Cryogenic Electron Microscopy enables Drug Research for Membrane Protein Targets

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel

von

Dongchun Ni

Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Henning Stahlberg

Prof. Dr. Camilo Perez

Prof. Dr. Andrea Ablasser

Basel, November 17, 2020

Prof. Dr. Martin Spiess

The Dean of Faculty

Contents

1 Chapter I: Introduction

Proteins are the molecules of life, performing the essential biological functions for all kinds of organisms on the earth. Membrane proteins are a common class of proteins that are integrated or anchored with cell membranes. More than half of drug targets are membrane proteins (Overington et al., 2006). The study of protein function is an essential step and the basis for the novel drug discovery. Obtaining three-dimensional atomic models of proteins is an important approach for capturing the essence of understanding of the basis of multiple biological processes. Nowadays, cryo-electron microscopy (cryo-EM) single-particle analysis (SPA) is widely used for high-resolution structure determination of both transmembrane and water-soluble proteins.

1.1 Protein structure

Structure biology is a field that combines a variety of approaches, including molecular biology, protein chemistry and biophysics. It is the basic knowledge of protein architecture, which is essential for the understanding of protein function in detail (Moore, 2017). Amino acids are the basic components of proteins. They are organic compounds containing amine (-NH2) and carboxyl (-COOH) functional groups, and a variable side chain specifically attached to each amino acid (Butterworth, 2005). Proteins are polypeptides mainly consisting of up to several thousands of L-type amino acids. D-type amino acids are not common in natural proteins, but some of specific bacterial polypeptides, such as gramicidin and valinomycin, are made by a mixture of D- and L-amino acids (Ketchem et al., 1993) (Rose and Henkens, 1974). Sometimes D-type amino acids can be also found in some special disease conditions, like aging or cancers (Bastings et al., 2019; Fujii et al., 2018).

The sequence of a polypeptide is considered to be the primary structural information of a protein molecule, which can be converted directly from the sequences of cDNA or mRNA. The sequence information can be used to determine various physical and chemical parameters such as the theoretical PI and molecular weight. In principle, the secondary structure can be predicted precisely based on the primary sequence, especially for the transmembrane (TM) alpha-helices (Chen et al., 2002). In addition, several types of functional motifs can be recognized in the polypeptide sequence, such as the protein post-translational modification sites, localization sequences (leader peptides) and protease cleavage sites (Wong et al., 2015).

1.1.1 Protein secondary structure

Protein secondary structure is the hydrogen-bond dependent primary level of three-dimensional atom arrangement in the local polypeptide chain. Most common secondary structure motifs include beta-sheets (or beta-strands) and alpha-helixes. The secondary structure can be defined according to the distribution of hydrogen-bonds between carbonyl oxygen atoms and amine hydrogen atoms or simply based on the Torsion Angles of the backbone peptide chain. As mentioned above, the performance of predicting the protein secondary structure is relatively accurate (roughly 80% of accuracy) (Petersen et al., 2000), in particular for the prediction of TM helix topology (Chen et al., 2002).

1.1.2 Protein tertiary and quaternary structure

With multiple secondary structural elements, the polypeptides fold into specific three-dimensional domains, called protein tertiary structures, which are the basis for conducting specific biological functions inside and outside of the cells. The accuracy of protein tertiary structure predictions is normally much lower than in the case of secondary structure. This is the case because the protein chemical environment and the interaction networks between multiple domains are extremely complicated compared to a single backbone. The tertiary structures are represented as the 3D models of protein and the coordinates contain the spatial information of each amino acid and every atom. Protein complexes or multi-domain proteins containing more than one structural unit, are constituting the protein quaternary structure (Copeland, 1994). The protein-protein interaction is a common phenomenon and strategy in terms of involvement in many cellular pathways, such as the GPCRs signaling (Hilger et al., 2018). In some of specific cases, the protein molecules can form oligomers or higher-order structures, which can facilitate some special functions as well as the progression of diseases, such as Parkinson's disease (Wu and Fuxreiter, 2016).

1.1.3 Protein structure determination

Several methods have been developed for the determination of the accurate 3D atomic structures of protein molecules. These include X-ray crystallography, NMR and cryo-EM SPA. Each method has its own applications, advantages and the limitations. Since the recent improvements in cryo-EM SPA, it has become more and more popular and comparably efficient in the field of structural biology. Meanwhile, the current method development of cryo-EM is developing forward in the direction of high-resolution structure determination

at cellular level (in-situ). To date, the best resolution structure of a membrane protein using the in-situ cryo-electron tomography (cryo-ET) was achieved by Mikhail Kudryashev`s group (Sanchez et al., 2020). The resolution of 7Å was sufficient to visualize the transmembrane alpha-helices directly in the physiological cell membrane (Sanchez et al., 2020). It is also predicted that the high-resolution in-situ structure determination will be largely advanced in the upcoming years.

1.1.4 Protein data bank

The Protein Data Bank (PDB) is a resource of three-dimensional protein structures in the public domain, containing information regarding single proteins or protein complexes, such as sequence, the secondary structure, the 3D coordinates and others (Burley et al., 2017). The protein data bank has a very important role in the field of structural biology. Structural biologists are usually required to make a PDB deposition during manuscript submission or reviewing. For the time being, there are around 1142 unique known protein structures (3084 pdb files in total) of membrane proteins (https://blanco.biomol.uci.edu/mpstruc/) in the PDB since the first membrane protein structure has been revealed in 1985 (Newport et al., 2019).

1.2 Membrane proteins and their functions

Cellular membrane systems, which include the cell membrane and intracellular membranes, are the envelopes for living cells and cellular organelles. They consist of a lipid bilayer composed of various lipid molecules. Separation of cytoplasmic components and the reactions occurring outside the cell and inside of the organelles is necessary for the basic spatio-temporal independence of multiple cellular processes (Alberts et al., 2002; Hedin et al., 2011). Membrane proteins are generally integrated or associated with cellular membrane systems, which function as a macromolecule machine mediating a large number of biological events. These include ion flow, small molecule transport, biopolymer transport, signal transduction and many others. Transmembrane proteins can be divided into three major classes: single-pass transmembrane protein, alpha-helical and beta-barrel membrane proteins (Figure 1) (Hiller et al., 2008; Okada et al., 2002; Yan et al., 2020), physiologically facilitates most biochemical reactions and mediates the cellular pathway in multiple-levels (Vinothkumar and Henderson, 2010).

1.2.1 Single-pass transmembrane proteins

The single-pass transmembrane proteins are a large class of membrane proteins that contain only one transmembrane helix anchoring them to the cell membrane system. Examples include the human ACE2 (Figure 1) or the beta-coronavirus spike protein (Yan et al., 2020). The single transmembrane helix is functionally relevant for protein localization and membrane association, but sometimes it also has some special biological functions, such as the cell signaling (Bugge et al., 2016). Misfolding of single-pass transmembrane proteins may sometimes cause diseases. For example, the Amyloid precursor protein (APP) is an unknown functional membrane-bound protein containing a predicted single transmembrane helix nearby its C-terminus. The soluble domain can be cleaved by a membrane enzyme called gamma secretase, which leads to the massive production of amyloid-beta peptides (A-Beta) and induces the amyloid fibrils formation. The cleaved A-Beta deposited in the brain tissue is supposed to be the pathologic basis for Alzheimer's disease (Tabaton and Tamagno, 2007).

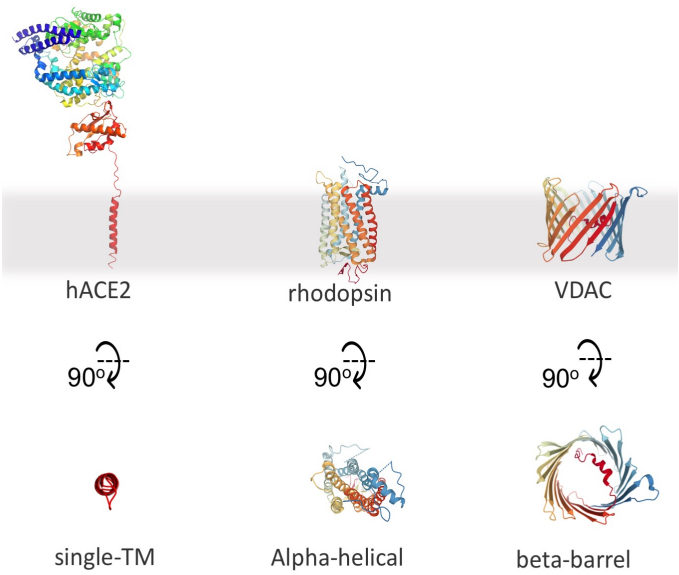


Figure 1 Cartoon representation of the structures of three major classes of transmembrane proteins. Single-pass transmembrane protein, human ACE2 (left); alpha-helical transmembrane protein, bovine rhodopsin (middle); beta-barrel membrane protein, human VDAC (right).

1.2.2 Alpha-helical membrane proteins

The proteins containing two or more transmembrane alpha-helices are generally folded into monomeric or oligomeric functional units and are widely distributed in the cells. Multiple transmembrane-helical proteins can be found in almost all kinds of cell membrane systems and account for the majority of membrane protein drug targets. Examples include GPCRs, ABC transporters, cell receptors and membrane enzymes (Cross et al., 2013; Overington et al., 2006). Structures of alpha-helical transmembrane proteins have been widely investigated since decades using different approaches. The Nobel Prize in chemistry in 1988

was awarded for the structure determination of the light-sensing complex in (https://www.nobelprize.org/prizes/chemistry/1988/summary/), which was the first time to determinate a high-resolution structure for alpha-helical membrane protein (Deisenhofer and Michel, 1991). X-ray crystallography has been the most common approach of determining high-resolution structures of proteins for the past 20 years. Nowadays, however, cryo-EM single particle analysis is taking over its position.

1.2.3 Beta-barrel membrane protein

Beta-barrel is a description of a particular protein tertiary structure, exemplified by the fold of green fluorescent protein (GFP) (Yang et al., 1996). Meanwhile, it is also a specific tertiary structural unit for beta-barrel transmembrane proteins. Hydrogen bonds can mediate the formation of beta-barrel architecture for both water-soluble and transmembrane proteins. The water-soluble beta-barrel proteins are composed of several antiparallel beta-sheets and folded into a closed barrel-sharped structure. Ideally, residues in the central cavity are hydrophobic, which facilitates the binding of hydrophobic ligands, such as streptavidin (Le Trong et al., 2011) and retinol (Newcomer et al., 1984). In contrast to the water-soluble forms, the transmembrane beta-barrel proteins are quite unique. The amino acids in the lumen of the beta-barrel are hydrophilic while the hydrophobic residues are exposed to the membrane lipids interface (Schulz, 2000).

1.2.3.1 Beta-barrel membrane protein structure and biochemistry

The architecture of beta-barrel membrane proteins is generally conserved, with the exception of some of the mitochondrial outer membrane proteins such as the VDAC and TOM40. Bacterial outer membrane proteins are composed of an even number of beta-sheets in the range from 8 to 60, with one or multiple polypeptide chains involved. Exceptionally, mitochondrial proteins VDAC (Figure 1) (Hiller et al., 2008) and TOM40 (Araiso et al., 2019) contain 19 beta-strands. The extracellular facing loops are usually larger than the intracellular loops and form extracellular cap, which is believed to be necessary for maintaining of cellular permeabilization in gram-negative bacteria (Nikaido, 2003). The structures of beta-barrel membrane proteins always exhibit a right-handed helical structure due to the chirality of their amino acids (Schulz, 2002).

Beta-barrel membrane proteins are very thermally stable compared to alpha-helical membrane proteins. A unique characteristic called “heat modifiable gel-shifts” of beta-barrel membrane protein is generally used as a useful tool to characterize the specific folding

behavior of beta-barrel membrane domains (Noinaj et al., 2015). Some of the beta-barrel membrane protein can be refolded in the solution of detergents or lipid environment by dialysis or rapid dilution, as in the case of BamA and VDAC (Hiller et al., 2008; Ni et al., 2014).

1.2.3.2 Lipid-induced beta-barrel membrane proteins

The beta-barrel membrane proteins are generally found in outer membrane systems, such as the Mitochondria, Chloroplast and the Gram-negative bacterial outer membrane as mentioned above. For this reason, they are also named as outer membrane proteins (OMPs). In addition, there is a special group called lipid-induced pore-forming proteins (PFPs). PFPs include pore forming toxins (PFTs), which are produced either in bacteria or human cells, are mainly associated with membrane permeabilization (Andreeva-Kovalevskaya et al., 2008; Dal Peraro and van der Goot, 2016). PFPs form homo-oligomers after the addition of lipid-like molecules in vitro or by coming into contact with cell membranes. Particular membrane-associated sequences from PFPs are essential for the conformational state transition from a pre-pore state (soluble form) to a pore-forming state (membrane-bound form) (Panchal et al., 2002).

Bacterial PFTs usually function as virulence factors that create holes on cell surface, leading to the uncontrollable leakage of cellular components and further killing of host cells (Dal Peraro and van der Goot, 2016). In contrast to the bacterial PFTs, the PFPs produced by human cells are usually involved in the immune response and inflammation. For example, the complement membrane attack complex (MAC) is an important protein machine that regulated by the innate immune system, which potentially forms unregulated transmembrane channels on the surface of pathogens, resulting in their death (Menny et al., 2018; Moreno-Hagelsieb et al., 2017). Another interesting example of a human PFP is Gasdermin D, which causes pyroptotic cell death upon the detection of foreign DNA in the cytosol (Feltham and Vince, 2018; Liu et al., 2016). Additionally, some studies have shown the potential pore-forming capacity of amyloid proteins, such as the amyloid beta (A-Beta) and the alpha-synuclein (Serra-Batiste et al., 2016; Tosatto et al., 2012). Amyloid proteins are believed to be organized as oligomeric ion channels in multiple cellular membrane systems, potentially causing a variety of neurodegenerative diseases as a result of the homeostatic dysfunction of cellular components.

1.3 Structural insights into outer membrane biogenesis.

1.3.1 Outer membrane protein (beta-barrel membrane protein) biogenesis

Gram-negative bacteria have two membrane bilayers surrounding their cytosol: the inner membrane (IM) and outer membrane (OM). As mentioned above, IM proteins belong to the alpha-helical or single-pass classes of transmembrane proteins. The heterotrimeric SecYEG complex is responsible for the insertion and assembly of muti-pass alpha-helical transmembrane proteins. In cooperation with Yidc, the SecYEG-Yidc super-complex catalyzes the preproteins insertion for single-pass transmembrane proteins as well. In addition, the SecYEG complex is the key molecular machinery for the translocation of the nascent polypeptides of OM proteins, which directly associates with the membrane bound ribosomes and incorporates the translation of mRNA (Tsirigotaki et al., 2017). At the outer membrane, the beta-barrel assembly machinery (BAM complex) is responsible for the catalyzing of insertion, folding and assembly of β-barrel membrane proteins. The BAM complex consists of five protein subunits: BamA, BamB, BamC, BamD and BamE (Figure 2) (Konovalova et al., 2017). BamB-E are the lipoprotein components, attached to the inner leaflet of bacterial outer membrane with a lipid-anchor group. The lipoprotein subunits are supposed to be essential for the substrate recruitment and the functional regulation of BAM complex when the beta-barrel protein assembly is occurring. BamA is the membrane-integrated omp85 family protein that containins a transmembrane beta-barrel domain and five polypeptide-transport-associated (POTRA) domains. BamA-dependent outer membrane protein folding is a fascinating biological process in gram-negative bacteria. The fundamental role of BamA is to provide the beta-barrel folding platform by deploying a unique laterally opened gating mechanism to promote the insertion of nascent OMP peptides and the formation of partially or completely folded beta-barrels (Iadanza et al., 2016). The structures of the BAM complex have been resolved by both cryo-EM or by x-ray crystallography. The five protein subunits are organized in a molar ratio of 1:1:1:1:1 (BamA:B:C:D:E) and BamA is located in center of the complex, facing toward the extracellular side. A recent study has described a cryo-EM structure of BamABCDE in complex with a truncated BamA and a partially folded beta-barrel substrate, revealing that BamA is laterally opened and trapped with an uncompleted substrate partially inserted into the OM (Tomasek et al., 2020). However, there are still many aspects remaining unclear. For example, how the beta-barrel directionality is determined and how the extracellular or intracellular domains are folded. These are the mysteries in this specific field and definitely require further investigations.

Interestingly, there are counterparts of the Bam complex in mitochondria (the Sorting and Assembly Machinery or SAM complex) and chloroplasts (Translocon on the outer chloroplast membrane or TOC complex). The composition of the beta-barrel assembly complexes varies slightly between different systems. In chloroplasts, for example, the BamA counterpart Toc75 potentially forms a heterotrimeric complex with two other functional-regulatory subunits Toc 34/33 and Toc 159 (Andrès et al., 2010; Sommer et al., 2011). Similarly, the mitochondrial SAM complex core comprises of three subunits: Sam 50, Sam 35 and Sam37 (Diederichs et al., 2020). A recent study has reported a high-resolution cryo-EM structure of lipid nanodisc-reconstituted SAM complex, which was purified from Myceliophthora thermophila. Interestingly, there is no structural or sequence similarity between these additional subunits and any components of the BAM complex. Structurally, the Sam50 beta-barrel domain resembles the bacterial homolog BamA, exhibiting the typical omp85 fold. The preproteins of VDAC and TOM40 are actually substrates of the SAM complex (Diederichs et al., 2020).

1.3.2 Lipopolysaccharides (LPS)

The outer membrane of Gram-negative bacteria is a conserved asymmetrical bilayer, composed of phospholipids and lipopolysaccharides (LPS). LPS molecules are specifically distributed in the outer leaflet and provide protection necessary for bacterial survival. Lipopolysaccharides are large sugar molecules with a varying molecule weight in the range from 50 to 100 kilodalton (Jann et al., 1975). In general, LPS molecules are composed of three distinct parts: Lipid A, core oligosaccharides and O-antigen. In some Gram-negative pathogens, such as Neisseria gonorrhoeae and Haemophilus, the O-antigen is missing (Moran et al., 1996). LPS is also called endotoxin, as it causes illnesses that result from a severe immune response. The pathogenesis mechanism of LPS-related diseases has been shown to be mainly mediated by the activation of Toll-like receptors (TLRs), which leads to a strong activation of the innate immune response (Poltorak et al., 1998). These discoveries were awarded the Nobel Prize in Physiology or Medicine in 2011 (https://www.nobelprize.org/prizes/medicine/2011/press-release/).

1.3.3 Structural basis of lipopolysaccharide transporting across the Gram-negative cell envelope.

The assembly of outer membrane lipopolysaccharide is a relatively complicated biological process that requires transport across both membranes of Gram-negative bacteria. As

described previously, the lipopolysaccharide is usually composed of three parts: Lipid A, core oligosaccharides and O-antigen. Multiple molecules are involved in the transport of LPS precursors and mature LPS. The whole sequence, which starts at the inner leaflet of the IM and ends at the outer leaflet of the OM, includes the following components: MsbA, WzM/T, LptB2FG complex, LptC, LptA and LptD/E. MsbA is a typical homo-dimeric ABC exporter, which functions as an exporter of the core-oligosaccharide lipid A. It performs the flipping of substrate molecules from the cytosolic to the periplasmic side of IM (Greenfield and Whitfield, 2012; Whitfield and Trent, 2014). The first cryo-EM structure of substrate-bound MsbA has been resolved by single particle cryo-EM analysis in 2017 (Mi et al., 2017), revealing a structure in an inward-facing state with the core-oligosaccharide lipid A molecule captured in the central cavity. The O-antigen chain is also synthesized in the cytosolic side of the IM as a lipid-anchored intermediate that is then transported to the periplasmic space by WzM/T. The high-resolution crystal structure of WzM/T, which has been reported by Jochen Zimmer et al in 2018, revealed a hetero-tetrameric ABC transporter structure organized in a molar ratio of 2:2 (WzM:WzT). It is still unclear how the O-antigen interacts with the transporter, as there was no density for the O-antigen in the published crystal structure (Bi et al., 2018).

Once in the periplasm, the O-antigen is ligated to the core-oligosaccharide lipid A by the protein WaaL (Greenfield and Whitfield, 2012).

Once the complete LPS molecule is formed on the outer leaflet of IM, it needs to be extracted and further loaded into the LPS transporting bridge. The LptB2FG complex is responsible for the LPS extrusion and the process is coupled with ATP hydrolysis. Again, according to the crystal structure as well as cryo-EM data, a particular structural arrangement of a hetero-tetrameric ABC transporter has been revealed. The complex is organized in a molar ratio of 2:1:1 (LptB:LptF:LptG) (Li et al., 2019; Luo et al., 2017). In addition, high-resolution cryo-EM structures also provided the structural basis of substrate interaction, where the trapped LPS molecule was bound in the periplasmic facing cavity of the transporter (Li et al., 2019). The LptC protein is a molecular adaptor for the connection between LptB2FG and the periplasmic LPS transporting bridge. LptA functions as a LPS transporting bridge by deploying a continuous molecular stream of LPS according to their concentration-gradient from the IM to the bacterial surface (Li et al., 2019). LptA has a unique elongated jellyroll fold architecture that was revealed by x-ray crystallography. It is believed that a oligomer composed of multiple copies of LptA paves the road across the

periplasmic space, connecting to the N-terminal domain of LptD (Sperandeo et al., 2019). LptD is the core of outer membrane LPS assembly machinery, arranged in a heterodimeric complex with the lipoprotein LptE and also deploying a lateral-opened gating to drive the LPS insertion, similar to the Bam complex (the lateral opening structure of LptD/E complex is one of the main findings of this PhD thesis and is discussed in detail in the Results section). LptE forms a chaperone of LptD by plugging the transmembrane beta-barrel region of LptD. LptE knockout can lead to impaired cell growth for certain Gram-negative bacteria. The LptD N-terminal domain resembles the structure of LptA or LptC with a similar jellyroll fold, likely assembling as an elongated glove together and forming the LPS transporting bridge within the periplasmic space (Figure 2).

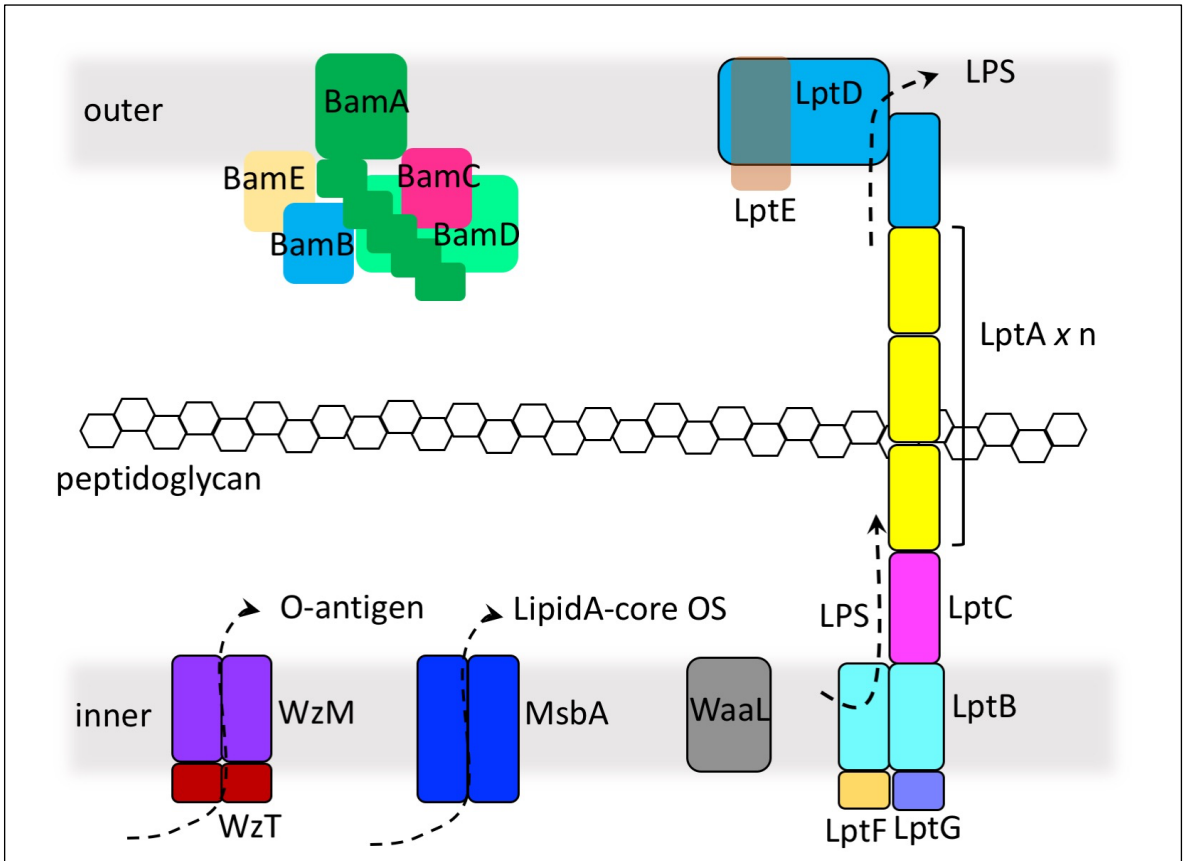


Figure 2 LPS translocation across the two-membrane system (inner and outer membrane) of gram-negative bacteria, from cytosol to extracellular side. First, the pre O-antigen chains are transported by WzM/T complex while the translocation of LipidA-core oligosaccharides (OS) are done by MsbA. Then the ligations are supposed be proceeded by a membrane associated enzyme called WaaL. Once the LPS are formed, they have to be extracted from the outer leaflet of inner membrane by complex of LptB2FG and further delivered to the LPS transporting bridge. Finally, the LPS are assembled in the outer leaflet of outer membrane by another complex of LptD/E. The Bam complex (ABCDE) is responsible for the beta-barrel membrane proteins assembly, including the LptD as well.

1.4 ATP-binding cassette (ABC) transporters and their functional roles.

ATP-binding cassette (ABC) transporters are a large superfamily of transmembrane proteins that use the energy of ATP binding to drive the translocations of a large numbers of physiological substances or xenobiotics across membranes. Their substrates include

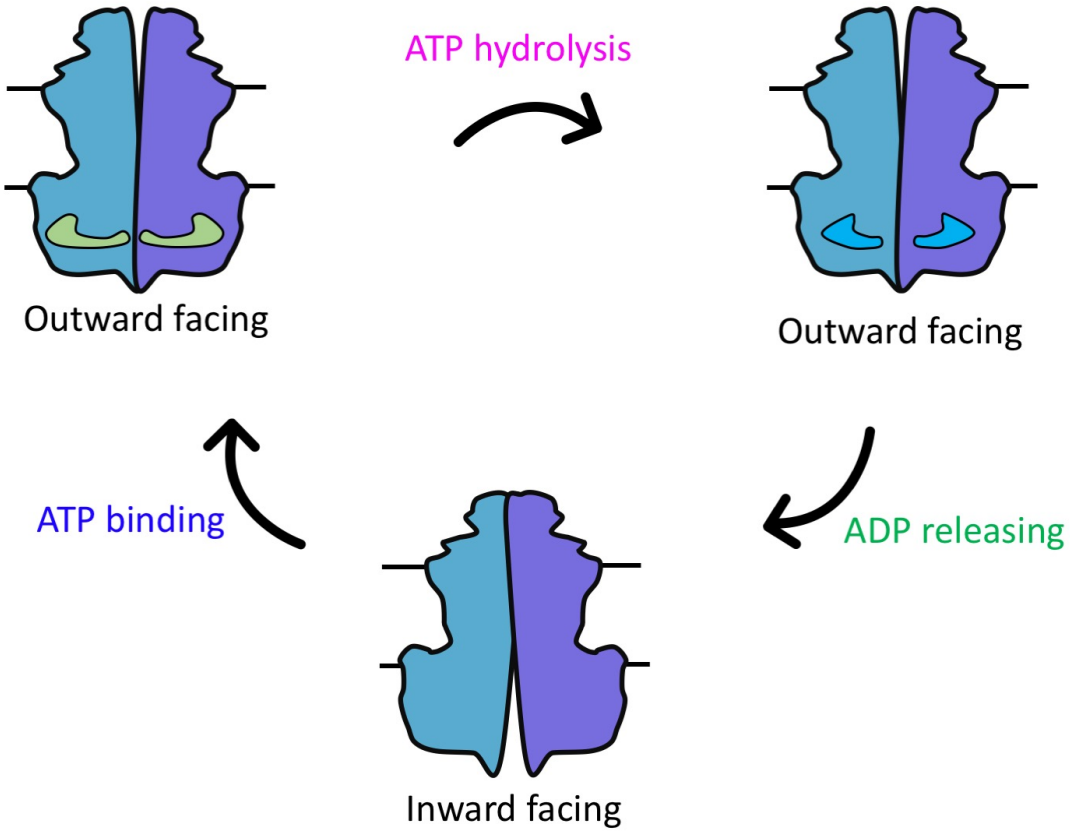


Figure 3 Conformation changes of ATP-binding cassette transporters associated with ATP hydrolysis during a functional cycle. Transporters adapt the conformations from inward facing to outward facing state by coupling the nucleotide binding, hydrolysis and release.

amino acids, ions, lipids, sugars, polypeptides, polysaccharides, anti-cancer drugs, antibiotics and and many others. ABC transporter are found in all living organisms and typically function as one of major energy conversion machineries that couple ATP binding, hydrolysis and ADP release to multiple cellular processes (Figure 3) (Thomas and Tampé, 2020). In mammalian cells, the ABC transporter superfamily is divided into six major classes: ABCA, ABCB, ABCD, ABCE, ABCF and ABCG. Most human ABC transporters are structurally well-conserved, sharing a typical fold of type IV ABC transporters. The sub-classes ABCA and ABCG have different architecture, while the ABCE and ABCF subclasses do not contain transmembrane domains (Figure 4) (Vasiliou et al., 2009).

Physiologically, the dysfunction of ABC transporters may lead to multiple diseases such as retinal degeneration, lipid trafficking disorders, cystic fibrosis and hypercholesterolemia. Multidrug resistance (MDR) is mainly mediated by a few ABC exporters. Human ABCB1

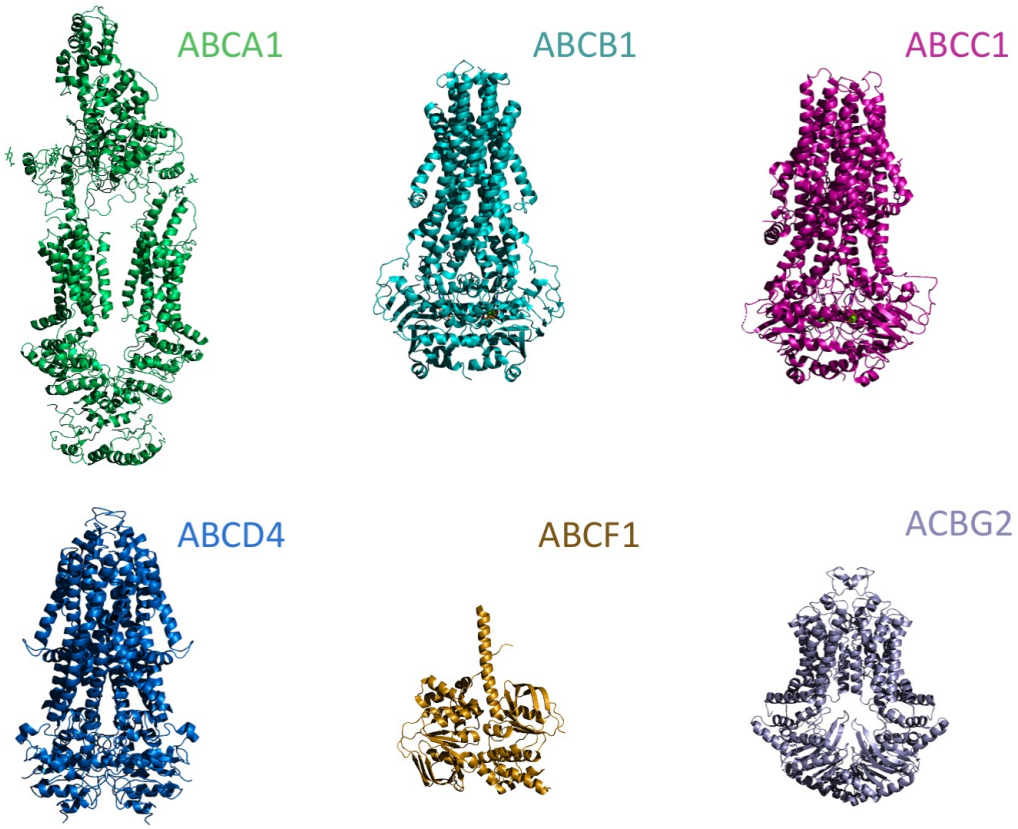


Figure 4 Structures of the mammalian subfamilies of ATP-binding cassette transporters. The cartoon representation of currently known structures. The structures are showed as side view and colored as follows: ABCA1, green (PDB-ID: 5XJY); ABCB1, cyan (PDB-ID: 6C0V); ABCC1, pink (PDB-ID: 6UY0); ABCD4, blue (PDB-ID: 6JBJ); ABCF1, yellow (PDB-ID: 5ZXD) and ABCG2, slate blue (PDB-ID: 6FFC).

and ABCG2 are proposed to be largely involved in the resistance to multiple anti-cancer chemotherapeutic drugs, leading to the failure of cancer treatment (Thomas and Tampé, 2020). For example, human ABCB1 is supposed to be a key molecular machine for pumping Paclitaxel (one of front-line anti-cancer drug) from cytosol to the extracellular space, dramatically reducing the efficiency of chemotherapy (Alam et al., 2019). The first high-resolution structure of the ABC exporter Sav1866 from Staphylococcus aureus has been resolved by Kaspar locher et al in 2006 (Dawson and Locher, 2006). With the new approach of cryo-EM SPA, the structural insights into this protein family have been rapidly advancing in recent years, consequently leading to successes of structure determination of major drug targets, including human ABC exporters ABCG2 as well as ABCB1 (P-glycoprotein) (Figure 4) (Alam et al., 2019; Taylor et al., 2017).

1.5 Main approaches of high-resolution structures determination of membrane proteins

Several technologies of structure determination of membrane proteins have been developed, including NMR, Crystallography and cryo-EM SPA. Each method has its own scope of applications, advantages and limitations in the structural studies of membrane protein samples. For example, in terms of target protein molecular weight, NMR is a good method for small proteins, limited to approximately 30-40 KDa for membrane protein samples. Crystallography covers almost all ranges of molecular weight, while cryo-EM single particle analysis is suitable for solving structures of relatively large membrane proteins.

1.5.1 NMR

As mentioned above, NMR spectroscopy is good to study structures of small membrane proteins, especially beta-barrel transmembrane protein samples. Representative work in this context is the successful structure determination of human VDAC channel, a well-known mitochondrial outer membrane protein with a molecular weight of 31KDa (Figure 1) (Hiller et al., 2008). The molecular weight is a major bottleneck for NMR due to the complexity of the spectroscopic peak-tracking. The largest NMR structure of a membrane protein solved so far is the mitochondrial homo-pentameric calcium uniporter (MCU), which has a molecular weight of 100 kDa (Oxenoid et al., 2016). However, several structures of MCU solved using cryo-EM and X-ray crystallography have been released since then that seem to contradict the previous NMR structure (Wang et al., 2019).

1.5.2 Crystallography

Crystallography uses the general property of protein molecules as well as other substances in nature to form crystals. In solution, protein molecules can pack with each other and form lattices in three-dimensions. A couple of approaches have been developed to obtain 3D crystals of membrane proteins, such as the lipidic cubic phase (LCP), Bicelles and 2D crystals (Ishchenko et al., 2017). A 2D crystal is a particular form of a 3D-crystal, consisting of a single layer of the reconstituted protein-lipid molecules. As long as crystals can be obtained, they can be illuminated with a high energy x-ray or electron beam and the diffraction patterns can be recorded sequentially in certain angles. The 3D electron-density map can be calculated by the processing of those diffraction data with providing the phase

information and then followed by model building against to the experimental map (Dauter and Wlodawer, 2016).

Crystallography is a powerful tool to investigate protein structures since the early days of the structural biology field. It was also the main approach of obtaining high-resolution structures of membrane proteins until the recent breakthrough in cryo-EM single particle analysis. Theoretically, there is no limitation in terms of the protein size and biochemical properties for membrane protein crystallography. However, to crystallize and obtain a well-diffracting crystal of a membrane protein is a big challenge for structural biologists. Nevertheless, crystallography is still an important method to determinate the structures of small membrane proteins, especially smaller than 60kDa (Gavira, 2016; Yeates and Kent, 2012).

1.5.3 Cryo-EM single particle analysis (SPA)

Cryo-EM SPA is a cutting-edge electron microscopy technology of studying biological samples in an unusual ice condition. For the purpose of cryo-EM studies, amorphous ice (also called vitreous water) needs to form in pre-cooled liquid ethane environment. In these conditions, freezing water molecules do not have enough time to form crystalline ice, as the temperature is decreased very fast (Dubochet et al., 1988). In 2017, three scientists had been awarded the Nobel Prize in Chemistry for their contributions to the method development of cryo-EM (Shen, 2018). Cryo-EM has been widely applied in the field of structural biology since a long time ago. Previously, the resolution of the obtained cryo-EM structures was generally limited by the sensitivity of cameras used to record images (CCD cameras in most cases).

