UNIVERSITÀ DEGLI STUDI DI TRIESTE

Università Ca'Foscari Venezia

UNIVERSITÀ DEGLI STUDI DI TRIESTE

e

UNIVERSITÀ CA' FOSCARI DI VENEZIA

XXXVIII CICLO DEL DOTTORATO DI RICERCA IN CHIMICA

From bacterial resistance to vulnerability: structural characterization of a *P. aeruginosa* flippase and a phage depolymerase targeting *K. pneumoniae*.

Settore scientifico-disciplinare: CHIM/03

DOTTORANDO
Marco **DE CONTO**

COORDINATORE
Prof. Enzo **ALESSIO**

SUPERVISORE DI TESI
Prof.ssa Rita **DE ZORZI**

CO-SUPERVISORE DI TESI
Prof.ssa Sara **FORTUNA**

ANNO ACCADEMICO 2024/2025

Abstract

Antimicrobial resistance (AMR) is one of the most pressing threats to global health, underscoring the need for novel strategies to combat multidrug-resistant bacteria. This thesis presents two structural biology projects addressing this challenge through the study of key bacterial and bacteriophage proteins. The first part focuses on the O-antigen flippase Wzx from *Pseudomonas aeruginosa*, an integral membrane protein essential for lipopolysaccharide biosynthesis and bacterial virulence. The protein was recombinantly expressed and purified, and its biochemical behavior was investigated using biophysical analyses. Crystallization trials and cryo-electron microscopy experiments were initiated to characterize its structure. Unfortunately, both structural techniques presented challenges that could not be overcome at this time. However, interesting insights into the protein stability and mechanism of action were obtained from molecular dynamics simulations of an AlphaFold model. The second part of the thesis concerns the structural and functional characterization of the endoglycosidase of bacteriophage ϕBO1E, active against the capsular polysaccharide (CPS) of *Klebsiella pneumoniae* KpB-1, a highly virulent strain. The enzyme was successfully expressed, purified, and crystallized. X-ray diffraction data collected at the Elettra synchrotron allowed the determination of its structure at 1.3 Å resolution. Structural comparison with homologous phage depolymerases identified a potential catalytic site located at the interface between monomers. Complementary cryo-electron microscopy analyses of both the apo and substrate-bound forms of the endoglycodisase are underway to further elucidate its catalytic mechanism. Together, these studies contribute to the understanding of bacterial surface assembly mechanisms and phage-derived enzymatic systems, providing a structural framework for the future design of novel antimicrobial agents based on membrane protein inhibition or phage enzyme engineering.

Riassunto

La resistenza antimicrobica (AMR) rappresenta una delle minacce più urgenti per la salute globale, ed è evidente la necessità di sviluppare nuove strategie per contrastare i batteri multiresistenti. Questa tesi presenta due progetti di biologia strutturale che affrontano tale sfida attraverso lo studio di proteine batteriche e fagicche di fondamentale rilevanza. La prima parte è dedicata alla flippasi dell'O-antigene Wzx di *Pseudomonas aeruginosa*, una proteina di membrana integrale essenziale per la biosintesi dei lipopolisaccaridi e per la virulenza batterica. La proteina è stata espressa e purificata, e il suo comportamento biochimico è stato analizzato mediante tecniche biofisiche. Sono stati avviati tentativi di cristallizzazione ed esperimenti di crio-microscopia elettronica per caratterizzarne la struttura. Sfortunatamente, entrambe le tecniche strutturali hanno presentato difficoltà che non è stato possibile superare in questa fase. Tuttavia, simulazioni di dinamica molecolare basate su un modello AlphaFold hanno fornito interessanti indicazioni sulla stabilità e sul possibile meccanismo d'azione della proteina. La seconda parte della tesi riguarda la caratterizzazione strutturale e funzionale dell'endoglicosidasi del batteriofago ϕBO1E, attiva contro il polisaccaride capsulare (CPS) di *Klebsiella pneumoniae* KpB-1, un ceppo altamente virulento. L'enzima è stato espresso, purificato e cristallizzato con successo. I dati di diffrazione ai raggi X raccolti presso il sincrotrone Elettra hanno permesso di determinarne la struttura con una risoluzione di 1.3 Å. Il confronto strutturale con depolimerasi fagicche omologhe ha permesso di identificare un potenziale sito catalitico localizzato all'interfaccia tra i monomeri. Analisi complementari di crio-microscopia elettronica, condotte sia sulla forma apo che su quella legata al substrato dell'endoglicosidasi, sono attualmente in corso per chiarirne ulteriormente il meccanismo catalitico. Nel complesso, questi studi contribuiscono alla comprensione dei meccanismi di assemblaggio delle superfici batteriche e dei sistemi enzimatici derivati da fagi, fornendo una base strutturale per la progettazione futura di nuovi agenti antimicrobici basati sull'inibizione di proteine di membrana o sull'ingegnerizzazione di enzimi fagici.

Table of Contents

Introduction

Antimicrobial resistance (AMR) has emerged as one of the most critical threats to global health in the 21st century. Once easily cured by antibiotics, infections are becoming difficult or impossible to treat due to the rise of drug-resistant bacteria. Across the world, multidrug-resistant pathogens are causing prolonged illnesses, higher healthcare costs, and increased mortality. The World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC, USA) have warned that without urgent action we may be approaching a "post-antibiotic" era in which common infections and minor injuries can once again be deadly [1].

Drug resistance

Antibiotics revolutionized medicine in the 20th century, drastically reducing deaths from infectious diseases. However, their success has eventually been undermined by the ability of microbes to develop resistance. AMR is nowadays a major global health crisis, threatening the effectiveness of a wide range of antibacterial therapies. A recent comprehensive analysis estimated that bacterial AMR directly caused 1.27 million deaths in 2019 and was associated with nearly 5 million deaths worldwide [1], making drug-resistant infections a leading cause of death. Earlier forecasts from the O'Neill report, a 2016 review funded by the Wellcome Trust and the UK Department of Health, warned that, without intervention, by 2050 up to 10 million people per year could die from resistant infections, surpassing cancer mortality (Figure 1). Although such long-term estimates involve some degree of uncertainty, the trajectory is undeniably alarming.

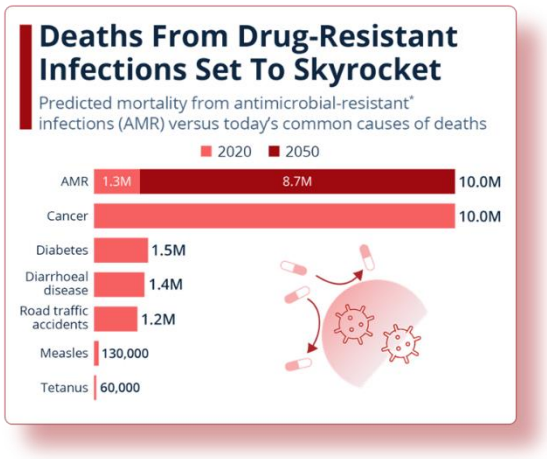


Figure 1 – An article from Statista, a global data and business intelligence platform based in Germany, sounding the alarm on the spread of drug-resistant infections. The reported prediction relies on an estimation of the UN 2023 report 'Bracing for Superbugs' (extracted from statista.com/chart/3095/drug-resistant-infections/)

AMR causes not only deaths but also major economic harm. The treatments for resistant infections require long hospitalizations, expensive drugs, and involve elaborate safety

precautions. The rise of drug-resistant bacteria represents a risk for many modern medicine procedures, such as surgery, chemotherapy, and organ transplants, whose outcomes rely heavily on effective antibiotics. From an economic point of view, the World Bank estimates that, if antibiotic resistance continues to increase, by 2030 its impact on the world economy could amount to trillions of dollars, with the heaviest impact on low- and middle-income countries that have fewer resources to fight it [1]. The main factors driving AMR arise from inadequate public health policies that lead to the misuse of antibiotics. The common practice of prescribing antibiotics for viral infections, using broad-spectrum drugs unnecessarily, or the noncompliance of patients not finishing their treatments, and the extensive use of antibiotics in animal breeding, all play a role in creating selective pressure that speeds up the spread of resistance. At the same time, good public health policies of preventive medicine, such as vaccination, sanitation, and hospital hygiene, can lower the number of infections, reducing the need for antibiotics and addressing AMR at its source.

The "ESKAPE"

Over the past decades, several bacterial pathogens have developed resistance mechanisms against multiple drugs, becoming a serious public health concern. In particular, the "ESKAPE" group [2] was singled out for its notorious ability to escape antibiotic action and cause infections that are especially difficult to treat. ESKAPE is the acronym of six highly virulent and antibiotic-resistant pathogens: *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter spp.* [3] Bacteria in this group combine high virulence and remarkable resistance to multiple antibiotics, making them a leading cause of deadly hospital infections (Figure 2).

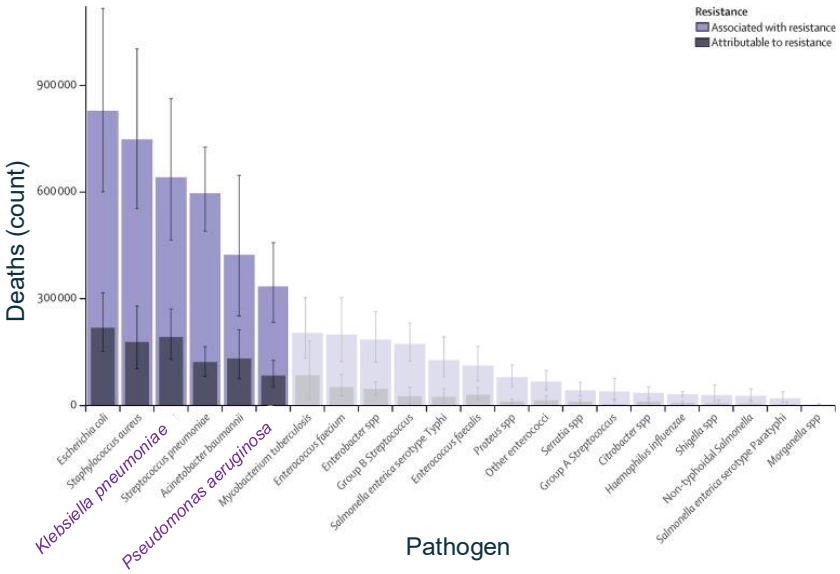


Figure 2 - Death caused by pathogens (from the report of Lancet 2022)

Mechanisms of resistance

Bacteria develop resistance to the action of antibiotics through several mechanisms, which are either intrinsic, i.e., inherent to some physiological characteristics of the strain, or acquired via a genome change, i.e., mutation or horizontal gene transfer. Resistance strategies include drug inactivation, a reduced uptake of the drug, activation of the efflux pump, mutation of the drug target, or biofilm formation (Figure 3). To avoid the antibiotic effect, a single organism can set up multiple mechanisms in parallel.

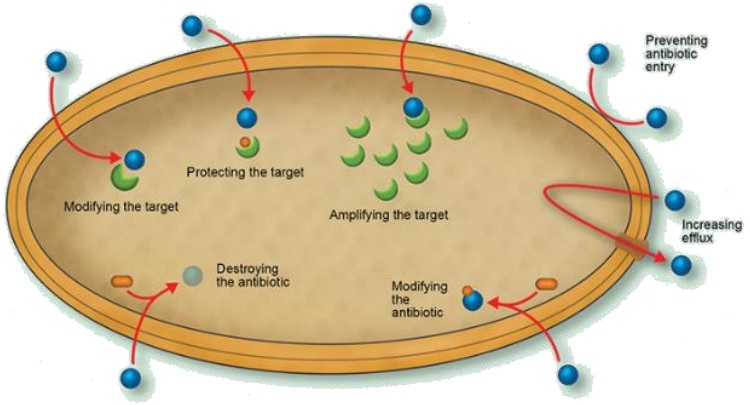


Figure 3 - An overview of the mechanisms of antibiotic resistance: protecting, modifying, and amplifying the target, preventing antibiotic entry, increasing efflux, and modifying or destroying the antibiotic (from www.open.edu/openlearncreate/mod/oucontent/)

In mechanisms involving the inactivation of the drug, bacteria become resistant to a specific antibiotic by producing enzymes that degrade or modify the antibiotic molecule, making it ineffective. There are a variety of known enzymes able to perform this task. The most emblematic example is the class of β-lactamases. These enzymes confer resistance to β-lactamic antibiotics and are present particularly in Gram-negative bacteria. β-lactamases modify penicillins, cephalosporins, and similar drugs by hydrolyzing their β-lactam ring, thus producing a harmless derivative. Other mechanisms that bacteria use to reduce the cytoplasmic concentration of antibiotics are the reduction of the permeability of the membrane or the activation of efflux pumps, which are membrane proteins (MPs) that expel a broad class of compounds by active transport. Some efflux pumps are selective for a single class of toxic compounds, while others can eject a wide range of structurally different molecules. A different class of resistance mechanisms involves the target of the antibiotic, rather than the drug itself. Among these, the modification of the antibiotic target through genetic mutations or the overexpression of the target protein represent common strategies. An example comes from the modification of the ribosome complex, preventing the binding of macrolides, tetracyclines, or oxazolidinones. From a therapeutic perspective, mechanisms such as the reduction of drug uptake or the activation of efflux pumps do not make the drug

completely ineffective, but require an increase in the dose of antibiotic to be administered. The bacterial cell wall plays a key role in these resistance strategies.

Bacterial cell wall

The bacterial cell wall has evolved to protect cells from environmental stress, maintain their shape, and regulate the exchange of molecules with the surrounding environment. Bacteria are classified as Gram-positive or Gram-negative, following the staining method developed by Christian Gram in 1884. Albeit the main difference between the two classes involves the cell wall, the distinction was made before its structure was known in detail and relies on the response of the colonies when stained with the Gram dye, also known as crystal violet. While the first class, the Gram-negative bacteria, does not appear colored upon removal of the dye, displaying a light pink hue, the second class retains the drug and shows a purple coloration. The reasons for this different behavior became clear once the characteristics of the membrane were elucidated. Gram-positive bacteria have a single phospholipid membrane surrounded by a thick peptidoglycan layer (30–100 nm) (Figure 4), mainly formed by N-acetylglucosamine and N-acetylmuramic acid units cross-linked by peptides.

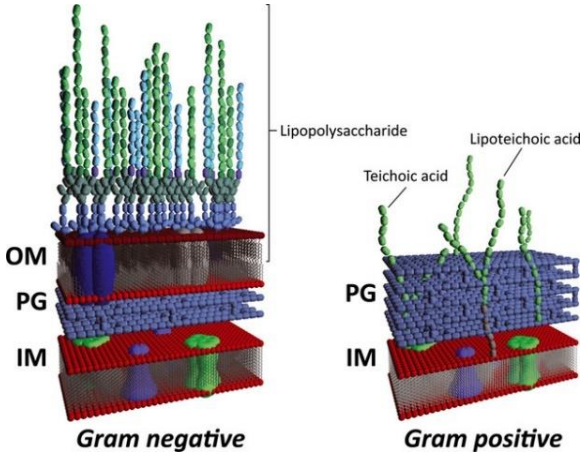


Figure 4 - Cell envelope in Gram-negative and Gram-positive bacteria. IM, inner membrane; PG, peptidoglycan; OM, outer membrane (from Islam et Lam, 2013).

The peptidoglycan layers on the surface of Gram-positive bacteria are the feature that allows the retention of the Gram dye. Embedded within this matrix are teichoic acids, which are polymers of glycerol phosphate and carbohydrate units linked together by phosphodiester bonds. Teichoic acids contribute to structural stability, ion regulation, and antibiotic resistance. The cell wall also contains various surface proteins involved in adhesion, nutrient acquisition, phage binding, and evasion of host defenses, making it a very complex and highly regulated system. The envelope of Gram-negative bacteria, on the other hand, is composed of an inner membrane, a thinner peptidoglycan layer located in the periplasmic space, and an outer membrane (Figure 4). The outer membrane is a highly asymmetric structure, characterized by lipopolysaccharides (LPS) anchored to the outer leaflet and phospholipids forming the inner one. This barrier limits

the entrance of toxic compounds and is linked to the peptidoglycan by lipoproteins. Proteins present in the outer membrane, namely porins, mediate the selective transport of molecules and ions. The periplasmic space hosts a dense network of enzymes, chaperones, and other proteins involved in nutrient processing and protection, while the inner membrane carries out essential functions such as energy production, lipid synthesis, and protein secretion. All the structural adaptations of Gram-positive and Gram-negative bacteria aim at allowing them to thrive in diverse and often hostile environments, balancing protection with the selective exchange of molecules necessary for survival. Considering their different structure, the limitation of the uptake of drugs in Gram-positive and negative bacteria follows different routes. In Gram-negative bacteria, the LPS acts as a barrier against exogenous and potentially dangerous molecules. Large or hydrophobic molecules cannot penetrate this layer, while small hydrophilic ones are usually let in through porin channels. Thus, Gram-negative bacteria can control the drug entrance by reducing the number of porins or altering their specificity.[4], [5] On the other hand, Gram-positive bacteria can regulate the thickness of their cell wall to block the entrance of antibiotics (e.g., vancomycin)[5], [6].

The LipoPolySaccharide

Lipopolysaccharide (LPS) is a fundamental structural and functional component of the outer leaflet of the outer membrane of Gram-negative bacteria. It represents a primary defense barrier, shielding the cell from toxic hydrophobic compounds such as antibiotics, bile salts, and antimicrobial peptides [7]. The LPS contributes to environmental persistence, host colonization, and pathogenesis, making it one of the most critical factors in bacterial survival and virulence. The human innate immune system is highly sensitive to this molecule, and LPS is recognized as the main trigger of endotoxic shock during Gram-negative infections [8]. The LPS is composed of three major structures: the lipid A, an oligosaccharide core, and the O-antigen (Figure 5).

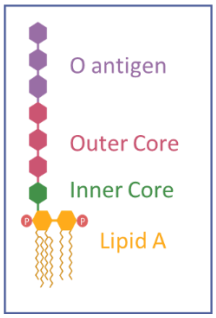


Figure 5 - Schematic representation of LPS molecule

Lipid A is a glucosamine disaccharide carrying six or seven acyl chains that anchor the LPS to the outer membrane. It represents the endotoxic moiety of the whole LPS molecule and is highly conserved among Gram-negative species [8]. Lipid A is responsible for causing strong inflammatory responses via macrophages and dendritic cells. Chemical modifications of its structure, such as altered acylation or the addition

of positively charged residues, can decrease immune recognition and enhance bacterial resistance to antimicrobial peptides [9]. The lipid A is synthesized in the inner membrane of bacteria, where it becomes the acceptor of the core oligosaccharide. The core oligosaccharide connects lipid A to the O-antigen and is composed of a conserved inner core, typically containing heptose and 3-deoxy-D-manno-octulosonic acid (Kdo) residues [10], and a variable outer core. Its biosynthesis involves a series of glycosyltransferases and phosphorylation reactions, mediated by WaaA and WaaP, and essential for LPS transport and outer membrane assembly. The O-antigen is a polysaccharide chain formed by repeated oligosaccharide units of 3-8 sugars [11] and represents the most exposed portion of the bacterial surface. Its structure varies significantly among different species, and even strains of the same species, which allows the classification known as bacterial serotyping. The O-antigen enhances virulence by preventing complement-mediated responses and phagocytosis, promoting adhesion, and protecting against oxidative stress. LPS molecules are densely packed, often stabilized by the presence of ion bridges (Mg^{2+}), creating a strong impermeable outer layer. Bacterial strains producing an LPS lacking the O-antigen, known as *rough* strains, exhibit greater susceptibility to environmental stress and antimicrobial agents. The hydrophobic resistance at the level of the outer membrane combined with the hydrophilic exclusion at the inner membrane creates the asymmetric permeability of the Gram-negative envelope, a highly efficient barrier that limits the entrance of exogenous molecules (e.g., antibiotics). When the LPS synthesis or assembly is disrupted, the membrane stability collapses, leading to protein misfolding, activation of stress response pathways and, in some cases, bacterial death. This dependency is exemplified by *Pseudomonas aeruginosa*, which cannot proliferate without a functional LPS [12]. Given the role of the LPS in bacterial viability and resistance, the inhibition of its biosynthesis or assembly is a promising antibacterial strategy. Targeting this pathway could destabilize the bacterial envelope, enhance antibiotic susceptibility, and provide new therapeutic strategies against multidrug-resistant Gram-negative pathogens.

Pseudomonas aeruginosa

Pseudomonas aeruginosa is a non-fermentative, Gram-negative, rod-shaped bacterium that survives in a wide range of environments. After its isolation from blue-green pus in wound infections, this bacterial species was first described in the 1880s by Gessard, who recognized it as a major opportunistic pathogen [13]. From the early 20th century, *P. aeruginosa* has been strongly associated with hospital-acquired infections (HAIs) and, at present, it represents the fourth cause of nosocomial diseases [14], particularly among immunocompromised individuals such as AIDS, cancer, or cystic fibrosis (CF) patients, or patients presenting extensive burns [15], [16], [17]. The clinical spectrum of *P. aeruginosa* infection is wide and includes intra-abdominal and urinary tract infections [18], [19], ocular infections such as ulcerative keratitis linked to contact lens use [20], [21], bacteremia in burn patients, and chronic pulmonary infections in CF patients [22],

[23]. The ability of *P. aeruginosa* to colonize different surfaces and materials, such as hospital equipment, water supplies, and implanted biomaterials, further increases the risk of transmission through both contact with environmental contamination and patient-to-patient interaction [24], [25]. The environmental persistence of this species, combined with a remarkable resistance to antimicrobial therapy, contributes to the spread of multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains in healthcare facilities. In a prospective study from the University Hospital in Wroclaw, Poland, *P. aeruginosa* accounted for 13.5% of all healthcare device-associated infections, most frequently ventilator-associated pneumonia, followed by urinary tract infections and central line-associated bloodstream infections [18]. These findings are consistent with broader European surveillance data reporting the presence of *P. aeruginosa* in Intensive Care Units (ICUs), showing its involvement in approximately 20.8% of healthcare-associated infections (ECDC, 2016). Notably, the ECDC surveillance from 2016–2020 highlighted persistently high carbapenem resistance rates in *P. aeruginosa*, with the number of occurrences on an increasing trend in several southern and eastern European countries, highlighting a growing east–west gap in resistance levels (19,20). The clinical impact of these infections is considerable: *P. aeruginosa* carries an estimated overall mortality of around ~20%, which rises to ~30% in ventilator-associated pneumonia and up to ~50% in cases of bacteremia [26]. In the U.S. alone, MDR *P. aeruginosa* caused an estimated 32,600 infections among hospitalized patients in 2017, resulting in approximately 2,700 deaths [21]. Fighting antibiotic resistance becomes, therefore, not only a pressing medical problem but also an economic and societal challenge. In this context, the latest research trends are directed towards developing novel antimicrobials or adjuvants that target bacterial virulence rather than viability. Instead of directly killing bacteria or inhibiting their growth, efforts are increasingly focused on disarming pathogens by interfering with their virulence factors. To this aim, the characterization of proteins engaged in lipopolysaccharide (LPS) biosynthesis may provide novel targets for antivirulence therapies. Among the deadliest strains of *Pseudomonas aeruginosa*, PAO1 (sometimes misspelled PA01) is a chloramphenicol-resistant mutant of the original PAO strain (earlier called "*P. aeruginosa* strain 1"), and is the common reference strain for *P. aeruginosa* worldwide [27].

The Wzx/Wzy synthetic pathway

The Wzx/Wzy-dependent pathway is the predominant system for bacterial polysaccharide synthesis [28]. Given its fundamental role in cell viability, the Wzy-dependent pathway is well conserved across both Gram-negative and Gram-positive bacteria [29], [30]. This biosynthetic pathway is responsible for the production of various surface polysaccharides, including the O antigen, exopolysaccharides, and capsular polysaccharides. The sequence variability of the enzymes involved in this pathway is connected to the structural diversity of the polysaccharides produced by different

bacterial strains. In the model proposed by Whitfield and confirmed by subsequent studies, synthesis begins on the cytoplasmic side of the inner membrane, where a phosphoglycosyltransferase catalyzes the addition of a sugar molecule to the undecaprenyl phosphate (UndPP) carrier. This initiates a cascade involving other glycosyltransferases that form an ordered sequence of sugar residues, i.e., the O-antigen repeating unit (RU), linked to the UndPP carrier [31]. This lipid-linked oligosaccharide is then flipped across the inner membrane to the periplasmic side by the flippase Wzx, which is an integral membrane protein that is thought to act as a cation antiporter (Figure 6, step 1) [31], [32]. Once translocated to the periplasm, the O-antigen units are polymerized by the Wzy polymerase, another integral membrane enzyme. Wzy extends the growing polysaccharide chain by transferring new RUs to the non-reducing end of the polymer through α- or β-glycosidic linkages (Figure 6, step 2) [33]. The chain length of the O-antigen is regulated by the polysaccharide co-polymerase Wzz, which acts as a molecular "ruler". In this hypothesis, Wzy and Wzz interact dynamically during polymerization, with Wzz helping to orient the growing chain until steric constraints cause its release (Figure 6, step 3) [34]. Once polymerization is complete, the ligase WaaL catalyzes the attachment of the O-antigen to the lipid A–core oligosaccharide complex (Figure 6, step 4), forming mature lipopolysaccharide (LPS) [35]. The fully assembled LPS molecule is then transported to the outer membrane via the lipopolysaccharide transport (Lpt) pathway [36].

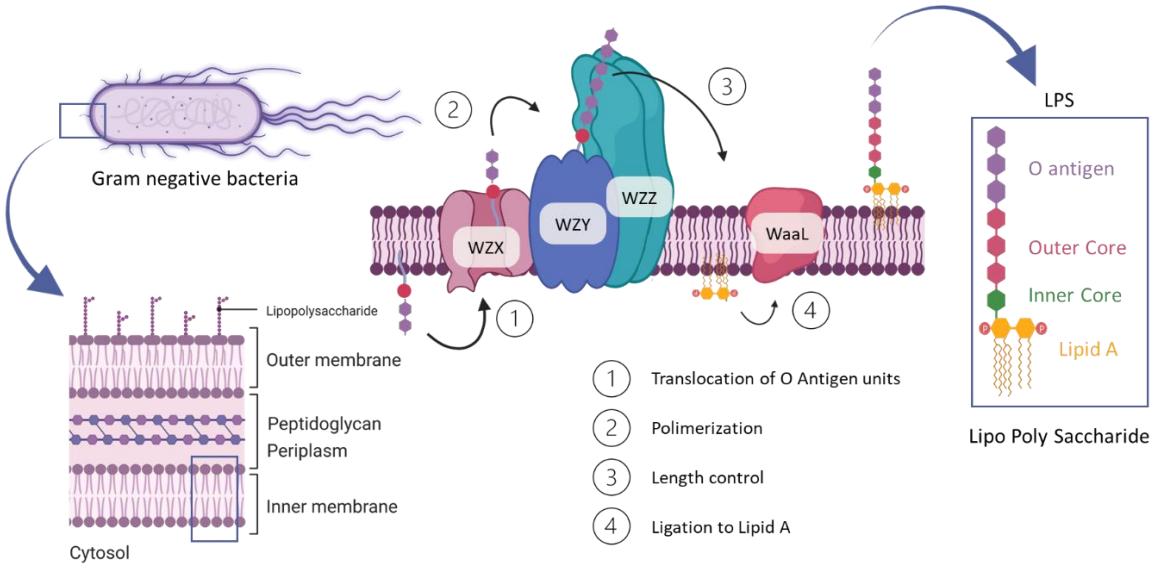


Figure 6 - Schematic representation of the Wzx/Wzy-dependent pathway (adapted from Islam et al. 2014)

The flippase protein Wzx

The role of the protein Wzx in O-antigen biosynthesis was identified for the first time by the Reeves group, who detected the accumulation of single O-antigen units on the surface of everted membrane vesicles derived from a wzx knockout mutant. Further analyses of genetic knockouts confirmed the presence of wzx genes in several bacterial species, including *Pseudomonas aeruginosa* [37], [38], but the sequence homology

between orthologs in different organisms appeared quite low, suggesting a high specificity for the substrate of these proteins, i.e., the different O antigen RUs. Despite its relatively low sequence similarity, the conformation predictions of Wzx sequences suggest that this protein is a member of the MATE (multidrug and toxic compound extrusion) family of inner membrane (IM) efflux proteins, classified as part of the polysaccharide transporter (PST) family, within the MOP superfamily (multidrug/oligosaccharidyl-lipid/polysaccharide exporter) [39]. Biochemical evidence suggests that Wzx acts as a flippase, translocating lipid-linked molecules between membrane leaflets. In particular, Wzx enables the transfer of negatively charged oligosaccharide O-units, linked to the undecaprenyl pyrophosphate (Und-PP) carrier, from the cytoplasmic to the periplasmic leaflet of the inner membrane [40]. Further analyses aimed at elucidating the structural arrangement of Wzx. The sequence of Wzx was fused with the sequence of the fluorescent reporter protein GFP and expressed in *P. aeruginosa* PAO1 strain to determine the orientation of the C-terminus of Wzx within the inner membrane. The resulting construct exhibited fluorescence, indicating that the C-terminus of Wzx faces the cytoplasm, as the reporter GFP does not exhibit fluorescence in the periplasmic space. The application of topology prediction data, together with experimental evidence from truncated mutants, suggested that 12 TransMembrane Segments (TMS) are present between the cytoplasmic N- and C-termini of the O antigen flippase (Figure 7) [41].

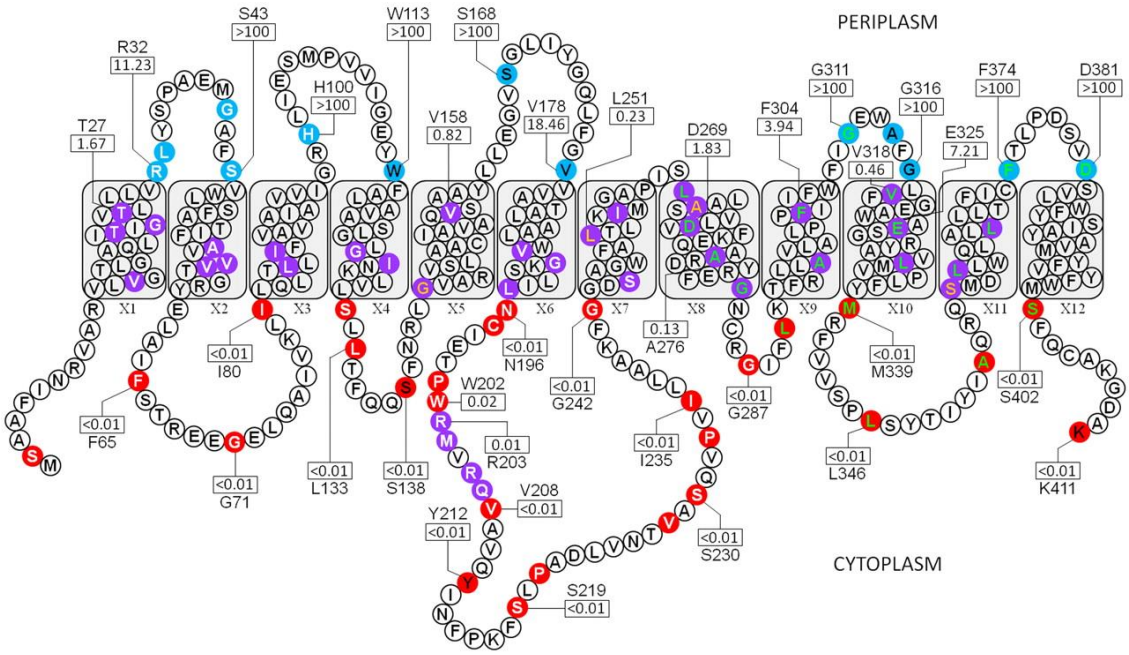


Figure 7 - Topological map of Wzx from the P. aeruginosa strain PAO1 (from Islam et al.2010). Colored residues represent the amino acid positions of each truncation used. Residue colors denote the subcellular localization of a given truncation: blue, periplasm; purple, TM; and red, cytoplasm. TMS are labeled from X1 to X12. N-terminus on the left (M residue), C-terminus on the right (K residue).

According to the estimated topology model, the cytoplasmic region of Wzx from *P. aeruginosa* PAO1 strain is composed of three large loops. The length of these loops and their predicted secondary structure indicate a complex tertiary structure architecture.

Given its considerable size, the cytoplasmic loop 3 (which connects transmembrane segments 6 and 7) may play a key role in the translocation process, potentially acting as part of the substrate recognition site in the cytoplasmic side of the protein. This evidence supports the proposed function of Wzx in transferring lipid-linked oligosaccharides, a process that starts in the cytoplasm [41]. On the opposite side, the periplasmic domain of the protein has only 3 shorter loops with possible functional importance, suggesting a less complex tertiary structure in the periplasmic surface of the protein. Site-directed mutagenesis of 102 charged and aromatic residues allowed the identification of several essential residues for flippase function (Figure 8, black residues). Notably, activity could be restored when substitutions of these functionally important residues preserved the original charge or aromatic properties [34]. Also, some of these residues are expected to face the cationic lumen of Wzx, consistent with a potential role in substrate interaction and recognition during translocation [34]. However, the molecular determinants of the translocation mechanism need further investigation, as the whole amino acid sequence may be involved in substrate binding specificity, besides the electrostatic potential of the protein surface [41].

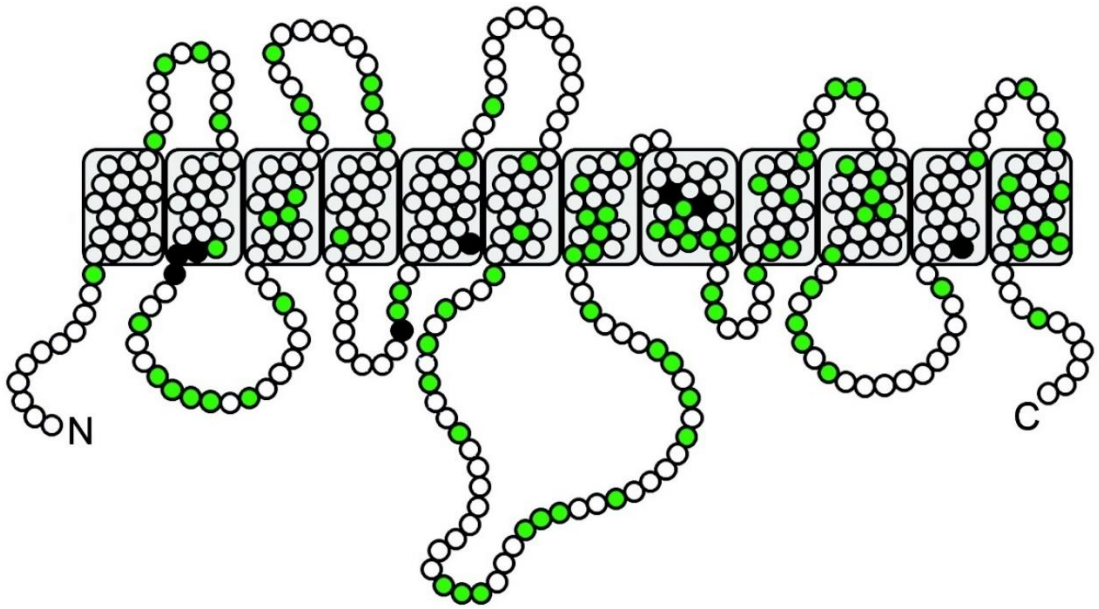


Figure 8 – Residues involved in mutagenetic studies, reported on the topology map of Wzx from P. aeruginosa PAO1. Black dots identify positions with functionally important residues, green dots are used for positions that do not display any effect on function [32].

Nine of the twelve transmembrane segments (TMS) of this protein contain up to seven charged residues. This unusual distribution of charged amino acids within the membrane region is further evidence of the formation of a recognition site essential for the translocation mechanism [41]. Initially, biochemical studies led to the hypothesis that the substrate selection and translocation are most likely determined by the first sugar of the RU, as the LPS synthesis in *P. aeruginosa* proceeded even in the presence of non-native subunits, as long as the first sugar of the RU was conserved [41]. However, subsequent studies proved that Wzx enzymes can be specific for the whole sequence of the RU [42]. As regards the mechanism of translocation, a first hypothesis by Lam and

co-workers suggested an active proton-coupled antiport mechanism, i.e., the coupling between substrate translocation and a proton movement across the membrane in the opposite direction. This work concluded that, in the case of the Wzx protein of *P. aeruginosa* PAO1, the translocation of the negatively charged O-antigen subunits, composed of two mannuronic acid and one fucosamine residues, is not spontaneous. While the transmembrane helices of Wzx, forming a positively charged pore-like channel, are able to facilitate substrate binding and translocation of the anionic sugar moieties, the energetics of the oligosaccharide translocation are unfavorable unless the movement is coupled with a proton transfer according to the gradient [41]. At present, no experimental structures of the Wzx flippase from *Pseudomonas aeruginosa* have been reported, and until recently, no structures were available for any of its homologs. However, homology modeling studies and molecular dynamics simulations took advantage of the structures of related transporters of the MATE family, such as NorM [43]. Although these models are based on homologous proteins and not on experimental data, they provide the only available structural insight into a possible mechanism for the function of the Wzx protein of *P. aeruginosa* PAO1 [32]. In particular, these studies suggest that the protein should be able to alternate between two distinct conformational states, one open toward the periplasm and the other toward the cytoplasm (Figure 9) [43]. In the hypothesized mechanism of translocation, the cytoplasm-facing protonated state of the transporter has high binding affinity for the UndPP-linked O antigen repeated unit (Figure 9, panel E). The binding of this substrate to the transporter via the TMS1-TMS9 helices (panel F) triggers a positional reset of TMS8 and TMS11, which causes the deprotonation of a residue in the transmembrane portion (panel G). In turn, this induces a conformational rearrangement, tightening the binding of the substrate (panel H). The relaxation of this system leads to a conformational change towards the periplasm-facing deprotonated state (panel A). Due to the high proton concentration in the periplasmic space, the TM domain of Wzx is again protonated (panel B), decreasing the affinity for the substrate that diffuses laterally into the periplasmic leaflet through TMS1 and TMS9 (panels C and D). A final conformation rearrangement brings the protein in the initial state, i.e., the cytoplasm-facing protonated conformation (panel A), thus starting again the translocation cycle. By comparison with other known structures of transporters of the MATE family, Lam and co-workers identify the site of the protonation in the D269-D359 dyad of residues, belonging to helices TMS8 and TMS11 respectively, with an assisting role of E61 (TMS2).

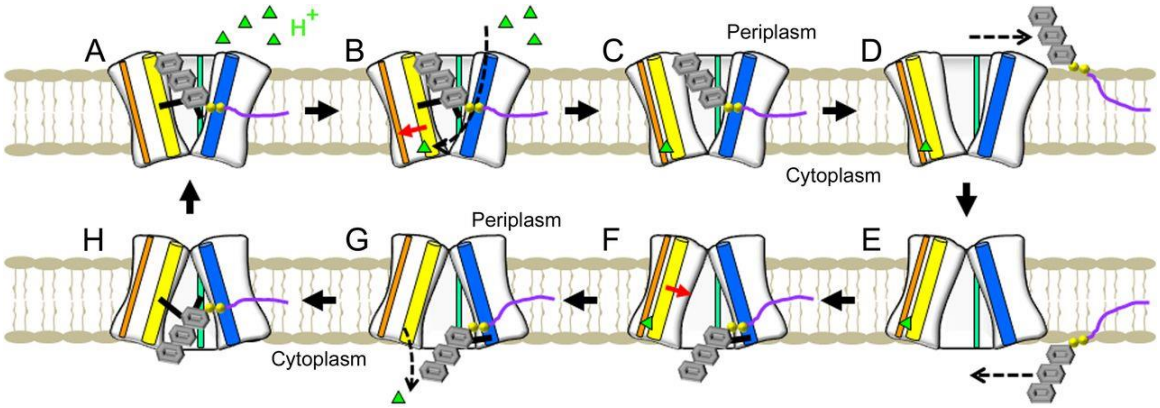


Figure 9 - Proposed mechanism of Wzx translocation (from Islam et al. 2013). (A) Substrate-bound, periplasm-facing, deprotonated state of Wzx. (B) The protonation of a TM residue causes (C) the disengagement of the substrate, (D) allowing for lateral diffusion of the flipped UndPP-linked O unit to the outer leaflet. (E) Wzx transitions to a cytoplasm-facing protonated state, without substrate. (F) The loading of a new UndPP-linked O unit from the cytoplasmic leaflet triggers (G) the deprotonation of the TM portion. (H) Substrate-bound, cytoplasm-facing, deprotonated state transitions (A) to a periplasm-facing conformation, resulting in flipping of the UndPP-linked O unit and in the restart of the translocation cycle.

