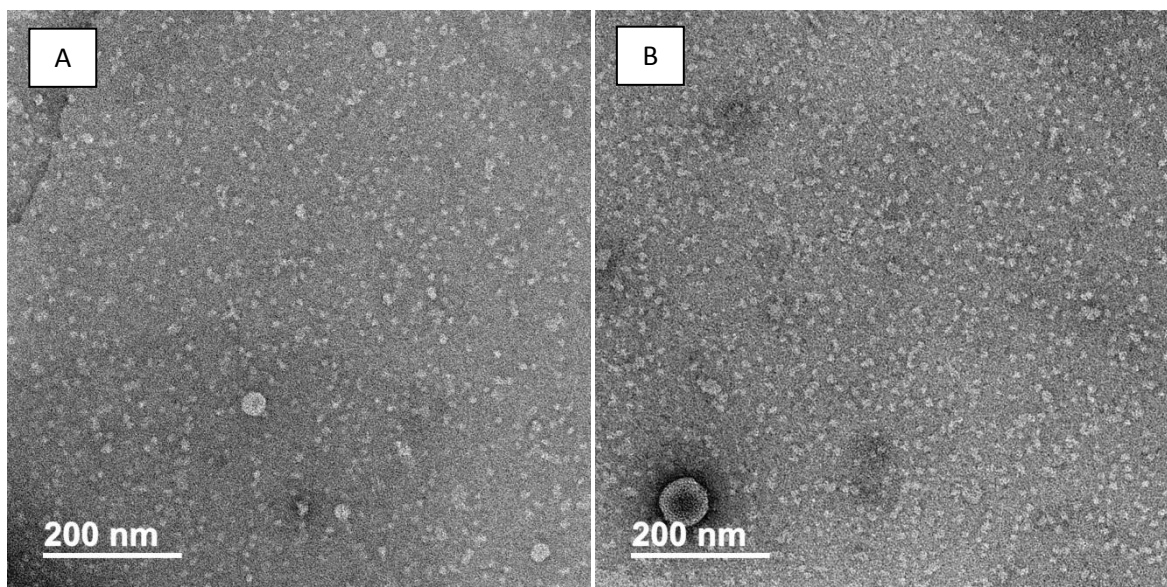


Figure 3.31: Images obtained from the negative staining EM analysis of the Wzy:Cytochrome O Ubiquinol Oxidase sample. TEM was a JEOL JEM2100 LaB6, pictures were recorded using a GATAN 2K × 2K cameras.

The EM analysis confirms the homogeneity of the protein sample and the absence of protein aggregates. The shape of the protein is globular, consistent with the predicted structure of large TM domains embedded in detergent micelles. A rough estimation of the particle dimension leads to a value of 170-200 kDa, compatible with a Wzy-GFP:Cytochrome O Ubiquinol Oxidase complex. Considering the absence of smaller particles, the EM analysis confirms the hypothesis that the recombinant protein from *P. aeruginosa* interacts with the *E. coli* protein, possibly improving Wzy stability. This interaction has to be further investigated to determine whether it forms in physiological conditions.

3.5 CONCLUSIONS AND FUTURE PERSPECTIVES

The polymerase protein Wzy of *P. Aeruginosa* has a key role in the formation of the O-specific antigen chain of the LPS. LPS is involved in bacterium-host interaction, including the activation of the host immune system, and in the formation of the polysaccharide capsule that confers to this pathogen a high resistance against antibiotic therapies. Therefore, my study focused on the characterization of Wzy considering its role as a potential target for a future medical treatment.

In view of the structural characterization of this protein, Wzy was amplified from PAO1 genome, inserted in the pWaldo plasmid, and expressed as a recombinant protein with GFP and 10xHis tag in *E. coli*. After multiple tests, conditions that led to a high yield of Wzy were identified. The protein recovered from bacterial membranes was solubilized in a range of surfactants, including non-ionic and zwitterionic detergents. Results from these tests indicate DM and LDAO as optimal detergents for Wzy solubilization.

The solubilized protein was then purified with different chromatographic methods, including affinity, ion exchange and size exclusion chromatography, but a contaminant was always co-eluted with Wzy recombinant protein, even in different detergent conditions. Interestingly, a highly selective method such as the purification with a nanobody, in this case an anti-GFP antibody, could not separate the contaminant from the protein, suggesting a close association between the two. Mass spectrometry identified the unknown protein as Cytochrome O Ubiquinol Oxidase from *E. coli*, a membrane protein involved in the respiratory chain of the bacterium. A further, harsher purification protocol was successful in separating a fraction of the protein sample containing the Cytochrome O Ubiquinol Oxidase and a fraction free from the contaminant.

Different techniques, i.e. CD analysis, fluorescence spectroscopy and Raman spectroscopy, were used to test the protein stability and the influence of the presence of Cytochrome O Ubiquinol Oxidase. From CD and fluorescence spectroscopy analysis a melting temperature of 55-60° C was obtained for the protein-contaminant sample. With Raman spectroscopy the two samples could be compared, revealing the higher stability of Wzy in presence of the Cytochrome O Ubiquinol Oxidase. Spectroscopic data indicate a role for the contaminant in stabilizing Wzy, but further tests must be conducted to verify if the association between these two proteins is present also in the native conditions of *P. aeruginosa* membranes, possibly involving the closely related homolog of the *E. coli* Cytochrome O Ubiquinol Oxidase.

Crystallization screens were performed on co-purified proteins. Optimization of the initial hits allowed to obtain crystals suitable for diffraction experiments, performed at the XRD2 beamline of the Elettra Synchrotron, yielding 4 datasets at low resolution. However, for one of these crystals a structure at 6 Å resolution could be solved. Unfortunately, electron density maps revealed that this crystal is formed only by the Cytochrome O Ubiquinol Oxidase, without Wzy. However, crystals obtained in LDAO detergent had a different

morphology and color and crystallographic data allowed to determine another set of unit cell parameters, suggesting that they could be formed by the Wzy protein. Further crystallization tests are in progress, in order to obtain crystals containing Wzy alone or in complex with the Cytochrome O Ubiquinol Oxidase

The co-purified protein sample was analyzed with negative staining Electron Microscopy at the Sorbonne University in Paris. Images of the specimen showed a homogeneous sample, suggesting the presence of only one species To evaluate the hypothesis of a tight interaction between Wzy and the Cytochrome O Ubiquinol Oxidase, and to obtain structural data about these proteins, different negative staining EM experiments will be performed, in order to optimize the conditions for further Cryo-EM analysis.

REFERENCES

- 1) Abeyrathne, Priyanka D. et al. 2005. "Functional Characterization of WaaL, a Ligase Associated with Linking O-Antigen Polysaccharide to the Core of *Pseudomonas Aeruginosa* Lipopolysaccharide." *Journal of Bacteriology* 187(9): 3002–12.
- 2) Abramson, Jeff et al. 2000. "The Structure of the Ubiquinol Oxidase from *Escherichia Coli* and Its Ubiquinone Binding Site." *Nature Structural Biology* 7(10): 910–17.
- 3) Alexander, C., and E. Th Rietschel. 2001. "Bacterial Lipopolysaccharides and Innate Immunity." *Journal of Endotoxin Research* 7(3): 167–202.
- 4) Anderson, A. B., and C. R. Robertson. 1995. "Absorption Spectra Indicate Conformational Alteration of Myoglobin Adsorbed on Polydimethylsiloxane." *Biophysical Journal* 68(5): 2091–97.
- 5) Annual report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). 2015. Quantum Field Theory and Gravity *Antimicrobial Resistance Surveillance in Europe Annual*.
- 6) Arsenault, T. L. et al. 1991. "Structural Studies on the Polysaccharide Portion of 'A-Band' Lipopolysaccharide from a Mutant (AK1401) of *Pseudomonas Aeruginosa* Strain PAO1." *Canadian Journal of Chemistry* 69(8): 1273–80.
- 7) Bäckhed, Fredrik et al. 2003. "Structural Requirements for TLR4-Mediated LPS Signalling: A Biological Role for LPS Modifications." *Microbes and Infection* 5(12): 1057–63.
- 8) Batchelor, R. A., G. E. Haraguchi, R. A. Hull, and S. I. Hull. 1991. "Regulation by a Novel Protein of the Bimodal Distribution of Lipopolysaccharide in the Outer Membrane of *Escherichia Coli*." *Journal of Bacteriology* 173(18): 5699–5704.
- 9) Berry, Matthew C., Gayle C. Mcghee, Youfu Zhao, and George W. Sundin. 2009. "Effect of a WaaL Mutation on Lipopolysaccharide Composition, Oxidative Stress Survival, and Virulence in *Erwinia Amylovora*." *FEMS Microbiology Letters* 291(1): 80–87.
- 10) Beutler, B., and A. Cerami. 1988. "Tumor Necrosis, Cachexia, Shock, and Inflammation: A Common Mediator." *Annual Review of Biochemistry* 57: 505–18.
- 11) Bhat, R., A. Marx, C. Galanos, and R. S. Conrad. 1990. "Structural Studies of Lipid A from *Pseudomonas Aeruginosa* PAO1: Occurrence of 4-Amino-4-Deoxyarabinose." *Journal of Bacteriology* 172(12): 6631–36.
- 12) Bollati, Michela et al. 2015. "Crystal Structure of LptH, the Periplasmic Component of the Lipopolysaccharide Transport Machinery from *Pseudomonas Aeruginosa*." *FEBS Journal* 282(10): 1980–97.
- 13) Bray, D., and P. W. Robbins. 1967. "The Direction of Chain Growth in *Salmonella Anatum* O-Antigen Biosynthesis." *Biochemical and Biophysical Research Communications* 28(3): 334–39.
- 14) Bystrova, Olga V. et al. 2006. "Structures of the Core Oligosaccharide and O-Units in the R- and SR-Type Lipopolysaccharides of Reference Strains of *Pseudomonas Aeruginosa* O-Serogroups." *FEMS Immunology and Medical Microbiology* 46(1): 85–99.
- 15) Chang, G., and C. B. Roth. 2001. "Structure of MsbA from *E. Coli*: A Homolog of the Multidrug

Chapter 3

- Resistance ATP Binding Cassette (ABC) Transporters." *Science* 293(5536): 1793–1800.
- 16) Chevalier, Sylvie et al. 2017. "Structure, Function and Regulation of Pseudomonas Aeruginosa Porins." *FEMS Microbiology Reviews* 41(5): 698–722.
 - 17) Chitkara, Y. K., and T. C. Feierabend. 1981. "Endogenous and Exogenous Infection with Pseudomonas Aeruginosa in a Burns Unit." *International Surgery* 66(3): 237–40.
 - 18) Chng, Shu Sin et al. 2010. "Characterization of the Two-Protein Complex in Escherichia Coli Responsible for Lipopolysaccharide Assembly at the Outer Membrane." *Proceedings of the National Academy of Sciences of the United States of America* 107(12): 5363–68.
 - 19) Collins, Richard F. et al. 2017. "Full-Length, Oligomeric Structure of Wzz Determined by Cryoelectron Microscopy Reveals Insights into Membrane-Bound States." *Structure* 25(5): 806-815.e3.
 - 20) Comolli, James C., Leslie L. Waite, Keith E. Mostov, and Joanne N. Engel. 1999. "Pili Binding to Asialo-GM1 on Epithelial Cells Can Mediate Cytotoxicity or Bacterial Internalization by Pseudomonas Aeruginosa." *Infection and Immunity* 67(7): 3207–14.
 - 21) Cuthbertson, L., V. Kos, and C. Whitfield. 2010. "ABC Transporters Involved in Export of Cell Surface Glycoconjugates." *Microbiology and Molecular Biology Reviews* 74(3): 341–62.
 - 22) Cuthbertson, Leslie, Jacqueline Powers, and Chris Whitfield. 2005. "The C-Terminal Domain of the Nucleotide-Binding Domain Protein Wzt Determines Substrate Specificity in the ATP-Binding Cassette Transporter for the Lipopolysaccharide O-Antigens in Escherichia Coli Serotypes O8 and O9a." *Journal of Biological Chemistry* 280(34): 30310–19.
 - 23) Daniels, Craig, Corrie Griffiths, Bryony Cowles, and Joseph S. Lam. 2002. "Pseudomonas Aeruginosa O-Antigen Chain Length Is Determined before Ligation to Lipid A Core." *Environmental Microbiology* 4(12): 883–97.
 - 24) Dasgupta, T. et al. 1994. "Characterization of Lipopolysaccharide-Deficient Mutants of Pseudomonas Aeruginosa Derived from Serotypes O3, O5, and O6." *Infection and Immunity* 62(3): 809–17.
 - 25) Dinarello, C. A. 1991. "Interleukin-1 and Interleukin-1 Antagonism." *Blood* 77: 1627–52.
 - 26) Doerrler, William T., Henry S. Gibbons, R. Christian, and H. Raetz. 2004. "MsbA-Dependent Translocation of Lipids across the Inner Membrane of Escherichia Coli." *Journal of Biological Chemistry* 279(43): 45102–9.
 - 27) Ernst, Robert K. et al. 1999. "Specific Lipopolysaccharide Found in Cystic Fibrosis Airway Pseudomonas Aeruginosa." *Science* 286(5444): 1561–65.
 - 28) Ernst, Robert K. et al. 2003. "Pseudomonas Aeruginosa Lipid A Diversity and Its Recognition by Toll-like Receptor 4." *Journal of Endotoxin Research* 9(6): 395–400.
 - 29) Ernst, Robert K. et al. 2006. "The Pseudomonas Aeruginosa Lipid A Deacylase: Selection for Expression and Loss within the Cystic Fibrosis Airway." *Journal of Bacteriology* 188(1): 191–201.
 - 30) Fleiszig, S. M.J., T. S. Zaidi, and G. B. Pier. 1995. "Pseudomonas Aeruginosa Invasion of and Multiplication within Corneal Epithelial Cells in Vitro." *Infection and Immunity* 63(10): 4072–77.
 - 31) Fleiszig, S. M.J., T. S. Zaidi, R. Ramphal, and G. B. Pier. 1994. "Modulation of Pseudomonas Aeruginosa Adherence to the Corneal Surface by Mucus." *Infection and Immunity* 62(5): 1799–1804.
 - 32) Galanos, Chris et al. 1985. "Synthetic and Natural Escherichia Coli Free Lipid A Express Identical

Chapter 3

- Endotoxic Activities." *European Journal of Biochemistry* 148(1): 1–5.
- 33) Galentine, Paul G. et al. 1984. "Corneal Ulcers Associated with Contact Lens Wear." *Archives of Ophthalmology* 102(6): 891–94.
- 34) Ghanei, Hamed, Priyanka D. Abeyrathne, and Joseph S. Lam. 2007. "Biochemical Characterization of MsbA from *Pseudomonas Aeruginosa*." *Journal of Biological Chemistry* 282(37): 26939–47.
- 35) Hancock, R. E W et al. 1983. "Pseudomonas Aeruginosa Isolates from Patients with Cystic Fibrosis: A Class of Serum-Sensitive, Nontypable Strains Deficient in Lipopolysaccharide O Side Chains." *Infection and Immunity* 42(1): 170–77.
- 36) Hirayama, Hiroshi et al. 2013. "ATPase Activity of Human ABCG1 Is Stimulated by Cholesterol and Sphingomyelin." *Journal of Lipid Research* 54(2): 496–502.
- 37) Hong, Yaoqin et al. 2015. "Three Wzy Polymerases Are Specific for Particular Forms of an Internal Linkage in Otherwise Identical O Units." *Microbiology (United Kingdom)* 161(8): 1639–47.
- 38) Islam, Salim T. et al. 2011. "Dual Conserved Periplasmic Loops Possess Essential Charge Characteristics That Support a Catch-and-Release Mechanism of O-Antigen Polymerization by Wzy in *Pseudomonas Aeruginosa* PAO1." *Journal of Biological Chemistry* 286(23): 20600–605.
- 39) Islam, Salim T., Steven M. Huszczyński, et al. 2013. "Conserved-Residue Mutations in Wzy Affect O-Antigen Polymerization and Wzz-Mediated Chain-Length Regulation in *Pseudomonas Aeruginosa* PAO1." *Scientific Reports* 3(3441): 1–9.
- 40) Islam, Salim T., and Joseph S. Lam. 2014. "Synthesis of Bacterial Polysaccharides via the Wzx/Wzy-Dependent Pathway." *Canadian Journal of Microbiology* 60(11): 697–716.
- 41) Islam, Salim T, Paul D W Eckford, et al. 2013. "Proton-Dependent Gating and Proton Uptake by Wzx Support O-Antigen-Subunit Antiport Across the Bacterial Inner Membrane." *moBio* 4(5): 1–8.
- 42) Islam, Salim T, Véronique L Taylor, Meng Qi, and Joseph S Lam. 2010. "Membrane Topology Mapping of the O-Antigen Flippase (Wzx), Polymerase (Wzy), and Ligase (WaaL) from *Pseudomonas Aeruginosa* PAO1 Reveals Novel Domain Architectures." *mBio* 1(3): 1–10.
- 43) Kadurugamuwa, J. L., J. S. Lam, and T. J. Beveridge. 1993. "Interaction of Gentamicin with the A Band and B Band Lipopolysaccharides of *Pseudomonas Aeruginosa* and Its Possible Lethal Effect." *Antimicrobial Agents and Chemotherapy* 37(4): 715–21.
- 44) Kalynych, Sergei, Miguel A. Valvano, and Mirosław Cygler. 2012. "Polysaccharide Co-Polymerases: The Enigmatic Conductors of the O-Antigen Assembly Orchestra." *Protein Engineering, Design and Selection* 25(11): 797–802.
- 45) Karnaukhova, Elena et al. 2014. "Characterization of Heme Binding to Recombinant A1-Microglobulin." *Frontiers in Physiology* 5(Nov): 1–11.
- 46) King, Jerry D., Dana Kocíncová, Erin L. Westman, and Joseph S. Lam. 2009. "Lipopolysaccharide Biosynthesis in *Pseudomonas Aeruginosa*." *Innate Immunity* 15(5): 261–312.
- 47) Kintz, Erica N., and Joanna B. Goldberg. 2011. "Site-Directed Mutagenesis Reveals Key Residue for O-Antigen Chain Length Regulation and Protein Stability in *Pseudomonas Aeruginosa* Wzz." *Journal of Biological Chemistry* 286(51): 44277–84.
- 48) Knirel, Yuriy A. et al. 2006. "Review: Conserved and Variable Structural Features in the