In recent years, however, cryo-EM has experienced a “resolution revolution”, started by the introduction of a new generation of Direct Detection Device (DDD) cameras. Now it is possible to record samples with better quality and sensitivity, leading to high-resolution structure determination of large protein samples by using the common image processing software (Kühlbrandt, 2014). Method development of cryo-EM data processing has been also progressed rapidly along with the increasing knowledge of the unique physical effects and behavior involved in the interactions between biological samples and electrons. Nowadays, with commonly used programs such as the Relion, cisTEM or CryoSPARC, the resolution of cryo-EM maps is comparable with crystallography, making it possible to investigate protein structures in side-chain resolution or even higher, sufficiently for the de novo model building. This is especially useful for membrane protein samples and large

protein super-complexes (Grant et al., 2018; Punjani et al., 2017; Scheres, 2012). One of the advantages of cryo-EM SPA is the capability to achieve high-resolution structures without the need of crystallization and solving the phasing problems. Interestingly, in contrast to other methods, cryo-EM SPA has a higher chance to visualize bound lipid molecules and the detergent micelles in the obtained 3D maps, which can be an excellent approach to understand the functional regulation of protein-lipid interactions (Thonghin et al., 2018). Sample preparation is one of the most challenging steps during a cryo-EM project, although the requirements of sample purity and concentration are largely reduced. For cryo-EM studies, the amount of protein is approximately 10-20 times less than for the other methods, like X-ray crystallography or NMR. The protein size of the samples still has to be large in order to be sufficient for particles alignment (Cheng et al., 2015). To date, the smallest protein sample that has been resolved at high-resolution by cryo-EM SPA is a symmetrical dimer in a 60KDa molecular weight, the SARS-CoV-2 ORF3a, without a protein binder (Figure 5a) (Kern et al., 2020).

1.6 Cryo-EM SPA of membrane proteins.

Protein production is the first and usually the key step for protein structure determination by cryo-EM or any other approach. Membrane proteins are normally over-expressed by bacterial or eukaryotic cells transfected with vector carrying the cDNA of interest. For the purification of membrane proteins, the detergents are essential for solubilization of cellular membranes, as the protein-lipid membrane interface is hydrophobic and has to be covered with detergent micelles or nanodiscs to prevent protein precipitation in solution (Thonghin et al., 2018).

Protein sample is isolated from the cells with a reasonable purity and concentration, the sample can be flash-frozen into vitreous ice for the subsequent TEM imaging in high-end electron microscope equipped with a DDD camera. DDD camera have better performance than CCD or CMOS cameras as the high energy electrons can be directly detected with a better signal to noise ratio (SNR), leading to a dramatic improvement in resolution (Milazzo et al., 2011, 2010). The cryo-EM datasets are normally recorded as multi-frame movies that can me later corrected for beam-induced particle motion (Li et al., 2013), with a variable defocus range of a few microns. Subsequently, those image sequences need to be aligned, gain-normalized, drift-corrected, dose-weighted and contrast transfer function (CTF) corrected before image processing (Fernandez-Leiro and Scheres, 2017).

Particle picking is an operation to identify the protein locations in the original 2D micrographs based on the diameter or shape of targeted protein 2D projections, and can be done either manually or automatically after providing a 2D reference. It is now possible to apply machine learning strategies to particle picking (Bepler et al., 2019). This approach seems to be especially useful in the case of in situ cryo-EM of liposome reconstituted samples or virions. Once the particle locations are well-defined and cropped into small images that only contain single proteins, the particle stacks are ready to be subject to 2D and 3D classification. This allows the removal of bad quality images and wrong particle picks from the initial particle set. A 3D reference can be generated either by providing a low-pass filtered 3D volume derived from other reported structures or by ab-initio 3D reconstruction. After 3D refinement, a cryo-EM map at a side-chain resolution can be reconstructed if the data quality is sufficient (Fernandez-Leiro and Scheres, 2017).

1.7 Protein binders used in cryo-EM SPA and the applications for membrane protein structural studies

During the sample preparation of small membrane proteins, one of the largest issues is the particle misalignment due to the limited protein size and high flexibility. For transmembrane proteins, the membrane associated domains are highly covered with detergents or lipid nanodiscs. The detergent micelle is a negative factor in terms of cryo-EM image processing (Thonghin et al., 2018). The micelles mostly represent noise, as detergent molecules are not specifically binding to the transmembrane domains and create issues with particle alignment, especially for small membrane proteins with no additional features in their soluble regions (Singh and Sigworth, 2015). In addition, the proteins composed of multiple flexible domains are highly dynamic, which introduces various distinct conformation heterogeneity in the cryo-EM data and complicates high-resolution structure determination. Several strategies have been developed in order to achieve high-resolution structures of small membrane proteins using cryo-EM SPA. Different protein binders have been widely applied previously for membrane protein crystallization to expand the crystal packing surface (Griffin and Lawson, 2011; Monroe et al., 2011; Röthlisberger et al., 2004). These include Fab fragments, nanobodies and protein fusions. In cryo-EM studies, using high-specific protein binders as well as imaging entire assembled biological complexes has been very helpful in gaining resolution (Koide, 2019). These approaches not only enlarge the protein molecular weight but also decrease protein flexibility.

Fragment antigen-binding (Fab) is a part of an antibody composed of one variable and one constant domain derived from one heavy chain and one light chain. The variable domain contains an antigen binding site, which interacts with the target protein (Wu et al., 2012) (Rougé et al., 2020). The Fab helps to align the protein particles more accurately as it largely

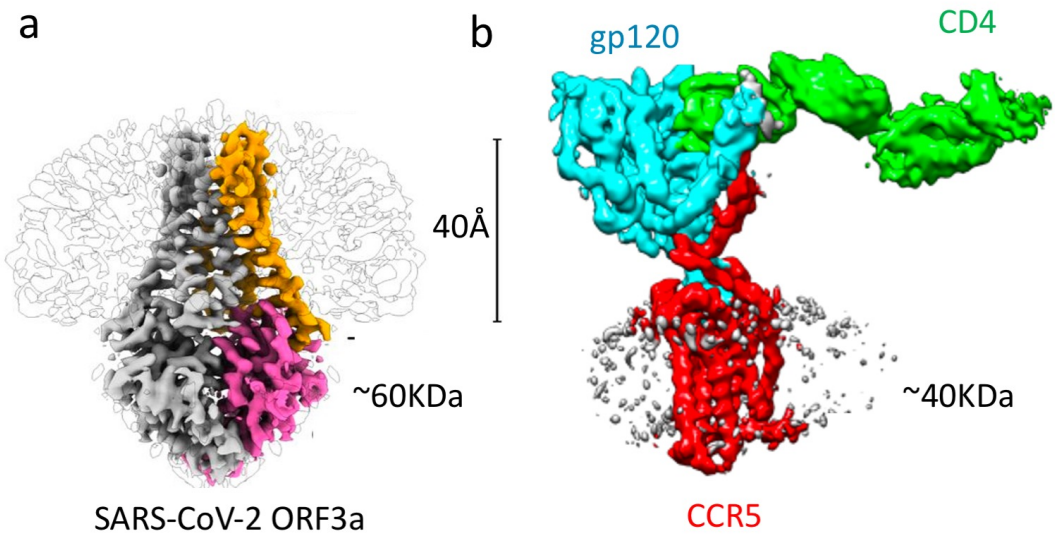


Figure 5 Cryo-EM structures of small transmembrane proteins. (a) The high-resolution structure of SARS-CoV-2 ORF3a was resolved by cryo-EM SPA with an overall resolution of 3 Å. The density is showed as surface. (b) The high-resolution structure of CCR5-gp120-CD4 complex. The density is showed as surface.

expands the water-soluble regions, creates more features and relatively stabilizes one specific conformation of protein molecules. Nanobodies, also called single-domain antibodies, only consists of a single variable domain, which has a molecular weight around 15KDa. Unlike Fabs, nanobodies normally do not build up a comparably large particle in solution. Binding of nanobodies also reduces sample flexibility by blocking protein molecular motion (Uchański et al., 2020, 2019).

Creating a usable and highly rigid antibody fragment is, however, not trivial. Alternatively, assembling a protein complex is a relatively simple way to enlarge the sample size for small membrane proteins, provided there is a biological complex existing for a given molecule. A good example of this approach are the cryo-EM studies of the CCR5-CD4-HIVgp120 complex (Figure 5b) (Shaik et al., 2019).

1.8 References:

Alam A, Kowal J, Broude E, Roninson I, Locher KP. 2019. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. Science 363:753–756. doi:10.1126/science.aav7102

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2002. Membrane Proteins. Molecular Biology of the Cell 4th edition.

Andreeva-Kovalevskaya ZI, Solonin AS, Sineva EV, Ternovsky VI. 2008. Pore-forming proteins and adaptation of living organisms to environmental conditions. Biochemistry Biokhimiia 73:1473–1492. doi:10.1134/s0006297908130087

Andrès C, Agne B, Kessler F. 2010. The TOC complex: Preprotein gateway to the chloroplast. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, Molecular Chaperones and Intracellular Protein Transport 1803:715–723. doi:10.1016/j.bbamcr.2010.03.004

Araiso Y, Tsutsumi A, Qiu J, Imai K, Shiota T, Song J, Lindau C, Wenz L-S, Sakaue H, Yunoki K, Kawano S, Suzuki J, Wischnewski M, Schütze C, Ariyama H, Ando T, Becker T, Lithgow T, Wiedemann N, Pfanner N, Kikkawa M, Endo T. 2019. Structure of the mitochondrial import gate reveals distinct preprotein paths. Nature 575:395–401. doi:10.1038/s41586-019-1680-7

Bastings JJAJ, van Eijk HM, Olde Damink SW, Rensen SS. 2019. d-amino Acids in Health and Disease: A Focus on Cancer. Nutrients 11. doi:10.3390/nu11092205

Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, Berger B. 2019. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nature Methods 16:1153–1160. doi:10.1038/s41592-019-0575-8

Bi Y, Mann E, Whitfield C, Zimmer J. 2018. Architecture of a channel-forming O-antigen polysaccharide ABC transporter. Nature 553:361–365. doi:10.1038/nature25190

Bugge K, Lindorff-Larsen K, Kragelund BB. 2016. Understanding single-pass transmembrane receptor signaling from a structural viewpoint—what are we missing? The FEBS Journal 283:4424–4451. doi:10.1111/febs.13793

Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. 2017. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. Methods in Molecular Biology (Clifton, NJ) 1607:627–641. doi:10.1007/978-1-4939-7000-1_26

Butterworth PJ. 2005. Lehninger: principles of biochemistry (4th edn) D. L. Nelson and M. C. Cox, W. H. Freeman & Co., New York, 1119 pp (plus 17 pp glossary), ISBN 0-7167-4339-6 (2004). Cell Biochemistry and Function 23:293–294. doi:10.1002/cbf.1216

Chen CP, Kernytsky A, Rost B. 2002. Transmembrane helix predictions revisited. Protein Sci 11:2774–2791.

Cheng Y, Grigorieff N, Penczek PA, Walz T. 2015. A primer to single-particle cryo-electron microscopy. Cell 161:438–449. doi:10.1016/j.cell.2015.03.050

Copeland RA. 1994. Introduction to Protein Structure In: Copeland RA, editor. Methods for Protein Analysis: A Practical Guide for Laboratory Protocols. Boston, MA: Springer US. pp. 1–16. doi:10.1007/978-1-4757-1505-7_1

Cross TA, Murray DT, Watts A. 2013. Helical membrane protein conformations and their environment. European biophysics journal: EBJ 42:731–755. doi:10.1007/s00249-013-0925-x

Dal Peraro M, van der Goot FG. 2016. Pore-forming toxins: ancient, but never really out of fashion. Nature Reviews Microbiology 14:77–92. doi:10.1038/nrmicro.2015.3

Dauter Z, Wlodawer A. 2016. Progress in protein crystallography. Protein and Peptide Letters 23:201–210. doi:10.2174/0929866523666160106153524

Dawson RJP, Locher KP. 2006. Structure of a bacterial multidrug ABC transporter. Nature 443:180–185. doi:10.1038/nature05155

Deisenhofer J, Michel H. 1991. High-resolution structures of photosynthetic reaction centers. Annual Review of Biophysics and Biophysical Chemistry 20:247–266. doi:10.1146/annurev.bb.20.060191.001335

Diederichs KA, Ni X, Rollauer SE, Botos I, Tan X, King MS, Kunji ERS, Jiang J, Buchanan SK. 2020. Structural insight into mitochondrial β-barrel outer membrane protein biogenesis. Nature Communications 11:3290. doi:10.1038/s41467-020-17144-1

Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowall AW, Schultz P. 1988. Cryo-electron microscopy of vitrified specimens. Quarterly Reviews of Biophysics 21:129–228. doi:10.1017/s0033583500004297

Fan C, Fan M, Orlando BJ, Fastman NM, Zhang J, Xu Y, Chambers MG, Xu X, Perry K, Liao M, Feng L. 2018. X-ray and cryo-EM structures of the mitochondrial calcium uniporter. Nature 559:575–579. doi:10.1038/s41586-018-0330-9

Feltham R, Vince JE. 2018. Ion Man: GSDMD Punches Pores to Knock Out cGAS. Immunity 49:379–381. doi:10.1016/j.immuni.2018.08.026

Fernandez-Leiro R, Scheres SHW. 2017. A pipeline approach to single-particle processing in RELION. Acta Crystallographica Section D, Structural Biology 73:496–502. doi:10.1107/S2059798316019276

Fujii Noriko, Takata T, Fujii Norihiko, Aki K, Sakaue H. 2018. D-Amino acids in protein: The mirror of life as a molecular index of aging. Biochimica Et Biophysica Acta Proteins and Proteomics 1866:840–847. doi:10.1016/j.bbapap.2018.03.001

Gavira JA. 2016. Current trends in protein crystallization. Archives of Biochemistry and Biophysics 602:3–11. doi:10.1016/j.abb.2015.12.010

Grant T, Rohou A, Grigorieff N. 2018. cisTEM, user-friendly software for single-particle image processing. eLife 7. doi:10.7554/eLife.35383

Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. Carbohydrate Research 356:12–24. doi:10.1016/j.carres.2012.02.027

Griffin L, Lawson A. 2011. Antibody fragments as tools in crystallography. Clin Exp Immunol 165:285–291. doi:10.1111/j.1365-2249.2011.04427.x

Hedin LE, Illergård K, Elofsson A. 2011. An Introduction to Membrane Proteins. J Proteome Res 10:3324–3331. doi:10.1021/pr200145a

Hilger D, Masureel M, Kobilka BK. 2018. Structure and dynamics of GPCR signaling complexes. Nature Structural & Molecular Biology 25:4–12. doi:10.1038/s41594-017-0011-7

Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G. 2008. Solution Structure of the Integral Human Membrane Protein VDAC-1 in Detergent Micelles. Science 321:1206–1210. doi:10.1126/science.1161302

Iadanza MG, Higgins AJ, Schiffrin B, Calabrese AN, Brockwell DJ, Ashcroft AE, Radford SE, Ranson NA. 2016. Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. Nature Communications 7:12865. doi:10.1038/ncomms12865

Ishchenko A, Abola EE, Cherezov V. 2017. Crystallization of Membrane Proteins: An Overview. Methods in Molecular Biology (Clifton, NJ) 1607:117–141. doi:10.1007/978-1-4939-7000-1_5

Jann B, Reske K, Jann K. 1975. Heterogeneity of lipopolysaccharides. Analysis of polysaccharide chain lengths by sodium dodecylsulfate-polyacrylamide gel electrophoresis. European Journal of Biochemistry 60:239–246. doi:10.1111/j.1432-1033.1975.tb20996.x

Kern DM, Sorum B, Hoel CM, Sridharan S, Remis JP, Toso DB, Brohawn SG. 2020. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. bioRxiv. doi:10.1101/2020.06.17.156554

Ketchem RR, Hu W, Cross TA. 1993. High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. Science 261:1457–1460. doi:10.1126/science.7690158

Koide S. 2019. Repurposing off-the-shelf antihelix antibodies for enabling structural biology. Proceedings of the National Academy of Sciences of the United States of America 116:17611–17613. doi:10.1073/pnas.1912643116

Konovalova A, Kahne DE, Silhavy TJ. 2017. Outer Membrane Biogenesis. Annual Review of Microbiology 71:539–556. doi:10.1146/annurev-micro-090816-093754

Kühlbrandt W. 2014. Biochemistry. The resolution revolution. Science (New York, NY) 343:1443–1444. doi:10.1126/science.1251652

Le Trong I, Wang Z, Hyre DE, Lybrand TP, Stayton PS, Stenkamp RE. 2011. Streptavidin and its biotin complex at atomic resolution. Acta Cryst D 67:813–821. doi:10.1107/S0907444911027806

Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, Gubbens S, Agard DA, Cheng Y. 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods 10:584–590. doi:10.1038/nmeth.2472

Li Y, Orlando BJ, Liao M. 2019. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. Nature 567:486–490. doi:10.1038/s41586-019-1025-6

Liu X, Zhang Z, Ruan J, Pan Y, Magupalli VG, Wu H, Lieberman J. 2016. Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores. Nature 535:153–158. doi:10.1038/nature18629

Luo Q, Yang X, Yu S, Shi H, Wang K, Xiao L, Zhu G, Sun C, Li T, Li D, Zhang X, Zhou M, Huang Y. 2017. Structural basis for lipopolysaccharide extraction by ABC

transporter LptB2FG. Nature Structural & Molecular Biology 24:469–474. doi:10.1038/nsmb.3399

Menny A, Serna M, Boyd CM, Gardner S, Joseph AP, Morgan BP, Topf M, Brooks NJ, Bubeck D. 2018. CryoEM reveals how the complement membrane attack complex ruptures lipid bilayers. Nature Communications 9:5316. doi:10.1038/s41467-018-07653-5

Mi W, Li Y, Yoon SH, Ernst RK, Walz T, Liao M. 2017. Structural basis of MsbA-mediated lipopolysaccharide transport. Nature 549:233–237. doi:10.1038/nature23649

Milazzo A-C, Cheng A, Moeller A, Lyumkis D, Jacovetty E, Polukas J, Ellisman MH, Xuong N-H, Carragher B, Potter CS. 2011. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. Journal of Structural Biology 176:404–408. doi:10.1016/j.jsb.2011.09.002

Milazzo A-C, Moldovan G, Lanman J, Jin L, Bouwer JC, Klienfelder S, Peltier ST, Ellisman MH, Kirkland AI, Xuong N-H. 2010. Characterization of a direct detection device imaging camera for transmission electron microscopy. Ultramicroscopy 110:744–747. doi:10.1016/j.ultramic.2010.03.007

Monroe N, Sennhauser G, Seeger MA, Briand C, Grütter MG. 2011. Designed ankyrin repeat protein binders for the crystallization of AcrB: plasticity of the dominant interface. Journal of Structural Biology 174:269–281. doi:10.1016/j.jsb.2011.01.014

Moore PB. 2017. Structural biology: Past, present, and future. New Biotechnology 38:29–35. doi:10.1016/j.nbt.2016.07.010

Moran AP, Prendergast MM, Appelmelk BJ. 1996. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. FEMS immunology and medical microbiology 16:105–115. doi:10.1111/j.1574-695X.1996.tb00127.x

Moreno-Hagelsieb G, Vitug B, Medrano-Soto A, Saier MH. 2017. The Membrane Attack Complex/Perforin Superfamily. Journal of Molecular Microbiology and Biotechnology 27:252–267. doi:10.1159/000481286

Newcomer ME, Jones TA, Aqvist J, Sundelin J, Eriksson U, Rask L, Peterson PA. 1984. The three-dimensional structure of retinol-binding protein. The EMBO journal 3:1451–1454.

Newport TD, Sansom MSP, Stansfeld PJ. 2019. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. Nucleic Acids Research 47:D390–D397. doi:10.1093/nar/gky1047

Ni D, Yang K, Huang Y. 2014. Refolding, crystallization and preliminary X-ray crystallographic studies of the β-barrel domain of BamA, a membrane protein essential for outer membrane protein biogenesis. Acta Crystallogr F Struct Biol Commun 70:362–365. doi:10.1107/S2053230X14003008

Nikaido H. 2003. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiol Mol Biol Rev 67:593–656. doi:10.1128/MMBR.67.4.593-656.2003

Noinaj N, Kuszak AJ, Buchanan SK. 2015. Heat Modifiability of Outer Membrane Proteins from Gram-Negative Bacteria. Methods Mol Biol 1329:51–56. doi:10.1007/978-1-4939-2871-2_4

Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y. 2002. Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA 99:5982. doi:10.1073/pnas.082666399

Overington JP, Al-Lazikani B, Hopkins AL. 2006. How many drug targets are there? Nature Reviews Drug Discovery 5:993–996. doi:10.1038/nrd2199

Oxenoid K, Dong Y, Cao C, Cui T, Sancak Y, Markhard AL, Grabarek Z, Kong L, Liu Z, Ouyang B, Cong Y, Mootha VK, Chou JJ. 2016. Architecture of the mitochondrial calcium uniporter. Nature 533:269–273. doi:10.1038/nature17656

Panchal RG, Smart ML, Bowser DN, Williams DA, Petrou S. 2002. Pore-forming proteins and their application in biotechnology. Current Pharmaceutical Biotechnology 3:99–115. doi:10.2174/1389201023378418

Petersen TN, Lundegaard C, Nielsen M, Bohr H, Bohr J, Brunak S, Gippert GP, Lund O. 2000. Prediction of protein secondary structure at 80% accuracy. Proteins 41:17–20.

Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, Freudenberg M, Ricciardi-Castagnoli P, Layton B, Beutler B. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science (New York, NY) 282:2085–2088. doi:10.1126/science.282.5396.2085

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods 14:290–296. doi:10.1038/nmeth.4169

Rose MC, Henkens RW. 1974. Stability of sodium and potassium complexes of valinomycin. Biochimica et Biophysica Acta (BBA) - General Subjects 372:426–435. doi:10.1016/0304-4165(74)90204-9

Röthlisberger D, Pos KM, Plückthun A. 2004. An antibody library for stabilizing and crystallizing membrane proteins - selecting binders to the citrate carrier CitS. FEBS letters 564:340–348. doi:10.1016/S0014-5793(04)00359-X

Rougé L, Chiang N, Steffek M, Kugel C, Croll TI, Tam C, Estevez A, Arthur CP, Koth CM, Ciferri C, Kraft E, Payandeh J, Nakamura G, Koerber JT, Rohou A. 2020. Structure of CD20 in complex with the therapeutic monoclonal antibody rituximab. Science (New York, NY) 367:1224–1230. doi:10.1126/science.aaz9356

Sanchez RM, Zhang Y, Chen W, Dietrich L, Kudryashev M. 2020. Subnanometer-resolution structure determination in situ by hybrid subtomogram averaging - single particle cryo-EM. Nature Communications 11:3709. doi:10.1038/s41467-020-17466-0

Scheres SHW. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology 180:519–530. doi:10.1016/j.jsb.2012.09.006

Schulz GE. 2002. The structure of bacterial outer membrane proteins. Biochimica Et Biophysica Acta 1565:308–317. doi:10.1016/s0005-2736(02)00577-1

Schulz GE. 2000. beta-Barrel membrane proteins. Current Opinion in Structural Biology 10:443–447. doi:10.1016/s0959-440x(00)00120-2

Serra-Batiste M, Ninot-Pedrosa M, Bayoumi M, Gairí M, Maglia G, Carulla N. 2016. Aβ42 assembles into specific β-barrel pore-forming oligomers in membrane-mimicking environments. Proc Natl Acad Sci U S A 113:10866–10871. doi:10.1073/pnas.1605104113

Shaik MM, Peng H, Lu J, Rits-Volloch S, Xu C, Liao M, Chen B. 2019. Structural basis of coreceptor recognition by HIV-1 envelope spike. Nature 565:318–323. doi:10.1038/s41586-018-0804-9

Shen PS. 2018. The 2017 Nobel Prize in Chemistry: cryo-EM comes of age. Analytical and Bioanalytical Chemistry 410:2053–2057. doi:10.1007/s00216-018-0899-8

Singh SK, Sigworth FJ. 2015. Cryo-EM: Spinning the Micelles Away. Structure (London, England: 1993) 23:1561. doi:10.1016/j.str.2015.08.001

Sommer MS, Daum B, Gross LE, Weis BLM, Mirus O, Abram L, Maier U-G, Kühlbrandt W, Schleiff E. 2011. Chloroplast Omp85 proteins change orientation during evolution. PNAS 108:13841–13846. doi:10.1073/pnas.1108626108

Sperandeo P, Martorana AM, Polissi A. 2019. The Lpt ABC transporter for lipopolysaccharide export to the cell surface. Research in Microbiology 170:366–373. doi:10.1016/j.resmic.2019.07.005

Tabaton M, Tamagno E. 2007. The molecular link between beta- and gamma-secretase activity on the amyloid beta precursor protein. Cellular and molecular life sciences: CMLS 64:2211–2218. doi:10.1007/s00018-007-7219-3

Taylor NMI, Manolaridis I, Jackson SM, Kowal J, Stahlberg H, Locher KP. 2017. Structure of the human multidrug transporter ABCG2. Nature 546:504–509. doi:10.1038/nature22345

Thomas C, Tampé R. 2020. Structural and Mechanistic Principles of ABC Transporters. Annual Review of Biochemistry 89:605–636. doi:10.1146/annurev-biochem-011520-105201

Thonghin N, Kargas V, Clews J, Ford RC. 2018. Cryo-electron microscopy of membrane proteins. Methods (San Diego, Calif) 147:176–186. doi:10.1016/j.ymeth.2018.04.018

Tomasek D, Rawson S, Lee J, Wzorek JS, Harrison SC, Li Z, Kahne D. 2020. Structure of a nascent membrane protein as it folds on the BAM complex. Nature 583:473–478. doi:10.1038/s41586-020-2370-1

Tosatto L, Andrighetti AO, Plotegher N, Antonini V, Tessari I, Ricci L, Bubacco L, Dalla Serra M. 2012. Alpha-synuclein pore forming activity upon membrane association. Biochimica Et Biophysica Acta 1818:2876–2883. doi:10.1016/j.bbamem.2012.07.007

Tsirigotaki A, De Geyter J, Šoštaric N, Economou A, Karamanou S. 2017. Protein export through the bacterial Sec pathway. Nature Reviews Microbiology 15:21–36. doi:10.1038/nrmicro.2016.161

Uchański T, Masiulis S, Fischer B, Kalichuk V, Wohlkönig A, Zögg T, Remaut H, Vranken W, Aricescu AR, Pardon E, Steyaert J. 2019. Megabodies expand the

nanobody toolkit for protein structure determination by single-particle cryo-EM. bioRxiv 812230. doi:10.1101/812230

Uchański T, Pardon E, Steyaert J. 2020. Nanobodies to study protein conformational states. Current Opinion in Structural Biology 60:117–123. doi:10.1016/j.sbi.2020.01.003

Vasiliou V, Vasiliou K, Nebert DW. 2009. Human ATP-binding cassette (ABC) transporter family. Hum Genomics 3:281–290. doi:10.1186/1479-7364-3-3-281

Vinothkumar KR, Henderson R. 2010. Structures of membrane proteins. Quarterly Reviews of Biophysics 43:65–158. doi:10.1017/S0033583510000041

Wang Y, Nguyen NX, She J, Zeng W, Yang Y, Bai X-C, Jiang Y. 2019. Structural Mechanism of EMRE-Dependent Gating of the Human Mitochondrial Calcium Uniporter. Cell 177:1252-1261.e13. doi:10.1016/j.cell.2019.03.050

Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annual Review of Biochemistry 83:99–128. doi:10.1146/annurev-biochem-060713-035600

Wong A, Gehring C, Irving HR. 2015. Conserved Functional Motifs and Homology Modeling to Predict Hidden Moonlighting Functional Sites. Front Bioeng Biotechnol 3. doi:10.3389/fbioe.2015.00082

Wu H, Fuxreiter M. 2016. A Structure and Dynamics Continuum of Higher-Order Assemblies: Amyloids, Signalosomes and Granules. Cell 165:1055–1066. doi:10.1016/j.cell.2016.05.004

Wu S, Avila-Sakar A, Kim J, Booth DS, Greenberg CH, Rossi A, Liao M, Li X, Alian A, Griner SL, Juge N, Yu Y, Mergel CM, Chaparro-Riggers J, Strop P, Tampé R, Edwards RH, Stroud RM, Craik CS, Cheng Y. 2012. Fabs enable single particle cryoEM studies of small proteins. Structure (London, England: 1993) 20:582–592. doi:10.1016/j.str.2012.02.017

Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367:1444–1448. doi:10.1126/science.abb2762

Yang F, Moss LG, Phillips GN. 1996. The molecular structure of green fluorescent protein. Nature Biotechnology 14:1246–1251. doi:10.1038/nbt1096-1246

Yeates TO, Kent SBH. 2012. Racemic protein crystallography. Annual Review of Biophysics 41:41–61. doi:10.1146/annurev-biophys-050511-102333

Bouvette J, Liu H-F, Du X, Zhou Y, Sikkema AP, Da Fonseca Rezende E Mello J, Klemm B, Huang R, Schaaper RM, Borgnia MJ, Bartesaghi A. 2020. Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. bioRxiv 2020.09.24.294983. doi:10.1101/2020.09.24.294983

2 Chapter II: Aims of my thesis

This thesis is a collection of my main projects performed over the course of my PhD studies in the past three years. The Results section (Chapters III to V) of this thesis are manuscripts prepared for submission.

Due to global health crises and the growth of pharmaceutical market, drug discovery is in high demand. It is more and more clear that the protein structure determination is an important aspect of the early stages of drug research and development. Furthermore, membrane proteins constitute the majority of druggable targets. Structure determination of membrane proteins is quite challenging with well-established approaches such as NMR and X-ray crystallography. Taking this into account, the goal of my thesis is to apply the new approach of cryogenic electron microscopy single particle analysis in real cases to study membrane proteins as potential drug targets.
Here I have presented three case studies to bring the perspective: How cryo-EM is dealing with the membrane protein and drug researches.

Chapter I is a general introduction for the field of membrane protein structure biology, the biological relevance for the following sub-projects and the applications of cryo-EM SPA.

Chapter III introduces the studies on the interactions of ABCG2, a human membrane protein involved in the resistance of multiple anti-cancer compounds, with a number of inhibitors.

Chapter IV introduces our findings concerning the Neisseria gonorrhoeae LPS outer membrane assembly machinery, and aims at providing the structural information for potential antibiotic discovery.

Chapter V introduces a third project, which analyzes structurally the interactions between the SARS-CoV2 spike protein and its receptor, human ACE2. This study is potentially useful in the context of the on-going global pandemic of COVID-19, which has started early this year.

Chapter VI summarizes and concludes the thesis, proposes directions and projects for new PhD students if the directions are followed, and introduces the outlook in the field of cryo-EM based drug researches.

3 Chapter III: Cryo-EM structures of human ABCG2 in complex with anti-cancer drugs

This section is a manuscript prepared for submission.
Running Title:

Structural basis of drug recognition by multidrug transporter ABCG2

Julia Kowal1*#, Dongchun Ni2*, Scott M Jackson1, Ioannis Manolaridis1, Henning Stahlberg2#, Kaspar P Locher1#

1 Institute of Molecular Biology and Biophysics, ETH Zurich, Switzerland
2 Center for Cellular Imaging and NanoAnalytics (C-CINA), Biozentrum, University of Basel, Switzerland

* These authors contributed equally to this work.

Correspondence should be addressed to J.K (julia.kowal@mol.biol.ethz.ch), H.S (henning.stahlberg@unibas.ch) and K.P.L (locher@mol.biol.ethz.ch)

Keywords: ABCG2, BCRP, topotecan, mitoxantrone, tariquidar, anticancer drugs, ABC transporter, cavity 1, cryo-EM

In this section, I describe a structural study on human ABCG2 (a human membrane protein involved in the multiple anti-cancer drugs resistance) with the binding of a number of anti-cancer drugs, including G2-MITO, G2-TOPO and G2-TAR. These structures have been resolved to high-resolution by cryo-EM single particle analysis, revealing the unique features of several anti-cancer drugs in the central cavity.

My contribution: Cryo-EM dataset acquisition; image processing; model building; coordinates refinement; interpreting the structures; contributing to manuscript writing and editing.

3.1 Abstract

ABCG2 is an ATP-binding cassette (ABC) transporter whose function affects the pharmacokinetics of drugs and contributes to multidrug resistance of cancer cells. While its interaction with the endogenous substrate estrone-3-sulfate (E1S) has been elucidated at a structural level, the recognition of exogenous drugs is not understood at sufficiently high resolution. Here we present three single-particle cryo-EM structures of nanodiscreconstituted, human ABCG2 with anticancer drugs topotecan, mitoxantrone and tariquidar. The structures were obtained in the presence of 5D3-Fab fragment, which lowers the ATPase activity of ABCG2 but does not abolish the stimulation by bound drugs. The structures revealed a single molecule of the substrates bound in cavity 1, previously identified as the recognition site of E1S. Biochemical analysis of ABCG2 variants confirmed the contributions of specific side chains to drug recognition. Using 3D variability analyses, we found a correlation between substrate binding to the transmembrane domains (TMDs) and reduced dynamics of the nucleotide binding domains (NBDs), suggesting a structural explanation for drug-induced ATPase stimulation. Our findings provide structural insight into how ABCG2 differentiates between inhibitors and substrates and may guide a rational design of new modulators or modified substrates.

3.2 Introduction

ABCG2, also known as breast cancer resistance protein (BCRP), is an ATP-binding cassette (ABC) transporter expressed in many tissues and tissue barriers, including the blood–brain, blood–testis and maternal–fetal barriers (Crowley, McDevitt, & Callaghan, 2010; Diestra et al., 2002; Fetsch et al., 2006; Robey et al., 2009; Thomas & Tampe, 2020). It actively transports endogenous substrates such as uric acid or estrone-3-sulfate (E1S), but also removes exogenous, cytotoxic compounds from cells (Ishikawa, Aw, & Kaneko, 2013). Like other multi-drug ABC transporters such as ABCB1 (P-glycoprotein) or ABCC1 (MRP1), ABCG2 is overexpressed in certain cancer cells, which contributes to drug resistance and affects chemotherapeutic intervention (Szakacs, Paterson, Ludwig, Booth-Genthe, & Gottesman, 2006). ABCG2 also affects the oral uptake rate and pharmacokinetics of many commonly used drugs (International Transporter et al., 2010). Both ABCG2 and ABCB1 are highly expressed at the blood-brain barrier (Dauchy et al., 2008), where they prevent the entry of a broad variety of xenobiotics, thereby limiting drug delivery to the brain (Begley, 2004; Ghosh, Puvenna, Gonzalez-Martinez, Janigro, & Marchi, 2011; Mahringer & Fricker, 2016).

ABCG2 recognizes and actively transports chemically distinct compounds, most of which are relatively flat, polycyclic and hydrophobic (Bakhsheshian et al., 2013; Imai et al., 2003; Mao & Unadkat, 2015; Mo & Zhang, 2012; Vlaming, Lagas, & Schinkel, 2009). Among them, anti-cancer drugs used during therapy are of particular interest. They include topoisomerase inhibitors, tyrosine kinase inhibitors, and antimetabolites (Mo & Zhang, 2012). There is a substantial overlap between the compounds recognized by different multidrug ABC transporters. Intriguingly, there are also compounds that are substrates of one, but inhibitors of another multidrug ABC transporter (Stacy, Jansson, & Richardson, 2013). A key example is tariquidar, a strong inhibitor of ABCB1 that was reported to be transported by ABCG2, albeit at low rates and low concentration (Bauer et al., 2013; Kannan et al., 2011; Kohler & Wiese, 2015; Ochoa-Puentes et al., 2013; Pick, Klinkhammer, & Wiese, 2010; Puentes et al., 2011; Roe et al., 1999).

Several structures of human ABCG2 have been determined over the past years (Jackson et al., 2018; Manolaridis et al., 2018; Taylor et al., 2017). The structure containing the physiological substrate E1S bound in a central, cytoplasm-facing binding pocket helped identify key residues contributing to substrate binding (Manolaridis et al., 2018). A

comparison with structures bound to the inhibitory compounds MZ29 or MB136 (Jackson et al., 2018) suggested that the binding sites of substrates and inhibitors overlap. While a recent study reported ABCG2 structures with bound anti-cancer drugs imatinib, mitoxantrone, and SN38 (Orlando & Liao, 2020), it remained unclear which ABCG2 residues are involved in specific drug-ABCG2 interactions and how larger drugs can be accommodated and transported by ABCG2. To address these questions, we determined three single-particle cryo-EM structures of nanodisc-reconstituted, substrate-bound human ABCG2 in a pre-translocation state. The drugs we selected were the topoisomerase I inhibitor topotecan (421.5 Da) used to treat ovarian, cervical and small-cell lung cancer (Pommier, 2006), the topoisomerase II inhibitor mitoxantrone (444.5 Da) supporting leukemia and multiple sclerosis therapies (Evison, Sleebs, Watson, Phillips, & Cutts, 2016), and the anthranilamide derivative tariquidar (647 Da), an ABCB1 inhibitor also used in cancer clinical trials (Bauer et al., 2013; Kohler & Wiese, 2015; Ochoa-Puentes et al., 2013; Pick et al., 2010; Puentes et al., 2011; Roe et al., 1999). Our functional studies and high-resolution structures reveal how these drugs are accommodated in the binding pocket and which ABCG2 residues contribute to drug recognition. We demonstrate that substrates are not locked as firmly as inhibitors but are likely shifting in the binding cavity. By analyzing the structural variability between the single particles contributing to the 3D reconstructions, we found that the structural flexibility of the nucleotide binding domains (NBDs) is reduced by substrate binding despite the fact that the substrate retains some mobility inside the drug-binding pocket. Our results provide the basis for understanding drug-ABCG2 interactions and for the future development of modulators and inhibitors.