An interesting new perspective on the translocation mechanism was opened by Le Bas *et al.* in January 2025 with the determination of the first experimental structures of a homolog of Wzx from *E. coli*, known as WzxE [44]. The authors were able to characterize multiple conformations of *E. coli* WzxE at resolutions between 2.3 Å and 2.8 Å, in complex with specific nanobodies able to stabilize different states. The structures confirmed that the protein is composed of 12 transmembrane helices, consistent with the architecture of MATE transporters, and show periplasm-facing and cytoplasm-facing states, proving the involvement of both conformations in the substrate flipping mechanism. While the transporter does not require ATP, lacking a nucleotide-binding motif [44], the energetic drive of the translocation is still not clear. Functional analysis of *E. coli* WzxE excluded the proton-coupled antiport mechanism previously proposed for *P. aeruginosa* Wzx, or an ion-coupled mechanism. The positively charged channel is lined by lysine and arginine residues, which are not expected to promote the transport of ions or protons. No negatively charged path was identified for the movement of positive charges across the membrane [44]. However, the difference in free energy of lipid III, the substrate of WzxE, between the inner and outer leaflet is estimated to be minimal, while the activation energy of the translocation is large, suggesting that the spontaneous translocation does not occur, and a source of energy is required. Yet, no evidence has been gathered so far to point to a specific energy-powered mechanism. However, results obtained on the *E. coli* WzxE protein cannot be directly related to its homologous Wzx from *P. aeruginosa*, as the two protein share only ~23% of sequence identity (obtained with ClustalO v.1.2.4 [45]), highlighting the need for further investigation on these proteins. Given the central role of Wzx proteins for the biosynthesis of the O antigen and other bacterial polysaccharides and the absence of mammalian homologs, Wzx represents an appealing target for selective antibacterial drugs with minimal side effects. The low sequence conservation of these proteins among bacterial species could enable the development of strain-specific inhibitors or adjuvants.

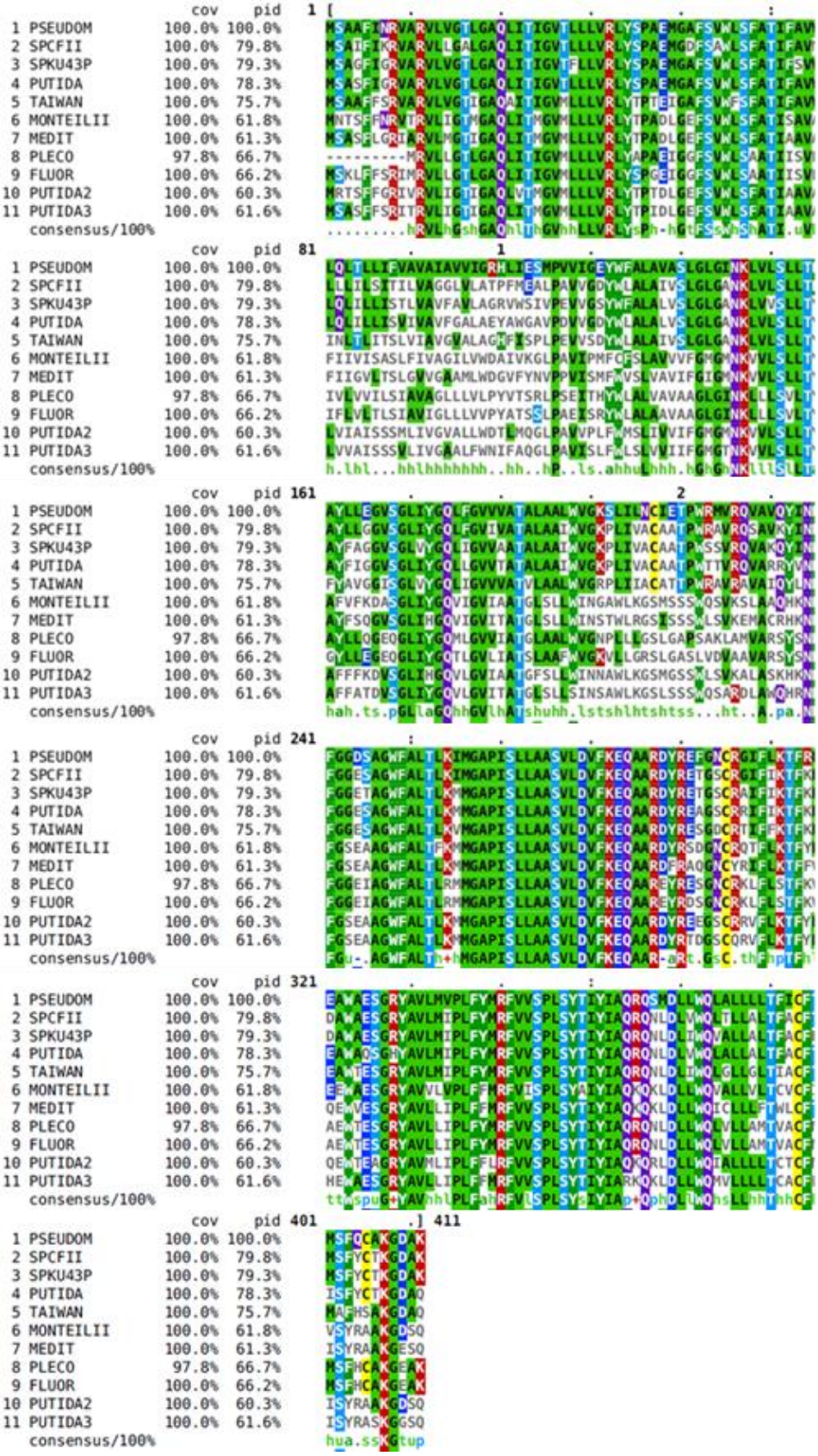


Figure 10 – Sequence alignment of Wzx proteins from different species of the genus Pseudomonas. PSEUDOM: P. aeruginosa. SPCFII: P. sp. CFII68. SPKU43P: P. sp. KU43P. PUTIDA: P. putida. TAIWAN: P. taiwanensis. >MONTEILII: P. monteilii. MEDIT: P. mediterranea. PLECO: P. plecoglossicida. FLUOR: P. fluorescens. PUTIDA2: P. putida. PUTIDA3: P. putida. This alignment has been performed with the ClustalO (1.2.4) algorithm searching in the NCBI database (National Center for Biotechnology Information, U.S.)

Klebsiella pneumoniae

Another member of the ESKAPE group, *Klebsiella pneumoniae* is a Gram-negative bacterium belonging to the *Enterobacteriaceae* family. It is widely distributed in the environment, found in soil, water, and on various surfaces, but it is also part of the human microbiota, populating the gastrointestinal tract and the upper respiratory mucosa [46]. Under normal conditions, *Klebsiella pneumoniae* is a harmless commensal; however, when host defenses are weakened by disease, medical treatment, or hospitalization, it becomes an opportunistic pathogen. In the past, infections were mostly confined to immunocompromised patients, but the emergence of hypervirulent (hvKp) lineages has broadened the clinical spectrum. These strains infect and cause severe pathologies even in healthy individuals, among which are liver abscesses, meningitis, and disseminated infections [47]. In hospital settings, *K. pneumoniae* is a major cause of nosocomial diseases, such as pneumonia, urinary tract infections, and bacteremia. In recent decades, *Klebsiella pneumoniae* has emerged as one of the most alarming bacterial threats to global health, largely due to the spread of multidrug-resistant (MDR) and carbapenem-resistant (CRKP) strains [3], hence its inclusion in the ESKAPE pathogens. *K. pneumoniae* is a striking example of microbial adaptability, as this pathogen can elude the action of both antimicrobial compounds and the host immune responses. Fundamental to its resilience is the capsule, and particularly the capsular polysaccharide (CPS), which plays a crucial role in immune evasion, biofilm formation, and persistence in the environment [48]. The capsule, primarily composed of polysaccharides and polypeptides, constitutes an envelope of the bacterium [49]. Not only does it enhance survival in hostile environments, but it also promotes virulence by increasing the persistence in tissues and blood [49]. Moreover, this polysaccharide-based envelope is responsible for the regulation of biofilm dimensions. In addition to the capsule, *K. pneumoniae* employs another protective mechanism, i.e., the production of the biofilm, a structured matrix composed of polysaccharides, proteins, and extracellular DNA, that embeds the cells in a highly protected microenvironment [48]. The CPS forms a dense, gel-like layer enveloping the bacterial surface and acting as a physical and chemical shield against phagocytosis, desiccation, and the action of antimicrobial peptides. The CPS shields the bacteria from both immune system attacks and antimicrobial agents, thus creating reservoirs of infection that are difficult to eradicate. Shielded by this matrix, *K. pneumoniae* exhibits increased tolerance to antibiotics and oxidative stress, while evading recognition and response by the immune system. *K. pneumoniae* relies on the Wzx/Wzy pathway for the synthesis of the CPS [50]. The CPS and the biofilm constitute the principal structural and functional components of an integrated defense system, giving *Klebsiella pneumoniae* the remarkable resilience that makes it such a formidable pathogen. Its propensity to form biofilms on medical devices, such as catheters, ventilators, and prosthetic implants, contributes to the chronicity of infections and antibiotic tolerance. The National Institute of Health (NIH,

U.S.) estimates that nearly 70% of chronic bacterial infections involve biofilm formation, with *K. pneumoniae* among the most prominent examples [51]. Its remarkable ability to evade treatment through several defense mechanisms underscores the growing threat of multidrug-resistant bacteria and highlights the urgent need for new and effective therapeutic strategies.

Phages

Bacteriophages, or phages, are viruses that infect bacteria, requiring the bacterial machinery for replication and spread of the infection. Each phage contains either a DNA or RNA genome, encapsulated in a protein shell called a “capsid”, whose structure defines the phage morphology and classification [52]. To infect the host cell, phages recognize specific receptors on the bacterial surface and, once docked, inject their genetic material, breaking the bacterial envelope. Inside the cell, new phage DNA is synthesized using the bacterial replication machinery and the bacteriophage genome as a template. Replication proceeds through either a lytic cycle, in which the host cell is lysed to release new viral particles, or a lysogenic cycle, with the phage genome integrating into the bacterial chromosome and remaining silent but capable of evolving to the lytic phase [52]. During the lytic cycle, transcription and translation of phage macromolecules proceed using the host machinery, leading to the assembly of new phage particles. Finally, the release of the bacteriophages requires the disruption of the cell wall and membrane, resulting in the death of the bacterial host cell [53]. Phages are extremely specific, often targeting only a single bacterial strain. However, it is to be presumed that each bacterial strain can be infected by at least one phage. From a therapeutic perspective, the high selectivity of phages, combined with their bactericidal capacity, makes them promising alternatives to conventional antibiotics, in particular against resistant pathogens. Among the key advantages provided by a selective therapy, there are its potential to preserve the human natural microbiota and a low propensity for the development of resistance strategies. In addition, the relatively simple large-scale production of phages has suggested their clinical use [54]. The study of phages with a therapeutic approach began in the first half of the 20th century, when phage preparations were developed to treat a variety of bacterial infections [55]. However, the use of bacteriophages in therapy was abandoned, as the newly developed antibiotics were considered a more convenient treatment option, carrying less potential risks [56]. Thus, while not definitely disproved as a therapeutic option, the application of phages in the treatment of infections still lacks a sufficient number of clinical trials. With the rise of antibiotic resistance and the spread of multidrug-resistant pathogens, the interest in bacteriophages as alternative antimicrobials has been renewed. Studies have shown that phages can effectively target virulent, resistant, and biofilm-forming *K. pneumoniae* strains [57]. Despite this promising potential, clinical application remains limited to a few specialized centers, particularly in Eastern Europe, where approved formulations are used to treat chronic and antibiotic-resistant infections. In Western countries, phage

therapy is still only experimental, though compassionate-use cases and ongoing clinical trials are progressively moving it toward wider regulatory approval. Several challenges remain to be addressed, including the lack of robust safety and efficacy clinical data, the need for highly strain-specific formulations, and the risks associated with the development of bacterial resistance or phage instability. Consequently, no phage therapy products have yet received formal approval for human use in either the European Union or the United States.

EndoGlycosidases (EG)

The success of lytic bacteriophages depends on their ability to infect a bacterial host, replicate, and release new viral particles. In *Caudovirales*, i.e., tailed phages, the most common virus type affecting human pathogens [58], infection requires breaching the bacterial cell wall twice, the first time to inject the genome, and the second to lyse the cell and release the newly synthesized virions. When bacteria are shielded by biofilms, additional barriers such as extracellular proteins, lipids, DNA, and polysaccharides must also be overcome. Phages employ different enzymes depending on the barriers that they have to breach. For example, different strategies must be deployed to cross the thick peptidoglycan layer of Gram-positive bacteria and the outer membrane of Gram-negative bacteria. Infection begins with host recognition, mediated by viral receptor-binding proteins (RBPs) that are able to attach to the bacterial surface receptors [59]. Once the target is recognized, bacteriophages employ specialized depolymerase enzymes to break down bacterial surface barriers, thus facilitating the injection of the genetic material into the host cell. Phage depolymerases comprise a broad group of enzymes that degrade polysaccharide structures on the bacterial surface, facilitating phage adsorption and infection. These enzymes are broadly classified as either hydrolases, which break covalent bonds through hydrolysis (e.g., rhamnosidases, sialidases, LPS deacetylases, etc.), or lyases, which cleave the polysaccharide linkages via β-elimination (such as K-specific lyases) [60]. The viral depolymerases degrade surface components like LPS, CPS, or biofilm exopolysaccharides, exposing the bacterial surface to phage attachment. Among them, endoglycosidases represent a subclass of depolymerases that hydrolyze internal glycosidic bonds of the capsule polysaccharides. After that, virion-associated lysins (VALs) are released and are able to digest locally the peptidoglycan layer to enable genome injection. Depolymerases also have a role at the end of the lytic cycle: once assembly of the virions is complete, intracellular lytic enzymes cause extensive disruption, leading to bacterial lysis and the release of the newly synthesized phage particles. Because of their high substrate specificity, new research efforts are aiming at developing recombinant endoglycosidases capable of selectively degrading the capsules of single bacterial species, or even strains. These engineered enzymes, called enzybiotics, can both recognize bacterial receptors and degrade protective barriers, thereby reducing bacterial virulence and enhancing antibiotic effectiveness [60]. Their therapeutic potential could

lead to the development of phage-derived enzymes targeting multidrug-resistant infections. In this thesis, one of these enzymes was studied, specifically, a hydrolase derived from the bacteriophage ΦBO1E [61] selected for its capability of degrading the CPS of the *Klebsiella pneumoniae* strain KpB-1, belonging to the clonal group 258 (CG258, Sequence Type 512), one of the most diffuse carbapenemases-resistant *K. pneumoniae* in several countries, including Italy [62].

Aim of the work

This work focuses on the P. aeruginosa Wzx flippase, a key component of the Wzy pathway responsible for the assembly of the O-antigen. Given its central role in virulence and surface antigen presentation, Wzx represents a potential drug target for next-generation therapeutics aimed at overcoming *P. aeruginosa* drug resistance. As a second objective of this thesis we aim at characterizing a Endoglycosidase, a bacteriophage protein involved in the disruption of the *K. pneumoniae* capsular polysaccharide, thus providing insights into this new approach for targeting bacteria.

General aspects of this thesis

This thesis presents the two main projects I carried out during my PhD and is divided into two parts: the first focuses on the characterization of Wzx, a bacterial flippase, and the second on the characterization of a viral endoglycosidase. Both works represent the continuation of projects initiated by other researchers, whose efforts I had the honor to build upon. The Wzx project began some time ago, with the goal of characterizing a potential target for a new class of antibacterial agents. The project was particularly challenging because the protein of interest is physiologically embedded in bacterial membranes, which makes its study especially difficult. Moreover, purification was problematic, as some contaminants consistently remained in the sample (possibly due to interaction with the target protein). Finally, assessing the functionality and thus the proper folding of the protein was not possible, as direct activity assays are technically unfeasible. Fortunately, thanks to the work of Dr. I. Marcovich, under the supervision of Prof. De Zorzi (the supervisor of this thesis), I was able to start with valuable background information, established protocols, experimental results, and overall prior expertise. The initial approach was to continue the previous efforts while integrating a computational strategy to optimize purification. This idea was formalized into a project (later extended into a second one) approved by CINECA, which provided access to supercomputers for a remarkable total of 93,000 computing hours. However, this approach turned out to be too slow and ultimately ineffective, although it did yield interesting insights into the molecules involved. For the structural characterization of the protein, we proceeded with X-ray diffraction, and as a complementary technique, we introduced cryo-electron microscopy. Despite the progress made, much work remains to be done in order to achieve a higher-resolution structural determination. Given the potential of cryo-EM for structural studies, I was then involved in the second project. Viral endoglycosidases are enzymes produced by viruses to facilitate their entry into bacteria. The group of Prof. Cescutti, and particularly Dr. M. Zaro, developed a novel and innovative approach to exploit their potential. Although structural determination was not essential for the development of this project, it proved highly valuable for clarifying the mechanism of action and for possible enzyme engineering toward future applications.

Structural characterization of proteins

The Protein Data Bank

The Protein Data Bank (PDB), established in 1971 with an initial set of seven protein structures, has evolved into the primary global repository for structural data of biological macromolecules. As of September 2025, it contains over 242000 entries derived from X-ray crystallography, cryo-electron microscopy, and NMR spectroscopy, encompassing proteins, protein complexes, protein-nucleic acid complexes, and large macromolecular assemblies responsible for concerted functions in cells. By providing open access to experimentally determined atomic coordinates, the PDB serves as a critical resource for structural biology, supporting comparative analyses, validation of experimental models, and the development of computational approaches such as Molecular Dynamics simulations and AI-based prediction tools. It is therefore not only an archive, but also an indispensable platform that integrates structural data into the broader framework of modern biomedical research. The following analysis of the PDB was performed using its Application Programming Interface (API), which allows users to query and analyze every record in the database. Custom scripts were implemented in Python with the *rcsbsearchapi* library, and results were plotted with *matplotlib* (plots were updated in September 2025). Nowadays, most structures deposited in PDB correspond to entries of soluble proteins (96.5%), while membrane proteins account for only 3.5% (Figure 11). This imbalance highlights the persistent challenge of obtaining high-resolution structures for membrane proteins, primarily due to difficulties in expression, purification, and crystallization. Analyzing the distribution of techniques used for structural characterization, X-ray crystallography remains the primary technique for soluble protein structure determination (78.9%), underscoring its historical role as the leading method. Nevertheless, cryo-EM (11.5%) makes a substantial contribution. On the contrary, if the structures of membrane proteins are considered, X-ray crystallography contributes 1.4%, while cryo-EM accounts for 2.1% of the PDB records. As confirmed by Figure 12, electron microscopy (EM) is currently the most used technique for Membrane Protein (MP) structure determination.

All the PDB entries:

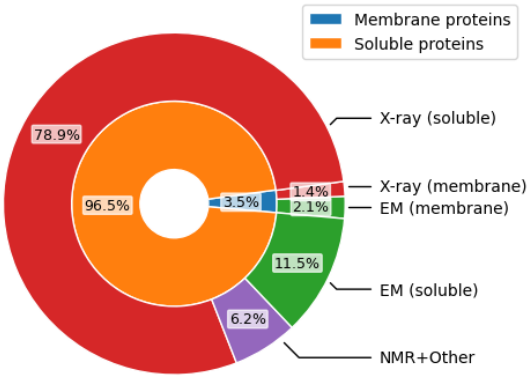


Figure 11 - Ratio between membrane and soluble protein structures, reported according to the experimental method used for their characterization

Membrane protein entries:

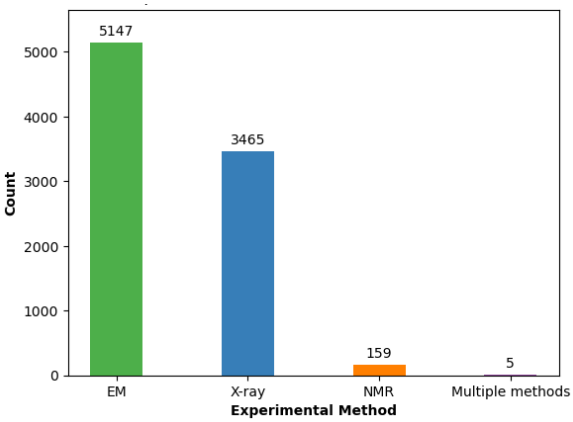


Figure 12 - Experimental methods for membrane protein characterization in the Protein Data Bank

Focusing on membrane proteins, some interesting trends emerge in the choice of experimental methods. One of the factors defining the success rate of experimental techniques for the structural characterization of membrane proteins appears to be their molecular weight (Figure 13). For small proteins (<100 kDa), X-ray crystallography dominates over EM. In contrast, for medium-sized proteins (100–150 kDa), EM becomes the main technique, and its dominance remains clear for larger macromolecules (≥150 kDa). NMR and other methods are relevant only for proteins smaller than 50 kDa. Overall, this plot illustrates the complementarity of methods for the structural characterization.

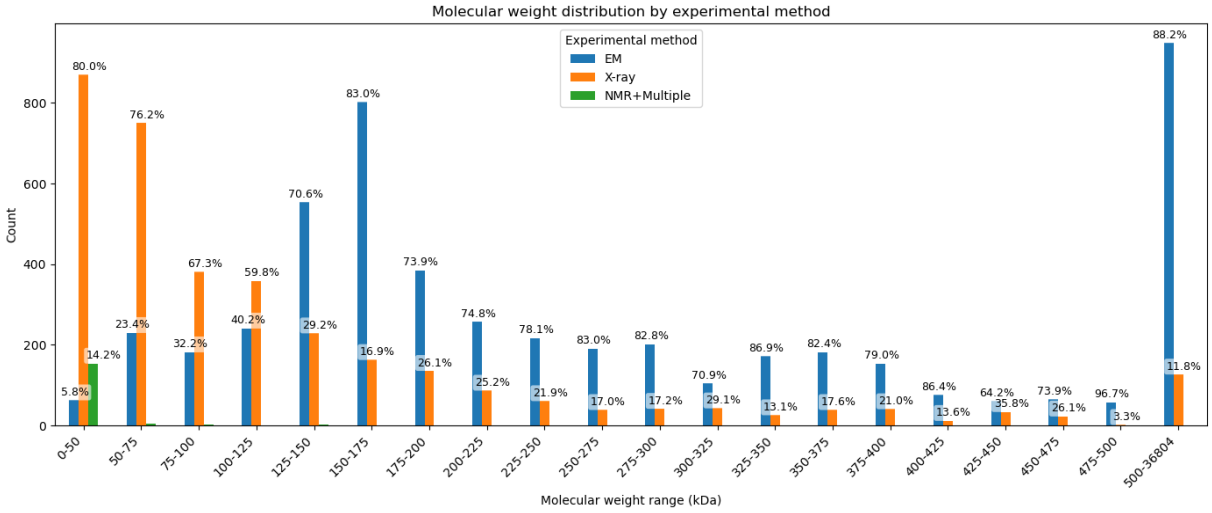


Figure 13 - Molecular weight distribution of membrane proteins entries by experimental method. Percentages above bars indicate the relative contribution of each technique within the molecular weight range.

Membrane proteins are a particularly challenging target for crystallography due to their amphipathic nature. First, the high purity required for structural studies necessitates the use of detergents or lipid mimetics for solubilization during purification steps. In addition, even when a pure sample can be obtained and the crystallization process is successful, crystals of membrane proteins are particularly fragile and often have low diffraction power, owing to their mostly hydrophobic nature and the weak protein-protein

contacts present in the crystals. In this analysis, the historical advantage of X-ray crystallography should be considered, as it was an established technique for structural biology several decades before EM became a high-resolution method. An analysis of the number of structures of membrane proteins deposited in the PDB in time, from the first deposition in 1991 until 2025, reveals that X-ray crystallography (orange curve in Figure 14) dominated the field for the first two decades, reaching a peak of about 200 depositions per year. By contrast, EM (blue curve) remained marginal until around 2013, when major technological advances triggered the so-called “cryo-EM revolution”. From 2015 onwards, EM grew exponentially, reaching over 1000 new structures deposited in 2022. In recent years, however, the number of structures determined with both methods has shown a decline. This trend reflects a changing landscape in structural biology, although the drop in depositions after 2022 remains surprising. In this respect, it is noteworthy that on July 22, 2021, DeepMind released AlphaFold, a new tool that provided more than 380000 predicted protein structures.

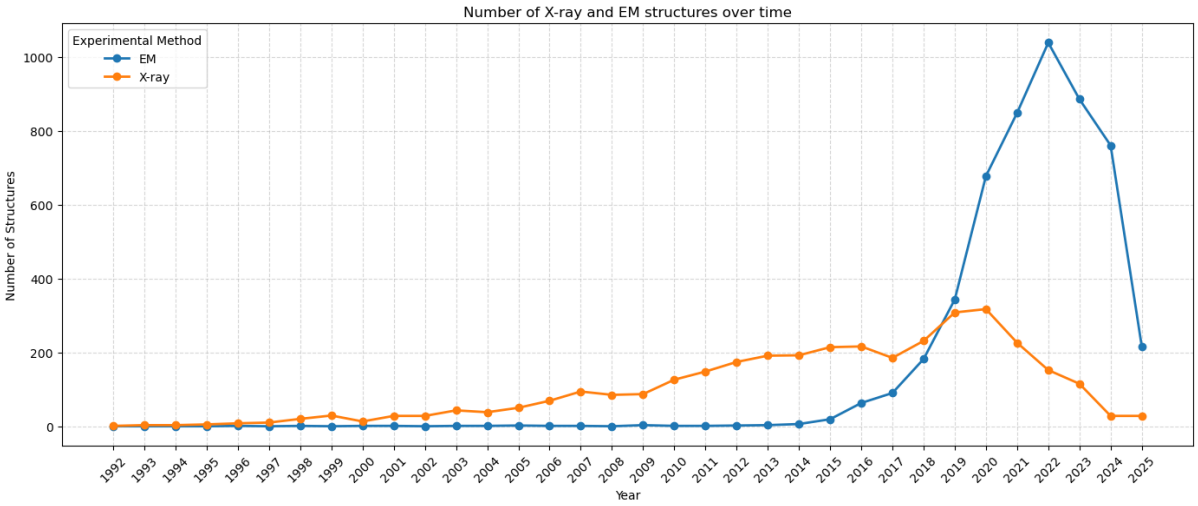


Figure 14 – Analysis of the number of X-ray and EM membrane protein structures deposited in the PDB over time.

X-rays Crystallography

In 1912, Max von Laue demonstrated that crystals diffract X-rays, revealing both the wave nature of the light and the periodicity of crystals. Soon after, William Henry and Lawrence Bragg formulated Bragg’s law ($n\lambda = 2d \sin\theta$), a breakthrough that earned them the 1915 Nobel Prize in Physics and laid the foundation for modern crystallography. Initially applied to problems in physics and chemistry, the method revolutionized biology in the mid-20th century when John Kendrew and Max Perutz solved the first protein structures, myoglobin and hemoglobin (Nobel Prize in Chemistry, 1962). That same year, crystallographic data also supported the double helix model of DNA proposed by Watson, Crick, Wilkins, and Franklin, cementing crystallography as a cornerstone of molecular biology. The power of X-ray crystallography lies in the interaction between X-rays and matter. With wavelengths of about 1 Å, comparable to interatomic distances, X-rays are ideally suited to probe the location of atoms in a molecule. When a crystal is

irradiated, X-rays scatter from the electron clouds of its atoms. Due to the periodicity of the crystal lattice, constructive interference produces diffraction spots, each representing a two-dimensional slice of information. By rotating the crystal and collecting multiple patterns, a three-dimensional dataset can be assembled. Fourier transformation of these data yields an electron density map, in which atomic coordinates can be modeled. Despite its potential for high-resolution structural characterization, crystallography depends on a single, often frustrating requirement: the availability of well-ordered crystals. Crystallization can be understood as a phase transition from solution to solid state, competing with amorphous precipitation. Unfortunately, the natural flexibility of proteins and the weakness of protein-protein contacts make it difficult for the regular lattices essential for diffraction to form. Success depends on several factors, including sample purity, sequence and conformational homogeneity, correct folding, and sufficient protein amount. In addition, not all crystals are suitable for diffraction: size, order, and single-crystal integrity are essential. Even crystals that appear ideal under the microscope may diffract poorly, often making the optimization indispensable. The crystallization process involves two main stages: nucleation, in which small ordered aggregates are formed, and crystal growth, during which these nuclei expand into larger, well-ordered crystals. Thermodynamically, this reflects a balance between weak protein–protein interactions, entropic penalties from the ordering of protein molecules, and entropic gains from releasing bound water. The crystallization process is commonly visualized using a phase diagram, which takes into account protein and precipitant concentrations (Figure 15). When the combination of these two parameters lies below the solubility curve, the protein is soluble; otherwise, the system is in a condition of supersaturation. The supersaturation region is further divided into a metastable region, where the formation of crystalline aggregates of molecules, i.e., nuclei, is hampered by kinetic factors, and a labile region, where nucleation occurs at a rate compatible with the timeframe of the experiment.

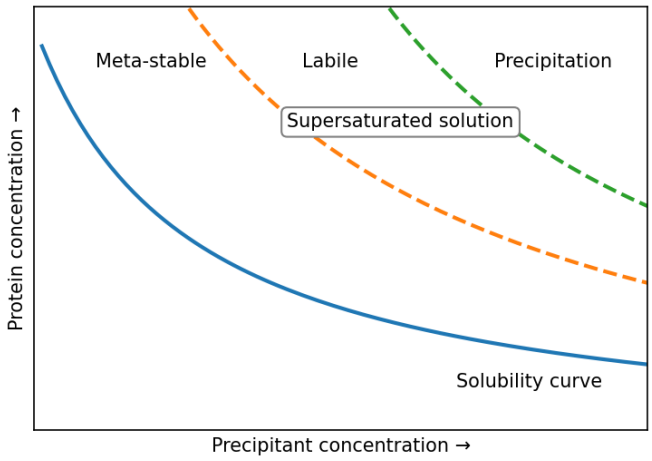


Figure 15 - Crystallization phase diagram

Ideally, the crystallization experiment should start just above the curve dividing the labile and metastable regions. In this condition, only a few nuclei are formed. Their formation,

in turn, reduces the protein concentration and brings the system into metastable conditions, where the previously formed nuclei grow into crystals suitable for the diffraction experiment. A variety of methods are available to set up crystallization experiments. One of the most popular techniques applied in crystallization experiments on proteins is the vapor diffusion method, in either the sitting or hanging drop configurations. However, batch crystallization under oil, microdialysis, and, for membrane proteins, crystallization in the lipidic cubic phase are also employed. Parameters such as protein and precipitant concentrations, salt type, buffer composition, additives, and temperature must be varied systematically to optimize the process. Commercial crystallization screens and robotic systems now enable automatic testing of hundreds of conditions. However, reproducibility remains a challenge because even small variations in protein preparations or experimental setup may significantly affect the outcome, and success is never guaranteed. Protein crystals are fragile, highly hydrated, and can be easily damaged by radiation, particularly when powerful X-ray sources such as synchrotrons are used. The development of synchrotron radiation in the late 20th century revolutionized the field, as synchrotrons generate extremely bright and tunable X-ray beams. Cryogenic cooling has therefore become a standard step in crystallography, although it introduces additional risks such as ice formation. Once crystals are obtained, the diffraction experiment is set up by carefully harvesting, cryoprotecting, and mounting one of them on the diffractometer. After data collection, the so-called phase problem has to be solved to obtain a map of the electron density of the crystal. Diffraction patterns collected on the detector during the crystallographic experiment provide the amplitudes of the diffracted beams, but not phases. However, both these pieces of information are essential for reconstructing the electron density maps. The absence of the phases of diffraction beams is known as the “phase problem”. Different methods are available to solve the phase problem in protein crystallography. Among them, the Molecular Replacement method uses a homologous structure as a model and, nowadays, is the most popular method, owing to the huge number of structures on the PDB and the availability of reliable computational models. Experimental phasing methods are also available, but often they require multiple data collections and several crystals. Once the phase problem is solved, model building and iterative refinement allow the determination of the atomic structure, provided the resolution of the data is sufficient (<3-4 Å). At high resolution (<1.5 Å), crystallography can even reveal alternative side-chain conformations and ordered water networks. Despite allowing the determination of only static structures that are an average of the conformations present in the whole crystal, the impact of crystallography is undeniable. As of 2025, the majority of the structures in the Protein Data Bank were solved by X-ray crystallography (as reported in Figure 11). For over half a century, crystallography has been the principal method for unveiling the architecture of biological macromolecules, and it continues to provide atomic-level insights that underpin modern molecular biology and biomedical research.

Data collection at the Elettra Synchrotron facility

During this PhD project, several crystallization trials were carried out for the proteins under investigation. Unfortunately, attempts to obtain crystals of the Wzx protein from *P. aeruginosa* were unsuccessful. In the case of the phage endoglycosidase capable of hydrolyzing the capsular polysaccharide of *K. pneumoniae*, crystallization attempts on a construct (EGNΔ8) yielded some crystals. High-quality diffraction data were obtained from one of these crystals using synchrotron radiation. Single-crystal X-ray diffraction experiments have been performed at the XRD2 beamline of the Elettra synchrotron (Trieste, Italy). The crystal was mounted on a loop, cryoprotected with 25% glycerol, and maintained at cryogenic temperatures throughout data acquisition to minimize radiation damage. For the EGNΔ8 protein, diffraction images were collected using radiation at 1.00 Å wavelength, with a focused beam of 100 × 100 µm in size. A total of 1440 images were acquired, each with an oscillation angle of 0.25°, covering a complete 360° rotation of the crystal. The first step in data processing was indexing, which assigns Miller indices (hkl) to each diffraction spot and enables the determination of the crystal system and unit-cell parameters (a, b, c, α, β, γ). Following indexing, integration was performed to extract spot intensities. This involved summing partial reflections, subtracting background noise, and applying corrections for the geometrical and optical properties of the experiment. Indexing and integration were carried out using *iMosflm* [63] (within the CCP4-8.0.019 suite [64]). Subsequently, the data were scaled, and corrections were applied to account for experimental variations affecting intensity measurements, such as fluctuations in beam intensity, absorption effects, and radiation damage. Symmetry-related reflections and repeated measurements of the same reflection were compared to determine appropriate scaling factors. After scaling, reflections were merged by averaging all equivalent symmetry-related reflections, weighted according to their estimated errors. The R_{merge} value was calculated to assess the internal consistency of the dataset. This parameter takes into account the differences between equivalent reflections:

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{N} \left| I_{hkl,i} - \langle I_{hkl} \rangle \right|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl,i}}$$

where $I_{hkl,i}$ is the intensity of a single reflection and the *i* sum runs within each group of equivalent reflections having an average intensity $\langle I_{hkl} \rangle$. Lower R_{merge} values indicate greater agreement among equivalent reflections. The R_{merge} values were calculated for the whole data collection and within shells of resolution. Another key parameter is completeness, which represents the percentage of unique diffraction intensities measured at a given resolution compared with the total number theoretically possible. While completeness below 90% may indicate experimental limitations, randomly missing reflections generally do not compromise the final structure solution. In this study, scaling and merging were performed using the *Aimless* [65] tool of the CCP4i suite [64]. The Molecular Replacement method was employed to solve the crystallographic

phase problem. This method relies on the availability of the structural model of a homologous protein. When such a model is available, the key step of the Molecular Replacement method is the determination of its orientation and position within the crystal unit cell. From this placement, initial estimates of the structure factor phases can be calculated, which are refined during later stages of structure determination. In this study, the initial search model was generated from an AlphaFold2 prediction of the whole sequence. However, considering the possible errors in the predicted protein conformation, only segments having a per-residue confidence score over 50% were included in the model used for phasing. The resulting model, representing approximately 50% of the EGN-Δ8 sequence, was used as input for Molrep[66] (CCP4i2 suite [64]). Since the model obtained by molecular replacement was incomplete, the missing protein residues were initially built manually in the electron density map and subsequently extended using the *Buccaneer* [67] software (CCP4i2 suite [64]). Structure refinement was performed to obtain the final model of the protein. In this step, parameters such as atomic coordinates and thermal displacement factors (beta-factors) are adjusted to minimize the differences between the observed and the calculated structure factor, using a least-squares approach. This process is known as reciprocal space refinement. Minimization cycles of reciprocal space refinement were alternated with manual or automated fitting of the model into the electron density map, known as real-space refinement. In this study, *Refmac5* [68] (CCP4i2 suite [64]) was employed for reciprocal-space refinement, while *Coot* [69] was used for manual model building and real-space refinement. In the final model, water molecules were added automatically using the *Refmac5* tool and subsequently verified manually.

The quality of the final model was evaluated using the *R-factor* parameter, which provides a numerical measure of the agreement between experimental diffraction data and intensities calculated from the structural model:

$$Rfactor = \frac{\sum_{hkl} \left| |F_{hkl,obs}| - |F_{hkl,calc}| \right|}{\sum_{hkl} |F_{hkl,obs}|}$$

where the modulus of the observed structure factor $|F_{hkl,obs}|$ is calculated as the square root of the experimental diffraction intensities I_{hkl}, while the modulus of the calculated structure factor $|F_{hkl,calc}|$ is obtained from the protein model. After structure refinement, the final model was validated by considering several indicators, including the Ramachandran plot, unmodeled electron density features, difference map peaks, water molecule placement, and other parameters. The validation process was performed using tools available in *Coot*.

Electron Microscopy

Transmission Electron Microscopy (EM) is an well-established technique for investigating molecular structures. Compared to X-ray crystallography, EM offers the significant advantage of a lower protein concentration, without the need for target

crystallization. In structural biology, EM is employed for the preliminary screening and sample quality assessment of a protein sample, and for high-resolution structure determination, in the latter case using the cryo-electron microscopy technique (Cryo-EM). In all EM approaches, the sample is deposited on a small metallic grid and introduced into the high-vacuum column of the microscope. An electron gun generates a beam that is focused on the sample by magnetic condenser lenses. As the beam interacts with the specimen, electrons follow three main pathways: some pass through the sample without interacting (unscattered electrons), others undergo inelastic scattering with energy loss, and some undergo elastic scattering (Figure 16). Elastic scattering provides structural information by producing phase shifts in the electron waves that, through interference, generate image contrast on the detector. Inelastic scattering is generally considered negligible but contributes to the noise of the image.

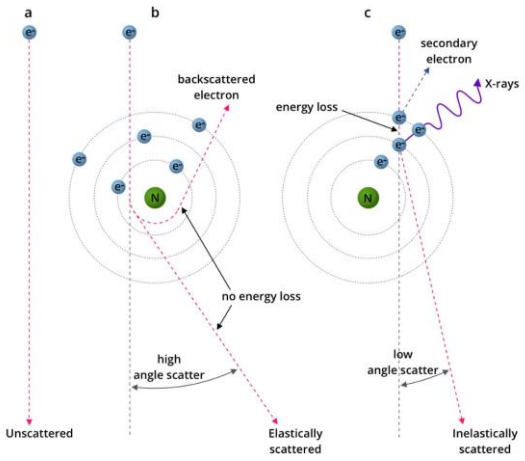


Figure 16 – Schematic representation of the interaction of electrons with atoms (from www.myscope.training)

In non-cryogenic EM, grids are commonly treated before imaging to prevent sublimation of biological material in vacuum and reduce electron damage of the proteins. Due to their small size and low atomic numbers, proteins and protein complexes scatter electrons very weakly, thus exhibiting very low contrast. To visualize such specimens, contrast is usually enhanced by heavy-metal staining, although this procedure can introduce artifacts. Two staining approaches are possible: in positive staining, the stain binds directly to the specimen and the particle appears darker than the background; in negative staining, particles are embedded in a thin layer of stain and appear bright against a dark background. This second approach is usually applied with heavy-metal staining. Typical stains include heavy-metal salts such as uranyl acetate or phosphotungstic acid, whose high atomic numbers provide strong electron scattering. Experimental parameters, including staining time and blotting conditions, strongly affect image quality and must therefore be carefully optimized for each sample. The negative-staining technique enables the collection of two-dimensional images of proteins, and it is a valuable preliminary method for evaluating protein quality, homogeneity, and monodispersity. However, its main limitation is the possibility of introducing distortions

to the native structure of the target protein. Negative staining grids used for this thesis were prepared following three protocols with slightly different parameters. In all cases, the grids were pre-treated before sample application by ozone treatment, plasma cleaning or glow discharge. These methods are used to clean the carbon film, making it more hydrophilic. Then, a small amount of protein solution was applied, followed by a small amount of staining.