Chapter 3

- Lipopolysaccharide of *Pseudomonas Aeruginosa*." *Journal of Endotoxin Research* 12(6): 324–36.
- 49) Kropinski, A. M.B., V. Lewis, and D. Berry. 1987. "Effect of Growth Temperature on the Lipids, Outer Membrane Proteins, and Lipopolysaccharides of *Pseudomonas Aeruginosa* PAO." *Journal of Bacteriology* 169(5): 1960–66.
- 50) Kulshin, Vladimir A. et al. 1991. "Structural Characterization of the Lipid A Component of *Pseudomonas Aeruginosa* Wild-type and Rough Mutant Lipopolysaccharides." *European Journal of Biochemistry* 198(3): 697–704.
- 51) Lam, J. S. et al. 1992. "Ultrastructural Examination of the Lipopolysaccharides of *Pseudomonas Aeruginosa* Strains and Their Isogenic Rough Mutants by Freeze-Substitution." *Journal of Bacteriology* 174(22): 7159–67.
- 52) Lam, M. Y C et al. 1989. "Occurrence of a Common Lipopolysaccharide Antigen in Standard and Clinical Strains of *Pseudomonas Aeruginosa*." *Journal of Clinical Microbiology* 27(5): 962–67.
- 53) Liu, P. V., and S. Wang. 1990. "Three New Major Somatic Antigens of *Pseudomonas Aeruginosa*." *Journal of Clinical Microbiology* 28(5): 922–25.
- 54) Lyczak, Jeffrey B., Carolyn L. Cannon, and Gerald B. Pier. 2000. "Establishment of *Pseudomonas Aeruginosa* Infection: Lessons from a Versatile Opportunist." *Microbes and Infection* 2(9): 1051–60.
- 55) Magill, Shelley S et al. 2014. "Multistate Point-Prevalence Survey of Health Care–Associated Infections for the Emerging Infections Program Healthcare-Associated Infections and Antimicrobial Use Prevalence Survey Team * Centers for Disease Control and Prevention (. " *St. Paul (R.L.); Connecticut Department of Public Health Oakland (J.N. Decatur (S.M.R 370(13): 1198–1208.*
- 56) Masinick, Sharon A., Christopher P. Montgomery, Paul C. Montgomery, and Linda D. Hazlett. 1997. "Secretory IgA Inhibits *Pseudomonas Aeruginosa* Binding to Cornea and Protects against Keratitis." *Investigative Ophthalmology and Visual Science* 38(5): 910–18.
- 57) McGroarty, E. J., and M. Rivera. 1990. "Growth-Dependent Alterations in Production of Serotype-Specific and Common Antigen Lipopolysaccharides in *Pseudomonas Aeruginosa* PAO1." *Infection and Immunity* 58(4): 1030–37.
- 58) Merino, Susana, Victor Gonzalez, and Juan M. Tomás. 2016. "The First Sugar of the Repeat Units Is Essential for the Wzy Polymerase Activity and Elongation of the O-Antigen Lipopolysaccharide." *Future Microbiology* 11(7): 903–18.
- 59) Mulcahy, Lawrence R., Vincent M. Isabella, and Kim Lewis. 2014. "*Pseudomonas Aeruginosa* Biofilms in Disease." *Microbial Ecology* 68(1): 1–12.
- 60) Narita, Shin ichiro, and Hajime Tokuda. 2009. "Biochemical Characterization of an ABC Transporter LptBFGC Complex Required for the Outer Membrane Sorting of Lipopolysaccharides." *FEBS Letters* 583(13): 2160–64.
- 61) Nath, Pratiti, and Renato Morona. 2015. "Detection of Wzy/Wzz Interaction in *Shigella Flexneri*." *Microbiology (United Kingdom)* 161(9): 1797–1805.
- 62) Novy, Robert, and Barbarra Morris. 2001. "Use of Glucose to Control Basal Expression in the PET System." *inNovations* 13(1): 13–15.
- 63) Oladepo, Sulayman A. et al. 2012. "UV Resonance Raman Investigations of Peptide and Protein

Chapter 3

- Structure and Dynamics." *Chemical Reviews* 112(5): 2604–28.
- 64) Pier, Gerald B., Martha Grout, and Tanweer S. Zaidi. 1997. "Cystic Fibrosis Transmembrane Conductance Regulator Is an Epithelial Cell Receptor for Clearance of *Pseudomonas Aeruginosa* from the Lung." *Proceedings of the National Academy of Sciences of the United States of America* 94(22): 12088–93.
- 65) Polissi, Alessandra, and Paola Sperandio. 2014. "The Lipopolysaccharide Export Pathway in *Escherichia Coli*: Structure, Organization and Regulated Assembly of the Lpt Machinery." *Marine Drugs* 12(2): 1023–42.
- 66) Pupo, Elder, and Eugenio Hardy. 2009. "Complexity and Solutions to the Isolation Problem of Gram Negative Lipopolysaccharides' Bacteria Molecular Species." *Biotechnologia Aplicada* 26(1): 1–15.
- 67) Raetz, Christian R. H., and Chris Whitfield. 2002. "Lipopolysaccharide Endotoxins." *Annual Review of Biochemistry* 71(1): 635–700.
- 68) Rhee, Sang Hoon. 2014. "Lipopolysaccharide: Basic Biochemistry, Intracellular Signaling, and Physiological Impacts in the Gut." *Intestinal Research* 12(2): 90–95.
- 69) Rocchetta, H. L., L. L. Burrows, and AJ. S. Lam. 1999. "Genetics of O-Antigen Biosynthesis in *Pseudomonas Aeruginosa*." *Microbiology and molecular biology reviews* 63(3): 523–53.
- 70) Rocchetta, Heather L., Lori L. Burrows, Jennifer C. Pacan, and Joseph S. Lam. 1998. "Three Rhamnosyltransferases Responsible for Assembly of the A-Band D- Rhamnan Polysaccharide in *Pseudomonas Aeruginosa*: A Fourth Transferase, WbpL, Is Required for the Initiation of Both A-Band and B-Band Lipopolysaccharide Synthesis." *Molecular Microbiology* 28(6): 1103–19.
- 71) Ruan, Xiang et al. 2012. "The WaaL O-Antigen Lipopolysaccharide Ligase Has Features in Common with Metal Ion-Independent Inverting Glycosyltransferases." *Glycobiology* 22(2): 288–99.
- 72) Ruffin, Manon, and Emmanuelle Brochiero. 2019. "Repair Process Impairment by *Pseudomonas Aeruginosa* in Epithelial Tissues: Major Features and Potential Therapeutic Avenues." *Frontiers in Cellular and Infection Microbiology* 9(MAY): 1–18.
- 73) Ruiz-Garbajosa, P, and R Cantón. 2017. "Epidemiology of Antibiotic Resistance in *Pseudomonas Aeruginosa*. Implications for Empiric and Definitive Therapy." *Revista espanola de quimioterapia : publicacion oficial de la Sociedad Espanola de Quimioterapia* 30 Suppl 1: 8–12.
- 74) Ruiz, Natividad, Daniel Kahne, and Thomas J Silhavy. 2009. "Road of Discovery." *Nature Reviews Microbiology* 7(9): 677–83.
- 75) Sachdeva, Shivangi, Raghuvamsi V. Palur, Karpagam U. Sudhakar, and Thenmalarchelvi Rathinavelan. 2017. "E. Coli Group 1 Capsular Polysaccharide Exportation Nanomachinery as a Plausible Antivirulence Target in the Perspective of Emerging Antimicrobial Resistance." *Frontiers in Microbiology* 8(Article 70): 1–19.
- 76) Sadikot, Ruxana T., Timothy S. Blackwell, John W. Christman, and Alice S. Prince. 2005. "Pathogen-Host Interactions in *Pseudomonas Aeruginosa* Pneumonia." *American Journal of Respiratory and Critical Care Medicine* 171(11): 1209–23.
- 77) Sadovskaya, Irina et al. 1998. "Structural Elucidation of the Lipopolysaccharide Core Regions of the Wild-Type Strain PAO1 and O-Chain-Deficient Mutant Strains AK1401 and AK1012 from *Pseudomonas*

Chapter 3

- Aeruginosa Serotype O5." *European Journal of Biochemistry* 255(3): 673–84.
- 78) Sato, Hisaaki, Kota Okinaga, and Hiroshi Saito. 1988. "Role of Pili in the Pathogenesis of *Pseudomonas Aeruginosa* Burn Infection." *Microbiology and Immunology* 32(2): 131–39.
- 79) Schlegel, Susan et al. 2012. "Optimizing Membrane Protein Overexpression in the *Escherichia Coli* Strain Lemo21(DE3)." *Journal of Molecular Biology* 423(4): 648–59.
- 80) Sperandio, Paola et al. 2007. "Characterization of LptA and LptB, Two Essential Genes Implicated in Lipopolysaccharide Transport to the Outer Membrane of *Escherichia Coli*." *Journal of Bacteriology* 189(1): 244–53.
- 81) Stanislavsky, Eugene S., and Joseph S. Lam. 1997. "Pseudomonas Aeruginosa Antigens as Potential Vaccines." *FEMS Microbiology Reviews* 21(3): 243–77.
- 82) Stefani, S. et al. 2017. "Relevance of Multidrug-Resistant *Pseudomonas Aeruginosa* Infections in Cystic Fibrosis." *International Journal of Medical Microbiology* 307(6): 353–62.
- 83) Subedi, Dinesh, Ajay Kumar Vijay, and Mark Willcox. 2017. "Overview of Mechanisms of Antibiotic Resistance in *Pseudomonas Aeruginosa*: An Ocular Perspective." *Clinical and Experimental Optometry* 101(2): 162–71.
- 84) Tabor, Stanley. 2001. "Expression Using the T7 RNA Polymerase/Promoter System." *Current Protocols in Molecular Biology* Chapter 16(Unit 16.2).
- 85) Taylor, Véronique L. et al. 2016. "A Bacteriophage-Acquired O-Antigen Polymerase (Wzy β) from *P. Aeruginosa* Serotype O16 Performs a Varied Mechanism Compared to Its Cognate Wzy α ." *Frontiers in Microbiology* 7(MAR): 1–15.
- 86) Trent, M. Stephen, Wendy Pabich, Christian R.H. Raetz, and Samuel I. Miller. 2001. "A PhoP/PhoQ-Induced Lipase (PagL) That Catalyzes 3-O-Deacylation of Lipid A Precursors in Membranes of *Salmonella Typhimurium*." *Journal of Biological Chemistry* 276(12): 9083–92.
- 87) Villavicencio, Raphael T. 1998. "The History of Blue Pus." *Journal of the American College of Surgeons* 187(2): 212–16.
- 88) Walsh, Andrew G. et al. 2000. "Lipopolysaccharide Core Phosphates Are Required for Viability and Intrinsic Drug Resistance in *Pseudomonas Aeruginosa*." *Molecular Microbiology* 35(4): 718–27.
- 89) Wei, Yang, Aby A. Thyparambil, and Robert A. Latour. 2014. "Protein Helical Structure Determination Using CD Spectroscopy for Solutions with Strong Background Absorbance from 190–230 Nm." *Biochimica et Biophysica Acta* 1844(12): 2331–37.
- 90) Woodward, Robert et al. 2010. "In Vitro Bacterial Polysaccharide Biosynthesis: Defining the Functions of Wzy and Wzz." *nature chemical biology* 6(6): 418–23.
- 91) Wyckoff, Timna J.O. et al. 1998. "Hydrocarbon Rulers in UDP-N-Acetylglucosamine Acyltransferases." *Journal of Biological Chemistry* 273(49): 32569–32372.
- 92) Yokota, Shin-ichi et al. 1987. "Characterization of a Polysaccharide Component of Lipopolysaccharide from *Pseudomonas Aeruginosa* IID 1008 (ATCC 27584) as D-rhamnan." *European Journal of Biochemistry* 167(2): 203–9.
- 93) Zaidi, Tanweer S., Jeffrey Lyczak, Michael Preston, and Gerald B. Pier. 1999. "Cystic Fibrosis Transmembrane Conductance Regulator-Mediated Corneal Epithelial Cell Ingestion of *Pseudomonas*

Chapter 3

- Aeruginosa* Is a Key Component in the Pathogenesis of Experimental Murine Keratitis." *Infection and Immunity* 67(3): 1481–92.
- 94) Zhang, Ge, Timothy C. Meredith, and Daniel Kahne. 2013. "On the Essentiality of Lipopolysaccharide to Gram-Negative Bacteria." *Current Opinion in Microbiology* 16(6): 779–85.
- 95) Zhao, F. et al. 2015. "Heterologous Production of *Pseudomonas Aeruginosa* Rhamnolipid under Anaerobic Conditions for Microbial Enhanced Oil Recovery." *Journal of Applied Microbiology* 118(2): 379–89.

Chapter 4

Materials and methods

4.1 CLONING

Molecular cloning techniques allow for isolation and amplification of genes that code for target proteins and their further insertion into vectors for propagation.

In order to obtain the amount of protein required for activity assays or structural analysis, the target protein is usually cloned in a vector that contains an adequate system for protein expression at high yields and expressed as recombinant protein (Hartley 2006).

Vectors for overexpression are circular DNA sequences, called plasmids, that carry all the necessary information for their replication, for transcription of the protein and for specific resistances (e.g. to antibiotics) that allow the selection of the host.

In our study we used the yeast *S. cerevisiae* to overexpress P-IV ATPases and *E. coli* as bacterial host for the membrane protein Wzy.

4.1.1 Vectors

- *P-IV ATPase expression in S. cerevisiae*

For P-IV ATPases cloning and expression, the pDDGFP_LEU2D plasmid was chosen for its useful characteristics (D. Drew et al. 2008). pDDGFP_LEU2D (**Figure 4.1**) allows for the selection of auxotrophic transformed colonies through their ability to survive in media lacking essential nutrients (leucine or uracil). In this plasmid, the gene of the target protein is under control of the strong inducible GAL1 promoter as described in Neo1 Results and Discussion. Additionally, this plasmid carries also a selective gene for bacterial hosts that confers ampicillin resistance. The gene is inserted after the GAL1 promoter in the multiple cloning site in which sequences for cleavage by endonuclease restriction enzymes (i.e. SmaI) are present. The protein construct obtained by overexpression with this plasmid is a chimera

with GFP-8 histidine tag (GFP-8H-tag) and a cleavage site for Tobacco Etch Virus protease (TEV). The GFP helps in detection of the protein, while the His-tag allows for its purification with chromatographic methods.

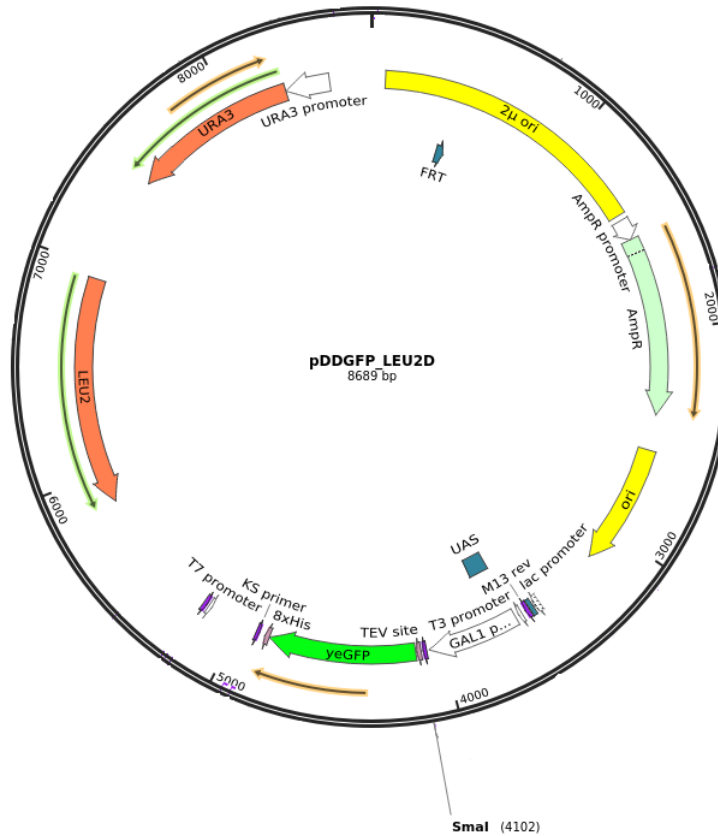


Figure 4.1: pDDGFP_LEU2D plasmid from Addgene.

- *Wzy* expression in *E. Coli*

The sequence coding for the Wzy protein of the PAO1 strain of *P. aeruginosa* (Wzy_{PA}) was inserted in the pWaldo_GFP vector. pWALDO_GFP plasmid possesses the following features: a T7 RNA polymerase promoter (whose role was described in detail in Wzy Results and Discussion), a gene for kanamycin resistance and a GFP-His-tag, to be expressed at the C-terminus of the target protein. The pWaldo_GFP original construct from Addgene (**Figure 4.2**) was previously mutated to increase the size of the His-tag from 8 histidine residues to 10, by insertional mutagenesis. As in the case of P-IV ATPases, the GFP tag allows for detection of the protein and the His-tag for purification of the recombinant protein.

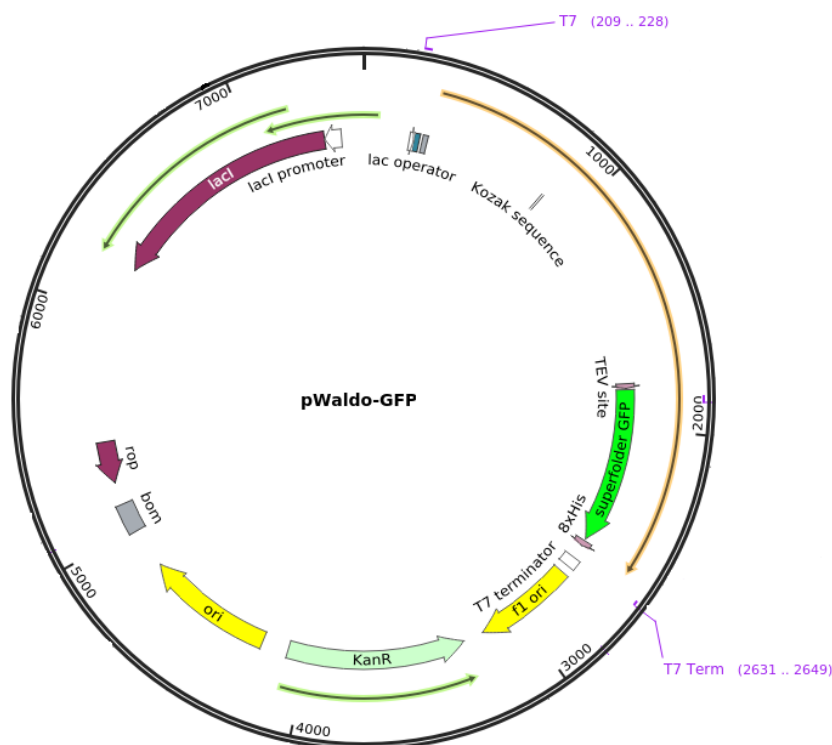


Figure 4.2: pWaldo_GFP plasmid from Addgene.