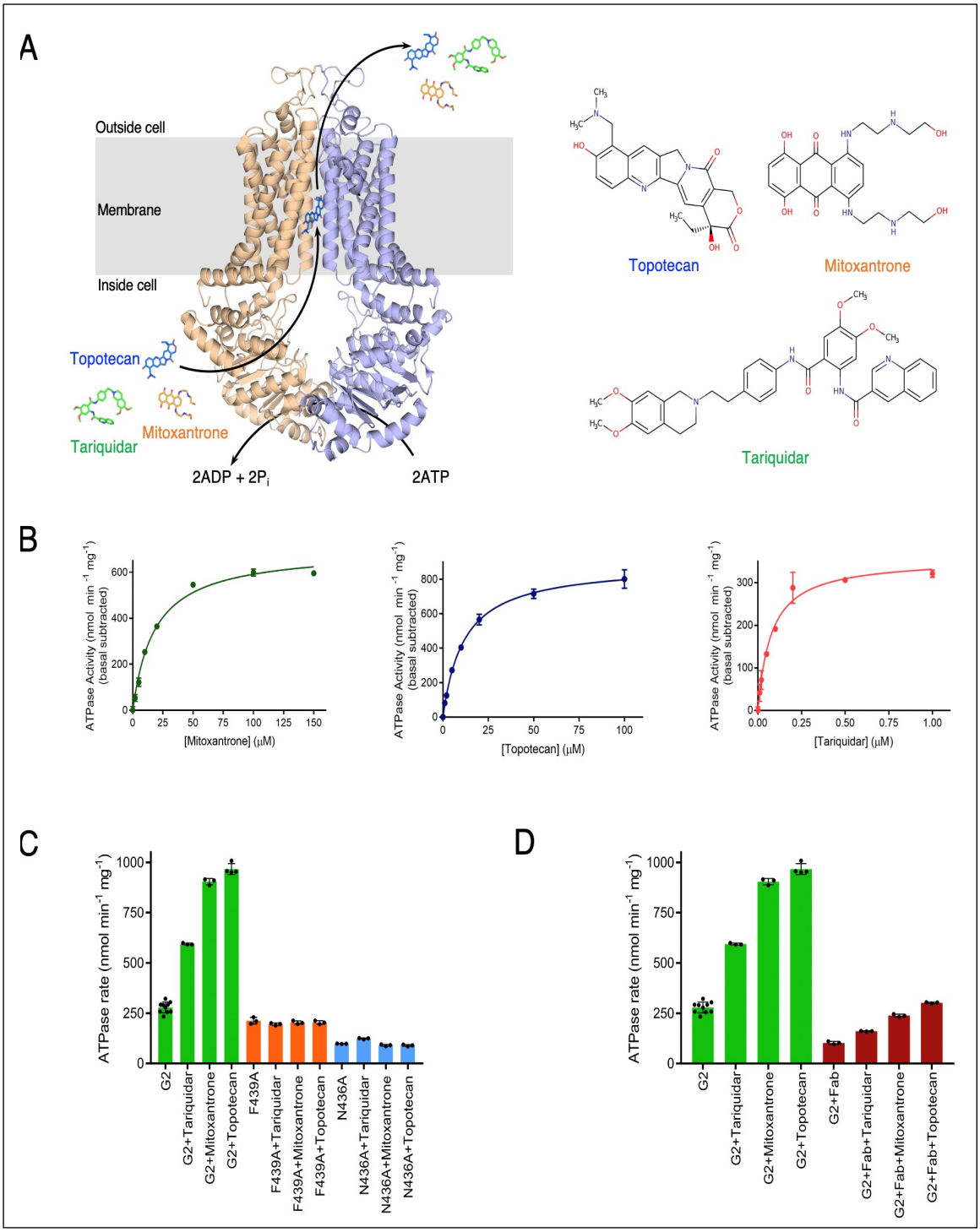


Figure 6 Structure and substrate transport of the ABCG2.

(A) Cartoon representation of ABCG2 with bound substrate topotecan. ABCG2 monomers are colored salmon and blue. The substrates topotecan, mitoxantrone and tariquidar are shown as sticks and colored blue, orange and green, respectively. The substrate transport cycle is indicated by arrows. Bound 5D3-Fab is omitted for clarity. Chemical structures of the substrates are shown on the right.

Figure 6 Structure and substrate transport of the ABCG2.

(B) Determination of the EC50 of ATPase stimulation of ABCG2 in proteoliposomes by mitoxantrone, topotecan or tariquidar. The basal ATPase activity was normalized to zero. Each point represents the mean rate derived from technical replicates, for mitoxantrone and tariquidar n = 3, for topotecan n = 4.

(C) ATPase activity of liposome-reconstituted wild-type and mutant ABCG2 (F439A, N436A) in the presence and absence of 0.5 µM tariquidar, 100 µM mitoxantrone, or 50 µM topotecan. The bars show means; error bars show standard deviations; and dots show rates derived from each technical replicate (same batch of liposomes).

(D) Effect of 5D3-Fab on ATPase stimulation of liposome-stimulated ABCG2. ATPase rates were determined in the presence or absence of 5D3-Fab and either 0 or 0.5 µM tariquidar, 0 or 100 µM mitoxantrone, and 0 or 50 µM topotecan. Bars show means and dots show the rates derived from each technical replicate (same batch of liposomes). Error bars show the standard deviation.

3.3 RESULTS

3.3.1 In vitro characterization of ABCG2 function

We first reconstituted ABCG2 in liposomes and performed ATPase assays in the presence of substrates. We observed a concentration-dependent stimulation of the ATPase activity by topotecan, mitoxantrone and tariquidar (Figure 6), in agreement with previous observations(Hernandez Lozano et al., 2020; Imai et al., 2003; Kannan et al., 2011; Mao & Unadkat, 2015; Mo & Zhang, 2012; Vlaming et al., 2009). The half-maximal effective concentration values (EC50) of substrate-induced ATPase stimulation were used to guide the substrate concentrations used for EM sample preparations. The EC50 values of mitoxantrone- or topotecan-induced ATP stimulation of ABCG2 were 18.3 µM, and 11.7 µM, respectively, which is three orders of magnitude greater than the EC50 observed for tariquidar, which was 0.08 µM (Table 1). The EC50 of E1S-induced ATPase stimulation of ABCG2 alone or in the presence of 5D3-Fab were 15.7 µM and 18 µM, respectively(Manolaridis et al., 2018), which is in the same range as now observed for mitoxantrone and topotecan.

We subsequently investigated ABCG2 variants carrying the mutations F439A and N436A, since these amino acids are involved in substrate binding(Manolaridis et al., 2018). The ATPase rates of these variant proteins were reduced, and they showed no stimulation by the drugs studied, indicating that interactions between the substrates and the side chain of F439 (π-stacking between phenyl ring and polycyclic molecules) and N436 (hydrogen bond formation) are essential for substrate binding (Figure 6C). For our structural studies, we used 5D3-Fab for structural studies to increase particle mass and thus the final volume

resolution. This antibody fragment binds to the extracellular side of ABCG2, inhibits its transport activity and slows down its ATPase rate(Taylor et al., 2017). To assess whether the antibody fragment interfered with drug binding to ABCG2, we measured the stimulation of ATPase activity of liposome-reconstituted ABCG2 by drugs in the presence and absence of 5D3-Fab. As shown in Figure 6D and Table 2, 5D3-Fab did not interfere with the ATPase stimulation, suggesting that drug binding to ABCG2 is not affected by bound antibody.

3.3.2 Cryo-EM structures of drug-bound ABCG2

In order to obtain the single particle cryo-EM structures, we added 5D3-Fab to nanodisc reconstituted ABCG2 and incubated the complex with either 100 µM topotecan, 150 µM mitoxantrone, or 1 µM tariquidar for 10 minutes at room temperature prior to plunge-

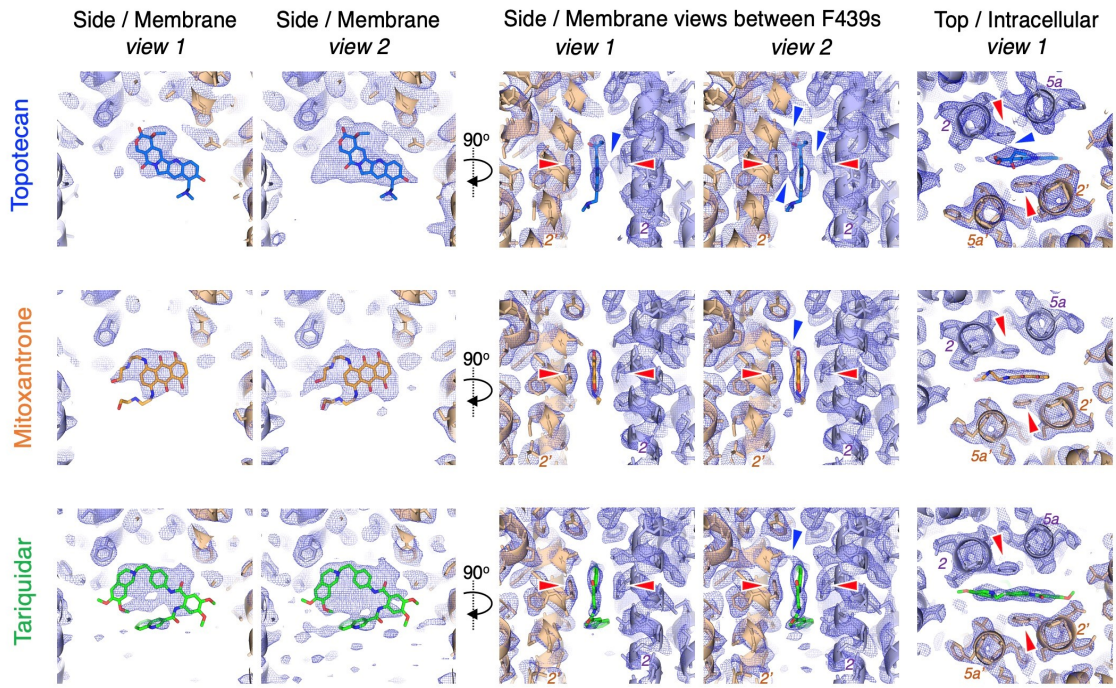


Figure 7 Drugs bound to ABCG2. The binding pocket (cavity 1) of ABCG2 and the EM density maps are shown from the membrane plane and from the extracellular side. Non-symmetrized electron microscopy maps of ABCG2 with substrates are shown as blue mesh at two different thresholds (views 1 and 2). In view 1 the map levels of 6.0σ/ 5.0σ/ 9.0σ, as defined by Pymol, were used to display the maps of ABCG2 with topotecan/mitoxantrone/tariquidar, respectively; in view 2 the map levels were set to 3.5σ/ 3.5σ/ 6.0σ, respectively. The map display levels were adjusted to show similar EM densities around amino acids in binding pocket in all maps. Bound topotecan, mitoxantrone and tariquidar molecules are shown as blue, orange and green sticks, respectively. Side chains of F439 are marked with red arrows. Blue arrows mark sites where the EM density covering side chains and bound drugs are connected. TM helices are labeled 2, 2', 5a and 5a'.

freezing the grids. The obtained cryo-EM data were processed with Relion and CryoSPARC software suites. During data processing we noticed significant dynamics of the NBDs of ABCG2 among the 3D classes. We observed that 35-70% of the particles contained separated and/or poorly resolved NBDs. After numerous rounds of 3D classifications, only the most stable conformations were selected and further refined to high-resolution. The cryo-EM structures of topotecan-bound, mitoxantrone-bound and tariquidar-bound ABCG2 were resolved at overall resolutions of 3.39 Å, 3.51 Å and 3.12 Å for non-symmetrized (i.e. C1) maps, respectively (Table 3, supplementary figures).

The refined structures revealed inward-facing conformations of ABCG2, with drugs bound in cavity 1 in a similar location where bound E1S was observed in ABCG2(Jackson et al., 2018; Manolaridis et al., 2018; Taylor et al., 2017). ABCG2 has two-fold symmetry but each substrate molecule was bound only in one copy, as shown in Figure 7. It was not possible to simultaneously fit two molecules in any of the EM densities because their polycyclic structures would sterically clash. We noticed that the ABCG2-topotecan C1-map showed additional weak density, consistent with a substrate molecule bound in a 180° rotated orientation at the same location. This is likely due to a subpopulation of the particles rotated around the two-fold symmetry axis of the transporter (Figure 7, Figure 18). A similar observation was made in the structure of mitoxantrone-bound ABCG2.

The TMDs, including the drug-binding pockets, were very well resolved (resolution higher than 3 Å), allowing accurate fitting of all side chains of key amino acids involved in substrate interactions. The strongest and most compact EM densities for substrates were observed for topotecan and mitoxantrone, where good fitting of the molecule models was despite the finding that the density covering the drug substrates is of lower quality than that covering the side chains. Intriguingly, despite tariquidar displaying the lowest EC50 of the three compounds, the density covering the drug was less-well-defined and more fragmented, suggesting that in spite of its higher affinity for ABG2, tariquidar is more flexible or bound in multiple orientations (Figure 7, Figure 19). Upon reprocessing the data in CryoSPARC, we obtained a more prominent density for tariquidar, albeit at the expense of a slightly lower

overall resolution of the transporter (Figure 20). In all EM maps, we observed EM density features connecting substrates and ABCG2 side chains (Figure 7), as detailed below.

3.3.3 Substrate-ABCG2 interactions

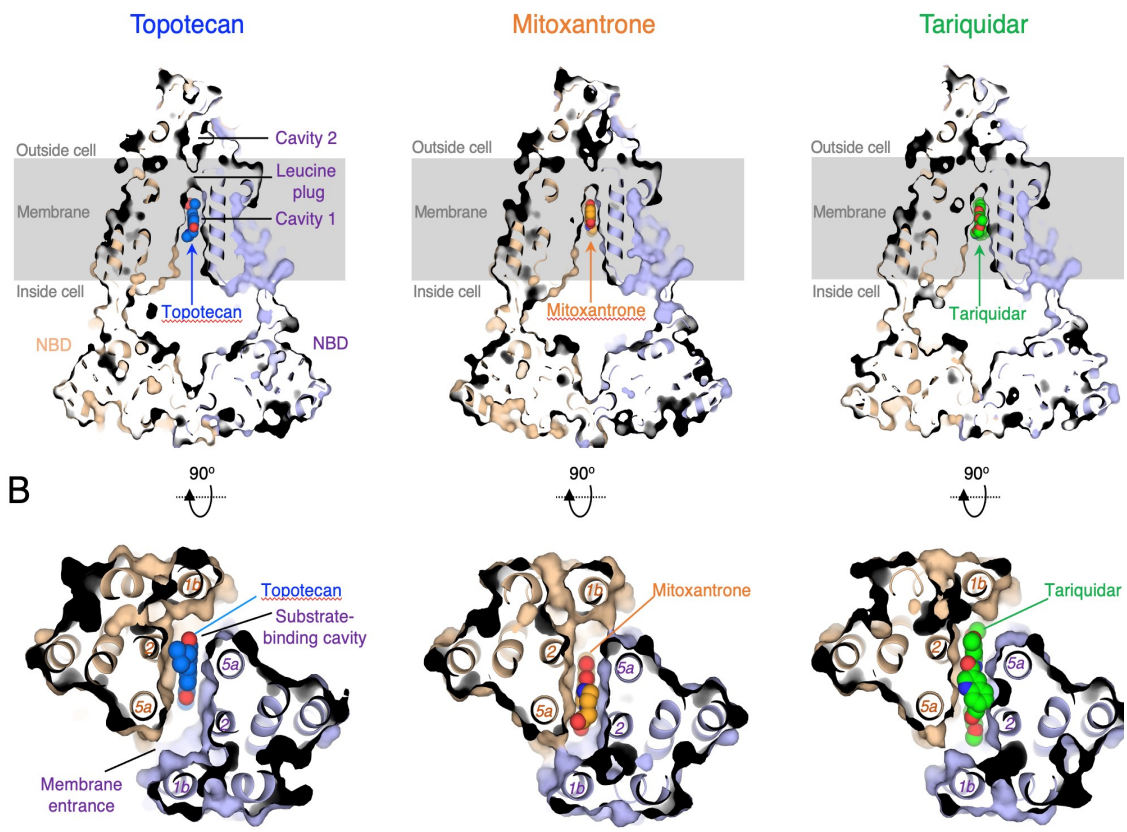


Figure 8 Structures of topotecan-, mitoxantrone- and tariquidar- bound ABCG2.

(A) Vertical slices through a surface representation of ABCG2 with bound substrates shown as blue (topotecan), orange (mitoxantrone) and green (tariquidar) spheres and labeled. Cavities 1 and 2 and the leucine plug, as well as membrane and NBD regions are indicated. 5D3-Fab is omitted for clarity.

(B) Cavity 1 viewed from the cytoplasm, with NBDs of ABCG2 removed for clarity and with substrates as in a and labeled. Structures reveal the fit of substrates in the slit-like binding cavity. TM helices are labeled 1b, 2 and 5a.

Topotecan, mitoxantrone and tariquidar all bind in the same multidrug slit-like binding cavity (cavity 1) located at the level of the center of the membrane, below the leucine plug of ABCG2 and between TM helices 1b, 2 and 5a (Figure 8). The drug binding pocket lies on the two-fold molecular symmetry axis of the ABCG2 dimer. In contrast to the single copies of drug substrates, the Ko143-derived inhibitor MZ29 was previously found to bind

in two copies shifted away from the two-fold symmetry axis, thereby almost entirely filling cavity 1. Another inhibitor, the tariquidar-derived MB136, was found to bind as one copy in a stretched conformation, also filling cavity more completely than the substrates(Jackson et al., 2018). The shape of ABCG2's binding pocket causes flat compounds to be preferred,

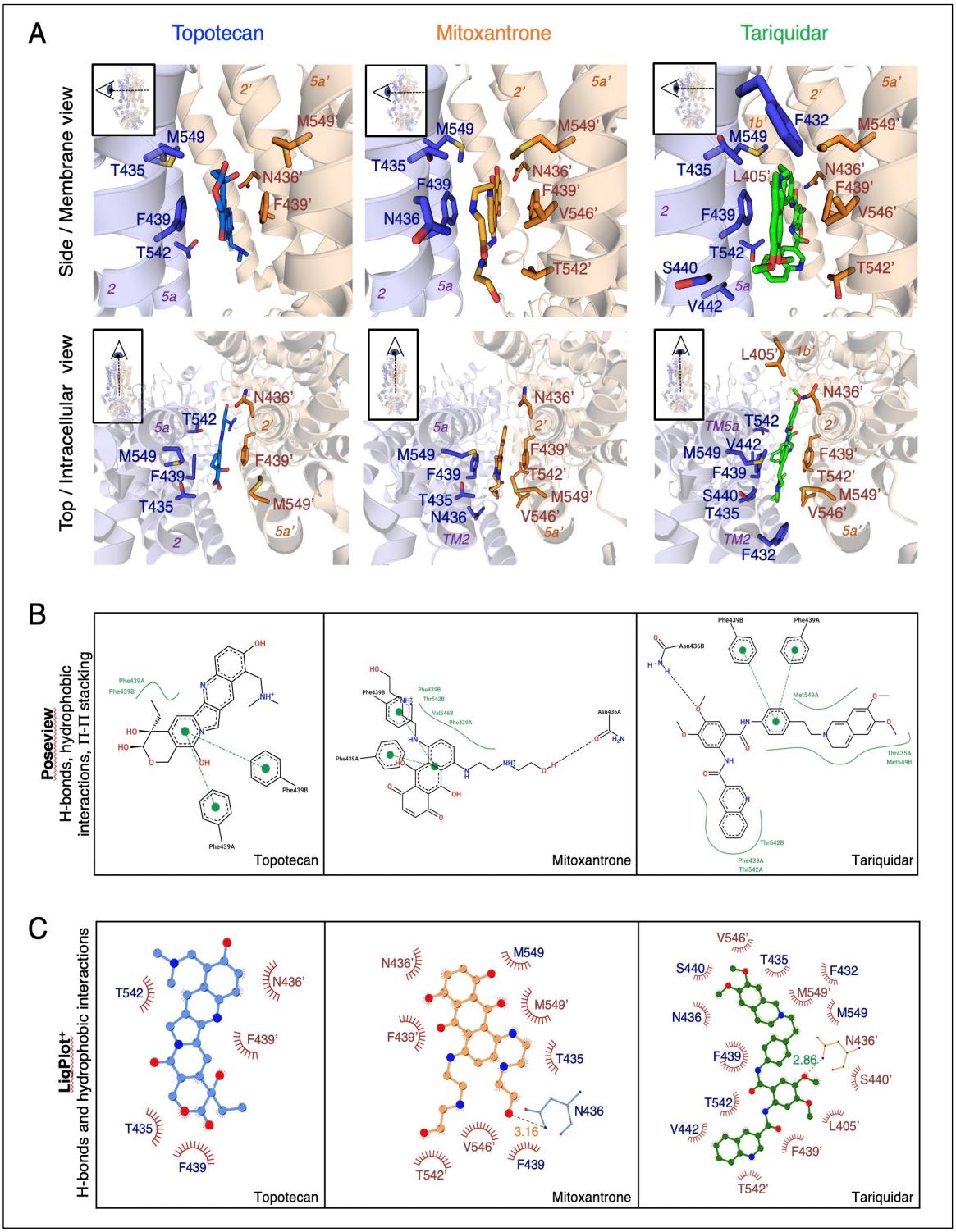


Figure 9 ABCG2-drug interactions.

(A) Specific interactions between substrates topotecan (blue), mitoxantrone (orange) or tariquidar (green) and ABCG2 residues in cavity 1, viewed from the membrane (upper line) and from the extracellular site towards the cytoplasm (bottom line), respectively. Ribbon diagram of the ABCG2 homodimer, with individual ABCG2 monomers is colored blue and salmon. Interacting residues are shown as sticks and labeled. Substrates are shown as sticks. TM helices are labeled 1b', 2, 2', 5a and 5a'.

(B) Two-dimensional visualizations of protein-substrate interactions generated from three-dimensional inputs. Hydrogen bonds are shown as black dashed lines, hydrophobic interactions as smooth contour lines between the respective amino acids and the substrate molecule, and p- p stacking as green dots connected by dashed line. Generated with software Poseview(Stierand & Rarey, 2010).

(C) Interactions between three substrates and ABCG2 side chains. Nonbonded interactions are represented by spoked red arcs and hydrogen bonds are indicated by dashed green lines. The amino acid residues are as shown as singleletter abbreviations.

in contrast to the globular molecules favored by ABCB1(Alam, Kowal, Broude, Roninson, & Locher, 2019; Alam et al., 2018) or ABCC1 transporters(Johnson & Chen, 2017).

The well-resolved and detailed maps of all three ABCG2-substrate structures allowed us to understand the specific requirements of substrate binding in cavity 1. Each substrate was polycyclic and bound between a pair of opposing phenyl rings of F439 by π-stacking interactions (Figure 9). In addition, tariquidar and mitoxantrone formed hydrogen bonds with the side chain of N436, while for topotecan this interaction was absent due to the smaller size of the drug. The distance between the carboxamide group of N436 and the oxygen atom of the methoxy group of tariquidar was 2.9 Å, that between N436 and the hydroxyl-ethylamine tail of mitoxantrone was
3.2 Å. At the bottom of cavity 1, EM density suggests that the hydrophobic side chain of M549 forms strong Van der Waals interactions with all three substrates, possibly an interaction between the sulfur of the methionine side chain and the edges of the aromatic ring systems of the drugs. Additionally, Van der Waals interactions involve T542 (all substrates), V546 (mitoxantrone and tariquidar), and S440 and L405 (tariquidar) (for a full list see Table 4).

A recent study reported cryo-EM structures of ABCG2 bound to chemotherapeutic drugs (Orlando & Liao, 2020). Similar to our results, Orlando et al. observed an inward-facing

conformation and found density for bound drugs in cavity 1. However, the resolution of these previous structures was lower because no 5D3-Fab was used. While the general location of the drugs between the phenyl rings of F439 is similar, we find distinct orientations of bound mitoxantone and SN38 (depicted in Figure 21). The mitoxantrone molecule in our structure is rotated almost 90° compared to the previous study, and the topotecan molecule is flipped 180° to the structurally similar SN38 molecule in Orlando et al. Given our functional data, we can rule out a 5D3-Fab-induced artifact in our structures. The significant discrepancy may therefore be due the fact that the reported 3D volumes(Orlando & Liao, 2020) suffered from lower resolution and from C2-symmetrization.

3.3.4 Movement of drug substrates in the binding pocket

We investigated the conformational flexibility of drug-bound ABCG2 by performing 3D variability analyses(Punjani, Rubinstein, Fleet, & Brubaker, 2017). This allowed the visualization of movements of the substrates inside the drug-binding cavity. The main observation was that the substrates kept sliding inside the slit-like cavity (Videos 1-3). We also noted structural differences akin to vibrations in the NBDs. Intriguingly, we found a

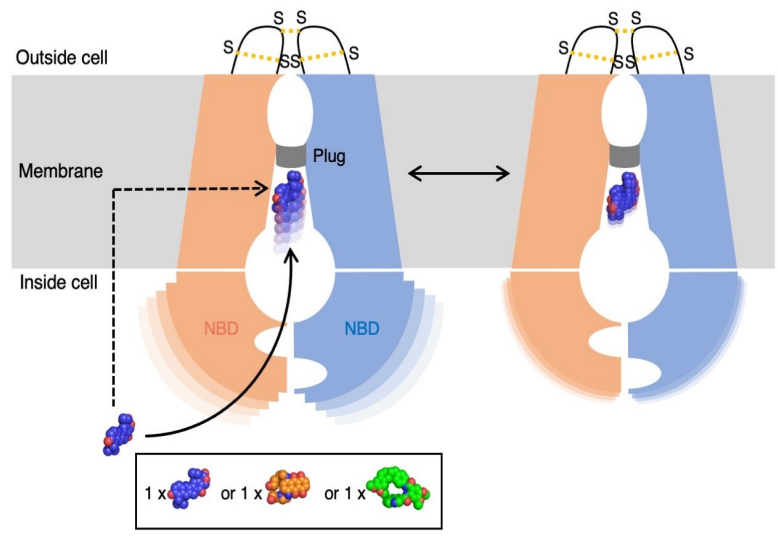


Figure 10 Effect of substrate binding on NBD structure. Schematic of ABCG2 binding the substrate molecule. The binding of single substrate molecule (blue spheres) in the center of cavity 1 is a continuous process which requires a lot of movements. Substrate molecule is neither static nor locked in the cavity 1. ABCG2 monomers are colored salmon and blue, disulfide bridges located at EL3 are indicated by yellow dashed lines connecting sulfur (S) atoms, and the leucine plug is shown as a gray bar between the cavities.

correlation between the density covering the NBDs and observed densities of substrates between the TMDs. When the NBDs are more separated, the EM density of substrate is not well resolved in cavity 1. Conversely, when the NBDs are closer together, the density for substrate is better defined (Videos 1-3). The mobility of the NBDs was particularly visible during 3D classifications, but a gradual selection of the most stable 3D class allowed refinement of the most stable conformation to high-resolution (Figure 13, Figure 15, Figure 17). We conclude that when substrate molecules access the binding pocket, the NBDs become ready to hydrolyze ATP and facilitate transport (Figure 10).

Upon comparing our EM maps, the most "mobile" NBDs and the least-well defined drug density were observed in the ABCG2-tariquidar structure. Tariquidar was reported to adopt at least three different conformations(Labrie et al., 2006). Both our Relion(Kimanius, Forsberg, Scheres, & Lindahl, 2016) and CryoSPARC(Punjani et al., 2017) EM maps (solved at 3.12 Å and 3.59 Å resolution, respectively) suggested that tariquidar adopt at least two distinct conformations (Figure 7, Figure 19, Figure 20), which are referred to as C- or U-conformation.

3.4 DISCUSSION

Among the endogenous substrates of ABCG2 are uric acid (168 kDa) and estrone-3-sulfate (E1S, 350 kDa). Both are relatively small, heterocyclic compounds that contain two or four aromatic rings, respectively. The slit-like substrate binding cavity in ABCG2 is well-suited to recruit these two compounds and sandwich them between the phenyl groups of the F439 side chains. We found that despite their larger sizes and more complex structures, the three anti-cancer drugs mitoxantrone, topotecan, and tariquidar interact with ABCG2 in a similar way. This can be rationalized because these drugs also contain flat, polycyclic ring scaffolds and therefore interact with the similar set of amino acids in ABCG2(Manolaridis et al., 2018).

Both topotecan and mitoxantrone intercalate between two DNA bases at the active sites of topoisomerases I or II, respectively, such that the enzymes are trapped in the cleavage complex bound to the DNA, ultimately leading to a cell death. When comparing the interactions of these anticancer drugs either with DNA or F439 residues in the ABCG2 binding pocket, we observed that the distance between base pairs and the drugs is about 0.5 Å smaller than between phenylalanine residues and the drugs (Figure 22). This can rationalize the much weaker interactions between drug substrates and ABCG2, which

allows the molecules to move within the binding pocket and, upon ATP binding to ABCG2, get transported.

Given that the two F439 residues can in principle clamp any of the scaffold rings of the substrates, it is not surprising that all three drugs revealed a certain flexibility in the binding site, which is reflected in the EM density that extends beyond the substrate structure in two dimensions, as defined by the F439 clamp. Our expectation had been that the higher the affinity of a substrate to ABCG2, the less flexibility there would be and therefore the better-defined an EM density we would observe. Contrary to our expectation, tariquidar revealed the least well-defined EM density despite displaying the highest affinity (EC50 for ATPase stimulation in the sub-micromolar range) of the three drugs. How can this observation be rationalized? Tariquidar (647 Da) is an inhibitor of ABCB1 and was initially assumed to inhibit other multidrug ABC transporters as well. However, it was later shown that ABCG2 can actually transport tariquidar, albeit at a very low rate(Kannan et al., 2011). Using proteolipsomes containing purified ABCG2, we have shown earlier(Jackson et al., 2018) that tariquidar can inhibit ABCG2-catalyzed E1S transport in vitro, which is compatible with tariquidar either being an inhibitor or a high-affinity but slowly transported substrate. We have been unable to demonstrate transport of radiolabeled tariquidar in our proteoliposome-based transport assays. Instead, derivatives of tariquidar such as MB136, URME22-1(Kuhnle et al., 2009) and HM30181 analogues(Kohler & Wiese, 2015) have been successfully developed as inhibitors of ABCG2. We have earlier shown that one of these compounds, MB136, rather than binding in a coiled conformation such as tariquidar, adopts an extended conformation, thereby generating additional contacts with residues V401 and L405' (Figure 11) from opposing ABCG2 monomers(Jackson et al., 2018). This suggests that the inhibitory characteristics of MB136 is associated with its extended conformation and interactions with the side walls of cavity 1, whereas the substrate-like characteristics of tariquidar are associated with the fact that it could assume a coiled, C- or U-shaped conformation when lodged in cavity 1.

Unlike inhibitors, which act as wedges to prevent ATP-induced NBD dimerization, substrates are compatible with the previously proposed, peristaltic extrusion mechanism through the center of the transporter(Manolaridis et al., 2018). Importantly, while the size of tariquidar is rather large for a substrate, its ability to adopt a C-shaped conformation therefore makes it a (slowly transported) substrate. If the classification of compounds into inhibitors or substrates is not a sharp distinction but rather represents a continuum, tariquidar

is an example of a compound around the middle of this continuum. This molecular interpretation could be useful in guiding the design of larger inhibitors that do not adopt a C-shape like tariquidar but maintain a set of interactions with residues located on the TM1b helices as observed in tariquidar derivatives(Antoni et al., 2020).

By performing 3D variability analyses, we obtained insight into the dynamics of ABCG2

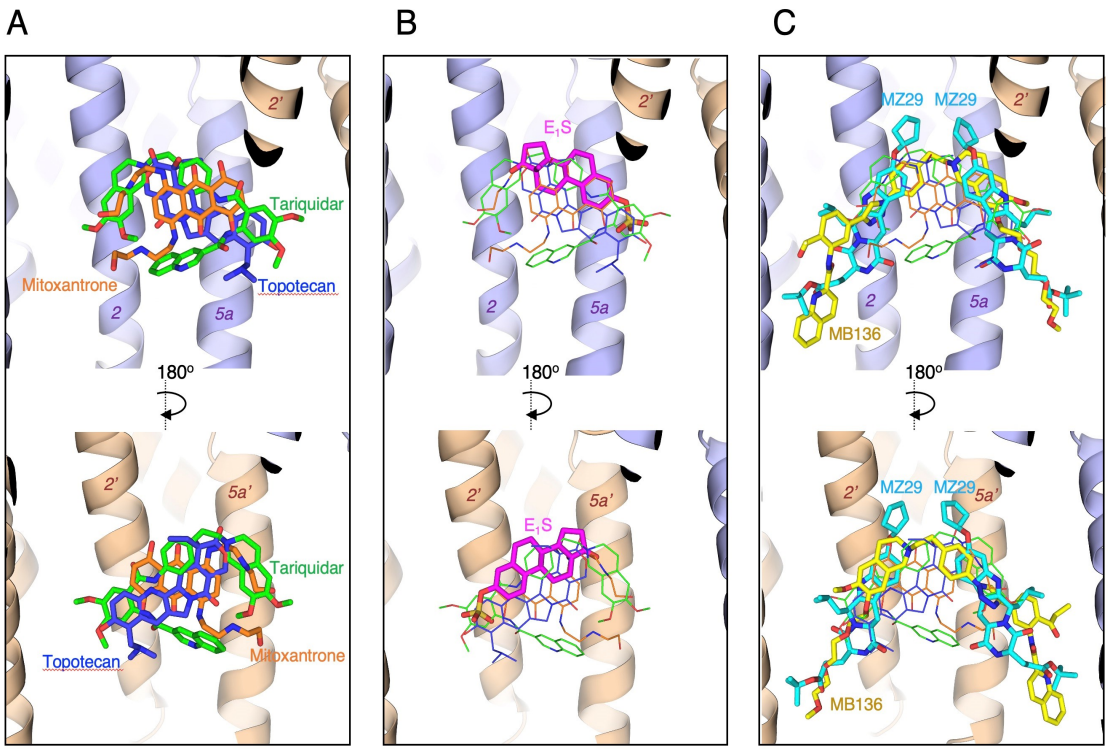


Figure 11 Substrates and inhibitors in cavity 1 of ABCG2. Overlay of bound ligands from different ABCG2 structures after superposition of the TMDs. Two membrane views are shown. TM helices are labeled 2, 2' and 5a, 5a'.

(A) Overlay of the three substrates topotecan, mitoxantrone and tariquidar (this work) shown as blue, orange and green sticks, respectively.

(B) Substrates from (A) shown as thin sticks and estrone 3-sulfate (E1S) (thick, magenta sticks, from PDB 6HCO).

(C) Overlay of substrates from (A) shown as thin sticks and ABCG2 inhibitors MZ29 (cyan sticks, from PDB 6ETI) and MB136 (yellow sticks, from PDB 6FEQ).

during substrate binding. The density features representing bound substrates among a continuous family of generated 3D structures (Videos 1-3) allowed us to conclude that the presence of substrate (i) alters the tilt of TM1, thereby slightly narrowing the entrance to the cavity, and (ii) correlates with an ordering of the NBDs, thereby providing a structural explanation of the observed ATPase stimulation in the presence of drugs (Figure 11). We hypothesize that similar effects may take place in other ABC-transporters.

The presented ABCG2-drug structures explain the structural basis of substrate binding in ABCG2. However, despite such progress, further improvements in obtaining high resolution and well-defined ligand densities as well as markedly increased structure determination throughput would be required to make thorough structure-based drug design campaigns viable. Even at this advanced stage of structural studies, we conclude that total stabilization of a substrate molecule in the ABCG2 binding pocket is questionable. However, by understanding the requirements of substrates and how they differ from those of inhibitory compounds might make campaigns at designing better ABCG2 inhibitors successful.

3.5 METHODS

3.5.1 Expression and purification of wild-type human ABCG2 and 5D3-Fab

Human ABCG2, containing an amino (N)-terminal Flag tag, was expressed in HEK293-EBNA (Thermo Fisher Scientific) cells and purified as described previously(Jackson et al., 2018; Manolaridis et al., 2018; Taylor et al., 2017).

5D3 hybridoma cells, producing the 5D3 monoclonal antibody, were obtained from B. Sorrentino. The cells were cultured in WHEATON CELLine Bioreactors, according to the manufacturer's protocol, and 5D3-Fab was then purified from the supernatant, as described in the Fab Preparation Kit protocol (Thermo Fisher Scientific).

3.5.2 Nanodisc preparation of ABCG2

The membrane scaffold protein (MSP) 1D1 was expressed and purified and ABCG2 was reconstituted into brain polar lipid (BPL)/cholesterol hemisuccinate (CHS) nanodiscs as described previously(Jackson et al., 2018; Manolaridis et al., 2018; Taylor et al., 2017). To generate the ABCG2-Substrate samples for cryo-EM studies, ABCG2 was first mixed with a threefold molar excess of 5D3-Fab before reconstitution. After size-exclusion chromatography (SEC) using a Superdex 200 10/300 column (GE Healthcare), the ABCG2-Fab complex was incubated with 100 µM topotecan, 150 µM mitoxantrone or 1 µM tariquidar for 10 min at room temperature before cryo-EM grids preparation.

3.5.3 ABCG2-liposome preparation

A BPL/cholesterol (BPL/chol) (Avanti Polar Lipids) mixture was prepared at a 4:1 (w/w) ratio as described previously(Geertsma, Nik Mahmood, Schuurman-Wolters, & Poolman, 2008). Briefly, the BPL/chol mixture was extruded through a 400-nm polycarbonate filter and destabilized with 0.17% (v/v) Triton X100. Detergent-purified ABCG2 was then mixed

with BPL/chol at a 100:1 (w/w) lipid/protein ratio. Detergent was removed with BioBeads, and proteoliposomes were spun at 100,000g, resuspended in buffer 25 mM HEPES, pH 7.5, 150 mM NaCl at a final lipid concentration of 10 mg/ml, and the reconstitution efficiency was determined(Schaffner & Weissmann, 1973).

3.5.4 ATPase assays and determination of the EC50 of mitoxantrone, topotecan or tariquidar stimulation

ATP-hydrolysis activity was measured using a previously described technique(Chifflet, Torriglia, Chiesa, & Tolosa, 1988). Proteoliposome samples were pre-incubated with the specified concentrations of mitoxantrone, topotecan or tariquidar for 10 min at 37 °C prior to the addition of 2 mM ATP and 10 mM MgCl2 to start the hydrolysis reaction. To assess the effect of 5D3-Fab, ABCG2 proteoliposomes were freeze-thawed five times in the presence of a three-fold molar excess of 5D3-Fab to ensure that 5D3-Fab was inside the proteoliposomes before extrusion(Jackson et al., 2018; Taylor et al., 2017). Data were recorded at four-time intervals (0, 5, 15 and 30 min) and subsequent ATPase rates were determined using linear regression in GraphPad Prism 7.00. Rates were corrected for the orientation of ABCG2 in proteoliposomes(Taylor et al., 2017). To determine the EC50 of mitoxantrone, topotecan and tariquidar stimulation, we plotted the ATPase rates against the substrate concentration, and generated curves using the nonlinear regression Michaelis–Menten analysis tool in GraphPad Prism 7.00.