1) The Wzx protein from *P. aeruginosa* was analyzed at the CIMA, University of Trieste, with the following protocol:
    - Sample dilutions: 0.1–0.5 mg/ml.
    - Sample quantity: 2µL.
    - Staining solution: uranyl acetate, saturated solution (2µL).
    - Grid preparation method: 5 min in ozone machine.
    - Sample and staining droplet were added at the same time on the grid. Incubation time: 2 min.
    - Blot excess sample with filter paper.
    - Air-drying of the grid.
2) The Wzx protein from *P. aeruginosa* was analyzed at the Cryo-EM Facility, INL, Braga, Portugal, with the following protocol:
    - Sample dilutions: 0.5, 1, 2 mg/ml.
    - Sample quantity: 2µL.
    - Staining solution: UranyLess (Delta Microscopies) (2µL).
    - Grid preparation method: plasma cleaner (support film facing up) for 3 s.
    - Sample droplets were deposited on the grid. Incubation time: 2 min.
    - The excess sample was blotted with filter paper.
    - A stain droplet was added to the grid. Incubation time: 10 s.
    - The excess stain was blotted with filter paper.
    - Air-drying of the grid.
3) The phage Endoglycosidase protein against *K. pneumoniae* was analyzed at the CryoEM CNB-CSIC facility, Madrid, Spain, with the following protocol:
    - Enzyme dilutions: 0.1, 0.5, 1, 2 mg/ml.
    - Substrate dilution (for enzyme concentrations of 0.5 mg/mL): 0.5, 1, 2 mg/mL
    - Sample quantity: 2µL.
    - Staining solution: Uranyl acetate 2% (2µL).
    - Grid preparation method: plasma cleaner for 90 s.
    - Sample droplets were deposited on the grid. Incubation time: 1 min.
    - The excess sample was blotted with filter paper.
    - A stain droplet was added to the grid. Incubation time: 1 min.
    - The excess stain was blotted with filter paper.
    - Air-drying of the grid.

In this thesis, three TEM microscopes were used for negative staining experiments:

- Philips EM 208, 100 kV, at the Interdepartmental Center for Advanced Microscopies (CIMA), University of Trieste
- JEOL JEM 2100, 100 kV, at the Cryo-EM Facility, INL, Braga, Portugal
- JEOL JEM1400, 120 kV, at the CryoEM CNB-CSIC facility, Madrid, Spain

Cryo-EM

Over the last decade, cryogenic electron microscopy has emerged as a mainstream technique in structural biology. Nowadays, advances in instrumentation and computational capacity allow the determination of protein structures at atomic resolution from micrographs of cryo-preserved samples, making cryo-EM a powerful complement to X-ray crystallography, as recognized by the 2017 Nobel Prize in Chemistry awarded to Jacques Dubochet, Joachim Frank, and Richard Henderson for the development of cryo-electron microscopy. The development of Direct Detecting Devices (DDDs, or Direct Electron Detectors, DEDs) led the cryo-EM revolution of the last 15 years. While CCD (Charge-Coupled Device) cameras used before required electrons to be converted into photons by a scintillator, introducing noise and limiting the resolution, DDDs measure electrons directly, greatly improving the signal-to-noise ratio and, thus, the resolution of the images. Furthermore, the high sensitivity of these cameras allows the use of low electron doses, which, together with their fast-reading capability, enable the acquisition of movies consisting of multiple subframes. By analyzing the subframes, beam-induced sample motion can be computationally corrected, thus further enhancing the final resolution. The significant increase in the achievable resolution of the images opened the cryo-EM technique, once restricted to very large macromolecules (i.e., above 300 kDa), to smaller proteins, and today proteins as small as 70 kDa can be investigated with this technique. Cryo-EM is highly versatile: it can be used to acquire 2D images, perform 3D single-particle reconstruction, conduct cryo-electron tomography experiments by tilting the specimen, and up to cellular-scale imaging. In this thesis, the technique was used for reconstruction by Single Particle Analysis (SPA), which remains the most widely employed approach for high-resolution structural determination. Although the theoretical resolution limit of a 200 kV microscope is 0.01 Å, technical limitations restrict the practical resolution limit to around 1.2-1.5 Å, a resolution that is close to the limit of X-ray crystallography. However, in comparison with both X-ray crystallography and negative-staining EM, cryo-EM allows the determination of protein structures in a more native environment, i.e., in solution, making it particularly valuable for macromolecules difficult to crystallize and avoiding, at the same time, stain artifacts. For cryo-EM, metal grids are typically coated with a holey carbon film, allowing the formation of a thin layer of vitrified ice containing the protein in the holes. Preferred grid metals are gold and copper. After applying the protein solution to the grid and removing the excess, the sample is vitrified by rapidly plunging the grid into (metastable) liquid ethane, kept -196 °C, i.e. the boiling temperature of liquid nitrogen. Ethane provides an extremely fast and uniform cooling due to a heat capacity higher than that of liquid

nitrogen, thus producing amorphous ice instead of crystalline water. On the contrary, in liquid nitrogen, the formation of an insulating vapor layer (i.e., Leidenfrost effect) slows down the cooling rate, favoring water crystallization rather than vitrification. The vitrification process is crucial to preserving the specimen in a near-native state and preventing its sublimation in the microscope vacuum. The vitrified ice thickness should remain below ~100 nm, as it critically affects particle distribution and data quality. Considering that cryo-EM is a time-consuming technique and instrument schedules are typically very crowded, sample quality is carefully assessed before data collection (e.g., by negative staining) to maximize the chances of success. A clear experimental plan, including the optimization of sample dilution, grid type, and blotting conditions, is pivotal to obtaining high-resolution data.

For this thesis, two cryo-EM instruments were employed:

- the Glacios 200 kV Cryo-TEM (ThermoFisher), with a high-brightness field electron gun (X-FEG), a ThermoFisher Falcon 4i direct electron detector (DED), at the Cryo-EM Facility, INL, Braga, Portugal;
- the Talos Arctica microscope 200 kV (ThermoFisher, previously FEI), equipped with a Falcon 4i DED detector, at the CryoEM CNB-CSIC facility, Madrid, Spain.

For cryo-EM experiments on the Wzx protein sample, analyzed at the Cryo-EM facility in Portugal, the following vitrification protocol was used:

- Automated grid preparation with the Vitrobot Mark IV instrument (FEI).
- Grid type: Quantifoil Cu/Rh (Copper/Rhodium) R 0.6/1 (hole size=0.6 µm, spacing 1.0 µm).
- Grid preparation method: glow discharge for 30 seconds
- Sample dilution: 0.5, 1.0, 2.0 mg/mL; sample volume: 3µL.
- Blotting time: 3 seconds; blotting force: 0.
- Humidity: 90%

For cryo-EM experiments on the phage endoglycosidase sample, analyzed at the Cryo-EM facility in Spain, the following vitrification protocol was used:

- Automated grid preparation with the Vitrobot Mark IV (FEI).
- Grid type: Quantifoil R 0.6/1 Cu/Rh and Au.
- Grid preparation method: glow discharge for 30 seconds (Cu/Rh grids) or 1 minute (Au grids).
- Enzyme dilution: 0.5, 1.0, 2.0 mg/mL
- Substrate dilution, for enzyme concentration of 0.5 mg/mL: 0.5, 1.0, 2.0 mg/mL. Total sample volume: 3 µL.
- As regards the Vitrobot parameters, several conditions of blotting time (1 to 3s) and force (0 to 5) have been screened
- Humidity: 95%.

After vitrification, the grids are stored in liquid nitrogen until loaded in the autoloader rack inside the microscope for data acquisition.

Single Particle Reconstruction

The procedure for cryo-EM Single Particle Analysis (SPA) follows a series of computational steps aimed at extracting high-resolution structural information from noisy raw data. In the vitrified grid, proteins are randomly embedded in vitreous ice, allowing particles to adopt all possible orientations. Each particle captured by the detector represents a two-dimensional projection of the same protein from a different angle. According to the mathematical principles of tomography, in particular the central slice theorem, a three-dimensional object can be reconstructed from a sufficient number of projections collected at orientations that cover the whole 3D space. SPA is based on the following assumptions:

1) All particles in the specimen have an identical structure. When the sample contains contaminants or aggregates, a frequent occurrence in real protein samples, the analysis can still achieve a reliable reconstruction provided that the target particle is easily discriminated from the contaminant.
2) Particles undergo only 3D rigid body transformations (rotations, translations). This assumption, however, cannot be made for flexible proteins that can assume multiple conformations.
3) Particle images are interpreted as the projection of the common structure plus noise. Noise and aberrations are removed by applying mathematical correction in this crucial step.

Thanks to the development of DDD cameras, images are acquired as movies composed of multiple frames, which are subsequently aligned during motion correction to reduce blurring due to electron beam-induced sample motion. Once corrected, the contrast transfer function (CTF) of each micrograph is estimated to account for the phase contrast and aberration effects introduced by the microscope optics. The next stage in the SPA protocol is particle picking, in which individual protein projections are identified and extracted from micrographs. In a process known as 2D classification, similar views are classified and averaged, which significantly enhances the signal-to-noise ratio. The resulting classes are selected, removing the noisiest and those eventually containing contaminants, and the angular orientation of each averaged projection, expressed in terms of Euler angles, is determined. The selected particles are combined into a consistent three-dimensional reconstruction, taking advantage of the analysis in Fourier space. By combining the slices in the Fourier space obtained by transforming images of particles in different orientations, the full 3D Fourier space of the protein can be assembled. An inverse Fourier transform of this dataset yields a three-dimensional Coulomb potential map, which is used to build the atomic model of the protein. When the reconstruction is complete, potential maps are deposited in the EMDB database, while raw image data can be shared through EMPIAR and related repositories. EM data collected for this thesis were analyzed using the suite CryoSPARC (v. 4.7.1) [70]. This integrated suite provides an intuitive user interface with tools suitable for all stages of

single-particle reconstruction. In addition, the software is optimized for GPU-based computation and has an active community supporting new users. For particle picking, the deep-learning-based software Topaz [71] was used, too. The general workflow used in this thesis includes:

-Importing raw micrographs, later corrected using CryoSPARC motion correction.

-CTF estimation and correction: the microscope defocus and astigmatism values were estimated and used to perform Contrast Transfer Function (CTF) correction. Micrographs with poor CTF fits or low estimated resolution were identified at this stage for potential exclusion.

-Dataset curation: micrographs were filtered using metrics such as the results of CTF fitting, the resolution, or the ice thickness; in some cases, individual micrographs were inspected; low-quality data were removed.

-Particle picking: this task was performed in three different stages.

(1) Blob Picker: this tool was used to identify automatically particle images on the micrographs. The Blob Picker is a fast, generic picker that uses size and shape thresholds to find potential particles across the micrographs, providing an initial set of coordinates useful for a first assessment of particle number and data quality. The initial coordinates are later used to refine interactively the picking parameters or to generate templates for more focused picking approaches.

(2) Manual particle picking: a limited number of micrographs are manually picked to generate a high-confidence reference set of particle coordinates, selected by visual inspection. Although manual picking is time-consuming and therefore performed on only a small subset of micrographs, these curated picks can be a valuable input for subsequent automated methods, as templates or training data.

(3) Deep Learning-Based Picking (Topaz Training): Topaz is a particle picker based on neural networks, which uses a deep learning approach to improve particle picking. During data analysis, picks of potential particles previously obtained by manual or template-based picking were used as training data for Topaz. Once trained, the Topaz model was applied to all micrographs, typically resulting in a more comprehensive and accurate set of particle coordinates.

-Pick inspection and selection: After particle automatic selection, the results were manually reviewed to remove clear false positives. This curation step ensures that only a high-quality subset of particle coordinates is kept for subsequent extraction.

-Particle Extraction from Micrographs: individual particle images were extracted from the micrographs using the curated coordinates, producing a stack later used as input for 2D classification. The extraction box was defined to be at least twice the particle diameter to include all relevant structural information.

-2D Classification: Extracted particles underwent 2D classification. In this step, similar views are aligned and averaged to generate representative class images with a

higher signal-to-noise ratio. Besides providing an initial look at the particle structural features, this process helps identify and discard particles from blurry classes, which are unlikely to contain the target protein but probably represent noise or artifacts. In addition, it can reveal heterogeneity or preferred orientations within the dataset. After classification, particles belonging to the best-resolved classes are retained; all others are discarded. By iteratively repeating this operation, the particle stack is progressively cleaned to retain only well-defined particles suitable for 3D reconstruction.

-Ab-Initio 3D Reconstruction: Using the cleaned particle set, initial three-dimensional *de novo* reconstructions were generated. The CryoSPARC Ab-Initio Reconstruction tool starts from 2D particle images to compute one or more low-resolution 3D volumes without any external reference model. When multiple models are generated, this process helps identify the presence of different conformations. In addition, it can be used as a further clean-up step, excluding heterogeneous particles that produce blurry, noisy and undefined 3D maps. The ab-initio results provide starting volumes for subsequent high-resolution refinement.

-3D Refinement: The initial 3D models were subsequently refined to enhance resolution and map quality. During refinement, all particles are aligned to the initial 3D map, which is iteratively updated to include the contributions of aligned particles. Homogeneous refinement is the standard refinement applied when a single protein conformation is assumed. However, Non-Uniform refinement (NU-refinement) is commonly employed as its adaptive regularization algorithm helps identify flexible regions of the map by iteratively focusing on well-resolved features and down-weighting signal from disordered areas. Typically, NU-refinement performs better on flexible targets or targets with large disordered domains (e.g., micelles) but it is slower than the homogeneous refinement algorithm. The result of the 3D refinement step is a high-resolution 3D reconstruction of the specimen, which can be used for biological interpretation and atomic modeling.

Molecular dynamics

Molecular dynamics (MD) is a computational method that simulates the behavior of molecular systems by integrating the motion equations, providing a description of how atoms and molecules evolve in time under the influence of physical forces [72]. The systems that can be studied with MD range from simple molecules of a few atoms to large protein complexes, including their hydration shells. The simulation can be performed at different levels of complexity, starting from the simpler coarse-grain model of the system, which includes groups of atoms as rigid objects, to a highly detailed quantum-mechanical (QM) treatment of atoms. The computational cost of the simulation increases with system complexity and with the level of detail included in the model (Figure 17). In this thesis, MD was employed to investigate the time-dependent behavior of systems, using the classical mechanics description of polarizable force

fields, which represents a medium level of complexity (Figure 17). Starting from initial conditions including the atomic coordinates, volume, temperature, and pressure, the simulation predicts the trajectories of all atoms according to the classical laws of motion. When performed with sufficient sampling time frequency, MD provides a realistic representation of molecular evolution, allowing the approximation of equilibrium states and the extraction of thermodynamic parameters.

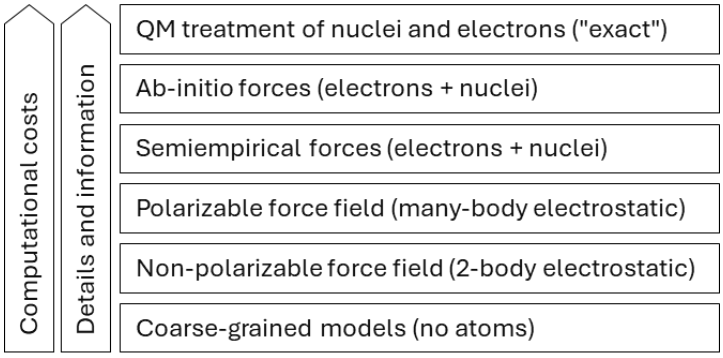


Figure 17 - Levels of complexity in MD simulations

This approach is particularly powerful for studying the behavior of proteins. Proteins are dynamic molecules whose biological function depends on shape, flexibility, and their ability to adopt different conformations over time. The study of conformational changes is, therefore, key to understanding protein activity at the molecular level and how this activity leads to observable biological effects. Based on the structure-activity relationship principle, this approach seeks to correlate the chemical and structural features of a molecule with its biological activity. MD has become a useful tool beyond the static protein structures provided by experimental methods such as X-ray crystallography and cryo-EM. MD allows the investigation of the stability of macromolecules of biological relevance, the interactions in which they are involved (protein-protein, protein-ligand, protein-DNA, etc.), and the effects of mutations. It has also been used as a tool for guiding drug design and, at the highest level of complexity, elucidating reaction mechanisms. The classical MD approach used in this thesis is based on Newton's laws of motion. Each atom in the system is described by its mass, position, and velocity, while the forces acting on it are obtained from the gradients of the potential energy function. By integrating Newton's second law of motion, the trajectory of every atom can be described in time. Numerical integration methods such as the Verlet algorithm are required to perform these calculations and optimize computational efforts. Typically, timesteps are set on the order of femtoseconds (10^{-15} s) to capture even the fastest molecular vibrations, such as the O-H bond stretching in water. At the beginning of a simulation, velocities are assigned according to the Maxwell–Boltzmann distribution at a desired temperature. The following iterative steps of this process require forces to be recalculated and atomic positions updated, generating the trajectory of the MD simulation. Statistical Mechanics (SM), which links the microscopic behavior of individual particles to the macroscopic properties of a system, is essential to interpreting

the results of molecular dynamics simulations. The analysis of simulations as statistical ensembles allows the comparison with experiments that measure bulk properties (e.g., pressure, temperature, or energy). Thus, the trajectory of the simulation represents a collection of all possible states that a system adopts under given thermodynamic conditions, each atomic configuration corresponding to a microstate. A central concept in MD is the ergodic hypothesis, which considers the difference between simulations and experiments. An experimental protein solution contains an enormous number of molecules (in the order of Avogadro's number), and any measurement reflects their average behavior. In MD, however, only a single protein or molecular system is considered. The ergodic hypothesis states that the average properties of a system obtained by following one molecule over a sufficiently long time are equivalent to the average over many molecules at a single time point. This allows the extraction of thermodynamic and structural properties from a single trajectory, as long as the simulation spans a sufficient timeframe and the system has reached equilibrium. Different ensembles of conditions are used depending on which parameters are kept constant. For instance, the NVE ensemble fixes energy, volume, and particle number, while the NVT ensemble fixes temperature and volume. The NPT ensemble, commonly used in biomolecular simulations, maintains constant pressure and temperature of the system while allowing volume to change. The NPT ensemble is particularly valuable because it best reflects typical experimental conditions and was used for all the simulations performed in this thesis. Thermostats and barostats are applied to control temperature and pressure during the simulation. A thermostat regulates temperature by adjusting atomic velocities, while a barostat maintains pressure by modifying the simulation box volume. Together, they allow the system to keep realistic equilibrium conditions, ensuring that the quantities calculated from the simulation can be directly compared with experimental observations. The definition of the system topology is a prerequisite for all classical molecular dynamics simulations. The topology includes atomic masses, bond lengths and angles, charges, heteroatoms, and other relevant parameters of the system. In addition, the topology relies on the definition of amino acid parameters within the selected force field. In this work, the Amber99SB-ILDN force field with improved side-chain torsion potentials was chosen, as it is widely used for protein simulations and provides reliable parameters for all standard amino acids. Once the topology has been created, a simulation box filled with solvent molecules is defined. The box is characterized by its dimensions and shape and is usually centered on the protein center of mass. Boxes can be either cubic or rhombic dodecahedral; the latter choice is often preferred because it reduces the system volume by ~29% compared to a cube of the same side, thus lowering computational cost while still allowing sufficient space for molecular motion. On the other hand, a cubic box simplifies the subsequent trajectory analyses. In this work, a cubic box was preferred, with a minimum distance of 1 nm between any protein atom and the box boundary to avoid interactions with its periodic images. To ensure the reliability of the simulation, every protein system must be

solvated. The box size and geometry determine the number of solvent molecules to be included. The solvent is typically water, often supplemented with salts to mimic physiological conditions. In MD simulations, solvent molecules are explicitly represented, and the chosen water model must be consistent with the force field. Common models include SPC, SPC/E, TIP3P, and TIP4P. After solvation, the electrical neutrality of the system is ensured by replacing selected water molecules with ions, most commonly Na^+ or Cl^-, to balance the net charge of the protein. In all simulations performed for this thesis, an ion concentration of 150 mM was applied, and the system was then neutralized. Even in the described setup, simulating a single finite box of particles is insufficient to reproduce the properties of a real protein solution. To overcome this limitation, periodic boundary conditions (PBCs) are employed. With PBCs, atoms that exit the simulation box re-enter from the opposite side, and the system is effectively surrounded by its copies in all directions. This approach enables the simulation to approximate the bulk behavior of macroscopic systems. To ensure physical accuracy, the protein must never interact with one of its own periodic images, as this would introduce artifacts into the calculated forces. This condition is known as the minimum image convention. The last two preparation steps before production of the molecular dynamics simulations are energy minimization and equilibration of the system. At this stage, the protein is solvated, neutralized, and placed under PBCs. However, during the setup, it is not uncommon to introduce artificial strains into the system. For example, a distance between two atoms not sharing a bond shorter than the sum of their van der Waals radii causes an unnatural increase in energy. The energy minimization (EM) is performed to remove steric clashes and geometrical artifacts by relaxing the system to the nearest local energy minimum before the simulation begins. Different algorithms can be employed for minimization depending on the characteristics of the system, the most common being the steepest descent and the conjugate gradient methods. The equilibration step prepares the system for stable dynamics and is typically performed in two stages. In the first, the system is equilibrated in the NVT ensemble, i.e., the number of particles, volume, and temperature are kept constant. Here, the system is coupled to a thermostat that gradually brings the solvent molecules to the target temperature by changing their kinetic energies. At this stage, protein motion is not completely restricted, but deviations from the starting structure are energetically penalized, ensuring that solvent equilibration proceeds without major protein rearrangements. In the second stage, the system is equilibrated in the NPT ensemble, also known as the isothermal–isobaric ensemble, i.e., the number of particles, pressure, and temperature are kept constant. In this phase, the system is coupled to a thermostat (to control the temperature) and a barostat (to regulate the pressure). The application of external pressure allows the system to reach the correct density, thus reproducing more realistic environmental conditions for the simulation. Once the system is equilibrated at the desired temperature and pressure, the production run starts. In MD simulations of biological systems, production of microstates is generally carried out in the NPT

ensemble, with the same thermostat and barostat used during equilibration. Each microstate, also termed a snapshot, that is produced by the simulation refers to the configuration of the system at a specific time point. The complete sequence of snapshots generated along the simulation constitutes the trajectory, which contains the time evolution of the system and provides the basis for subsequent analyses. Data generated by the production run are analyzed to evaluate the stability and dynamics of the system, taking into account some key parameters:

- Potential energy: Usually, the potential energy of the system decreases during minimization until it stabilizes at a plateau (typically ~10^5-10^6 kJ/mol), indicating relaxation into a local minimum. Rarely, two or more stable states can be observed at different energy values, corresponding to different protein conformations.
- RMSD (Root Mean Square Deviation): This parameter tracks conformational changes by measuring the average atomic displacement from a reference structure over time.
- RMSF (Root Mean Square Fluctuation): This parameter represents the average in the timeframe of the simulation of the position of each atom or residue and is used to quantify the fluctuations and identify flexible versus rigid regions.
- Radius of gyration (Rg): The radius of gyration reflects the overall protein compactness, calculated as the mean square distance of atoms from the center of mass. It is used to monitor folding or unfolding during the simulation.

In the standard protocol applied in this thesis, a steepest descent algorithm was used for energy minimization of the protein chain. Minimization was stopped either when the maximum force fell below 1000.0 kJ/(mol·nm) or after 50,000 steps, whichever occurred first. The energy step size was set to 0.005 kJ/mol, using the Verlet cut-off scheme with short-range electrostatics and van der Waals interactions truncated at 1.0 nm. Simulations were carried out with the AMBER99SB-ILDN force field [73], placing the solvated protein complex in a cubic box, using the TIP3P model for water molecules. Equilibration was performed first in the NVT ensemble and then in the NPT ensemble, with the protein backbone constrained for 100 ps. This was followed by a NPT production run at 300 K. The integration time step was set to 2 fs, using the Verlet integrator and the LINCS algorithm to constrain hydrogen bond lengths, thereby avoiding explicit recalculation of hydrogen positions at each step. Periodic boundary conditions (PBCs) were applied in all three dimensions. For analysis, RMSD and RMSF values were calculated from configurations sampled every 10 ps (unless otherwise stated). All simulations and analyses were performed using GROMACS [74] v.2020.3, and computations were run on the Leonardo HPC facility (CINECA, Italy) [75]. In some occurrences, the standard protocol proved ineffective, particularly when simulating surfactant micelles. Due to their peculiar structural features, these systems did not relax in the same manner as more conventional biomolecular systems such as proteins. In these cases, the minimization step frequently ended with errors and failed to generate a

stable minimized conformation. In several instances, the systems became unstable and exploded, likely due to atomic overlaps correlated with unrealistically high energies. To address this issue, a more advanced minimization protocol was adopted, designed to allow the system to relax gradually. This method, originally developed and validated by Roe and Brooks [76], was adapted to GROMACS for this thesis.

The Roe's protocol

The protocol described by Roe & Brooks is a systematic method to set up biomolecular systems solvated in explicit water for stable MD simulations. This ten-step protocol combines energy minimization with short restrained MD simulations, progressively releasing restraints to allow the system to relax in a controlled manner. The steps reported here refer to the version adapted for GROMACS:

1) Initial energy minimization for solvent and ions, without changes in the macromolecular system: 1000 steps of steepest descent minimization, restraining all non-hydrogen atoms of macromolecules (proteins, nucleic acids, lipids) with a strong force constant (5000 kJ/mol·nm^2).
2) First solvent relaxation: 15 ps of MD equilibration in the NVT ensemble, 1 fs timestep. Solvent and ions move freely, while non-hydrogen atoms of the macromolecule remain restrained (5000 kJ/mol nm^2).
3) First energy minimization of the macromolecule: 1000 steps of steepest descent minimization with medium restraints (2000 kJ/mol nm^2) on non-hydrogen atoms of the macromolecular system, to gently release internal strain.
4) Second minimization of the macromolecule: 1000 additional minimization steps with weak restraints (500 kJ/mol nm^2), letting the structure further relax without collapsing.
5) Unrestrained minimization: 1000 minimization steps without restraints, to allow the stabilization of the whole system.
6) Initial macromolecule relaxation: 5 ps of restrained NPT MD equilibration (1 fs timestep), allowing pressure and density to adjust. Non-hydrogen atoms of the macromolecular system are restrained with a force constant of 1000 kJ/mol nm^2.
7) Continued macromolecule relaxation: 5 ps of additional NPT MD equilibration, weakening the restraints (500 kJ/mol nm^2) on non-hydrogen atoms of the macromolecule.
8) Relaxation of side chains: 10 ps of NPT MD equilibration (1 fs timestep), applying restraints only to protein backbone atoms (1000 kJ/mol nm^2). This step allows the side chains to relax while preserving the global secondary structure.
9) Unrestrained equilibration: 10 ps of NPT MD equilibration (2 fs timestep) without restraints, allowing the system to adjust.
10) Final density stabilization: additional NPT MD equilibration steps (in 1 ns increments) until the system density reaches a plateau. This step usually takes less than 1 ns. This ensures that the system is stable before MD production.

The strength of this protocol is its ability to progressively relax the system. First, only the solvent is allowed to equilibrate, while the macromolecular system is restrained. Then, restraints on the macromolecule are gradually released, starting from non-hydrogen atoms of the side chains, then the backbone, and finally the whole structure. This stepwise process permits resolving local steric clashes without compromising the overall conformation. Equilibration is assessed when the density of the system plateau, providing an objective and quantitative measure that the system has reached a sufficiently stable state to proceed to MD production. Roe and Brooks tested the reliability of this approach with nearly 400 systems of varying size and complexity, including proteins, nucleic acids, protein-DNA complexes, lipid bilayers, and carbohydrates, with system sizes ranging from a few thousand to over 850,000 atoms. Across the systems tested, the protocol effectively prevents common instabilities, such as blow-ups caused by atomic overlaps, improper solvation, or unrealistic densities, that often occur with simpler equilibration procedures. Importantly, this can be achieved with good computational efficiency: in most of the tested cases, density stabilization was reached within ~180 ps, enabling production runs to begin without unnecessary delays. Besides stability and efficiency, the protocol also enhances reproducibility. The clear and standardized workflow replaces the often inconsistent equilibration steps reported in the literature, thus ensuring that the MD simulation starts from a well-defined and reproducible point. In addition to the authors' validation, this protocol was applied and validated in two different systems, described in peer-reviewed publications by Guercia et al. [77] and by Miolo et al. [78].

Machine learning and AlphaFold

Machine learning is a branch of computer science that enables computers to learn, i.e., extrapolate information, from large datasets without explicit programming. Machine learning algorithms are trained to find patterns within dataset by adapting a defined number of internal parameters, often using features selected by programmers. Deep learning, which is a subset of machine learning, uses a network composed of thousands of layers, hence the adjective *deep*, that automatically learns complex patterns directly from raw data. AlphaFold is a series of deep-learning models developed by DeepMind (Alphabet) that predicts a 3D protein structure from its amino acid sequence with remarkable accuracy [79]. The first version of AlphaFold (2018) won the CASP13 competition for the best structure prediction algorithm. In this challenge, different software algorithms are applied to predict the structure of proteins whose structures, albeit solved by experimental methods, are unknown to the participants and kept on hold in the PDB. In 2021, the second version of AlphaFold2 (AF2) was released, introducing an innovative approach [80], [81], and in 2024, Alphafold3 further advanced the method by integrating predictions for ligands, RNA, and DNA [82]. The basic idea behind the AF algorithms is that protein evolution is mostly neutral as regards the overall conformation: mutations usually do not significantly affect the protein structure and thus its function.

For example, the 3D structure of a protein that changes by 80% through evolutionary steps often remains almost the same mainly unvaried [81], as the structure is usually more conserved than the sequence. An important observation is that a residue conserved in a specific position usually implies a role in protein function. Moreover, co-evolution of two residues, i.e. two mutations in homolog proteins, suggests an interaction between them. These ideas are the basis for the AF2 algorithm. While the implementation is quite complex, the application of this approach in AlphaFold can be summarized in simple terms. First, the deep learning model is trained on the PDB database using structural information from experimental data. Then, the trained model estimates the structure of a given sequence, in three main steps:

1- AF2 conducts a multiple sequence alignment (MSA) by searching databases for homologs of the sequence. Then, AF2 generates an MSA representation and pair representations, which report in matrices the structural interactions of related proteins with known structures.
2- Subsequently, this information is encoded and passes through a deep learning module (called Evoformer).
3- The output of Evoformer is decoded into the 3D structure of the protein.

In the post-processing step, AF2 applies relaxation algorithms to resolve clashes without impacting the prediction accuracy. Amber99sb force field is used for this process. As a result, the post-processing yields several, usually five, relaxed 3D structures of the protein with a relative ranking, and metadata encoding for prediction reliability and sequence coverage. AF is a proprietary technology working with a large-size model. Thus, the weights, i.e., the trained parameters, of the Evoformer model are not publicly available. However, DeepMind releases a lighter version of the full software that can be run through a web application. Scientists can use AF2 in a Google Colab notebook [83], [84], and AlphaFold3 via its own web app (alphafoldserver.com). In this thesis, AF3 was applied for the prediction of the structure of the Wzx-Bril chimera and AF2, in Colab, was used to predict the structures of other constructs of the Wzx protein. In addition, the structures of several constructs of the endoglycosidase protein were estimated using the CINECA HPC [75], as the size and the complexity of the estimation exceeded the computational resources available through Colab. In this case, a JupyterLab server was activated on the LEONARDO computational facility [85], and Colab was used via a remote runtime connection. In line with the conservative hypothesis regarding the structure of homolog proteins, a question may arise: in Figure 10 ten homologs of Wzx are aligned, revealing a consensus sequence; it can be assumed that residues conserved across all the proteins are essential for the flippase activity, on the other hand, we can hypothesize that the non-conserved ones might determine the substrate specificity. Unfortunately, this theory is difficult to verify experimentally.

Materials & Protocols

Plasmids

Plasmid of the protein chimera between Wzx from *P. aeruginosa* and GFP

A plasmid for the expression of the protein Wzx from *P. aeruginosa* was already available in the Laboratory of Biocrystallography of the Department of Chemical and Pharmaceutical Sciences (University of Trieste) before the beginning of this thesis (Figure 18). In this plasmid, for which the vector pWaldo (Addgene) was used, the wzx gene from PAO1, followed by the gene of the Green Fluorescent Protein (GFP) and a tag formed by 10 histidine residues (10xHis) are present after a T7 RNA-polymerase promoter. For membrane proteins, tags such as the GFP protein or the histidine sequence are usually placed at the C-terminus to avoid interference with the N-terminal signal peptide, which is essential for proper protein inclusion in the membrane. The GFP tag was included in the construct to enable direct monitoring of the protein during both expression and purification, by using the GFP fluorescence signal [86]. Moreover, loss of fluorescence might indicate misfolding, aggregation, or precipitation of the protein. The 10xHis tag was introduced with the aim to purify the protein by affinity chromatography using divalent metal ions such as Ni^{2+} or Co^{2+}, a purification technique also known as Immobilized Metal Affinity Chromatography (IMAC). In addition, the construct contains the gene sequence for the thrombin cleavage site, located at the C-terminus of the protein and at the N-terminus of the GFP sequence. This protease recognition site was inserted to allow easy tag removal by enzymatic reaction. The pWaldo vector includes the antibiotic resistance gene for Kanamycin, useful for colony selection (Figure 18).

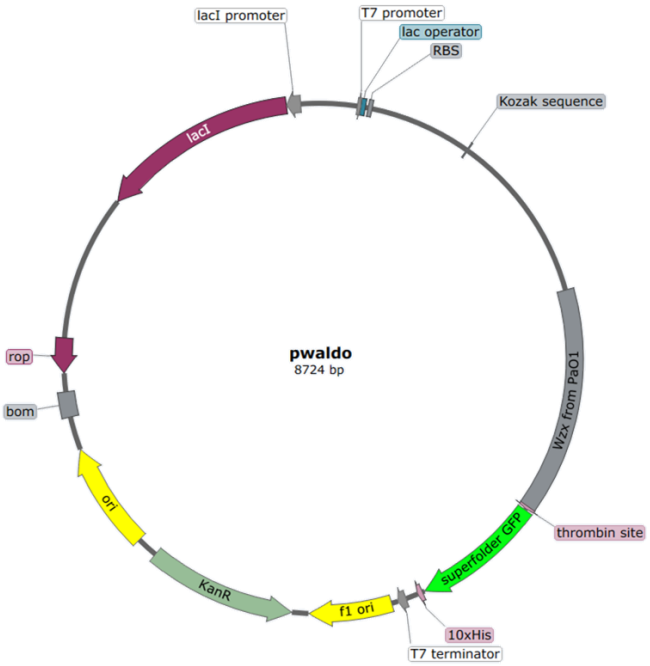


Figure 18 – Schematic representation of the pWaldo plasmid for protein expression, containing the wzx-GFP chimera.

The plasmid containing the wzx gene is suitable for protein expression in a bacterial strain containing the DE3 prophage. This system affords a fine tuning of the protein expression by modulating the concentration of IPTG (β-D-1-thiogalactopyranoside), an analog of a lactose metabolite, the allolactose (Figure 19). This compound is able to bind the *LacI* protein, an inhibitor of the *lac* promoter expressed by *E. coli* cells. In the absence of IPTG, *LacI* binds to the *LacO* sequence, thus blocking transcription of the target gene. By binding to IPTG, the *LacI* is removed from the promoter located in the sequence of the prophage integrated in the bacterial genome of the DE3 cells. The activation of the *lac* promoter allows expression of the T7 RNA-polymerase. This enzyme, in turn, is able to bind to the T7 promoter present on the pWaldo plasmid and activate the transcription of the following gene.

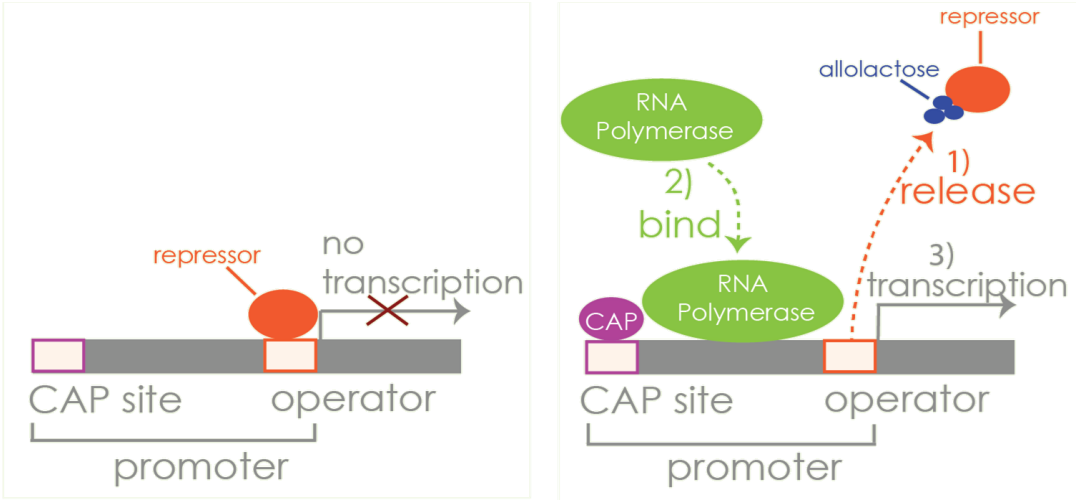


Figure 19 – Activation mechanism of the lac promoter. Both the allolactose and the IPTG, a non-metabolizable analog of allolactose, are able to activate the transcription of the following gene. on the left, the behavior in absence of allactose (or IPTG), on the right the activation in presence of IPTG or allolactose. (from goldbio.com)

Plasmid of the protein EG from *K. pneumoniae*

A plasmid containing the gene of the phage endoglycosidase protein against *P. aeruginosa* was obtained by Demi Vattovaz Ph.D. and Michela Zaro Ph.D, in the Laboratory of Biocrystallgoraphy of the Department of Chemical and Pharmaceutical Sciences (University of Trieste). To produce this plasmid, the gene was inserted into a pET302/NT-His vector (InvitroGen). The construct encoded by this plasmid has a N-terminal tag of 6 histidine residues (6xHis). Previous results showed that, in this case, only the N-terminal tag produces a viable protein with enzymatic activity. When the histidine tag was fused with the C-terminus of the protein sequence, the enzyme was not active, probably due to an altered protein folding, as suggested by AlphaFold generated models. A Tobacco Etch Virus (TEV) protease cleavage site was included between the 6xHis tag and the endoglycosidase sequence, to enable the removal of the affinity tag. The pET302 vector contains a gene associated with Ampicillin resistance for colony selection.

Cloning and primer design

Cloning of the Wzx(*P. aeruginosa*)-TEV cleavage site-GFP-10xHis construct

The initial construct of the Wzx protein was expressed, solubilized, and purified. However, when the enzymatic reaction for tag removal was attempted, the results were poor. A possible reason for this failed attempt was identified in the presence of the detergent, required for protein retrieval from the membrane and purification: the activity of the chosen protease, i.e., thrombin, may be affected by the denaturing effect of the detergent. Thus, to improve the efficiency of tag removal, the cleavage site was modified and the Tobacco Etch Virus (TEV) protease was selected as a more reliable alternative to thrombin. TEV protease shows an exceptionally high sequence specificity, recognizing the ENLYFQ↓S motif and cleaving only in this specific position. For this reason, the original thrombin cleavage site (ThrbClSt) was replaced with the TEV recognition sequence. To generate the new construct, the plasmid containing the Wzx-ThrbClSt-GFP-10xHis construct was used as a template and new cleavage site sequence was introduced via a PCR method, using a site-directed mutagenesis approach. Synthetic oligonucleotides (primers) were designed carrying the new nucleotide sequence and a sequence annealing to the template plasmid. During the PCR reaction these primers were extended by the DNA polymerase incorporating the modified sequence. A first attempt was made by designing two primers that contained the whole sequence corresponding to the TEV cleavage site, split in two halves (Figure 20). The online tool NEBaseChanger (www.nebasechanger.neb.com) was used to optimize the primer design.