4.1.2 Polymerase Chain Reaction

The genes coding for the target proteins, P-IV ATPases and Wzy_{PA}, were amplified from the genomes of *S. cerevisiae* and *P. aeruginosa* PAO1, respectively, via Polymerase Chain Reaction (PCR) and inserted into the plasmids described above. PCR is a molecular biology technique that allows amplification of a specific DNA fragment. The reaction is based on the use of a thermostable DNA polymerase that synthesizes new DNA sequences in the 5'→3' direction starting from a template. The DNA polymerase recognizes short fragments of single strand nucleic acids, known as oligonucleotides or primers, generally around 20-30 nucleotides in length, designed to be complementary to the target sequence on the template. Usually two primers are used in the reaction: the forward (For), complementary to the 3' end of the target sequence on one DNA strand, and the reverse (Rev), complementary to the 3' end of the sequence on the other strand. It is important for the primers to possess a balance between guanine and cytosine residues and a low tendency to form stable secondary structures (Green and Sambrook 2018). Additionally, deoxynucleotides

triphosphates (dNTPs) are required for the DNA polymerase to synthesize the new DNA sequence.

The reaction solution undergoes several cycles of thermal treatment in a thermocycler, allowing for sequence amplification. Every cycle is divided in steps (**Figure 4.3**): i) during denaturation, the solution is heated to 95-98°C in order to separate the double stranded DNA; ii) in the annealing phase, primers pair to the target DNA sequences according to their melting point temperature (45-65°C); iii) during elongation, the temperature is raised to 72°C to allow identification of the primers by the DNA polymerase and subsequent addition of nucleotides to a new developing DNA strand. New synthesized fragments then pair to a complementary sequence to re-create double stranded DNA. At each cycle the number of target DNA fragments is doubled. A final elongation step at 72°C is required to ensure full elongation of all strands. DNA absorbs at 260 nm, therefore for its quantification UV methods can be used.

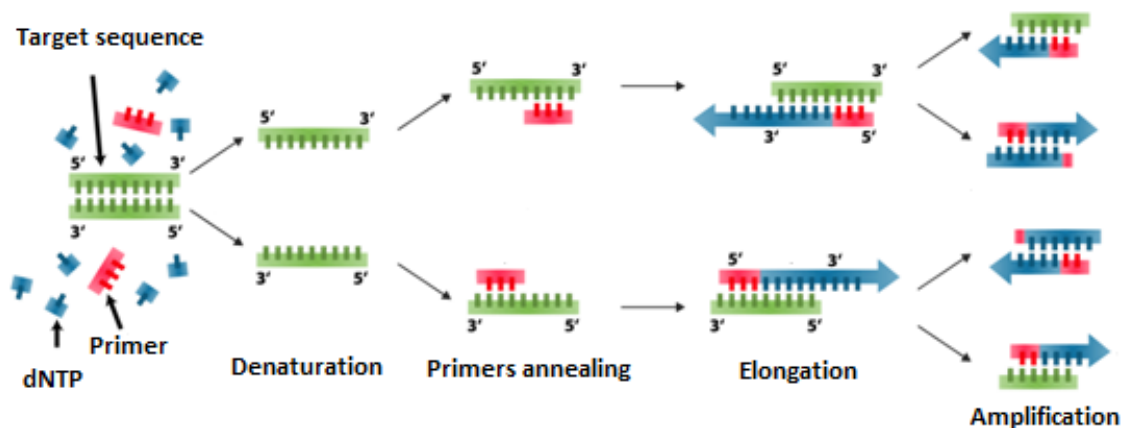


Figure 4.3: PCR steps. In the denaturation step the DNA strands are separated; in the annealing step primers anneal with the single stranded DNA; in the elongation step the DNA polymerase adds nucleotides to a new developing DNA strand. From www.sciencemusicvideos.com.

The amplified sequence was checked with DNA electrophoresis on agarose gel.

DNA electrophoresis on gel is an easy technique to visualize and analyze a DNA sample. DNA is negatively charged and therefore migrates towards the positively charged anode when subjected to an electric field. Agarose is a polymer of agarabiose, a sugar composed of L- and D-galactose, non-covalently associated to form a matrix in which DNA fragments migrate according to their size. A DNA-binding dye is added to the gel to visualize the

fragments. The dye intercalates between the DNA strands, allowing for their visualization under UV light (P. Y. Lee et al. 2012).

The amplified sequences coding for the target proteins were inserted in pDDGFP_LEU2D via homologous recombination or in pWALDO_GFP via RF cloning.

- **Homologous recombination in yeast:** Homologous recombination is possible only in eukaryotic organisms and exploits host mechanisms to insert the target gene into the plasmid, previously linearized with a restriction enzyme (in pDDGFP_LEU2D case, SmaI enzyme) (D. Drew et al. 2008). This technique avoids the use of the classical method of double digestion.
- **Restriction-Free cloning in *E. coli*:** Restriction-Free (RF) cloning is a method to insert large DNA fragments in a DNA vector using two PCR reactions (**Figure 4.4**). The sequence of interest is amplified by the first PCR, using primers designed to be complementary to both the target sequence and the destination plasmid. The large DNA fragment produced in the first PCR is purified from agarose gel with commercial kits, and then used as a “megaprimer” in the second PCR for the insertion into the plasmid. Unlike the normal PCR, RF cloning is not an amplification of the target DNA sequence, therefore the final DNA yield is usually low.

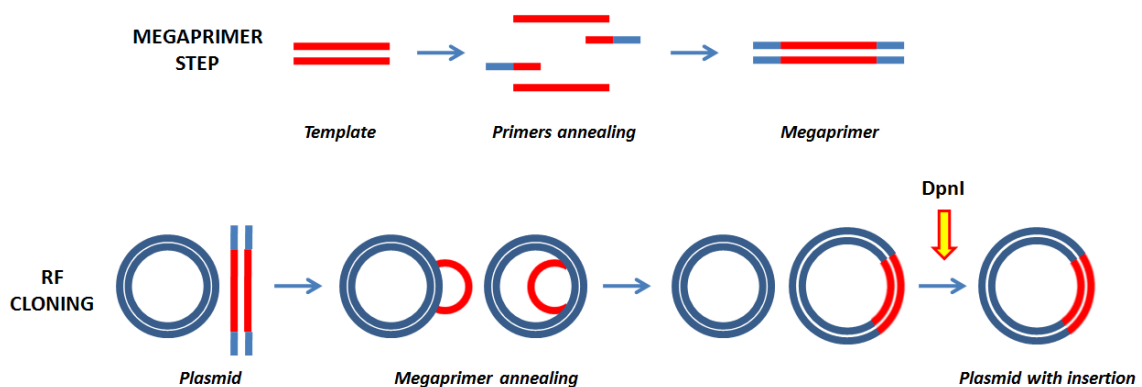


Figure 4.4: Workflow of the RF cloning. The gene is isolated and amplified from a template with 5' and 3' ends complementary to the plasmid, for the creation of the megaprimer. The megaprimer is, then, inserted in the plasmid. Image is adapted from Mathieu J. et al. (2014), *MethodsX* (J. Mathieu, Alvarez, and Alvarez 2014)

The PCR product is finally digested with the enzyme DpnI that is able to eliminate all the methylated parental plasmid that does not contain the new insertion (Van Den Ent and Löwe 2006).

4.1.3 *Plasmid transformation into E. coli or S. cerevisiae*

The introduction of the plasmid into the host is referred to as transformation. Specific manipulation is required to induce plasmid crossing of the host membranes. This technique allows for production of transformed cells.

For the yeast *S. cerevisiae* one of the most used techniques is the Lithium-acetate method. This technique involves an incubation of the *S. cerevisiae* cells with lithium acetate, polyethylene glycol and single stranded DNA from salmon sperm, while the cells are in their mid-log phase, i.e. at their productivity apex. Once the cells are competent, a heat shock at 42°C allows plasmid access (Kawai, Hashimoto, and Murata 2010).

For *E. coli*, a widely used technique is the rubidium (or calcium) chloride method. This method involves two steps: adsorption and binding of plasmid DNA to the cellular membrane of the host at 0°C in presence of high concentrations of cations, and a short heat shock at 42°C to allow DNA penetration through the membrane (Roychoudhury, Basu, and Sengupta 2009). After heat shock the cells are grown in a not selective broth for a short period and then seeded on selective agarose plates.

4.1.4 *Selection of positive colonies*

For the selection of yeast transformed colonies with plasmids containing the P-IV ATPase genes, the autotrophic information for uracil or leucine production carried by the pDDGFP_LEU2D plasmid was exploited. When transformed in auxotrophic yeast strains for these two nutrients, only the transformed cells are able to survive and grow in a medium lacking these molecules (Pronk 2002).

For selection of *E. coli* colonies transformed with the Wzy-containing plasmid, the information for the production of the kanamycin resistance protein carried by the pWaldo_GFP plasmid was used, allowing transformed cells to survive in media supplemented with the antibiotic kanamycin.

After selection, positive colonies are further grown in liquid medium, the plasmid is then extracted by cell lysis and isolated from different DNA fragments present in these cells with commercial kits. The extracted plasmid is checked via PCR (control PCR) and positive plasmids are sequenced for further confirmation of the correct gene insertion.

4.1.5 Site-directed mutagenesis

Site-directed mutagenesis is widely used to create insertions, deletions or small changes in selected sequences. For single-residue mutations the method is based on a standard PCR amplification using primers carrying the mutation of interest. The primers can be designed so as to be completely complementary to each other (overlapping), carrying the mutation in their middle, or not overlapping (back to back) carrying the mutation only on one of the primers (**Figure 4.5**). As for RF cloning, parental plasmid is digested with the enzyme DpnI to eliminate eventual non-mutated plasmid (Castorena-Torres, Peñuelas-Urquides, and Bermúdez de León 2016). The plasmid is then transformed in competent cells, extracted and sent for sequencing.

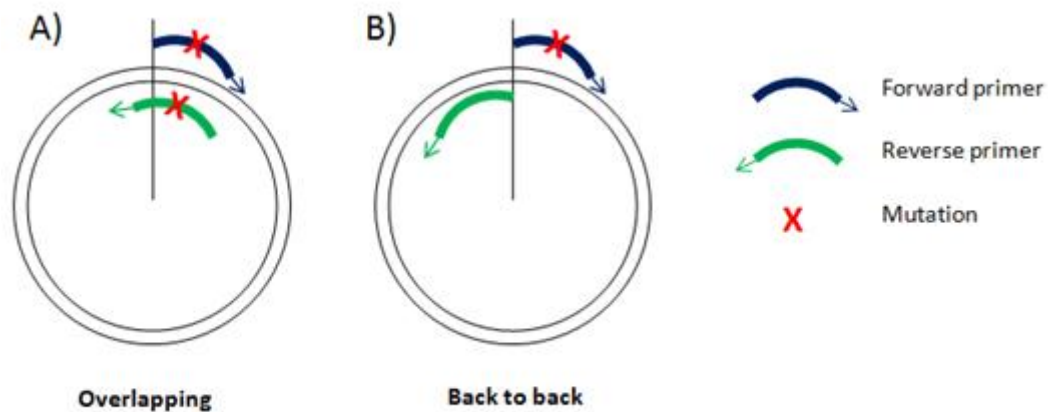


Figure 4.5: Site-directed mutagenesis primers. **A)** Overlapping primers with mutation on both primers; **B)** Back to back primers with mutation only on the forward primer.

4.2 MEMBRANE PROTEIN OVEREXPRESSION AND PURIFICATION

The plasmid carrying the target gene is used to transform the host selected for overexpression of the protein. Initially a small-scale expression test is performed to select the best condition that leads to the highest protein yield. Then, a scale-up allows to obtain high amounts of protein. Cells are then lysed and membranes with target proteins are recovered through ultracentrifugation.

The identification of the best detergent for membrane solubilization is one of the most critical steps in membrane protein purification protocols. Once the optimal detergent is selected, a purification protocol with suitable chromatographic techniques is performed in order to obtain pure protein sample.

4.2.1 *Small-scale expression*

The expression and purification of functional membrane proteins is not an easy task due to their possible toxic effects for the host, that decrease protein yields. For this reason, a tight control of expression allows to maximize protein production and reduce toxicity. In this preliminary step, parameters as temperature, host cell strain, and induction time are explored.

The expression induction follows different routes in yeast and bacteria.

- ***Yeast S. cerevisiae***: The gene coding for the protein of interest in the pDDGFP_LEU2D plasmid is located downstream the strong GAL1 promoter. As already discussed in Neo1 Results and Discussion the GAL1 promoter allows for protein transcription only in presence of galactose, while is inhibited in presence of glucose (J. L. Parker and Newstead 2014). Small-scale overexpression of P-IV ATPases was performed in INVSc1 or FGY217 *S. cerevisiae* strains. The main characteristics of these strains are described in Neo1 results and discussion
- ***Bacterium E. coli***: In *E. coli* the expression is performed in strains that are engineered to contain the λ DE3 lysogen. In these cells the T7 RNA polymerase is produced under the control of the lac promoter as described in Wzy Results and Discussion.

The cell growth rate is monitored by the optical cellular density measured at the wavelength of 600 nm (OD_{600}). This parameter is related to protein expression: when the cells are under high stress due to overexpression of a toxic protein, they reduce their replication rate if compared to non-induced control cells .

4.2.2 *Large-scale expression*

With the conditions identified in the small-scale overexpression tests, a scale-up of several culture liters is performed to produce high amounts of protein.

4.2.3 *Detergent solubilization*

One of the major challenges in membrane protein purification is the solubilization step (Stetsenko and Guskov 2017). Membrane proteins are generally highly hydrophobic and need a lipid surrounding to maintain their shape and functionality. Isolation from the membrane is allowed by creating a micelle around the hydrophobic transmembrane portion of the protein. Detergents mimic the phospholipid environment since they own a polar headgroup and a hydrophobic tail, thus forming micelles in aqueous solutions. The minimum concentration for each detergent to spontaneously form micelles is the CMC and varies according to temperature, ionic strength, pH and presence of other lipids and proteins in the solution. Detergent extraction of the proteins from the membrane often destabilizes their native structure. Therefore, the selection of the correct detergent is a critical point for the whole purification procedure.

Detergents are divided in four main families (A. M. Seddon, Curnow, and Booth 2004) (**Figure 4.6**):

- **Ionic detergents:** the headgroup of ionic detergents has a net charge, while the tail contains a steroidal backbone or a hydrocarbon chain. Ionic detergents can be further divided in cationic and anionic, depending on positivity or negativity of the headgroup charge.

Ionic detergents are classified as harsh and are extremely effective in membrane protein solubilization. However, they have the tendency to denature the protein and

are therefore generally avoided for membrane protein structure determination. An example of this class of detergents is the sodium-dodecyl-sulfate (SDS).

- **Non-ionic detergents:** the headgroup of non-ionic detergents is formed by polyoxyethylene or glycosidic uncharged groups. While ionic detergents tend to break protein-protein interactions, non-ionic detergents are only able to break lipid-lipid or lipid-protein interactions, being thus mild and relatively non-denaturing. For these properties non-ionic detergents are preferred for membrane protein purification. Examples of this class are *n*-octyl- β -D-glucopyranoside (OG), *n*-decyl- β -D-maltoside (DM) and *n*-dodecyl- β -D-maltoside (DDM), octaethylene-glycol-monododecyl-ether (C12E8) and 2,2-didecylpropane-1,3-bis- β -D-maltopyranoside (LMNG). In particular, DDM and DM possess the same maltose headgroup, differing only for the length of the carbon hydrophobic tail. Similarly, LMNG headgroup is formed of two maltose residues, while its hydrophobic tail is formed of two carbon chains, resembling a double DDM molecule.
- **Zwitterionic detergents:** they possess a combination of non-ionic and ionic detergent properties. Zwitterionic detergents are generally less harsh than ionic detergents but more denaturing than non-ionic ones. Examples of this class are *n*-dodecyl-N,N-dimethylamine-N-oxide (LDAO) and 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS).
- **Bile acid salts:** this class of detergents represents a subgroup of ionic detergents, as the headgroup of bile acid salts possesses a positive or a negative charge, but their backbone is composed by a rigid steroidal group. This structure generates a polar and a non-polar face, rather than a well-defined headgroup and tail. They are relatively milder and less denaturing than ionic detergents. A member of this group of detergent is sodium deoxycholate.

Besides detergent nature, an important factor is its CMC. Low detergent CMC usually leads to low solubilization yields, whereas high CMC leads to protein structure destabilization.

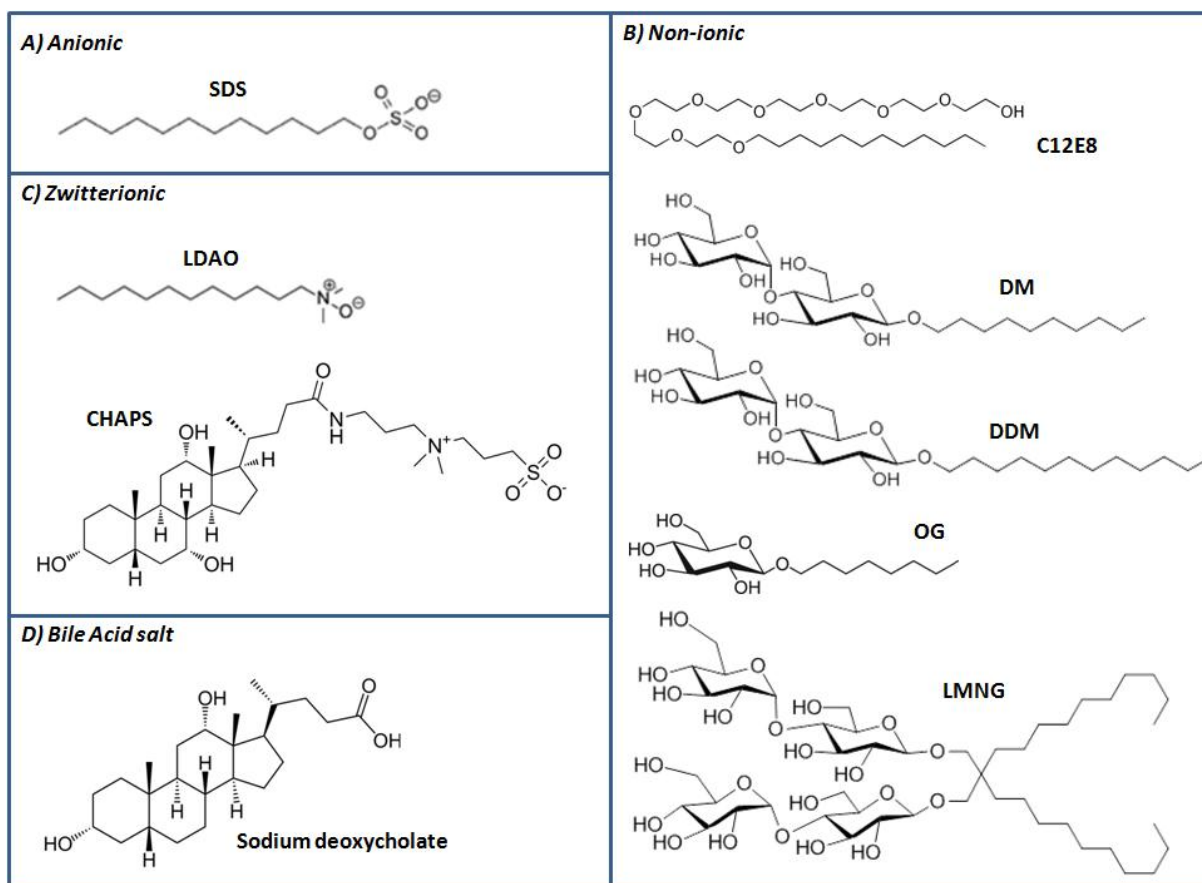


Figure 4.6: Detergents for membrane protein solubilization. A) Anionic: sodium-dodecyl-sulfate SDS; B) Non-ionic: octaethylene-glycol-monododecyl-ether (C12E8), n-decyl- β -D-maltoside (DM) and n-dodecyl- β -D-maltoside (DDM), n-octyl- β -D-glucopyranoside (OG), 2,2-didecylpropane-1,3-bis- β -D-maltopyranoside (LMNG); C) Zwitterionic: n-dodecyl-N,N-dimethylamine-N-oxide (LDAO) and the 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS); D) Bile acid salt: Sodium deoxycholate.