3.5.5 Sample preparation and cryo-EM data acquisition

All cryo-EM grids of wild type ABCG2 with topotecan, mitoxantrone or tariquidar were prepared with Vitrobot mark IV (FEI) with the environmental chamber set at 100% humidity and 4°C. Aliquots of 4 µL of purified ABCG2-Fab complexes with substrates at a protein concentration of approximately 0.5 mg/mL were placed onto Quantifoil carbon grids (R1.2/1.3, 300 mesh, copper) that had been glow-discharged for 45 seconds at 25 mA using Pelco easiGlow 91000 Glow Discharge Cleaning System. Grids were blotted for 1.5-2.0 s and flash-frozen in a mixture of liquid ethane and propane cooled by liquid nitrogen.

The final data set of ABCG2-tariquidar-Fab was composed of 5894 super-resolution movies. The ABCG2-tariquidar-Fab grids were imaged with a Titan Krios (FEI) electron microscope operated at 300 kV, equipped with a Gatan K3 direct electron detector and Gatan Quantum-LS energy filter (GIF), with a slit width of 20 eV to remove inelastically

scattered electrons. Movies were recorded semi-automatically with EPU software (Thermo Fisher Co.), in super-resolution counting mode with a defocus range of –0.4 to –2.5 µm and a super-resolution pixel size of 0.33 Å/pixel. Movies were 1.0 s long, dose-fractionated into 40 frames, with a frame exposure rate of 2.0 e–/Å2. All stacks were gain-normalized, motion-corrected, dose-weighted and then binned 2-fold with MotionCor2(Zheng et al., 2017). The defocus values were estimated on the non-dose-weighted micrographs with Gctf(Zhang, 2016).

The ABCG2-topotecan-Fab dataset was composed of 3403 movies. The movies were collected on Titan Krios microscope (300kV), equipped with a Gatan Quantum-LS energy filter (20 eV zero loss filtering) on a K2 Summit detector with the automation of SerialEM(Mastronarde, 2005). Movies were recorded in counting mode with a physical pixel size of 0.82 Å/pixel, in a defocus range from -0.8 to -2.8 µm, with a frame exposure rate of 1.55 e–/Å2 and 40 frames for each movie (8s exposures). All the movie stacks were gain-normalized, motion-corrected and dose-weighted with the program MotionCor2(Zheng et al., 2017).

The ABCG2-mitoxantrone-Fab data was composed of 6192 movies in total. The movie data was collected on a Titan Krios microscope (300kV) with an automation software SerialEM(Mastronarde, 2005), equipped with a Gatan Quantum-LS energy filter (20 eV zero loss filtering). The movies were recorded with a K2 Summit direct electron detector (counting mode) with a pixel size of 0.64 Å/pixel. Movies were recorded in a defocus range from -0.6 to -2.5 µm, with a frame exposure rate of 1.50 e–/Å2 (40 frames for each movie, 8s in total). All the movie stacks were gain-normalized, motion-corrected and dose-weighted with the program MotionCor2(Zheng et al., 2017). The micrographs were sorted according to different parameters such as: Iciness, Sample Drift, Defocus, Astigmatism and Resolution of CTF Fit (in Focus(Biyani et al., 2017)), in order to remove the bad images.

3.5.6 Image processing

Image processing of the ABCG2-tariquidar-Fab complex was performed both in Relion 3.1(Kimanius et al., 2016) and CryoSPARC2(Punjani et al., 2017) software. From the 5,894 selected dose-weighted micrographs, a total of 1,326,686 particles were picked within Relion. After five rounds of 2D classifications and selection, we obtained a final set of 737,525 particles which were 3D classified within 3 classes using as reference a low-passed

initial model obtained from the small dataset processed in CryoSPARC2 (ab-initio reconstruction). The particles from the best 3D class, containing 644,542 (87.4%) of particles, were re-extracted (384 x 384 box) without binning, and 3D refined with soft mask covering the full density of the complex. The resolution of this map was 3.62 Å. Subsequently, the series of 3D refinements including CTF refinement with anisotropic magnification correction and beam-tilt correction, per-particle CTF refinement, Bayesian polishing and refinement with a soft mask without detergent belt and constant domain of Fab were performed and yielded the 3.27 Å resolution map. To further improve the map resolution, the particles were applied for another 3D classification without alignments, without a mask. At this point we noticed that only 27% (166,727) of particles contributed to the 3D class with stable NBDs. This class was further 3D refined in the Relion with an angular sampling rate of 0.5 degree and local searches set up to 1.8 degrees, and with C1 or C2 symmetry applied (maps at 3.12 Å and 2.98 Å resolution, respectively, with automatically determined Bfactors of −74 Å2 and −77.5 Å2). In parallel, the same set of particles was processed in CryoSPARC v2.14.2 where 2D classification and homogeneous 3D refinement (map at 3.59 Å resolution, with B-factor adjusted to −100 Å2), and 3D variability analyses were performed.

All of aligned movies of the ABCG2-topotecan-Fab and ABCG2-mitoxantrone-Fab were prepared in Focus(Biyani et al., 2017) and then imported into CryoSPARC2(Punjani et al., 2017). The CTF and defocus values were estimated on the dose-weighted micrographs. For the ABCG2-topotecanFab sample, to generate a template for particles picking, 3000 particles were picked manually and 2D classified. The second round of particle picking was performed using the templates and resulted in 775,993 particles. After two rounds of 2D classifications 417,949 particles were selected. The 3D template was created using the ab-initio 3D reconstruction (with C1 symmetry, similarity 0.1). The initial 3D classification was followed by 3D refinement of the best 3D class containing 203,789 particles. The resulted map (C1 symmetry) had an overall 3.50 Å resolution. Because of the possible heterogeneity within the ABCG2-topotecan-Fab sample another round of ab-initio 3D reconstruction (similarity 0.65) plus 3D hetero-refinement with three classes were performed and the final 3D class contained 133,832 particles. The final subset was subjected to 3D refinement jobs (Homogeneous Refinement > Non-uniform Refinement > Local Refinement) applying C1 and C2 symmetry (maps at 3.39 Å and 3.14 Å resolution, respectively).

For ABCG2-mitoxantrone sample, a similar processing was performed as described above. However, the main problem which we were facing during the processing, was the misalignment of the particles resulting in an averaged fake density. Therefore, we performed two independent ab-initio 3D reconstruction jobs, with the similarities of 0.1 for C2 map, in order to achieve better resolution, and of 0.65 for C1 map to determine the substrate density located at the two-fold symmetry axis of the transporter, followed by the heterogenous 3D refinements. Final refinements resulted in EM density maps at 3.51 Å and 3.35 Å resolution, for C1- and C2-symmetrized maps, respectively.

The particle stacks contributing into best 3D maps of ABCG2 with drugs topotecan, mitoxantrone and tariquidar were used for further 3D Variability Analysis, which is a tool in CryoSPARC 2.9 for exploring both discrete and continuous heterogeneity in single particle cryo-EM data sets. In order to remove the signal coming from the micelle the masks without detergent belt made in Chimera were used. The simple linear "movies" of volumes were generated for 3 eigenvectors. Outputs were low-pass filtered during optimization to 4 Å and visualized with 3D Variability Display tool in CryoSPARC 2.9.

3.5.7 Model building and refinement

For the generation of an initial model of human ABCG2-tariquidar-Fab, we used a post-processed non-symmetrized map from Relion at an overall resolution of 3.12 Å. Guided by the structure of ABCG2-MB136-Fab (PDB: 6FEQ)(Jackson et al., 2018), we performed manual building and fitting using the software Coot(Emsley, Lohkamp, Scott, & Cowtan, 2010). The EM density was of excellent quality in the transmembrane region and variable domain of the antibody region and allowed the unambiguous building of ABCG2. In NBD region, we carried out manual fitting and modifications where the resolution allowed. Subsequently, the manual fitting of the tariquidar molecule into the density was performed. Four possible conformations of tariquidar in the EM density were saved. In parallel, the initial ABCG2-tariquidar-Fab model was fitted to the 3.59 Å map generated in CryoSPARC2. Further fitting of the tariquidar molecule into CryoSPARC2 EM density was performed in Coot(Emsley et al., 2010) and two possible conformations of tariquidar were found. Geometry restraints for tariquidar, cholesterol and phospholipid molecules were generated in eLBOW(Moriarty, Grosse-Kunstleve, & Adams, 2009). The complete models

were refined against the working maps in PHENIX(Adams et al., 2010) using real space refinement. The quality of the final models was analyzed by MolProbity.

For the ABCG2-topotecan and -mitoxantrone structures model building, we used the previous ABCG2EQ-estrone-3-sulfate structure (PDB 6HCO) as the initial template. The PDB files were fitted into the resolved EM map low-passed to 6 Å in Chimera. The model building was performed manually in Coot. Geometry restraints for mitoxantrone and topotecan molecules were downloaded from PDB. After rebuilding, the atomic model was refined in PHENIX(Adams et al., 2010) using the option real space refinement with a standard parameter for geometry minimization including the global real-space refinement, Non-crystallographic symmetry (NCS) and secondary structure, Ramachandran plot and Rotamer restraints. The complete models were refined against the working maps in PHENIX(Adams et al., 2010) using real space refinement.

3.5.8 Figure preparation

Figures were prepared using the programs PyMOL (PyMOL Molecular Graphics System, DeLano Scientific), ChimeraX and GraphPad Prism 7.00 (GraphPad Software).

3.5.9 Data availability

Atomic coordinates for ABCG2–topotecan-Fab, ABCG2–mitoxantrone-Fab and ABCG2-tariquidar-Fab (including only the variable domain of 5D3-Fab) were deposited in the Protein Data Bank under accession codes XXXX, YYYY and ZZZZ, respectively. Electron microscopy data for the three structures were deposited in the Electron Microscopy Data Bank under accession codes EMD-XXXX (ABCG2–topotecan-Fab), EMD-YYYY (ABCG2–mitoxantrone-Fab) and EMD-ZZZZ (ABCG2–tariquidar-Fab).

3.5.10 Acknowledgements

This work was in part supported by the Swiss National Science Foundation (NCCR TransCure).

3.5.11 Author contributions

I.M expressed and purified ABCG2 and 5D3-Fab. S.M.J performed ATPase activity assays and proteoliposomes experiments. S.M.J and I.M. reconstituted ABCG2 into liposomes and lipidic nanodiscs. J.K prepared all cryo-grids and with I.M. collected cryo-EM data for

ABCG2tariquidar-Fab. J.K determined the structure of ABCG2-tariquidar-Fab. D.N. and H.S collected cryo-EM data and determined the structures of ABCG2-topotecan-Fab and ABCG2-mitoxantroneFab. J.K. and K.P.L refined and validated the structure ABCG2-tariquidar-Fab. D.N. refined and validated the structures ABCG2-topotecan-Fab and ABCG2-mitoxantrone-Fab. K.P.L, J.K., S.M.J and I.M conceived the project. K.P.L, J.K., S.M.J and I.M planned the experiments. J.K. and K.P.L wrote the manuscript; all authors contributed to revisions.

3.5.12 Competing financial interests

The authors declare no competing financial interests.

3.6 References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr, 66(Pt 2), 213-221. doi:10.1107/S0907444909052925

Alam, A., Kowal, J., Broude, E., Roninson, I., & Locher, K. P. (2019). Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. Science, 363(6428), 753-756. doi:10.1126/science.aav7102

Alam, A., Kung, R., Kowal, J., McLeod, R. A., Tremp, N., Broude, E. V., . . . Locher, K. P. (2018). Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1. Proc Natl Acad Sci U S A, 115(9), E1973-E1982. doi:10.1073/pnas.1717044115

Antoni, F., Bause, M., Scholler, M., Bauer, S., Stark, S. A., Jackson, S. M., . . . Bernhardt, G. (2020). Tariquidar-related triazoles as potent, selective and stable inhibitors of ABCG2 (BCRP). Eur J Med Chem, 191, 112133. doi:10.1016/j.ejmech.2020.112133

Bakhsheshian, J., Hall, M. D., Robey, R. W., Herrmann, M. A., Chen, J. Q., Bates, S. E., & Gottesman, M. M. (2013). Overlapping substrate and inhibitor specificity of human and murine ABCG2. Drug Metab Dispos, 41(10), 1805-1812. doi:10.1124/dmd.113.053140

Bauer, S., Ochoa-Puentes, C., Sun, Q., Bause, M., Bernhardt, G., Konig, B., & Buschauer, A. (2013). Quinoline carboxamide-type ABCG2 modulators: indole and quinoline moieties as anilide replacements. ChemMedChem, 8(11), 1773-1778. doi:10.1002/cmdc.201300319

Begley, D. J. (2004). ABC transporters and the blood-brain barrier. Curr Pharm Des, 10(12), 1295-1312. doi:10.2174/1381612043384844

Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N., & Stahlberg, H. (2017). Focus: The interface between data collection and data processing in cryo-EM. J Struct Biol, 198(2), 124-133. doi:10.1016/j.jsb.2017.03.007

Chifflet, S., Torriglia, A., Chiesa, R., & Tolosa, S. (1988). A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: application to lens ATPases. Anal Biochem, 168(1), 1-4. doi:10.1016/0003-2697(88)90002-4

Crowley, E., McDevitt, C. A., & Callaghan, R. (2010). Generating inhibitors of P-glycoprotein: where to, now? Methods Mol Biol, 596, 405-432. doi:10.1007/978-1-60761-416-6_18

Dauchy, S., Dutheil, F., Weaver, R. J., Chassoux, F., Daumas-Duport, C., Couraud, P. O., . . . Decleves, X. (2008). ABC transporters, cytochromes P450 and their main transcription factors: expression at the human blood-brain barrier. J Neurochem, 107(6), 1518-1528. doi:10.1111/j.1471-4159.2008.05720.x

Diestra, J. E., Scheffer, G. L., Catala, I., Maliepaard, M., Schellens, J. H., Scheper, R. J., . . . Izquierdo, M. A. (2002). Frequent expression of the multi-drug resistance-associated protein BCRP/MXR/ABCP/ABCG2 in human tumours detected by the BXP-21 monoclonal antibody in paraffin-embedded material. J Pathol, 198(2), 213-219. doi:10.1002/path.1203

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr, 66(Pt 4), 486-501. doi:10.1107/S0907444910007493

Evison, B. J., Sleebs, B. E., Watson, K. G., Phillips, D. R., & Cutts, S. M. (2016). Mitoxantrone, More than Just Another Topoisomerase II Poison. Med Res Rev, 36(2), 248-299. doi:10.1002/med.21364

Fetsch, P. A., Abati, A., Litman, T., Morisaki, K., Honjo, Y., Mittal, K., & Bates, S. E. (2006). Localization of the ABCG2 mitoxantrone resistance-associated protein in normal tissues. Cancer Lett, 235(1), 84-92. doi:10.1016/j.canlet.2005.04.024

Geertsma, E. R., Nik Mahmood, N. A., Schuurman-Wolters, G. K., & Poolman, B. (2008). Membrane reconstitution of ABC transporters and assays of translocator function. Nat Protoc, 3(2), 256-266. doi:10.1038/nprot.2007.519

Ghosh, C., Puvenna, V., Gonzalez-Martinez, J., Janigro, D., & Marchi, N. (2011). Blood-brain barrier P450 enzymes and multidrug transporters in drug resistance: a synergistic role in neurological diseases. Curr Drug Metab, 12(8), 742-749. doi:10.2174/138920011798357051

Hernandez Lozano, I., Bauer, M., Wulkersdorfer, B., Traxl, A., Philippe, C., Weber, M., . . . Langer, O. (2020). Measurement of Hepatic ABCB1 and ABCG2 Transport Activity with [(11)C]Tariquidar and PET in Humans and Mice. Mol Pharm, 17(1), 316-326. doi:10.1021/acs.molpharmaceut.9b01060

Imai, Y., Asada, S., Tsukahara, S., Ishikawa, E., Tsuruo, T., & Sugimoto, Y. (2003). Breast cancer resistance protein exports sulfated estrogens but not free estrogens. Mol Pharmacol, 64(3), 610-618. doi:10.1124/mol.64.3.610

International Transporter, C., Giacomini, K. M., Huang, S. M., Tweedie, D. J., Benet, L. Z., Brouwer, K. L., . . . Zhang, L. (2010). Membrane transporters in drug development. Nat Rev Drug Discov, 9(3), 215-236. doi:10.1038/nrd3028

Ishikawa, T., Aw, W., & Kaneko, K. (2013). Metabolic Interactions of Purine Derivatives with Human ABC Transporter ABCG2: Genetic Testing to Assess Gout Risk. Pharmaceuticals (Basel), 6(11), 1347-1360. doi:10.3390/ph6111347

Jackson, S. M., Manolaridis, I., Kowal, J., Zechner, M., Taylor, N. M. I., Bause, M., . . . Locher, K. P. (2018). Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. Nat Struct Mol Biol, 25(4), 333-340. doi:10.1038/s41594-018-0049-1

Johnson, Z. L., & Chen, J. (2017). Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. Cell, 168(6), 1075-1085 e1079. doi:10.1016/j.cell.2017.01.041

Kannan, P., Telu, S., Shukla, S., Ambudkar, S. V., Pike, V. W., Halldin, C., . . . Hall, M. D. (2011). The "specific" P-glycoprotein inhibitor Tariquidar is also a substrate and an inhibitor for breast cancer resistance protein (BCRP/ABCG2). ACS Chem Neurosci, 2(2), 82-89. doi:10.1021/cn100078a

Kimanius, D., Forsberg, B. O., Scheres, S. H., & Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife, 5. doi:10.7554/eLife.18722

Kohler, S. C., & Wiese, M. (2015). HM30181 Derivatives as Novel Potent and Selective Inhibitors of the Breast Cancer Resistance Protein (BCRP/ABCG2). J Med Chem, 58(9), 3910-3921. doi:10.1021/acs.jmedchem.5b00188

Kuhnle, M., Egger, M., Muller, C., Mahringer, A., Bernhardt, G., Fricker, G., . . . Buschauer, A. (2009). Potent and selective inhibitors of breast cancer resistance protein (ABCG2) derived from the p-glycoprotein (ABCB1) modulator tariquidar. J Med Chem, 52(4), 1190-1197. doi:10.1021/jm8013822

Labrie, P., Maddaford, S. P., Fortin, S., Rakhit, S., Kotra, L. P., & Gaudreault, R. C. (2006). A comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of anthranilamide derivatives that are multidrug resistance modulators. J Med Chem, 49(26), 7646-7660. doi:10.1021/jm060239b

Mahringer, A., & Fricker, G. (2016). ABC transporters at the blood-brain barrier. Expert Opin Drug Metab Toxicol, 12(5), 499-508. doi:10.1517/17425255.2016.1168804

Manolaridis, I., Jackson, S. M., Taylor, N. M. I., Kowal, J., Stahlberg, H., & Locher, K. P. (2018). Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. Nature, 563(7731), 426-430. doi:10.1038/s41586-018-0680-3

Mao, Q., & Unadkat, J. D. (2015). Role of the breast cancer resistance protein (BCRP/ABCG2) in drug transport--an update. AAPS J, 17(1), 65-82. doi:10.1208/s12248-014-9668-6

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol, 152(1), 36-51. doi:10.1016/j.jsb.2005.07.007

Mo, W., & Zhang, J. T. (2012). Human ABCG2: structure, function, and its role in multidrug resistance. Int J Biochem Mol Biol, 3(1), 1-27. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22509477

Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallogr D Biol Crystallogr, 65(Pt 10), 1074-1080. doi:10.1107/S0907444909029436

Ochoa-Puentes, C., Bauer, S., Kuhnle, M., Bernhardt, G., Buschauer, A., & Konig, B. (2013). Benzanilide-Biphenyl Replacement: A Bioisosteric Approach to Quinoline Carboxamide-Type ABCG2 Modulators. ACS Med Chem Lett, 4(4), 393-396. doi:10.1021/ml4000832

Orlando, B. J., & Liao, M. (2020). ABCG2 transports anticancer drugs via a closed-to-open switch. Nat Commun, 11(1), 2264. doi:10.1038/s41467-020-16155-2

Pick, A., Klinkhammer, W., & Wiese, M. (2010). Specific inhibitors of the breast cancer resistance protein (BCRP). ChemMedChem, 5(9), 1498-1505. doi:10.1002/cmdc.201000216

Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. Nat Rev Cancer, 6(10), 789-802. doi:10.1038/nrc1977

Puentes, C. O., Hocherl, P., Kuhnle, M., Bauer, S., Burger, K., Bernhardt, G., . . . Konig, B. (2011). Solid phase synthesis of tariquidar-related modulators of ABC transporters preferring breast cancer resistance protein (ABCG2). Bioorg Med Chem Lett, 21(12), 3654-3657. doi:10.1016/j.bmcl.2011.04.094

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods, 14(3), 290-296. doi:10.1038/nmeth.4169

Robey, R. W., To, K. K., Polgar, O., Dohse, M., Fetsch, P., Dean, M., & Bates, S. E. (2009). ABCG2: a perspective. Adv Drug Deliv Rev, 61(1), 3-13. doi:10.1016/j.addr.2008.11.003

Roe, M., Folkes, A., Ashworth, P., Brumwell, J., Chima, L., Hunjan, S., . . . Charlton, P. (1999). Reversal of P-glycoprotein mediated multidrug resistance by novel anthranilamide derivatives. Bioorg Med Chem Lett, 9(4), 595-600. doi:10.1016/s0960-894x(99)00030-x

Schaffner, W., & Weissmann, C. (1973). A rapid, sensitive, and specific method for the determination of protein in dilute solution. Anal Biochem, 56(2), 502-514. doi:10.1016/0003-2697(73)90217-0

Stacy, A. E., Jansson, P. J., & Richardson, D. R. (2013). Molecular pharmacology of ABCG2 and its role in chemoresistance. Mol Pharmacol, 84(5), 655-669. doi:10.1124/mol.113.088609

Stierand, K., & Rarey, M. (2010). Drawing the PDB: Protein-Ligand Complexes in Two Dimensions. ACS Med Chem Lett, 1(9), 540-545. doi:10.1021/ml100164p

Szakacs, G., Paterson, J. K., Ludwig, J. A., Booth-Genthe, C., & Gottesman, M. M. (2006). Targeting multidrug resistance in cancer. Nat Rev Drug Discov, 5(3), 219-234. doi:10.1038/nrd1984

Taylor, N. M. I., Manolaridis, I., Jackson, S. M., Kowal, J., Stahlberg, H., & Locher, K. P. (2017). Structure of the human multidrug transporter ABCG2. Nature, 546(7659), 504-509. doi:10.1038/nature22345

Thomas, C., & Tampe, R. (2020). Structural and Mechanistic Principles of ABC Transporters. Annu Rev Biochem, 89, 605-636. doi:10.1146/annurev-biochem-011520-105201

Vlaming, M. L., Lagas, J. S., & Schinkel, A. H. (2009). Physiological and pharmacological roles of ABCG2 (BCRP): recent findings in Abcg2 knockout mice. Adv Drug Deliv Rev, 61(1), 14-25. doi:10.1016/j.addr.2008.08.007

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J Struct Biol, 193(1), 1-12. doi:10.1016/j.jsb.2015.11.003

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods, 14(4), 331-332. doi:10.1038/nmeth.4193

3.7 SUPPLEMENTARY FIGURES

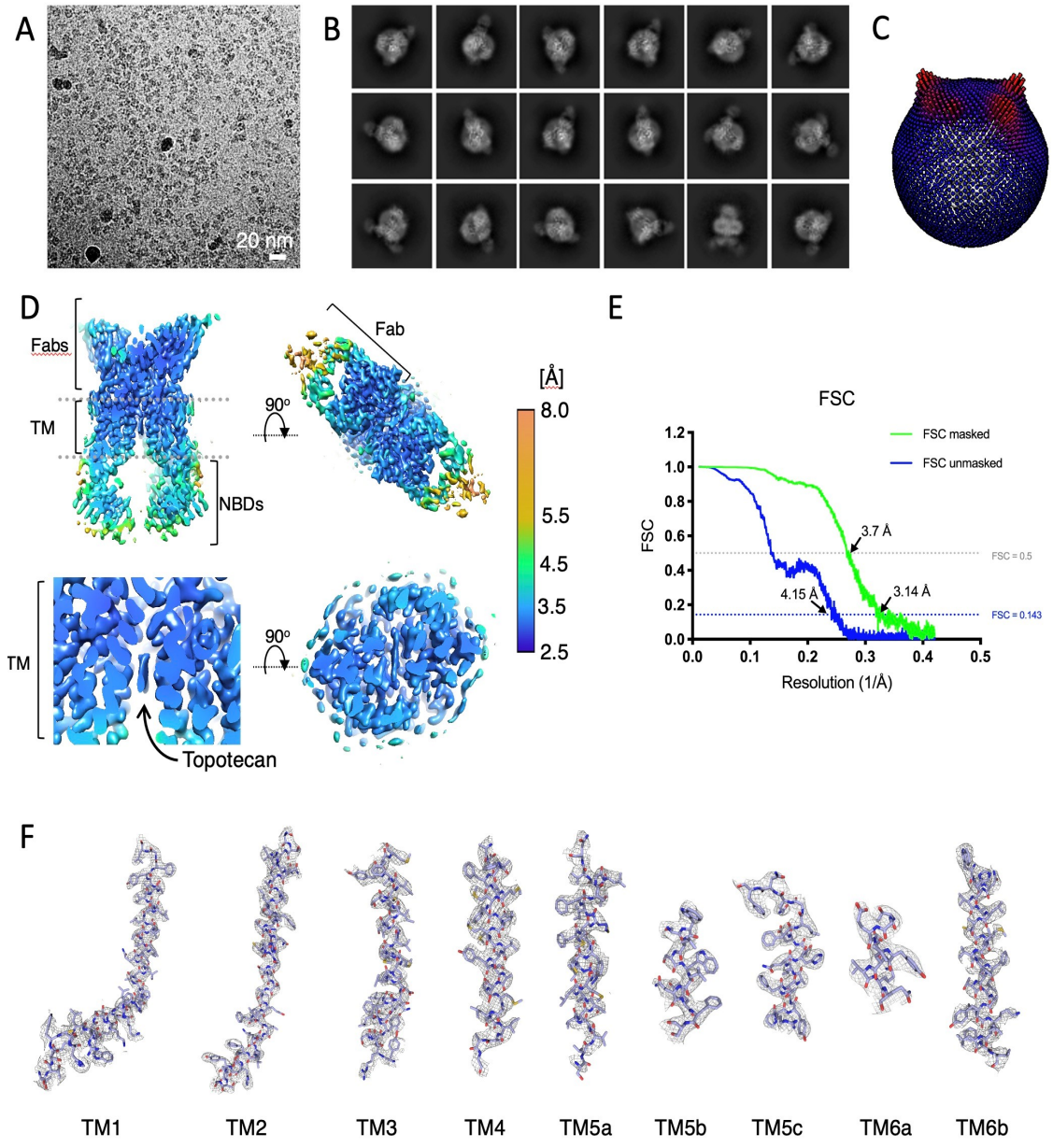


Figure 12 Cryo-EM map of ABCG2-topotecan-Fab generation and resolution estimation. (A) An example micrograph (drift-corrected, dose-weighted, and low-pass filtered) of the nanodisc-reconstituted ABCG2-topotecan-Fab data set. White scale bar, 20 nm. (B) Eighteen representative 2D class averages of the final round of 2D classification, sorted in decreasing order by the number of particles assigned to each class. (C) Angular distribution plot for the final reconstruction. (D) Full view of the final CryoSPARC B-factor-sharpened map of ABCG2-topotecan-Fab coloured by local resolution in Å, with the clipping plane in the middle of the molecule. On the top right corner is shown 5D3-Fabs region, in the bottom line the TM region with topotecan substrate density. (E) FSC from the CryoSPARC2 auto-refine procedure of the C2symmetrized unmasked half-maps (blue) and the half-maps after

masking (green). Horizontal dotted lines (blue and black) are drawn for the FSC = 0.143 and FSC = 0.5 criterion, respectively.

For both the unmasked and the masked FSC curves, their intersection with the FSC = 0.143 and the FSC = 0.5 lines are marked by arrows, and the resolutions at these points are indicated. (F) Fitting of the all TM helices of ABCG2 in the EM density map. A region of up to 2 Å around the atoms is shown. TM helices are shown as sticks. Density is shown as a grey mesh.

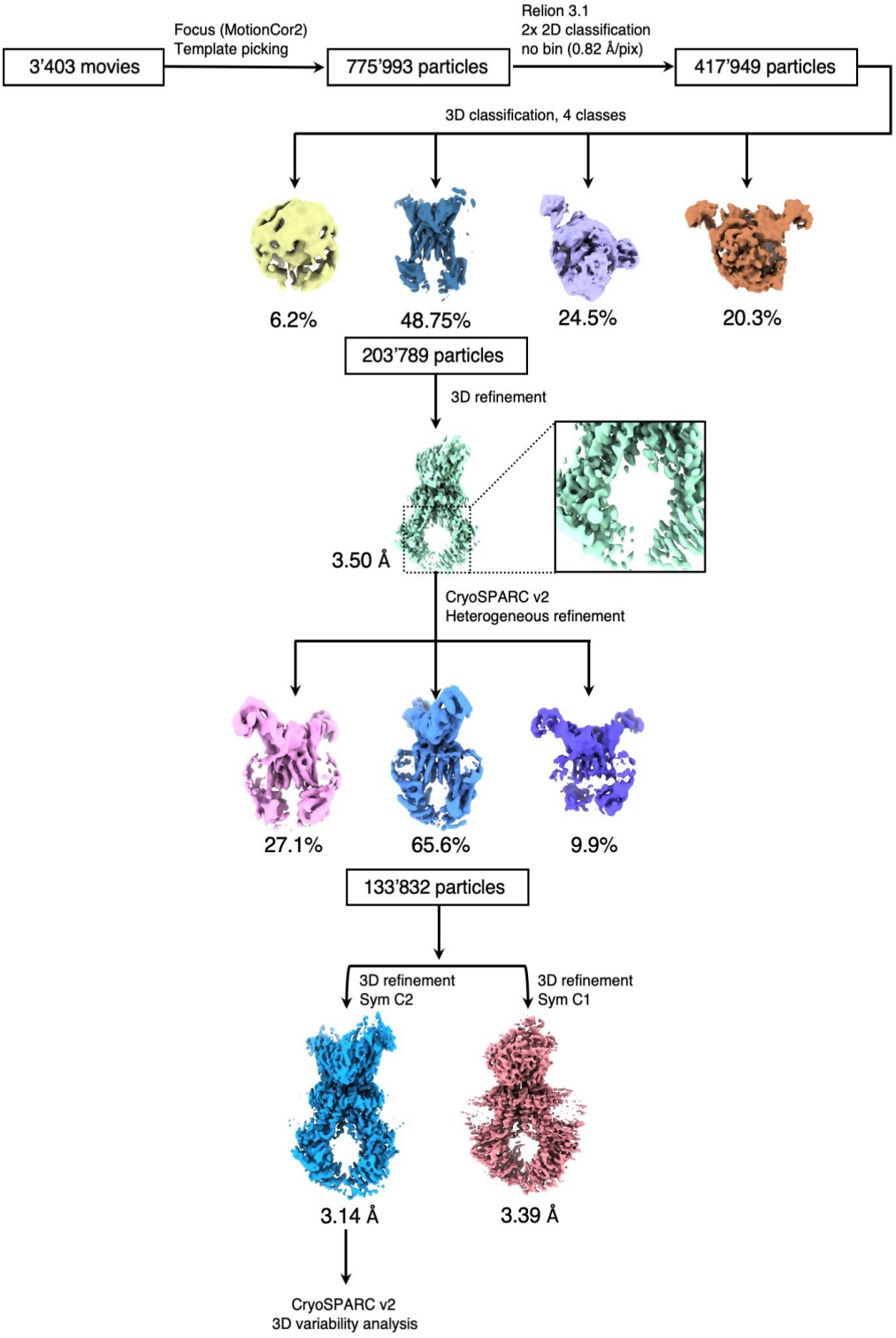


Figure 13 Flow chart for cryo-EM data processing and structure determination of the ABCG2-topotecan-Fab complex.

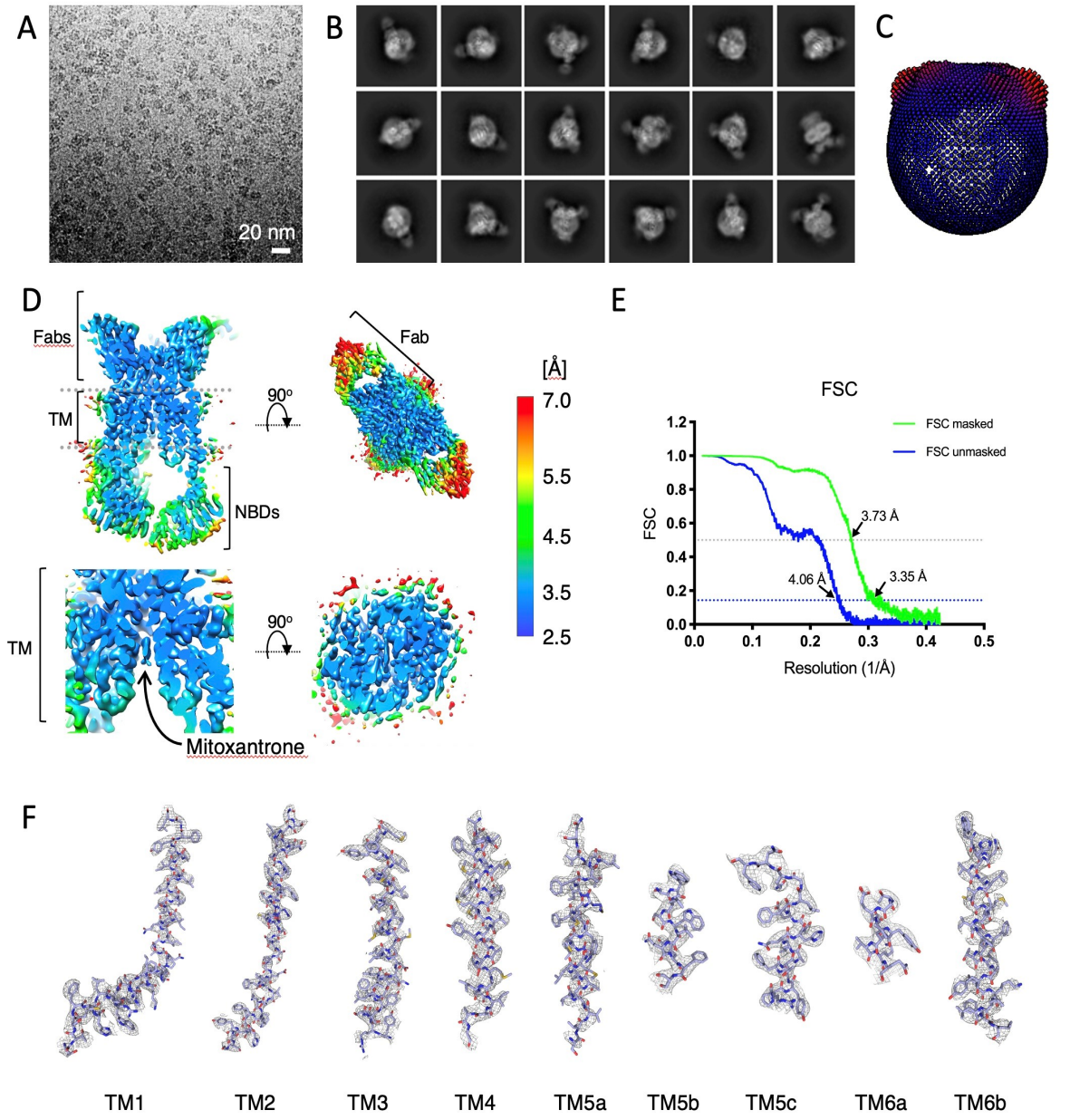


Figure 14 Cryo-EM map of ABCG2-mitoxantrone-Fab generation and resolution estimation. (A) An example micrograph (drift-corrected, dose-weighted, and lowpass filtered) of the nanodisc-reconstituted ABCG2-mitoxantrone-Fab data set. White scale bar, 20 nm. (B) Eighteen representative 2D class averages of the final round of 2D classification, sorted in decreasing order by the number of particles assigned to each class. (C) Angular distribution plot for the final reconstruction. (D) Full view of the final CryoSPARC B-factor-sharpened map of ABCG2-mitoxantrone-Fab coloured by local resolution in Å, with the clipping plane in the middle of the molecule. On the top right corner is shown 5D3-Fabs region, in the bottom line the TM region with mitoxantrone substrate density. (E) FSC from the CryoSPARC2 auto-refine procedure of the C2-symmetrized unmasked half-maps (blue) and the half-maps after masking (green). Horizontal dotted lines (blue and black) are drawn for the FSC = 0.143 and FSC = 0.5 criterion, respectively. For both the unmasked and the masked FSC curves, their intersection with the FSC = 0.143 and the FSC = 0.5 lines are marked by arrows, and the resolutions at these points are indicated. (F) Fitting of the all TM helices of ABCG2 in the EM density map. A region of up to 2 Å around the atoms is shown. TM helices are shown as sticks. Density is shown as a grey mesh.