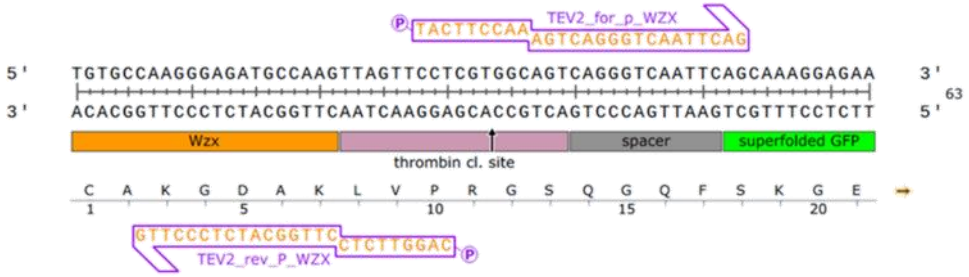


Figure 20 - Primers design for the substitution of the thrombin cleavage site introducing the TEV cleavage site in a single step. (This picture is obtained from SnapGene visualization software)

Phosphorylated primers at the 5' end were used to facilitate the subsequent ligation reaction. All primers used in this thesis were purchased from Merck as custom-made oligonucleotides. The PCR reaction was performed using the Q5 High-Fidelity DNA Polymerase kit (New England Biolabs), following the protocol suggested by the manufacturer, for a total volume of 25 µL:

- 5 µL of 5x Q5 reaction buffer;
- 0.5 µL of 10 mM dNTPs solution (200 µM);
- 1.25 µL of 10 µM forward primer (0.5 µM);
- 1.25 µL of 10 µM reverse primer (0.5 µM);
- 0.5 µL template DNA (25 ng);

- 0.5 µL of Q5 High Fidelity DNA Polymerase (0.04 U/µL);
- Ultrapure H_2O to reach a final volume of 25 µL.

The thermal cycler program was set as follows:

- Preheating of the lid at 105°C, essential to avoid solution condensation on the lid;
- Initial denaturation at 98°C for 30 seconds;
- 35 cycles:
    - Denaturation at 98°C for 10 seconds;
    - Annealing at a temperature depending on primer sequence for 30 seconds;
    - Elongation at 72°C for 90 seconds (10 s/kb)
- Final elongation at 72°C for 120 seconds;
- Final hold at 8°C.

Annealing temperature for each primer was calculated with NEBTmCalculator (tmcalculator.neb.com) and a temperature of 67°C was selected for the PCR reaction. The PCR procedure generates a non-methylated, linear, double-stranded DNA. After the reaction, the product was treated with the KLD enzyme mix (New England Biolabs), which contains a kinase to phosphorylate the PCR product, a ligase to join the two ends of the product in a circular plasmid, and the restriction enzyme DpnI. The DpnI enzyme is able to digest only methylated DNA fragments and is used to remove the template plasmid, which was previously amplified in bacteria and is, hence, methylated. 10 µL of the digested product were used to transform 50 µL of *E. coli* DH5α competent cells (the transformation protocol is reported below). The transformed cells were plated on LB agar supplemented with Kanamycin to select colonies containing the modified plasmid. After incubation, colonies were picked and transferred in 7 mL of fresh LB medium supplemented with kanamycin and incubated at 37°C overnight. Plasmid DNA was purified using an extraction kit (GenEluteTM Plasmid Miniprep Kit, Merck). To confirm the correct introduction of the nucleotide sequence corresponding to the TEV cleavage site, the cloning product was analyzed by sequencing (Eurofins Genomics) employing the T7 term primer. Unfortunately, the first attempt at introducing the new sequence was unsuccessful. A possible reason for the failure of this reaction could be the length of the DNA segment to be inserted. A second attempt was performed in two steps, thus reducing the number of bases introduced in each PCR reaction (Figure 21). The protocol for each step was the same as previously described. The primers designed for the two reactions (Figure 20 and Figure 21) had an annealing temperature, calculated with the NEBTmCalculator tool, of 70 and 65°C, respectively. The initial plasmid, containing the Wzx-ThrbClSt-GFP-10xHis construct was used as template DNA in the first reaction, while the product of the first was used as template for the second. After the two-step cloning attempt, sequencing revealed the correct insertion of the cleavage site for the TEV protease.

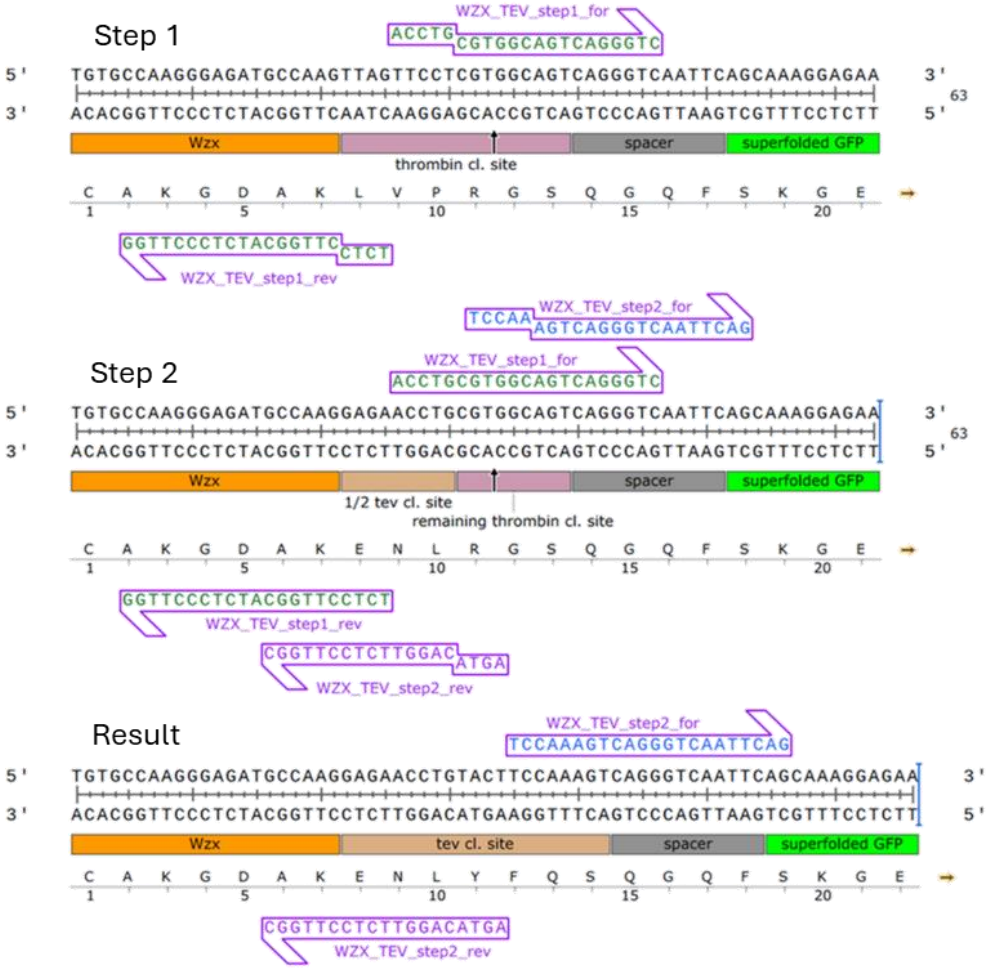


Figure 21 -Representation of primer for introducing the mutation (TEV cleavage site), in two subsequent steps. (This picture is obtained from SnapGene visualization software)

Cloning of the Wzx-BRIL-Thrombin cleavage site-GFP-10xHis construct

AlphaFold generated models revealed that the Wzx protein could display a limited hydrophilic surface area for crystal formation and an intrinsic tendency to adopt multiple conformations. These features make its crystallization challenging. Moreover, the presence of the detergent micelle, which surrounds the protein and is required for solubilization, can interfere with specific protein–protein interactions, further complicating crystal formation. To facilitate crystallization, the BRIL scaffold was fused to the C-terminus of Wzx. The BRIL scaffold is based on the sequence of the apocytochrome b_{562}RIL, a small, four-helix bundle protein commonly used to facilitate the crystallization of membrane proteins and previously applied on transporters, receptors, and GPCRs [86]. BRIL provides a rigid and hydrophilic folded domain, increasing the available surface area for crystal contacts and exposing regions capable of forming intermolecular interactions within the crystal. To generate the Wzx-BRIL fusion protein, the gene encoding BRIL was inserted into the previously available pWaldo plasmid containing the thrombin cleavable construct, using the Restriction-Free (RF) cloning method (Figure 22).

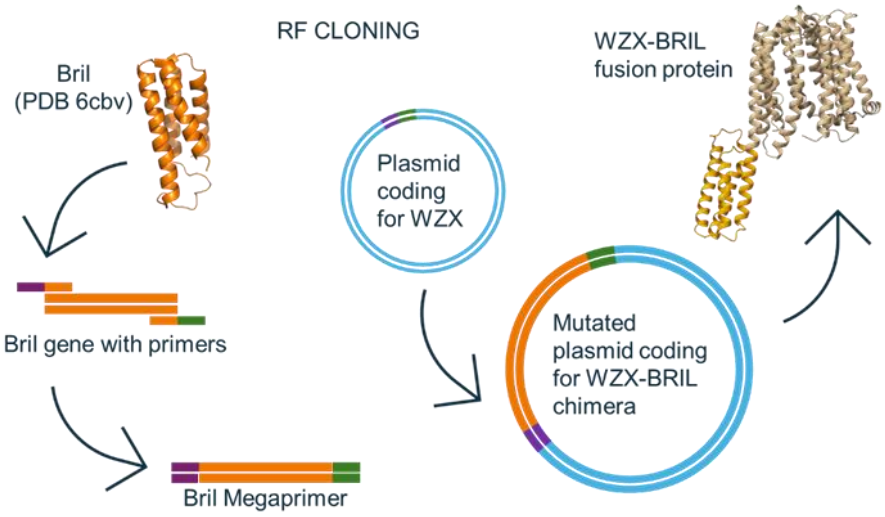


Figure 22 - Schematic representation of the RF cloning procedure followed to insert the BRIL fragment into the Wzx-Thombin cleavage site-GFP-10xHis construct.

Unlike conventional techniques, RF cloning allows large vector modifications without restriction enzymes but taking advantage of the PCR. The procedure involves two sequential PCR reactions. In the first, a high-fidelity DNA polymerase amplifies the DNA fragment to be inserted, attaching terminal tails at both 5' and 3' complementary to the sequence of the vector where the fragment has to be inserted. This PCR produces a fragment, called megaprimer (MP), which is used as primer in the second reaction, in which it anneals to the vector template. The result of this second PCR is a circular plasmid containing the inserted segment, with two nicks in different positions of each strand. The strands are later repaired when the DNA is amplified in a bacterial culture. The *bril* gene sequence coding for the protein in the PDB structure 1M6T [87] were evaluated, as codon-optimized for *E. coli* expression by the tools from GenSmart and Twist Bioscience. The final construct designed for this work is based on the Twist Bioscience sequence, which is reported in "Appendix - BRIL gene". The selected sequence was purchased from Twist Bioscience. The primers for the first step of the RF protocol were designed to have complementary sequences to both sequence to be inserted and the target vector. The 3' end of each primer, i.e., forward and reverse, anneals to the gene of interest, enabling megaprimer formation, while the opposite end is complementary to the vector [88]. The online tool rf-cloning.org was used to design the primers. The first PCR was set up as follows:

- 10 µL of 5x Q5 reaction buffer;
- 1 µL of 10 mM dNTPs solution (200 µM);
- 2.5 µL of 10 µM forward primer (0.5 µM);
- 2.5 µL of 10 µM reverse primer (0.5 µM);
- 25 ng DNA template (*bril* gene);
- 0.5µL Q5 High Fidelity DNA Polymerase (0.02 U/µL);
- Ultrapure H_2O to reach a final volume of 50 µL.

Thermal cycler conditions:

- Preheating the lid at 105°C;
- Initial denaturation at 95°C for 30 seconds;
- 35 cycles:
    - Denaturation at 95°C for 10 seconds;
    - Annealing at the 59°C for 20 seconds;
    - Extension at 72°C for 10 seconds;
- Final elongation at 72°C for 2 minutes;
- Final hold at 8°C.

The PCR product was separated in a 0.8% agarose gel. Agarose gels were prepared in-house by dissolving 400 mg of agarose in 50 mL of TAE buffer (Tris 40 mM, acetic acid 20 mM, EDTA 1 mM, pH 8). The mixture was heated until complete dissolution, then cooled above room temperature. Before solidification of the gel, 2.5 µL of intercalating agent (Eurosafe, from EuroClone) was added, then the solution was poured into a casting tray and further cooled at room temperature to allow gel formation. Samples were mixed with loading buffer (Gel Loading Dye, Purple 6x, New England Biolabs) for tracking the DNA during the run, and loaded in the gel. After electrophoresis (45 minutes at 100 V) the band corresponding to the megaprimer was identified under UV light and excised from the gel. The exposure of the gel to UV light was limited to avoid DNA damage. The DNA was collected by using the gel extraction kit (GenElute, Merck). The quantity and the purity of the MP were estimated using a SimpliNano spectrophotometer (BioChrom), measuring the sample absorbance at 260 nm. The isolated MP was used in the second PCR:

- 4 µL of 5x Q5 reaction buffer;
- 0.4 µL of 10 mM dNTPs solution (200 µM);
- ~2 µg of Megaprimer;
- 100 ng Wzx-Thrombin cleavage site-GFP-10xHis plasmid;
- 0.2 µL Q5 High Fidelity DNA Polymerase;
- Ultrapure H_2O to reach a final volume of 20 µL.

Thermal cycler conditions:

- Preheating the lid at 105°C;
- Initial denaturation at 95°C for 30 seconds;
- 15 cycles:
    - Denaturation at 95°C for 10 seconds;
    - Annealing and extension at 72°C for 90 seconds;
- Final elongation at 72°C for 2 minutes;
- Final hold at 8°C.

The product was used to transform 50 µL of *E. coli* DH5α and the amplified plasmid was purified, as previously described. The plasmid was sequenced to confirm the correct insertion of the *bril* gene.

Bacterial transformation

Protocol for the transformation with plasmids containing the Wzx protein

The protein Wzx from *P. aeruginosa* PAO1 was produced through recombinant expression in the *E. coli* Lemo21(DE3) bacterial strain (New England Biolabs), which revealed the highest protein expression level in a preliminary expression test, despite the fact that for membrane proteins a high expression level does not necessarily correlate with correct protein folding. In Lemo21(DE3) system, the activity of the T7 RNA polymerase can be precisely controlled by tuning the production of its natural inhibitor T7 lysozyme encoded in the plasmid pLemo carried by the selected strain [89]. The pLemo plasmid contains a gene for resistance to the antibiotic Chloramphenicol. Competent cells were transformed with the following protocol:

- 50 µL of Lemo21(DE3) *E. coli* competent cells (New England BioLabs) were thawed at 4°C. 1 µL of plasmid (75 ng/µL) was added.
- Cells were then incubated for 30 minutes on ice, followed by 10 seconds at 42°C (heat shock).
- The suspension was incubated in ice for 5 minutes and then 950 µL of LB broth (Tryptone 10 g/L, 5 g/L Yeast extract, 5 g/L NaCl, all reagents were purchased from Merck), was added to the mixture, allowing cells to express antibiotic resistance genes.
- After 60 minutes at 37°C, shaking vigorously, cells were diluted in 10 mL LB supplemented with Chloramphenicol (30 µg/mL) to maintain the pLemo plasmid and Kanamycin (50 µg/mL) to maintain the modified pWaldo plasmid, and placed in a shaking incubator overnight.

When expression tests were performed with other strains, the same protocol was followed with two changes: the heat shock time was increased to 40 seconds, and the Chloramphenicol was omitted from the medium since resistance to this antibiotic is conferred by the pLemo plasmid.

Protocol for the transformation with plasmids containing the endoglycosidase protein

The endoglycosidase protein was overexpressed in *E. Coli* BL21(DE3) cells. The production protocol was developed by Michela Zaro, Ph.D. The transformation protocol involved the following steps:

- 50 µL of competent cells was thawed in ice for 20 minutes, 1µL of plasmid ~50 ng/µL was added.
- Cells were placed on ice for 30 minutes. A heat shock at 42 °C for 45 seconds was followed by 5 minutes on ice.
- 350 µL of LB broth was added to the cell suspension, which was incubated at 37 °C with shaking (180 rpm) for 1 hour.
- The cells were then diluted in 10 mL of LB broth with 50 µg/mL of Ampicillin. The mixture was then incubated overnight at 37 °C with shaking at 180 rpm.

Protein expression

Expression protocol for the Wzx protein from *P. aeruginosa*

Protein expression was optimized by systematically varying parameters such as host strain, induction time, temperature, inducer concentration, etc. For expression test experiments, cells were grown in 10 mL of LB under various conditions and protein production was evaluated by optical density to estimate cell growth and fluorescence to measure GFP expression. The following protocol was used for protein expression:

- Transformed cells were inoculated 5:500 in LB broth supplemented with 50 µg/ml Kanamycin and, only for tests with Lemo21 cells, 30 µg/ml Chloramphenicol at 37°C until OD_{600} = 0.4 -0.5. When pre-warmed medium was used for the initial dilution, this step took about 3 hours.
- The protein expression was induced with 0.4 mM IPTG and the flasks were incubated at 18°C overnight (14-18 hours).
- After expression, the cells were harvested by centrifugation at 5000xg for 10 min.
- For washing, the pellet was resuspended in 100 mL of resuspension solution (TBS pH 8.0, PMSF 1 mM) and collected again by centrifugation at 4430x g for 20 minutes.
- If not immediately processed, the pellet was stored at -80°C until further use.

For protein solubilization and purification, the bacterial pellet corresponding to 3 L of culture was thawed and resuspended in 50 mL of lysis solution (TBS pH 8.0, PMSF 1 mM, Lysozyme 1 mg/mL, DNAse I 100 µg/mL, $MgCl_2$ 25 µM). The suspension was then incubated at room temperature for 15 minutes to enable the catalytic reactions of lysozyme and DNAse to occur. This step improved the lysis of the cell wall of bacteria, while digesting the DNA released in the process. Afterwards, a homogenizer was used to disrupt the cells by applying a 120 MPa pressure while keeping the cell suspension on ice. The homogenization procedure was repeated for several consecutive cycles. While usually 5 cycles were sufficient to obtain a satisfactory lysis of the cells, the homogenization is highly dependent on the viscosity and density of the suspension, therefore the number of cycles was corrected according to sample properties. The lysate was centrifuged at 4430xg for 30 min at 4°C to remove unbroken cells and large debris.
The supernatant was further centrifuged using an ultracentrifuge at 100 000xg, for 60 minutes at 4°C, to collect the membrane fraction. Membranes were collected after carefully removing the bottom portion of the pellet, which contained leftover cell debris, inclusions bodies and other bacterial components. Membranes were homogenized with a glass tissue grinder, suspending them in 1-2 mL of TBS and PMSF 1 mM solution, and stored at -80°C.

Expression protocol for the phage endoglycosidase protein against *K. pneumoniae*

The transformed bacteria were diluted 1:500 in LB broth with 50 µg/mL of Ampicillin and incubated at 37°C, with shaking at 180 rpm, until reaching OD_{600} = 0.6. Finally, protein expression was induced by adding IPTG to a final concentration of 0.5 mM. The culture

was incubated overnight at 18°C, with shaking. Cells were harvested by centrifugation, the pellet was resuspended in HEPES 50 mM pH 7.4, NaCl 400 mM, TRITON-X100 0.5% v/v, and lysozyme. After 15 minutes of incubation at 25°C, the suspension was lysed by sonication on ice at 45 W, in five cycles of 60 seconds with 1 minute of rest. Cellular debris was removed by centrifugation at 6000xg for 1 hour. The supernatant, containing the soluble enzyme, was collected and used for the following protein purification step.

Solubilization in detergents (Wzx)

Detergents play a critical role in the characterization of membrane proteins. Thanks to their amphipathic nature, detergents are able to interact with transmembrane regions of membrane proteins, mimicking the lipid bilayer and maintaining them in solution. Selecting the appropriate detergent, or sometimes detergent mixture, is crucial as it strongly influences protein yield and stability during purification and structural characterization. The chemical structure of a detergent molecule determines its properties and suitability for protein solubilization. Ionic detergents are capable of solubilizing most membrane samples due to their charged groups but, on the other hand, they often denature proteins by disrupting their native secondary structure. Non-ionic detergents, such as DDM (n-dodecyl-β-D-maltoside) or DM (n-decyl-β-D-maltopyranoside) are milder and usually preserve protein conformation and activity, making them a popular choice for structural studies. Zwitterionic detergents (e.g., LDAO, Lauryl dimethylamine oxide) carry both positive and negative charges but have an overall neutral charge, offering an intermediate solubilization strength. Harsh detergents, i.e., detergents able to disrupt lipid–protein interactions, are usually applied to extract proteins from membranes in an efficient way. Milder detergents, such as non-ionic and zwitterionic detergents, are generally preferred for membrane protein purification as they preserve structural integrity and maintain a native-like protein conformation. DDM and DM have been extensively used to purify transporters, channels, and receptors [90]. Recently, neopentyl glycol (NG)-based detergents became popular as detergents for protein solubilization for structural studies, as they offer greater stabilization compared to other mild detergents [90]. Besides selecting the detergent, its concentration must also be carefully considered. The CMC (critical micelle concentration) defines the minimal detergent concentration of a specific detergent required for micelle formation. At lower concentration, the detergent is dispersed as a single molecule. To prevent micelle dissociation and protein aggregation, the detergent concentration should remain above the CMC at all times. However, an excessively high detergent concentration can destabilize the protein structure or interfere with downstream analyses. In standard protocols, high detergent concentration, usually 10 times the CMC, is used during the solubilization step to maximize the protein extraction from the lipid membrane, while purification is performed at concentrations equal or slightly above the CMC to minimize the denaturing effects. For Wzx protein purification, the membrane fraction from 3 L of culture was thawed on ice and resuspended in a TBS solution at pH 8, supplemented

with 1 mM PMSF and the chosen detergent. Usually, a total volume of 40 mL of solubilizing buffer was used, taking care of dissolving the membrane at a nearly constant concentration of both salt and detergent, thus avoiding exposure of the membrane proteins to high detergent concentration, such that of the detergent stock solution. In the standard protocol, the membrane suspension was added slowly and gradually to the detergent solution, allowing a sufficient time for each membrane suspension drop to dissolve. To maximize yield, the suspension was further incubated for one hour at 4°C under gentle rotation. Finally, the sample was separated by centrifugation to remove undissolved pellet at 4430xg for 20 min.

Protein purification

Membrane proteins are purified as protein–lipid–detergent complexes. Since these complexes are soluble in aqueous buffers, the purification can be performed using chromatographic techniques similar to those applied for soluble proteins. The key difference is that detergents must be included in all buffers, since protein–detergent complexes are in dynamic equilibrium and micelles would otherwise disaggregate, leading to protein aggregation or denaturation. Both the phage endoglycosidase protein against *K. pneumoniae* and the Wzx membrane protein of *P. aeruginosa* were purified using a two-step approach, with an affinity chromatography (IMAC) followed by a gel filtration (SEC). Additional purification steps were employed to remove residual contaminants or tags after enzymatic cleavage.

IMAC

Immobilized Metal Affinity Chromatography (IMAC) is an affinity purification method in which proteins are separated according to their ability to interact with metal ions (e.g., Zn^{2+}, Co^{2+}, or Ni^{2+}), immobilized on an inert polymer matrix. Amino acids residues present in proteins, such as histidine, tryptophan, and cysteine, can coordinate these exposed metal ions. This technique is especially effective for the purification of recombinant proteins carrying poly-histidine tags. The most used matrices are those chelating Co^{2+} or Ni^{2+} ions. A buffer containing imidazole was used to displace the tagged proteins from the matrix. For the IMAC purification step of the Wzx protein, the Talon Superflow (GE Healthcare) resin was washed with water to remove the 20% ethanol storage solution, and, subsequently, with an equilibration buffer containing TBS, 1mM PMSF and the detergent at the solubilizing concentration. 1 mL of resin was used for every liter of cell culture. The solubilized membrane sample was incubated for 60 minutes at 4°C with the equilibrated resin to allow the interaction of the tagged protein with the metal ions in the matrix. The incubation was carried out in a centrifuge tube with a gentle rotation to avoid foam formation due to the presence of the detergent. The suspension was transferred into a column to collect the resin and discard the flow-through (FT). To remove weakly bound contaminants, the resin was washed with 15 mL of Wash Buffer 1 (TBS 1x, 10xCMC detergent, 1mM PMSF) (WB1), corresponding to 5 column volumes (CV). An

additional wash step was performed with 5 CV of the same buffer supplemented with 5 mM of imidazole (Wash Buffer 2, WB2). Finally, the target protein was collected in ~500 µL fractions using 4 CV of elution solution containing 150 mM of imidazole in (TBS 1x, 10xCMC detergent, 1mM PMSF, 150 mM imidazole). A sample from each step of the purification (FT, WB1, WB2 and elution fractions) was later analyzed by SDS-PAGE. Finally, after elution, the resin was washed with water, with a buffer solution containing MES 20 mM pH 5, NaCl 100 mM, and again with water. The resin was stored in 20% ethanol. For the purification of the endoglycosidase protein, 6 mL of Talon resin was used for the amount of lysate from 200 mL of culture. After removing the ethanol, the resin was equilibrated with 20 CV of HEPES 50 mM pH 7.4, NaCl 400 mM. To remove leftover debris, the lysed bacterial culture was filtered with a gauze, and then incubated with the pre-treated cobalt resin for 1 hour at 4°C under agitation. After incubation, the suspension was transferred into a column. The unbound material was removed by gravity (FT), and the resin was sequentially washed with two buffers: the first containing HEPES 50 mM pH 7.4, NaCl 400 mM (10 CV, WB1), and the second HEPES 50 mM pH 7.4, NaCl 400 mM supplemented with 5 mM imidazole (10 CV, WB2). The endoglycosidase protein was eluted using HEPES 50 mM pH 7.4, NaCl 400 mM, containing 150 mM imidazole, and collected in five aliquots. The presence of the target protein in the eluted fractions was verified by SDS-PAGE.

SDS-PAGE

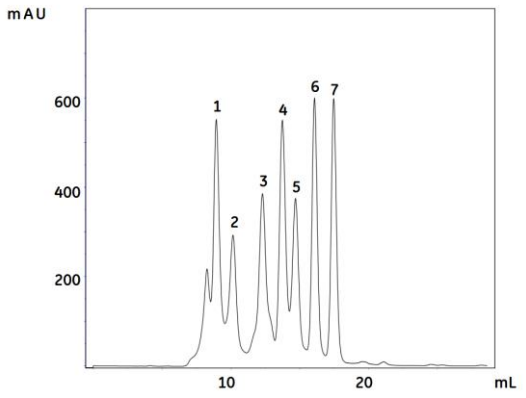
SDS–PAGE (Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis) allows fast protein analysis by separating polypeptides according to their molecular weight. Samples are mixed with a loading buffer containing SDS that denatures the proteins and covers the unfolded polypeptide chains with negative charges. Samples are then loaded onto a polyacrylamide gel and a tension (in this work a 190V was applied for SDS-PAGE separation) is applied to the gel. Protein migrate through the gel matrix, whose pore size depends on the acrylamide concentration, toward the anode. Lower acrylamide concentrations produce gels with larger pores that favor the migration of high molecular weight proteins, whereas higher concentrations yield smaller pores suitable for resolving low molecular weight species. Gels were prepared in-house. Since all constructs containing the Wzx protein had an estimated molecular weight of 70-75 kDa (excluding the bril derivate), a 10% polyacrylamide gel was used for the separation of these protein samples. Each sample was prepared by mixing 15 µL of protein solution with 5 µL of 4x loading buffer (Tris 120 mM pH 6.8, SDS 8%, glycerol 40%, β-mercaptoethanol 14%, bromophenol blue 0.1%) and loaded into the wells of the gel. Additionally, 2 µL of molecular weight marker (mPAGE, Millipore) was included as a reference ladder. SDS-PAGE gels used to analyze the endoglycosidase protein samples were prepared using a 12% polyacrylamide concentration, due to the lower molecular weight of the polypeptide chain (50 kDa). Samples were prepared as described above, but with an additional denaturation step obtained by boiling the samples at 100°C for 10 minutes. After

electrophoresis, gels were stained with ProBlue Safe Stain (Giotto Biotech), which contains the Coomassie Brilliant Blue dye that binds to hydrophobic regions of denatured proteins.

SEC

To further purify the protein samples, fractions obtained from the IMAC chromatography underwent another step of separation by Size Exclusion Chromatography (SEC) using an FPLC system (AKTA-Go FPLC, Cytiva). In gel filtration chromatography, proteins are separated according to their molecular size; therefore, the choice of the column depends on the size of the target protein. In the case of the Wzx protein, a Superdex 200 Increase 10/300 GL (Cytiva) was used, which is suitable for the separation of proteins in the range of 100-300 kDa (Figure 23). While the molecular mass of the polypeptide chain is expected to be in the range of 70-75 kDa, the protein is surrounded by a detergent micelle that increases its hydrodynamic radius and, thus, increases its apparent molecular weight. Prior to separation, the sample was concentrated to approximately 700 µL using a 10 kDa molecular weight cutoff concentrator (Vivaspin, Sartorius). The solution was centrifuged at 12000xg for 10 minutes and subsequently loaded onto the column. Gel filtration was performed using TBS buffer, pH 8.0, containing detergent as the mobile phase, at a flow rate of 0.33 mL/min. The chromatographic separation was monitored by measuring the absorbance of the solution at 280 nm to identify fractions containing proteins. Based on the standardization of the chosen column, the Wzx protein was expected to elute with a retention volume of 11-12 mL.

Superdex 200 Increase 10/300 GL



Sample (100 *µL)*:

1. Thyroglobulin (Mr 669 000) 3 mg/mL
2. Ferritin (Mr 440 000) 0.3 mg/mL
3. Aldolase (Mr 158 000) 3 mg/mL
4. Conalbumin (Mr 75 000) 3 mg/mL
5. Ovalbumin (Mr 44 000) 3 mg/mL
6. Carbonic anhydrase (Mr 29 000) 3 mg/mL
7. Ribonuclease A (Mr 13 700) 3 mg/mL

Eluent: 0.01 M phosphate buffer, 0.14 M NaCl, pH 7.4
Flow rate: 0.75 mL/min, room temperature
Detection: 280 nm

Figure 23 - Chromatogram of the standardization of the Superdex 200 Increase column, as reported in the manufacturer's manual.

In the purification of the endoglycosidase protein, a SEC chromatographic separation was carried out on the same ÄKTA-Go instrument but using a Superdex 200 10/300 GL column (Cytiva). This column has a similar range as the Increase column described above and is therefore suitable to purify this enzyme, which is a trimer in its native state. A solution of HEPES 50 mM pH 7.4 and 400mM of NaCl was used as mobile phase.

Protein quantification with BCA assay

The bicinchoninic acid assay (BCA) is a colorimetric method used to quantify protein concentrations. It relies on the redox reaction between peptide bonds and Cu^{2+} ions, which are oxidized to Cu^{+} under alkaline conditions. The Cu^{+} ions are, in turn, chelated by the bicinchoninic acid forming a purple complex. The absorbance of the complex measured at 562 nm is proportional to the concentration of the peptide bonds in the sample, which is approximately proportional to the protein concentration, expressed as mass over volume (mg/mL). The BCA method was chosen as quantification method because it is not sensitive to the presence of detergents commonly used for membrane protein extraction and solubilization. In addition, the BCA assay is a reliable quantification method as the color development is stable for several hours, and it is easily adaptable for high-throughput applications. Its sensitivity ranges between 200 µg/mL to 1000 µg/mL covering the protein concentrations used in this thesis. Protein quantification was performed in 96 wells plates using a standard curve prepared with bovine serum albumin (BSA) ranging from 0 µg/mL (buffer only) to 1000 µg/mL. Each measurement was performed in technical triplicate by mixing 25 µL of the protein solution (either the BSA standard or the sample to be quantified) with 200 µL of BCA reagent (Bicinchoninic Acid Protein Assay Kit, Sigma-Aldrich). The plate was sealed with a transparent film to minimize evaporation and incubated for 30 min at 37 °C. The absorbance at 562 nm of the standard solutions was measured to obtain the calibration curve. By interpolation, the absorbance of the unknown sample solution was used to determine its concentration.

Tag removal and reverse IMAC purification

Considering that the GFP tag represents a flexible domain within the construct, an enzymatic cleavage reaction was carried out to remove the C-terminal GFP–His tag from the Wzx construct, purified by size-exclusion chromatography. This step aimed to increase the likelihood of obtaining well-ordered crystals for structural studies. The Wzx_ThrbClSt_GFP_10xHis cleavage reaction was set up by mixing the protein with thrombin protease at a typical enzyme-to-substrate ratio of 1 U per mg of protein in thrombin cleavage buffer (TBS 1x pH8.0, 1x CMC of LMNG or 10xCMC LDAO, supplied by 2.5 mM $CaCl_2$). The mixture was incubated at 18°C overnight. The reaction was then stopped removing thrombin through size exclusion chromatography, where also the cleaved tag and any uncut protein were separated from the target protein. Cleavage efficiency and sample purity were assessed by SDS–PAGE analysis. The reaction for Wzx_TEVClSt_GFP_10xHis was performed using TEV protease produced in our laboratory, which carries a C-terminal His tag. This design enables the efficient removal of the protease itself, as well as any non-cleaved Wzx and the released GFP fragment, by Reverse Immobilized Metal Affinity Chromatography (Reverse IMAC). The reverse IMAC served as a buffer-exchange procedure, effectively removing the TEV cleavage buffer

(TBS 1x pH8.0, 1xCMC of LMNG or 10xCMC LDAO, supplied by 0.5 mM EDTA and 1 mM DTT). Prior to cleavage, the concentration of Wzx_TEV_GFP was determined by BCA assay to accurately calculate the enzyme-to-substrate ratio. The reaction was set up at a 1:10 molar ratio (TEV:Wzx) and incubated overnight at 4 °C to ensure complete digestion. Following incubation with Co^{2+}-functionalized TALON resin, the cleaved Wzx protein (without tag) was recovered in the flow-through, while the TEV protease, cleaved GFP–His fragment, and the uncleaved Wzx were retained on the resin. To monitor the efficiency of the enzymatic cleavage and assess sample purity, SDS–PAGE analysis was performed under semi-denaturing conditions (i.e., without boiling the samples) on both the flow-through and the bound fractions.

Circular Dichroism

Circular Dichroism (CD) spectroscopy is a technique widely used to study the secondary structure and folding of proteins. CD measures the differential absorption of right- and left-circularly polarized light by optically active molecules, resulting in elliptically polarized transmitted light. Since distinct secondary structures interact differently with circularly polarized light between 180 and 240 nm, characteristic spectra can be used to infer the secondary structural content of a protein (Figure 24). More specifically, α-helices show negative bands at 208 and 222 nm and a positive one near 193 nm, β-sheets display a negative band around 218 nm and a positive one near 195 nm, whereas disordered structures exhibit low ellipticity above 210 nm with a negative band around 195 nm, as shown in Figure 24.

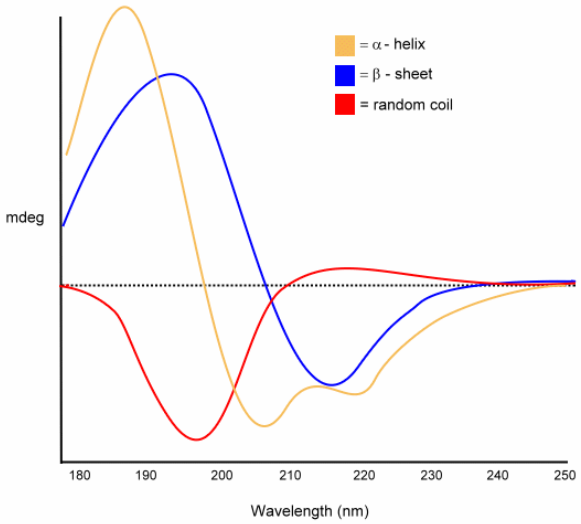


Figure 24 - Typical CD spectra for specific secondary structures [91]

In this study, the thermal stability of the Wzx protein was evaluated by estimating its melting temperature (Tm) from CD spectra at different temperatures, providing insight into Wzx folding stability. A Jasco J-815 spectropolarimeter equipped with a photomultiplier detector and a Peltier temperature controller was used to detect the ellipticity. A 1 mg/mL protein solution in 10xCMC LDAO and TBS buffer (pH 8) was placed in a 0.5 mm quartz cuvette (Hellma Analytics). CD spectra were recorded from 190 to 250

nm over a temperature range between 20°C and 80°C, acquiring one spectrum per degree. After removing the buffer contribution, the melting temperature was estimated using the second derivative of the ellipticity against the temperature at specified wavelengths (222nm) and averaging the results obtained for the different wavelengths. This method was implemented in Python 3.11 using numpy, pandas, and matplotlib libraries.

Dynamic Light Scattering

Dynamic Light Scattering (DLS) measurements were performed using a Malvern Zetasizer Nano Series instrument (Malvern Panalytical). DLS allows the determination of the size distribution and homogeneity of particles in solution by measuring fluctuations in the intensity of scattered light caused by the Brownian motion of the molecules. In particular, a He–Ne laser (λ = 633 nm) illuminates the sample, and the detector measures the backscattered light at a fixed angle of 173°. Analysis of the correlation function of the signal enables estimation of the hydrodynamic diameter and the polydispersity index (PDI) of the sample. Measurements were carried out in quartz cuvettes, filled to approximately one-third of their total volume with the sample solution. Each sample was measured in five independent replicates, with an acquisition time of 10 seconds per run, for a total duration of about 50 seconds per measurement. The physical parameters of the solvent were set to standard values for aqueous solutions: refractive index (RI) = 1.333 and viscosity = 0.8872 cP at 25 °C. Analyses were performed using three detergents (DM 10xCMC, DDM 10xCMC, and LDAO 10xCMC) to evaluate the average micelle size and to assess the correlation between particle diameter and salt concentration (varied between 100mM to 200mM).

Crystallization

Crystallization attempts of the Wzx protein

A fundamental requirement for protein crystallization is a protein sample at high purity and concentration close to the saturation value, which is especially challenging for membrane proteins, often expressed at low yield and with demanding purification protocols. Owing to their weaker packing interactions, membrane proteins crystals tend to be even more fragile than soluble protein and, hence, more difficult to handle. Furthermore, crystal formation, which is mainly driven by the entropic contribution resulting from the release of ordered water molecules from hydrophilic protein surfaces when proteins pack together, is more difficult for proteins with large hydrophobic transmembrane regions embedded in disordered detergent micelles. Detergents such as OG, DDM, DM, and LDAO are commonly used, as they promote the growth of crystals with improved internal order and diffraction quality [92]. In this study, crystallization trials were carried out using the MemStart, MemSys and MemGold (both 1 and 2) screens from Molecular Dimensions. 96 wells plates were used to set up the crystallization experiments, using 80 µL of reservoir solution in each well. The purified protein was

concentrated up to 3 mg/mL, and crystallization in sitting drops was set up using a Mosquito robot, mixing 0.2 µL of protein with 0.2 µL of precipitant solution, or alternatively by hand mixing up to 1 µL of protein in a 1:1 ratio with the precipitant solution. Plates were sealed and incubated at either 18°C or 4°C and monitored periodically. Suitable crystals were collected with nylon loops and cryoprotected with the reservoir solution supplemented with 20% glycerol, which prevented ice formation during flash-freezing. The frozen crystals were stored in liquid nitrogen for diffraction analysis. X-ray diffraction data were collected at the XRD2 beamline of the Elettra Synchrotron (Trieste, Italy).

Crystallization of the endoglycosidase protein

Crystallization trials for the EG-NΔ8 protein were carried out using the vapor diffusion method in a sitting-drop configuration. This technique relies on the gradual diffusion of water vapor from a small drop containing the protein solution into a larger reservoir well containing a higher concentration of precipitant. Typically, the protein solution and reservoir solution are mixed in a 1:1 ratio to initiate equilibration and promote crystal growth. Purified EG-NΔ8 was concentrated to approximately 3.2 mg/mL, and crystallization screening was performed at 18 °C using commercial kits (Index and PEG-Ion, Hampton Research). Crystallization plates were set up at the Elettra facility in Trieste using a Mosquito nanoliter dispenser, combining 200 nL of protein solution with an equal volume of reservoir solution. After approximately 2–3 months, a crystal suitable for X-ray diffraction was obtained in well D5 of the Index screen, containing 0.1 M sodium acetate trihydrate (pH 4.5) and 20% (w/v) polyethylene glycol (PEG) 3350. Cryoprotection was achieved by briefly soaking the crystal in the reservoir solution supplemented with 20% (v/v) glycerol. The crystal was then flash-frozen in liquid nitrogen and stored until data collection.