4.2.4 Chromatographic methods for membrane protein purification

Recombinant proteins are often expressed with tags that allow for their selection on different chromatographic resins. Besides the presence of tags, physical and chemical properties such as size and isoelectric point can be exploited.

- **Affinity chromatography:** Affinity chromatography is often the first purification step for both soluble and membrane proteins. Immobilized metal affinity chromatography (IMAC) exploits the presence on the recombinant protein of a histidine tag, normally formed by 6 to 10 histidine residues. The bivalent metal ions immobilized on the IMAC resin (typically Co^{2+} or Ni^{2+}) bind to the protein His-tag. After the binding, the protein can be eluted from the resin with a buffer containing imidazole, that binds to

the resin with stronger affinity. The histidine chain length and position affects the protein purification. Usually, longer His-tags lead to stronger binding to the IMAC resin, but they also require higher imidazole concentration for elution, increasing the risk of imidazole-related protein denaturation. Finally, constructs with a high number of histidine in the tag have the tendency to form oligomers that represent a problem in further purification steps (Mohanty and Wiener 2004).

Other types of affinity resins are used with different tags. In particular, some resins are designed with highly specific antibodies linked to the matrix for immunoprecipitation (IP). An example of this is the GFP-TrapA resin, specifically designed for binding to GFP chimeric proteins.

- ***Size exclusion chromatography (SEC)***: this technique is generally used to separate macromolecules according to their hydrodynamic volume, defined by the Stoke radius. Moving through the porous polymer beads of the column, proteins are retained in the pores of the stationary phase. Small proteins or molecules are able to penetrate all the pores of the polymer and require more time than large proteins. SEC is usually the final step in protein purification for structural studies, as it allows to assess the aggregation state prior to crystallization (Kunji et al. 2008).
- ***Ion exchange chromatography (IEX)***: this chromatographic method is often used to separate and purify proteins by exploiting their isoelectric point (pI). IEX depends on the reversible adsorption of charged molecules on groups of opposite charge, immobilized on an inert resin. Above the isoelectric point the protein surface is negatively charged, while below it is positively charged. Mono-Q and Mono-S IEX are widely used as ion exchange matrices since they bind to negatively or positively charged proteins, respectively. During elution, the surface charge of the bounded protein is altered by an increase of the ionic strength or a variation of the pH of the buffer (Khan 2012).

4.2.5 *Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis SDS-PAGE*

Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is the elective technique for protein separation and identification. In this techniques, proteins are solubilized in SDS and run in an electric field through the porous matrix generated by the polyacrylamide. Since they are surrounded by the negative charges of the detergent, proteins can be separated only according to their molecular weight (MW). Different polyacrylamide concentrations generate a porous matrix with tighter or larger meshes.

This technique is widely used to evaluate the molecular size of soluble proteins, but for membrane proteins an anomalous migration on SDS-PAGE is observed with respect to the soluble globular proteins used as standards. This phenomenon is related to high hydrophobicity of membrane proteins and to their tendency to aggregate. In order to facilitate their migration through the polyacrylamide gel and reduce aggregation, membrane proteins can be solubilized in higher amounts of SDS and should not be thermally treated before loading on the gel (Rath and Deber 2013).

4.2.6 *Use of Green Fluorescent Protein for membrane proteins detection*

The green fluorescent protein (GFP) from *Aequorea victoria* is a 238-residue protein with MW around 30 kDa. It is widely used as detection method when expressed as a tag of the target protein. GFP contains a chromophore, constituted by the oxidized form of three residues, emitting green light under blue light excitation. Maturation of the GFP chromophore occurs within 2-4 hours after synthesis (Gerdes and Kaether 1996). Absorption and emission wavelengths depend on the GFP-type. The fluorescence of the e-GFP (enhanced GFP) used in these studies has been measured at 512 nm using a excitation wavelength of 488 nm (D. Drew et al. 2008).

It has been demonstrated that the emission of green light is a marker for correct protein folding, even for recombinant membrane proteins, since fluorescence is detectable only if the GFP protein is correctly folded. As the GFP is linked to the protein of interest at its C-terminal, incorrect protein synthesis or folding leads to incorrect folding of the GFP (D. E. Drew et al. 2001). In addition, analysis of the fluorescence of membrane and soluble components of the cell lysate can confirm protein localization on the membrane.

Thanks to its properties, the GFP tag is widely used as label during overexpression, solubilization and purification of membrane proteins. Moreover, the fluorescence of GFP-tagged protein is visible in SDS-polyacrylamide gels under UV light, allowing identification of the target protein without the need for long and laborious techniques such as Western Blot. The detection limit for each GFP-tagged protein band is of 5 ng, making this method highly sensitive for protein identification (D. Drew et al. 2008).

4.2.7 *Membrane protein quantification*

Standard protein quantification methods are used on membrane proteins taking into account the presence of detergents can possibly interfere with some assays (Olson and Markwell 2007).

- **UV absorbance:** the UV absorbance method is the simpler way to quantify a protein. Aromatic residues such as tyrosine and tryptophan absorb UV light at 280 nm, while phenylalanine absorbs at 260 nm. UV analysis can be performed on a classic spectrophotometer or with instruments, like Simpliciano or Nanodrop, that require microvolumes of sample, preventing waste of precious specimens. Requisite of this analysis is a high purity of the sample in order to apply the Beer-Lambert law. The absorption coefficient of a protein can be easily calculated with online tools such as ProtParam starting from the protein sequence.
- **BCA assay:** this quantification method is based on the complexation between the bicinchoninic acid (BCA) and Cu^+ ions, obtained from the reaction of the Cu^{2+} -protein Biuret complex in alkaline solutions. Cu^{2+} is reduced by protein peptide bonds to Cu^+ , in a temperature dependent reaction. The BCA- Cu^+ complex has a strong absorbance at 562 nm. The unknown protein concentration is calculated from a standard curve, generally prepared by testing bovine serum albumin (BSA) standard solutions.

4.3 BIOPHYSICAL CHARACTERIZATION OF MEMBRANE PROTEINS

Prior to structural studies, the protein stability has to be evaluated. Different techniques such as circular dichroism and Raman spectroscopy give information about the correct folding of the protein and its melting temperature.

4.3.1 *Circular dichroism on membrane proteins*

Circular dichroism (CD) is a method for the evaluation of protein secondary structure content. In this technique a light beam is polarized through a prism, resulting in a linearly polarized light formed by two components, a left-handed and a right-handed circularly polarized components, 90° out of phase. When these two circularly polarized lights interact with an asymmetric molecule, they are absorbed and refracted in different manner. This results in a rotation of the polarization plane and in an elliptically polarized transmitted light. Proteins absorb the polarized light, since not only they are formed by chiral residues, but also their 3D arrangement is asymmetric. Different structural elements possess particular CD spectra but the most evident effect is given by α -helices and β -sheets (**Figure 4.7**). α -helical proteins have a positive maximum around 193 nm and two negative minima around 208 nm and 222 nm. Proteins that are mostly β -sheets have positive maxima around 195 nm and a single negative minimum at 218 nm. Disordered proteins display very low positive signals above 210 nm and a negative minimum around 195 nm (Greenfield 2006). The behavior of membrane proteins in CD analysis is slightly different from that of soluble proteins. In fact, in membrane proteins with large α -helical domains, the presence of the detergent micelle and their high hydrophobic composition produce a 1-3 nm red shift of the minimum at 222 nm and an intensity reduction of the signal at 190 nm. Therefore caution must be exercised when analyzing CD spectra of membrane proteins (Glaser and Singer 1971).

The presence in the CD spectra of features related to the secondary structure of the protein can be exploited also for the estimation of the melting point, by applying a temperature gradient to the sample. In fact, below the melting point the secondary structure of the protein is clearly visible, while after thermal denaturation signals due to the secondary structure disappear. During protein melting several adjustments in the secondary structure occur, resulting in a sigmoid in the temperature/signal intensity plot.

Circular Dichroism Spectra

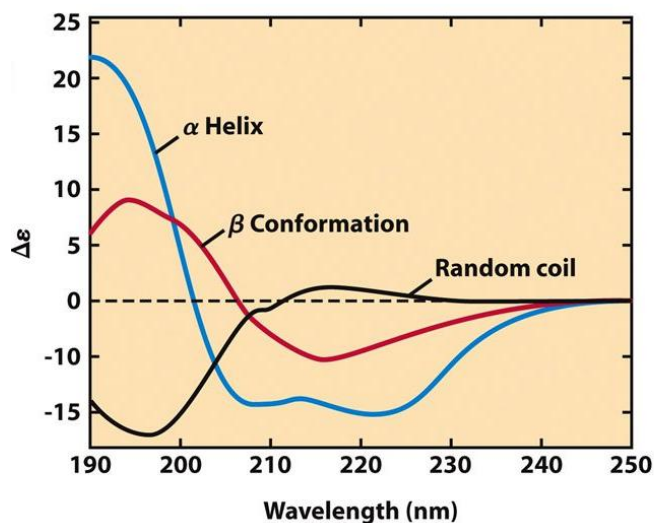


Figure 4.7: CD protein spectra related to secondary structure elements. Image from <https://slideplayer.com/slide/12130416/>.

4.3.2 Fluorescence spectroscopy

Fluorescence spectroscopy allows for tryptophan detection. Tryptophan excitation wavelength is 295 nm while its emission peak is around 320-350 nm. The Wzy recombinant protein has 7 tryptophan residues, 5 of which in the TM domain according to predictions. During a temperature gradient, when the protein reaches its melting temperature, the secondary structure starts to unfold, exposing the tryptophan residues to the aqueous environment surrounding the protein. This results in a variation in the tryptophan total signal in the spectrum (Mozo-Villarías 2002).

4.3.3 Raman spectroscopy for membrane proteins

Raman spectroscopy can be used to obtain protein structural information. In Raman spectroscopy the sample is excited by a laser and a spectrum is generated by recording intensities bands between 500 cm^{-1} and 1800 cm^{-1} . The Raman scattering spectrum gives a complementary information compared to infrared (IR) absorption spectroscopy. In the Raman technique, in order to generate a signal, each molecule needs to undergo a change in its polarizability (Thomas 1999). This is true for vibrations involving multiple bonds or

electron rich groups, such as C=O, C=N, C=C or sulfur-containing groups. Therefore, the main features of the Raman spectrum can be assigned to peptide chains, amines, aromatic side chains and sulfur-containing structures. Some residues display transitions at specific wavelengths: the tryptophan band at 878 cm^{-1} , the phenylalanine band at 1003 cm^{-1} , the tyrosine and tryptophan doublet at 1618 cm^{-1} and the amide I band at 1666 cm^{-1} .

As for CD analysis, applying a temperature gradient to the sample it is possible to follow protein denaturation since the intensity of typical bands decreases (**Figure 4.8**). Usually, the most evident effect of secondary structure loss, indicative of protein denaturation, is the decrease of the intensities related to the peptide backbone such as amide III at 1238 cm^{-1} and amide I at 1666 cm^{-1} .

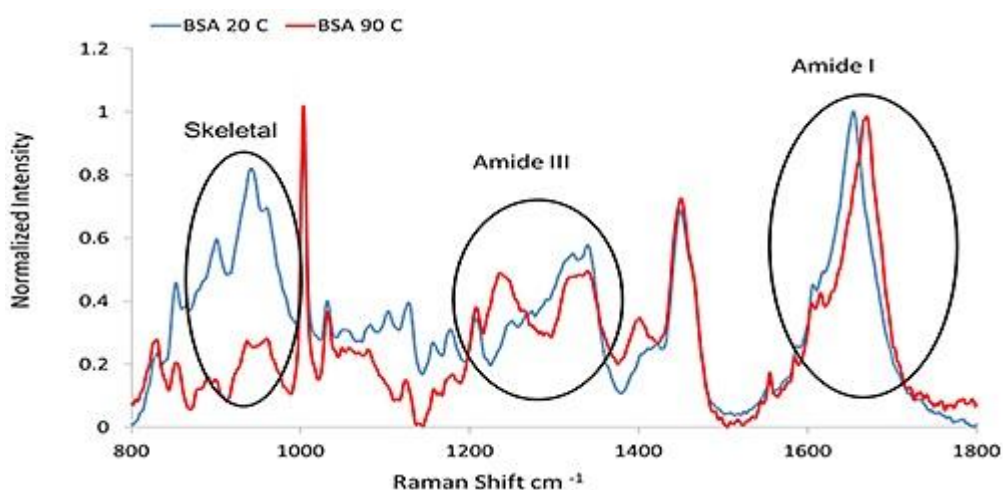


Figure 4.8: Comparison between BSA Raman spectra at 20°C and 90°C . The image shows the decrease and shift in typical peaks occurring after denaturation. Image from <https://www.azom.com/article.aspx?ArticleID=11142>

4.4 STRUCTURAL STUDIES

Once the stability of the purified protein has been determined, structural studies can be performed. The method of choice depends on the type and size of protein. In this study, I focused on X-ray crystallography and Negative Staining EM analysis.

4.4.1 Crystallization of membrane proteins

X-ray crystallography is the most used technique to study protein structures at the atomic level (**Figure 4.9**).

The crystalline solid structure is generated when molecules dispose in a regular periodic array. The smallest repetitive unit inside a crystal is called unit cell. In the diffraction experiments, crystals act as amplifiers for the scattering signal, obtained when a molecule is irradiated by an X-ray source, increasing it of orders of magnitude (M. W. Parker 2003).

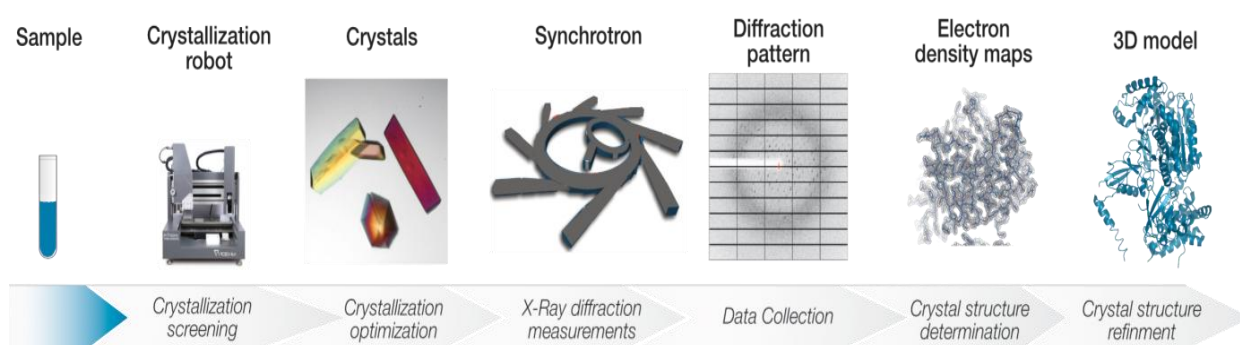


Figure 4.9: X-ray crystallography workflow. The protein is expressed or extracted, purified, crystallized and then analyzed under the X-ray beam of a synchrotron radiation source; the diffraction patterns are processed through software, generating an electron density map from which the structure can be determined. From <http://probiostructures.com/technology/>.

The 3D unit cell is defined by three length dimensions, a , b and c , and three angles, α , β and γ . Translation vectors generate an infinite array of points that represent the nodes of the crystal lattice, called Bravais lattice. According to the relations between their parameters, unit cells can be divided in seven crystal systems that are compatible with the formation of a regular crystal. These systems range from the most regular cubic system, where $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$, to the most asymmetric triclinic system, where $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$. In the unit cell molecules can dispose according to a particular symmetry, defined by the space group elements. Proteins, that are asymmetric objects due to the chirality of their residues, can dispose in 65 possible space groups, since mirror planes and inversion centers cannot be present.

In order to obtain a single crystal of a protein, the high purity degree of the sample is crucial, together with specific conditions of temperature, concentration, pH and precipitants (e.g.

salts or polyethylene glycol). Commercial screen kits are available to test a large number of crystallization conditions.

The crystallization process can be described in a phase diagram (**Figure 4.10**): at low protein and precipitant concentration (undersaturation) the protein is completely soluble; when concentration increases (e.g. through evaporation) the protein becomes less soluble (saturation or supersaturation) and undergoes aggregation or formation of crystalline nuclei. Initial crystal formation occurs in the nucleation zone of the phase diagram. With the formation of nuclei, the concentration of the protein in solution decreases. This variation brings the system in the metastable zone of the phase diagram, the size of crystalline nuclei grows.

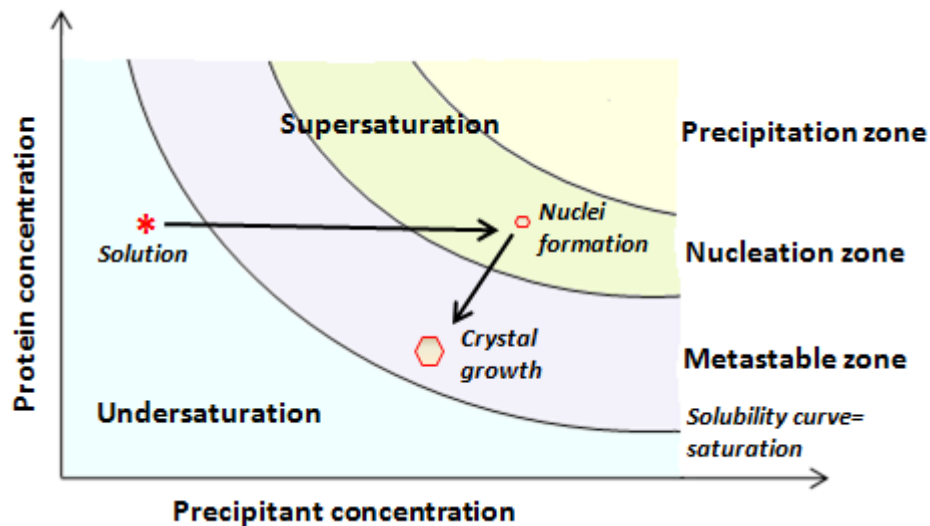


Figure 4.10: Phase diagram describing the crystallization of a protein.