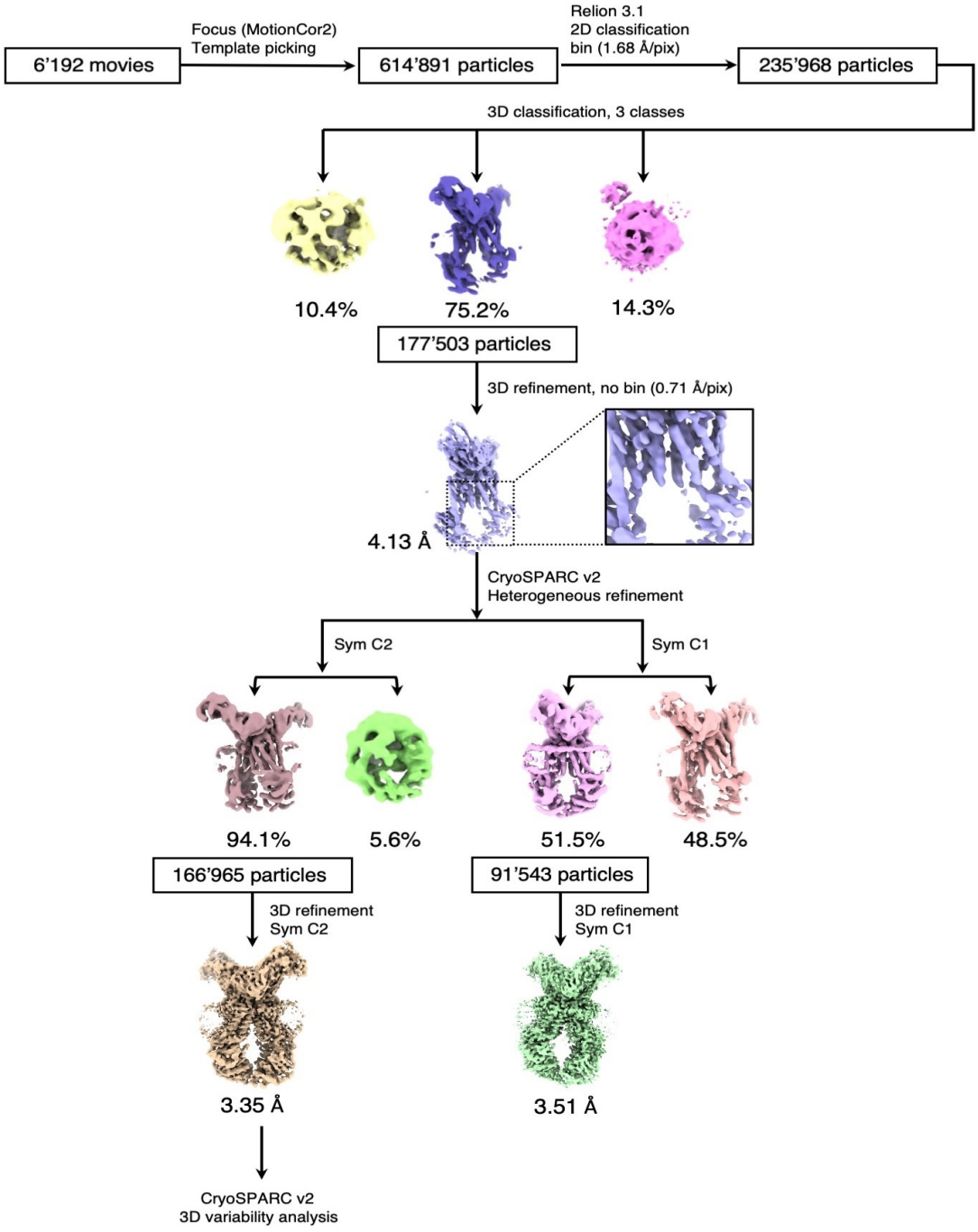


Figure 15 Flow chart for cryo-EM data processing and structure determination of the ABCG2-mitoxantrone-Fab complex.

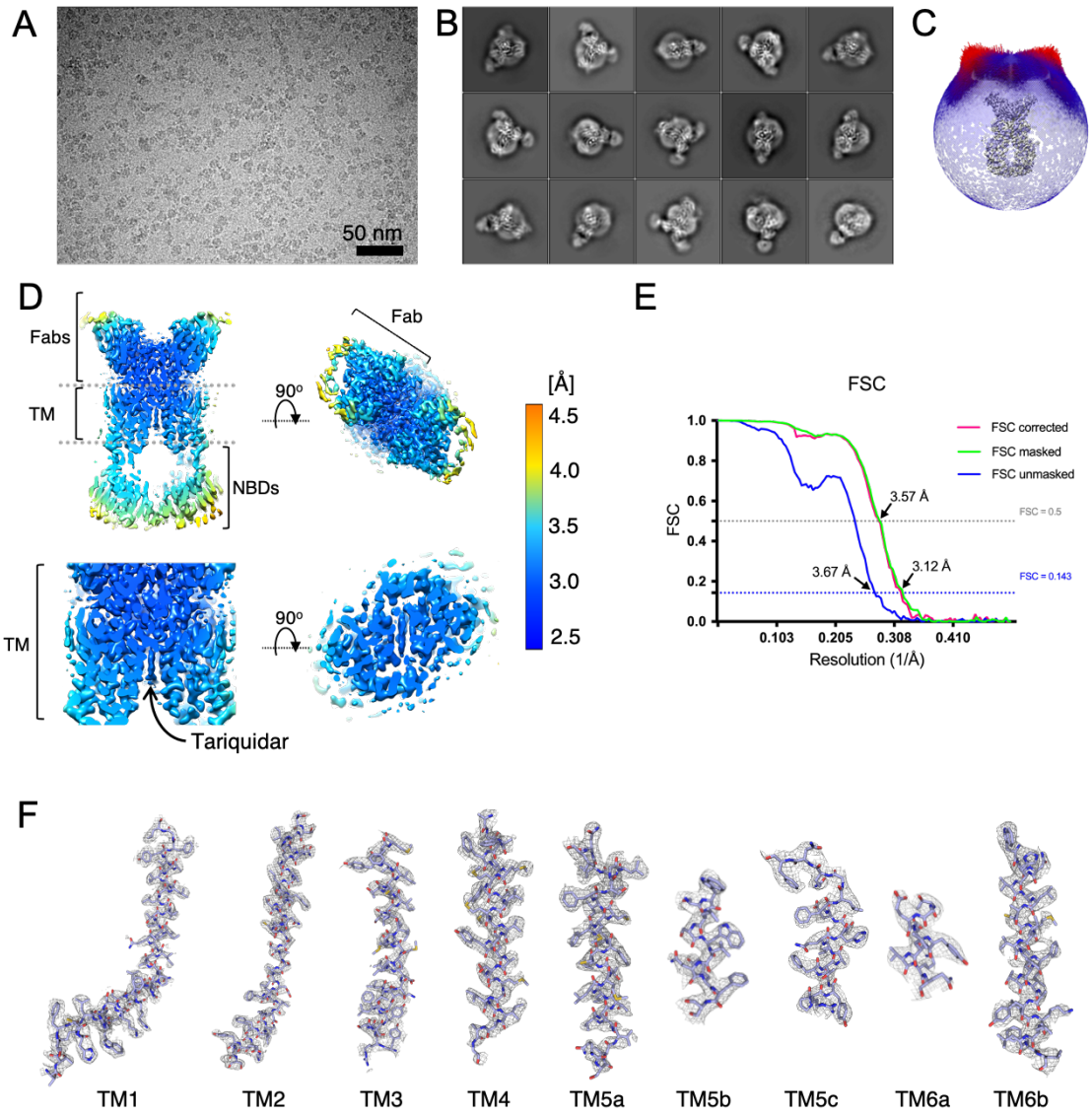


Figure 16 Cryo-EM map of ABCG2-tariquidar-Fab generation and resolution estimation. (A) An example micrograph (drift-corrected, dose-weighted, and low-pass filtered) of the nanodisc-reconstituted ABCG2-tariquidar-Fab data set. Black scale bar, 50 nm. (B) Fifteen representative 2D class averages of the final round of 2D classification, sorted in decreasing order by the number of particles assigned to each class. (C) Angular distribution plot for the final reconstruction. (D) Full view of the RELION local-resolution-filtered map of ABCG2-tariquidar-Fab, coloured by local resolution in Å as calculated in Relion 3.1, with the clipping plane in the middle of the molecule. On the top right corner is shown 5D3-Fabs region, in the bottom line the TM region with tariquidar substrate density. (E) FSC from the RELION auto-refine procedure of the non-symmetrized unmasked half-maps (blue), the half-maps after masking (green), and the half-maps after masking and correction for the influence of the mask (pink). Horizontal dotted lines (blue and black) are drawn for the FSC = 0.143 and FSC = 0.5 criterion, respectively. For both the unmasked and the corrected FSC curves, their intersection with the FSC = 0.143 and the FSC = 0.5 lines are marked by arrows, and the resolutions at these points are indicated. (F) Fitting of the all TM helices of ABCG2 in the EM density map. A region of up to 2 Å around the atoms is shown. TM helices are shown as sticks. Density is shown as a grey mesh.

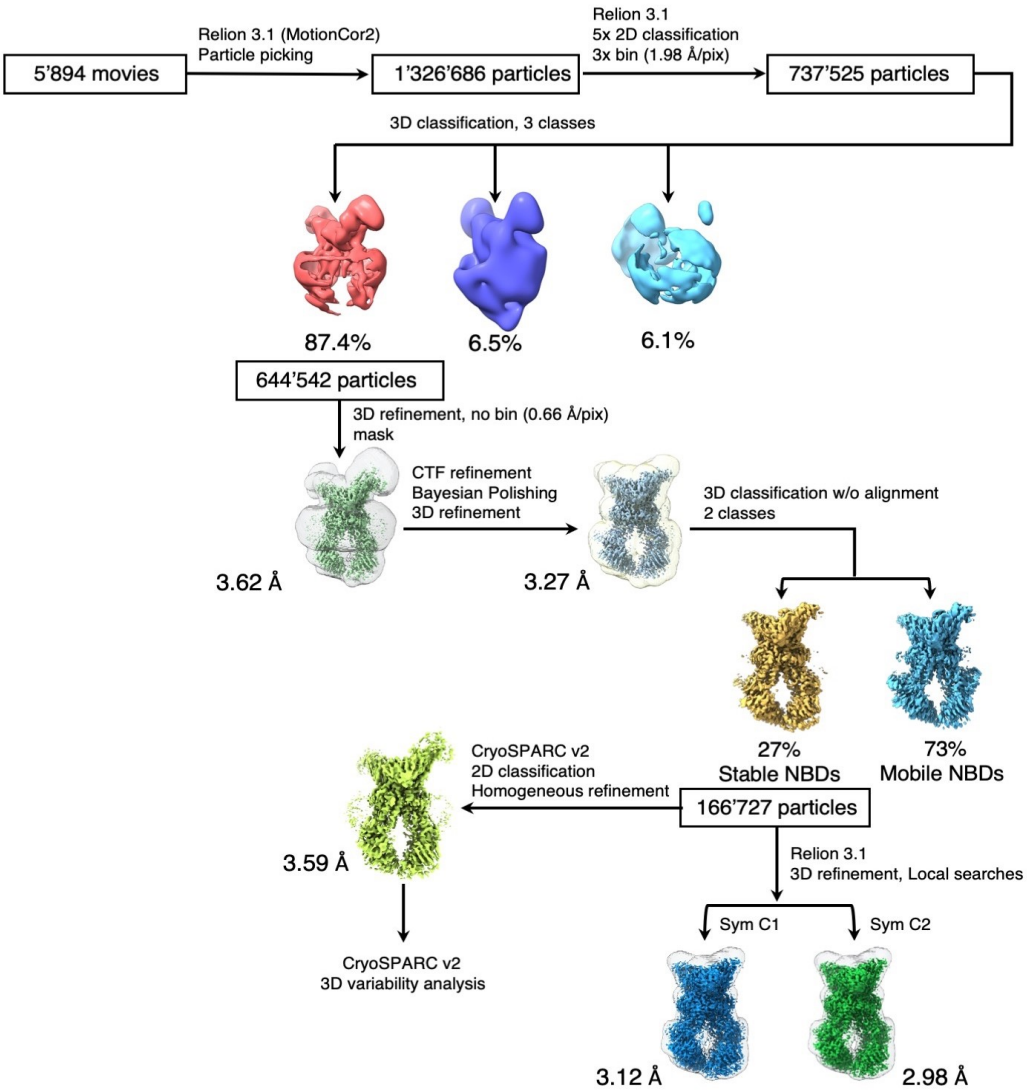


Figure 17 Flow chart for cryo-EM data processing and structure determination of the ABCG2-tariquidar-Fab complex.

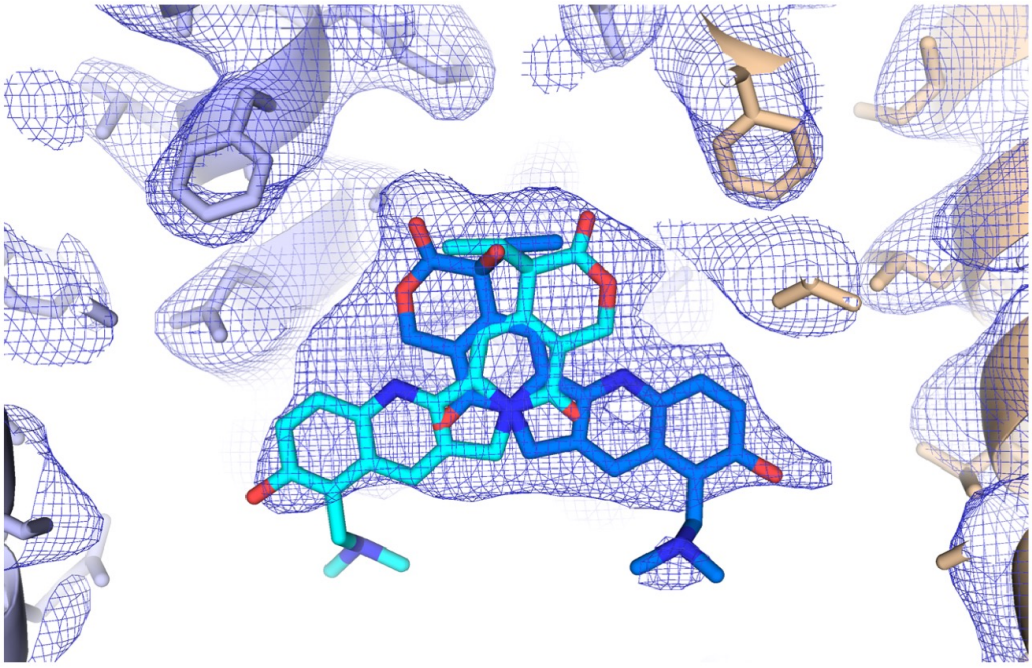


Figure 18 Substrate-binding cavity of ABCG2 with topotecan molecule. Non-symmetrized EM density of ABCG2-topotecan; bound topotecan molecule (blue and cyan sticks) is shown in two possible orientations, rotated by 180° along y-axis. The map level was set to 3.5σ, as defined by Pymol.

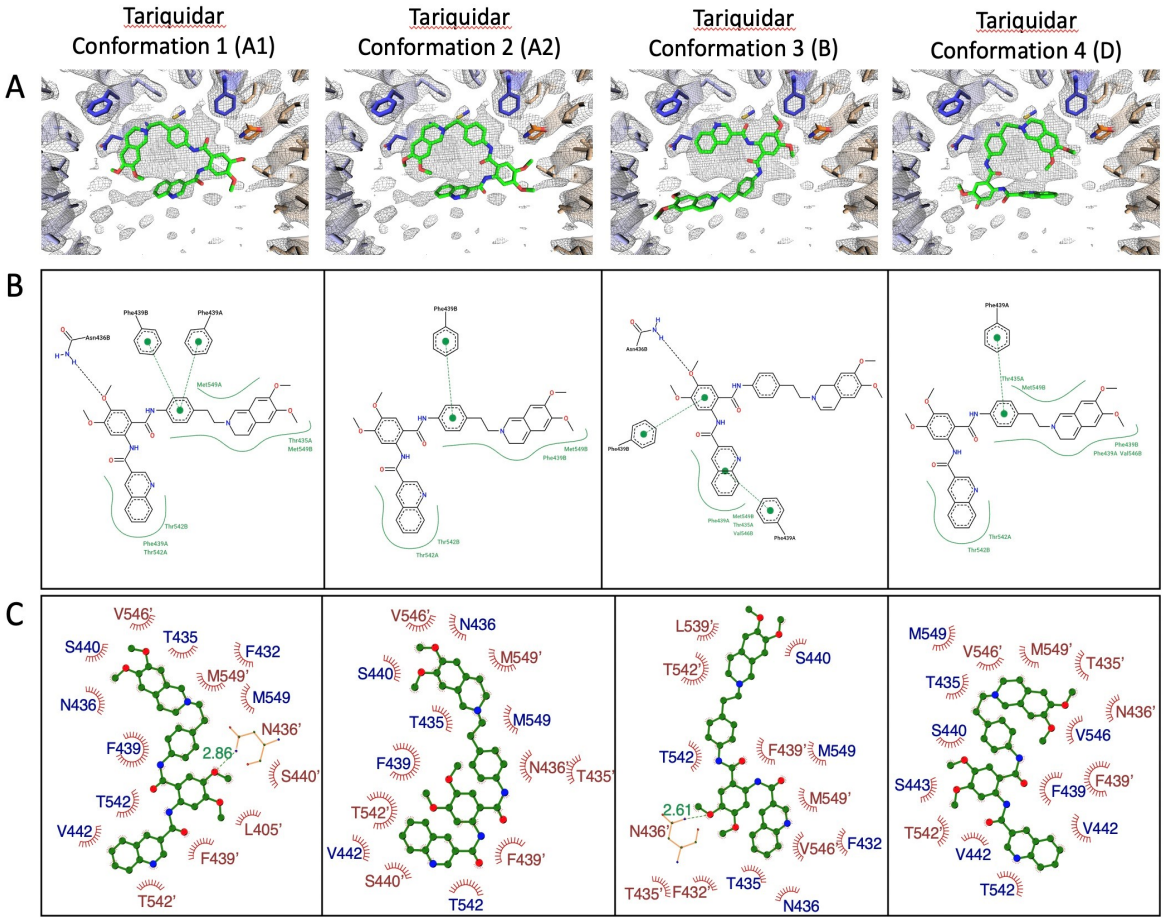


Figure 19 Possible conformations of tariquidar molecule in ABCG2 substrate binding cavity. (A) Four possible conformations of tariquidar fitted into the ABCG2tariquidar map. Tariquidar molecule shown as green sticks, map as gray mesh. Conformation 1 is presented in Figures 1-4. (B) Two-dimensional visualisations of protein-substrate interactions generated from three-dimensional inputs. Hydrogen bonds are shown as black dashed lines, hydrophobic interactions as smooth contour lines between the respective amino acids and the substrate molecule, and p-p stacking as green dots connected by dashed line. (C) Interactions between tariquidar in four different conformations and ABCG2 side chains. Nonbonded interactions are represented by spoked red arcs and hydrogen bonds are indicated by dashed green lines. The amino acid residues are as shown as single-letter abbreviations.

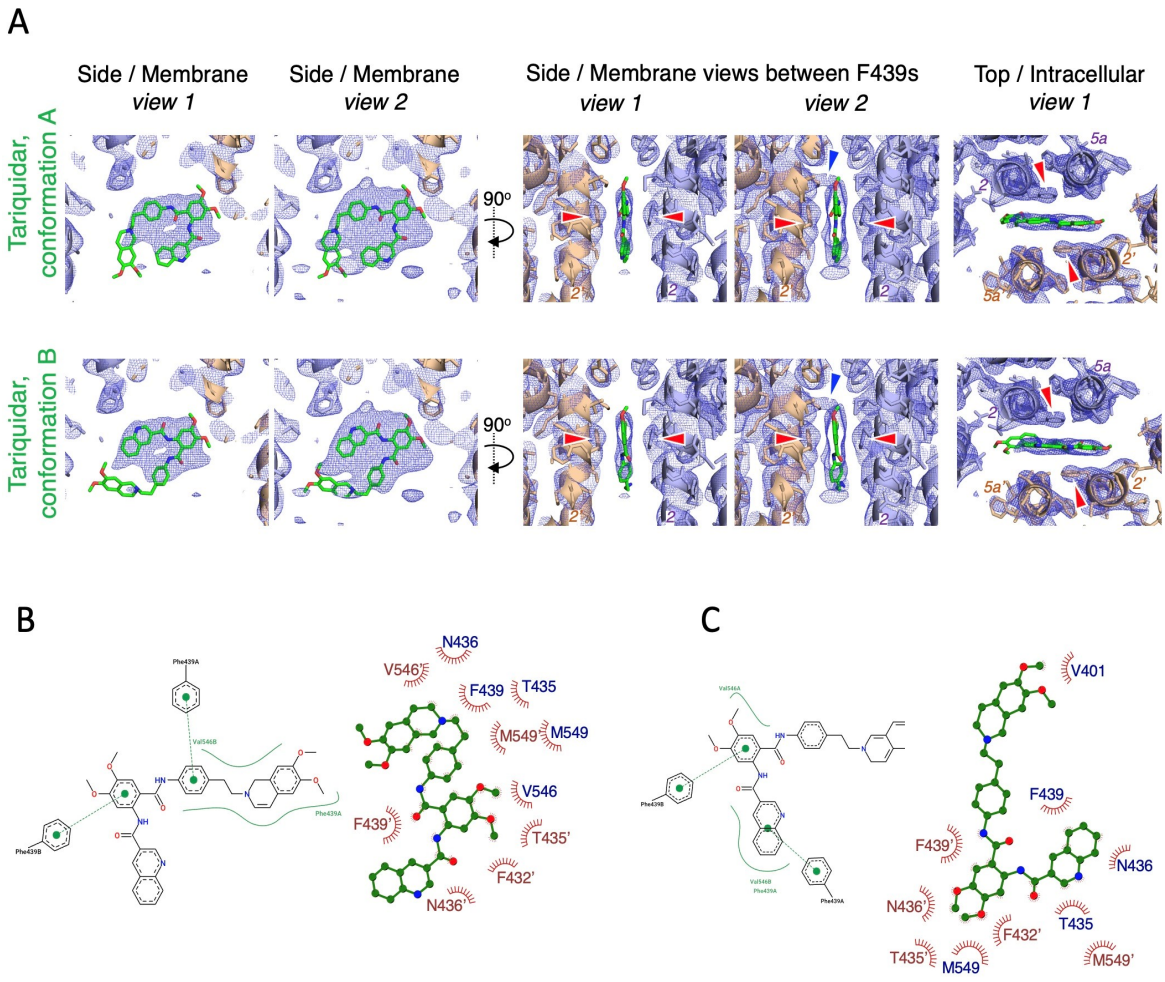


Figure 20 Possible conformations of tariquidar molecule in ABCG2tariquidar-Fab map obtained with CryoSPARC2. (A) Two possible conformations (labeled A and B) of tariquidar fitted into the ABCG2-Tariquidar-Fab map. Tariquidar molecule shown as green sticks, map as blue mesh. Labels as in Figure 7. (B) Two-dimensional visualisations of protein-substrate interactions generated from three-dimensional input for conformation A with Poseview and LigPlot+. (C) Two-dimensional visualisations of protein-substrate interactions generated from three-dimensional input for conformation B with Poseview and LigPlot+.

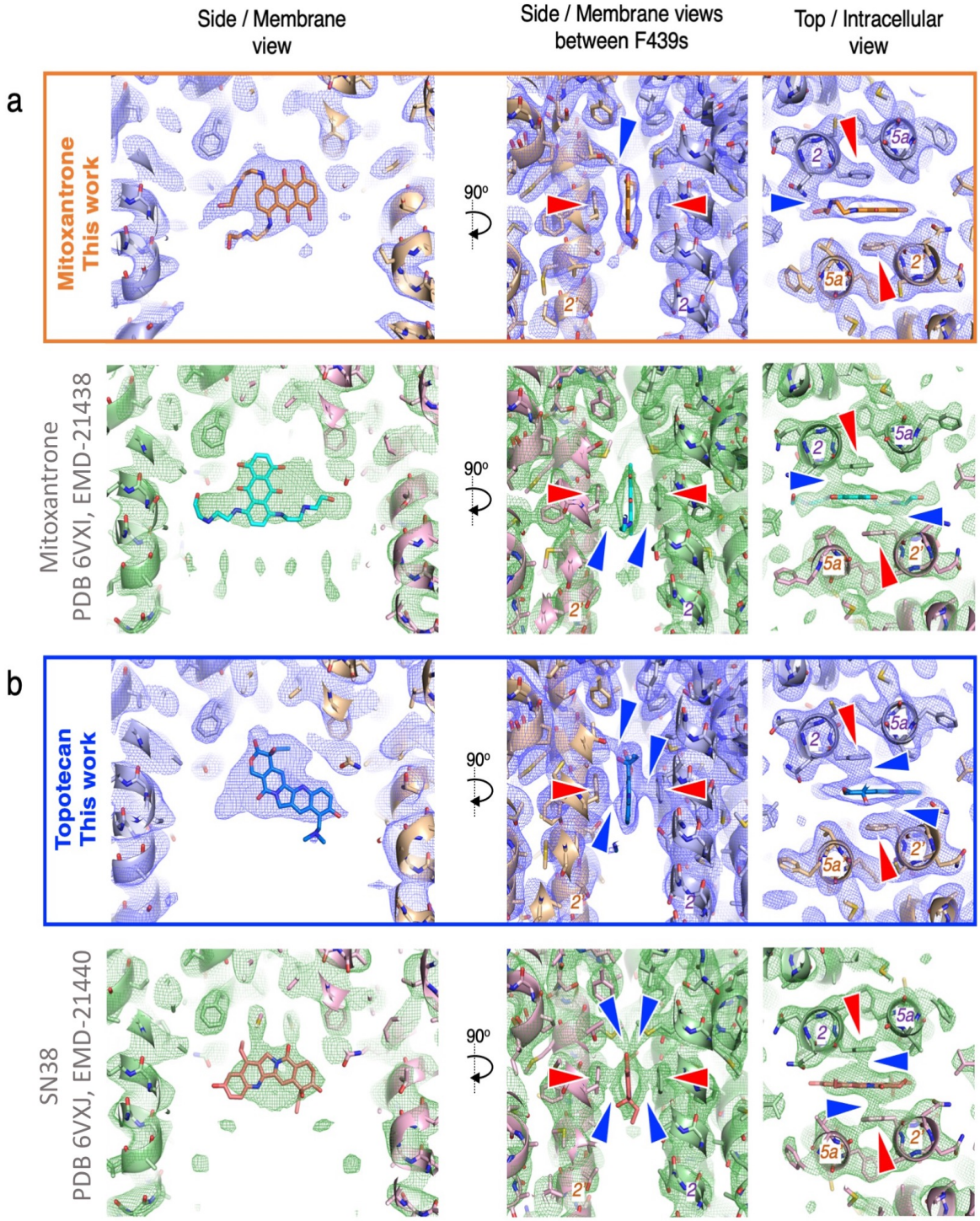


Figure 21 Comparison of the cryo-EM maps and models of ABCG2 with mitoxantrone, topotecan and SN38 anticancer drugs. Current ABCG2-mitoxantrone and ABCG2-topotecan models were fitted into the EM density maps (blue mesh) and compared with recently published maps/models of ABCG2-mitoxantrone and ABCG2-SN38 (green mesh). Two membrane views and intracellular view are presented for each of the structure. (A) EM densities and models of ABCG2-mitoxantron. Top line, the model from current work (mitoxantrone is orange). Bottom line, the ABCG2-mitoxantrone C2-symmetrized map from Orlando et al. (PDB 6VXI, EMD-21438, mitoxantrone is cyan). Mitoxantrone

presented in Orlando et al.(Orlando & Liao, 2020) is rotated almost 90° relative to our structure since its anthracene ring system is positioned vertically in cavity 1, while its long hydroxyl-ethylamine chains are horizontally stretched on both sides of cavity 1. In our ABCG2-mitoxantrone structure, the anthracene rings are oriented horizontally in the transporter and one of the hydroxyl-ethylamine tails forms a hydrogen bound with N436. The EM density is compact and does not suffer from C2symmetrization effects. The map levels were readjusted to display similar EM densities around binding pocket of ABCG2. Levels 4σ and 6σ, as defined by Pymol, were used to display the EM maps (blue mesh – this work, green mesh – Orlando et al. work), respectively. Labels as on the Figure 7. (B) EM densities and models of ABCG2-topotecan (this work, top line) and ABCG2SN38 (PDB 6VXJ, EMD-21440, bottom line, molecule is pink). Labels/adjustments as in (A) and on Figure 7. SN38 molecule, whose chemical structure is similar to topotecan is located in the binding cavity upside down (180° rotation along x axis) in comparison to our model.

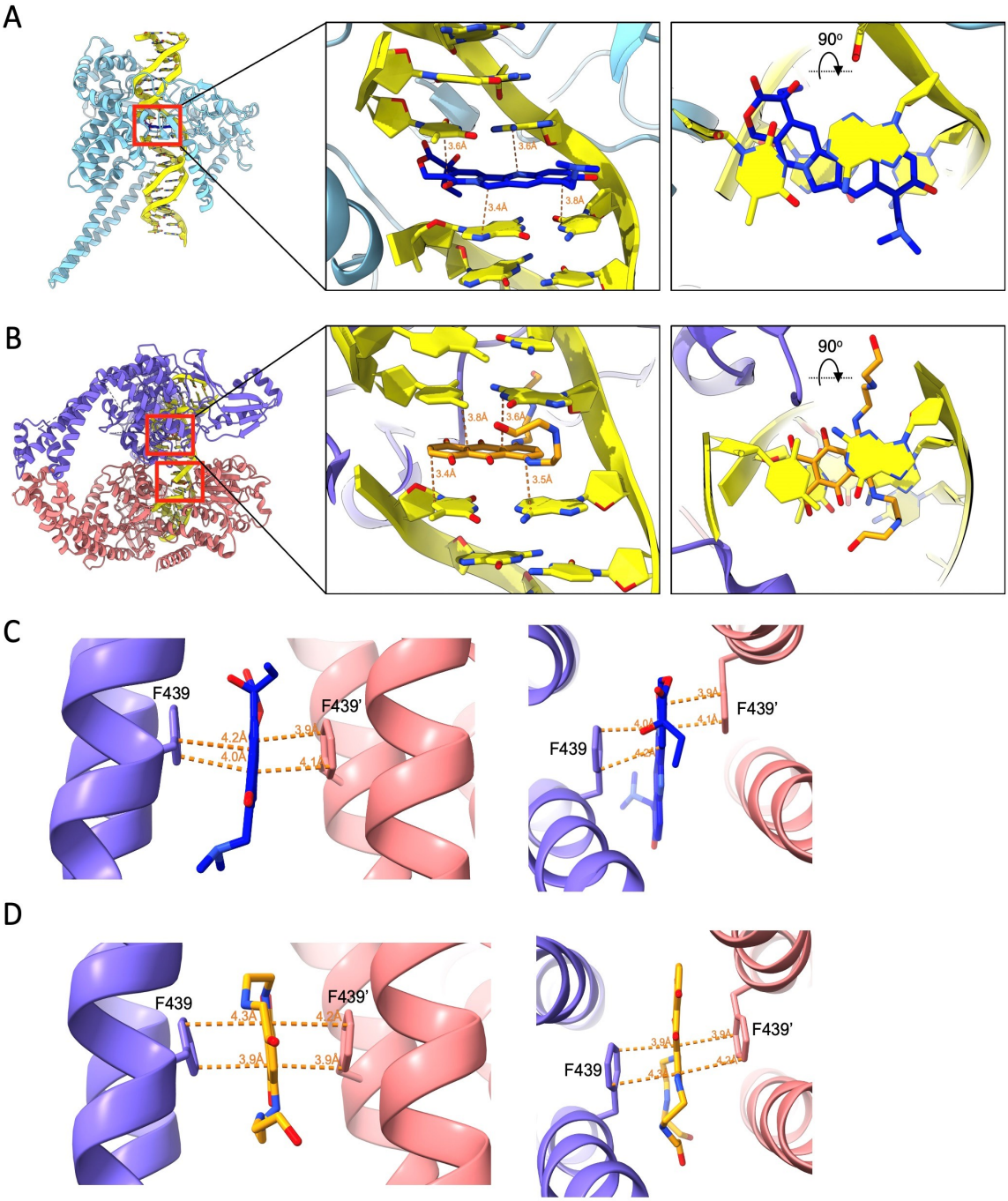


Figure 22 Interactions of topotecan and mitoxantrone with topoisomerases and ABCG2.

(A), Model of human topoisomerase I in the complex with topotecan and DNA (PDB 1K4T). The topoisomerase I is light blue, DNA yellow and topotecan blue. The region with inhibitor bound was marked with red rectangle and zoomed in to show both side and top views. The distances between aromatic rings of inhibitor and base pairs of nucleotides are shown as dashed lines and colored orange.

(B), Model of human topoisomerase II beta in the complex with two mitoxantrones and DNA (PDB 4G0V). Two monomers of topoisomerase II were colored purple and salmon,

DNA is yellow and two mitoxantrone molecules orange. Regions with mitoxantrone bound are marked with red rectangles and show as in a.

(C), Distances between topotecan and F439s in binding pocket of ABCG2 transporter. ABCG2 monomers are colored salmon and purple.

(D), Distances between mitoxantrone and F439s in ABCG2. Colored as in (C).

3.8 SUPPLEMENTARY TABLES

Table 1 The EC50 of substrate ATPase stimulation determined using all curves from Figure 1B with the error of the fit (standard deviation) shown.

Substrate	EC50 (μM) (ATPase)
Mitoxantrone	18.3 ± 2.2
Topotecan	11.7 ± 0.3
Tariquidar	0.08 ± 0.01

Table 2 The fold stimulation in the ATPase activity of ABCG2 caused by the addition of substrate in the presence and absence of Fab. Calculated from data on Figure 1D.

Substrate	G2 fold stimulation	G2+Fab fold stimulation
+ Tariquidar	2.1	1.6
+ Mitoxantrone	3.2	2.3
+ Topotecan	3.5	2.9

Table 3 Cryo-EM data collection, refinement and validation statistics.

Cryo-electron microscopy data collection and processing			
Microscope	FEI Titan Krios	FEI Titan Krios	FEI Titan Krios
Voltage (keV)	300	300	300
Camera	Gatan K2-Summit	Gatan K2-Summit	Gatan K3
Electron exposure (e–/Å^2/frame)	1.55	1.5	2.0
Energy filter slit width (eV)	20 (Gatan Quantum-LS (GIF))	20 (Gatan Quantum-LS (GIF))	20 (Gatan Quantum-LS energy filter (GIF))
Pixel size (Å)	0.82	0.64	0.66
Defocus range (μm)	(−0.8) - (−3.0)	(−0.6) - (−2.0)	(−0.4) - (−2.5)
Magnification (nominal)	60’975x (165kx)	78’125x (215kx)	76’000 (130kx)
Number of frames per movie	40	40	40
Number of good micrographs	3’403	6’192	5’894
Initial particles	775’993	614’891	1’326’686
Final particles	133’832	166’965 (91’543 for C1 map)	166’727
Symmetry imposed	C1, /C2 (model building)	C1, /C2 (model building)	C1 (model building), / C2
Map resolution (Å)	3.39 / 3.14	3.51 / 3.35	3.12 / 2.98
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	20 -2.70	20 -2.60	20-2.60
EMDB ID	EMD-XXXX	EMD-XXXX	EMD-XXXX

Refinement			
Non-hydrogen atoms	12393	12394	12529
Protein residues	1582	1582	1582
Ligands	TTC:1 NAG:	MIX:1 NAG:2	TAR:1 NAG:2 CLR:3 PEE:1
RMSD bonds (Å)	0.006	0.005	0.009
RMSD angles (°)	0.723	0.757	0.752
Ramachandran favored (%)	90.63	90.56	92.11
Ramachandran Allowed (%)	9.37	9.31	7.89
Ramachandran Outliers (%)	0.00	0.13	0.00
Rotamer outliers (%)	0.30	0.00	5.56
Cβ outliers (%)	0.00	0.00	0.00
All-atom clashscore	12.62	20.23	15.39
MolProbity Score	2.14	2.33	2.75
PDB ID	XXXX	XXXX	XXXX

Table 4 ABCG2 residues within 4 Å that interact with anticancer drugs in the ABCG2-substrateFab structures as shown on Figure 4. The prime (') corresponds to the second half of the ABCG2, the (^{2}) to both halves of transporter.

Compound	TM α-helix	Residue
Topotecan	1b	-
	2	T435, N436', F439^2
	5a	T542^2, M549^2
Mitoxantrone	1b	-
	2	T435, N436^2, F439^2
	5a	T542', V546', M549^2
Tariquidar	1b	L405'
	2	F432, T435, N436^2 , F439^2 , S440^2 , V442
	5a	T542^2, V546', M549^2

3.9 SUPPLEMENTARY MOVIES

Video 1

Visualization of the 3D variability analysis of ABCG2-topotecan-Fab structure performed in CryoSPARC2. The nanodisc density was masked.

Video 2

Visualization of the 3D variability analysis of ABCG2-mitoxantrone-Fab structure performed in CryoSPARC2. The nanodisc density was masked.

Video 3

Visualization of the 3D variability analysis of ABCG2-tariquidar-Fab structure performed in CryoSPARC2. The nanodisc density was masked.

4 Chapter IV: Cryo-EM structure of the Ng. LptDE complex

In this section, I present a structure study of Ng. LptDE complex. Several structures have been obtained in this study by cryo-EM.

The main findings:

1) A structure of LptD/E with the luminal gate in the open state has been determined by cryo-EM at an overall resolution of 3.4 Å.

2) Another high-resolution structure containing a C-terminal plug.

3) A structure with completely opened lateral gate, which is the first visualization of the laterally opening of the LPS assembly complex, providing direct evidence of the current proposed model for LPS biogenesis.

4) A dimeric structure of LptD/E(2:2) has been resolved by cryo-EM, suggesting a potential model for the LPS transporting bridge.

My contribution: Protein expression and purification; cryo-EM sample preparation; dataset acquisition; image processing; model building; interpreting the structures; supporting manuscript writing and editing.

This section is prepared from a manuscript in preparation.