Results and Discussion (Wzx flippase)

AlphaFold modelling of the Wzx protein of *P. aeruginosa*

AlphaFold2 [80] was used to prepare a model of the Wzx construct expressed and purified in this doctoral work. While a model of the Wzx protein from *P. aeruginosa* is available in the literature [32], the whole construct of the chimeric protein was used as input to evaluate the flexibility of the construct and identify possible issues for protein stability and crystallization. The construct analyzed by the Deep Learning tool includes, from N-terminus to C-terminus: the Wzx protein of *P. aeruginosa*, the Thrombin cleavage site, the GFP protein sequence, and a 10xHis tag. AlphaFold2 predicted five possible models of the Wzx-GFP fusion protein (Figure 25). Two structurally different domains are visible in the predicted models: a domain consisting of a compact bundle of α-helices can be clearly identified as the transmembrane portion of the Wzx protein, while a region characterized by a β-barrel fold shows the typical structure of the green fluorescent protein (GFP). The 10x His tag is predicted to form a separate α-helix. The model confirms the well-characterized structure of GFP, and supports the insights obtained from both theoretical and experimental analyses of Wzx [32].

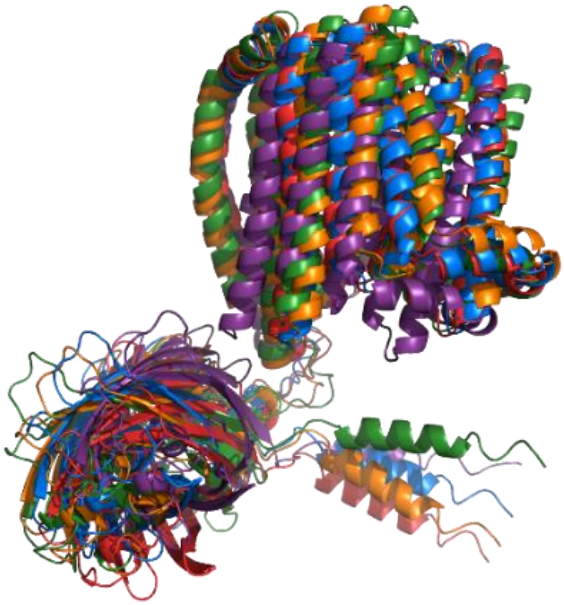


Figure 25 - Five superimposed models estimated by AlphaFold2. Blue one, which had the best rank, was selected for next analysis.

AlphaFold2 assigns to each residue of the predicted model a pLDDT (predicted Local Distance Difference Test) score, a value that estimates the reliability of the structural prediction of the software (Figure 26). Scores above 90 indicate very high confidence, typically corresponding to well-folded or experimentally validated regions. Values between 70 and 90 denote good reliability, while scores below 50 correspond to unstructured or flexible segments, that are generally correlated to higher uncertainty. All the models predicted for the Wzx-GFP construct display a similar level of confidence along the protein sequence, with the minimum values of the plDDT index corresponding

to the N-terminal segment of the Wzx protein, the linker between the two domains and the C-terminal histidine tag.

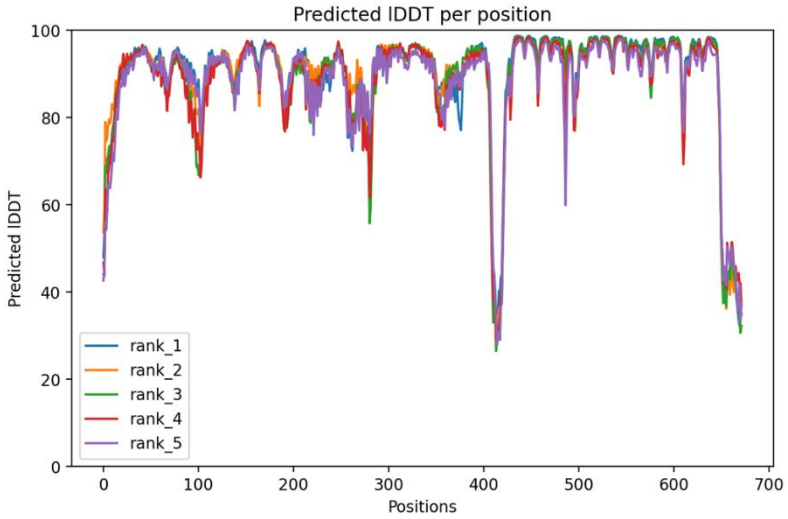


Figure 26 – Predicted local Difference Distance Test (plDDT) calculated for each position on the protein construct. Values above 90 represent regions of the protein model estimated with high confidence. Minima indicate low confidence in the prediction. Colors match with models in Figure 25.

Figure 27 shows a graphical representation of the pLDDT scores for the highest ranking model predicted for the Wzx-GFP construct. The color scale ranges from blue (high confidence) to red (low confidence), which allows a visual assessment of model quality in the different protein domains and segments. For most of the model in Figure 27, the blue color confirms a high confidence in the prediction, in particular as regards the transmembrane helices of Wzx and the β-barrel of GFP. On the other hand, red and orange regions, localized mainly in the linker connecting Wzx and GFP (which is also the protease cleavage site) and in the polyhistidine tag, correspond to low-confidence areas. These are indeed expected to be highly flexible or intrinsically disordered regions.

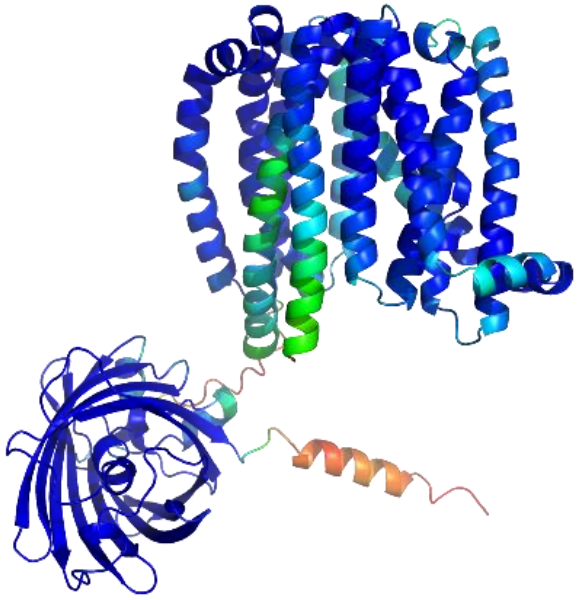


Figure 27 – Highest ranking model obtained from AlphaFold2 for the Wzx-GFP construct, colored according to the pLDDT scores, from blue (high confidence) to red (low confidence).

The structural superposition of all GFP and all Wzx domain, respectively, shows that the highest-scoring models of AlphaFold predict the same arrangement for the single domains, with only small variations. In particular, the model of the Wzx domain obtained here is consistent with the structural predictions reported by Islam *et al.* [32]. The high uncertainty estimated by AlphaFold for the linker between the two domains, on the other hand, suggests that the relative orientation between them remains poorly defined. The positioning of GFP with respect to Wzx should be interpreted with caution. The low plDDT scores for residues of the linker and the comparison of the 5 models produced by AlphaFold (Figure 25) suggests a high degree of flexibility between the domains, with a resulting conformational heterogeneity in the protein sample. However, the prediction by AlphaFold2 for the Wzx domain aligns with the biological information and expected features of the construct, offering a valuable starting structural model for subsequent computational analyses. The topological analysis of the transmembrane (TM) domain reveals the presence of twelve α-helices arranged in a globular bundle. In Figure 28, the helices are colored according to sequence order, from blue (residue 1) to red (residue 411), providing a visual gradient that allows to easily track the protein topology. For clarity, in the right picture of the Figure, the helices are represented as cylinders.

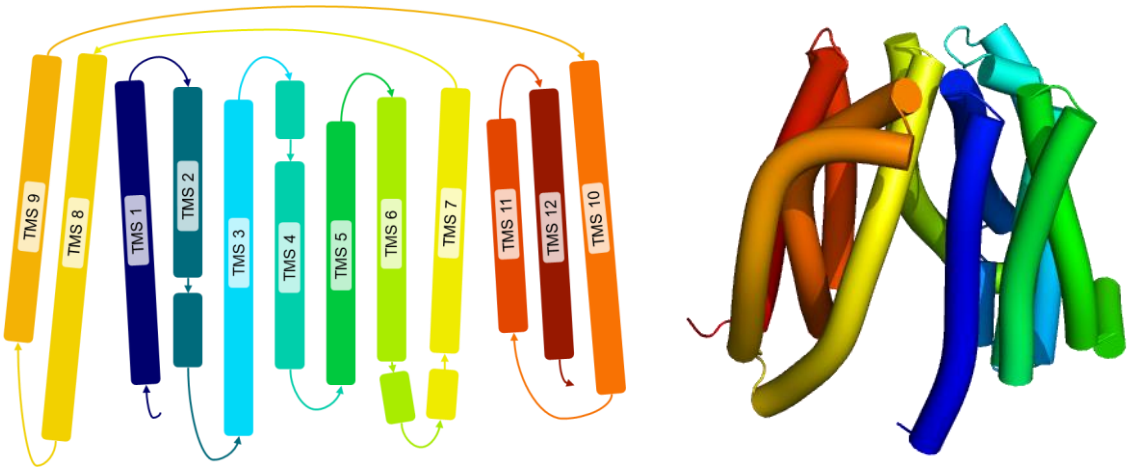


Figure 28 – (a) Topological representation of the Wzx protein. Residues are colored from blue (N-terminus) to red (C-terminus) allowing the identification of the 12 Trans Membrane Segments. (b) Representation of the arrangement of the 12 α-helical segments, shown as cylinders, in the predicted structure. Coloring corresponds to the topological map in (a).

The organization of the helices suggests a bilobate topology, consisting of two distinct helical bundles, the first from TMS 1 to TMS 7 and the second from TMS 8 to TMS 12, connected by a periplasmic loop between TMS 7 and TMS 8 (Figure 28, right panel). Based on previous experimental and computational studies, both the N- and C-termini of the Wzx protein are predicted to be on the cytoplasmic side of the membrane. According to this hypothesis, the model in Figure 28 is oriented with the cytoplasmic face in the lower part of the figure and the periplasmic face at the top. Thus, the conformation of the model represents a cytoplasmic-facing state of the transporter, i.e., inward-facing. In this model, the central cavity of the flippase is open toward the cytoplasmic side and closed toward the periplasm, suggesting that this conformation may correspond to the

substrate-loading state of the transport cycle, based on the model proposed by Islam *et al.* [32]. This interpretation is further supported by the observed divergence of the two helical halves, which creates two lateral clefts on opposite sides of the transmembrane bundle, one located between TMS1 and TMS8, and the other between TMS7 and TMS11. These clefts are likely embedded within the membrane bilayer and could represent the entrance through which the substrate accesses the internal cavity of the flippase during translocation, in agreement with the mechanism proposed by Islam *et al.* [32].

Building a membrane model

To investigate the conformational dynamics of the Wzx protein, the AlphaFold-generated model, without the GFP tag, was embedded into a membrane bilayer model, as starting point for Molecular Dynamics (MD) simulations. Before building the system, the orientation of the protein with respect to the membrane plane was verified using the PPM 2.0 server [93], which employs data from the OPM (Orientations of Proteins in Membranes) database [93]. Consistent with the topology predicted by the AlphaFold2 model, both the N- and C-termini were positioned on the cytoplasmic side of the membrane, corresponding to an inward-open (cytoplasmic-facing) conformation. The protein was then centered in the simulation box, ensuring proper embedding of all transmembrane segments within the hydrophobic core of the bilayer (Figure 29). System preparation was performed using the CHARMM-GUI Membrane Builder online interface [94], which provides an automated pipeline for embedding proteins into lipid bilayers with a precise membrane composition.

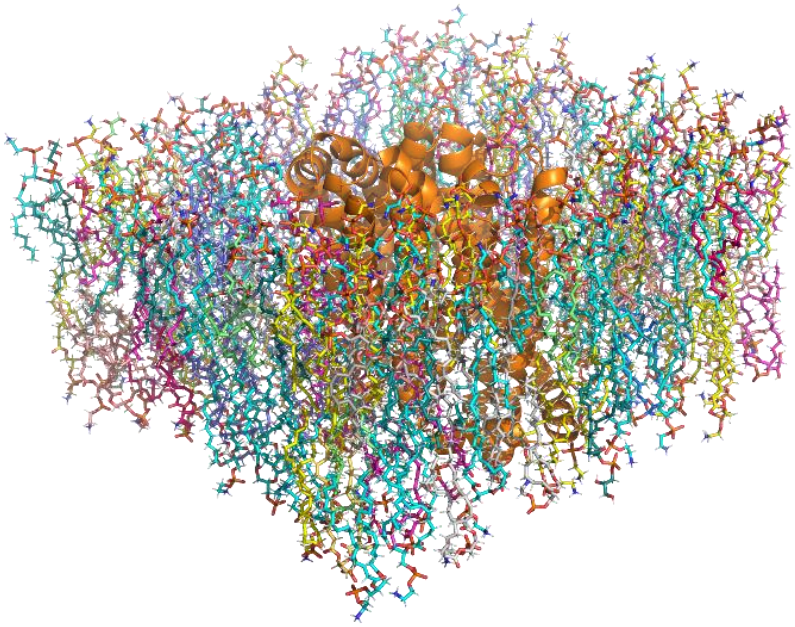


Figure 29 - Model of Wzx embedded in membrane bilayer in the simulation box. The protein is represented as orange cartoon and the lipids as colored sticks. Water molecules and ions, added to the simulation box during the preparation, are omitted for clarity.

The composition of the lipid environment of Wzx is expected to influence the

conformational behavior and dynamics of the flippase. To reproduce a biologically relevant environment, the membrane composition was designed to mimic the inner membrane of Gram-negative bacteria, with the lipid ratios determined by Ishinaga *et al.*, reported in Table 1. The bilayer was designed as a heterogeneous mixture of phosphatidylethanolamine (POPE, 70-74% of the weight of the membrane) and phosphatidylglycerol (POPG, 23-25%), in the experimentally determined proportions, with the addition of approximately 3-5% of cardiolipin. The lipid distribution was symmetrically arranged across the bilayer. Lipids overlapping with the transmembrane regions were removed, and the remaining lipids were repacked to achieve a homogeneous distribution. Following membrane embedding, the system was solvated with explicit water molecules, and 150 mM NaCl was added to both neutralize the system and reproduce physiological ionic strength, thus approximating the electrostatic environment of the bacterial cytoplasm. The resulting system provides a realistic representation of Wzx embedded in a native-like lipid bilayer, suitable for production of MD simulations aimed at exploring the conformational variability of the protein and characterizing the accessible conformational space.

Table 1 - Phospholipid composition of the membranes of E. coli K-12, grown at 17°C and 37°C (from Ishinaga et al. 1979). The first column reports the acyl chains in the first and second positions of the glycerol-3-phosphate scaffold. PE: phosphatidylethanolamine; PG: phosphatidylglycerol; 14:0 myristoyl; 14:1 myristoleoyl; 16:0 palmitoyl; 16:1 palmitoleyl; 17:0 cis-9,10-methylenehexadecanoyl; 18:1 cis-vaccenoyl; 19:0 cis-11,12-methyleneoctadecanoyl.

Molecular Species	**Cytoplasmic membrane (%)**				Outer membrane (%)			
	PE		**PG**		PE		PG	
C1 - C2	17°C	**37°C**	17°C	**37°C**	17°C	37°C	17°C	37°C
16:1-16:1	8	3	5	3	4	1	5	2
18:1-16:1	15	11	16	15	7	4	10	11
18:1-18:1	2	1	6	3	1	T	4	3
16:0-14:1	5	2	3	2	5	2	3	2
16:0-16:1	49	31	43	38	49	35	42	42
16:0-18:1	8	9	24	17	12	9	26	14
16:0-14:0	Traces	4	Traces	3	1	9	2	5
16:0-16:0	Traces	1	Traces	Traces	Traces	Traces	Traces	Traces
16:0-17:0	13	37	3	18	21	39	8	20
16:0-19:0	Traces	1	Traces	1	Traces	1	Traces	1

Building micelle models

In parallel to the study of Wzx in the bacterial membrane, the behavior of the protein within the detergent micelle was analyzed through computational tools. The first step of this work was to build a reliable detergent micelle model and, on this model, optimize the simulation parameters for the subsequent studies of Wzx embedded in detergent. A series of preliminary MD simulations was performed on empty detergent micelles to define the optimal setup parameters, including micelle dimension and equilibration protocol, before introducing the membrane protein into the system. Detergent micelles were obtained with the same tool (CHARMM-GUI) used for designing the membrane

system. Simulations were carried out for several detergents commonly used in membrane protein solubilization, including DM, DDM, and LDAO. Since the micelle size could not be estimated with high certainty, multiple micelles were constructed for each detergent type varying the number of detergent molecules. An expansion of the micelle dimension was observed during the run. Once the simulation parameters were optimized, this approach allowed the relaxation of the system to a stable micellar structure, thus enabling the estimation of the equilibrium dimension. Simulations were performed using the adapted Roe protocol, which prevents simulation instabilities or crashes associated with equilibration. As an example, Figure 30 shows the fluctuations of the radius of gyration (calculated by gromacs Gyrate tool) of DDM micelles during 100 ns of simulation, for three different systems containing 50 (blue line), 75 (orange line), and 100 (green line) detergent molecules, respectively. As expected, the radius of gyration does not increase linearly with the number of molecules, since the micelle tends to adopt a spherical geometry. For clarity, the data have been smoothed using a Gaussian filter (σ = 3), while the raw values are shown as grey lines. The average dimensions of each system are summarized in Figure 31, where black dots represent the average values, grey bars indicate the maximum and minimum radii, and black lines represent the standard deviation.

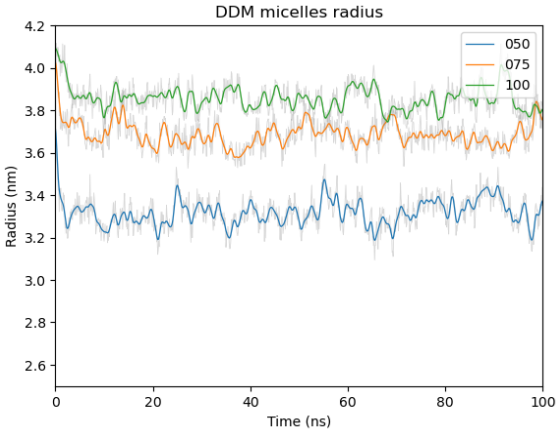


Figure 30 - Radius of gyration of DDM micelles in systems containing 50 (blue line), 75 (orange li ne) and 100 (green line) detergent molecules. Each simulation was run for 100 ns. For each system, the radius (gray lines) has been averaged with gaussian smoothing with σ=3 (colored lines).

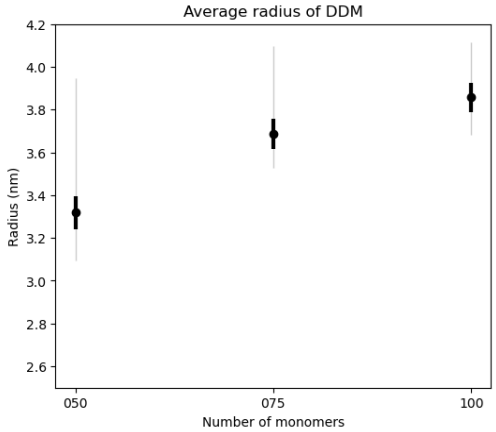


Figure 31 - Average radii of DDM micelles during 100 ns simulations (black dots), reported against the number of detergent molecules. Maximum and minimum radii are reported as grey bars, standard deviations as black bars.

To further analyze the stability of the system, the RMSD values of each detergent molecule over the course of a 200 ns MD simulation was considered. The RMSD quantifies the displacement of each molecule from its reference position, in this case the minimized system conformation, and therefore reflects the movement of monomers within the micelle. It is important to note that an increase in RMSD does not imply an expansion of the micelle volume, as molecules can move significantly within the micelle while remaining part of it. Figure 32 reports, as an example, the RMSD plot obtained for the LDAO system containing 75 detergent molecules. The plot shows that most

molecules maintain low RMSD fluctuations, consistent with continuous rearrangement within a stable micellar structure. However, three outlier lines display an increase in RMSD values. These correspond to three LDAO molecules that leave the micelle, diffusing into the solvent. This observation is particularly relevant, as it demonstrates that the simulation and force-field parameters successfully reproduce the natural equilibrium between the aggregated (micellar) and monomeric (dispersed) states of the detergent. Interestingly, the overall trajectory also shows that the detached molecules can be reincorporated into the micelle during the simulation, confirming the reversible nature of this equilibrium. The fact that these outlier RMSD lines do not return to the average values, is due to periodic boundary condition artifacts introduced during trajectory post-processing. Nevertheless, the curves reach an almost stable plateau, indicating that the reincorporated molecules adopt a semi stable configuration in the micelle.

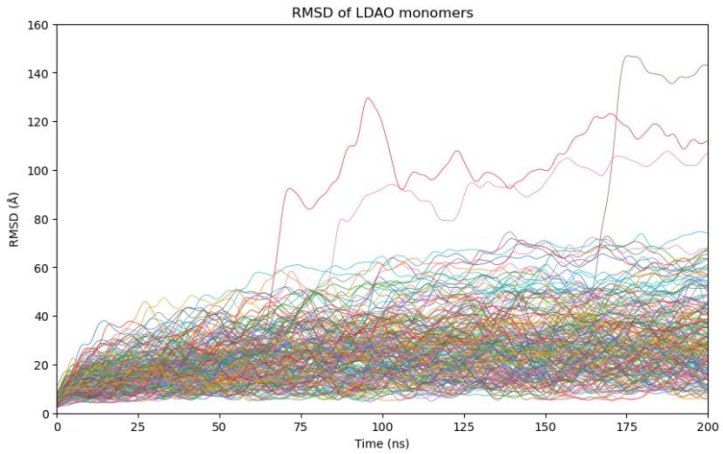


Figure 32 – RMSD plot of single detergent molecules in a LDAO micelle in the system prepared with 75 molecules.

Interestingly, the LDAO system containing 100 detergent molecules exhibited a markedly different behavior from the LDAO system with 75 molecules. During the simulation, the micelle rapidly converged to an elongated shape (Figure 33), which is consistent with observations reported in the literature as high concentration of LDAO notoriously leads to the formation of elongated micelle. In fact, the shape of a micelle depends on the surfactant packing parameters, and LDAO is known to form cylindrical or elongated micelles under specific conditions.

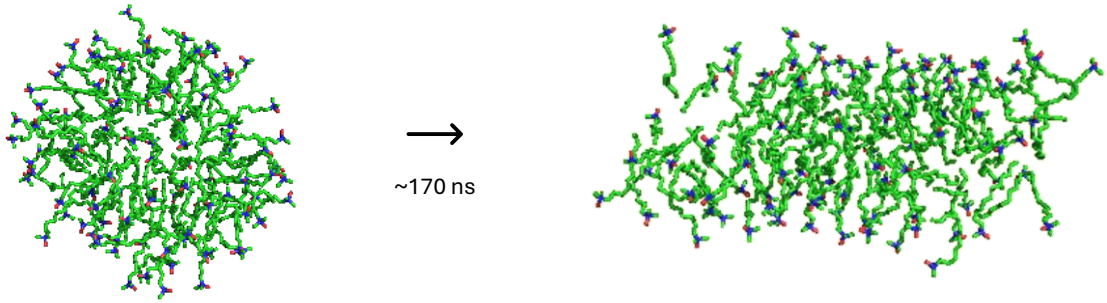


Figure 33 – Evolution of the LDAO micelle containing 100 molecules during the simulation, from the initial spherical aggregate to an elongated shape.

In all the simulation conducted with detergents, the force field and simulation parameters employed were capable of reproducing realistic micellar behavior. The systems evolved from their initial configurations to aggregates characterized by stable diameters specific for each detergent type and number of molecules, indicating moreover that the micelle packing was not an artifact due to the initial state of the model but the result of self-assembly dynamics. A confirmation of the simulations were able to reproduce the experimental behavior of detergents comes from the dynamic equilibrium between monodisperse molecules and the aggregated micellar state. During the 100 molecules LDAO simulation, monomers were able to leave and reinsert into the micelle during the trajectory, which reflects the natural exchange equilibrium of micellar systems in solution. Notably, the LDAO micelle also underwent a spontaneous structural remodeling, adopting an elongated shape without any applied bias. This behavior is consistent with the intrinsic properties of LDAO and confirms that the simulation setup can capture the self-organizing nature of detergent micelles under physiological conditions.

DLS data validation

A dynamic light scattering (DLS) analysis was performed to experimentaly evaluate the dimensions of micelles and corroborate the computational data. Five LDAO solutions were prepared with increasing NaCl concentrations, from 100 mM to 200 mM, while maintaining the detergent concentration constant at 10x CMC (corresponding to 10 mM or 0.23% in weight). The range of salt concentration was chosen as its intermediate value of 150 mM corresponds to the concentrations used in experiments and, roughly, with the physiological conditions in cells. The DLS experiments proved that the average micelle size increases linearly with salt concentration within the range tested (Figure 34). LDAO micelles formed in NaCl 150 mM had an average diameter of approximately 5.7 nm (highlighted in green in Figure 34). In the same conditions, i.e., LDAO 10 mM, NaCl 150 mM, a measurement was performed on the Wzx–LDAO complex. In this case, the average particle diameter increased to approximately 7.0 nm (highlighted in blue in Figure 34), corresponding to an increase of roughly 13 Å compared to the empty micelles. This observation indicates that the incorporation of the Wzx protein within the detergent micelles induces a measurable increase in hydrodynamic radius. In addition, the size distribution of the micelles containing Wzx is remarkably narrow, suggesting a highly uniform population. However, this result should be taken with caution, as the analyzed protein sample had undergone multiple purification steps, including size-exclusion chromatography, and SDS-PAGE analysis demonstrates that co-solubilized proteins are present even in purified Wzx preparations. This suggests that the detergent micelle averaging around 7.0 nm in diameter incorporated both the target protein and the contaminant.

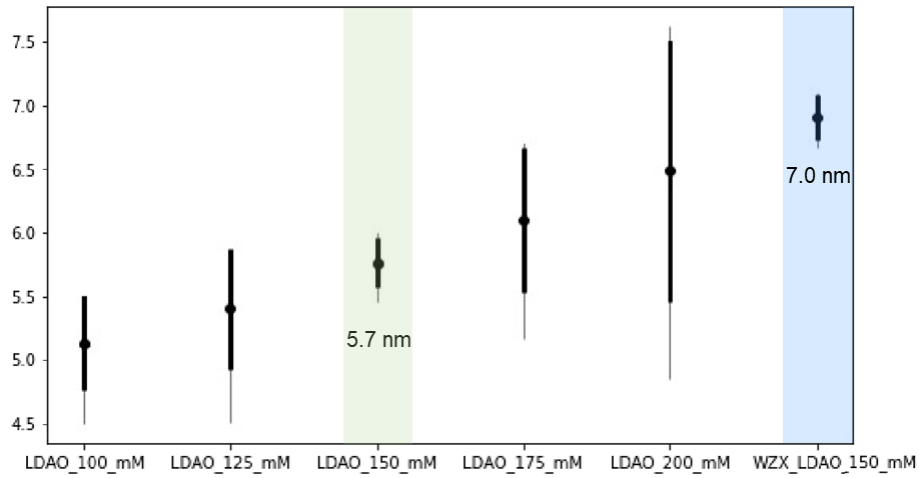


Figure 34 – LDAO micelle dimensions measured using DLS at different salt concentrations. Buffer solution containing 10x CMC LDAO, NaCl 150 mM are highlighted in green as they represent the experimental conditions used in Wzx experiments, reported on the right in blue. Black dots represent the average values, black bars represent the standard deviations.

Another notable feature of the DLS profile for the Wzx-LDAO sample is the absence of a population corresponding to empty micelles (5.7 nm). A possible explanation is that the previous SEC purification step removed the empty micelles, effectively reducing the detergent concentration in solution. Since all detergent molecules present in the sample are engaged in the interaction with the Wzx protein, the equilibrium with empty micelles was not restore during the experiment. Overall, these results provide an estimation of the average dimension of the detergent micelles and the expected dimensions of the micelle containing Wzx to be used for simulations. Data obtained from DLS, indicating that the average diameter of LDAO micelles in NaCl 150 mM is approximately 5.76 nm, were compared with the computational model, to assess the reliability of the latter. Five molecular dynamics simulations were performed for LDAO micelles composed of 50, 65, 75, 85, and 100 detergent molecules. The largest system (100 molecules) was excluded from the analysis due to its anomalous behavior that led to an elongate-shaped aggregate, as previously discussed. In the remaining four simulations, a linear relationship was observed between the average micelle radius and the number of detergent molecules used in the simulation (Figure 35), indicating that the size of the micelle obtained in the computational simulation depends on the number of monomers included in the initial state. The estimation of the average number of monomers in a LDAO micelle in NaCl 150 mM was obtained by interpolating the experimental micelle size obtained from DLS measurements (dark green point in Figure 35, right panel) with the computational data, showing that a micelle of the size measured by DLS contains approximately 63 ± 3 detergent molecules.

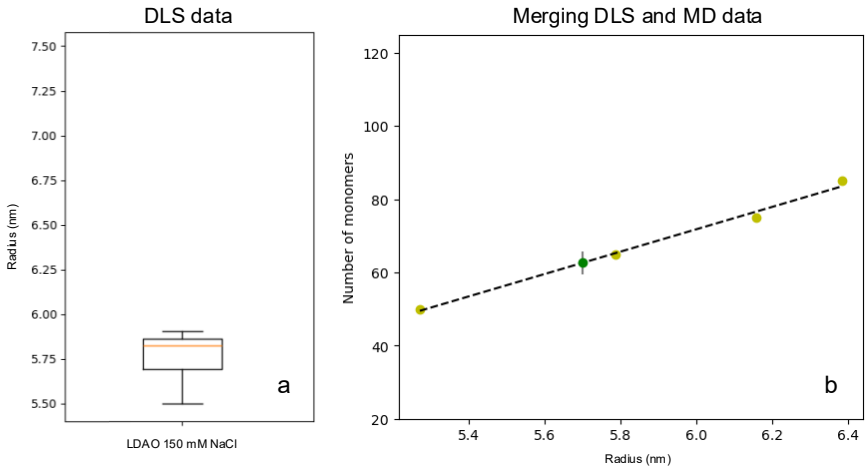


Figure 35 – Comparison between the micelle size measured with DLS at a NaCl concentration of 150 mM, left panel, and data from the MD simulations, right panel. A linear correlation was observed between the number of molecules included in the MD simulations and the average radius of the micelle (light green dots). The comparison with computational data allows to estimate a number of 63 detergent molecules per micelle in the experimental setup (dark green dot in the right panel).

Molecular dynamics simulations

MD simulation of the Wzx protein model in the lipid bilayer

To investigate the dynamic behavior of Wzx in different environments, MD simulations were performed with the protein embedded in both a native-like lipid bilayer and a detergent micelle. Each simulation was run for 200 nanoseconds, a timeframe that provides a sufficient sampling to evaluate the local flexibility and stability of the protein structure. However, the timespan of the simulation remains insufficient to capture large-scale conformational changes, such as the hypothesized conformational change of the protein from a cytoplasmic- to a periplasmic-facing state, which are typically associated with a longer timescale. It should also be noted that the simulation does not include the natural substrate of the flippase. Furthermore, the simulations were not biased by external forces to drive conformational switching. As a result, the trajectories primarily describe the protein fluctuations and the overall stability of the inward-open conformation modeled by AlphaFold, providing valuable insights into local rearrangements rather than a complete picture of the transport cycle. The MD simulation of Wzx embedded in the membrane was performed using the previously described membrane model (Figure 36).

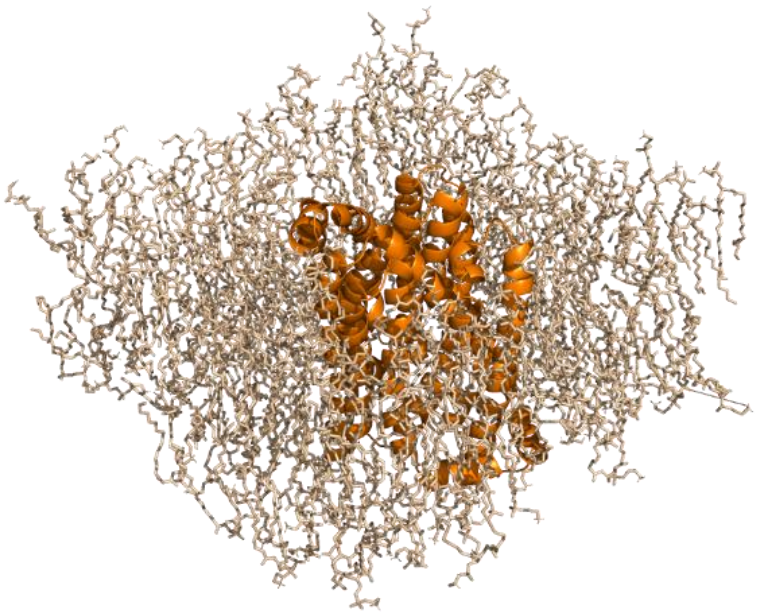


Figure 36 – Representative snapshot of the simulation box of the Wzx protein (orange cartoon) embedded in the membrane bilayer (brown sticks).

The RMSD of each residue compared to the initial model was calculated during the first 170 ns of MD simulation of Wzx embedded in the membrane bilayer (Figure 37). This parameter represents the atomic displacement of each residue relative to its initial position, allowing a detailed overview of local flexibility along the protein sequence and simulation time. The analysis reveals that the N-terminal segment of Wzx exhibits high RMSD values, indicating a high flexibility. This behavior is not unusual for terminal segments of proteins, suggesting the presence of segments that are not tightly integrated within the protein core. In contrast, it is notable that the C-terminal region of the Wzx

sequence exhibits a very low RMSD. This uncommon behavior can be explained considering that this region is less exposed to the solvent and is more deeply embedded within the lipid leaflet, which reduces its mobility. Additionally, lower peaks are observed along the sequence, presumably corresponding to loop regions connecting the transmembrane helices (TMS), which are less constraint and usually more flexible. Aside from these flexible regions, the overall RMSD remains below 2 Å for most residues throughout the whole simulation, demonstrating that the protein remains structurally stable preserving its overall fold within the membrane bilayer.

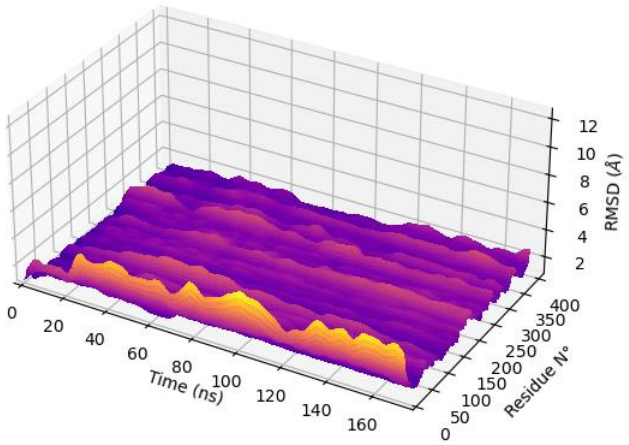


Figure 37 - Visual representation of per-residue RMSD along time during the MD simulation of the Wzx protein model from AlphaFold in the membrane bilayer. The chart was made with two python3 libraries: RMSD was recursively calculated for each frame of the run using MDAnalysis and plotted with matplotlib. This representation provides an innovative and more informative way to visualize both RMSD and RMSF. By flattening the surface along the time axis, the overall RMSD of the system can be obtained, while flattening it along the residue axis generates the root mean square fluctuation (RMSF) profile.

Figure 38 shows two representative frames from the molecular dynamics simulation: the initial configuration (left panel) and the system after approximately 170 ns (right panel).

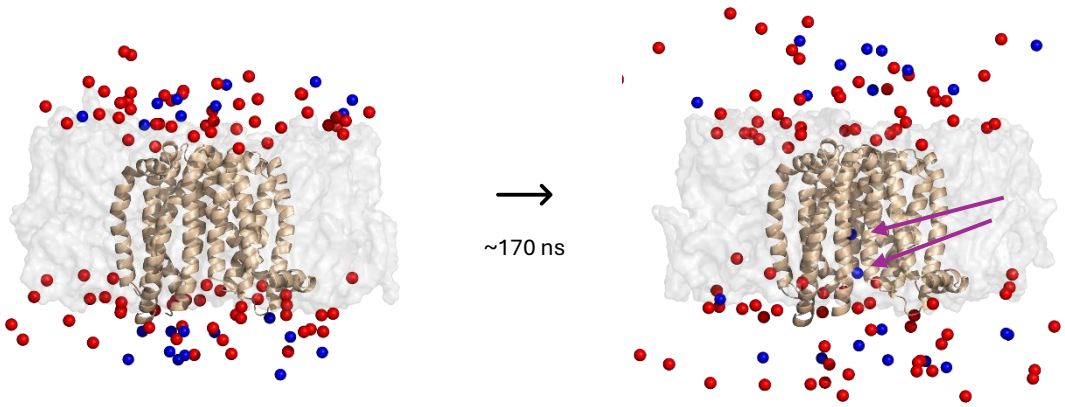


Figure 38 – Results of the MD simulation of the protein Wzx embedded in a lipid bilayer with composition similar to that of the inner membrane of E. coli. The protein is represented as tan cartoon, embedded in a semi-transparent lipid bilayer; ions are shown as red spheres (sodium) and blue spheres (chloride), while water molecules are omitted for clarity. In the right side panel, two arrows indicate the movement of two chloride ions that enter the cavity of the flippase and remain confined there for the remaining of the duration of the simulation.

Along the simulation, a notable movement of two chloride ions was observed. These negatively charged particles entered the cavity of the flippase and remained confined

there for the remaining duration of the simulation. Their stable localization within the cavity suggests that the positively charged lumen of the Wzx protein provides a favorable electrostatic environment for negatively charged species. This finding supports the hypothesis that electrostatic interactions drive the binding of the anionic substrates, i.e., the lipid-linked O-antigen molecules, to the transmembrane cavity of Wzx, in agreement with the proposed mechanism of the Wzx transport [32]. Furthermore, the entrance of chloride ions from the cytoplasmic side is consistent with the hypothesis of an inward-open conformation of the model, in which water and ions can access the cavity from the cytoplasmic side but not from the periplasmic side. Moreover, the absence of significant conformational drift during the simulation indicates that the cytoplasmic-open conformation of Wzx remains structurally stable even after the electrostatic interaction of the ions in the cavity of the transporter. The apparent displacement of ions from the protein cytoplasmic and periplasmic surfaces, observed in the right side panel of Figure 38, is an artifact of trajectory post-processing. In this final step of the computational analysis, the protein was centered in the simulation box to maintain a reference position, causing the shift of the surrounding solvent molecules outside the box due to periodic boundary conditions. This apparent explosion effect does not reflect a physical event but it should be instead considered as the reconstructed trajectory of atoms moving across box boundaries. The positions of the phospholipids forming the membrane bilayer was also analyzed. During the simulation, two lipid molecules from the cytoplasmic (inner) leaflet penetrated the cavity interface of the flippase through the two diametrically opposed entry sites (Figure 39), i.e., the opening formed between helices TMS8 (yellow in the left panel of the figure) and TMS1 (blue), and the cleft between helices TMS7 (light green in the right panel of the figure) and TMS11 (orange). The snapshots shown in Figure 39 were taken after ~150 ns of simulation. The first phospholipid molecule (purple in the left panel of the figure) penetrates into the TMS1-TMS8 interface, inserting one of its acyl chains into the opening while its polar headgroup remains close the cytoplasmic surface, a configuration consistent with the previously proposed initial docking mechanism of the substrate. This observation is particularly interesting when compared with the model proposed by Islam *et al*. [32], which suggests that substrate entry occurs through a portal formed by helices TMS1 and TMS9, rather than TMS1 and TMS8 as in the present study. Interestingly, the second lipid molecule (gray in the right panel of Figure 39) establishes a similar interaction with helices TMS7 and TMS11, suggesting the presence of an additional lateral entrance on the opposite side of the transporter.