For proteins, the typical starting concentration range is from 2 mg/ml to 50 mg/ml, this concentration should be below the solubility curve of the sample. The most used crystallization techniques are vapor diffusion and batch crystallization (**Figure 4.11**) (Dessau and Modis 2011). In the vapor diffusion technique, a drop obtained by mixing a defined volume of the protein solution and a volume (usually equal) of the precipitant solution is set to equilibrate against a larger amount of the precipitant solution in a closed chamber, until the system reaches the equilibrium. Two experimental setups can be chosen: the *hanging-drop* and the *sitting-drop* depending on the position of the drop containing the protein. The vapor diffusion leads to slow dehydration of the drop until protein and precipitant

concentrations reach the nucleation zone of the phase diagram. In the batch method protein is mixed directly with the adequate amount of precipitant to reach the nucleation zone. In this method the drop is covered with a layer of oil to avoid further dehydration. Another technique is the crystallization in dialysis in which the protein solution is separated from surrounding precipitant solution by a semi-permeable dialysis membrane.

Once formed, the small protein crystals are collected with loops made of a not-refracting materials that vary in shape, size and composition. Crystals are flash-frozen in liquid nitrogen to avoid crystal damage prior to and during data collection. Additionally, a cryoprotectant is usually added to the solution in order to protect the crystal during the flash-freezing and avoid formation of ice crystals.

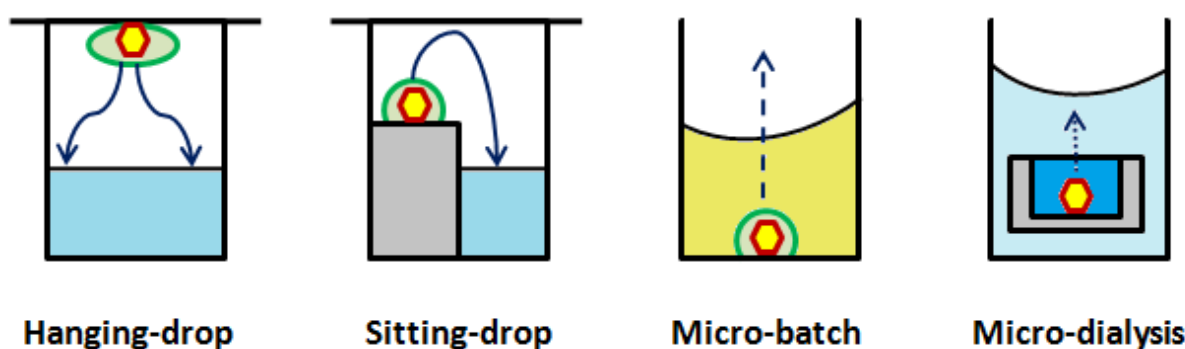


Figure 4.11: Crystallization techniques: hanging-drop vapor diffusion, sitting-drop vapor diffusion, micro-batch under oil layer and micro-dialysis.

For data collection, X-ray beams with wavelength of about 0.1 nm (1 Å) are commonly used for atomic-resolution structural determination of protein samples.

Due to the periodicity of the crystal lattice, the interference of scattering contributions from the atoms in the crystal generates a total scattered wave that has the form of a delta function, with high intensity only in specific directions, generating a typical diffraction pattern.

To each diffracted spot the Miller indexes h , k and l of the family of planes that generated the spot can be associated. The diffracted spots constitute the nodes of the lattice defined by the Miller indexes in the reciprocal space, a Fourier transform of the space of the crystal.

Spots are visible in each diffraction pattern when they satisfy the Bragg's law: $2d\sin\theta = n\lambda$, where d is the distance between two lattice planes, θ is the incidence angle, n is a positive integral number and λ is the wavelength. A useful depiction of this system uses a geometric

construction, known as the Ewald sphere, that has a radius equal to the wavelength of the incident radiation and is centered in the reciprocal lattice. While the crystal rotates, the reciprocal lattice rotates with respect to the Ewald sphere. At each rotation angle, nodes of the reciprocal lattice that intersect the Ewald sphere are in diffraction condition according to Bragg's law and, therefore, produce a diffraction beam.

For protein crystals it is common to use the synchrotron radiation as X-ray source due to its higher brilliance compared to other sources. In synchrotrons, electrons are accelerated close to light speed and then injected in a ring where they are maintained in a circular orbit by magnetic fields. During their run, electrons emit electromagnetic energy for every acceleration, i.e. change of their linear path. The light emitted by electrons is conveyed into beamlines where monochromators allow the selection of radiation of the wavelength of interest.

From diffraction data, the 3D structure of the sample can be obtained after the phase problem is solved. For each diffraction spot, the interaction of the radiation with the electron density of the crystal can be described using a structure factor that takes into account the contents of the unit cell and the direction of the incident radiation. Structure factors are imaginary numbers. While their module can be obtained from the intensity of the diffraction spots, information regarding their phase ϕ is lost. However, in order to reconstruct the electron density of the unit cell, and therefore the 3D structure of the target molecule, both the information regarding the amplitude and the phase of the structure factors is required. To overcome this problem, different methods are available: i) multiple isomorphous replacement (MIR), ii) multi-wavelength anomalous dispersion (MAD), or iii) molecular replacement. In MIR technique, different data sets from the native protein and at least two heavy-atom derivatives are compared. Heavy atom derivatives are obtained by soaking the native crystal in the solution of a heavy atom. In the MAD method, the anomalous scattering generated by the inner shell electrons of molecules when hit by an incident radiation is analyzed. In the molecular replacement method, the structure of a known protein, that shares at least 25% identity with the unknown sample, is correctly positioned in the unit cell of the unknown sample and an initial set of phases is calculated.

Once the phase problem has been solved, i.e. once an initial set of phases is available, an electron density map can be calculated and atoms of the protein are located in the maxima of the electron density. The structure determined from the electron density is used to

calculate new phases, from which an improved electron density map can be obtained. Manual optimization of the structure are alternated to refinement cycles, in which the parameters of the structure are varied in order to obtain a better fit of the experimental data.

The detail of the electron density map depends on the maximum angle of diffraction, from which, in turn, the resolution d_{\min} of the structure can be calculated as the minimal distance between planes of the same family that generate diffraction spots, according to Bragg's law. In general, a resolution from 8 Å to 3.5 Å is considered low and allows for visualization only of the overall shape of the protein. At a medium resolution, between 3.5 Å and 2.5 Å, the side chains of amino acids can be identified. At a resolution of 2.5 Å and below, individual atoms of the protein and of well-ordered solvent molecules can be seen in the electron density (M. W. Parker 2003).

Besides the resolution, other parameters are used to evaluate the reliability of a crystallographic structure, namely the R factor and the R free. The first calculates the difference between the experimentally obtained intensities of the diffracted beams and the values calculated from the structural model. For high-resolution protein structures this value is typically around 20%. However, during refinement some bias is introduced due to the necessity to provide calculated phases in order to produce the electron density map. The R free parameter takes into account the bias by calculating the differences between experimental and calculated intensities for diffraction spots (usually about 5% of the whole dataset) that are purposely kept out during the refinement cycles. For good protein models R value and R free are similar, even if R free always tends to be slightly higher (Kleywegt and Jones 1997).

Membrane proteins: The first requisite for protein crystallization is to work with high concentrations of pure sample. Since membrane proteins have poor expression yields and are difficult to purify, this requisite is not always easy to achieve. Moreover, membrane protein crystals are fragile and difficult to handle, if compared to those of soluble proteins, due to the large solvent content, the relatively weak protein-protein contacts and the presence of disordered detergent. In membrane protein crystallization, only the hydrophilic surfaces of the protein are involved in protein-protein interactions that form the regular structure of the crystal. The most commonly used detergents for protein crystallization are OG, DDM, DM and LDAO since they lead to formation of crystals with a good internal order

and a consequent good diffraction. Different commercial screens kits, such as MemGold 1-2, MemSyst and MemStart (from Molecular Dimensions), are available to test a high number of crystallization conditions (J. L. Parker and Newstead 2016).

4.4.2 *Negative Staining Electron Microscopy for membrane proteins*

Protein negative staining combined with transmission electron microscopy analysis allows for protein visualization on a carbon grid. In a TEM microscope, an electron gun produces a beam, conveyed over the sample by different magnetic lenses, known as condenser lenses. When the beam hits the sample, scattered waves are combined by the objective lens, another magnetic lens, generating on the detector placed after the sample a magnified image of the sample. The sample, inserted in the vacuum column of the microscope, can be moved mechanically in x, y and z directions or rotated to obtain the best image possible (Nixon 1971).

In the negative staining technique, the sample is usually placed on a carbon grid and coated with a thin layer of heavy metal salts. Atoms of the stain have a higher scattering power compared to the protein, resulting in a black background where the protein can be seen as a white object (**Figure 4.12**). The main limitations of this technique are represented by protein dimensions, as the protein has to be large enough to be visible, low resolution due to the dimension of the stain grains, and structural deformation of the sample due to the staining (Zhang et al. 2013). In addition, the presence of the stain induces the protein to adopt a preferential orientation on the microscopy grid. However, in order to reconstruct the 3D model of the protein, images from different orientations are required. Therefore, in negative staining EM, the sample is rotated in order to obtain other views of the protein, in a technique known as Random Conical Tilt.

Several computation tools are available to combine different orientation images and produce a three-dimensional structure of the molecule of interest (Ohi et al. 2004).

For membrane proteins, EM analysis represents an appealing technique since no crystal formation is needed and protein concentration required by this technique is lower than for crystallographic studies.

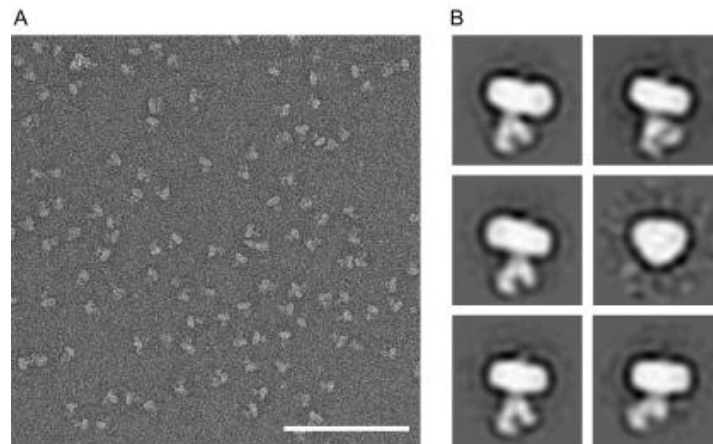


Figure 4.12: A) Negative staining of the membrane protein SERCA surrounded by the Saposin-Lipoprotein. **B)** Classification of images of SERCA in different orientations and averaging increases the contrast of the images and increases their level of detail. From Lyons J. A. et al. (2017), *Methods in Enzymology* (Lyons et al. 2017).

4.5 COMPUTATIONL MODELS

Protein alignment with BLOSUM62 matrix and computational model generation with SwissModel

Many free online tools are available to compare amino acid sequences of different proteins. Software like Lalign or Blast (Pearson 2013; X. Huang and Miller 1991), that allow sequence alignment for both amino acid or DNA sequences, are widely used. In this software, the alignment is based on the BLOSUM matrix, that gives to each residue a score number when present on both the compared sequences. The rarer the residue the higher its score, since it is less likely that two sequences are randomly aligned with these residues in the same position. The matrix gives a score for each possible alignment of the protein sequences of two or more proteins (Eddy 2004). Sequence alignments can be used to create a three-dimensional model of an unknown structure from known structures with high sequence similarity.

Swiss-model is a free web-based automated model facility that allows generation of a 3D protein model only from its amino acid sequence. Model building is possible only if at least one experimental structure of a protein that possesses a significant amino acid sequence similarity with the unknown protein is available. The model can be generated from the

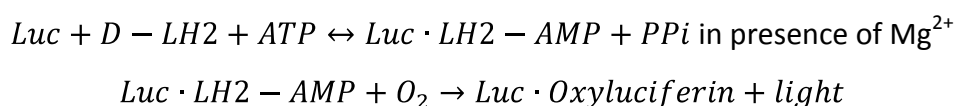
server by automatic selection of suitable templates, user submission of the template or user submission of a complete model.

Templates are weighted according to their sequence similarity with the unknown protein, therefore highly diverse regions are impossible to model. The evaluation of the model quality follows two parameters: the QMEAN factor, that has to be above -4.0 for a good model, and the GMQE factor, that has to be above 0.6 (Schwede et al. 2003).

4.6 BIOCHEMICAL ASSAYS

ATP activity assay based on the luciferin/luciferase chemoluminescence

Proteins such as P-ATPases or ABC transporters require ATP for their enzymatic activity. Usually, for the ATP hydrolysis the enzyme needs also an Mg^{2+} ion, that works as co-factor. Various assays are available to detect the ATP consumption by a protein. The most common is the enzymatic bioluminescent luciferin/luciferase assay (**Figure 4.13**). In this assay, a molecule of luciferin binds to the luciferase protein in presence of ATP and Mg^{2+} , forming an enzyme-bound luciferyl adenylate. This molecule then reacts with oxygen forming an excited complex that emits light according to the reaction:



where Luc is luciferase, D-LH2 is D-luciferin and PPi is the inorganic pyrophosphate.

Light can be easily detected with any luminescence reader instrument (DeLuca and McElroy 1974). To detect ATP consumption, highly sensitive commercial kits are available.

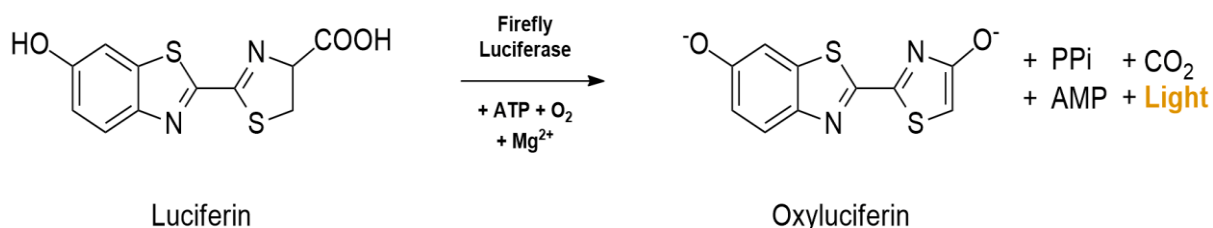


Figure 4.13: Luciferin/Luciferase bioluminescent reaction. The molecule of Luciferin reacts with Firefly Luciferase consuming ATP in presence of Mg^{2+} and O_2 ; the generation of Oxyluciferin emits light.

4.7 MATERIALS

Materials:

Primers: primers for homologous recombination in P-IV ATPase experiments were purchased from Thermo Fisher Scientific, all other primers were from Sigma-Aldrich.

Cell strains: *S. cerevisiae* **FGY217** [*MATa, ura3-52, lys2Δ201, pep4Δ*] unmarked strain (Kota, Gilstring, and Ljungdahl 2007) and **INVSc1** [*MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52*] Thermo Fisher; *E. coli* **XL1-Blue** [*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proABlacIq ZΔM15 Tn10 (Tetr)]*] kindly given from the Elettra BioLab, **BL21(DE3)pLysE** [*F - ompThsdSB(rB - mB -) gal dcm (DE3) pLysE (CmR)*] Sigma-Aldrich, **Lemo21(DE3)** [*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/pLemo (Cam^R) λ DE3=λ sBamHloΔEcoRI-B int:: (lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*] New England Biolabs.

Plasmids: pDDGFP_LEU2D and PWaldo_GFP from Addgene.

Enzymes: Smal Roche, DpnI, Lysozyme from chicken egg white powder and Q5 High Fidelity polymerase from New England Biolabs.

Lipids: all lipids were purchased from AvantiPolar lipids.

Affinity Resins: Talon Superflow GE Healthcare, Ni-NTA Agarose Invitrogen, GFP-TrapA ChromoTek.

Columns: Superdex200 Increase 10/300 GL and Superdex 200 10/300 GL were purchased from GE Healthcare, while MonoQ 5/50 GL and MonoS 5/50 GL were purchased from Amersham Biosciences.

Kits: Q5 Site-Directed mutagenesis kit New England Biolabs, GeneElute plasmid Miniprep kit, GeneElute Gel extraction kit, BCA bicinchoninic acid kit for protein determination Sigma, ADP-Glo kinase assay kit and Cell-Titer-Glo 2.0 kit from Promega, MemGold 1-2, MemSyst and MemStart crystallization screens were purchased from Molecular Dimension.

Materials: BSA, Bradford reagent, Phenylmethanesulfonyl-fluoride PMSF, protease inhibitor cocktail, Agarose, Tryzma-BASE, Deoxyribonucleic acid sodium sulfate from salmon testis, Galactose, Glucose, Yeast nitrogen base without amino acids, Yeast synthetic drop-out medium supplements without uracil or leucine, Chloramphenicol, Ampicillin sodium salt, Isopropyl-β-D-thiogalactopyranoside IPTG, Kanamycin sulfate, Manganese (II) chloride tetrahydrate, Potassium acetate (KAc), 3-(N-Morpholino)propansulfonic acid 4-Morpholine

Materials and methods

propane sulfonic acid (MOPS), Glycine, calcium acetate, cobalt acetate, brilliant blue, dNTPs, Ammonium Persulfate (APS), protein Marker, nitrocellulose 0.22 μm filter membranes were purchased from Sigma; Acrylamide Acryl/BIS 40% w/v and Tryptone were purchased from VWR; Yeast extract and lithium acetate were purchased from Acros; Peptone de caseine was purchased from Fisher bioscientific; L-rhamnose was purchased from New England Biosciences; Sodium chloride NaCl was purchased from Carlo Erba; Imidazole was purchased from TitolChimica; Rubidium chloride RbCl was purchased from Aldrich chemistry; Glycerol and Calcium chloride dihydrate $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ were purchased from Riedel-de Haën; Zinc acetate; Bromophenol Blue, Polyethylene glycol PEG 400 and 4000 were purchased from Fluka analytical; PEG 300 and 3350 were purchased from Aldrich; EDTA was purchased from Amersham Biosciences; Pro Blue safe stain was purchased from Giotto Biotek; Temed was purchased from PlusOne; DNA marker Gel Pilot was purchased from Qiagen; ECL star enhanced chemiluminescent substrate Euroclone; Filter papers were purchased from Whatman; AmiconUltra 100K 15 ml and 4 ml were purchased from Millipore; Vivaspin 50K 6 ml or 20 ml were purchased from Sartorius; High precision Cell Quartz 0.1 mm or 0.2 mm light path were purchased from Helma analytics; Filtropur 0.2 μm , Elisa plate white Med Bind F, Elisa plate black Med Bind F and all plastics were purchased from Starstedt, TunAir shake flasks 2500 ml were purchased from Sigma-Aldrich, adhesive film for microplates from VWR and CrystalQuick 24 plate from Greiner bio-one.