Running Title:

Insight into the lipopolysaccharide translocation mechanism by cryo-EM structures of a LptDE transporter in complex with Pro-Macrobodies

Mathieu Botte1*, Dongchun Ni2*, Stephan Schenck1‡*, Iwan Zimmermann3, Mohamed Chami2, Nicolas Bocquet1, Denis Bucher1, Pascal Egloff3, Matilde Trabuco1, Robert K.Y. Cheng1, Janine Brunner4‡, Markus A. Seeger3, Henning Stahlberg2 and Michael Hennig1

1 leadXpro AG, Park Innovaare, 5234 Villigen, Switzerland
2 C-CINA, Biozentrum, University of Basel, Mattenstr. 24, 4058 Basel, Switzerland
3 Institute of Medical Microbiology, University of Zurich, Gloriastasse 28/30, 8006 Zürich
4 Paul Scherrer Institute (PSI), 5232 Villigen, Switzerland

*Authors contributed equally

‡ Present address: VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium

Subject terms: Gram-negative bacteria, antibiotics drug target, β-barrel outer membrane protein, lipoprotein, lipopolysaccharide (LPS) transport, LptDE complex, cryo-EM

4.1 Abstract:

Molecular chaperones like nanobodies are widely used to facilitate the structure determination of challenging proteins by stabilization of their conformation. Here we describe the computational design and creation of fusion proteins of nanobodies with maltose binding protein, called Pro-Macrobodies (PMb). Because of their rigidity and high molecular weight, PMb facilitate the particle picking and particle alignment in cryo-EM image processing. Complexation of the Neisseria gonorrhoeae LptDE transporter with two PMb allowed the first structure determination by cryo-EM of a 110kDa β-barrel membrane protein at 3.4 Å resolution. Our work gives new insights into the mechanism of insertion of lipopolysaccharide (LPS) into the outer membrane of Gram-negative bacteria, provides a structural basis for blocking of LPS transporters that could be used for novel antibiotics, and describes a novel, highly rigid and widely applicable chaperone scaffold to enable cryo-EM structure determination of challenging protein targets.

4.2 Introduction:

Multi-drug resistant bacteria present a growing concern for human health. Among these, Gram-negative bacteria protect themselves from antibiotics with an outer membrane (OM) that establishes a tight barrier for several antibiotics. Glycolipid lipopolysaccharides (LPS) form the outer leaflet of the OM as an essential component of the bacterial defense.

LPS is synthesized in the cytosol via the Lpt pathway, consisting of the proteins LptA, B, C, D, E, F and G, which form a trans-envelope complex that transports LPS across the periplasmic space to the outer leaflet of the OM (Whitfield & Trent, 2014; Freinkman et al., 2011, 2012). (Narita & Tokuda, 2009). The structures of LptDE from multiple species were determined by X-ray crystallography. These provide insight into the transport mechanism of LPS to the outer leaflet of the bacterial outer membrane, which represents an essential shield for the Gram-negative bacteria to hinder penetration of antibiotics (Nikaido, 2003; Ruiz et al., 2009). The beta-jellyroll domains of LptC, LptA and the N-terminal domain of LptD assemble into a slide with a continuous hydrophobic groove spanning the periplasm, which allows the lipid A portion of LPS to passage towards the C-terminal domain of LptD in the outer membrane (Botos et al., 2016; Villa et al., 2013). Next, the lipid A portion is inserted into the hydrophobic membrane bilayer, while the saccharidic portions of LPS are translocated through the hydrophilic lumen of the LptD β-barrel domain (Dong et al., 2014; Gu et al., 2015; Li et al., 2015; Qiao et al., 2014a). Disrupting the assembly of the outer membrane by inhibiting the lipopolysaccharide (LPS) transport pathway (Lpt) is an attractive strategy for novel antibiotic therapeutics (Lehman & Grabowicz, 2019; Ruiz et al., 2008). Among the constituents of the Lpt pathway, only LptD faces the extracellular milieu and is most accessible for inhibitors as exemplified with the synthetic peptidomimetic murepavadin (POL7080) that was identified as outer membrane assembly inhibitor acting on the target LptD(Andolina et al., 2018; Srinivas et al., 2010). Murepavadin showed efficacy as antibiotic in clinical studies phase I and II (Zha et al., 2016).

Here we present the cryo-electron microscopy (cryo-EM) structures of Neisseria gonorrhoeae LptDE (NgLptDE), a 110 kDa beta-strand barrel protein complex. High resolution of the analysis was enabled by formation of a complex with two PMbs. PMbs are fusion proteins of natural or synthetic nanobodies (sybodies) with maltose binding protein designed to form rigid molecules to increase molecular weight and facilitate particle picking and particle alignment in cryo-EM image processing. The structure provides yet unknown

conformations of the transporter and gives insight into the mechanism of LPS insertion into the outer membrane by opening of the luminal loops, which allows the saccharide part of LPS to enter the LptD lumen, while lipid A is inserted into the outer membrane. Another structure shows the fully opened lateral gate between β-1 and β-26 of the beta-strand barrel that allows the core oligo saccharides and the O-antigen to reach the extracellular side of the outer membrane through the lumen of LptD in step with lipid A. A third structure suggests that the luminal gate could be blocked by the C-terminus of LptD, a possible auto regulative mechanism. Together these findings are significant not only for understanding the mechanism of LPS insertion into the outer membrane but also to facilitate the development of novel antibiotics against multi resistant Gram-negative bacteria.

4.3 Results

4.3.1 4.Protein generation

Full length LPTD and LPTE from Neisseria gonorrhoeae together with Escherichia coli YIFL peptide were co-expressed in E. coli and purified to homogeneity for sybody generation, biophysical and structural studies. The exact function of YifL is yet unknown but Yihua Huang and co-workers demonstrated that co-expression of E. coli YIFL lipoprotein was key to improve the expression yield, purification quality and stability of LptDE (personal communication). Surprisingly and contrary to previously published purifications of LptDE complexes (Qiao et al., 2014; Dong et al., 2014; Botos et al., 2016), the solubilized NgLptDE complex exists as a monomer and dimer in solution (Figure 30a-b). Monomeric and dimeric NgLptDE could be isolated from each other and the dimeric form was stable (Figure 30c). Monomeric NgLptDE was used for sybody generation as the dimer formation could mask some important epitopes.

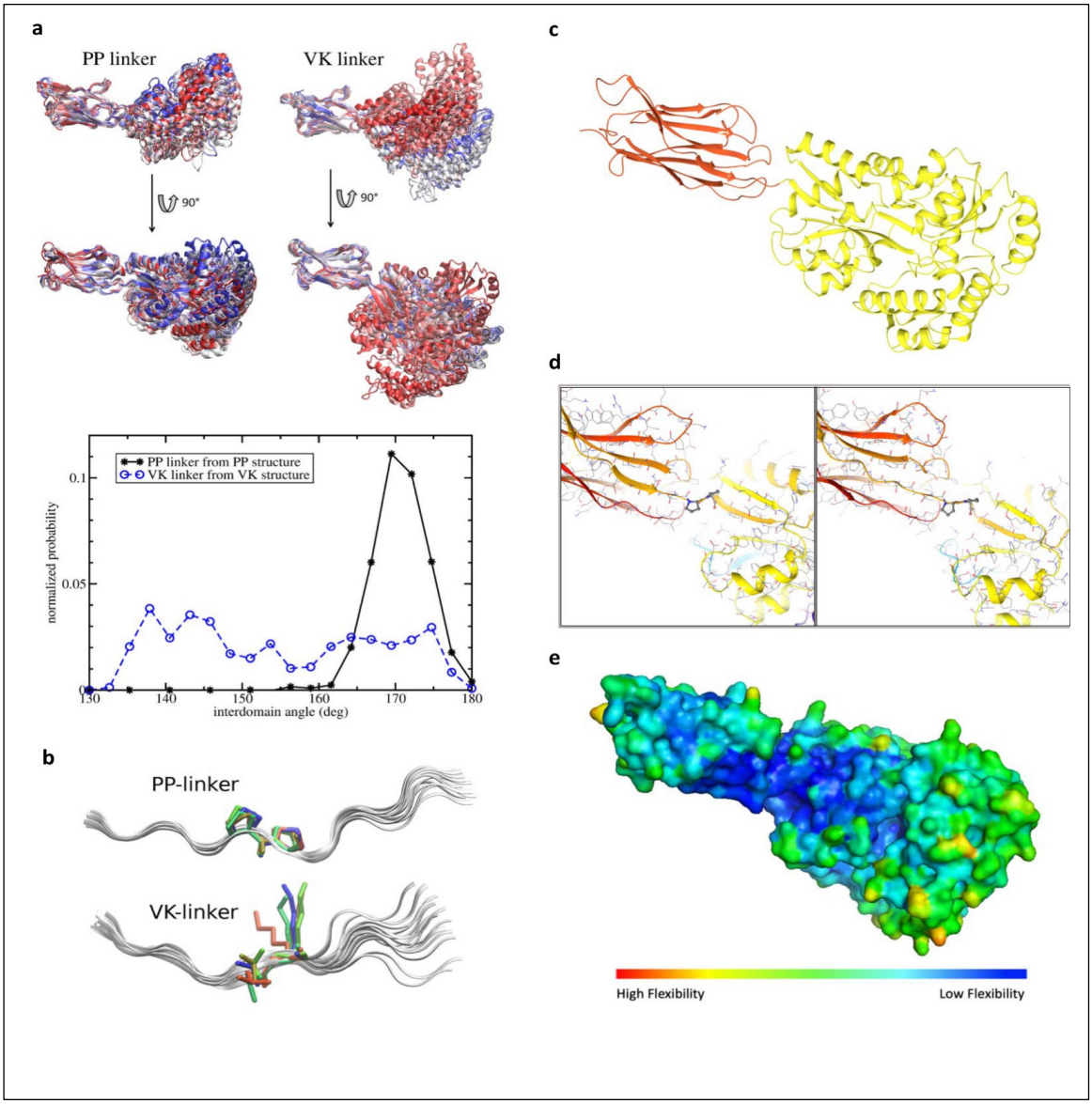
4.3.2 Generation of sybodies as molecular tools for structure determination

To obtain molecular chaperones for structure determination, we generated sybodies via a recently established in vitro selection platform (Figure 31a-c). As little as 100 µg of purified and biotinylated recombinant full length NgLptDE complex expressed in E. coli were consumed as target protein for selection (Figure Figure 1) an amount of protein insufficient for an immunization campaign (Zimmermann et al., 2018, 2020). Characterization of ten sybodies by grating-coupled interferometry showed binding affinities in the range of 800 pM and 1000 nM and size exclusion chromatography revealed two sybodies (called 21 and 51) that can bind simultaneously to purified NgLptDE (Figure 32a-b and Figure 33). They

exhibit affinities in the low nM range, and slow off-rates (Figure 32c-f). The sybodies did not bind E. coli overexpressing NgLptDE, indicating that their epitopes are not accessible in the context of the outer membrane (Figure 31d). Therefore, it was not surprising that none of the sybodies exhibited antibiotic activity against Neisseria gonorrhoeae, when tested alone or in synergy with the antibiotic vancomycin (Figure 31e).

4.3.3 Design and engineering of the Pro-Macrobody

To create molecules that combine the unique protein conformation stabilization properties and easy production of nanobodies with the size and orientation marker for cryo-EM image analysis as powerful as Fab, we designed a rigid fusion linker of synthetic nanobodies



(sybodies) with maltose binding protein. To estimate the mobility of the two moieties

Figure 23 Generation of the Pro-Macrobodies.

(a) (Top): Flexibility of designed chimeric fusion polypeptides (macrobodies and Pro-macrobodies) in MD simulations. Conformational space explored during a 500 ns MD trajectory initiated from an Xrays structure solved with a double proline, PP-linker (left), and (right) from a simulation model with the linker computationally mutated to a valine and a lysine (VK-linker). Both trajectories were aligned on the VHH structure (residues 1-120). Snapshots are taken every 100 ns and overlaid. (Bottom): Analysis of macrobody interdomain (VHH to MBP) angle during MD trajectories. The interdomain angle is defined as the angle between the geometrical centers of residue 1-120 (Nb), residue 121-122 (linker), and residue 123-486 (MBP). Distribution of interdomain angles seen in 2 different MD simulations started from two different X-rays structures solved with either a PP-linker (full circles), or a VK-linker (empty circles, 6HD8.pdb). (b) Close-up view on the two linkers showing the larger conformational flexibility of the VK-linker by overlaying several snapshots taken from simulations. (c) X-ray structure of the Pro-macrobody 21. The nanobody domain is colored in orange whereas the MBP domain is represented in yellow. (d) Close-up comparison on the linker regions.(Left) Linker region of the X-ray structure of Pro-macrobody 21. (Right) Linker region of the Macrobody model from 6HD8.pdb simulated with di-proline-linker. Domains are colored as in (c) and the linker is shown in gray sticks. Note the very similar overall conformation of the X-ray structure and the simulation. (e) Rigidity of the Pro-macrobody. The molecule in a surface representation is colored according to the relative flexibility as determined by experimental structure determination.

(sybody or nanobody and MBP) relative to each other, we performed all-atom molecular dynamics (MD) simulations of the macrobody structure (extracted from PDB entry 6HD8) (Figure 23a). The simulation revealed a high flexibility of the molecule and bending at the linker of over 40°. Systematic evaluation of alternative residues at positions Val122 and Lys123 in the linker revealed that replacement by two consecutive prolines was best suited to reduce flexibility (Figure 23a-b). Here, the MD simulation converged within 10ns to a linker rotation toward the equilibrium angles of prolines in the cis-configuration, the maltose binding protein moiety rotating by over 170 degrees relative to the macrobody structure (PDB-ID 6HDC), and the PMb forming a linearly extended straight molecule (Figure 23a). Thus, using sybody 21 as a first test example, the di-proline version of the macrobody (termed Pro-Macrobody 21, PMb21) was generated and purified. Affinity and binding kinetics of the PMbs are very similar compared to their parent sybodies and confirm that the enlargement does not alter the binding properties (Figure 32g-j). The X-ray crystal structure of PMb21 was determined at 2Å resolution (Figure 23c), which confirmed the extended conformation predicted in the simulation and showed particularly low temperature factors at the di-proline linker (Figure 23d-e). Unlike alternative engineering approaches

showing higher flexibility of the domains (Laverty et al., 2019; Uchański et al., 2019), the properties of PMbs albeit connected only by a single peptide stretch appeared very well suited for particle enlargement and for aiding particle alignment in cryo-EM. PMb51 and PMb21 were selected (Figure 32) to form complexes with NgLptDE and analyzed by SEC analysis showing large shifts indicating complex formation with NgLptDE (Figure 33). In addition, the ternary complex of NgLptDE with PMb21 was analyzed by electron microscopy, which allowed recognizing the location of PMb51 and PMb21 in micrographs of the quaternary complex (Figure 24), and helped quantifying the benefits of adding the second PMb to the analysis.

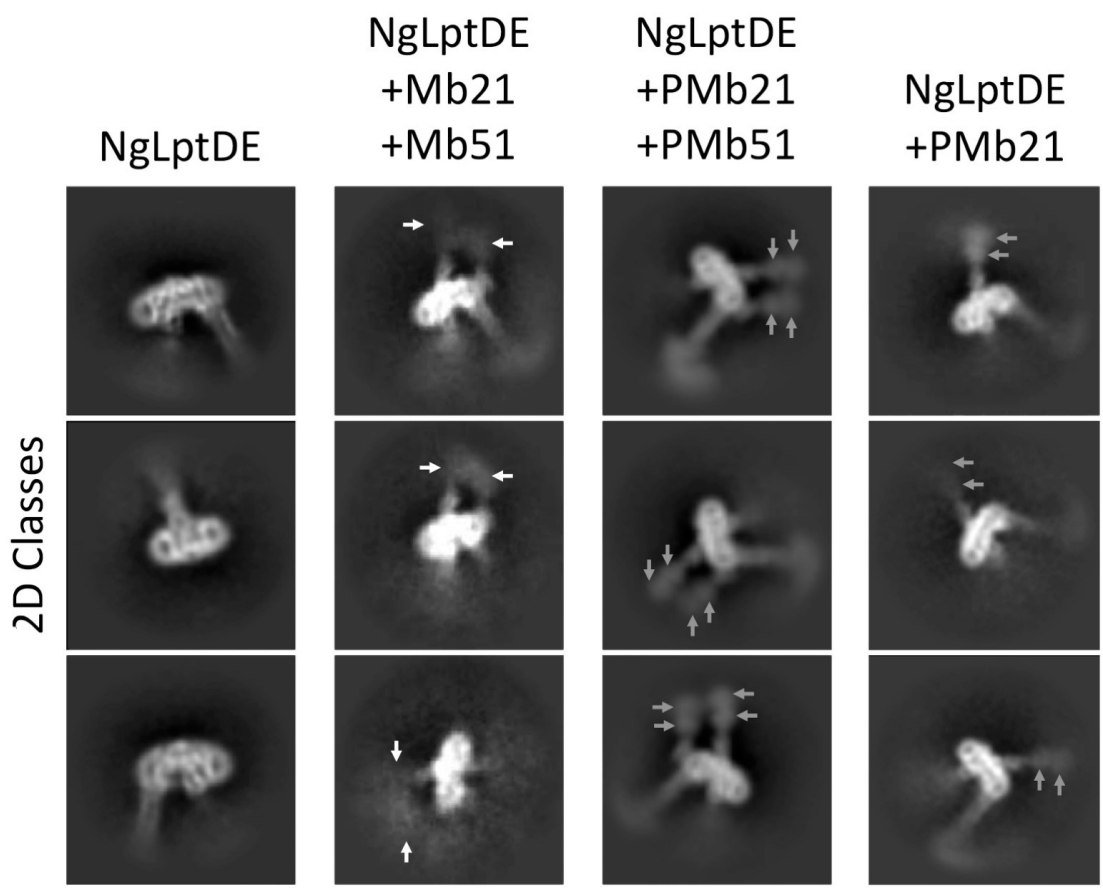


Figure 24 Pro-Macrobodies rigidity observed by cryo-EM.

Representative 2D class averages obtained by cryo-EM for NgLptDE in complex with Macrobodies or Pro-Macrobodies. The fuzzy MBP densities are identified by a white arrow and distinct MBP lobs are indicated by a blue arrow.

To investigate the improvement of rigidity of the PMb, Macrobody-versions of the very same sybodies 21 and 51 with the original Val-Lys linker were generated and analyzed. In 2D class averages of cryo-EM data, the MBP moiety following the Val-Lys-linker showed

high flexibility and no clearly defined features that could be used for subclass identification in automated particle picking (Figure 24). In contrast, both the cryo-EM densities of PMb21 and PMb51 show excellent contrast, providing evidence for the universality of this scaffold (Figure 24 and Figure 25a). In addition, the examination of the 2D classes obtained with the Pro-Macrobodies allows to clearly distinguish the two lobes of the MBP domain which was not possible with the Macrobodies due to the high flexibility of the Val-Lys-linker (Figure 24).

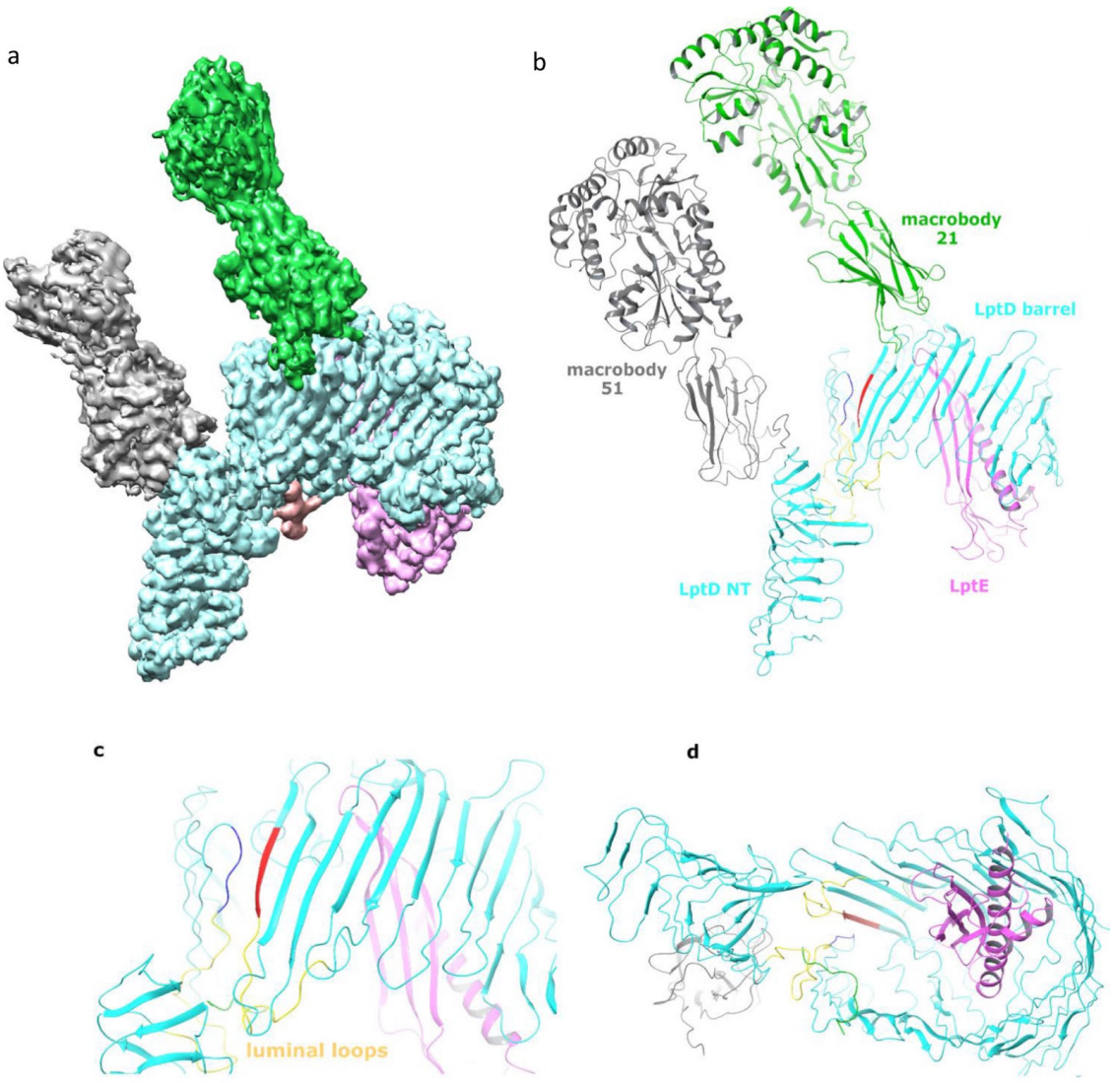


Figure 25 The architecture of the quarternary LptDE complex. (a) The cryo-EM map of the full quaternary complex shows the rigidity of the PMb. (b) Ribbon representation of the high resolution cryo-EM structure of NgLptDE. Luminal loops 1 and 2 are colored yellow, and β1 and β26 are marked blue and red, respectively. The PMbs containing sybodies 21 and 51 are shown in green, and grey representations, respectively. PMb21 binds the C-terminal β-strands of the LptD NT jellyroll domain. PMb51 binds extracellular loop 12 (L12) of the β-barrel domain. (c) Close-up representation of the LptD barrel, and (d) view from the interior side of the out-membrane.

4.3.4 High resolution structure of NgLptDE

Initial attempts at solving the structure of recombinant NgLptDE from the pathogenic Gram-negative bacterium Neisseria gonorrhoeae (NgLptDE) resulted in cryo-EM maps at 4.6 Å resolution, only (Figure 34). Consequently, molecular details such as side chains for confident tracing of the sequence and the identification of potential ligands could not be visualized.

Complexing LptDE with the PMbs 21 and 51 improved the resolution of the cryo-EM analysis to 3.4 Å, which allowed recognizing several molecular details relevant to the understanding of the mechanism of action of the transporter (Figure 25 and Figure 35). The resulting cryo-EM structure of the complex showed the same overall LptDE architecture observed also in X-ray structures from LptDE of other Gram-negative bacteria (Grabowicz et al., 2013 ,Qiao et al 2014, Botos et al 2016) (Figure 25). However, the N-terminal 63 residues of LptD (88 with the cleaved signal sequence) were not visible in the cryo-EM map, possibly because NgLptD in our study lacked the Cys residue near the N-terminus that fixes the N-terminal helix to the periplasmic turn between β24 and β25 in Shigella flexneri LptDE (SfLptDE) and Klebsiella pneumoniae LptDE (KpLptDE). Despite the missing di-sulphide bridge, the orientation of the N-terminal domain was the same as in KpLptDE, whereas that of SfLptdE is rotated by approximately 20° (Figure 26a). In our map, the LptE helix H2 extends into the periplasmic space and the N-terminal segment reaches over the wall of the barrel to anchor the lipid-modified N-terminal cysteine in the membrane.

A closer analysis of the recorded cryo-EM data allowed the identification of subsets of particles that allowed solving additional structures of hitherto undescribed conformations.

In our cryo-EM structure of NgLptDE, the conformations of the lateral and luminal gates are more open compared to KpLptDE and much more open compared to the X-ray structures of the SfLptDE complexes (Botos et al., 2016; Dong et al., 2014; Qiao et al., 2014a). Only three hydrogen bonds were observed between β-sheets β1 and β26, which increases the separation of the two gating strands at their periplasmic end in NgLptDE by 3 Å compared to SfLptDE, while KpLptDE has five hydrogen bonds between those strands (Figure 26a and b). Luminal loop 1 preceding β1 adopts a conformation in NgLptDE that does not obstruct the luminal gate. Those residues of luminal loop 2 that are resolved in the cryo-EM map indicate an open conformation of this loop. This conformation results in a direct connection between the hydrophobic groove of the N-terminal domain and the lumen of the

β-barrel, which can be described as an opened luminal gate (Figure 26b). The gate has a diameter of 12Å as determined by the smallest distance between Leu250-Asp251 and Asp768-Leu769, which could allow passage of the LPS core and O-antigen from the periplasmic space to the lumen of the barrel. And indeed, an LPS molecule (extracted from PDB-ID 3FXI without further adjustments) could be fitted into this open conformation of

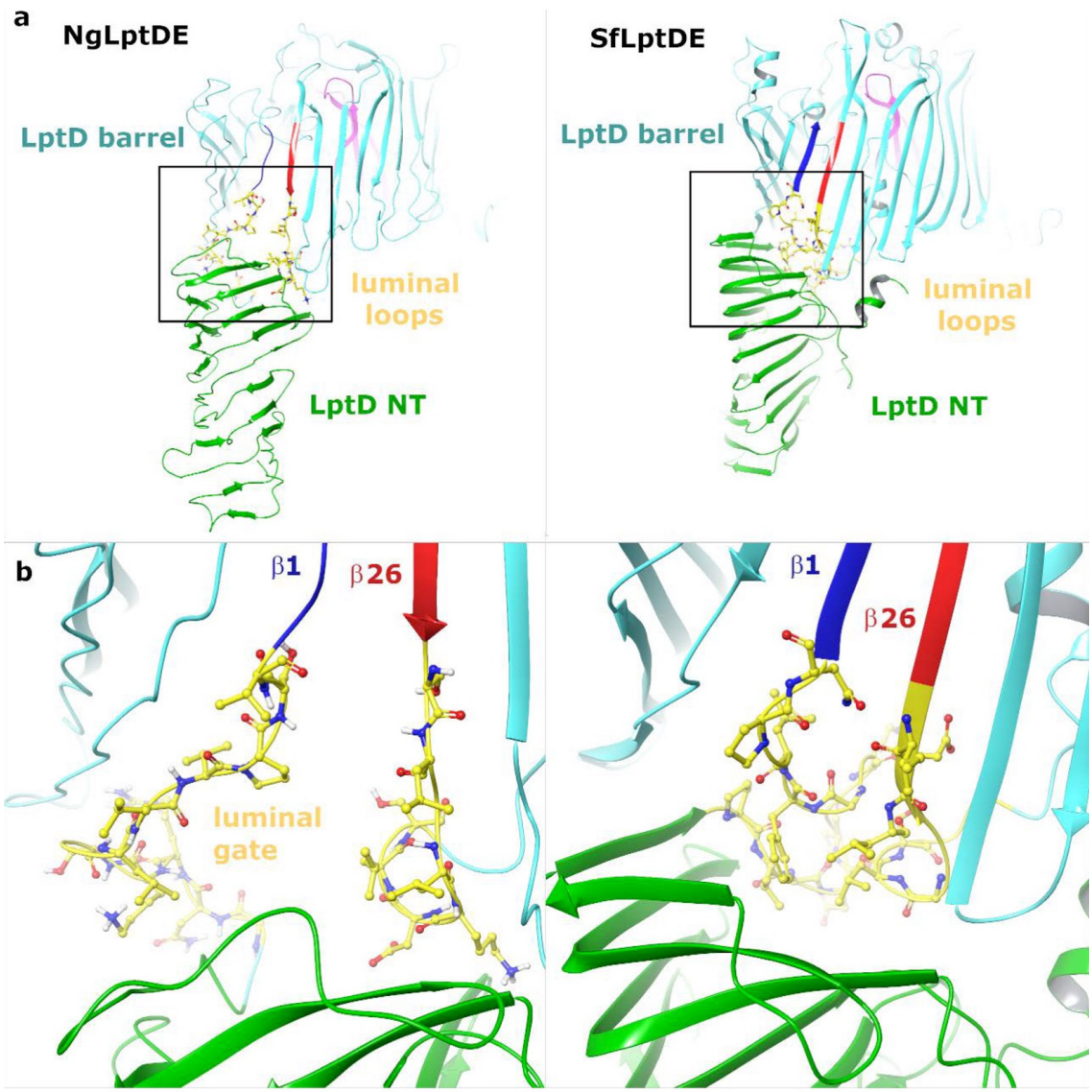


Figure 26 The luminal gate of NgLptDE

(a) Comparison of NgLptDE (left) with SfLptDE PDBID 4q35 (right) showing the increased opening between β1 and β26 and between the luminal loops in NgLptDE. The N-terminal domain (green) has the same orientation as in KpLptDE, wheraeas that in SfLptdE it is rotated by approximately 20° (b) Close-up representation of the resulting opening of the luminal gate. There is no constriction observed between the jellyroll domain and the barrel in NgLptDE in contrast to SfLptDE.

NgLptD without atomic clashes, whereas the closed conformation observed in SfLptD does not provide enough space for LPS.

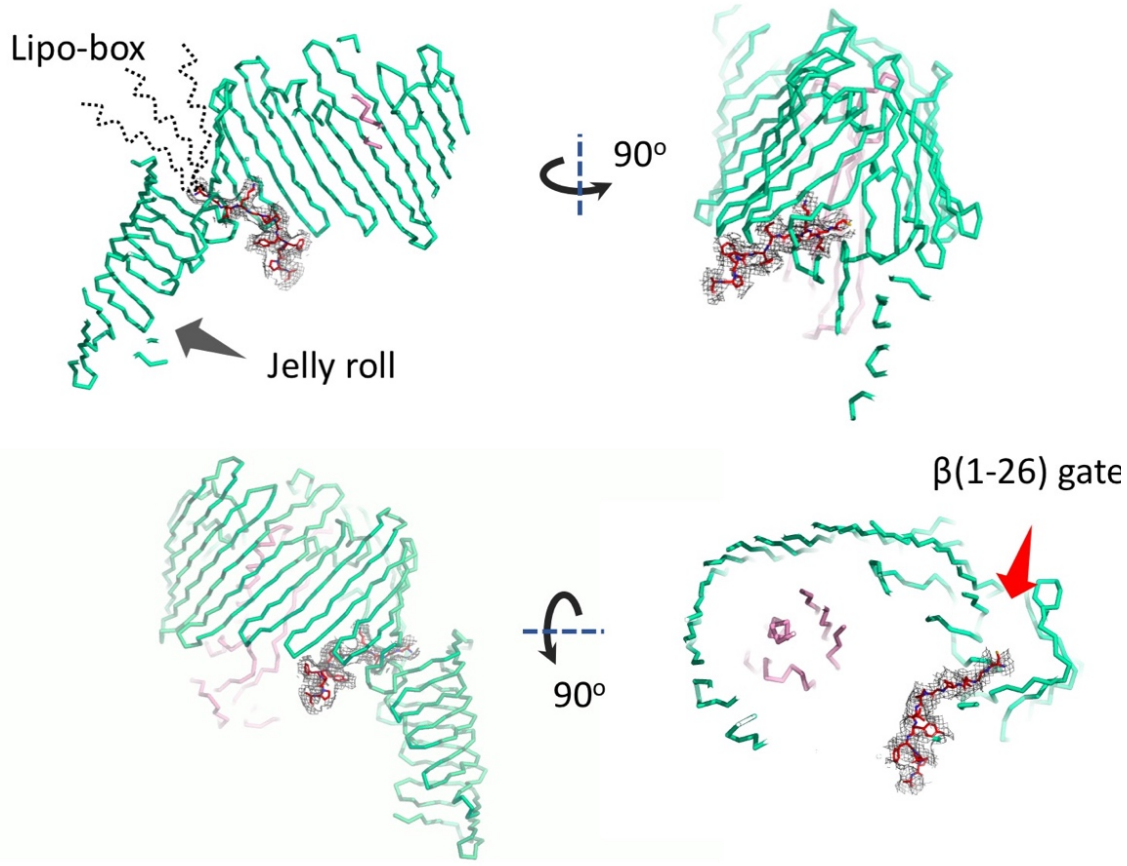


Figure 27 Unknown function (substrate like molecule) small lipoprotein YifL binding to the region close to LptD-barrel beta (1-26) gate. The electron microscopy density is around the bound peptide with the structure of LptDE complex. This peptide derived from the E.coil unknown functional lipoprotein called YifL. The peptide is colored red and showed as atoms. The small lipoprotein is inserted towards the detergent micelles through beta(1-26) gate.

At the position where the LPS core saccharide would be expected to enter the lumen of the β-barrel, we observed an elongated density between the periplasmic ends of β1 and β26 across β2 and along β3 and β4 towards the periplasmic rim of the barrel. This density does not fit the structure of the saccharide core of LPS, but it could be reliably interpreted as the lipoprotein YIFL from E. coli that was co-expressed with LPTDE to improve the expression yield and the purification quality of LptDE. We therefore modelled this density as peptide with the sequence of the YIFL lipoprotein (Figure 27). According to this interpretation, YIFL was binding in a shallow grove of the beta-barrel structure with multiple direct interactions stabilizing the local conformation in this area.

4.3.5 C-terminal loop structure

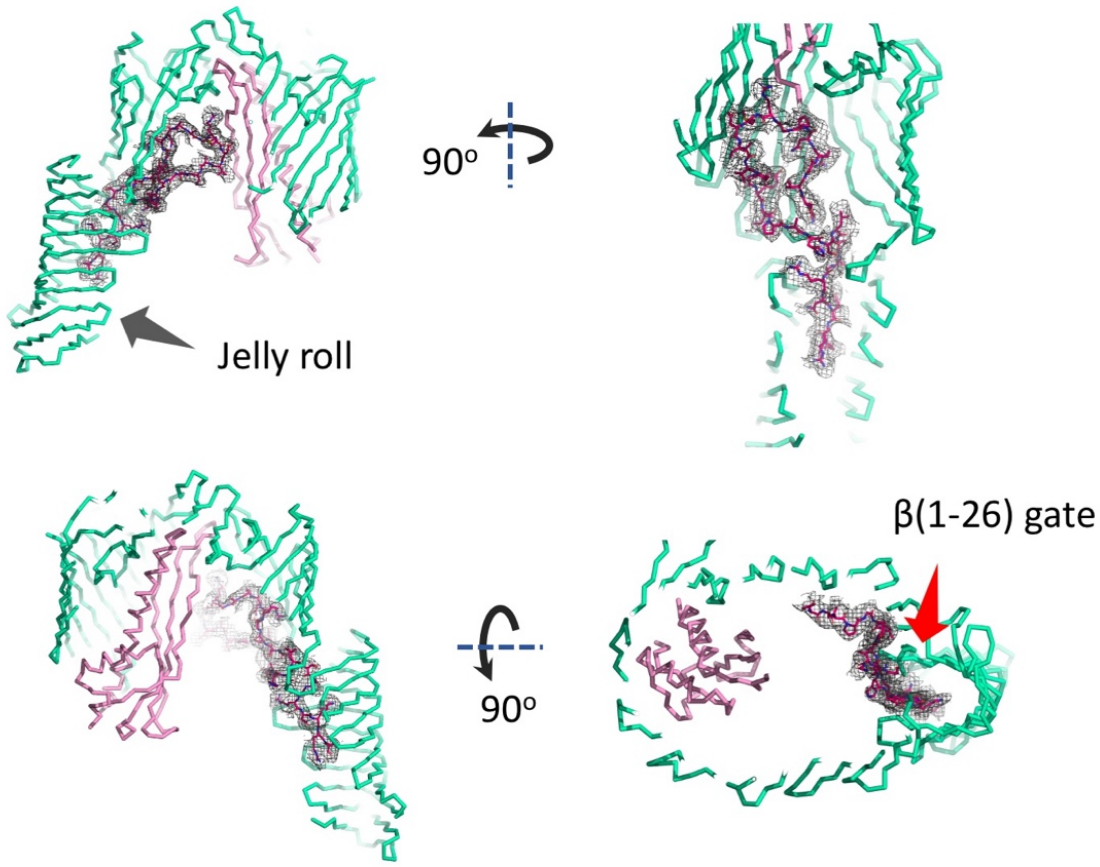


Figure 28 Structure of LptD C-terminus. The electron microscopy density is around C-terminus of LptD with the structure of the entire complex of LptDE. The LptD C-terminal loop is colored red and showed as atoms. The C-terminal tail is around the luminal gate and finally inserted into N-terminal Jellyroll domain deeply.

The 25 C-terminal residues of the LptD barrel domain and luminal loop 2 were not visible in our highest resolution cryo-EM map. In order to get a better understanding of the function of the luminal gate in LPS transport, we re-analyzed the cryo-EM data and focused the refinement processing to the area within a tight mask encompassing the N-terminal domain as well as β1-β26 region of LptD, only. The resulting cryo-EM map at 3.4 Å resolution allowed to trace the complete chain of the C-terminal region of LptD up to its C-terminus (Figure 28). The C-terminal region following β26 extends deeply into the lumen of the β-barrel towards the restriction separating the two lobes of the barrel and in proximity to LptE. The C-terminal residues from Asn798 to Pro801 bind into the groove of the N-terminal

jellyroll domain. In contrast to the crystal structures of SfLptD and KpLptDE, no helical region is observed in luminal loop 2.