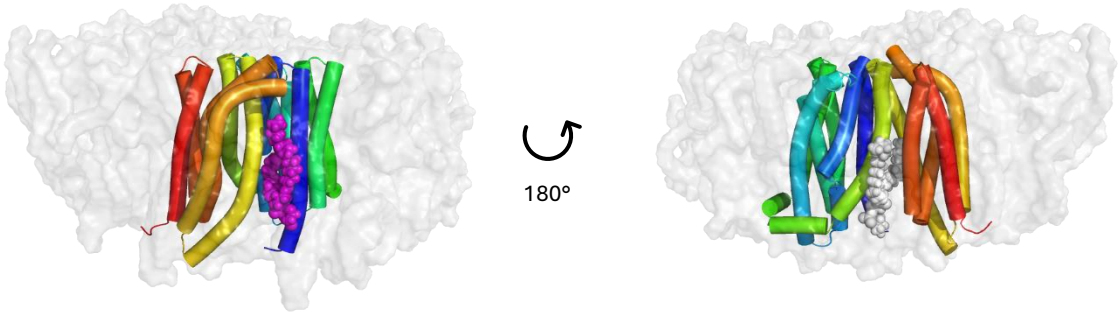


Figure 39 – Views of a snapshot obtained after ~150 ns MD simulation of the behavior of the protein Wzx embedded in a model of the membrane. The second view was taken by rotating the model of 180° along the vertical axis of the first. Color coding of TMS helices of Wzx as reported in Figure 28. In particular, helices TMS8, TMS1, TMS7, and TMS11, forming the two putative clefts, are shown as yellow, blue, light green, and orange cylinders, respectively. The lipid molecules that penetrate the clefts from the inner membrane are shown as purple and grey spheres.

This bilateral accessibility of the transmembrane bundle may indicate a flexible gating mechanism, allowing multiple entry routes for the lipid-linked substrate. However, it cannot be excluded that the reduced lateral membrane pressure in the simulation environment leads to a divergence of the two helical bundles and thus the transient opening of a secondary cleft. In any case, the spontaneous interaction of phospholipids with the cavity interfaces supports the hypothesis that the cytoplasmic-facing configuration of Wzx is naturally predisposed to recruit lipid components, consistent with its proposed physiological role in binding the lipid-linked O-antigen substrate from the inner leaflet. The finding that the main entrance portal differs from the one proposed in earlier studies does not imply a fundamentally different transport mechanism. Moreover, it is important to note that the simulation system lacks the natural substrate, and therefore the observed protein-lipid interactions might not reflect the exact docking mechanism of the undecaprenyl pyrophosphate ligand.

MD simulation of the Wzx protein model in the detergent micelle

The molecular dynamics simulation of Wzx embedded in an LDAO micelle was performed using the same setup as that employed for the “empty” micelle simulations. The simulation box was defined to be at least 1 nm larger than the molecule in all dimensions, with periodic boundary conditions applied. A fully explicit three-point water model was used, and the system was neutralized with ions to reach a physiological salt concentration of 150 mM NaCl. Energy minimization was carried out using the ROE protocol, followed by a 200 ns production run under standard NPT conditions. (Figure 40).

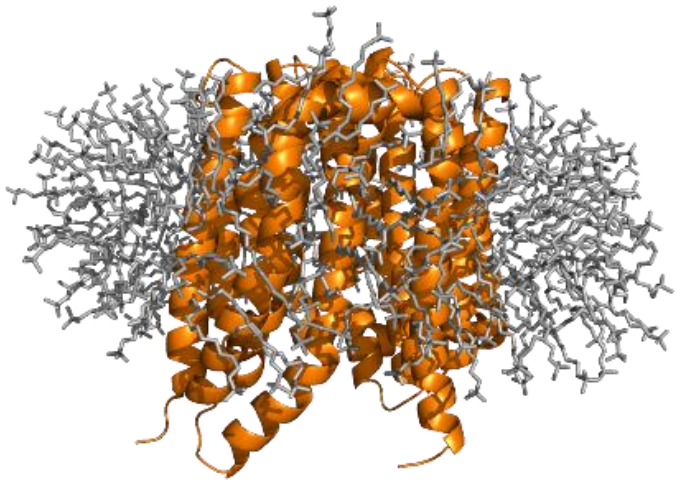


Figure 40 - Representative snapshot of the simulation box of the Wzx protein (orange cartoon) embedded in a LDAO detergent micelle (gray sticks).

Analysis of the per-residue RMSD surface plot for the Wzx protein model embedded in the detergent micelle (Figure 41) reveals an overall pattern quite different compared to the corresponding plot obtained for the protein in the lipid bilayer (Figure 37). In particular, several regions of the protein display considerably higher RMSD values, indicating an increase in molecular mobility in detergent conditions. Notably, the C-terminal region, which was relatively rigid when embedded in the membrane, exhibits enhanced flexibility in the micellar system. This increased mobility might result from the less constrained nature of the detergent micelle compared to the more organized lipid bilayer.

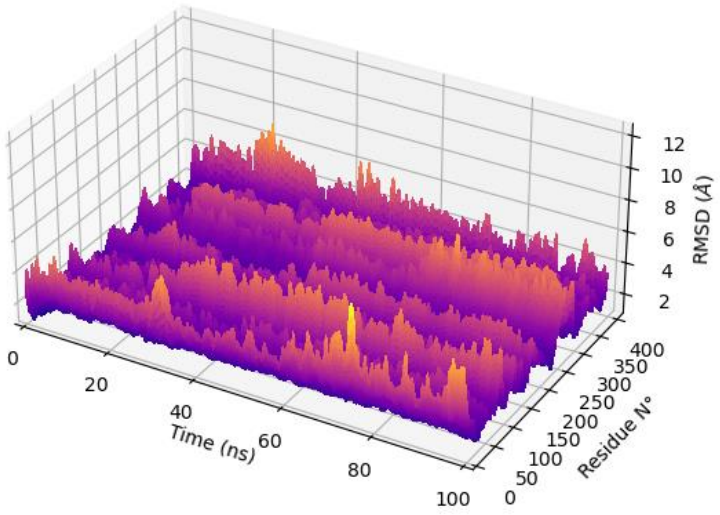


Figure 41 - Visual representation of per-residue RMSD along time during the MD simulation of the Wzx protein model from AlphaFold in the LDAO micelle. The chart was made with two python3 libraries: RMSD was recursively calculated for each frame of the run using MDAnalysis and plotted with matplotlib.

The plots represented in Figure 37 and Figure 41 allows for a clear distinction between flexibility and displacement. Flexibility corresponds to fluctuations of each residue of a specific region around an average position, showing a RMSD plot with a high variability but that alternating high and low values for each residue. In contrast, displacement

indicates a continuous motion of a protein region to a new position, leading to a consistently high RMSD that remains stable over time. Thus, while both behaviors can be associated with comparable RMSD values in the two dimensional plots of RMSD per residue, their temporal patterns are different: flexible regions show an oscillating behavior, whereas when coordinated conformational changes are detected, the RMSD value increases to reach a plateau. In the MD simulation of the Wzx-LDAO micelle system (Figure 40), the large number of residues displaying high RMSD values suggests higher flexibility of the model and, possibly, the structural instability of the protein within the micellar environment. In order to further analyze the behavior of the Wzx protein in detergent, a structural model of the whole Wzx–GFP construct used for protein expression and purification in the laboratory was embedded in a detergent micelle (Figure 42). Specifically, this model was generated to evaluate the accessibility of the thrombin cleavage site, which, considering the poor results of the enzymatic cleavage experiments, was hypothesized to be partially shielded by detergent molecules. A large micelle was constructed around the protein, intentionally exceeding the 7.0 nm diameter estimated by Dynamic Light Scattering experiments. Although the resulting system was too large to be used in MD simulations, the static structural model provides useful preliminary insights. In fact, it suggests that the protease cleavage site is exposed to the solvent and appears relatively accessible. The micelle size does not appear to hinder access to the cleavage site. However, in the absence of dynamic movement, it is not possible to exclude the formation of a steric hindrance in the cleavage region, thus preventing protease access.

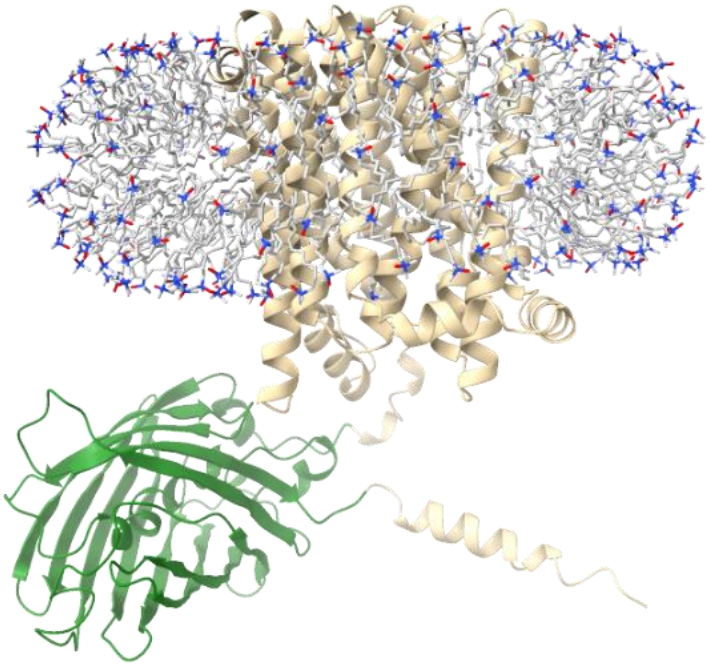


Figure 42 - Model of the construct used for laboratory experiments, which includes the model predicted by AlphaFold2 for the whole Wzx construct, i.e., the Wzx sequence, the thrombin cleavage site, GFP, and the 10x His tag, and a micelle of LDAO.

Design of experiments for the Overexpression of the O-antigen flippase Wzx from *P. aeruginosa*

Overexpression of recombinant proteins is usually a multivariable problem that requires intensive testing, particularly for proteins expressed at low yield such as membrane proteins. To speed up the testing step and maximizing at the same time the number of parameters taken into account, in this thesis, a Design of Experiments (DoE) approach was applied for the optimization of expression conditions. DoE is a statistical method used to conduct and analyze experiments with a large number of variables, to identify those that affect more significantly the experimental outcome, determine their relationship to other variables, and find the optimal conditions to obtain a desired result. In a protein expression experiment, the main variables are the bacterial strain, the level of bacterial growth at the moment of induction (expressed by the Optical Density, OD, measured as the absorbance at 600 nm), the concentration of the inducing agent, and the temperature at which protein expression is carried out. In addition, for some engineered bacterial strains, additives may be added, activating regulatory biochemical pathways. For the expression of the Wzx protein construct containing a thrombin cleavage site, the GFP sequence and a His tag, expression tests were conducted using a single bacterial strain, i.e., *E. coli* Lemo21(DE3), under different induction conditions. In the Lemo21 system, the addition of rhamnose to the culture is able to suppress basal expression. Using the DoE approach, the OD of induction, the concentration of the inducing agent (IPTG), the temperature of expression, and the concentration of the additive rhamnose were taken into account. The length of induction, which is another variable that could affect protein yield, was only changed according to the expression temperature: when cells were grown at 37°C, cell pellet was collected after 5 hours from the induction, as the medium was depleted by the fast-growing bacteria; instead, cultures run at 18°C were kept overnight to allow for the slower kinetics of expression at this temperature. Results were monitored by measuring the OD of the culture and its fluorescence at the end of induction. The OD represents a measure of the cell growth, while the fluorescence can be correlated with protein expression, considering that the construct contains a GFP domain. Figure 43 shows the results obtained in optimization experiments for the overexpression of the Wzx-Thrombin cleavage site-GFP-10xHis construct. The experiments were run at two different temperatures, 18°C and 37°C, with concentrations of IPTG varying between 0 and 0.8 M, added when the OD value of the culture was 0.4, 0.6 or 0.8. In addition, a set of experiments was conducted in presence of rhamnose and another set without the additive. OD and fluorescence of the final cultures were measured using a Synergy spectrophotometer in 96-well plates, with each condition analyzed in technical triplicate. A non-transformed control culture was included in the plate as a reference ("CTR" in Figure 43). To facilitate the analysis of the results, all measurements of OD and fluorescence were normalized to the values

measured for the control culture.

		5h - 37°				18h - 18°C			
		OD		fluo		OD		fluo	
induction OD	[IPTG] (mM)	no RHA	si RHA	no RHA	si RHA	no RHA	si RHA	no RHA	si RHA
0.4	NI	0.617167	0.643333	10420	10429.67	0.840833	0.828833	16264.83	16830.67
	0.4	0.543833	0.600167	14649.67	11317	0.785833	0.879833	33076.83	17880.67
	0.8	0.534167	0.602	14237.5	11496.83	0.775167	0.832833	27752.17	17800.67
0.6	NI	0.628833	0.644167	10713.67	10856.83	0.868	0.9245	16977.33	16784.67
	0.4	0.5325	0.646167	14306.5	11085.83	0.7565	0.9555	26787	16440.83
	0.8	0.511833	0.629833	13670.17	10937.83	0.730167	0.959167	26138	17868.5
0.8	NI	0.6585	0.6845	9976.833	10529.33	0.898167	0.9435	16669.5	16618.33
	0.4	0.573833	0.660167	12880.17	10755	0.829333	0.898	25197.5	18211.5
	0.8	0.541333	0.641833	13329.83	11174.5	0.870167	0.936167	22556.5	17023.17
	CTR		0.656833		12518.67		0.571		23499.83

Figure 43 – Results of experiments testing the expression yield of the Wzx protein construct containing the thrombin cleavage site, the GFP sequence and a C-terminal histidine tag. To monitor protein expression, the OD (reported with a green gradient coloring) and the fluorescence of the culture (reported with a red gradient coloring) were measured. The last line of the tables represents the non-transformed cell culture, measured as a reference.

		5h - 37°				18h - 18°C			
		OD		fluo		OD		fluo	
induction OD	[IPTG] (mM)	no RHA	si RHA	no RHA	si RHA	no RHA	si RHA	no RHA	si RHA
0.4	NI	0.936331	0.976251	0.799438	0.806938	1.470141	1.452576	0.670299	0.696418
	0.4	0.821122	0.90854	1.16264	0.888034	1.377635	1.545667	1.367885	0.755129
	0.8	0.808489	0.911066	1.127303	0.907669	1.357143	1.461944	1.137961	0.749301
0.6	NI	0.954522	0.974735	0.841011	0.853006	1.523419	1.621194	0.717218	0.710993
	0.4	0.807479	0.977261	1.133904	0.876489	1.326698	1.676815	1.092869	0.692342
	0.8	0.770086	0.95048	1.074803	0.863708	1.276347	1.679157	1.075722	0.750419
0.8	NI	0.996463	1.034361	0.77191	0.815618	1.570843	1.65281	0.697387	0.693386
	0.4	0.868621	0.998484	1.010393	0.840056	1.456089	1.576698	1.033835	0.758782
	0.8	0.818595	0.971703	1.04764	0.884551	1.522834	1.644614	0.927907	0.713664
	CTR		1		1		1		1

Figure 44 – Results of expression tests of the Wzx protein construct. To monitor protein expression, the OD (reported with a green gradient coloring) and the fluorescence of the culture (reported with a red gradient coloring) were measured and normalized to the cultures of non-transformed cells, used as standards.

As expected, non-induced cultures consistently showed higher optical density (OD) values and lower fluorescence signals, reflecting a faster cell growth when protein production was not induced. Cultures incubated overnight at 18°C developed a remarkable higher cell density compared to those grown for 5 hours at 37 °C, indicating that the number of bacteria growths despite the slower metabolic rate. Interestingly, the highest fluorescence intensities (highlighted in darker red shades in Figure 44) were generally associated with lower OD values (lighter green), suggesting that strong protein expression induces a stress on the culture that compromises bacterial growth. In all cases, addition of rhamnose resulted in increased OD and reduced fluorescence values, confirming its activity as expression suppressor in Lemo21. Despite lowering the protein yield, the addition of rhamnose could be advantageous by potentially improving protein folding and stability, as previously reported [89]. However, in this study, for all conditions that included rhamnose the fluorescence of induced cultures was comparable to that of

non-induced cells, meaning that under this conditions the protein would likely not be produced in usable quantity. For this reason, rhamnose was excluded from the final expression protocol. Considering the results of the optimization experiments, the overexpression conditions chosen for Wzx production were as follows: induction at OD 0.4 using IPTG 0.4 mM , without rhamnose, followed by overnight expression at 18 °C. This protocol is associated with significantly higher protein production, as demonstrated by the highest fluorescence value highlighted in yellow in Figure 44. A similar experimental approach was applied to optimize the expression conditions for the Wzx construct containing the TEV protease cleavage site. As expected considering that the two constructs differ by only a few residues and the modified region is predicted by AlphaFold to be in an unstructured region, the expression tests yielded results similar to those obtained for the previous construct. Consequently, in the following experiments, the construct containing the Wzx sequence, the TEV protease cleavage site, the GFP, and a 10xHis tag was expressed by inducing the culture at an OD_{600} of 0.4 with 0.4 mM IPTG, followed by ON incubation at 18 °C.

Protein solubilization (Wzx)

Selecting an appropriate detergent for membrane protein purification is particularly challenging, as the detergent should at the same time keep the protein in its native state and maintain it in a solubilized form, preventing aggregation through hydrophobic interactions. Whether a protein sample is in its native conformation is difficult to evaluate and the available assessment methods, e.g., CD, are often only partially informative. Reliable evaluation of the protein stabilization for a given detergent requires a sufficient quantity of purified protein, which depends on a well-optimized purification protocol. For each detergent the solubilization concentration must be individually refined, as the detergent efficacy in solubilizing the membranes varies with the CMC and with the detergent nature, e.g. ionic versus non-ionic. At the same time, the minimum detergent amount able to stabilize the protein in solution should be evaluated, as a lower detergent concentration usually correlates with a longer-term protein stability. In this study, the reported purification procedure was designed with the aim of preparing the protein for crystallization. In order to obtain crystals suitable for diffraction analysis, the protein stability in the long-term is of paramount importance, but at the same time the experimental structure is only meaningful if it represents the native state of the protein. In addition, the presence of contaminants may interfere with crystal growth, thus hampering the crystallization trials. Another critical issue is reproducibility in both the membrane protein expression and the purification protocol. Unless a strong expression protocol is available, a significant variability is usually observed in bacterial cultures and in protein yields. Measuring the fluorescence of GFP fused proteins may appear as a good method to track the protein sample throughout the expression and purification steps, but it only provides relative quantification of expression levels and is subject to interference by other factors, such as the protein aggregation or GFP denaturation due

to the detergent. Finally, for membrane proteins, the efficiency of detergent solubilization varies depending on the composition and heterogeneity of the bacterial membranes, which are rarely constant across different bacterial cultures. After bacteria collection and cell lysis, membranes containing the overexpressed Wzx proteins were solubilized using different detergents, namely LDAO, DM, DDM, and OG, each applied at different concentrations corresponding to 10x their CMC: final concentrations of 10mM for LDAO, 18mM for DM, 1.2 for DDM, and 230mM for OG were used. Solubilization was carried out at 4 °C for 1 hour under gentle agitation. Following incubation with the detergent, the membrane sample containing the Wzx construct with the thrombin cleavage site was completely dissolved in LDAO. On the contrary, the solubilization experiments of the same membranes in the other detergents did not appear completely clear, suggesting less effective solubilization. In all cases, insoluble material was removed by centrifugation. Prolonged incubations were tested as well, but did not lead to significant improvements in solubilization efficiency. Consistent with the initial observations, the LDAO-solubilized sample exhibited a minimal amount of insoluble material, while for other detergents the residual material was significantly more abundant. The solubilized samples were diluted ten times and then loaded into an SDS-PAGE gel. After electrophoresis, the gel was stained with Coomassie blue dye. The protein bands appear comparable for each sample, although the sample solubilized in LDAO displays a slightly more intense band, suggesting a higher amount of solubilized protein (Figure 45). Bands of membrane proteins fused with GFP may show a fluorescence signal in the polyacrylamide gel, if the SDS denaturation during sample preparation is conducted in mild conditions, i.e., without boiling the samples. However, for the samples of solubilized membranes containing the Wzx construct no fluorescent band was visible. A plausible explanation is that the concentration of SDS used for sample preparation was too high, and may have disrupted the GFP chromophore. Based on the solubilization efficiency test, LDAO was selected as the detergent of choice for subsequent purification experiments for the Wzx construct with the thrombin cleavage site.

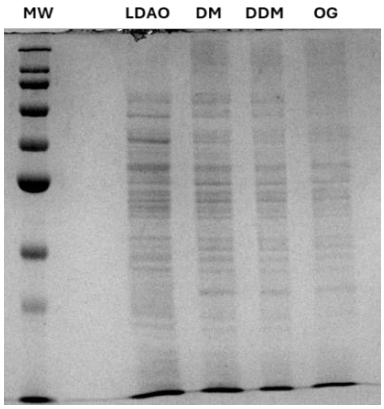


Figure 45 – SDS-PAGE analysis of the samples of membrane containing the overexpressed Wzx construct, solubilized with detergents LDAO, DM, DDM, and OG. In the first lane (MW), molecular weight markers were introduced. The gel was stained with Coomassie Blue staining.

After several unsuccessful attempts to optimize crystallization of Wzx samples solubilized in LDAO, an alternative detergent was evaluated. LMNG was selected as it has been reported to maintain membrane proteins in solution even at a concentration equal to the CMC. The low detergent concentration, in turn, may help preserve protein stability and reduce the risk of denaturation during crystallization. Despite the fact that LMNG proved to be less efficient in solubilizing the membrane samples, it was chosen to purify both the constructs of Wzx-GFP, containing the thrombin and the TEV protease cleavage sites.

Immobilized Metal Affinity Chromatography

Immobilized Metal Affinity Chromatography (IMAC) purification was employed as initial purification step, to remove contaminants from the solubilized membrane extract, in particular the membrane proteases often present in membrane samples. The presence of proteases in the sample often leads to proteolysis of the target protein, increasing the heterogeneity of the sample. The IMAC technique can be used not only to isolate the target protein from a highly heterogeneous sample, but also to concentrate the protein, reducing the sample volume. This is particularly important for membrane proteins, as protein purification usually starts from a large amount of membranes where the protein is present in low yield. In addition, the IMAC step can also be used to change detergent conditions towards a less denaturing solution, i.e., with a milder detergent or with a lower concentration. However, sudden changes in detergent type or concentration can cause protein misfolding and thus precipitation. Therefore, the detergent exchange should be performed in a stepwise manner. Detergent exchange using the IMAC resin, compared to the exchange in the gel filtration step, offers the significant advantage of that the resin can be easily washed and regenerated in case of protein precipitation, while the formation of aggregates or precipitates are usually very dangerous for the gel filtration matrices. In the present work, a reduction of the LMNG detergent concentration was performed while the protein remained bound to the IMAC resin. While the initial solubilization buffer contained an amount of LMNG equal to 10 times the CMC of this detergent (0.1mM), during IMAC wash steps the amount of detergent was reduced to the value of the CMC (0.01mM). The results of the IMAC purification of the construct containing a thrombin cleavage site are shown in Figure 46. In the left panel, the fluorescence of the elution fractions of the purification indicates the presence of the target protein. In addition, using a milder sample preparation protocol, the fluorescence could be observed also in the bands separated by SDS-PAGE analysis (right side panel in Figure 46).

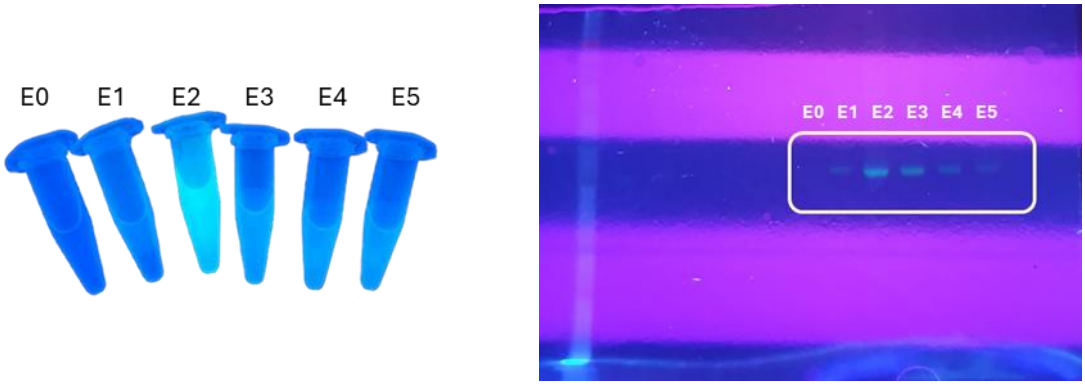


Figure 46 – (Right side panel) Fluorescence of the eluted fractions from the IMAC purification of the construct containing the thrombin cleavage site. Centrifuge tubes containing the fractions were visualized at the UV transilluminator. (Left side panel) In-gel fluorescence was observed in the SDS-PAGE analysis of the eluted fractions prior to Coomassie staining.

However, for both constructs tested, the eluted fractions consistently contained residual contaminant proteins, as clearly shown in SDS-PAGE analyses. It is possible that the contaminants carried during the first purification step have some affinity for the IMAC resin. In the standard protocol for IMAC purification, proteins with weak affinity for the resin should be removed during the second washing step with 5 mM imidazole (see Materials and Methods section). However, in this case, some contaminants were always found in the eluted fractions, together with the His-tagged protein, suggesting that either they possess a similar affinity for the resin or are tightly associated with the target protein, forming stable complexes that prevent their separation. In either case, an additional purification step is required to remove contaminant proteins as effectively as possible.

Size Exclusion Chromatography

Size Exclusion Chromatography (SEC) was chosen as second purification step. The SEC purification can separate proteins according to their size and is usually performed before setting up the crystallization experiments to remove protein aggregates. While the technique is able to provide also an estimation of the protein size, which correlates with the elution volume, in the case of membrane proteins, the expected elution volume is usually higher than the observed one due to the presence of the detergent micelle surrounding the protein. For the separation of the Wzx constructs, a SEC purification was performed on the samples after each IMAC purification. The separating column (Superdex 200 Inc.) was pre-equilibrated with a buffer solution containing the detergent used during prior purification steps. Both Wzx constructs were purified using LDAO detergent buffer (TBS pH 8.0 10xCMC). The SEC chromatogram shown in Figure 47 corresponds to a purification performed using LDAO. The blue line represents the absorbance at 280 nm of the total protein content in the elution, rather than that of Wzx specifically. The overlapping peaks indicate sample heterogeneity within individual fractions, suggesting that SEC could not efficiently separate the target protein. Fractions were collected and analyzed by SDS-PAGE. Each fraction was mixed with loading buffer and incubated for 10 minutes at room temperature. Avoiding sample boiling allowed the detection of GFP fluorescence even after electrophoresis. However, this incomplete denaturation led to an inaccurate correspondence with the molecular weight marker, as the folded protein migrated as if it were smaller in size. This behavior is evident in the gel shown in the Figure 47, where the most intense fluorescent bands in lanes 8, 9, and 10 indicate the presence of GFP, but at an apparent molecular weight lower than that expected for the Wzx_GFP construct.

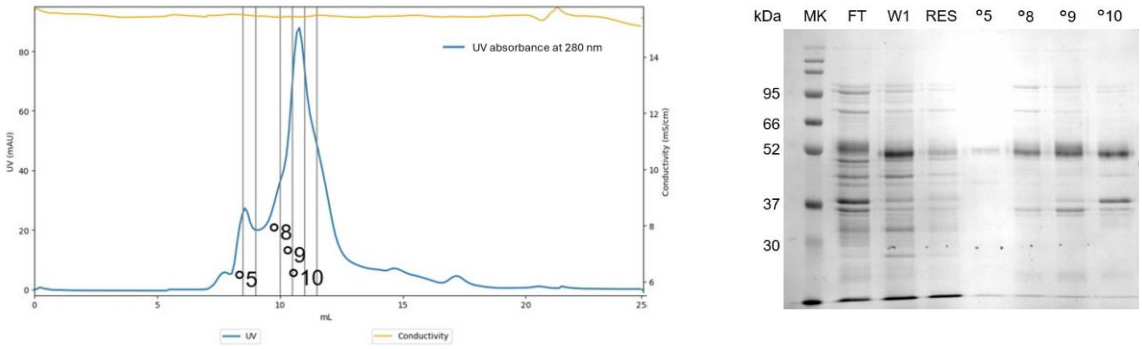


Figure 47 - On the left, SEC chromatogram of the Wzx_Thrb_GFP_10xHis purification on a 10xCMC buffer, with the elution fractions selected for SDS-PAGE analysis indicated. On the right, the corresponding SDS-PAGE 10% gel colored with Giotto dye. MK: MW marker; FT: flow-through; RES: TALON resin; numbered lanes correspond to the elution fractions collected from the SEC run.

The SEC profile, together with the corresponding SDS-PAGE analysis, supports the hypothesis previously suggested by the DLS and IMAC experiments that some contaminants might be strictly associated with the Wzx construct, co-eluting with it. Accordingly, in the SDS-PAGE gel, multiple protein species spanning a broad range of molecular weights are observed in correspondence with the eluted fractions (5, 8, 9, 10,

Figure 47, right panel). In particular, the SDS-PAGE gel reveals a main band corresponding to Wzx_GFP, identified by in-gel fluorescence under UV light, along with additional bands ranging from approximately 20 kDa to 100 kDa. After SEC separation, all components within a single fraction should, in principle, exhibit comparable hydrodynamic sizes. Three hypotheses may explain these observations: first, the gel reveals proteins with very different molecular weights but similar hydrodynamic radii, for instance, unfolded low-mass proteins and compact high-mass ones, although this scenario is rather unlikely. Second, the Wzx_GFP sample may have undergone partial degradation or aggregation during sample handling for electrophoresis, leading to higher-mass aggregates and lower-mass hydrolyzed fragments; however, this too appears improbable. A more plausible explanation is that Wzx forms detergent micelles that catch additional proteins, resulting in the co-elution of Wzx together with contaminant proteins of varying molecular mass. Although small quantities of contaminant proteins were still present after further optimization of the purification protocol, the sample was largely composed of Wzx, which appeared the most present specie in solution. For this reason, crystallization trials were nonetheless pursued.

Crystallization

Crystallization attempts of the Wzx_ThrbClSt protein

To obtain crystals of the Wzx_ThrbClSt_GFP_10xHis construct, several strategies were explored. This construct includes a GFP domain, which was initially retained during crystallization trials, as its hydrophilic surface was expected to promote crystal lattice formation for this membrane protein. Crystallization trials were performed under a large variety of conditions, testing dozen conditions using the Wzx protein solubilized in both LMNG and LDAO buffer solutions. The protein concentration varied between 0.2 mg/mL and 3.5 mg/mL, and all experiments were carried out using sitting-drop vapor diffusion plates. The drop volume ranged from 20 nL (dispensed using a SPTLabtech Mosquito) up to 1.5 µL. Crystallization plates were set up using various commercial screening kits to maximize the likelihood of crystal formation, including MemStart, MemSys, and MemGold from Molecular Dimensions. Despite the large number of attempts, no crystal formation was observed, even after prolonged incubation over several months. Based on this observation, the removal of the GFP tag was attempted, as the linking sequence connecting Wzx and the GFP tag structural flexibility to the construct. This added mobility was suspected to disrupt crystal packing and prevent proper lattice formation. Eliminating the tag was therefore expected to increase the chances of obtaining well-ordered crystals suitable for structural analysis. The efficiency of tag removal depends on the enzyme-to-substrate ratio, which must be adjusted according to the protein concentration. This optimization is difficult because diluted protein samples often fall below the detection range of BCA assays (~200 µg/mL), while concentrated fractions, which are easier to quantify, are typically available only in small volumes (30–90 µL). To overcome this limitation, the area under the curve (AUC) of the SEC profile can be used

as a comparative indicator of protein abundance between purifications. For instance, two samples showing comparable AUC values are expected to have similar protein concentrations; thus, if one is concentrated and quantified, its concentration factor can be used to estimate the protein abundance in the other. The cleavage procedure, however, presents a challenge related to the removal of the protease (in this case, thrombin) following digestion, as well as the subsequent purification of the cleavage mixture. Thrombin (Roche), used for tag removal, lacks any affinity tag that could allow its selective removal. Consequently, it can only be separated from the target protein by gel filtration chromatography. However, the SEC separation intrinsically dilutes samples, requiring an additional concentration step prior to setting up new crystallization trials, and this extra step introduces the risk of sample loss. Nevertheless, enzymatic cleavage was carried out to evaluate the efficiency of the process. Thrombin digestion was performed using an enzyme-to-substrate ratio of 1 unit of thrombin per milligram of SEC purified protein (2 mL). The reaction was tested under two conditions: overnight incubation at 4 °C and 4 hours at 18 °C. Each condition was evaluated for samples solubilized in either LDAO or LMNG detergent. After incubation, the samples were concentrated to 600 µL and purified again by gel filtration chromatography. However, the resulting SEC chromatograms revealed an insufficient amount of protein in the elution fractions, preventing subsequent concentration and crystallization of the purified material. It is possible that precipitation of the protein/protease mixture occurred during the concentration step prior SEC, likely due to the reduction of sample volume required for chromatographic injection.

Crystallization attempts of the Wzx_TEVClSt protein

The issues faced during the cleavage process led to the adoption of an alternative strategy designed to enhance proteolytic efficiency. Therefore, a new construct, Wzx_TEVClSt, was designed and produced. This variant includes a TEV protease cleavage site, allowing tag removal with TEV protease instead of thrombin. Although the general challenges associated with proteolytic cleavage remain similar (i.e. further purification steps and risk of sample loss), the use of TEV protease offers a key advantage: the enzyme carries a polyhistidine tag, enabling its selective removal by IMAC through a procedure known as reverse-IMAC. In this approach, after the cleavage reaction, the His-tagged protease and the cleaved GFP tag (which also retains its His-tag) bind to the resin, while the cleaved Wzx protein, lacking any affinity tag, is recovered in the flow-through fraction, separating it from the other components. The TEV cleavage reaction was tested on the SEC purified Wzx_TEVClSt protein in an LMNG-containing buffer using a 1:10 protease/sample ratio at 4°C overnight. After incubation a reverse IMAC was performed and the collected fractions were analyzed. However, the SDS-PAGE analysis of these samples did not reveal any protein bands in the flowthrough fraction, where a band corresponding to the cleaved Wzx protein (~55 kDa) was expected. Moreover, two bands appeared in the elution lane, migrating at approximately 30 kDa and 75 kDa. The first band corresponds to the molecular weight of both GFP and TEV

protease, while the second matches the uncleaved Wzx_TEVClSt_GFP molecular weight, indicating that proteolysis did not occur. Further optimization of the protocol was carried out by changing protocol parameters however, a sample suitable for crystallization could not be obtained.

Circular Dichroism

To investigate the secondary structure and thermal stability of the protein, circular dichroism (CD) spectroscopy was performed. For this analysis, the uncleaved Wzx_ThrbClSt_GFP_10xHis construct in LDAO 10xCMC buffer solution was used. Prior to measurement, the sample underwent IMAC and SEC purification and was concentrated to approximately 1 mg/mL. The sample was placed in a 0.5 mm quartz cuvette and analyzed using a Jasco J-815 spectropolarimeter. The CD spectrum at room temperature confirmed that the protein adopts a predominantly α-helical conformation, as evidenced by the characteristic negative minima near 208 and 222 nm (Figure 48). To assess the thermal stability of the protein, CD spectra were recorded over a temperature range from 18 °C to 80 °C (as reported in Figure 48 and Figure 49). This analysis allowed monitoring of the progressive decrease in the α-helical signal as the temperature increased, reflecting the loss of secondary structure associated with protein denaturation. The evolution of the ellipticity at 222 nm along the temperature gradient provided quantitative information on the thermal stability correlated with the unfolding of the protein, offering insights into its structural robustness under heat-induced stress. In Figure 48, the spectra show a progressive evolution of the signal, narrowing as the temperature increases from room temperature to 80 °C. Wavelengths below 200 nm are strongly influenced by buffer absorption, and given the high noise-to-signal ratio in this region, should not be considered reliable for analysis.

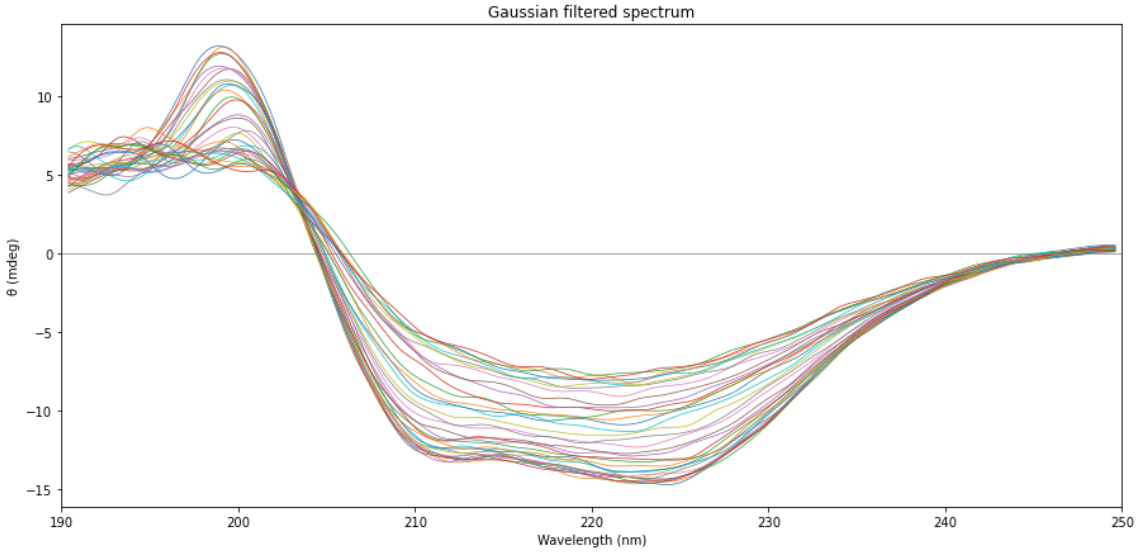


Figure 48 - CD spectra recorded between 18 °C and 80 °C. Each spectrum is smoothed using a Gaussian filter (sigma=3).

The melting temperature (T_m) of the protein was estimated as the temperature corresponding to the inflection point of the curve describing the variation of the CD signal

at a single wavelength as a function of temperature. In other words, the CD spectra reported in Figure 49 can be interpreted as chromatograms of ellipticity versus temperature, each representing the thermal denaturation profile at a given wavelength. To determine the position of the inflection point, the experimental data were fitted with a general sigmoidal polynomial function. Figure 50 shows the fitted curve for data collected at 222 nm, which corresponds to a characteristic α-helical signal of the protein and displays one of the largest amplitude variations in the dataset. The fitting procedure was performed using a Python script, which also computed the Tm value which is the temperature corresponding to the zero of the second derivative of the fitted curve. To evaluate the precision of the estimate, the script calculated the standard deviation of T_m values obtained across all analyzed wavelengths. However, by convention, the confidence interval was defined as ±4 °C. Accordingly, the melting temperature of the protein was determined to be 53.4 ± 4.0 °C.

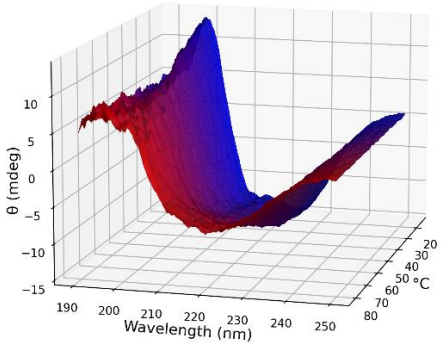


Figure 49 – 3D representation of CD spectra of Wzx recorded between 18 °C and 80 °C. The evolution of the signal is illustrated as a function of wavelength and temperature, with colors ranging from blue (low temperature) to red (high temperature).

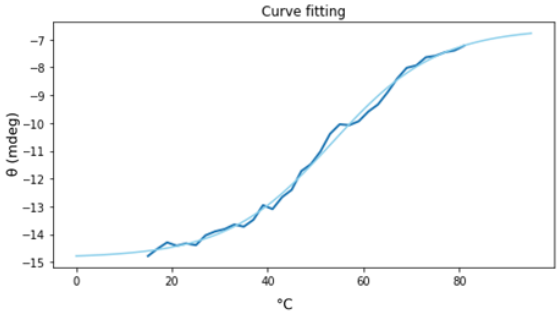


Figure 50 – CD signal at 222 nm as function of temperature. The blue line represents the experimental data, while the light-blue curve corresponds to a polynomial fit of the dataset. The inflection point (melting temperature T_m), is observed at 53.4 °C.

The protein showed unfolding at relatively mild temperatures, indicating only moderate thermal stability. Consequently, the detergent was changed, and purification was subsequently performed using LMNG.