Instruments: Tecan infinite m1000pro plate reader, Jasco j-815 CD spectrometer, Varian Cary eclipse fluorescence spectrophotometer, Biotek synergy h1 hybrid reader, VWR imager CHEMI premium, Simplinano GE, Shimadzu UV-1800 UV spectrophotometer, Hermle z326k centrifuge, Beckman L8-70 ultracentrifuge and Avanti j-26 XP centrifuge, Major science MP-300V electrophoresys, AKTA FPLC Amersham biosciences, Prime Thermal Cycler, APV Lab 2000 homogenizer, Philips EM208 with a QUEMESA camera, EM JEOL JEM2100 LaB6 with GATAN 2K \times 2K cameras, Mosquito TTP Labtech.

Media and buffers:

- **YPD medium:** 10 g l^{-1} yeast extract, 20 g l^{-1} peptone, 20 g l^{-1} glucose.
- **URA- and LEU- selective liquid media:** 6,7 g l^{-1} yeast nitrogen base without amino acids and no glucose, 2 g l^{-1} yeast synthetic drop out medium without uracil or leucine.

Materials and methods

- **URA- and LEU- agar plates:** 6,7 g l⁻¹ yeast nitrogen base without amino acids and no glucose, 2 g l⁻¹ yeast synthetic drop out medium without uracil or leucine, 20 g l⁻¹ glucose, 20 g l⁻¹ agar.
- **LB medium:** 10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 5 g l⁻¹ NaCl.
- **LB agar plates:** 10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 5 g l⁻¹ NaCl, 20 g l⁻¹ agar.
- **TBS buffer:** 20 mM Tris pH 7.5, NaCl 150 mM.
- **Neo1 buffer:** 50 mM Tris pH 7.5, NaCl 150 mM.
- **Neo1 resuspension buffer:** 50 mM Tris pH 7.5, NaCl 150 mM, PMSF 1 mM, 200 µl protease inhibitor cocktail for liter of culture.
- **Neo1 solubilization buffer:** 50 mM Tris pH7.5, NaCl 150 mM, detergent of choice.
- **Wzy buffer:** 20 mM Tris pH 8.2, 150 mM NaCl.
- **Wzy resuspension buffer:** 20 mM Tris pH 8.2, 150 mM NaCl, 1 mM PMSF.
- **Wzy solubilization buffer:** 20 mM Tris pH 8.2, 150 mM NaCl, 1 mM PMSF, detergent of choice.
- **Wzy SEC or IEX buffer:** 20 mM Tris pH 8.2, 150 mM NaCl, detergent of choice.

All media were sterilized before use. All buffers were filtered on 0.2 µm membranes. Buffers for SEC or IEX purification were degassed overnight.

Detergents: in order to solubilize and purify membrane proteins in this study C12E8, CHAPS, DM, DDM, LDAO, LMNG and OG detergents were used. Their properties are summarized in **Table 4.1**, together with the concentration used in experiments. All detergents were purchased from Anatrace.

Materials and methods

| Detergent | Formula Weight (Da) | CMC (mM) | Aggregation Number | Micelle size (kDa) | Family | Concentration used in this study (x CMC) |
|--------------|---------------------|----------|--------------------|--------------------|--------------|---|
| C12E8 | 538.8 | 0.09 | 90-120 | 66 | Non Ionic | 6 (purification), 200 (solubilization) |
| CHAPS | 614.9 | 8 | ~10 | 8-10 | Zwitterionic | 10 |
| DM | 482.6 | 1.8 | 69 | 40 | Non Ionic | 10 (purification), 10-20 (solubilization) |
| DDM | 510.6 | 0.12 | 78-149 | 65-72 | Non Ionic | 5 (purification), 10- 20 (solubilization) |
| LDAO | 229.4 | 1-2 | ~76 | 17-215 | Zwitterionic | 10 (purification), 10-22 (solubilization) |
| LMNG | 1005.19 | 0.01 | ~400 | 91-393 | Non Ionic | 3-10 (purification), 100 (solubilization) |
| OG | 292.4 | 18-20 | 27-100 | 8-29 | Non Ionic | 10 |

Table 4.1: Summary of detergents used in this PhD work, with their properties. Data from Anatrace website or from A. Stetsenko and A. Guskov (2017), *Crystals* (Stetsenko and Guskov 2017).

Agarose gel for DNA separation: Running buffer, TAE buffer (Tris-acetate-ethylenediaminetetraacetic acid (EDTA)): 40 mM Tris, 20 mM acetic acid, 1 mM EDTA at pH 8.0. Gels were prepared in the same buffer with 0.8% w/v agarose and DNA-binding dye. Gel electrophoresis system was set on 100 V and 80 mA.

SDS-PAGE for protein separation and identification: Polyacrylamide gels were made of two parts: stacking gel (125 mM Tris pH 6.8, 0.1% w/v SDS, 3.75% w/v acrylamide, ammonium persulfate (APS) 0.1% w/v and Tetramethylethylenediamine (Temed) 1.6 $\mu\text{l ml}^{-1}$) where samples were loaded, and running gel (Tris 0.4 M pH 8.8, SDS 0.1% w/v, APS 0.1% w/v, Temed 1 $\mu\text{l ml}^{-1}$ and acrylamide 8% w/v for P-IV ATPases analysis and 14% w/v for Wzy analysis) where proteins were effectively separated. Samples were premixed with 4X Laemmli loading buffer (8% w/v SDS, 40% w/v glycerol, 120 mM Tris pH 6.8, 14% w/v β -mercaptoethanol, 0.1% w/v bromophenol blue), with increased SDS concentration when required, and heated at 37°C for 10 min to avoid membrane protein aggregation. Running

buffer (192 mM glycine, 25 mM Tris pH 8.5, 3.5 mM SDS) was always freshly prepared. Gel electrophoresis system was set on 120-140 V and gels were stained with Pro Blue safe stain.

4.7.1 *Neo1* protocols

- Cloning protocol for P-IV ATPases:

PCR for gene amplification:

According to DNA polymerase manufacturer's protocol the reaction solution (0.2 mM dNTPs, 0.5 μ M primer For, 0.5 μ M primer Rev (**Table 4.2**), 20 ng template, Q5 high fidelity DNA polymerase 1 μ L, 5 X Q5 reaction buffer 5 μ l, H₂O up to 25 μ L) underwent the following thermal treatment in a thermal cycler:

- Initial denaturation: 5 min at 98°C
- Reaction cycle: i) Denaturation: 30'' at 98°C ii) Annealing: 20'' at 55°C; iii) Elongation: 3 min at 72°C. For a total of 35 cycles.
- Final elongation: 5 min at 72°C
- Final hold 4°C.

| Protein | Primer | Sequence |
|-------------|--------|--|
| Drs2 | For | ACCCCGGATTCTAGAACTAGTGGATCCCCCATGAATGACGACAGAGAAACC |
| | Rev | AAATTGACCTTGAAAATATAAATTTTCCCCTATATCAAATGAAATATCATCTCT |
| Dnf1 | For | ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTCTGGAACCTTTCATGGC |
| | Rev | AAATTGACCTTGAAAATATAAATTTTCCCATTGTTCTGTTGTGTTCCGAT |
| Dnf2 | For | ACCCCGGATTCTAGAACTAGTGGATCCCCCATGTCAAGTCCCTCCAAACCC |
| | Rev | AAATTGACCTTGAAAATATAAATTTTCCCACGGTCCCTGCTGCGCTGGCT |
| Neo1 | For | ACCCCGGATTCTAGAACTAGTGGATCCCCCATGCCTAACCTCCTTCATT |
| | Rev | AAATTGACCTTGAAAATATAAATTTTCCCCTGGAGTGGCAAACCTCTGCAC |

Table 4.2: For and Rev primers for *Drs2*, *Dnf1*, *Dnf2* and *Neo1* amplification from *S. cerevisiae* genome. Green: part of the primer complementary to *pDDGFP_LEU2D*; Light blue: part of the primer complementary to the gene of interest.

DNA amplified sequence size was checked on a 0.8% w/v agarose gel.

Transformation in FGY217 and INVSc1 S. cerevisiae strains via homologous recombination:

Competent cells preparation: yeast strains FGY217 and INVSc1 were inoculated in 5 ml YPD medium overnight at 30° C and 180 rpm shake. The culture was diluted to OD₆₀₀=0.1, measured with UV spectrophotometer, in 50 ml YPD and grown to OD₆₀₀=0.6 at 30°C and 180 rpm shake. Cells were then centrifuged at 3000 g for 5 min at 4°C. The resulting cellular pellet was resuspended in 25 ml sterile H₂O. A second centrifugation was performed, and the pellet was resuspended in 1 ml of 100 mM lithium acetate solution. Cells were centrifuged at 8000 g for 15'' at 4°C. Pellet was resuspended in 400 µl 100 mM lithium acetate solution (D. Drew et al. 2008).

Yeast transformation: to 50 µl of competent cell resuspension solution 240 µl PEG 3350 50% w/v, cooled on ice, were added and the resuspension was vortexed. 25 µl of 2 mg ml⁻¹ single-strand DNA, pre-heated for 10 min at 100°C, were added to cell resuspension that was then vortexed. Addition of 50 µl DNA mix (3 µl of 25 ng µl⁻¹ pDDGFP_LEU2D digested with SmaI enzyme, 5 µl of 150 ng µl⁻¹ amplified gene and 42 µl sterile H₂O) to the cell solution and vortex. In this step homologous recombination allows gene insertion into the plasmid exploiting cellular mechanisms, without the need of additional enzymes. Cells were incubated for 30 min at 30°C. To allow plasmid penetration, a heat-shock at 42°C for 25 min was performed. Cell resuspension was centrifuged at 8000 g for 15'' and the pellet was resuspended in 100 µl sterile H₂O. Cells were plated on selective URA- or LEU- agar plates and grown at 30°C for 2-3 days. In alternative cells were inoculated in 5 ml liquid selective medium with ampicillin 0.1 mg ml⁻¹ and grown overnight at 30°C and 180 rpm shake.

Plasmid yield increase in E. coli and gene insertion control:

To obtain a plasmid amount suitable for further transformations and sequence control, the plasmid was extracted with commercial kits from yeast cells and transformed in *E. coli* competent cells. Plasmid was then extracted from *E. coli* cells.

FGY217 or INVSc1 transformed cells were inoculated in 5 ml liquid selective medium with ampicillin 0.1 mg ml⁻¹ and grown overnight at 30°C under 180 rpm shake. From this culture, plasmid was extracted with a Miniprep kit, that allows for selective extraction of bacterial plasmid DNA. A control PCR was performed with For and Rev primers for homologous recombination procedure with the same parameters, and the resulting DNA fragment was

checked on a 0.8% w/v agarose gel. Only plasmids that carry the target sequence were used in the further steps.

E coli XL1-Blue competent cells were then transformed with the plasmid according to manufacturer's instructions. XL1-Blue cells are engineered for transformation with large plasmids but not for protein expression (Dilworth et al. 2018). Transformation protocol was:

1. Addition of the extracted plasmid to 50 μ l of XL1-Blue competent cells (no more than 5 μ l and no more than 1 μ g DNA).
2. 30' on ice.
3. 45'' heat-shock at 42°C followed by 5 min on ice.
4. 1 h at 37°C and 180 rpm shake.
5. Cells inoculation in 5 ml selective LB liquid medium with ampicillin 0.1 mg ml⁻¹; overnight growth at 37°C and 180 rpm shake.

Plasmid was then extracted from transformed cells with Miniprep kit as further described. The obtained plasmid was checked with control PCR as previously described. Additionally, plasmid sequence was analyzed via sequencing and the positive plasmid used to transform yeast cells for protein expression.

- Expression protocol for P-IV ATPases:

To induce protein production, 2% w/v galactose was added to the culture medium to activate the strong GAL1 promoter upstream the P-IV ATPase gene sequence. On the contrary, to inhibit the promoter, 2% w/v glucose was added to the culture medium (negative controls).

Small-scale overexpression protocol:

FGY217 or INVSc1 cells were transformed and inoculated in 5 ml liquid URA- selective medium with 2% w/v glucose and ampicillin 0.1 mg ml⁻¹. The culture was grown overnight at 30°C and 180 rpm shake. Cells were diluted to OD₆₀₀=0.12 in 10 ml liquid selective medium with 0.1% w/v glucose and ampicillin 0.1 mg ml⁻¹, and grown at 30°C and 180 rpm shake to OD₆₀₀=0.6. Protein production was induced with 2% w/v galactose and culture was grown for 22-24 h at 30°C and 180 rpm shake. The protein production was constantly followed with GFP fluorescence detection, while cells growth was checked through the OD₆₀₀ parameter.

Expression condition variations:

Materials and methods

- i. Addition of 2.5% w/v DMSO to selective liquid medium (D. Drew et al. 2008).
- ii. Addition of 100 μ M cobalt acetate, 5 mM zinc acetate or 200 mM calcium acetate to selective liquid medium (T. Pomorski et al. 2003).
- iii. Lowering of growth temperature to 18°C

Large-scale expression protocol:

FGY217 or INVSc1 transformed cells were inoculated in 5 ml liquid selective medium with 2% w/v glucose and ampicillin 0.1 mg ml⁻¹ and grown overnight at 30°C and 180 rpm shake. Cells were diluted in 50 ml liquid selective medium with 2% w/v glucose and ampicillin 0.1 mg ml⁻¹ and grown overnight at 30°C and 180rpm shake. Overnight culture was diluted to OD₆₀₀=0.12 in 1 L liquid selective medium with 0.1% w/v glucose and ampicillin 0.1 mg ml⁻¹ and cells were grown to OD₆₀₀=1.2 at 30°C and 180 rpm shake. Protein production was induced with 2% w/v galactose and culture was grown for 22 h at 30°C and 180 rpm shake. GFP fluorescence was followed in all the steps of the process to estimate protein yield.

- *Neo1 mutants protocol*

Mutants on Neo1 ATPase activity

Mutants were designed to substitute the original residue with one that leads to a reduced ATP consumption (**Table 4.3**) (Andersen et al. 2016). Elongation time 6 minutes, 25 cycles.

| Mutation | Primer type | Primer | Sequence | T elongation |
|--------------|----------------|--------|-----------------------------------|--------------|
| D503N | Phosphorylated | For | CCTTCTAAGCAACAAAACAGG | 57°C |
| | back to back | Rev | TATTC AATACGGCCTAAATC | |
| K504N | Phosphorylated | For | CCTTCTAAGCGACAACACAGG | 57°C |
| | back to back | Rev | TATTC AATACGGCCTAAATC | |
| K504R | Overlapping | For | TACCTTCTAAGCGACAGAACAGGGACGCTTACT | 64°C |
| | | Rev | AGTAAGCGTCCCTGTTCTGTCGCTTAGAAGGTA | |
| E310S | Phosphorylated | For | ACTAGATGGGAGCACTGACTGGAAAC | 57°C |
| | back to back | Rev | TGGTCAGTTTTGATAAATGAC | |

Table 4.3: Primers for mutants on ATPase activity. Mutations are highlighted in yellow.

Mutants on the lipid substrate binding site

All the primers were designed back to back (see **Figure 4.5**), therefore every amino acid had the same reverse primer for all the selected mutations (**Table 4.4**). Since all primers had different annealing temperature, the PCR was done in a temperature gradient from 54°C to 66°C, for 35 cycles. The elongation time was of 6.30 min. A double R946E/E1065R mutant was created with the same protocol starting from the R946E mutant plasmid.

| Residue | Mutation | Primer | Sequence |
|---------|----------|--------------------------|--------------------------------------|
| R946 | R946A | For | TGTCATGCATGCAAGGTTTAATCATTG |
| | R946D | For | TGTCATGCATGATGGTTTAATCATTGC |
| | R946E | For | TGTCATGCATGAAAGGTTTAATCATTG |
| | R946K | For | TGTCATGCATAAAAGGTTTAATCATTG |
| | | Rev | AATTGAGCCAATTTAGCAG |
| E1065 | E1065A | For | GGTTGTAAATGCATTAATAATGGTTG |
| | E1065D | For | GGTTGTAAATGATTTAATAATGGTTG |
| | E1065R | For 5' Phosphorylated | GGTTGTAAATAGATTAATAATGGTTGC |
| | | Rev | AAAGCAGTGAAGCTGATAG |
| R460 | R460D | For | TGTCTCTTTAGATGTTAATTTAGATTTAGCCAAAAG |
| | R460E | For | TGTCTCTTTAGAAAGTTAATTTAGATTTAGC |
| | R460K | For | TGTCTCTTTAAAAAGTTAATTTAGATTTAG |
| | | Rev | GGGATGATGGTAGAAAAC |
| R446 | R446A | For | TGACATTTTGCGTATCTGATATTGTTTTC |
| | R446D | For | TGACATTTTGATTATCTGATATTGTTTTCTACCATC |
| | R446E | For | TGACATTTTGAGTATCTGATATTGTTTTC |
| | R446K | For | TGACATTTGAAATATCTGATATTG |
| | | Rev | ATGTACCAGTCGTCATTG |

Table 4.4: Primers for mutations on substrate selection residues. Mutations are highlighted in yellow.

Mutagenesis protocol for Neo1:

Single amino acid site directed mutagenesis was performed with the Q5-site directed mutagenesis kit from New England Biolabs following the manufacturer's instructions.

The first step was the mutagenesis of the wild type (Wt) Neo1 pDDGFP_LEU2D plasmid (1 µl plasmid 1-25 ng, 12.5 µl Q5-site directed mutagenesis kit master mix, 0.5 µM For primer, 0.5

μ M Rev primer, 9 μ l H₂O, T annealing and cycles number: according to the primer in **Tables 4.3 and 4.4**). Plasmid amplification was checked on a 0.8% w/v agarose gel. The reaction mix was digested with DpnI enzyme (1 μ l PCR product, 5 μ l Q5-site directed mutagenesis kit reaction buffer, 1 μ l DpnI, 3 μ l H₂O) for 5 min at 25°C. XL1-Blue cells were used to increase the plasmid yield as previously described (*Plasmid yield increase in E. coli and gene insertion control*). With the plasmid extracted from XL1-Blue cells, FGY217 or INVSc1 cells were transformed according to the previous protocol for the transformation in *S. cerevisiae*, with 1 μ g plasmid in 49 μ l of sterile water. Small-scale overexpression was then tested as already described (*Small-scale overexpression protocol*).

Based on the results of the expression test, mutants were overexpressed in large-scale in 0.5 l LEU- with addition of 2% w/v galactose.

- Membranes recovery:

Cells from large-scale overexpression were centrifuged at 6000 g and 4°C for 10 min. The pellet was resuspended in TBS buffer, centrifuged twice to remove medium and stored at -80°C. Then the pellet was thawed on ice and cells were resuspended in Neo1 resuspension buffer. Resuspension was homogenized at 20000 psi for 3 times. Cell lysate was centrifuged for 20 min at 6000 g and 4°C to eliminate debris. The supernatant was centrifuged for 1 h at 110000 g and 8°C. Membranes were recovered and resuspended in Neo1 buffer (2 ml buffer for 1 l culture). Membranes were then shattered in a potter until obtaining a homogeneous resuspension. The membrane GFP fluorescence was measured and membranes were stored at -80°C.