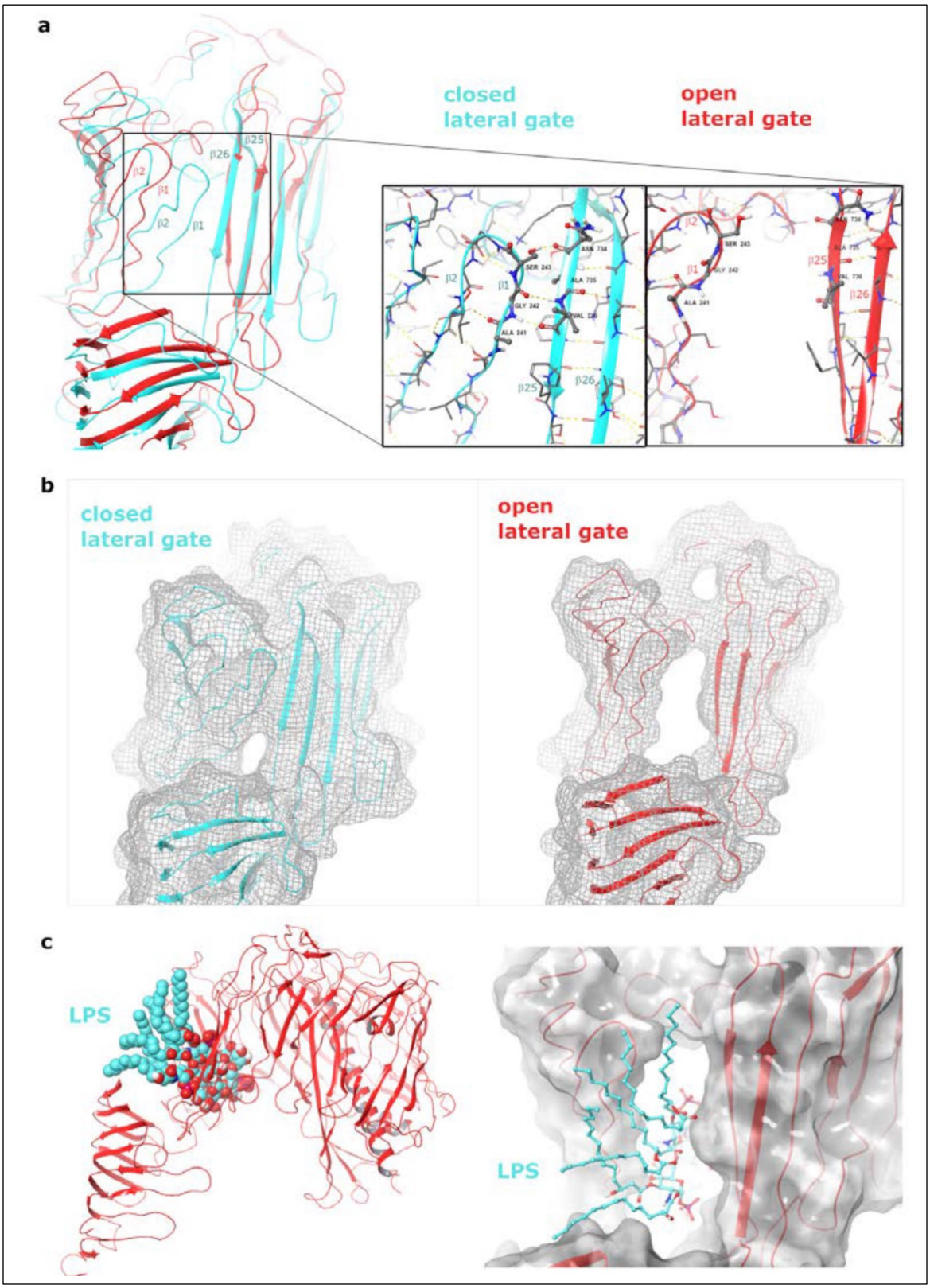


Figure 29 Opening of the lateral gate. (a) Overlay of open (red) and closed (cyan) structures with respect to the lateral gate between β1 and β26. The gate opening involves mainly the final N- and C-terminal strands of the barrel while strands β7 to β22 superimposed well with the high-resolution closed structure. The relative orientation of the jelly-roll domain does not change. The close-up view of the lateral gate formed by β1 and β26 shows hydrogen bonds that are broken in the open state between residues 241-243 on β1 and residues 734-736 on β26. (b) Calculated surface on the closed structure (cyan) and open (red) state. (c) LPS taken from the 3FXI crystal structure rigidly docked into the empty volume available in the open state, showing the availability of such volume only in the open conformation.

4.3.6 The opening of the lateral gate

Image processing assigned approx. one fifth of the particles on the cryo-EM grid to a structure with an open barrel that did not fit the conformation of the main particle population. We used this subset of particles to compute a cryo-EM map of NgLptDE with an open lateral gate at 4.72Å resolution. This map shows a full opening of the β-barrel with no H-bonds between β1 and β26, and a separation of 10Å between the extracellular ends and 15Å between the periplasmic ends of these strands. The overall diameter of the open barrel was not significantly different than the closed version. Only the six N-terminal and four C-terminal strands were shifted by more than 1.5Å (Figure 29a and b). Docking of LPS into this experimental structure indicated that the saccharidic portion of LPS likely enters the barrel and passes through it to the extracellular face, while lipid A would be inserted into the outer membrane (Figure 29c).

4.4 Discussion

4.4.1 PMb as structural chaperone tool

Pro-Macrobodies (PMb) were applied to elucidate the structure of the potential drug target LptDE that is involved in lipopolysaccharide (LPS) transport of Gram-negative bacteria. Up to now, beta-barrel proteins like LptDE with a molecular weight of about 100 kDa have been a significant challenge for structural analysis by cryo-EM, because the β-barrel architecture complicates the correct particle alignment in the cryo-EM image classification. The low molecular weight (120kDa) and pseudosymmetry of the 26-stranded β-barrel complicates accurate particle alignment for the LptDE complex alone, preventing successful cryo-EM studies as illustrated by the low resolution cryo-EM structure of NgLptDE (Figure 34). To overcome this challenge, we here generated sybodies that exhibit high affinities and address non-overlapping epitopes (Figure 32 and Figure 33). Suitable sybodies were converted to rigid PMbs of higher molecular weight, resulting in a quaternary complex of

LptD, LptE and two PMbs of approx. 240kDa, sufficient for cryo-EM analysis. This complex allowed structure determination by cryo-EM at 3.4Å (Figure 35), which allowed de-novo model building, positioning of side-chains and subclass structure elucidations such as the one for an opened lateral gate. As indicated by in-silico design and comparative 2D classification analysis (Figure 23 and Figure 24), the replacement of two linker residues with Prolines resulted in a much rigidified molecule compared to the previously described macrobodies (Brunner et al., 2020). As any nanobody can be converted into a PMb and produced readily in E. coli with high yield, the PMb scaffold represents a novel molecular chaperone for cryo-EM structure determination of small and challenging proteins, as well as crystallization facilitator for X-ray structure determination.

4.4.2 The structure of LPTDE and implications to transport function

LPS transport requires a transient lateral opening of the β-barrel that was proposed to allow the synchronized translocation of the hydrophobic and hydrophilic portions of the LPS molecule through the outer membrane (Dong et al., 2014). However, direct structural evidence for such a lateral opening was so far lacking.

From mutagenesis studies and MD simulations, the two conserved prolines Pro230 and Pro245 in YpLptD (corresponding in NgLptD to Pro261 and Pro275, respectively) were postulated to play a crucial role in opening (Botos et al., 2016). Mutating these prolines to alanines led to a decline in the bacteria colony count, and MD simulations of the double mutant showed the improved H-bonding between β1 and β26 hampering barrel opening. In addition to insights on the transport of LPS from numerous structures of Lpt pathway members and functional investigations, the two cryo-EM structures presented here, together with the dynamic connection of the N-terminal domain to the beta-barrel, potential regulatory roles of the LptD C-terminus (Figure 28), and most importantly the first observation of the opening of the lateral gate in LptD (Figure 29), corroborate the proposed mechanism as both prolines are found to perturb the β-strand geometry of β1 and β2, presumably due to the lack of an available NH group. The combined effect of the two prolines is to prevent the lower segment of β1 from forming typical β-sheet hydrogen bond interactions, thus facilitating strand separation, which is observed in the open form.

MD simulations on the LptD structure and mutational studies of the O-antigen and the core oligosaccharide of LPS support that LPS translocates through the lumen of the barrel domain, while the lipid portion of LPS would enter directly into the outer membrane through the opening of a transient lateral gate between β1 and β26 (Botos et al., 2016; Dong et al.,

2014; Qiao et al., 2014a). The saccharidic portion of LPS was proposed to enter the barrel lumen through the luminal gate formed by the two periplasmic loops adjacent to β1 and β26 (Gu et al., 2015). The experimental evidence of a laterally open structure of the LptDE complex (Figure 29) suggests that the opening of the transmembrane β-barrel that was postulated for the insertion of LPS into the outer leaflet of the bacterial outer membrane works similarly as observed in the other outer membrane β-barrel proteins FadL, PagP, OmpW, and OprG (Hearn et al., 2009; Hong et al., 2006; Khan & Bishop, 2009; Touw et al., 2010; van den Berg et al., 2004).

Overall, the LptDE structure presented here shows a wider diameter of the NgLptDE barrel lumen, likely reflecting the more open conformation of the luminal gate (distance between β1 and β26 increased by 3 Å at the periplasmic side), representing a specific and until now unobserved stage of LPS transport.

Qiao et al. reported in a previous study of the Shigella protein LptDE the backfolding of the N-terminal loop to the N-terminal domain of LptD (Qiao et al., 2014b). For NgLptDE the C-terminus of LptD can be detected in the EM density well showing back-folding to a very similar area on the N-terminal domain. We put forward the hypothesis that LptD protein termini are used for auto-regulation purposes of LPS transport. Such activity regulation sites could represent a potential new site for a drug interaction inhibiting the LPS transport (Brunner et al., 2020; Bloch et al., 2020; Butterwick et al., 2018; Manolaridis et al., 2018)(Qiao et al., 2014b; Botos et al., 2016).

E. coli YIFL lipoprotein could be identified as ligand in this high-resolution EM structure. The peptide binds to the luminal gate of LptD, suggesting that the LPTDE transporter might also be acting as transporter for lipoproteins or YIFL is involved in the regulation of LPS transport. The observed stabilization of the local conformation of the beta-barrel and much improved protein expression yield by the co-expression of YIFL support an overall stabilizing effect of the peptide (Figure 27). Mutagenesis studies at the binding sites of the C-termini of LptDE, as well as of YIFL are needed to shine more light onto their function.

Even though the monomeric fraction of NgLptDE was used for the complex formation with PMb21 and PMb51, the analysis of the cryo-EM data revealed a fraction of molecules in a dimeric oligomeric state with a dimer interface right at the top of the N-terminal LptD jelly-roll domain (Figure 36). In crystals of KpLptDE and SfLptDE, crystal contacts via the N-

terminal LptD were observed as well. However, these interactions were formed by the crystal lattices and no continuous groove between the two N-terminal domains was formed in those studies. In our cryo-EM structure, the dimer was formed by a pseudo 2-fold symmetry and generated a continuous groove between the two beta jelly roll domains. This further corroborated the oligomerization of jelly-roll domains of the Lpt-pathway members LptD, LptA and LptC as a construction principle of the continuous hydrophobic slide across the periplasm (Botos et al., 2016). However, probably due to species related specificity, we could not detect a stable complex of Neisseria gonorrhoeae LptD with endogenous Escherichia coli LptA for a direct observation of a heterodimer of jelly-roll domains . By performing a focused refinement using a soft-mask delimiting the dimerization interface, we were able to solve the structure of dimeric LptD N-terminal domains at 3.93Å resolution. In this structure, the residues Tyr88 to Val102 are not visible and the dimerization interface is formed by the residues Lys103 to Glu113 (Figure 36 and Figure 37). We hypothesized that the residues 88 to 102 would be displaced to allow formation of the dimer. As LptA is structurally highly similar to the N-terminal domain of LptD (Qiao et al, 2014) we therefore suggest that the residues Lys103 to Glu113 could putatively be involved in the complex formation with LptA in vivo.

In addition, from our cryo-EM density map, the bound detergent micelle is well visualized, leading to a direct observation for the membrane association of LptD N-terminal domain. By measuring the thickness of detergent micelle, we hypothesize that the LPS transporting bridge is likely accessible to the outer leaflet of outer membrane, consequently delivering and inserting the LPS molecules into the outer membrane outer leaflet directly. While the O-antigen chain is facing towards the extracellular side (Figure 38). Interestingly, we were able to observe an undefined extra density from all of our resulted cryo-EM maps, where it attached to the bottom of LptE, facing towards the periplasmic space (Figure 39). However, it is still unsure what it is really bound in our samples and it is likely a specific binding ligand. By docking an O-antigen molecular chain, we hypothesize that the unexpected density is possibly the O-antigen part of the E. coil endogenous lipopolysaccharides, while the further experiment to explore the real bound molecules and the O-antigen binding site will be highly required in the future investigations.

4.5 Methods:

4.5.1 Expression and purification of the Neisseria gonorrhoeae LptDE complex (NgLptDE)

N. gonorrhoeae wild type full length LptD (UniProt: Q5F651) LptE (UniProt: Q5F9V6), with a hex-histidine tag on the c-terminus, and E. coli YifL (UniProt: P0ADN6) were cloned into the expression vector pBAD22A (graciously provided by Professor Yihua Huang). NgLptDE complex was co transformed in SF100-deltaYifL E. coli cells (graciously provided by Professor Yihua Huang). After transformation, pre-culture was started, and cells were grown overnight at 37°C with 100µg/ml ampicilin and 50µg/ml kanamycin. Large LB broth culture were then inoculated at a starting OD600nm=0.05 and grown at 37°C with 100 µg/ml ampicillin. When OD600nm=0.8, temperature was switched to 20°C and induction was performed at OD600nm=1 with 0.4 % arabinose. After overnight induction at 20°C, cells were harvested and resuspended in lysis buffer consisting of 200 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM MgCl2, 10 µg/ml DNAse I, 100 µg/ml AEBSF [4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride] and protease inhibitors cocktail (Complete, Roche). Cells were cracked by multiple passages through a microfluidizer system, using a pressure of 18'000 psi, and the lysate was centrifuged at 7'500 x g for 10 minutes to remove the cell debris. The supernatant was collected, and inner membranes were solubilized by incubation with 2% Triton X-100 for 30 minutes at 4°C under gentle agitation. The outer membrane fraction was collected by centrifugation at 100'000 x g for 30 minutes at 4°C. The pellet containing the outer membrane fraction was resuspended in 200 mM NaCl, 50 mM Tris-HCl pH 7.5, 20 mM imidazole, 1 % LDAO [lauryldimethylamine oxide], supplemented with protease inhibitors cocktail (Complete, Roche) and incubated under gentle agitation for 12 hours at 4°C. Insoluble material was removed by centrifugation at 100'000 x g for 1 hour at 4°C. The supernatant was incubated in batch with ~ 5 ml NiNTA resin for 2 hours at 4°C under gentle agitation. The resin was subsequently washed by gravity flow with 10 column volumes (CV) of Wash buffer A (200 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole, 1 % LDAO), 10 CV of Wash buffer B (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 40 mM imidazole, 0.5 % LDAO), 10 CV of Wash buffer C (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 40 mM imidazole, 0.2 % LDAO) and 10 CV of Wash buffer D (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 40 mM imidazole, 0.1 % LMNG [lauryl maltose neopentyl glycol]). Elution was performed with 5 CV of Elution buffer (150 mM NaCl, 20 mM Tris-HCl pH 8.0, 300 mM imidazole, 0.01 % LMNG). Eluted material was desalted against 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.005% LMNG using disposable PD-10 desalting columns (GE Healthcare) and was concentrated

in 100 kDa centrifugal concentrators (Millipore). Concentrated sample was further purified by size exclusion chromatography on a Superdex 200 Increase column, equilibrated to 150 mM NaCl, 20 mM Tris-HCl pH8.0 and 0.005 % LMNG. The peak fractions corresponding to the monomeric or dimeric LptDE complex were kept separately, concentrated to ~ 1 mg/ml and use for sybody generation, as well as biophysical and structural studies.

4.5.2 Protein used for sybodies generation and grating coupled Interferometry characterization.

N. gonorrhoeae (Zopf) Trevisan (ATCC 700825) LptD (1-801)-3C-His10 and LptE (1-159)-AVI complex was produced, using E.coli SF100 cells, and purified as described above with an enyzmatic biotinylation step added after the size exclusion chromatography (SEC) step. Briefly, LptDE protein was concentrated to 10 µM and mixed with 40 µg of BirA E.coli enzyme, 5 mM ATP, 10 mM Magnesium acetate and 15 µM biotin. The mixture was incubated 16 hours at 4°C and a second SEC step was performed to desalt the sample and remove the BirA enzyme and free biotin in the following buffer: 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.005 % LMNG. Fractions corresponding to the monomeric complex peak were pooled, concentrated and flash frozen in liquid nitrogen for subsequent use.

4.5.3 Sybody generation

Sybodies were generated as described previously (Zimmermann et al., 2018, 2020) with one notable difference. NgLptDE could not be produced in large amounts, which did not allow to use a large excess of non-biotinylated NgLptDE for off-rate selection. Therefore, we pooled all the sybodies present after the first round of phage display and used them as competitors during the second round of phage display to perform an off-rate selection. To this end, the the phagemids (pDX_init vector) encoding the sybodies resulting from the first round of phage display were isolated. FX cloning was performed to transfer the sybody pool from the pDX_init to the pSb_init expression plasmid, using 2 µg of pDX_init pool and 1 µg of pSb_init. The cloning reaction was subsequently transformed into electrocompetent E. coli MC1061 cells (> 10 mio cfu). The sybody pools were expressed as described for single sybodies using the pSb_init construct (Zimmermann et al., 2020). After expression of the pools in 600 ml cultures of TB medium, the sybodies were extracted from the cells by periplasmic extraction, purified by IMAC and dialyzed overnight against Tris buffered saline (TBS). Precipitation was removed by centrifugation at 20’000 x g for 15 min. The

pools were used at a concentration of approximately 100 µM to perform an off-rate selection for 2 min in the second round of phage display.

4.5.4 Fluorescent labelling of sybodies

To perform site specific labelling of the sybodies, a cysteine mutation was introduced into the linker region between the sybody and the myc/his tag in the pSb_init construct, namely four amino acids after the C-terminus of the sybody. The sybodies were expressed and purified as previously described (Zimmermann et al., 2020), except that 1 mM DTT was added to all buffers used for purification. Subsequently, DTT was removed and the sybody was re-buffered to PBS using a PD10 desalting column and immediately mixing the sybody with Alexa Fluor 647 C2 maleimide (ThermoFisher scientific) at a molar ratio of 1:1.5. The labelling reaction was carried out overnight on ice. Excess label was removed by desalting the labelled sybody with a PD10 column.

4.5.5 Cellular binding assay

For cellular binding assays, overnight cultures of E. coli SF100 cells with and without overexpression of NgLptDE were used. The number of cells were normalized by adjusting 1 ml of culture to an OD600 of 3. The cells were harvested by centrifugation, washed three times with 500 µl PBS containing 0.5 % BSA (PBS-BSA), and subsequently blocked for 20 min in the same buffer. After an additional wash with 500 µl PBS-BSA, the cells were incubated for 20 min in 100 µl PBS-BSA containing 1 µM of the Alexa Fluor 647-labelled sybodies. After three washes with 500 µl PBS, cells were resuspended in 100 µl PBS and transferred to a microtiter plate with non-transparent walls. Fluorescence was measured in a plate reader with excitation of 651 nm and emission of 671 nm.

4.5.6 Antibiotic susceptibility assay

N. gonorrhoeae (Zopf) Trevisan (ATCC 700825) were streaked from a glycerol stock on blood agar and incubated for 24 hours at 37 °C with 5 % CO2 atmosphere. Colonies were scraped off the agar and resuspended in Fastidious broth at a density of McFarland 0.5 (Jacobson et al., 2019). The cells were further diluted 1:100 in Fastidious broth. In 96 well plates, dilution series of vancomycin with and without sybodies in Fastidious broth were prepared and mixed with the diluted culture. The plates were incubated without shaking at 37 °C with 5 % CO2 atmosphere for 24 hours. 100 µl of a 0.04 mg/ml resazurin stock

solution in PBS was added to the cells and incubated for one hour at 37 °C with 5 % CO2 atmosphere. Fluorescence was measured at 571 nm excitation and 585 nm emission.

4.5.7 Pro-Macrobody generation

Pro-Macrobodies (PMbs) were produced in E.coli as described earlier for the original macrobodies (Brunner et al. 2020). Briefly, two PCR-amplified fragments, the specific sybody and the C-terminal MBP were cloned in parallel into the expression vector pBXNPH3M (Addgene) using FX cloning (Geertsma, 2013). N-terminally of the resulting PMb insert, the plasmid expresses a pelB leader sequence followed by a deca-His tag, an MBP and a 3C site. The insert is fused during cloning through an overlapping proline-encoding CCG codon introduced by reverse and forward primers at the 3'- and 5'-end of the sybodies and MBP respectively and released by digestion with the Type IIS restriction enzyme SapI (NEB). The second proline of the linker is encoded in the forward primer of the MBP (3' of the overlapping CCG codon) and replaces the natural lysine. The resulting amino acid sequence of the linker is VTVPPLVI (VTV is the conserved C-terminus of sybodies, PP in italics/bold denotes the linker and underlined the truncated N-terminus of processed E.coli malE starting at Leu7). PMbs were expressed in terrific broth in MC1061 E.coli cells at 37 °C by induction with 0.02% arabinose at an OD600=0.7. After 3.5 hours, cells were harvested and resuspended in lysis buffer consisting of 150 mM NaCl, 50 mM Tris-HCl pH 8, 20 mM imidazole, 5 mM MgCl2, 10% glycerol, 10 µg/ml DNAse I and protease inhibitors (Complete, Roche). Cells were cracked and the lysate centrifuged for 30 min at 147'000 x g in a Beckman 45Ti rotor and the supernatant incubated in batch with ~ 3ml NiNTA resin/ 6L of culture. The resin was subsequently washed with 150 mM KCl, 40 mM imidazole pH 7.6, 10% glycerol and eluted with 150 mM KCl, 300 mM imidazole pH 7.6, 10% glycerol. The N-terminal MBP with the deca-His tag was removed by cleavage with 3C protease overnight during dialysis against 150 mM KCl, 10 mM Hepes-NaOH, 20 mM imidazole, pH 7.6, 10% glycerol. After removal of the His-tagged MBP by Re-IMAC, the unbound material was concentrated in 50 kDa centrifugal concentrators (Millipore) and subjected to SEC on a Superdex 200 Increase column (equilibrated 150 mM NaCl, 10 mM Hepes-Na, pH 7.6). The eluted fractions with the PMbs were supplemented with 20% glycerol, concentrated to 3-8 mg/ml and aliquots were flash-frozen in liquid N2 for subsequent use. Macrobody versions of sybodies 21 and 51 with the original VK-linker were expressed and purified in the same way.

4.5.8 SEC analysis/ purification of LptDE-PMb complexes

To identify ternary and quaternary complexes of NgLptDE with various PMbs against NgLptDE, monomeric NgLptDE was first separated from dimers by SEC on a Superdex 200 Increase column (10/300) in 150 mM NaCl, 20 mM Tris-Cl, pH 8, 0.005% LMNG. A complex was formed by addition of 3-4 folds molar excess of PMb51 to NgLptDE sample. The uncomplexed and PMb-bound samples were analyzed on an Agilent 1260 Infinity II HPLC using a Superdex 200 Increase (5/150) column and Trp-fluorescence detection after 10-15 min incubation time on ice. Binding was indicated by a shift to earlier elution volumes and increase in the UV-absorption and Trp-fluorescence. The quaternary complex was formed by adding PMb21 in 3-4 folds molar excess to the preformed NgLptDE-PMb51 complex. The quaternary NgLptDE-PMb21-PMb51 complex showed the most distinct shift and fluorescence increase, and was therefore chosen for scale up and cryo-EM analysis.

4.5.9 Binding analytics by Grating Coupled Interferometry

An initial screen with ELISA positive sybodies was performed at 20 °C on the Wave delta instrument from Creoptix in Tris 25 mM, pH 7.5, NaCl 300 mM, DDM 0.1% as running buffer. Biotinylated protein was immobilized on a 4PCP-S (streptavidin) chip, conditioned with 1 M NaCl, 0.1 M sodium borate, at levels between 600 and 700 pg/mm2 to avoid any mass transport limitation. One injection of each sybody (200 nM) was performed and binding responses were evaluated with the Wave control software. The ten best sybodies in term of slower Koff as well as quality of binding signals obtained were characterized further with dose-response analysis in order to determine accurate kinetic parameters. For each ten sybody, 8 concentrations were recorded (serial 2-fold dilution) in duplicate injections, and equilibrium as well as kinetic data were analyzed and fitted with 1:1 model. Fits were of high quality (black curves) and recapitulate the experimental data (red curves). At higher concentration of sybodies, bulk (RI) effects could be observed, but this effect did not influence data analysis. PMb 21 showed a slightly decreased affinity (3 to 4-fold) compared to the respective sybody 21. This loss can be explained by a faster Koff of the PMb21.

4.5.10 Sample preparation and Cryo-EM data acquisition

Quantifoil (1/2) 200-mesh copper grids were glow-discharged for 20 seconds prior to sample freezing. 3 µl of NgLptDE-PMb21-PMb51 complex at a concentration of 1 mg/ml were placed on the grid, blotted for 3.0 s and flash frozen in a mixture of liquid propane and

liquid ethane cooled with liquid nitrogen, using a Vitrobot Mak IV (FEI) operated at 4 °C and under 100% humidity.

The EM data collection statistics in this study are reported in Table 5. Data were recorded on a FEI Titan Krios transmission electron microscope, operated at 300 kV and equipped with a Quantum-LS energy filter (slit width 20 eV; Gatan Inc.) containing a K2 Summit direct electron detector. Data were automatically collected using the software SerialEM (Mastronarde, 2005). Dose-fractionated exposures (movies) were recorded in electron-counting mode, applying 60 electrons per square Angstrom (e-/Å2) over 45 frames, or 50 e-/Å2 over 35 frames for, respectively, the NgLptDE (apo) or the NgLptDE-PMb21-PMb51 samples. A defocus range of -0.8 to -2.8 µm was used and the physical pixel size was 0.64 Å/pixel for the NgLptDE and 0.82 Å/pixel for the NgLptDE-PMb21-PMb51 datasets. Recorded data were online analyzed and pre-processed using FOCUS (Biyani et al., 2017), which included gain-normalization, motion-correction, and calculation of dose-weighted averages with MotionCor2 (Zheng et al., 2017), as well as estimation of micrograph defocus with CTFFIND4 (Rohou & Grigorieff, 2015).

4.5.11 Image processing

The following processing workflows were used for the samples in the study. The aligned movies were imported into CryoSPARC V2 (Punjani et al., 2017). A set of aligned averages with a calculated defocus range of −0.6 to –3.0 µm was selected, from which averages with poor CTF estimation statistics were discarded. Automated particle picking in CryoSPARC V2 resulted in 815,057 particle locations for the NgLptDE-PMb21-PMb51 sample. After several rounds of 2D classification, 490,743 particles were selected and subjected to 3D classification using the multi class ab initio refinement process (5 classes; 0.4 similarity) and heterogenous refinement. The best resolved class consisting of 184,206 particles was finally subjected to 3D non-uniform refinement. The overall resolution of the resulted map was estimated at 3.4 Å based on the Fourier shell correlation (FSC) at 0.143 cutoff (Scheres & Chen, 2012). To visualize the LptD-NTD dimerization interface, another round of 3D heterogenous refinement was performed and a subset consisting of 80,140 particles was selected. Those particle coordinates were used for re-extraction of particle images with an increased box size. Particles were re-centered by 2D classification in order to process the dimeric LptDE complex. Ab initio reconstruction and non-uniform refinement on this set of particles resulted in a DIMERNgLptDE-PMb21-PMb51 map with an overall resolution of 5.27 Å. After further reanalyzing the data by applying a soft mask and performing a 3D

local refinement, a cryo-EM map focused on the dimerization interface was obtained at an overall resolution of 3.93 Å based on the Fourier shel correlation (FSC) at 0.143 cutoff. As for the opened state, a multi class ab initio refinement and heterogenous refinement was performed for the NgLptDE-PMb51-PMb21 sample. Of the five 3D classes, one class consisting of 93,151 particles was further refined by computationally removing the density corresponding to the detergent micelle with the particle subtraction tool within CryoSPARC V2, followed by 3D local refinement. The resulting map had an estimated overall resolution of 4.72 Å as judged by FSC at 0.143 cutoff. Analysis of the apo NgLptDE was performed similarly to the NgLptDE-PMb21-PMb51. Briefly, a set of aligned averages with a calculated defocus range of −0.6 to –3.0 μm was selected, from which averages with poor CTF estimation statistics were discarded. Automated particle picking in CryoSPARC V2 resulted in 1,395,392 particle locations for the NgLptDE. After several rounds of 2D classification, 196,182 particles were selected and subjected to 3D classification using the multi class ab initio refinement process (3 classes; 0.1 similarity) and heterogenous refinement for the 2 best classes. The best resolved class consisting of 119,115 particles was finally subjected to 3D non-uniform refinement. The overall resolution of the resulted map was estimated at 4.6 Å based on the Fourier shell correlation (FSC) at 0.143 cutoff

4.5.12 Model building and refinement

An initial LptDE model was generated using SWISS-MODEL, using as templates the KpLptD structure (PDB-ID 5IV9) and the EcLptE structure (PDB-ID 4RHB). The template for building the PMb21 coordinates into the EM map was based on an X-ray structure that was solved specifically for this study. The same structure was also used to model the maltose binding protein region of PMb51.

Rigid body fitting was initially done in Chimera (Pettersen et al., 2004) followed by manually rebuilding of the model in Coot (Emsley et al., 2010). Remaining clashes between sidechains were detected using Schrodinger version 2019-4, and remodeled using prime (Zhu et al., 2014). Manual inspection of missing H-bonds in the model was used to refine sidechain positions. Finally, real-space refinement was performed in Phenix version 1.17-3644 , applying Ramachandran plot restraints (Liebschner et al., 2019).

4.5.13 Figure preparation

Figures were prepared using the programs Chimera X (http://www.rbvi.ucsf.edu/chimerax/ (Goddard et al., 2018), Chimera (http://www.cgl.ucsf.edu/chimera/ (Pettersen et al., 2004), PyMOL (http://www.pymol.org/) and Schrodinger (www.schrodinger.com).

4.5.14 Data availability

EM raw image data were deposited in the Electron Microscopy Public Image Archive (EMPIAR) under accession number EMPIAR-XXXX. EM maps were deposited in the Electron Microscopy Data Bank under accession codes EMD-XXXX, EMD-XXXX, EMD-XXXX, EMD-XXXX and EMD-XXXX. Atomic coordinates for NgLptDE from the cryo-EM study were deposited in the Protein Data Bank under accession codes PDB-XXXX. The X-ray structure of PMb51 was deposited with accession codes PDB-XXXX. All other data are available from the corresponding authors upon reasonable request.

Acknowledgement

We thank Armin Ruf for editing and critical reading of the manuscript. Lubomir Kovaczik is acknowledged for help with automation of EM data collection. We thank Yihua Huang for very valuable scientific discussion of the project. The research was supported by an Innosuisse grant KTI 25864.1, and by the Swiss National Science Foundation (NCCR TransCure).

Author contribution

MB (Mathieu Botte): conceived the experiments, performed protein purification for EM experiments, analyzed data and co-wrote the initial draft of the manuscript

DN (Dongchun Ni): performed the sample preparation, recorded EM data, performed image processing and 3D reconstruction, supported the initial draft of the manuscript.

SS Design of Pro-Macrobody together with DB, generation and purification of the Pro-Macrobodies

MC (Mohamed Chami): Supported EM data collection.

NB (Nicolas Bocquet) Analysis of Sybody and Pro-Macrobody ligand binding with grating coupled interferometry

DB (Denis Bucher) Design of the linker for the Pro-Macrobodies, EM structure building and figure preparation.

RC (Robert Cheng) X-ray structure of the Pro-Macrobody 21

MT (Matilde Trabuco) X-ray structure of the Pro-Macrobody 21

JDB (Janine D. Brunner) Design and characterization of Macrobodies, proposed Macrobodies for cryo-EM studies

HS (Henning Stahlberg): Provided resources, supervised electron microscopy study.

MH (Michael Hennig): conceived the project, wrote grant application, supervision, data review and coordination of the project, co-wrote the initial draft of the manuscript

MAS (Markus Andreas Seeger): conceived the project, wrote grant application, interpreted data, co-wrote the initial draft of the manuscript, supervision of students/postdocs.

IZ (Iwan Zimmermann): Generated sybodies, established off-rate selection with purified sybodies to save LptDE, purified sybodies, established and performed cellular binding assay, established and performed N. gonorrhoeae growth inhibition assay.

PE (Pascal Egloff): supported MB to establish enzymatic biotinylation of NgLptDE.

All authors edited the manuscript and agreed with the content.