Transmission Electron Microscopy

Transmission electron microscopy (TEM) can provide valuable insights into both the quality of a protein sample and its structural features. In this work, TEM was employed as a preliminary technique to support subsequent cryo-EM characterization of Wzx construct, using UranyLess for negative staining. A key advantage of this approach is its ability to visualize proteins even in the presence of a small amount of contaminants. Since the Wzx protein, regardless of the purification strategy used, retained minor contaminant particles which could not be completely removed, it still represents a suitable candidate for electron microscopy analysis. Such impurities can be effectively excluded through particle classification and selection steps during data processing. The Wzx construct containing the thrombin cleavage site and the GFP was purified in a LMNG

buffer solution, concentrated to ~2mg/mL and analyzed with a transmission electron microscope. These analyses were carried out at the CNB Institute in Braga as part of the preliminary steps for cryo-EM data acquisition which was also performed at the same facility. The TEM image reported in Figure 51 provides an initial qualitative assessment of the grid prepared with Wzx_ThrbClSt_GFP_10xHis. The overall appearance suggests that the protein concentration used for grid preparation was likely too high, resulting in a continuous "carpet" of material that completely covers the grid surface. Under these conditions, it becomes difficult to evaluate whether the particles are monodispersed or aggregated, since overlapping particles obscure individual features and prevent an accurate evaluation of particle distribution and orientation. At this concentration, the excessive coverage may also prevent optimal particle embedding within the vitreous ice layer during cryo-EM grid preparation, leading to preferential orientation, particle stacking, or thick ice regions. Ideally, the optimal grid should display well-separated particles distributed homogeneously across the carbon film, without visible aggregation or clustering.

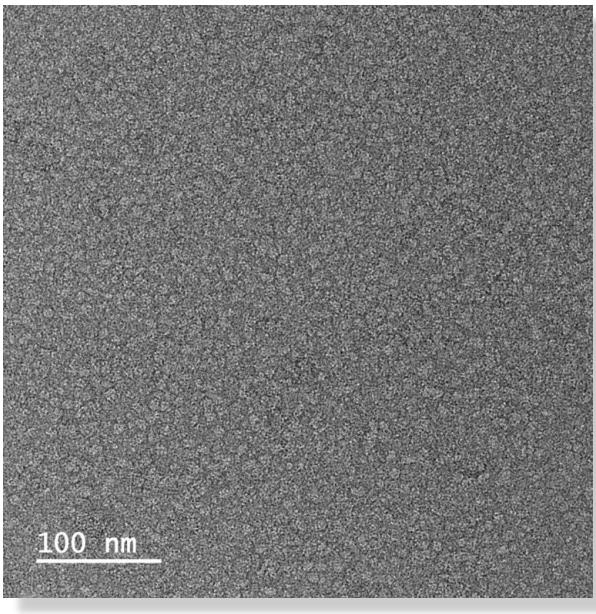


Figure 51 - TEM image of 2mg/mL Wzx_Thrb solubilized in LMNG, showing a highly crowded sample, which underscores the need for further dilution.

Overall, this grid confirms that the protein successfully adsorbed to the support film, providing useful preliminary information on sample behavior. However, the sample appears too concentrated for structural analysis, and dilution is recommended before proceeding with CryoEM vitrification. It is important to note that negative-stain TEM grid preparation differs substantially from that for CryoEM, mainly due to the use of heavy metal stains and the absence of sample vitrification. Nevertheless, such preliminary TEM analysis still offer valuable insights into particle quality, aggregation state, and sample homogeneity. Based on these observations, vitrification grids for CryoEM were subsequently prepared using diluted samples at 0.2 mg/mL and 0.1 mg/mL to optimize particle distribution.

Cryogenic Electron Microscopy

Given the unsuccessful crystallization attempts and the incomplete removal of contaminants, cryo-electron microscopy was selected as an alternative structural approach to investigate the structure of Wzx. For these experiments, the construct containing the thrombin cleavage site, dissolved in a LMNG buffer solution, was used. The GFP tag was intentionally left to improve visualization: maintaining the GFP, in fact, offers several advantages. GFP is a compact, electron-dense domain, which increases the overall molecular weight by approximately 45% and presumably also improves the contrast of the complex, thereby enhancing the visibility of the protein in cryo-EM micrographs. This is particularly relevant for small or low-contrast membrane proteins like Wzx, which might otherwise remain below the detection threshold. Nevertheless, the presence of GFP also introduces a potential disadvantage. The tag is connected to Wzx via a flexible linker region, which can lead to heterogeneous orientations and reduced particle alignment accuracy during image processing. Such conformational flexibility can blur the averaged signal and result in suboptimal reconstructions. Therefore, while GFP improves detectability, it simultaneously poses a challenge for high-resolution structure determination, requiring careful consideration in subsequent data analysis and classification steps. To carry out the analysis, the sample was transported to Braga, Portugal, where it was processed and vitrified for CryoEM analysis. After a rapid screening with a cryo microscope, the grid prepared with 0.1 mg/mL of protein displayed better ice quality and holes integrity, and was therefore selected for data collection. The overnight acquisition yielded 2343 movies, recorded at a pixel size of 0.6 Å with electrons accelerated at 200 kV and a total dose of 70 $e^{-}/Å^{2}$. Each movie consisted of 40 frames at a resolution of 8192x8192 pixels. At first inspection (Figure 52), the sample appeared remarkably different from that observed in the negative-stain TEM grid. In this case, no aggregates were visible, yet no distinct protein particles could be clearly identified.

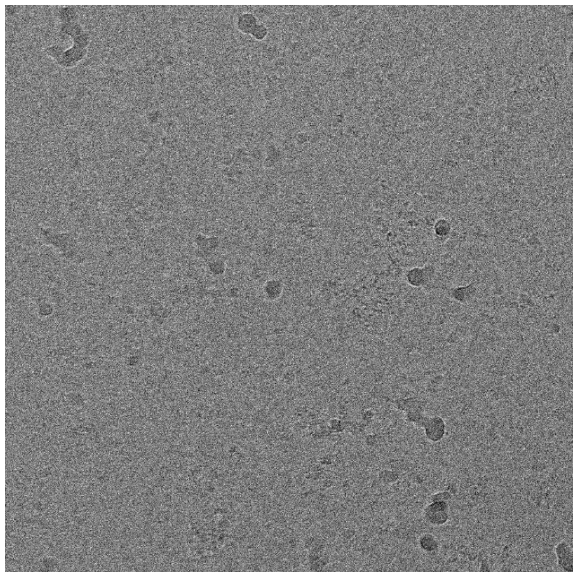


Figure 52 - Representative section from a cryo-EM micrograph of the Wzx_Thrb construct at 0.1 mg/mL

The dataset was transported from Braga to Trieste and processed using the computational resources of SISSA, using the CryoSPARC software suite The movies were first motion-corrected to compensate for beam-induced drift and thus minimizing blurring. Subsequent preprocessing steps included manual curation of the micrographs, in which 224 exposures (10% of the dataset) were excluded based on multiple quality parameters, such as signal intensity, defocus, astigmatism, CTF fit resolution, defocus tilt angle, ice thickness, motion distance, etc. to remove outliers and retain only the highest-quality images for further analysis. Since no individual particles were visually detectable in the micrographs, an automatic blob-picking procedure was employed to identify all potential protein projections. A subset of micrographs was first processed to identify potential projections of the target protein and to generate 2D templates from these views. Using these templates which appeared as blobs consistent with the expected dimensions of the protein, an automated search was then extended to the entire dataset, resulting in the identification of approximately 4 million candidate particles. Subsequently, manual curation was performed to remove aggregates, oversized or undersized blobs, and evident artifacts, reducing the dataset to roughly 100000 particles. The corresponding micrograph sections containing these particles were then extracted and subjected to 2D classification. Following this, a manual selection of well-defined classes was carried out to obtain the best subsets for preliminary 3D reconstruction. Despite several rounds of optimization, adjusting parameters at each stage of processing, no reliable 3D model of Wzx could be obtained. To address this limitation, a collaboration was initiated with Dr. Dmitry A. Semchonok (ITQB, Braga, Portugal), whose CryoEM processing expertise was fundamental to advancing the analysis; by refining the workflow parameters, he achieved well-resolved 2D class averages (reported in Figure 53).

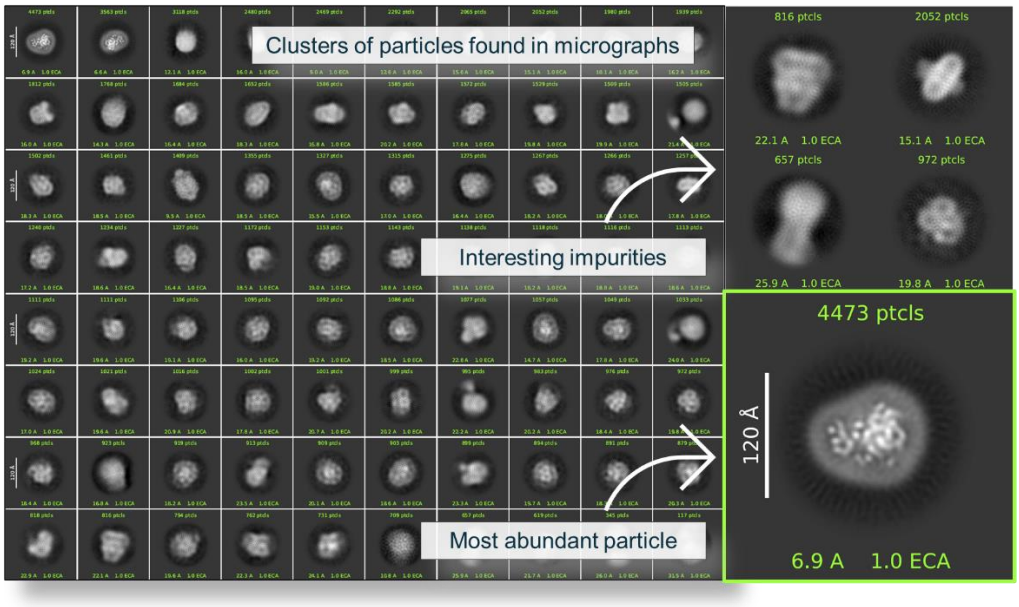


Figure 53 - 2D class averages from particle clustering, some interesting protein projections are reported in upper right part of the picture. The most aboundant particle is highlithed in green.

These averaged 2D classes, allow to make some interesting observations. The sample appears highly heterogeneous, as large amount of different projections are present. Some particles (indicated as “interesting impurities” by the arrow) likely correspond to some contaminant proteins, and one of these is even embedded within a detergent ring, confirming the presence of membrane associated material. The most interesting class, highlighted in green, is the one most likely representing a projection of the target protein, as it corresponds to the most abundant particle class. This class is characterized by a dense central core surrounded by ring-shaped region of lower density, which is consistent with the presence of a detergent micelle embedding a transmembrane protein. This class was used to reconstruct the electron density map of the protein–detergent complex, shown in Figure 54. From a first preliminary analysis, excluding the outer detergent ring, the inner density, which should correspond to the transmembrane segments of Wzx, appeared significantly larger than expected for a protein containing only twelve transmembrane helices. This observation suggested the possible presence of an additional domain or associated subunit, as schematically represented in Figure 55. Delimited by the outer detergent ring, there appears to be sufficient space to accommodate a second protein, represented as the second cylinder in the model. Moreover, the density map shows an extended, solvent-exposed region, which could correspond to the GFP tag, consistent with its expected position of the domain.

Figure 54 - Reconstructed 3D electron density map of the predominant protein population identified in the cryo-EM analysis, obtained by Dr. Semchonok (ITQB, Braga), likely representing Wzx protein embedded in detergent micelle.

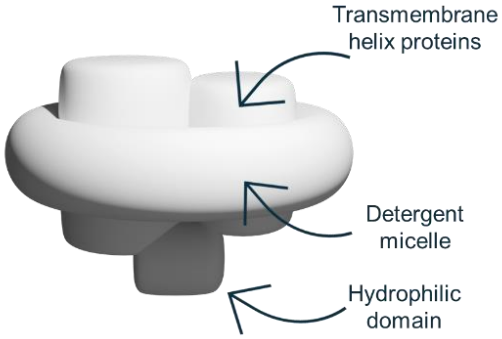


Figure 55 - Schematic representation of the hypothesized organization of the protein domains or subunit based on the density map. The TMSs are embedded within the detergent micelle, while the hydrophilic domain extends.

This interpretation is consistent with previous findings from IMAC and SEC purifications, which indicated that the detergent micelles may encapsulate multiple proteins, and that one or more contaminant proteins may be co-purified with Wzx, eluting together in each purification steps. However, after several attempts to fit the Wzx model into the obtained density, no satisfactory agreement was found, suggesting that the reconstructed map may correspond to a different protein. Considering prior results from a former PhD student in our laboratory, who characterized a contaminant while purifying a protein associated with Wzx, we hypothesize that the observed density could instead correspond to the same contaminant, known as Cytochrome O Ubiquinol Oxidase (pdbID:1FFT) [95]. When the structure of this protein was fitted into the density map, a

remarkably good correspondence was observed, supporting this interpretation. However, projections corresponding to Wzx were still missing. Therefore, to identify them, the search had to restart from the beginning. One could argue that, since a predicted structural model of the protein is available, it could be converted into a volumetric map, from which 2D projections at various Euler angles could be generated and used as templates to search for experimental matches within the dataset. Although this is a well-established approach, it was intentionally avoided at this preliminary stage to prevent introducing model bias into the analysis. Instead, the projection search was reinitiated, performing hundreds of iterations of parameter optimization in both blob picking and 2D classification. During this process, an interesting class was identified (Figure 56), as the appearance of this class was consistent with a plausible projection of Wzx_GFP, showing features compatible with the expected size and shape of the construct. This class was therefore used as a template to search for similar projections throughout the entire dataset.

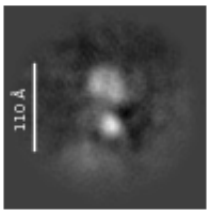


Figure 56 - Averaged class of Wzx projections, which was used as template for particle picking.

A limitation of this approach, however, lies in the orientation bias of the particles. The identified class represents only a singular side view of the protein, whereas top and bottom projections are missing. Although in theory a 3D reconstruction can be generated from a single projection, obtaining particle orientations from multiple directions is essential to improve map completeness and local resolution. The workflow, consisting of template definition, template-based particle picking, manual inspection of picks, micrograph extraction, 2D classification, iterative cleaning of poor-quality classes, and ab initio reconstruction, was repeated multiple times. Despite numerous optimization attempts, a high-resolution 3D model could not be obtained. For the final reconstruction, approximately 45000 particles were extracted from 2079 micrographs using a box size of 300 pixels. The extracted particles were then classified into 50 2D classes, and artifacts were removed in successive rounds of refinement until no additional classes required exclusion. The final dataset comprised roughly 12500 particles with an average of only 6 particles per micrographs. An ab initio 3D reconstruction performed on these particles generated two distinct density maps, as a single reconstruction displayed significant heterogeneity. One of these maps, derived from approximately 6000 particles, is shown in Figure 57, represented at different isosurface levels to illustrate the internal features of the density.

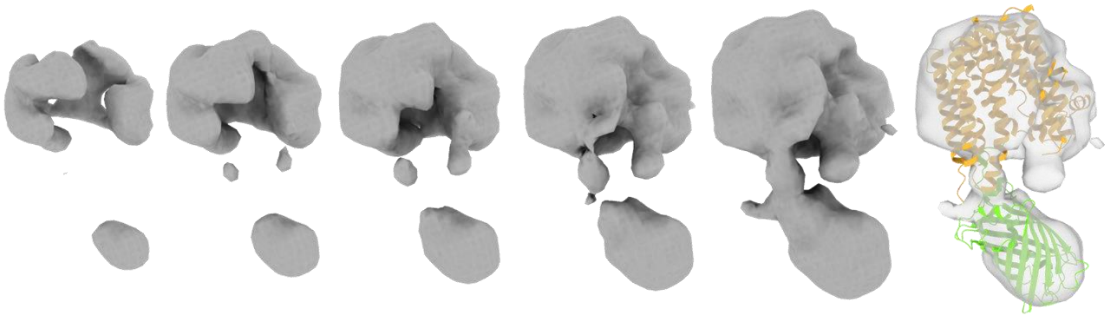


Figure 57 - The electron density map obtained from the final ab initio 3D reconstruction, which is hypothesized to correspond to the Wzx protein. On the right the AlphaFold model of the protein fitted in the density map.

The map in the picture reveals the internal structural organization of the reconstructed electron density. The overall morphology of the map is compatible with the expected dimensions of Wzx_GFP construct and the model can fit in the density map as shown in Figure 57. Despite this promising findings the limited number of well-defined 2D classes significantly reduced the available projections used for reconstructing the final map, thereby limiting the achievable resolution. Despite several additional rounds of refinement of the map, the maximum resolution obtained was approximately 8 Å, which does not allow reliable modeling of individual helices or side chains. While further refinement and validation could marginally improve map quality, the dataset is ultimately insufficient for high-resolution structural determination, which was the primary objective of this study. Nonetheless, these results demonstrate that the Wzx construct is suitable for cryo-EM visualization and that with improved sample quality and data collection, high-resolution reconstruction should be achievable in future work.

Wzx-Bril fusion protein

Following suboptimal results during the purification and enzymatic cleavage of previous constructs, a new optimization strategy was undertaken by introducing the BRIL (apocytochrome b562RIL) domain as a stabilizing scaffold. Two new constructs were designed: A Wzx–BRIL–GFP fusion, containing a thrombin cleavage site and a C-terminal His-tag, allowing both monitoring of expression through GFP fluorescence and tag removal prior to crystallization and a Wzx–BRIL construct in which BRIL replaced GFP, followed by a TEV protease cleavage site and a His-tag, designed to enhance the conformational stability of the membrane protein during crystallization. To identify the optimal construct and the most suitable linker length between Wzx and BRIL, AlphaFold3 was employed to predict the 3D structures and assess the most favorable configuration. Several models were generated, connecting BRIL to the C-terminus of Wzx via linkers composed of varying numbers of alanine residues. Structural predictions showed that even a single alanine residue significantly affected the relative orientation of the two domains. The resulting model is reported in *Figure 58*.

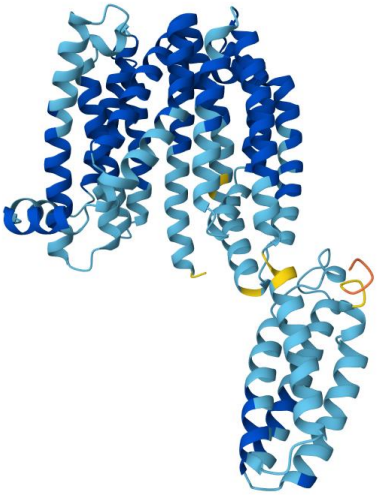


Figure 58 - AlphaFold3 model of Wzx_BRIL construct

Both constructs were cloned using the Restriction-Free (RF) cloning method, which allows seamless insertion of heterologous DNA sequences without restriction enzymes. While cloning of the Wzx–BRIL–GFP construct was successful, the Wzx–BRIL (TEV) variant did not yield satisfactory results. Subsequently, expression tests of the successful Wzx–BRIL–GFP construct were performed in E. coli BL21(DE3) and Lemo21(DE3) strains under different induction conditions (OD_{600} = 0.4 or 0.8, IPTG = 0.4 or 0.8 mM, at 18 °C or 37 °C). Fluorescence measurements revealed that expression at 18 °C overnight with OD_{600} = 0.8 and 0.8 mM IPTG resulted in the highest protein yield and culture density, representing the most favorable condition for large-scale production.

Results and Discussion (EG)

Bacteriophages have been regarded as a possible solution to bacterial infections since before the discovery of antibiotics. However, their use has been limited due to the possible risks connected with the injection of viral particles is a patient. Thus, in the Western world, phages have been used only as a last resort, when other more conventional therapies fail. While bacteriophage therapy may remain of limited application, the weapons that these viral particles use against bacteria are studied as a possible innovative solution to drug resistance in bacteria. The idea of using proteins of phages comes from their high specificity against a single species of bacteria, and sometimes even a single strain. This targeted therapy has the potential to avoid the clinical issues connected to the use of broad spectrum antibiotics. In this thesis, an enzyme from the bacteriophage ϕBO1E [96] that is able to cleave the glycoside bonds in the capsular polysaccharide of *Klebsiella pneumoniae* [97] was studied for its potential applications in medical therapies against drug resistant bacteria. The gene coding for this enzyme was identified by the group of Prof. Marco Maria D'Andrea (University of Roma Tor Vergata) and later expressed as a recombinant protein by the groups of Proff. Paola Cescutti and Rita De Zorzi (University of Trieste). The construct expressed in Trieste contained a C-terminal histidine tag and, thus, was purified by affinity chromatography with a IMAC resin and by size exclusion chromatography. However, the purification did not yield a stable sample and, more importantly, the enzyme did not display any catalytic activity against strains of *K. pneumoniae* producing the capsular polysaccharide.

AlphaFold model of the endoglycosidase enzyme

Several endoglycosidases have been previously characterized, including Depo32 a phage depolymerase which targets the K2 serotype *K. pneumoniae*. However, an initial sequence search performed using the sequence of the endoglycosidase protein from the ϕBO1E phage against *K. pneumoniae* revealed a very low identity with most of the already characterized enzymes of this class. More generally, the known enzymes belonging to this family exhibit low sequence similarity, probably as a result of substrate specificity. Among these, Depo32 (pdbID: 7VYV) (Figure 59), HK620 tailspike (pdbID: 2X6Y), KP32gp38 (pdbID: 6TKU), and APK14_gp49 (pdbID: 8OQ1) are all depolymerases from phage viruses targeting *Klebsiella* or other bacterial strains. However, conserved structural features are probably connected to the enzymatic function of these proteins, particularly as regards the architecture of the active site. The phage endoglycosidases that have been characterized adopt a central trimeric β-helix fold, in which precise substrate-binding modules are combined with acid/base residues crucial for the catalytic mechanism. The β-helical central domain (in blue in Figure 59) is always present in the reported structures and most of the endoglycosidases share also a C-terminal carbohydrate-binding module involved in the trimerization (in yellow in the figure), which in most cases is required for the activity of the enzyme. In addition, in some of the

enzymes of this family, other domains have been observed, such as a N-terminal α-helix bundle observed for Depo32 (Figure 59, in red). This domain, in particular, breaks the trimeric symmetry of the complex and probably constitutes a flexible tethering to the surface of the viral particles. However, the N-terminal portion is only present in some of the characterized endoglycosidases.

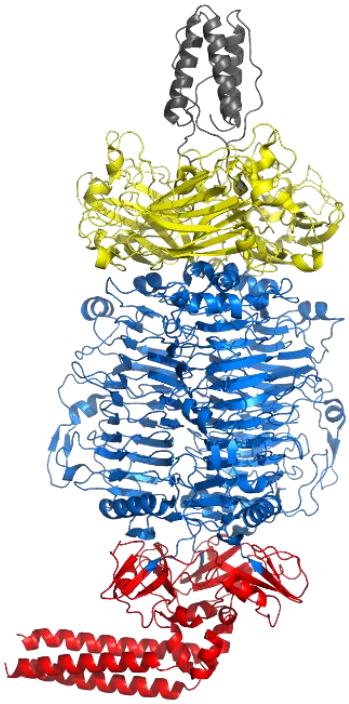


Figure 59 - Domain of Depo32. The N-terminal domain in shown in red, the central domain in blue, the C-terminal in yellow with terminal α-helices highlighted in gray.

Extensive biochemical investigations across various phage enzymes confirm a conserved catalytic paradigm. Each enzyme employs two carboxylate residues spaced approximately 5–7 Å for the cleavage. These acidic residues work in cooperation with aromatic residues, such as tryptophan and tyrosine, which provide stacking interactions that stabilize the substrate within β-helix grooves. Considering the low homology of the endoglycosidase identified by the group of Prof. D'Andrea, the AlphaFold tool was applied to elucidate the reasons for the lack of activity of the recombinant protein expressed in the laboratories of the University of Trieste. The sequence of the construct containing the C-terminal histidine tag was used as input for the DeepLearning software and, thanks to the options available on the second version of the program, a trimeric quaternary structure was imposed. The result shows an arrangement of each domain consistent with the known structures of endoglycosidases, even if the superimposition of the model with the available experimental structures was only partial. In addition, AlphaFold was used to generate also a trimeric structure of the protein sequence without the tag (Figure 60). The comparison between the two models showed that the tag significantly interferes with the stability of the trimeric folding, suggesting to remove it from the construct, or to change its position to the N-terminus. The second option was chosen, as it allows to maintain a IMAC step in the purification protocol.

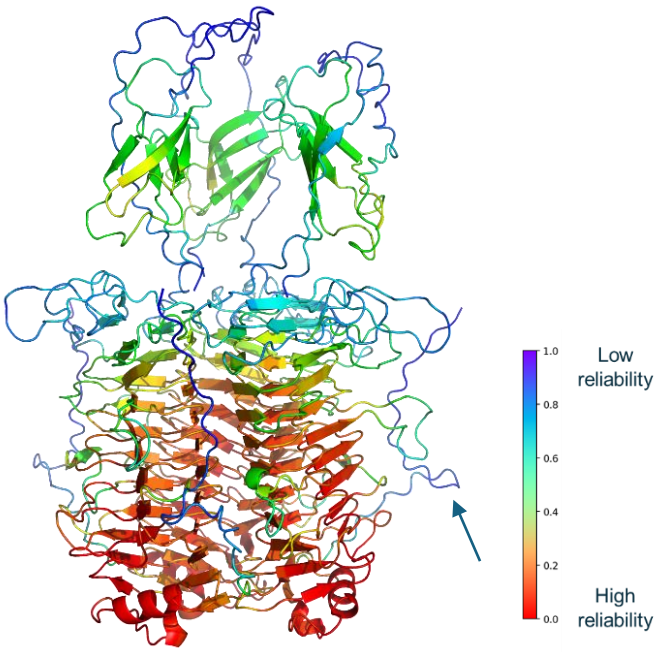


Figure 60 - Model of the endoglycosidase protein of the bacteriophage ϕBO1E calculated by AlphFold2, imposing a trimeric arrangement. Colors represent reliability of the estimation in terms of plDDT. The arrow identifies the N-terminal flexible region of the protein.

The construct with the N-terminal histidine tag was expressed and purified by Dr. Michela Zaro of the laboratory of Prof. Cescutti. Tests on *K. pneumonia* bacteria expressing the capsular polysaccharide showed that the enzyme was active and, albeit the cells were still alive, they lost their capsular envelope. The AlphaFold model was further analyzed in view of its structural characterization. In particular, the software predicted a low reliability for the N-terminal region (arrow in Figure 60), suggesting a highly flexible segment or an unstructured region. In the Laboratory of Prof. De Zorzi, a series of endoglycosidase constructs featuring progressive truncations at the N-terminus were designed to reduce the flexibility of the protein and, hence, improve the crystallization probability. These constructs contained also the N-terminal histidine sequence required for IMAC purification, followed by the TEV protease cleavage site, useful to enzymatically remove the protein tag. Among these variants, the construct lacking the first eight residues (EG-NΔ8) was successfully expressed and purified. The activity of the protein was analyzed on *K. pneumonia* bacteria, proving that the removal of the N-terminal tail did not affect protein function. After purification, the tag was removed through enzymatic reaction with the TEV protease, followed by two additional purification steps, a reverse affinity chromatography and a SEC purification.

Structural determination of EGN

Crystallization trials for the truncated enzyme EG-NΔ8 were performed using the vapor diffusion method in a sitting-drop configuration, set up with a Mosquito robotic system to enable high-throughput screening. Approximately 200 distinct crystallization

conditions were tested, among which only one produced a crystal with dimensions suitable for X-ray diffraction analysis (0.1 M sodium acetate trihydrate pH 4.5 and 20% w/v PEG 3350). The crystal, which reached its full size after nearly two months of incubation, was carefully harvested, cryoprotected with a solution containing glycerol 20%, and flash-frozen in liquid nitrogen prior to data collection. Diffraction data were collected at the XRD2 beamline of the Elettra Synchrotron (Trieste), yielding an initial dataset at 1.50 Å resolution. The EG-NΔ8 crystal was indexed in the trigonal system, space group P 3 2 1, unit cell parameters *a* = 83.77 Å and *c* = 130.71 Å, with an asymmetric unit containing a single protein monomer. The structure was solved by molecular replacement using a model obtained from the prediction of AlphaFold2 (reported in Figure 60). In the probe used for phasing, only the central core of the protein corresponding to the region with the highest confidence in the structural prediction was included, representing approximately 56% of the total sequence. Following multiple cycles of manual model building and reciprocal space refinement, the C-terminal portion of the protein remained incomplete. To address this, the Buccaneer automated model building pipeline (part of the CCP4i2 suite) was successfully employed, reconstructing the remaining regions. Given the absence of radiation damage and the high signal-to-noise ratio of the diffraction data, the same crystal was subsequently used for a second data collection, which extended the resolution limit to 1.10 Å, indicative of exceptional crystal quality. The functional trimer was generated by the crystallographic threefold symmetry axis (Figure 61). Electron density was well defined for nearly the entire chain, except for the first seven N-terminal residues (Ser1–Tyr7), which were not visible in the maps, likely due to their intrinsic conformational flexibility. Statistics of the scaling and refinement are reported in Table 2 (Appendix)

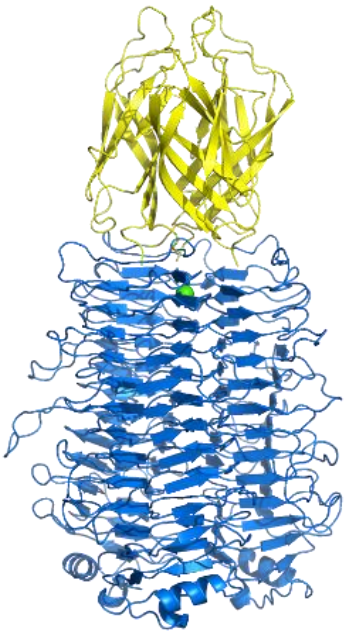


Figure 61 - Domain organization of the EG from ϕBO1E. The N-terminal central domain is shown in blue, while the C-terminal domain is highlighted in yellow.

The secondary structure of this enzyme closely resembles that of several previously characterized endoglycosidases including Depo32 a phage depolymerase that targets the K2 serotype of K. pneumonia (Figure 59). Notably, compared to Depo32, the endoglycosidase from ϕBO1E lacks the N-terminal α-helical bundle (highlighted in red in Figure 59) commonly observed in other phage depolymerases, and it also features a shorter C-terminal domain that is missing the extended α-helix sequence (shown in gray in Figure 59). In the EG-NΔ8 protein, the central domain (Thr8-Tyr371, in blue in Figure 59) comprises two short, flexible α-helices and a right-handed parallel β-helix fold composed of 13 complete turns. This β-helical structure is a conserved feature among phage-derived glycosidases (as shown in Figure 59 and Figure 61) and plays a crucial role in substrate recognition and catalysis. The first eleven turns consist of three β-strands, while the last two include two β-strands. A linking loop (Thr372-Pro389) connects the central β-helical domain to the C-terminal region (Ile390–Phe487, shown in yellow in Figure 61), which adopts a β-sandwich architecture composed of six β-strands. The structural characteristics of this C-terminal domain are typical of carbohydrate-binding modules, suggesting a role in substrate recognition or stabilization, even though the active site of other endoglycosidases was identified in the central β-helical domain. The biologically active form of the enzyme is a homotrimer of approximately 75 Å in diameter and 125 Å in length. The interfaces between monomers (shown in different colors in Figure 62) are primarily mediated by residues located in the β-strands of the central β-helix domain, and the quaternary structure is further stabilized by the coordination of a Ca2+ ion (visible in both top and bottom views in Figure 62 and in green in Figure 61), which in turn coordinates also three water molecules.

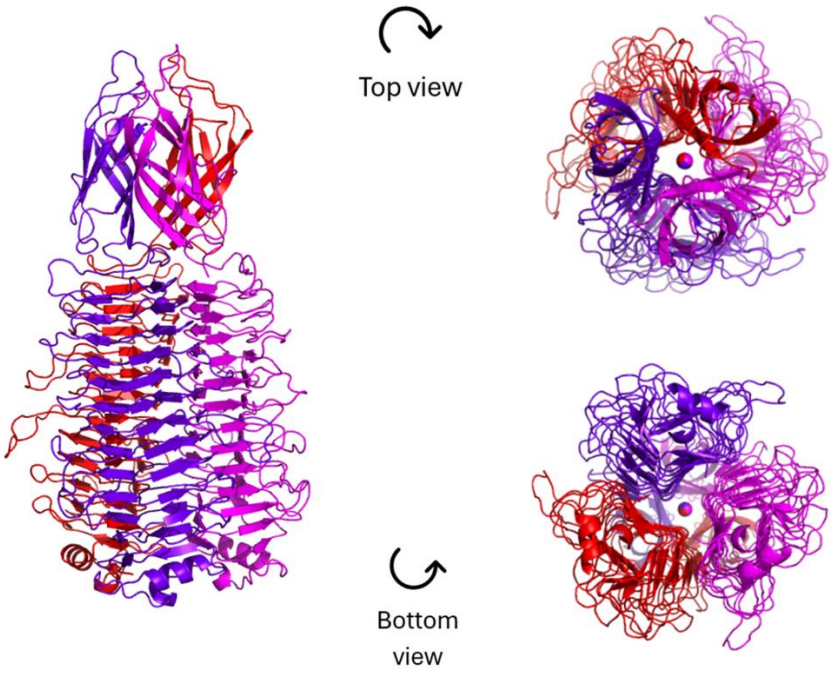


Figure 62 - Refined experimental structure of the endoglycosidase from bacteriophage ϕBO1E. Individual monomers are shown in distinct colors.

Endoglycosidases catalyze the hydrolysis of glycosidic bonds through an acid–base mechanism that typically involves Asp and Glu residues, assisted by aromatic amino acids such as Tyr and Trp, which stabilize the substrate through stacking interactions. Analyzing the positions of the Asp, Glu, Tyr, and Trp residues in the EG-NΔ8 structure, a potential catalytic site was identified at the interface between two subunits, within the central β-helix domain (Figure 63). In this region, the distance between Glu211 and Asp243 (~5.8 Å) is comparable to that observed between the catalytic residues Asp339 and Glu372 in the HK620 tailspike (5.5–6.0 Å) which has been characterized both in its *apo* and *holo* forms, thus suggesting a similar catalytic role [98]. Unlike HK620, where both catalytic residues belong to the same polypeptide chain, in EG-NΔ8 they are located in adjacent subunits. Nevertheless, the clefts containing the putative active sites exhibit a conformation similar to that of the HK620 enzyme.

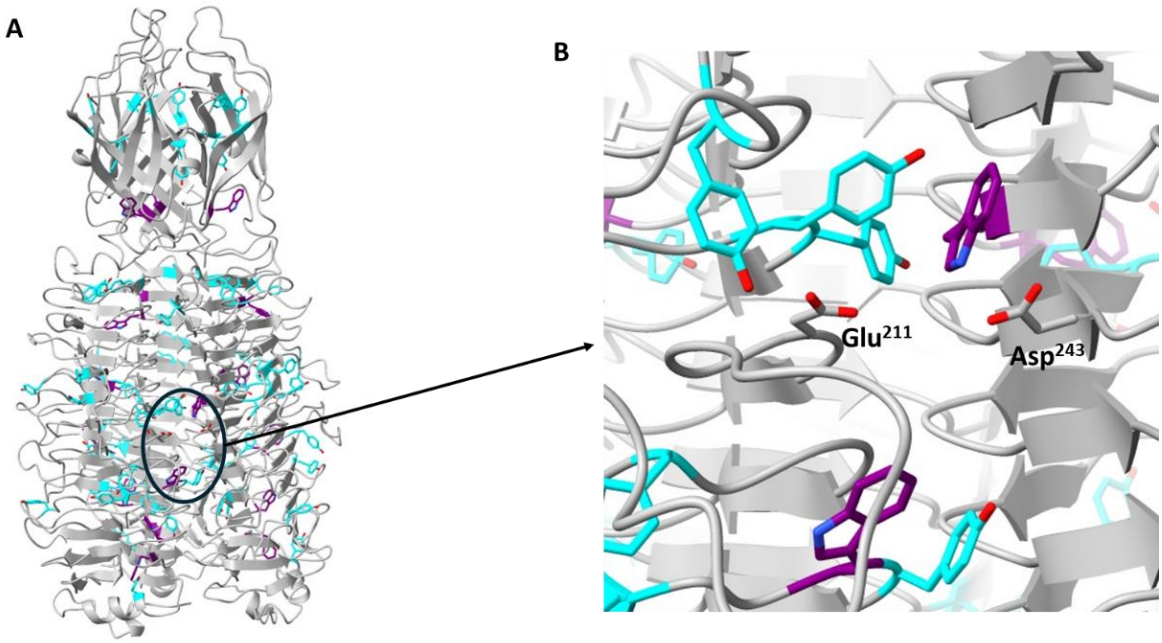


Figure 63 - Identification of the putative active site of the EG-N*Δ8 protein, at the interface between two polypeptide chains. (A) Representation of the protein structure highlighting the location of tryptophan (magenta) and tyrosine (cyan) residues. (B) Detail of the putative binding site showing the position of the glutamate and aspartate residues (sticks) and the aromatic side chains of tyrosine (cyan) and tryptophan (magenta) residues*

At this stage, it is worth examining the role of the structural model generated by AlphaFold. The AlphaFold model was crucial in revealing how the addition of a C-terminal tag influenced protein folding, probably hampering the formation of the active trimer and thereby preventing enzymatic activity. Furthermore, the predicted model was fundamental in solving the crystallographic phase problem through molecular replacement. However, once the experimental structure was available, the comparison with the predicted model shows that the machine learning–derived conformation differs substantially from the experimental structure, especially in the C-terminal domain. The structural characterization of this endoglycosidase represents an emblematic example of how artificial intelligence and experimental approaches should not be seen as alternatives, but rather as complementary tools.

CryoEM analysis of the endoglycosidase-substrate complex

To identify the position of the active site of the enzyme, a cryo-EM analysis was performed using a sample of the full-length endoglycosidase and a substrate. The *in vivo* substrate of this enzyme is a large and probably partially disordered polysaccharide, but previous analyses of the digestion products of the *in vitro* reaction revealed that the enzyme is able to hydrolyze the polymer to the smallest repetitive unit (unpublished results, Paola Cescutti's group), which in the case of the capsular polysaccharide of *K. pneumoniae* consists of a pentasaccharide backbone with a rhamnose side chain [97]. Therefore, a dimer of repetitive units, provided by Dr. Michela Zaro (Paola Cescutti's laboratory) was used in the Cryo-EM analysis. Before vitrification, the sample quality was carefully evaluated by negative-stain TEM analysis. This preliminary screening step, performed on the protein without substrate, was crucial to assess particle integrity, homogeneity, and dispersion. For cryo-EM grid preparation, the protein concentration was systematically adjusted between 0.5 to 2.0 mg/mL to identify the best conditions for achieving an even particle distribution while avoiding aggregation. The screening process was carried out using the same Talos Arctica microscope employed for the subsequent data collection (Figure 64). This procedure aimed at identifying the regions of the grid where particles were most evenly distributed. The grid squares (panel A) are classified based on their size and brightness (panel B), parameters that reflect the thickness of the vitreous ice. Squares containing damaged areas or ice crystals are identified at low magnification and manually deselected to be excluded (panel C). Each selected square contains multiple holes (panel E), which are then individually chosen to define the regions for data acquisition. Even at this preliminary screening stage, protein particles can already be observed as distinct projections within the holes, as illustrated in panel F.

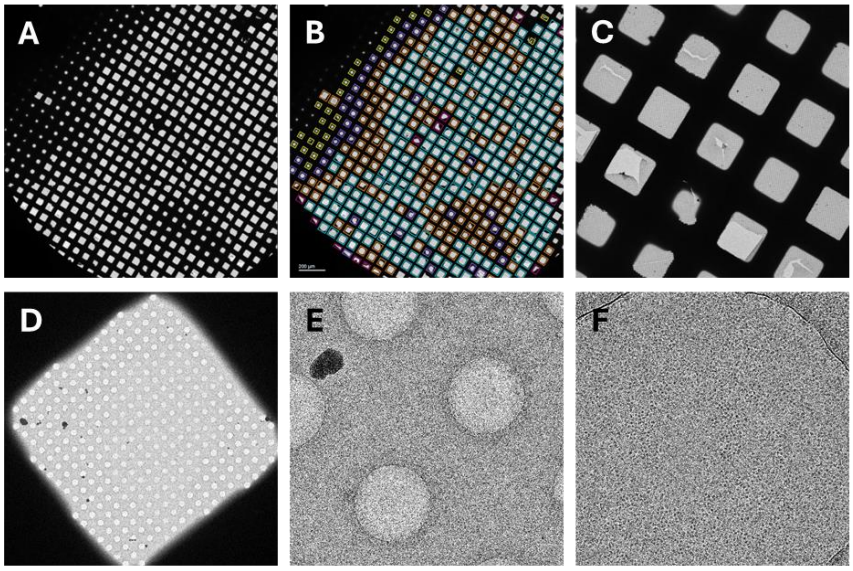


Figure 64 - Cryo-EM grid screening at different magnifications. (A) Overview of the vitrified grid showing the distribution of all squares. (B) Atlas view with automated ice-thickness analysis. (C) Close-up of a damaged region containing broken carbon film. (D) Representative square suitable for data collection. (E) Detail of individual holes. (F) High-magnification micrograph of a selected hole, where particles are clearly visible as small dark dots.