- Neo1 detergent test:

200 μ l Neo1 membranes (100 ml culture) were diluted to 2 ml in Neo1 solubilization buffer. Solubilization was performed 1 h or overnight at 4°C under slow agitation. Samples were then centrifuged at 12000 g for 1 h at 4°C and the supernatant was analyzed following the GFP fluorescence in solution and on SDS-PAGE with an 8% w/v gel. For the SDS-PAGE analysis samples were heated at 37°C with an additional 2% w/v SDS in order to avoid protein aggregation.

Detergents: OG 180 mM (10x CMC), DM 18 mM (10x CMC), DDM 1.2 mM (10x CMC), LDAO 20 mM (10x CMC) or 1% w/v (22x CMC), C12E8 1% w/v (200x CMC) (Hirayama et al. 2013). For detergent characteristics see **Table 4.1**.

- Protein purification:

In order to separate the target protein purification was performed using IMAC and SEC chromatography. At each purification membranes obtained from 1-2 l overexpression culture (corresponding to 2-4 ml membranes) were processed. Detergents were used as described in **Table 4.5**:

| | Solubilization | IMAC AND SEC Purification |
|---------------------------|--|--|
| LDAO purification | LDAO 10x CMC (20 mM) | LDAO 10x CMC (20 mM) |
| DDM purification | DM 20x CMC (36 mM) | DDM 5x CMC (0.6 mM) |
| C12E8 purification | C12E8 200x CMC (1% w/v) | C12E8 6x CMC (0,03% w/v) |
| LMNG purification | LMNG 100x CMC (1 mg ml ⁻¹) | IMAC purification in LMNG 10x CMC(0.1 mg ml ⁻¹), SEC purification in LMNG 3x CMC (0.03 mg ml ⁻¹) |

Table 4.5: Detergents used in Neo1 purification.

Membranes were solubilized in Neo1 solubilization buffer and 1 mM PMSF for 1 h. Solution was then centrifuged for 1 h at 70000 g and 8°C and the supernatant was recovered. 1 ml Talon Superflow resin was added for each ml of membrane suspension and incubated for 1 h at 4°C under slow agitation. Flow through was discharged and the resin was washed with 10 column volumes (CV) Neo1 solubilization buffer. A second wash with 10 CV Neo1 solubilization buffer and 5 mM imidazole was performed in order to remove non-specific binding. The protein was eluted with 0.5 CV of Neo1 solubilization buffer containing 150 mM imidazole. Elution was repeated 5 times until complete recovery of the protein.

The higher fluorescence elution fractions were pooled and the volume reduced to 800 µl. The specimen was then centrifuged at 12000 g for 15 min in order to remove protein

precipitate. Sample was injected into a Superdex 200 SEC column previously conditioned with Neo1 solubilization buffer. The SEC purification was performed at 0.4 ml min⁻¹ flow.

SEC fractions with higher fluorescence were collected and concentrated in a tube concentrator with cutoff 100 kDa. Protein concentration was determined both with UV spectroscopy at 280 nm and BCA protein assay.

Each step was checked with SDS-PAGE on an 8% w/v polyacrylamide gel and with GFP fluorescence.

The protocol was adjusted according to quality and quantity of the sample.

- Neo1 CD analysis

Neo1 purified protein 0.6-1-2 mg ml⁻¹ in DDM, LDAO and LMNG detergents, respectively, were placed in a high precision quartz cell with 0.1 mm path length and heated from 4°C to 80°C (Miles and Wallace 2016). Ellipticity was measured from 250 nm to 190 nm in order to follow α -helix signal. Five accumulations were collected every 2-5°C. Analyses were performed on a Jasco j-815 CD spectrometer.

- Neo1 crystallization trials

Neo1 purified protein at 0.8-1.2 mg ml⁻¹ in LMNG 0.03 mg ml⁻¹ solution was placed in 96-well plates for sitting drop with 80 μ l of MemGold1, MemGold2, MemSyst and MemStart kits (J. L. Parker and Newstead 2016). Protein solution drops of 0.2 μ l were dispensed using the Mosquito TTP Labtech. Plates were stored at 4°C and checked weekly.

- Neo1 negative staining and TEM analysis

Neo1 purified in LMNG 0.03 mg ml⁻¹ was diluted in a range from 40 μ g ml⁻¹ to 10 μ g ml⁻¹. 3 μ l protein solution were placed on a Formvar carbon grid for 30''. Formvar grids are specifically treated to enhance attachment of materials with low adhesion such as membrane proteins. The excess of solution was removed with a filter paper. 3 μ l of milliQ water were used to wash the grid. The procedure was repeated twice. 3 μ l of 2% w/v uranyl acetate solution

were placed on the grid for 1 min. The excess of solution was removed and grids were left to dry (Lyons et al. 2017).

The grids were analyzed with a TEM Philips EM208 with a QUEMESA camera using RADIUS software at 100 kV in defocus mode, images were taken at 50K, 80K or 100K magnification.

- *Neo1 activity assay*

Neo1 purified 0.03 mg ml⁻¹ solution in LMNG was tested with the ADP-Glo kinase assay kit Promega to determine its activity in presence of putative lipid substrates (PC, PE, PS). Crude *S. cerevisiae* membranes were used as control of Neo1 native substrate source as it is possible that its substrate differs from phospholipids used in this test (Montigny et al. 2017; Theorin et al. 2019).

25 µl of reaction solution contained: 25 ng or 50 ng purified Neo1, 150 mM NaCl, 50 mM Tris pH 7.5, 5 mM MgCl₂, 0.5 mM ATP, 1.2 mM DDM, 0.12 mg ml⁻¹ lipid or crude *S. cerevisiae* membranes.

Preliminary lipid exchange with the detergent for 20 min at room temperature (RT) under slow agitation was performed. The effective reaction was then triggered with addition of 0.5 mM ultrapure ATP, and solution was incubated 30 min at RT under slow agitation. Reaction was then stopped with addition of 25 µl ADP-Glo reagent. The obtained solution was incubated 40 min at RT under slow agitation. In this step the enzymes stop the ATP consumption by removing all the unreacted ATP. A further addition of 50 µl kinase detection reagent allowed for activity yield detection. In this step the reagent enzymes convert ADP to ATP. Solution incubation time was of 60 min at RT under slow agitation. The generated chemoluminescence was recorded on Tecan infinite m1000pro plate reader, with luminescence parameters of 100 ms integration time with automatic attenuation.

- *Mutants activity assays*

In order to test the effect of the selected mutations on Neo1 enzymatic activity, mutants in their native membranes were analyzed with different kits based on the luciferin/luciferase ATP reaction (DeLuca and McElroy 1974).

Total membrane protein concentration was determined through the BCA assay. GFP fluorescence was then recorded on solutions of the mutants at 1 mg ml⁻¹ total protein concentration.

The CellTiter-Glo 2.0 kit Promega was used to evaluate ATPase activity of the mutants in comparison with Neo1 wild type (WT) and *S. cerevisiae* membranes with no overexpressed protein. Orthovanadate Na₃VO₄, inhibitor of P-IV ATPases (Okumura and Kinoshita 2016), was used to set a “zero activity” level.

50 µl of reaction solution contained: 25 µg of total proteins, 150 mM NaCl, 50 mM Tris pH 7.5, 5 mM MgCl₂, 5 mM ATP, 1 mM Na₃VO₄ (when added) (Promega, n.d.).

The first step of the assay involves an incubation of 50 µl of reaction solution, with a previous 5 min incubation when Na₃VO₄ is added, for 40 min at 37°C under slow agitation. Then, 50 µl Cell-Titer Glo reagent were added to the solution and were incubated 10 min at RT under slow agitation. Luminescence was recorded on Biotek synergy h1 hybrid reader with default parameters.

4.7.2 *Wzy* protocols

- *RF cloning of Wzy from PAO1*

A first PCR reaction was performed to amplify *Wzy* gene from *P. aeruginosa* PAO1 genome and to create the megaprimer following the DNA polymerase manufacturer's instructions (0.2 mM dNTPs, 0.5 µM 5' phosphorylated For and Rev primers, shown in **Table 4.6**, 3 µl PAO1 genome as template, 1 µl Q5 high fidelity DNA polymerase, 5 µl 5X Q5 reaction buffer, H₂O up to 50 µl, T annealing 66°C, elongation time 2 min, 30 reaction cycles). The amplified DNA sequence was then checked and extracted from a 0.8% w/v agarose gel with Gel extraction kit. The obtained gene was inserted in pWaldo plasmid with PCR reaction (65 ng pWaldo_GFP, 10 mM dNTPs, 315 ng *Wzy* megaprimer, 1 µl Q5 high fidelity DNA polymerase, H₂O up to 25 µl, T annealing 62°C, elongation time 5 min, 30 cycles reaction). The obtained plasmid was then digested with DpnI enzyme for 2 h at 37°C, in order to remove all the parental plasmid with no *Wzy* insertion.

XL1-Blue *E. coli* competent cells were transformed with *Wzy* pWaldo_GFP plasmid (as described in Neo1 protocol *Plasmid yield increase in E. coli and gene insertion control*), and

plated on LB agar plates with 50 $\mu\text{g ml}^{-1}$ kanamycin. Plates were grown overnight at 37°C. Some colonies were selected and inoculated in liquid LB medium with 50 $\mu\text{g ml}^{-1}$ kanamycin, for plasmid extraction with Miniprep kit. The extracted plasmid was then checked for correct gene insertion with PCR using commercial T7 and T7-term primers.

All steps were controlled on an agarose 0.8% w/v gel.

| Primer | Sequence |
|------------|---|
| For | TTTTGTTTAACTTTAAGAAGGAGATATCTAGAATGTATATACTTGCTCGAGTCGACAGG |
| Rev | ACCCTGGAAGTACAGGTTTTTCACCTAGAGTTTTTCTAAAGACATCTTG |

Table 4.6: 5' phosphorylated For and Rev primers for Wzy RF cloning.

- *E. coli* competent cells

The rubidium chloride method was used to prepare competent *E. coli* cells (Roychoudhury, Basu, and Sengupta 2009).

Non competent cells were inoculated in 100 ml LB medium and grown at 37°C and 180 rpm to $\text{OD}_{600}=0.35-0.4$. Culture was then centrifuged at 3000 g and 4°C for 5 min, supernatant was discharged and cell pellet resuspended in 16 ml buffer A (12 g l^{-1} RbCl, 9.9 g l^{-1} $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 0.03 M KAc at pH 7.5, 1.5 g l^{-1} $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 150 g l^{-1} glycerol). Cells were then incubated for 15 min on ice. Centrifugation step was repeated, supernatant discharged and cells resuspended in 8 ml buffer B (0.01 M MOPS pH 6.8, 1.2 g l^{-1} RbCl, 11 g l^{-1} $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 150 g l^{-1} glycerol). The resulting solution was aliquoted into sterile tubes and flash frozen in liquid nitrogen. Cells were stored at -80°C. Cell competence was controlled through transformation with commercial pUC19 vector.

- Transformation in *E. coli* Lemo21(DE3) or BL21(DE3)pLysE cells

Transformation of competent cells was performed according to manufacturer's instruction. 1 μl of WzypWaldo_GFP plasmid (1-80 ng) was added to 50 μl competent Lemo21(DE3) or BL21(DE3)pLysE cells. Cells were incubated for 30 min on ice. Heat shock to allow plasmid transfer through the membrane was performed for 45'' at 42°C followed by 2 min incubation

on ice. 600 μ l LB medium were added to the cells, that were then grown 1 h at 37°C and 180 rpm shake. 100 μ l of cell resuspension was split in 5 ml LB with 50 μ g ml⁻¹ kanamycin and 34 μ g ml⁻¹ chloramphenicol. Cells were grown overnight at 37°C and 180 rpm shake, or plated on LB agar with 50 μ g ml⁻¹ kanamycin or 50 μ g ml⁻¹ kanamycin and 34 μ g ml⁻¹ chloramphenicol, for BL21(DE3)pLysE or Lemo21(DE3) culture, respectively.

Plasmids were extracted from cells with Miniprep kit and checked for correct plasmid insertion via PCR with commercial T7 and T7 term primers.

- Wzy expression

Small-scale expression

WzypWaldo_GFP plasmid was transformed in competent BL21(DE3)pLysE or Lemo21(DE3) cells following transformation protocol described before (*Transformation in E. coli Lemo21(DE3) or BL21(DE3)pLysE cells*).

Transformed cells were grown in 5 ml LB with kanamycin 50 μ g ml⁻¹ and chloramphenicol 34 μ g ml⁻¹, for BL21(DE3)pLysE and Lemo21(DE3) culture, at 37°C and 180 rpm shake overnight. Culture was then diluted in the same media at OD₆₀₀=0.1 and grown at 37°C and 180 rpm shake to OD₆₀₀=0.4 for Lemo21(DE3) and OD₆₀₀=0.6 for BL21(DE3)pLysE cells. BL21(DE3)pLysE cell medium was supplemented with 1% w/v or 2% w/v glucose, for basal protein expression inhibition, while Lemo21(DE3) medium was supplemented with different L-rhamnose concentrations (from 0 mM to 2 mM) for lysozyme production.

Protein expression was induced with IPTG addition in a range from 0.2 mM to 1 mM for BL21(DE3)pLysE cells, and with 0.4 mM IPTG for Lemo21(DE3).

Different growth temperatures were tested (20°C and 37°C) following the GFP fluorescence after few hours to overnight.

Large-scale expression in Lemo21(DE3) cells

Competent Lemo21(DE3) cells were transformed with Wzy pWaldo_GFP plasmid as previously described and grown overnight at 37°C and 180 rpm shake. Overnight culture was diluted in 500 ml LB with kanamycin 50 μ g ml⁻¹, chloramphenicol 34 μ g ml⁻¹ and L-rhamnose 0.25 mM and then grown to the optical density OD₆₀₀=0.4. Protein production was induced

with 0.4mM IPTG, and culture grown overnight at 20°C. The protein production yield was followed by GFP fluorescence analysis.

Membranes recovery

Cells from large-scale overexpression were centrifuged at 6000 g and 4°C for 10 min. Supernatant was discharged and the pellet resuspended in TBS buffer with 1 mM PMSF. Cells were centrifuged again and pellet was stored at -80°C. Then the pellet was thawed on ice and cells were resuspended in TBS buffer with 1 mM PMSF and lysozyme. Resuspension was incubated 30 min at RT.

Cells were homogenized at 20000 psi of pressure for 3 times. sample was centrifuged 20 min at 6000 g at 4°C to remove debris. Cell membranes were obtained by centrifugation of the recovered supernatant for 1 h at 110000 g and 8°C. Membranes were then resuspended in Wzy resuspension buffer. GFP fluorescence was recorded and membranes were stored at -80°C.

- *Wzy detergent test*

Solubilization test on 200 µl Wzy membranes (40 ml culture) was performed using the same protocol in *Neo1 detergent test*.

- *Wzy affinity chromatography test*

Wzy test in different affinity chromatography resins

To select the best purification, condition with the higher Wzy-binding efficiency were tested GFP-TrapA, Ni-NTA and Talon Superflow resins. Purification was performed according to manufacturer's instructions in presence of LDAO 10x CMC (20 mM) detergent in the solubilization buffer.

- GFP-TrapA: 1 ml membranes (0.2 l culture) were solubilized for 1 h at 4°C in 10 ml Wzy solubilization buffer. 75 µl of GFP-TrapA were added to the solution and incubated with slow agitation for 1 h at 4°C. Flow through was discharged and the resin was washed with 40 CV of Wzy solubilization buffer with 0.5 mM EDTA. A first elution step was done with 100 µl of GFP 0.13 mg ml⁻¹ and repeated for 4 times. Resin

was washed again with solubilization buffer. A second elution was performed with 50 μ l glycine 0.2 M pH 2.5 for 3 times. Elution solution was immediately diluted in 200 μ l of solubilization buffer with 0.5 mM EDTA in order to increase the acidic pH of the glycine solution.

- 300 μ l Ni-NTA resin were added to the protein solution and incubated under slow agitation for 1 h at 4°C. The resin was previously washed with 10 CV Wzy solubilization buffer and a second wash with 10 CV of the same buffer with 5 mM imidazole. Elution was performed in 1 CV Wzy solubilization buffer with 250 mM imidazole repeated 4 times.
- Talon Superflow resin: the procedure was the same as for Ni-NTA resin with a final elution in 1 CV solubilization buffer supplemented with 150 mM imidazole for 4 times.

Flow through, washes and eluates were analyzed with SDS-PAGE.

Test on different detergents with Co²⁺ resins

Different detergent conditions were tested on Talon Superflow resin (LDAO 20 mM, DM 18 mM, DDM 2.4 mM, CHAPS 80 mM and C12E8 1% w/v or 0.54 mM) (for detergent characteristics see **Table 4.1**). The detergent concentration was 10x CMC for all detergents during solubilization and purification, except for DDM that was 20x CMC and C12E8 that was 200x CMC (1% w/v) during solubilization and 6x CMC (0.03% w/v) for purification.

1 ml membranes (0.2 l culture) was solubilized in 10 ml of Wzy solubilization buffer (with detergent concentration as described above) for 1 h under slow agitation at 4°C. The purification procedure was performed as described before (*Wzy test in different affinity chromatography resins*).

- Wzy purification

Wzy purification was performed with the following detergent concentration in buffer:

- i) DM 10x CMC (18 mM);
- ii) DDM 5x CMC with a previous solubilization in LDAO 10x CMC (20 mM);
- iii) LDAO 10x CMC (20 mM)

Materials and methods

For each experiment 3-5 ml membranes were tested, corresponding to 0.6-1 l of culture.

Membranes were solubilized in Wzy solubilization buffer with 1 mM PMSF for 1 h at 4°C under slow agitation. The solution was then centrifuged for 1 h at 70000 g and 8°C, and the pellet was discarded.

0.4 ml Talon resin were added for each ml membrane and slowly shaken for 1 h at 4°C. Flow through was discharged. Washes were performed as described above (*Wzy test in different affinity chromatography resins*).

Wzy was then eluted with 1 x CV of Wzy solubilization buffer with 150 mM imidazole, elution was repeated 5 times until complete protein recovery.

Fluorescent fractions were pooled and concentrated. Sample was centrifuged at 12000 g for 15 min to pellet protein precipitates.

The sample was then injected in a Superdex200 SEC column previously conditioned with Wzy solubilization buffer (all buffers were filtered on a 0.2 µm membrane and degassed overnight). Purification was performed at 0.4 ml min⁻¹ flow.

Fluorescent fractions were collected and concentrated in tube concentrators, cutoff 50 kDa, to 1-3 mg ml⁻¹.