Two independent data collection sessions were performed. The first dataset of the enzyme at 0.5 mg/mL, without substrate, provided excellent results in terms of particle abundance, distribution, and concentration. Based on these findings, a second acquisition was undertaken to investigate the enzyme-substrate complex. For this purpose, additional grids were prepared containing the protein (0.5 mg/mL) supplemented with the dimer of repeated units at a concentration 75 times higher than the trimeric protein, which corresponds to 25 times the number of active sites. For each session, more than 5000 movies were recorded. The large datasets were processed using the CryoSPARC suite, following a workflow similar to the one employed for the Wzx reconstruction. However, these datasets presented remarkably different characteristics. The signal-to-noise ratio was significantly improved, with particles being clearly visible and each micrograph contained hundreds of well-defined particles. Despite these advantages, the datasets had the major limitation of strong preferential orientation. Most particles were aligned with their symmetry axis perpendicular to the grid plane, resulting in a predominance of top-bottom views, while side views were almost absent. This anisotropic distribution of particle orientations posed a challenge for obtaining a high resolution three-dimensional reconstruction. To overcome this issue, a manual particle picking step was added to the data processing protocol. The goal was to generate a side-view template of the protein, which was underrepresented in the dataset. Approximately 500 projections were manually selected and used to train a machine learning-based particle picker (TOPAZ) to automatically identify similar views across all micrographs. However, in this specific case, the standard template-based picker proved to be more effective. For the dataset of the *apo* protein, From around 2.5 million side-view particles (comprehensive of artifacts and false positive projections) were identified, and a subset of 100000 was extracted and subjected to 2D classification into 50 classes. After discarding poor-quality and artifact-containing classes, the classification was iteratively refined (over ten rounds) to progressively eliminate noisy classes. This process yielded a final set of approximately 60000 side view projections of the endoglycosidase, which was then merged with a class of 13000 top-bottom view particles for subsequent 3D reconstruction. After the *ab-initio* reconstruction, several rounds of refinement were carried out to maximize the local resolution of the map. During this process, particle projections were filtered at 4 Å, and an initial low-pass filter of 10 Å was applied to each input model. In addition, per-particle defocus refinement was performed, while CTF parameters were optimized for each group of particles originating from the same exposure. Finally, C3 symmetry was imposed during the final refinement step of the structure of the *apo* protein, yielding the map shown in Figure 65. The reconstructed volume reveals the overall architecture of the trimeric particle, characterized by an elongated shape with a wider base and a narrower top, consistent with the expected morphology of the protein. Moreover, the volume displays some secondary-structure features, particularly along the central β-helix core. The structural alignment with the X-ray structure, reported in Figure 66Figure 66, demonstrates a strong

correspondence between the crystallographic atomic model (depicted as blue cartoon) and the cryo-EM density map (shown as a semi-transparent gray surface). This fitting confirms the overall accuracy of the cryo-EM reconstruction and validates the structural consistency between the crystallographic and cryo-EM data. Some deviations are visible at the peripheral regions and might correspond to flexible loops or domain movements not fully resolved in the volume reconstruction.

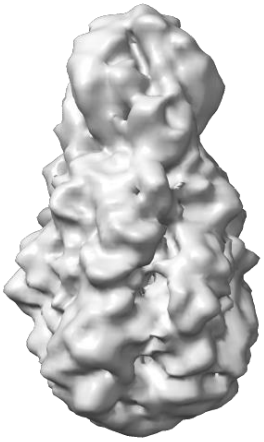


Figure 65 - Reconstructed map of the endoglycosidase protein without the substrate from cryo-EM data

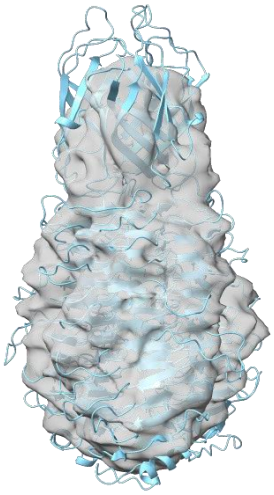


Figure 66 – Crystallographic structure fitted in the reconstructed map from cryo-EM data.

The reconstruction process is still ongoing at the University of Trieste, and the current map, which reaches at an approximate resolution of 3.5 Å, estimated by Gold-standard Fourier Shell Correlation (GSFSC) curve (Figure 67) should be considered preliminary. Further optimization and refinement are expected to improve the final resolution, although this will require substantial additional work. Nevertheless, this initial reconstruction provides a solid foundation for subsequent improvements, as the dataset exhibits excellent quality and strong potential for achieving a high-resolution structural model.

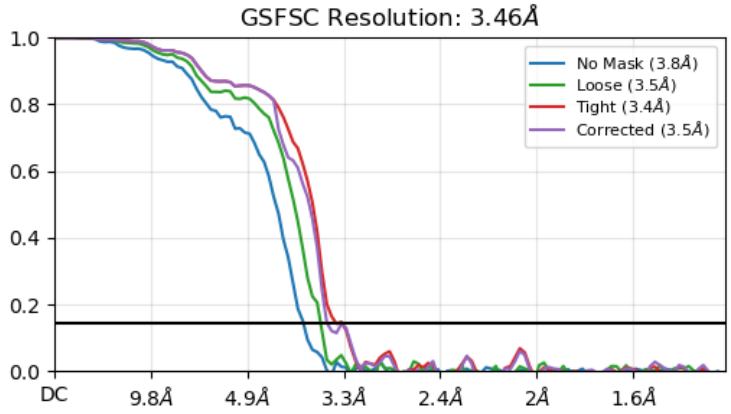


Figure 67 - Gold-standard Fourier Shell Correlation (GSFSC) curve of the reconstructed density map

The plot in Figure 67 reports the resolution estimation of the reconstructed cryo-EM map based on the Fourier Shell Correlation (FSC) between two independently refined maps. The FSC is calculated under several masking conditions, and the corresponding curves are represented in the plot. The intersection of the curves with the 0.143 FSC threshold (horizontal line) defines the resolution limit of the map, which, in this reconstruction, is approximately 3.5 Å. Although this 3D volume represents a reliable starting model, it may be necessary to repeat the particle picking and include additional projections, as suggested by the distribution plot shown in Figure 68. This heatmap represents the angular distribution of particle orientations used for the 3D reconstruction of the endoglycosidase enzyme. The color scale (on the right) indicates the number of particles assigned to each orientation. The plot shows a non-uniform angular distribution, with most particles clustered around specific azimuthal regions. This confirms the presence of preferential orientation in the vitrified grids which resulted in underrepresented particle views and consequently unbalanced classes in the 2D classification.

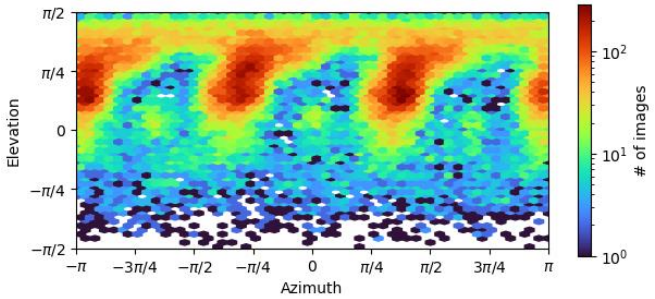


Figure 68 - Direction distribution plot, highlight the presence of non-uniform angular distribution of projections, likely due to underepresented views in particle classes

Regarding the second dataset, which includes movies of enzyme in the presence of its substrate, the identification of the active site still requires further analysis. Since the endoglycosidase trimer may contains three active sites, each positioned at the interface between two monomers, a possible asymmetric occupation of the sites should be considered. Thus, the application of C3 symmetry during reconstruction is not advisable in this case, as it could mask structural asymmetries that are crucial for understanding the catalytic mechanism. Preliminary reconstructions of the enzyme–substrate complex exhibited the same issue of preferential particle orientation observed in the apo form, resulting in underrepresented views within the dataset. Therefore, in the preliminary density map obtained for the complex, which exhibits a resolution lower than the *apo* map, the presence of the substrate could not be identified. To address this limitation, the particle-picking procedure for this system should also be further optimized to enhance particle diversity and improve the quality of future reconstructions.

Conclusions

In this thesis, the structural characterization of two key proteins connected to possible strategies against drug resistance in bacteria was pursued. Along the process, challenges encountered in the experimental work impelled the use of the state-of-the-art computational instruments both to aid overcoming the experimental challenges, and to seek biochemical insights into the dynamics of the systems. The study of the *Pseudomonas aeruginosa* Wzx flippase aimed to advance the structural and biochemical understanding of a key membrane protein involved in O-antigen biosynthesis, a process directly linked to bacterial virulence and antibiotic resistance. Due to its intrinsic hydrophobic nature and membrane embedding, Wzx presented significant challenges in terms of expression, purification, and structural characterization. The work began with the optimization of previously developed constructs and purification protocols, with the goal of improving protein yield and stability. Despite several purification attempts, residual contaminants and the limited stability of the protein, induced by detergent solubilization, limited the progress of the crystallographic analysis. To address these issues, new constructs were designed introducing the BRIL stabilizing domain, known to enhance membrane protein folding and crystallization efficiency. Computational modeling through AlphaFold3 guided the optimization of the linker length and the evaluation of construct stability. Experimental cloning using the Restriction-Free (RF) cloning technique led to the successful generation of a Wzx–BRIL–GFP construct. However, obtaining well-diffracting crystals remains challenging. Nevertheless, this work established the methodological and computational groundwork for future structural studies of Wzx, including the analysis through cryo-electron microscopy (cryo-EM), which is particularly suitable for facing issues in determining the structure of this membrane protein. Computational modelling of Wzx followed by Molecular Dynamics simulations were carried out in an environment closely resembling the physiological environment of the inner bacterial membrane. The results of the simulations confirmed the stability of the protein in the membrane and were compatible with the previously hypothesized mechanism of action of this protein. The stability of the protein was investigated also in a detergent micelle, suggesting that the higher degree of flexibility of the protein embedded in the micelle could be the main factor hampering experimental efforts to obtain structural information. The second part of this thesis focused on the structural and functional characterization of an endoglycosidase (EG) derived from bacteriophage ϕBO1E, active against the capsular polysaccharide (CPS) of *Klebsiella pneumoniae* KpB-1, a multidrug-resistant pathogen. This project was part of a collaborative effort aimed at exploring phage-derived enzymes as potential therapeutic tools against antimicrobial resistance. A truncated construct of the protein (EG-NΔ8) was designed based on an AlphaFold model that suggested a high flexibility of the N-terminal segment. The truncated protein, cloned into a construct containing a N-terminal histidine tag followed by a TEV protease cleavage site, was

successfully expressed and purified using affinity and size-exclusion chromatography. The N-terminal tag was removed by enzymatic reaction and the protein was further purified. Crystallization trials performed using the vapor diffusion method yielded a single crystal suitable for X-ray diffraction analysis. Data collected at the Elettra synchrotron allowed the determination of the protein structure at 1.3 Å resolution, representing the first detailed structural characterization of the ϕBO1E endoglycosidase. The enzyme adopts a right-handed parallel β-helix architecture typical of phage depolymerases and forms a trimeric assembly stabilized by a coordinated calcium ion. Structural comparison with homologous enzymes, such as Depo32 and the HK620 tailspike, identified a putative catalytic site located at the interface between monomers, and an aromatic groove potentially involved in substrate binding. Cryo-EM studies were initiated to verify the position of the active site. However, achieving a high-resolution structural reconstruction proved challenging and will require further efforts.

Both projects contribute to the growing field of structural biology aimed at developing innovative approaches to combat antimicrobial resistance. While the first study, on the flippase Wzx, addressed the complexities of a membrane-associated target relevant to bacterial envelope synthesis, the endoglycosidase project successfully elucidated the structure and function of a phage-derived enzyme with therapeutic potential. Together, these works illustrate how the integration of experimental techniques, such as X-ray crystallography and cryo-EM, with computational modeling can provide complementary insights into complex biological systems.

Appendice

Wzx protein parameters

UniProtID: G3XD19_PSEAE (G3XD19)
Description: O-antigen translocase
Organism: *Pseudomonas aeruginosa* (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)
Number of amino acids: 411
Theoretical pI: 8.89
Molecular weight: 45271.00
Amino acid composition:

Ala (A)	51	12.4%
Arg (R)	17	4.1%
Asn (N)	7	1.7%
Asp (D)	8	1.9%
Cys (C)	5	1.2%
Gln (Q)	15	3.6%
Glu (E)	14	3.4%
Gly (G)	27	6.6%
His (H)	1	0.2%
Ile (I)	29	7.1%
Leu (L)	60	14.6%
Lys (K)	10	2.4%
Met (M)	10	2.4%
Phe (F)	31	7.5%
Pro (P)	12	2.9%
Ser (S)	26	6.3%
Thr (T)	16	3.9%
Trp (W)	11	2.7%
Tyr (Y)	15	3.6%
Val (V)	46	11.2%

Total number of negatively charged residues (Asp + Glu): 22
Total number of positively charged residues (Arg + Lys): 27
Formula: $C_{2146}H_{3327}N_{507}O_{535}S_{15}$
Total number of atoms: 6530
Ext. coefficient: 82850 M-1 cm-1 (at 280 nm measured in water)
Abs 0.1% (=1 g/l) 1.830
Estimated half-life: The estimated half-life is: 30 hours (in vitro), >10 hours (Escherichia coli, in vivo).

BRIL gene

From TWIST Bioscience:
GCGGATCTGGAAGATAACTGGGAGACGCTGAACGATAATTTGAAAGTAATCGAAAAGGCT
GATAATGCGGCTCAGGTTAAGGACGCGTTAACCAAAATGCGGGCAGCAGCATTAGATGC
CCAAAAAGCCACTCCGCCAAAACTGGAAGATAAAAGCCCTGATAGTCCGGAAATGAAAGA
CTTTCGTCATGGTTTCGACATTCTTGTGGGGCAGATTGATGACGCGCTGAAACTCGCGAAT
GAGGGCAAAGTCAAAGAAGCGCAGGCTGCCGCAGAGCAGTTGAAAACAACCCGCAACG
CCTATATCCAAAAGTACCTG

Table 2 - Unit cell parameters and scaling statistics for the EG-N-Δ8 crystal.

Temperature [K]	100
Wavelength [Å]	1.00
Number of images	1440
Crystal-detector distance [mm]	150
Exposure time [s]	360
Δϕ [°]	0.25
Crystal system	Trigonal
Space group	P 3 2 1
a [Å], b [Å], c [Å]	83.77, 83.77, 130.71
α [°], β [°], γ [°]	90.0, 90.0, 120.0
Matthew's coefficient	2.40
V [Å3]	7.943·105
Resolution / outer shell [Å]	19.89 - 1.30 (1.32 - 1.30)
Rmerge / outer shell	0.074 / 1.251
Rpim / outer shell	0.024 / 0.402
Total number of reflections / outer shell	1385708 / 67322
Number of unique reflections / outer shell	129586 / 6357
I/I(σ) / outer shell	20.0 / 1.9
Correlation coefficient (CC(½)) / outer shell	0.999 / 0.710
Completeness / outer shell	100% / 100%

References

[1] “Antimicrobial resistance.”. Available: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

[2] L. B. Rice, “Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE,” *J. Infect. Dis.*, vol. 197, no. 8, pp. 1079–1081, Apr. 2008, doi: 10.1086/533452.

[3] W. R. Miller and C. A. Arias, “ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics,” *Nat. Rev. Microbiol.*, vol. 22, no. 10, pp. 598–616, Oct. 2024, doi: 10.1038/s41579-024-01054-w.

[4] A. Kumar and H. P. Schweizer, “Bacterial resistance to antibiotics: active efflux and reduced uptake,” *Adv. Drug Deliv. Rev.*, vol. 57, no. 10, pp. 1486–1513, July 2005, doi: 10.1016/j.addr.2005.04.004.

[5] P. A. Lambert, “Cellular impermeability and uptake of biocides and antibiotics in Gram-positive bacteria and mycobacteria,” *J. Appl. Microbiol.*, vol. 92 Suppl, pp. 46S-54S, 2002.

[6] W. R. Miller, J. M. Munita, and C. A. Arias, “Mechanisms of antibiotic resistance in enterococci,” *Expert Rev. Anti Infect. Ther.*, vol. 12, no. 10, pp. 1221–1236, Oct. 2014, doi: 10.1586/14787210.2014.956092.

[7] F. Di Lorenzo, K. A. Duda, R. Lanzetta, A. Silipo, C. De Castro, and A. Molinaro, “A Journey from Structure to Function of Bacterial Lipopolysaccharides,” *Chem. Rev.*, vol. 122, no. 20, pp. 15767–15821, Oct. 2022, doi: 10.1021/acs.chemrev.0c01321.

[8] X. Wang and P. J. Quinn, “Endotoxins: Lipopolysaccharides of Gram-Negative Bacteria,” in *Endotoxins: Structure, Function and Recognition*, X. Wang and P. J. Quinn, Eds., Dordrecht: Springer Netherlands, 2010, pp. 3–25. doi: 10.1007/978-90-481-9078-2_1.

[9] W. Li, F. Separovic, N. M. O’Brien-Simpson, and J. D. Wade, “Chemically modified and conjugated antimicrobial peptides against superbugs,” *Chem. Soc. Rev.*, vol. 50, no. 8, pp. 4932–4973, 2021, doi: 10.1039/D0CS01026J.

[10] E. Vinogradov, M. Cedzynski, A. Ziolkowski, and A. Swierzko, “The structure of the core region of the lipopolysaccharide from Klebsiella pneumoniae O3,” *Eur. J. Biochem.*, vol. 268, no. 6, pp. 1722–1729, 2001, doi: 10.1046/j.1432-1327.2001.02047.x.

[11] Determining glycosyltransferase functional order via lethality due to accumulated O-antigen intermediates, exemplified with Shigella flexneri O-antigen biosynthesis,”

[12] S. M. Huszczynski, J. S. Lam, and C. M. Khursigara, “The Role of Pseudomonas aeruginosa Lipopolysaccharide in Bacterial Pathogenesis and Physiology,” *Pathogens*, vol. 9, no. 1, p. 6, Jan. 2020, doi: 10.3390/pathogens9010006.

[13] R. T. Villavicencio, “The History of Blue Pus,” *J. Am. Coll. Surg.*, vol. 187, no. 2, p. 212, Aug. 1998, doi: 10.1016/S1072-7515(98)00137-9.

[14] B. Schwartz, K. Klamer, J. Zimmerman, P. B. Kale-Pradhan, and A. Bhargava, “Multidrug Resistant Pseudomonas aeruginosa in Clinical Settings: A Review of Resistance Mechanisms and Treatment Strategies,” *Pathogens*, vol. 13, no. 11, p. 975, Nov. 2024, doi: 10.3390/pathogens13110975.

[15] R. T. Sadikot, T. S. Blackwell, J. W. Christman, and A. S. Prince, “Pathogen–Host Interactions in Pseudomonas aeruginosa Pneumonia,” *Am. J. Respir. Crit. Care*

Med., vol. 171, no. 11, pp. 1209–1223, June 2005, doi: 10.1164/rccm.200408-1044SO.
[16] R. B. McFee, "Nosocomial or Hospital-acquired Infections: An Overview," *Dis. Mon.*, vol. 55, no. 7, pp. 422–438, July 2009, doi: 10.1016/j.disamonth.2009.03.014.
[17] S. Labovská, "Pseudomonas aeruginosa as a Cause of Nosocomial Infections," in *Pseudomonas aeruginosa - Biofilm Formation, Infections and Treatments*, IntechOpen, 2021. doi: 10.5772/intechopen.95908.
[18] "Epidemiology of antibiotic resistance in Pseudomonas aeruginosa. Implications for empiric and definitive therapy. | EBSCOhost."
[19] Y. Carmeli *et al.*, "Ceftazidime-avibactam or best available therapy in patients with ceftazidime-resistant Enterobacteriaceae and Pseudomonas aeruginosa complicated urinary tract infections or complicated intra-abdominal infections (REPRISE): a randomised, pathogen-directed, phase 3 study," *Lancet Infect. Dis.*, vol. 16, no. 6, pp. 661–673, June 2016, doi: 10.1016/S1473-3099(16)30004-4.
[20] M. D. Willcox, "Management and treatment of contact lens-related Pseudomonas keratitis," *Clin. Ophthalmol.*, vol. 6, pp. 919–924, June 2012, doi: 10.2147/OPTH.S25168.
[21] Y. Hilliam, S. Kaye, and C. Winstanley, "Pseudomonas aeruginosa and microbial keratitis," *J. Med. Microbiol.*, vol. 69, no. 1, pp. 3–13, 2020, doi: 10.1099/jmm.0.001110.
[22] E. Chiappini, G. Taccetti, and M. de Martino, "Bacterial Lung Infections in Cystic Fibrosis Patients: An Update," *Pediatr. Infect. Dis. J.*, vol. 33, no. 6, p. 653, June 2014, doi: 10.1097/INF.0000000000000347.
[23] P. Allen, J. Borick, and J. Borick, "Acute and Chronic Infection Management in CF," in *Cystic Fibrosis in Primary Care : An Essential Guide to a Complex, Multi-System Disease*, M. Lewis FAAFP, Douglas, Ed., Cham: Springer International Publishing, 2020, pp. 69–87. doi: 10.1007/978-3-030-25909-9_8.
[24] L. R. Mulcahy, V. M. Isabella, and K. Lewis, "Pseudomonas aeruginosa Biofilms in Disease," *Microb. Ecol.*, vol. 68, no. 1, pp. 1–12, July 2014, doi: 10.1007/s00248-013-0297-x.
[25] A. C. Abreu, R. R. Tavares, A. Borges, F. Mergulhão, and M. Simões, "Current and emergent strategies for disinfection of hospital environments," *J. Antimicrob. Chemother.*, vol. 68, no. 12, pp. 2718–2732, Dec. 2013, doi: 10.1093/jac/dkt281.
[26] D. Reynolds and M. Kollef, "The Epidemiology and Pathogenesis and Treatment of Pseudomonas aeruginosa Infections: An Update," *Drugs*, vol. 81, no. 18, pp. 2117–2131, 2021, doi: 10.1007/s40265-021-01635-6.
[27] J. Klockgether *et al.*, "Genome Diversity of Pseudomonas aeruginosa PAO1 Laboratory Strains," *J. Bacteriol.*, vol. 192, no. 4, pp. 1113–1121, Feb. 2010, doi: 10.1128/jb.01515-09.
[28] A. Ascari and R. Morona, "Recent insights into Wzy polymerases and lipopolysaccharide O-antigen biosynthesis," *J. Bacteriol.*, vol. 207, no. 4, pp. e00417-24, Mar. 2025, doi: 10.1128/jb.00417-24.
[29] C. K. Raymond *et al.*, "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas aeruginosa," *J. Bacteriol.*, vol. 184, no. 13, pp. 3614–3622, July 2002, doi: 10.1128/jb.184.13.3614-3622.2002.
[30] C. Whitfield, "Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli," *Annu. Rev. Biochem.*, vol. 75, no. Volume 75, 2006, pp. 39–68,

July 2006, doi: 10.1146/annurev.biochem.75.103004.142545.
[31] M. Weckener *et al.*, “The lipid linked oligosaccharide polymerase Wzy and its regulating co-polymerase, Wzz, from enterobacterial common antigen biosynthesis form a complex,” *Open Biol.*, vol. 13, no. 3, p. 220373, Mar. 2023, doi: 10.1098/rsob.220373.
[32] S. T. Islam and J. S. Lam, “Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria,” *Environ. Microbiol.*, vol. 15, no. 4, pp. 1001–1015, 2013, doi: 10.1111/j.1462-2920.2012.02890.x.
[33] F. A. Ganaie, M. B. Oliver, J. S. Saad, D. G. Glanville, A. T. Ulijasz, and M. H. Nahm, “Wzy 3D structural models correlate with inter-repeat unit glycosidic bond configuration in pneumococcal capsule polysaccharides,” *Microbiol. Spectr.*, vol. 13, no. 10, pp. e00328-25, Sept. 2025, doi: 10.1128/spectrum.00328-25.
[34] S. T. Islam and J. S. Lam, “Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway,” *Can. J. Microbiol.*, vol. 60, no. 11, pp. 697–716, Nov. 2014, doi: 10.1139/cjm-2014-0595.
[35] X. Ruan, D. E. Loyola, C. L. Marolda, J. M. Perez-Donoso, and M. A. Valvano, “The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases*,” *Glycobiology*, vol. 22, no. 2, pp. 288–299, Feb. 2012, doi: 10.1093/glycob/cwr150.
[36] A. Polissi and P. Sperandeo, “The Lipopolysaccharide Export Pathway in Escherichia coli: Structure, Organization and Regulated Assembly of the Lpt Machinery,” *Mar. Drugs*, vol. 12, no. 2, pp. 1023–1042, Feb. 2014, doi: 10.3390/md12021023.
[37] L. L. Burrows and J. S. Lam, “Effect of wzx (rfbX) Mutations on A-Band and B-Band Lipopolysaccharide Biosynthesis in Pseudomonas aeruginosa O5,” *J. Bacteriol.*, vol. 181, no. 3, pp. 973–980, Feb. 1999, doi: 10.1128/jb.181.3.973-980.1999.
[38] D. Liu, R. A. Cole, and P. R. Reeves, “An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase,” *J. Bacteriol.*, vol. 178, no. 7, pp. 2102–2107, Apr. 1996, doi: 10.1128/jb.178.7.2102-2107.1996.
[39] “Progress in Our Understanding of Wzx Flippase for Translocation of Bacterial Membrane Lipid-Linked Oligosaccharide.”
[40] “Review: Conserved and variable structural features in the lipopolysaccharide of Pseudomonas aeruginosa.”
[41] S. T. Islam, V. L. Taylor, M. Qi, and J. S. Lam, “Membrane Topology Mapping of the O-Antigen Flippase (Wzx), Polymerase (Wzy), and Ligase (WaaL) from Pseudomonas aeruginosa PAO1 Reveals Novel Domain Architectures,” *mBio*, vol. 1, no. 3, p. 10.1128/mbio.00189-10, Aug. 2010, doi: 10.1128/mbio.00189-10.
[42] Y. Hong and P. R. Reeves, “Diversity of O-Antigen Repeat Unit Structures Can Account for the Substantial Sequence Variation of Wzx Translocases,” *J. Bacteriol.*, vol. 196, no. 9, pp. 1713–1722, Apr. 2014, doi: 10.1128/jb.01323-13.
[43] S. T. Islam *et al.*, “Proton-Dependent Gating and Proton Uptake by Wzx Support O-Antigen-Subunit Antiport Across the Bacterial Inner Membrane,” *mBio*, vol. 4, no. 5, p. 10.1128/mbio.00678-13, Sept. 2013, doi: 10.1128/mbio.00678-13.
[44] “Structure of WzxE the lipid III flippase for Enterobacterial Common Antigen polysaccharide.”
[45] F. Sievers *et al.*, “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” *Mol. Syst. Biol.*, vol. 7, p. 539, Oct.

2011, doi: 10.1038/msb.2011.75.
[46] C. P. Kaur, J. Vadivelu, and S. Chandramathi, “Impact of Klebsiella pneumoniae in lower gastrointestinal tract diseases,” *J. Dig. Dis.*, vol. 19, no. 5, pp. 262–271, 2018, doi: 10.1111/1751-2980.12595.
[47] “Hypervirulent Klebsiella pneumoniae | Clinical Microbiology Reviews.” A https://journals.asm.org/doi/full/10.1128/cmr.00001-19
[48] S. Gao *et al.*, “Bacterial capsules: Occurrence, mechanism, and function,” *Npj Biofilms Microbiomes*, vol. 10, no. 1, p. 21, Mar. 2024, doi: 10.1038/s41522-024-00497-6.
[49] S. Gao *et al.*, “Bacterial capsules: Occurrence, mechanism, and function,” *Npj Biofilms Microbiomes*, vol. 10, no. 1, p. 21, Mar. 2024, doi: 10.1038/s41522-024-00497-6.
[50] O. Rendueles, “Deciphering the role of the capsule of Klebsiella pneumoniae during pathogenesis: A cautionary tale,” *Mol. Microbiol.*, vol. 113, no. 5, pp. 883–888, 2020, doi: 10.1111/mmi.14474.
[51] S. Sharma, J. Mohler, S. D. Mahajan, S. A. Schwartz, L. Bruggemann, and R. Aalinkeel, “Microbial Biofilm: A Review on Formation, Infection, Antibiotic Resistance, Control Measures, and Innovative Treatment,” *Microorganisms*, vol. 11, no. 6, p. 1614, June 2023, doi: 10.3390/microorganisms11061614.
[52] S. Sharma *et al.*, “Bacteriophages and its applications: an overview,” *Folia Microbiol. (Praha)*, vol. 62, no. 1, pp. 17–55, Jan. 2017, doi: 10.1007/s12223-016-0471-x.
[53] F. Zhang and W. Cheng, “The Mechanism of Bacterial Resistance and Potential Bacteriostatic Strategies,” *Antibiotics*, vol. 11, no. 9, p. 1215, Sept. 2022, doi: 10.3390/antibiotics11091215.
[54] D. Romero-Calle, R. Guimarães Benevides, A. Góes-Neto, and C. Billington, “Bacteriophages as Alternatives to Antibiotics in Clinical Care,” *Antibiotics*, vol. 8, no. 3, p. 138, Sept. 2019, doi: 10.3390/antibiotics8030138.
[55] S. Sharma *et al.*, “Bacteriophages and its applications: an overview,” *Folia Microbiol. (Praha)*, vol. 62, no. 1, pp. 17–55, Jan. 2017, doi: 10.1007/s12223-016-0471-x.
[56] S. McCallin, J. C. Sacher, J. Zheng, and B. K. Chan, “Current State of Compassionate Phage Therapy,” *Viruses*, vol. 11, no. 4, p. 343, Apr. 2019, doi: 10.3390/v11040343.
[57] B. M. Zaki, A. H. Hussein, T. A. Hakim, M. S. Fayez, and A. El-Shibiny, “Phages for treatment of *Klebsiella pneumoniae* infections,” in *Progress in Molecular Biology and Translational Science*, vol. 200, V. Singh, Ed., in Phage Therapy - Part A, vol. 200. , Academic Press, 2023, pp. 207–239. doi: 10.1016/bs.pmbts.2023.03.007.
[58] D. M. Lin, B. Koskella, and H. C. Lin, “Phage therapy: An alternative to antibiotics in the age of multi-drug resistance,” *World J. Gastrointest. Pharmacol. Ther.*, vol. 8, no. 3, pp. 162–173, Aug. 2017, doi: 10.4292/wjgpt.v8.i3.162.
[59] J. Klumpp, M. Dunne, and M. J. Loessner, “A perfect fit: Bacteriophage receptor-binding proteins for diagnostic and therapeutic applications,” *Curr. Opin. Microbiol.*, vol. 71, p. 102240, Feb. 2023, doi: 10.1016/j.mib.2022.102240.
[60] K. M. Danis-Wlodarczyk, D. J. Wozniak, and S. T. Abedon, “Treating Bacterial Infections with Bacteriophage-Based Enzybiotics: In Vitro, In Vivo and Clinical Application,” *Antibiotics*, vol. 10, no. 12, p. 1497, Dec. 2021, doi: 10.3390/antibiotics10121497.

[61] M. M. D'Andrea *et al.*, "ϕBO1E, a newly discovered lytic bacteriophage targeting carbapenemase-producing Klebsiella pneumoniae of the pandemic Clonal Group 258 clade II lineage," *Sci. Rep.*, vol. 7, no. 1, p. 2614, June 2017, doi: 10.1038/s41598-017-02788-9.
[62] B. Bellich, C. Lagatolla, R. Rizzo, M. M. D'Andrea, G. M. Rossolini, and P. Cescutti, "Determination of the capsular polysaccharide structure of the *Klebsiella pneumoniae* ST512 representative strain KPB-1 and assignments of the glycosyltransferases functions," *Int. J. Biol. Macromol.*, vol. 155, pp. 315–323, July 2020, doi: 10.1016/j.ijbiomac.2020.03.196.
[63] T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 271–281, Apr. 2011, doi: 10.1107/S0907444910048675.
[64] L. Potterton *et al.*, "CCP4i2: The new graphical user interface to the CCP4 proGram suite," *Acta Crystallogr. Sect. Struct. Biol.*, vol. 74, no. 2, pp. 68–84, Feb. 2018, doi: 10.1107/S2059798317016035.
[65] P. R. Evans and G. N. Murshudov, "How good are my data and what is the resolution?," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 69, no. 7, pp. 1204–1214, July 2013, doi: 10.1107/S0907444913000061.
[66] A. Vagin and A. Teplyakov, "Molecular replacement with MOLREP," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, no. 1, pp. 22–25, Jan. 2010, doi: 10.1107/S0907444909042589.
[67] "(IUCr) ModelCraft: an advanced automated model-building pipeline using Buccaneer." https://journals.iucr.org/d/issues/2022/09/00/qo5001/index.html
[68] "(IUCr) REFMAC5 for the refinement of macromolecular crystal structures." https://journals.iucr.org/d/issues/2011/04/00/ba5152/index.html
[69] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, "Features and development of Coot," *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 486–501, Apr. 2010, doi: 10.1107/S0907444910007493.
[70] "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination | Nature Methods." https://www.nature.com/articles/nmeth.4169
[71] T. Bepler, K. Kelley, A. J. Noble, and B. Berger, "Topaz-Denoise: general deep denoising models for cryoEM and cryoET," *Nat. Commun.*, vol. 11, no. 1, p. 5208, Oct. 2020, doi: 10.1038/s41467-020-18952-1.
[72] "Molecular Dynamics - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/chemistry/molecular-dynamics
[73] K. Lindorff-Larsen *et al.*, "Improved side-chain torsion potentials for the Amber ff99SB protein force field," *Proteins*, vol. 78, no. 8, pp. 1950–1958, June 2010, doi: 10.1002/prot.22711.
[74] S. Páll *et al.*, "Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS," *J. Chem. Phys.*, vol. 153, no. 13, p. 134110, Oct. 2020, doi: 10.1063/5.0018516.
[75] M. Turisini, M. Cestari, and G. Amati, "LEONARDO: A Pan-European Pre-Exascale Supercomputer for HPC and AI applications," *J. Large-Scale Res. Facil. JLSRF*, vol. 9, no. 1, Jan. 2024, doi: 10.17815/jlsrf-8-186.
[76] D. R. Roe and B. R. Brooks, "A protocol for preparing explicitly solvated systems for stable molecular dynamics simulations," *J. Chem. Phys.*, vol. 153, no. 5, p. 054123,

Aug. 2020, doi: 10.1063/5.0013849.
[77] E. Guercia *et al.*, “On the Cholesterol Raising Effect of Coffee Diterpenes Cafestol and 16-O-Methylcafestol: Interaction with Farnesoid X Receptor,” *Int. J. Mol. Sci.*, vol. 25, no. 11, p. 6096, May 2024, doi: 10.3390/ijms25116096.
[78] G. Miolo *et al.*, “Identification and Structural Characterization of a Novel COL3A1 Gene Duplication in a Family With Vascular Ehlers-Danlos Syndrome,” *Mol. Genet. Genomic Med.*, vol. 13, no. 4, p. e70095, Apr. 2025, doi: 10.1002/mgg3.70095.
[79] T. C. Terwilliger *et al.*, “AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination,” *Nat. Methods*, vol. 21, no. 1, pp. 110–116, Jan. 2024, doi: 10.1038/s41592-023-02087-4.
[80] Z. Yang, X. Zeng, Y. Zhao, and R. Chen, “AlphaFold2 and its applications in the fields of biology and medicine,” *Signal Transduct. Target. Ther.*, vol. 8, no. 1, p. 115, Mar. 2023, doi: 10.1038/s41392-023-01381-z.
[81] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.
[82] J. Abramson *et al.*, “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” *Nature*, vol. 630, no. 8016, pp. 493–500, June 2024, doi: 10.1038/s41586-024-07487-w.
[83] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, “ColabFold: making protein folding accessible to all,” *Nat. Methods*, vol. 19, no. 6, pp. 679–682, June 2022, doi: 10.1038/s41592-022-01488-1.
[84] G. Kim *et al.*, “Easy and accurate protein structure prediction using ColabFold,” *Nat. Protoc.*, vol. 20, no. 3, pp. 620–642, Mar. 2025, doi: 10.1038/s41596-024-01060-5.
[85] M. Turisini, M. Cestari, and G. Amati, “LEONARDO: A Pan-European Pre-Exascale Supercomputer for HPC and AI applications,” *J. Large-Scale Res. Facil. JLSRF*, vol. 9, no. 1, Jan. 2024, doi: 10.17815/jlsrf-8-186.
[86] S. Liu and W. Li, “Protein Fusion Strategies for Membrane Protein Stabilization and Crystal Structure Determination,” *Crystals*, vol. 12, no. 8, p. 1041, Aug. 2022, doi: 10.3390/cryst12081041.
[87] R. Chu *et al.*, “Redesign of a Four-helix Bundle Protein by Phage Display Coupled with Proteolysis and Structural Characterization by NMR and X-ray Crystallography,” *J. Mol. Biol.*, vol. 323, no. 2, pp. 253–262, Oct. 2002, doi: 10.1016/S0022-2836(02)00884-7.
[88] M. Forloni, A. Y. Liu, and N. Wajapeyee, “Megaprimer Polymerase Chain Reaction (PCR)-Based Mutagenesis,” *Cold Spring Harb. Protoc.*, vol. 2019, no. 6, June 2019, doi: 10.1101/pdb.prot097824.
[89] S. Wagner *et al.*, “Tuning Escherichia coli for membrane protein overexpression,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 38, pp. 14371–14376, Sept. 2008, doi: 10.1073/pnas.0804090105.
[90] H. J. Lee, H. S. Lee, T. Youn, B. Byrne, and P. S. Chae, “Impact of novel detergents on membrane protein studies,” *Chem*, vol. 8, no. 4, pp. 980–1013, Apr. 2022, doi: 10.1016/j.chempr.2022.02.007.
[91] N. J. Greenfield, “Using circular dichroism spectra to estimate protein secondary structure,” *Nat. Protoc.*, vol. 1, no. 6, pp. 2876–2890, 2006, doi: 10.1038/nprot.2006.202.

[92] R. Hussain and G. Siligardi, “Characterisation of Conformational and Ligand Binding Properties of Membrane Proteins Using Synchrotron Radiation Circular Dichroism (SRCD),” in *The Next Generation in Membrane Protein Structure Determination*, I. Moraes, Ed., Cham: Springer International Publishing, 2016, pp. 43–59. doi: 10.1007/978-3-319-35072-1_4.
[93] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, and A. L. Lomize, “OPM database and PPM web server: resources for positioning of proteins in membranes,” *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D370–D376, Jan. 2012, doi: 10.1093/nar/gkr703.
[94] S. Jo, T. Kim, V. G. Iyer, and W. Im, “CHARMM-GUI: A web-based graphical user interface for CHARMM,” *J. Comput. Chem.*, vol. 29, no. 11, pp. 1859–1865, 2008, doi: 10.1002/jcc.20945.
[95] J. Abramson *et al.*, “The structure of the ubiquinol oxidase from Escherichia coli and its ubiquinone binding site,” *Nat. Struct. Biol.*, vol. 7, no. 10, pp. 910–917, Oct. 2000, doi: 10.1038/82824.
[96] N. Ciacci *et al.*, “Characterization of vB_Kpn_F48, a Newly Discovered Lytic Bacteriophage for Klebsiella pneumoniae of Sequence Type 101,” *Viruses*, vol. 10, no. 9, p. 482, Sept. 2018, doi: 10.3390/v10090482.
[97] B. Bellich, C. Lagatolla, R. Rizzo, M. M. D’Andrea, G. M. Rossolini, and P. Cescutti, “Determination of the capsular polysaccharide structure of the Klebsiella pneumoniae ST512 representative strain KPB-1 and assignments of the glycosyltransferases functions,” *Int. J. Biol. Macromol.*, vol. 155, pp. 315–323, July 2020, doi: 10.1016/j.ijbiomac.2020.03.196.
[98] N. K. Broeker *et al.*, “Single amino acid exchange in bacteriophage HK620 tailspike protein results in thousand-fold increase of its oligosaccharide affinity,” *Glycobiology*, vol. 23, no. 1, pp. 59–68, Jan. 2013, doi: 10.1093/glycob/cws126.