All the steps were controlled on SDS-PAGE. Furthermore the GFP fluorescence was followed for each step in order to control the protein presence and yield. Final protein amount was determined with UV spectroscopy 280 nm on SimpliNano and with BCA protein assay.

Variation of this protocol:

- In order to remove the second band at 30 kDa, the sample was washed with Wzy solubilization buffer with an increasing gradient of imidazole from 0 mM to 145 mM, before the standard elution with 150 mM imidazole.
- Instead of SEC column, MonoQ IEX column was tested under a NaCl gradient from 150 mM up to 1 M in LDAO 10x CMC buffer.
- Wzy CD analysis

Wzy purified protein in a range 2-3.7 mg ml⁻¹ in LDAO was analyzed in a high precision quartz cell with 0.1 mm path length and heated from 4°C to 80°C. Ellipticity was measured from 250 nm to 190 nm in order to follow α-helix peaks. Five accumulations were recorded for each 2°C step. Analyses were performed on a Jasco j-815 CD spectrometer. The sample in LDAO

10x CMC was additionally analyzed at 3.7 ml^{-1} in a 0.2 mm path length cell in order to improve signal intensity at 222 nm.

- Wzy fluorometry analysis

Wzy at a concentration of $0.75\text{-}0.9 \text{ mg ml}^{-1}$ was placed into a 1 ml volume quartz cuvette and heated from 10°C to 75°C . Tryptophan was excited at 295 nm and spectra were collected in the range 300-460 nm. Data were recorded with a Varian Cary eclipse fluorescence spectrophotometer equipped with a Peltier cell.

- Wzy for Raman spectroscopy

To obtain structural data of Wzy, UV resonant Raman analyses were performed on purified protein at 19°C and 70°C following protocols from literature (D'amico et al. 2013; Oladepo et al. 2012).

Raman measurements were performed on 1 mg ml^{-1} purified Wzy in LDAO 10x CMC solution, analyzed in a Suprasil quartz cuvette. Intensity of UV excitation was $40 \mu\text{V}$. Spectra were collected with 244 nm excitation wavelength in backscattering geometry by using a triple stage spectrometer (Trivista, Princeton Instruments) with 8 cm^{-1} spectral resolution, for three hours. To prevent sample photodamage, the cuvette was horizontally oscillated during measurement. Cuvette was placed in a thermostatic bath and the sample was analyzed initially at 19°C and after heating to 70°C .

The data were averaged and processed removing background, empty cell and buffer signals. Background, cell and buffer spectra were recorded with the same parameters as the Wzy sample. Calibration with cyclohexane as a standard was performed. Software Igor Pro Version: 7.0.8.1 2019 (Wavemetrics Corporation) was used to analyze and plot the results.

Raman measurements were performed at the BL10.2_IUVS beamline at the Elettra Synchrotron in Trieste.

The main bands are summarized in **Table 4.7**.

| Residue / Amide | Peak (cm ⁻¹) | Vibrational mode |
|-----------------|--------------------------|---|
| Trp | 1010 | W16 (symmetric pyrrole out-of-phase breathing) |
| Tyr | 1173 | Y9a |
| Tyr | 1209 | Y7a |
| Amide III | 1238 | C-N + N-H |
| Trp | 1342 | W7 (pyrrole ring vibration) |
| Trp | 1362 | |
| Trp+ Amide II | 1559 | W3 (C2-C3 stretching mode of the pyrrole ring) |
| Trp + Tyr | 1623 | W2Trp (phenyl mode with some contribution from the pyrrole N1C8 stretching) + Y8 (in plane ring stretching) |
| Amide I | 1666 | C=O |

Table 4.7: Bands visible in the Raman UV resonance spectra of Wzy.

- Wzy crystallization trials

Wzy was concentrated after purification in DM 18 mM or LDAO 20 mM (both 10x CMC) in a range of 2.7-3.7 mg ml⁻¹. 0.2 µl of Wzy solution were dispensed by Mosquito TTP Labtech in 96-well plates for sitting drop crystallization. The crystallization screens used were MemGold 1, MemGold 2, MemSyst and MemStart. Plates were stored at 4°C and 18°C and checked weekly. Three conditions led to crystal formation:

- 1) NaCl 100 mM, Glicine 50 mM pH 9.5, PEG 300 30% w/v in DM 10x CMC buffer (18 mM);
- 2) NaCl 100 mM, Bicine 50 mM pH 9.0, PEG 300 30% w/v in DM 10x CMC buffer (18 mM);
- 3) Sodium acetate tryhydrate 200 mM, hepes 100 mM pH 7.0, PEG 3000 22% w/v, KCl 200 mM in LDAO 10x CMC buffer (20 mM).

Conditions that led to crystals formation were optimized with variations in pH, PEG and protein concentration with sitting drop and hanging drop vapor diffusion techniques.

Variations on condition 1: Glicine 50 mM pH 9.5, PEG 300 in a range 24-39% w/v, NaCl in a range 0.08-0.15 M, Wzy concentration in a range 2.7-3 mg ml⁻¹. Detergent for purification DM 10x CMC (18 mM).

Variations on condition 2: Bicine 50 mM pH 9.0, PEG 300 in a range 24-39% w/v, NaCl in a range 0.08-0.15 M, Wzy concentration in a range 2.7-3 mg ml⁻¹. Detergent for purification DM 10x CMC (18 mM).

Variations on condition 3: Sodium acetate trihydrate 200 mM, KCl 200 mM, Hepes 100 mM in a pH range 6.6-8, PEG 3000 in a range 16-30% w/v, Wzy concentration in a range 2.7-3.7 mg ml⁻¹. Detergent for purification LDAO 10x CMC (20 mM).

Once formed, crystals were collected with nylon loops and flash-frozen in liquid nitrogen. When necessary, 20% PEG 300 was added as cryoprotectant.

Crystals were analyzed at the XRD2 beamline of the Elettra synchrotron of Trieste, at a wavelength of 1 Å. The beamline detector is a Dectris Pilatus 6M. Data were analyzed with Mosflm software to obtain the cell parameters of each crystal (Battye et al. 2011).

- Negative staining and EM analysis of Wzy

3 µl of 15 µg ml⁻¹ purified Wzy in LDAO 20 mM were placed on a carbon grid and blotted on a filter paper after 10'' to remove excess of solution. 3 µl of 2% w/v uranyl acetate were added and removed after 30'' with a filter paper. The staining step was repeated twice and the grid was left to dry.

Grids were analyzed with EM JEOL JEM2100 LaB6, pictures were recorded using a GATAN 2K × 2K camera, at 32.5K magnification. The spherical aberration (Cs factor) value for this instrument is 1.6 mm. Recorded images have pixel dimension of 4.3 Å.

REFERENCES

- 1) Andersen, Jens P. et al. 2016. "P4-ATPases as Phospholipid Flippases-Structure, Function, and Enigmas." *Frontiers in Physiology* 7(JUL): 1–23.
- 2) Batty, T. Geoff G. et al. 2011. "IMOSFLM: A New Graphical Interface for Diffraction-Image Processing with MOSFLM." *Acta Crystallographica Section D: Biological Crystallography* 67(4): 271–81.
- 3) Castorena-Torres, Fabiola, Katia Peñuelas-Urquides, and Mario Bermúdez de León. 2016. "Site-Directed Mutagenesis by Polymerase Chain Reaction." *Intech*: 159–73.
- 4) D'amico, Francesco et al. 2013. "UV Resonant Raman Scattering Facility at Elettra." *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 703: 33–37.
- 5) DeLuca, Marlene, and William D. McElroy. 1974. "Kinetics of the Firefly Luciferase Catalyzed Reactions." *Biochemistry* 13(5): 921–25.
- 6) Dessau, Moshe A., and Yorgo Modis. 2011. "Protein Crystallization for X-Ray Crystallography." *Journal of Visualized Experiments* 9(47): 1–6.
- 7) Dilworth, Marvin V. et al. 2018. "Microbial Expression Systems for Membrane Proteins." *Methods* 147(March): 3–39.
- 8) Drew, David et al. 2008. "GFP-Based Optimization Scheme for the Overexpression and Purification of Eukaryotic Membrane Proteins in *Saccharomyces Cerevisiae*." *Nature Protocols* 3(5): 784–98.
- 9) Drew, David E., Gunnar Von Heijne, Pär Nordlund, and Jan Willem L De Gier. 2001. "Green Fluorescent Protein as an Indicator to Monitor Membrane Protein Overexpression in *Escherichia Coli*." *FEBS Letters* 507(2): 220–24.
- 10) Eddy, Sean R. 2004. "Where Did the BLOSUM62 Alignment Score Matrix Come From?" *Nature Biotechnology* 22(8): 1035–36.
- 11) Van Den Ent, Fusinita, and Jan Löwe. 2006. "RF Cloning: A Restriction-Free Method for Inserting Target Genes into Plasmids." *Journal of Biochemical and Biophysical Methods* 67(1): 67–74.
- 12) Gerdes, Hans Hermann, and Christoph Kaether. 1996. "Green Fluorescent Protein: Applications in Cell Biology." *FEBS Letters* 389(1): 44–47.
- 13) Glaser, Michael, and S. J. Singer. 1971. "Circular Dichroism and the Conformations of Membrane Proteins. Studies with Red Blood Cell Membranes." *Biochemistry* 10(10): 1780–87.
- 14) Green, Michael R., and Joseph Sambrook. 2018. "The Basic Polymerase Chain Reaction (PCR)." *Cold Spring Harbor Protocols* 5: 338–45.
- 15) Greenfield, Norma J. 2006. "Using Circular Dichroism Spectra to Estimate Protein Secondary Structure." *Nature Protocols* 1(6): 2876–90.
- 16) Hartley, James L. 2006. "Cloning Technologies for Protein Expression and Purification." *Current Opinion in Biotechnology* 17(4): 359–66.
- 17) Hirayama, Hiroshi et al. 2013. "ATPase Activity of Human ABCG1 Is Stimulated by Cholesterol and Sphingomyelin." *Journal of Lipid Research* 54(2): 496–502.

Materials and methods

- 18) Huang, Xiaoqiu, and Webb Miller. 1991. "A Time-Efficient, Linear-Space Local Similarity Algorithm." *Advances in Applied Mathematics* 12(3): 337–57.
- 19) Kawai, Shigeyuki, Wataru Hashimoto, and Kousaku Murata. 2010. "Transformation of *Saccharomyces Cerevisiae* and Other Fungi: Methods and Possible Underlying Mechanism." *Bioengineered bugs* 1(6): 395–403.
- 20) Khan, Hidayat Ullah. 2012. "The Role of Ion Exchange Chromatography in Purification and Characterization of Molecules." *Intech Ion exchan*: 331–42.
- 21) Kleywegt, Gerard J., and T. Alwyn Jones. 1997. "Model Building and Refinement Practice." *Methods in Enzymology* 277(1995): 208–30.
- 22) Kota, Jhansi, C. Fredrik Gilstring, and Per O. Ljungdahl. 2007. "Membrane Chaperone Shr3 Assists in Folding Amino Acid Permeases Preventing Precocious ERAD." *Journal of Cell Biology* 176(5): 617–28.
- 23) Kunji, Edmund R.S., Marilyn Harding, P. Jonathan G. Butler, and Pearl Akamine. 2008. "Determination of the Molecular Mass and Dimensions of Membrane Proteins by Size Exclusion Chromatography." *Methods* 46(2): 62–72.
- 24) Lee, Pei Yun, John Costumbrado, Chih Yuan Hsu, and Yong Hoon Kim. 2012. "Agarose Gel Electrophoresis for the Separation of DNA Fragments." *Journal of Visualized Experiments* (62): 1–5.
- 25) Lyons, Joseph A., Andreas Bøggild, Poul Nissen, and Jens Frauenfeld. 2017. 594 *Methods in Enzymology Saposin-Lipoprotein Scaffolds for Structure Determination of Membrane Transporters*. 1st ed. Elsevier Inc.
- 26) Mathieu, Jacques, Emilia Alvarez, and Pedro J.J. Alvarez. 2014. "Recombination-Assisted Megaprimer (RAM) Cloning." *MethodsX* 1(1): 23–29.
- 27) Miles, A. J., and B. A. Wallace. 2016. "Circular Dichroism Spectroscopy of Membrane Proteins." *Chemical Society Reviews* 45(18): 4859–72.
- 28) Mohanty, Arun K., and Michael C. Wiener. 2004. "Membrane Protein Expression and Production: Effects of Polyhistidine Tag Length and Position." *Protein Expression and Purification* 33(2): 311–25.
- 29) Montigny, Cédric et al. 2017. "Slow Phospholipid Exchange between a Detergent-Solubilized Membrane Protein and Lipid-Detergent Mixed Micelles: Brominated Phospholipids as Tools to Follow Its Kinetics." *PLoS ONE* 12(1): 1–16.
- 30) Mozo-Villarías, A. 2002. "Second Derivative Fluorescence Spectroscopy of Tryptophan in Proteins." *Journal of Biochemical and Biophysical Methods* 50(2–3): 163–78.
- 31) Nixon, W. C. 1971. "The General Principles of Scanning Electron Microscopy." *Philosophical Transactions of the Royal Society B: Biological Sciences* 261(837): 45–50.
- 32) Ohi, Melania, Ying Li, Yifan Cheng, and Thomas Walz. 2004. "Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy." *Biological Procedures Online* 6(1): 23–34.
- 33) Okumura, Masaki, and Toshinori Kinoshita. 2016. "Measurement of ATP Hydrolytic Activity of Plasma Membrane H⁺-ATPase from *Arabidopsis Thaliana* Leaves." *Bio-Protocol* 6(23): e2044.
- 34) Oladepo, Sulayman A. et al. 2012. "UV Resonance Raman Investigations of Peptide and Protein Structure and Dynamics." *Chemical Reviews* 112(5): 2604–28.
- 35) Olson, Bradley J.S.C., and John Markwell. 2007. "Assays for Determination of Protein Concentration."

Materials and methods

- Current Protocols in protein science* 48(1): 3.4.1-3.4.29.
- 36) Parker, Joanne L., and Simon Newstead. 2014. "Method to Increase the Yield of Eukaryotic Membrane Protein Expression in *Saccharomyces Cerevisiae* for Structural and Functional Studies." *Protein Science* 23(9): 1309–14.
 - 37) Parker, Joanne L, and Simon Newstead. 2016. "The Next Generation in Membrane Protein Structure Determination." *Advances in Experimental Medicine and Biology* 922: 61–72.
 - 38) Parker, M. W. 2003. "Protein Structure from X-Ray Diffraction." *Journal of Biological Physics* 29(4): 341–62.
 - 39) Pearson, William R. 2013. "An Introduction to Sequence ('Homology') Searching." *Current protocols in Bioinformatics* 1: 03.
 - 40) Pomorski, Thomas et al. 2003. "Drs2p-Related P-Type ATPases Dnf1p and Dnf2p Are Required for Phospholipid Translocation across the Yeast Plasma Membrane and Serve a Role in Endocytosis." *Molecular Biology of the Cell* 14: 1240–54.
 - 41) Promega, Corporation. "Pgp-Glo(TM) Assay Systems Technical Bulletin TB341." : 608–277.
 - 42) Pronk, Jack T. 2002. "Auxotrophic Yeast Strains in Fundamental and Applied Research." *Applied and Environmental Microbiology* 68(5): 2095–2100.
 - 43) Rath, Arianna, and Charles M. Deber. 2013. "Correction Factors for Membrane Protein Molecular Weight Readouts on Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis." *Analytical Biochemistry* 434(1): 67–72.
 - 44) Roychoudhury, Aryadeep, Supratim Basu, and Dibyendu N. Sengupta. 2009. "Analysis of Comparative Efficiencies of Different Transformation Methods of *E. Coli* Using Two Common Plasmid Vectors." *Indian Journal of Biochemistry and Biophysics* 46(5): 395–400.
 - 45) Schwede, Torsten, Jürgen Kopp, Nicolas Guex, and Manuel C. Peitsch. 2003. "SWISS-MODEL: An Automated Protein Homology-Modeling Server." *Nucleic Acids Research* 31(13): 3381–85.
 - 46) Seddon, Annela M., Paul Curnow, and Paula J. Booth. 2004. "Membrane Proteins, Lipids and Detergents: Not Just a Soap Opera." *Biochimica et Biophysica Acta - Biomembranes* 1666(1–2): 105–17.
 - 47) Stetsenko, Artem, and Albert Guskov. 2017. "An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization." *Crystals* 7(7): 1–16.
 - 48) Theorin, Lisa et al. 2019. "The Lipid Head Group Is the Key Element for Substrate Recognition by the P4 ATPase ALA2: A Phosphatidylserine Flippase." *Biochemical Journal* 476(5): 783–94.
 - 49) Thomas, George J. 1999. "Raman Spectroscopy of Protein and Nucleic Acid Assemblies." *Annual Review of Biophysics and Biomolecular Structure* 28: 1–27.
 - 50) Zhang, Lei, Huimin Tong, Mark Garewal, and Gang Ren. 2013. "Optimized Negative-Staining Electron Microscopy for Lipoprotein Studies." *Biochimica et Biophysica Acta - General Subjects* 1830(1): 2150–59.

ACKNOWLEDGEMENTS

I acknowledge all the people that helped me in this work during the last three years.

In particular I would like to acknowledge Professor Rita De Zorzi to be such a motivating and inspiring boss.

I acknowledge Professor Federica Tramer, together with Odeta Kalaja, of the Department of Life Sciences of the University of Trieste for the help in the enzymatic assays and for the illuminating discussions on many aspects of this thesis.

I thank Professor Chaterine Venien-Bryan for the opportunity she gave me to analyze my samples at the EM facility of the Pierre et Marie Curie University of the Sorbonne Université in Paris, and Doctor Paolo Bertoncin of the Department of Life Sciences of the University of Trieste for all the time he spent on the EM analysis of my samples in Trieste.

A special thanks is for the people of the XRD2 beamline of the Elettra Synchrotron, in particular to Dr Annie Heroux, and to Dr Nicola Demitri, always present and helpful in the moment of need.

I acknowledge Dr Silvia Di Fonzo, Maria Pachetti and all the people of the IUVS beamline of the Elettra Synchrotron for the Raman analysis. A special thanks is for Professor Attilio Cesaro, for his illuminating suggestions. I want also to thank Dr Giovanni Birarda and Dr Federica Piccirilli of the SSSI beamline of the Elettra Synchrotron for helpful discussions.

I thank Dr Riccardo Sgarra for the mass spectrometry analysis: without his results the identification of the contaminant would have been impossible.

I want to acknowledge the people in the Elettra Biolab for the use of different instruments and all the suggestions they gave me.

Finally, a special acknowledgement goes to Dr Laura Destefanis for being a great mentor at the beginning of my thesis and for all the preliminary work on P-IV ATPases, and to Dr Sabrina Semeraro for preliminary work on Wzy and for her motivational spirit that encouraged me in these years. I would also like to thank Dr Barbara Medagli, whose help has been precious for the success of this thesis.