4.6 References

Andolina, G., Bencze, L. C., Zerbe, K., Müller, M., Steinmann, J., Kocherla, H., Mondal, M., Sobek, J., Moehle, K., Malojčić, G., Wollscheid, B., & Robinson, J. A. (2018). A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from Pseudomonas aeruginosa. ACS Chemical Biology, 13(3), 666–675. https://doi.org/10.1021/acschembio.7b00822

Biyani, N., Righetto, R. D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K. N., & Stahlberg, H. (2017). Focus: The interface between data collection and data processing in cryo-EM. Journal of Structural Biology, 198(2), 124–133. https://doi.org/10.1016/j.jsb.2017.03.007

Bloch, J. S., Pesciullesi, G., Boilevin, J., Nosol, K., Irobalieva, R. N., Darbre, T., Aebi, M., Kossiakoff, A. A., Reymond, J.-L., & Locher, K. P. (2020). Structure and mechanism of the ER-based glucosyltransferase ALG6. Nature, 579(7799), 443–447. https://doi.org/10.1038/s41586-020-2044-z

Botos, I., Majdalani, N., Mayclin, S. J., McCarthy, J. G., Lundquist, K., Wojtowicz, D., Barnard, T. J., Gumbart, J. C., & Buchanan, S. K. (2016). Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. Structure, 24(6), 965–976. https://doi.org/10.1016/j.str.2016.03.026

Brunner, J. D., Jakob, R. P., Schulze, T., Neldner, Y., Moroni, A., Thiel, G., Maier, T., & Schenck, S. (2020). Structural basis for ion selectivity in TMEM175 K+ channels. ELife, 9, 1–24. https://doi.org/10.7554/eLife.53683

Butterwick, J. A., del Mármol, J., Kim, K. H., Kahlson, M. A., Rogow, J. A., Walz, T., & Ruta, V. (2018). Cryo-EM structure of the insect olfactory receptor Orco. Nature, 560(7719), 447–452. https://doi.org/10.1038/s41586-018-0420-8

Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W., & Dong, C. (2014). Structural basis for outer membrane lipopolysaccharide insertion. Nature, 511(7507), 52–56. https://doi.org/10.1038/nature13464

Freinkman, E., Chng, S. S., & Kahne, D. (2011). The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. Proceedings of the National Academy of Sciences of the United States of America, 108(6), 2486–2491. https://doi.org/10.1073/pnas.1015617108

Freinkman, E., Okuda, S., Ruiz, N., & Kahne, D. (2012). Regulated Assembly of the Transenvelope Protein Complex Required for Lipopolysaccharide Export. Biochemistry, 51(24), 4800–4806. https://doi.org/10.1021/bi300592c

Grabowicz, M., Yeh, J., & Silhavy, T. J. (2013). Dominant negative LptE mutation that supports a role for LptE as a plug in the LptD barrel. Journal of Bacteriology, 195(6), 1327–1334. https://doi.org/10.1128/JB.02142-12

Gu, Y., Stansfeld, P. J., Zeng, Y., Dong, H., Wang, W., & Dong, C. (2015). Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of lptd. Structure, 23(3), 496–504. https://doi.org/10.1016/j.str.2015.01.001

Hearn, E. M., Patel, D. R., Lepore, B. W., Indic, M., & van den Berg, B. (2009). Transmembrane passage of hydrophobic compounds through a protein channel wall. Nature, 458(7236), 367–370. https://doi.org/10.1038/nature07678

Hong, H., Patel, D. R., Tamm, L. K., & van den Berg, B. (2006). The outer membrane protein OmpW forms an eight-stranded β-barrel with a hydrophobic channel. Journal of Biological Chemistry, 281(11), 7568–7577. https://doi.org/10.1074/jbc.M512365200

Hunte, C., & Michel, H. (2002). Crystallisation of membrane proteins mediated by antibody fragments. Current Opinion in Structural Biology, 12(4), 503–508. https://doi.org/10.1016/s0959-440x(02)00354-8

Jacobson, R. K., Notaro, M. J., & Carr, G. J. (2019). Comparison of Neisseria gonorrhoeae minimum inhibitory concentrations obtained using agar dilution versus microbroth dilution methods. Journal of Microbiological Methods, 157, 93–99. https://doi.org/10.1016/j.mimet.2019.01.001

Khan, M. A., & Bishop, R. E. (2009). Molecular mechanism for lateral lipid diffusion between the outer membrane external leaflet and a β-barrel hydrocarbon ruler. Biochemistry, 48(41), 9745–9756. https://doi.org/10.1021/bi9013566

Kovari, L. C., Momany, C., & Rossmann, M. G. (1995). The use of antibody fragments for crystallization and structure determinations. Structure (London, England : 1993), 3(12), 1291–1293. https://doi.org/10.1016/s0969-2126(01)00266-0

Laverty, D., Desai, R., Uchański, T., Masiulis, S., Stec, W. J., Malinauskas, T., Zivanov, J., Pardon, E., Steyaert, J., Miller, K. W., & Aricescu, A. R. (2019). Cryo-EM structure of the human α1β3γ2 GABAA receptor in a lipid bilayer. Nature, 565(7740), 516–520. https://doi.org/10.1038/s41586-018-0833-4

Lehman, K. M., & Grabowicz, M. (2019). Countering gram-negative antibiotic resistance: Recent progress in disrupting the outer membrane with novel therapeutics. Antibiotics, 8(4). https://doi.org/10.3390/antibiotics8040163

Liebschner, D., Afonine, P. v, Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. v, … Adams, P. D. (2019). Macromolecular structure

determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallographica. Section D, Structural Biology, 75(Pt 10), 861–877. https://doi.org/10.1107/S2059798319011471

Li, X., Gu, Y., Dong, H., Wang, W., & Dong, C. (2015). Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the Escherichia coli outer membrane. Scientific Reports, 5(February), 1–8. https://doi.org/10.1038/srep11883

Malojčić, G., Andres, D., Grabowicz, M., George, A. H., Ruiz, N., Silhavy, T. J., & Kahne, D. (2014). LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface. Proceedings of the National Academy of Sciences of the United States of America, 111(26), 9467–9472. https://doi.org/10.1073/pnas.1402746111

Manolaridis, I., Jackson, S. M., Taylor, N. M. I., Kowal, J., Stahlberg, H., & Locher, K. P. (2018). Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. Nature, 563(7731), 426–430. https://doi.org/10.1038/s41586-018-0680-3

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology, 152(1), 36–51. https://doi.org/10.1016/j.jsb.2005.07.007

Narita, S. ichiro, & Tokuda, H. (2009). Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. FEBS Letters, 583(13), 2160–2164. https://doi.org/10.1016/j.febslet.2009.05.051

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited. 67(4), 593–656. https://doi.org/10.1128/MMBR.67.4.593

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G. F., Wohlkönig, A., Ruf, A., Muyldermans, S., Hol, W. G. J., Kobilka, B. K., & Steyaert, J. (2014). A general protocol for the generation of Nanobodies for structural biology. Nature Protocols, 9(3), 674–693. https://doi.org/10.1038/nprot.2014.039

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods, 14(3), 290–296. https://doi.org/10.1038/nmeth.4169

Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C., & Huang, Y. (2014a). Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. Nature, 511(7507), 108–111. https://doi.org/10.1038/nature13484

Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C., & Huang, Y. (2014b). Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. Nature, 511(7507), 108–111. https://doi.org/10.1038/nature13484

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. Journal of Structural Biology, 192(2), 216–221. https://doi.org/10.1016/j.jsb.2015.08.008

Ruiz, N., Gronenberg, L. S., Kahne, D., & Silhavy, T. J. (2008). Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. Proceedings of the National Academy of

Sciences of the United States of America, 105(14), 5537–5542. https://doi.org/10.1073/pnas.0801196105

Ruiz, N., Kahne, D., & Silhavy, T. J. (2009). Road of Discovery. Nature Reviews Microbiology, 7(9), 677–683. https://doi.org/10.1038/nrmicro2184.Transport

Scheres, S. H. W., & Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. In Nature methods (Vol. 9, Issue 9, pp. 853–854). https://doi.org/10.1038/nmeth.2115

Sperandeo, P., Villa, R., Martorana, A. M., Maria, S., Grandori, R., Deho, G., & Polissi, A. (2011). New Insights into the Lpt Machinery for Lipopolysaccharide Transport to the Cell Surface : LptA-LptC Interaction and LptA Stability as Sensors of a Properly Assembled Transenvelope Complex □. 193(5), 1042–1053. https://doi.org/10.1128/JB.01037-10

Srinivas, N., Jetter, P., Ueberbacher, B. J., Werneburg, M., Zerbe, K., Steinmann, J., van der Meijden, B., Bernardini, F., Lederer, A., Dias, R. L. A., Misson, P. E., Henze, H., Zumbrunn, J., Gombert, F. O., Obrecht, D., Hunziker, P., Schauer, S., Ziegler, U., Käch, A., … Robinson, J. A. (2010). Peptidomimetic antibiotics target outer-membrane biogenesis in pseudomonas aeruginosa. Science, 327(5968), 1010–1013. https://doi.org/10.1126/science.1182749

Steyaert, J., & Kobilka, B. K. (2011). Nanobody stabilization of G protein-coupled receptor conformational states. Current Opinion in Structural Biology, 21(4), 567–572. https://doi.org/10.1016/j.sbi.2011.06.011

Storek, K. M., Chan, J., Vij, R., Chiang, N., Lin, Z., Bevers, J., Koth, C. M., Vernes, J. M., Meng, Y. G., Yin, J. P., Wallweber, H., Dalmas, O., Shriver, S., Tam, C., Schneider, K., Seshasayee, D., Nakamura, G., Smith, P. A., Payandeh, J., … Rutherford, S. T. (2019). Massive antibody discovery used to probe structure–function relationships of the essential outer membrane protein lptd. ELife, 8, 1–20. https://doi.org/10.7554/eLife.46258.001

Touw, D. S., Patel, D. R., & van den Berg, B. (2010). The crystal structure of OprG from Pseudomonas aeruginosa, a potential channel for transport of hydrophobic molecules across the outer membrane. PLoS ONE, 5(11). https://doi.org/10.1371/journal.pone.0015016

Uchański, T., Masiulis, S., Fischer, B., Kalichuk, V., Wohlkönig, A., Zögg, T., Remaut, H., Vranken, W., Aricescu, A. R., Pardon, E., & Steyaert, J. (2019). Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. BioRxiv, 812230. https://doi.org/10.1101/812230

Uchański, T., Pardon, E., & Steyaert, J. (2020). Nanobodies to study protein conformational states. Current Opinion in Structural Biology, 60, 117–123. https://doi.org/10.1016/j.sbi.2020.01.003

van den Berg, B., Black, P. N., Clemons, W. M., & Rapoport, T. A. (2004). Crystal Structure of the Long-Chain Fatty Acid Transporter FadL. Science, 304(5676), 1506 LP – 1509. https://doi.org/10.1126/science.1097524

Vetterli, S. U., Zerbe, K., Müller, M., Urfer, M., Mondal, M., Wang, S. Y., Moehle, K., Zerbe, O., Vitale, A., Pessi, G., Eberl, L., Wollscheid, B., & Robinson, J. A. (2018). Thanatin targets the intermembrane protein complex required for

lipopolysaccharide transport in Escherichia coli. Science Advances, 4(11). https://doi.org/10.1126/sciadv.aau2634

Villa, R., Martorana, A. M., Okuda, S., Gourlay, L. J., Nardini, M., Sperandeo, P., Bolognesi, M., Kahne, D., & Polissi, A. (2013). The Escherichia coli Lpt Transenvelope Protein Complex for Lipopolysaccharide Export Is Assembled via Conserved Structurally. 195(5), 1100–1108. https://doi.org/10.1128/JB.02057-12

Whitfield, C., & Trent, M. S. (2014). Biosynthesis and Export of Bacterial Lipopolysaccharides. Annual Review of Biochemistry, 83(1), 99–128. https://doi.org/10.1146/annurev-biochem-060713-035600

Zha, Z., Li, C., Li, W., Ye, Z., & Pan, J. (2016). LptD is a promising vaccine antigen and potential immunotherapeutic target for protection against Vibrio species infection. Scientific Reports, 6(August), 1–13. https://doi.org/10.1038/srep38577

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nature Methods, 14(4), 331–332. https://doi.org/10.1038/nmeth.4193

Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R., & Pearlman, D. (2014). Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. Proteins, 82(8), 1646–1655. https://doi.org/10.1002/prot.24551

Zimmermann, I., Egloff, P., Hutter, C. A. J., Arnold, F. M., Stohler, P., Bocquet, N., Hug, M. N., Huber, S., Siegrist, M., Hetemann, L., Gera, J., Gmür, S., Spies, P., Gygax, D., Geertsma, E. R., Dawson, R. J. P., & Seeger, M. A. (2018). Synthetic single domain antibodies for the conformational trapping of membrane proteins. ELife, 7, 1–32. https://doi.org/10.7554/eLife.34317

Zimmermann, I., Egloff, P., Hutter, C. A. J., Kuhn, B. T., Bräuer, P., Newstead, S., Dawson, R. J. P., Geertsma, E. R., & Seeger, M. A. (2020). Generation of synthetic nanobodies against delicate proteins. Nature Protocols, 15(5), 1707–1741. https://doi.org/10.1038/s41596-020-0304-x

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Cryst D 66:486–501. doi:10.1107/S0907444910007493

Geertsma ER. 2013. FX cloning: a versatile high-throughput cloning system for characterization of enzyme variants. Methods in Molecular Biology (Clifton, NJ) 978:133–148. doi:10.1007/978-1-62703-293-3_10

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci 27:14–25. doi:10.1002/pro.3235

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi:10.1002/jcc.20084

4.7 SUPPLEMENTARY DATA

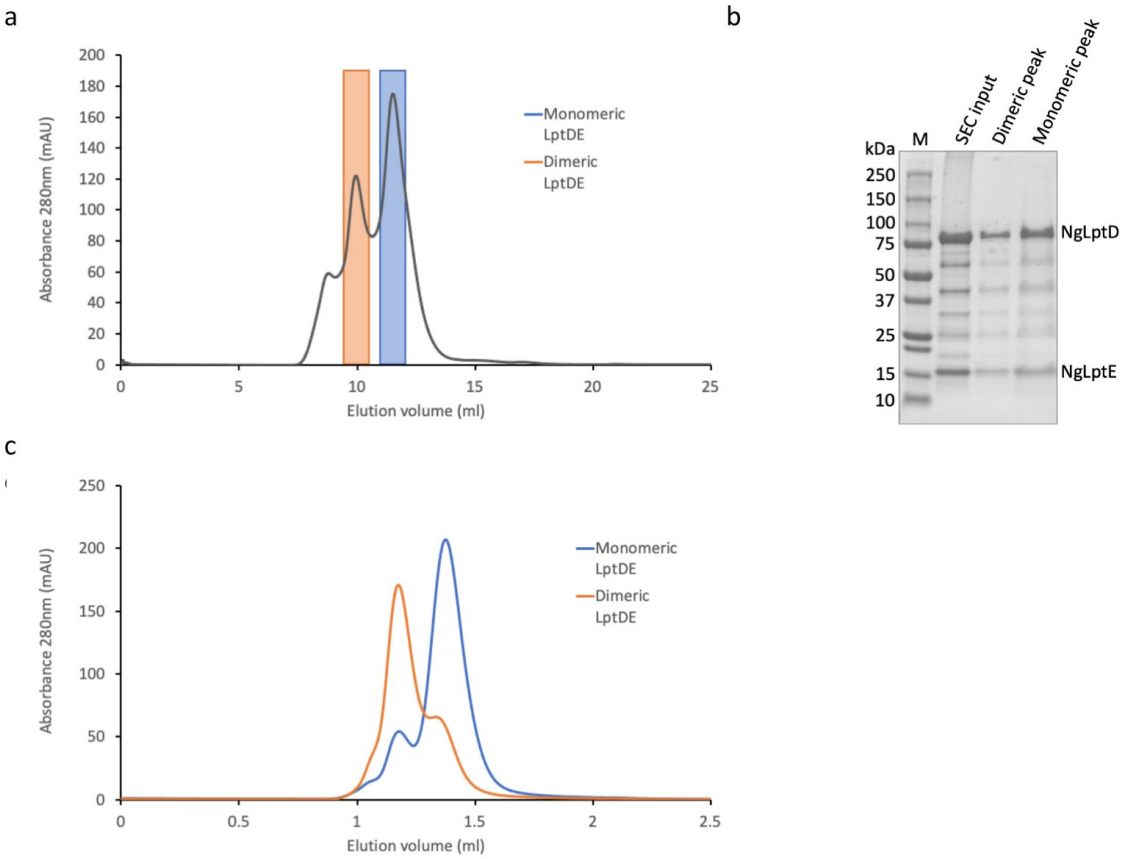


Figure 30 Isolation of Neisseria gonorrhoeae LptDE complex as monomer and stable dimer forms.

(a) Size exclusion chromatogram of purified NgLptDE. The curve represents the Abs280 as arbitrary units in function of the elution volume on a Superdex200 increase 10/300 GL. The dimeric peak is highlighted in orange whereas the monomeric is highlighted in blue. (b) Coomassie stained SDS-PAGE analysis of the eluted dimeric and monomeric peak fractions. The molecular weight of the standard marker is reported for comparison. LptD and LptE migration position is reported on the right of the gel. (c) Re-chromatography of the isolated monomeric and dimeric forms of NgLptDE. The peak fractions were re-analyzed by analytical size exclusion chromatography by injecting sample on Superdex200 increase 5/150. The curves represent the Abs280 as arbitrary units in function of the elution volume. The blue and the orange curve correspond respectively to the monomeric and the dimeric fractions.

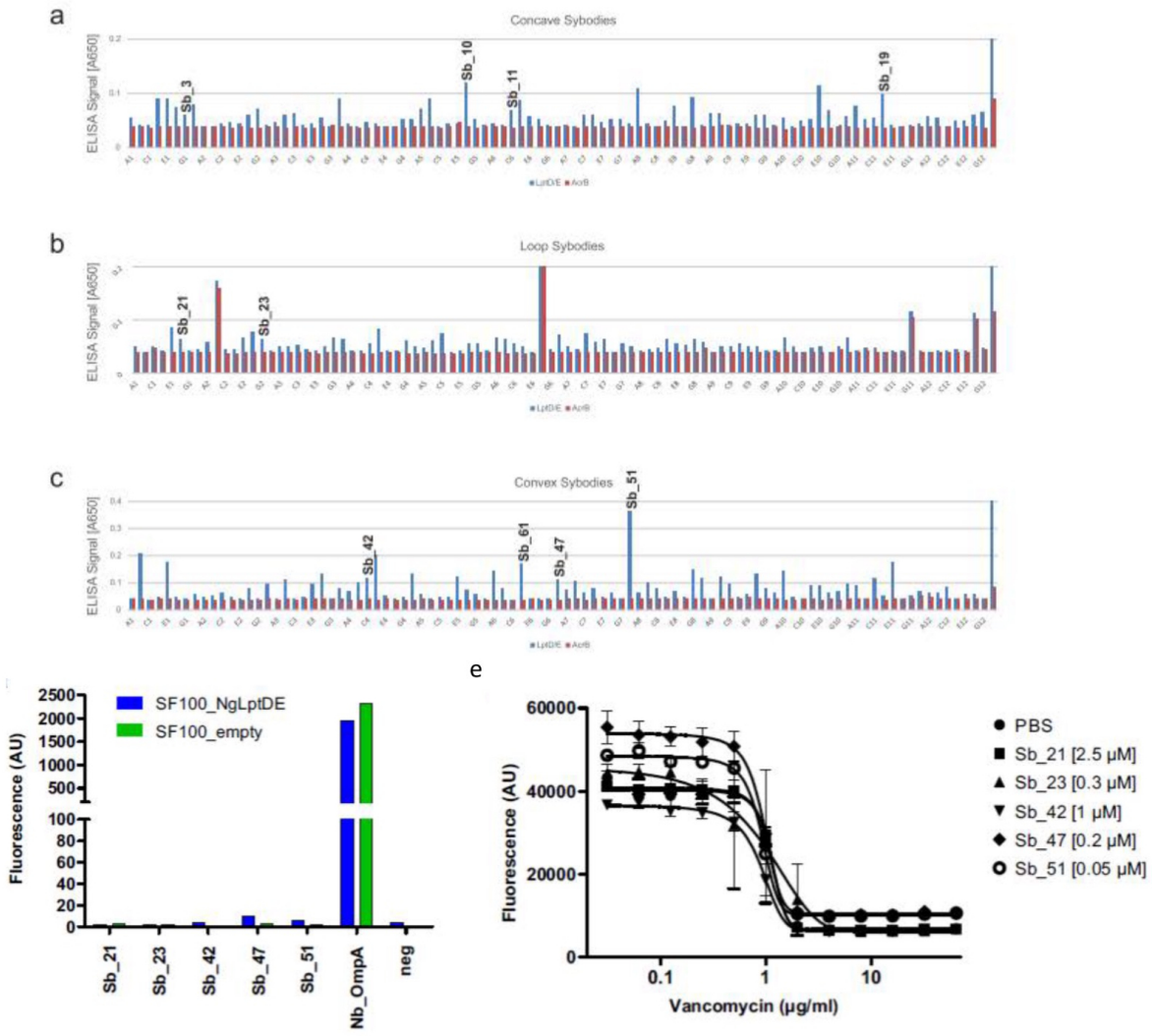


Figure 31 Generation of sybodies binding LptDE and their effect on bacterial survival

(a), (b), (c) Sybody identification by ELISA. ELISA signal intensities of all analyzed clones from concave (a), loop (b) and convex (c) sybody libraries. Well H12 represents the positive control of the ELISA. Sybodies whose affinities were determined by Grating Coupled Interferometry are labeled (Sb_3, Sb_10, etc.).

(d) Cellular binding assay. Alexa Fluor 647-labelled sybodies (Sb_21, Sb_23, Sb_42, Sb_47, Sb_51) were used to probe binding against E. coli SF100 cells overexpressing NgLptDE versus cells harbouring instead the empty plasmid. After washing, fluorescence of Alexa Fluor 647 was measured in a plate reader (see methods). A nanobody binding to E. coli OmpA (Nb_OmpA) was used as positive control (signal expected regardless of NgLptDE expression). The non-randomized loop sybody was used as negative control (neg).

(e) Test of antibiotic activity of sybodies against N. gonorrhoeae (Zopf) Trevisan (ATCC 700825) in synergy with the antibiotic vancomycin. N. gonorrhoeae was grown in Fastidious broth at increasing concentrations of vancomycin and sybodies added at concentrations > 10x KD (concentrations indicated in legend). Cell density was measured via resorufin fluorescence (see methods).

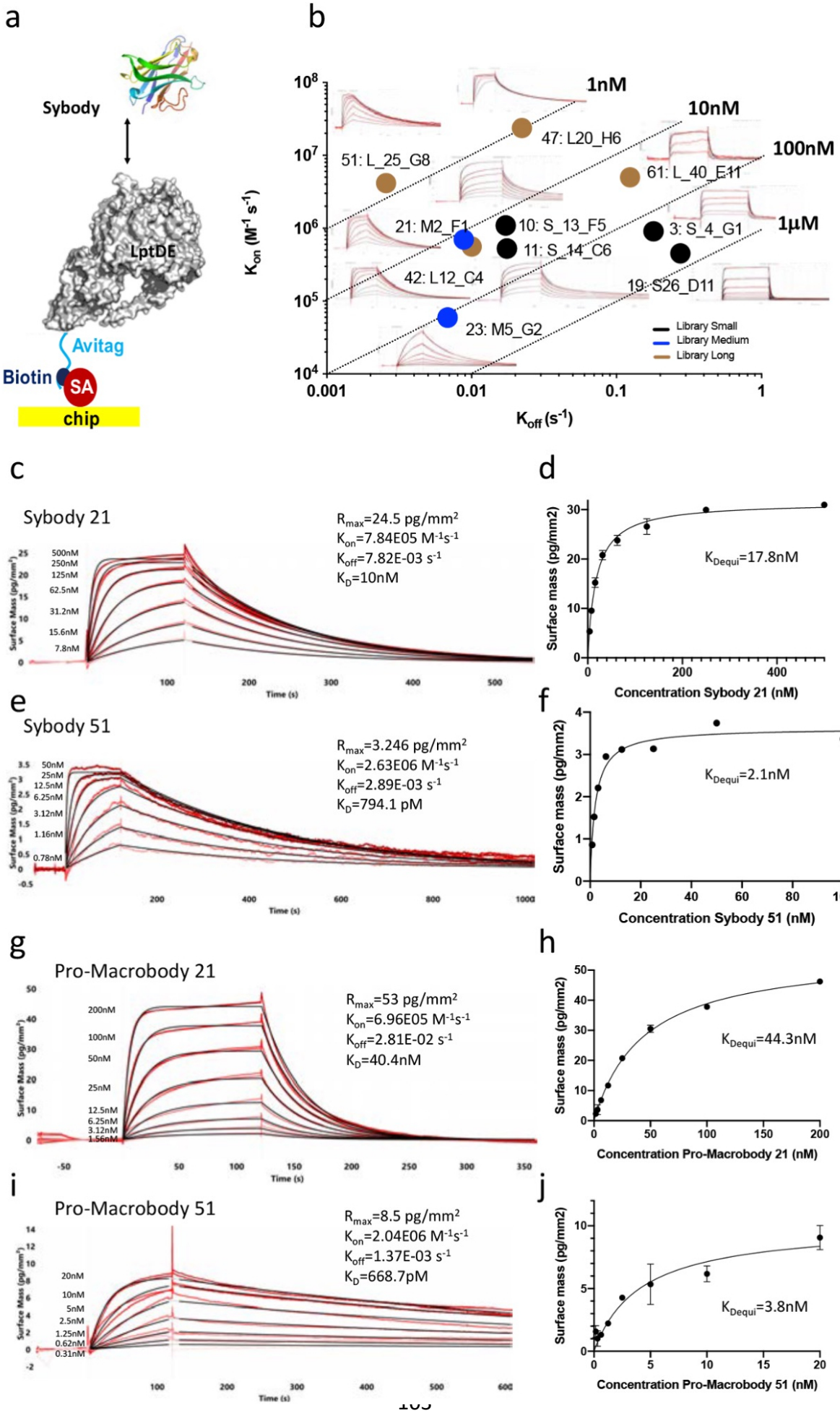

a
Sybody
LptDE
Avitag
Biotin
SA
chip
b
1nM
10nM
100nM
1µM
47: L20_H6
51: L_25_G8
61: L_40_E11
21: M2_F1
10: S_13_F5
11: S_14_C6
3: S_4_G1
42: L12_C4
19: S26_D11
23: M5_G2
Library Small
Library Medium
Library Long
K_{on} (M^{-1} s^{-1})
K_{off} (s^{-1})
c
Sybody 21
R_{max}=24.5 pg/mm²
K_{on}=7.84E05 $M^{-1}s^{-1}$
K_{off}=7.82E-03 s^{-1}
K_D=10nM
500nM
250nM
125nM
62.5nM
31.2nM
15.6nM
7.8nM
Surface Mass (pg/mm²)
Time (s)
d
K_{Dequi}=17.8nM
Surface mass (pg/mm2)
Concentration Sybody 21 (nM)
e
Sybody 51
R_{max}=3.246 pg/mm²
K_{on}=2.63E06 $M^{-1}s^{-1}$
K_{off}=2.89E-03 s^{-1}
K_D=794.1 pM
50nM
25nM
12.5nM
6.25nM
3.12nM
1.16nM
0.78nM
f
K_{Dequi}=2.1nM
Concentration Sybody 51 (nM)
g
Pro-Macrobody 21
R_{max}=53 pg/mm²
K_{on}=6.96E05 $M^{-1}s^{-1}$
K_{off}=2.81E-02 s^{-1}
K_D=40.4nM
200nM
100nM
50nM
25nM
12.5nM
6.25nM
3.12nM
1.56nM
h
K_{Dequi}=44.3nM
Concentration Pro-Macrobody 21 (nM)
i
Pro-Macrobody 51
R_{max}=8.5 pg/mm²
K_{on}=2.04E06 $M^{-1}s^{-1}$
K_{off}=1.37E-03 s^{-1}
K_D=668.7pM
20nM
10nM
5nM
2.5nM
1.25nM
0.62nM
0.31nM
j
K_{Dequi}=3.8nM
Concentration Pro-Macrobody 51 (nM)

Figure 32 Binding analysis of sybodies and PMbs by Grating Coupled Interferometry

(a) General set-up of the grating coupled interferometry assay. AVI tagged LpdTE complex is immobilized on a streptavidin (4PCP-S) chip. (b) Kinetic characterization of the top ten sybodies out of the initial screen. (c) Detailed kinetic characterization of sybody 21 with the corresponding equilibrium plot (d). (e) Detailed kinetic characterization of sybody 51 with the corresponding equilibrium plot (f). (g) Detailed kinetic characterization of PMb21 with the corresponding equilibrium plot (h). (i) Detailed kinetic characterization of PMb51 with the corresponding equilibrium plot (j).

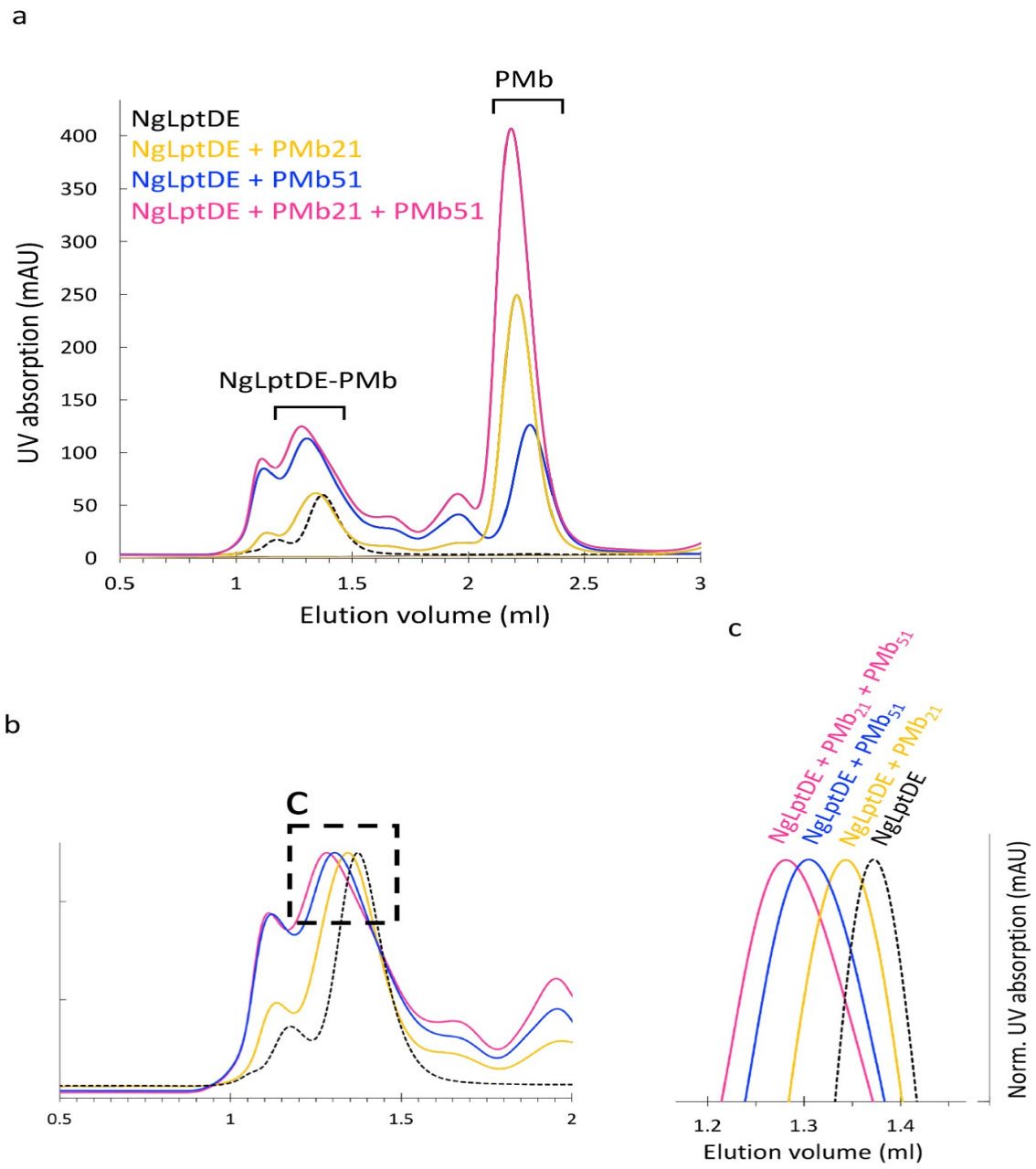


Figure 33 Complex formation between NgLptDE and ProMacrobodies.

a) Size exclusion chromatography was performed on monomeric LptDE, alone in complex with single PMb or with PMbs combination. The absorbance at 280nm in arbitrary unit was followed in function of the elution volume on a Superdex 200 increase 5/15 column. b) Normalized size exclusion chromatography chromatograms. c) Enlarged view of the peaks of interest shows the shift in elution volume caused by the different complex formations.

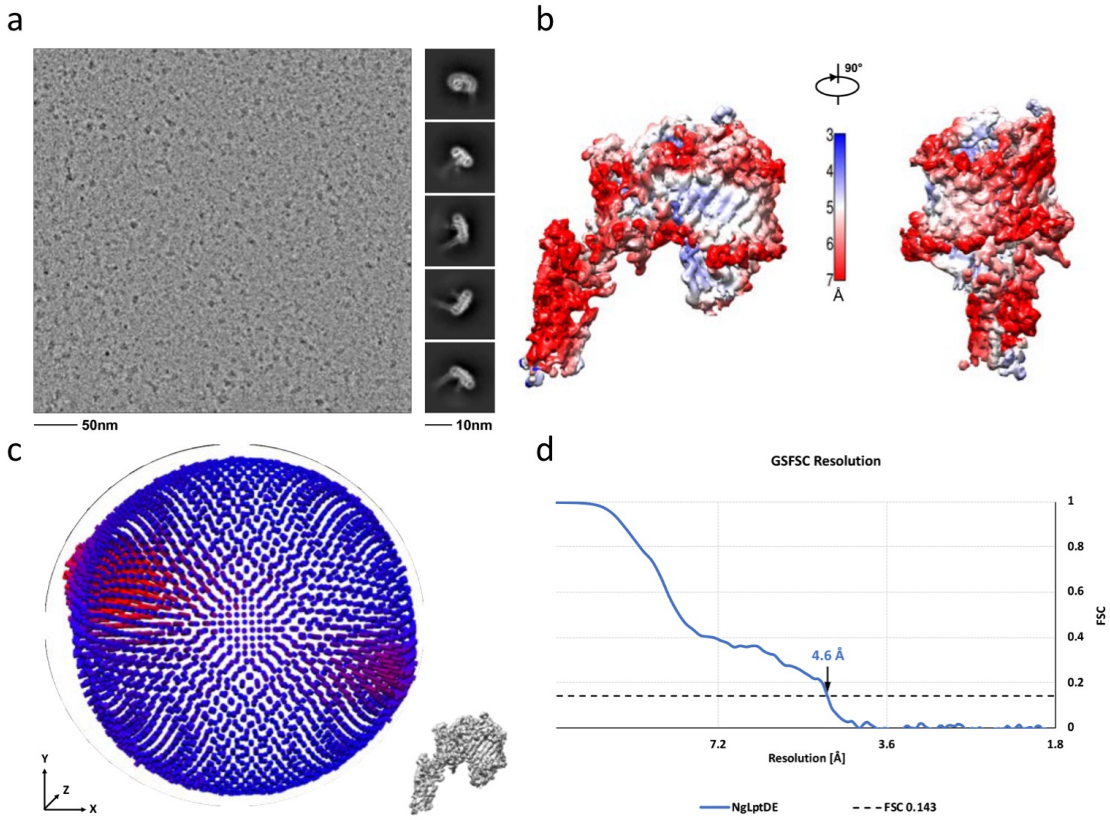


Figure 34 Cryo-EM analysis of the Apo NgLptDE complex.

(a) Typical electron micrograph and 2D classes. (b) Local resolution of the final cryo-EM map. (c) Euler angle distribution plot of the final 3D reconstruction. (e) Gold standard FSC curve of Apo NgLptDE.

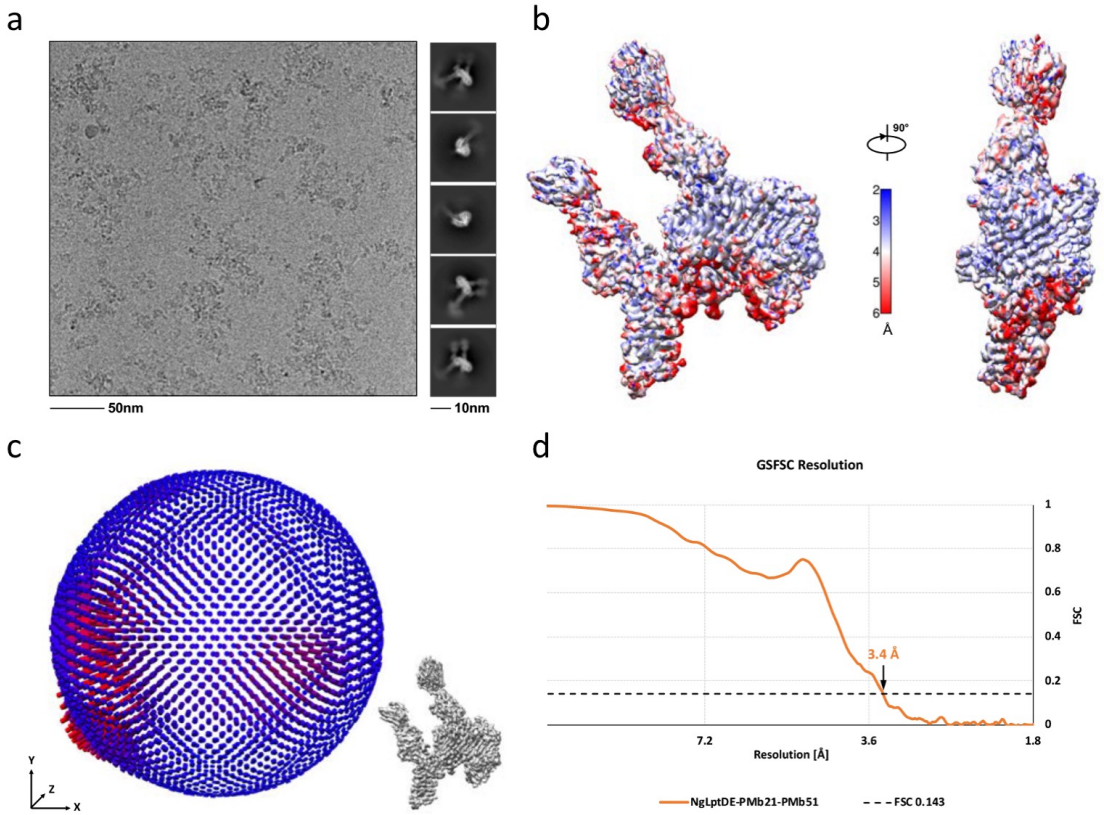


Figure 35 Cryo-EM analysis of the NgLptDE-PMb21-PMb51 complex. (a) Typical electron micrograph and 2D classes. (b) Local resolution of the final cryo-EM map. (c) Euler angle distribution plot of the final 3D reconstruction. (e) Gold standard FSC curve of NgLptDE-PMb21-PMb51.

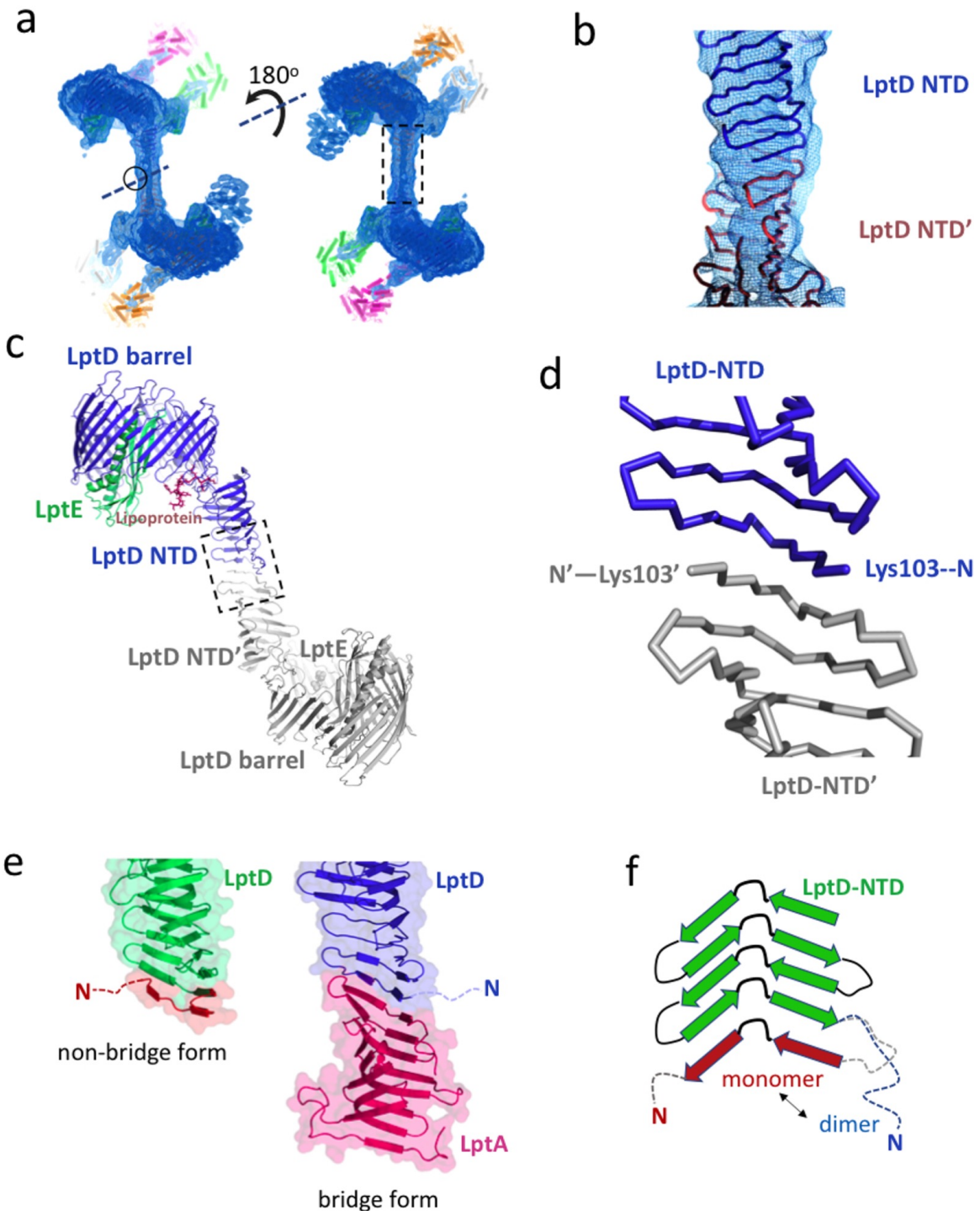


Figure 36 Observation of dimers of the N-terminal domain

(a) Cryo-EM map of dimer observed for the NgLptDE in complex with PMbs. Two models of the NgLptDE complex could be fitted in the cryo-EM map. (b) Fit of the model in the cryoEM density in the region delimited in (a). (c) Atomic model of the dimeric organization of LptDE. (d) Close up view of the dimerization interface delimited in (c). (e) Hypothetic model of the interaction the LptD N-terminal and LptA. (f) Cartoon representation of the transition between the monomeric and dimeric organization of the LptD N-terminal.

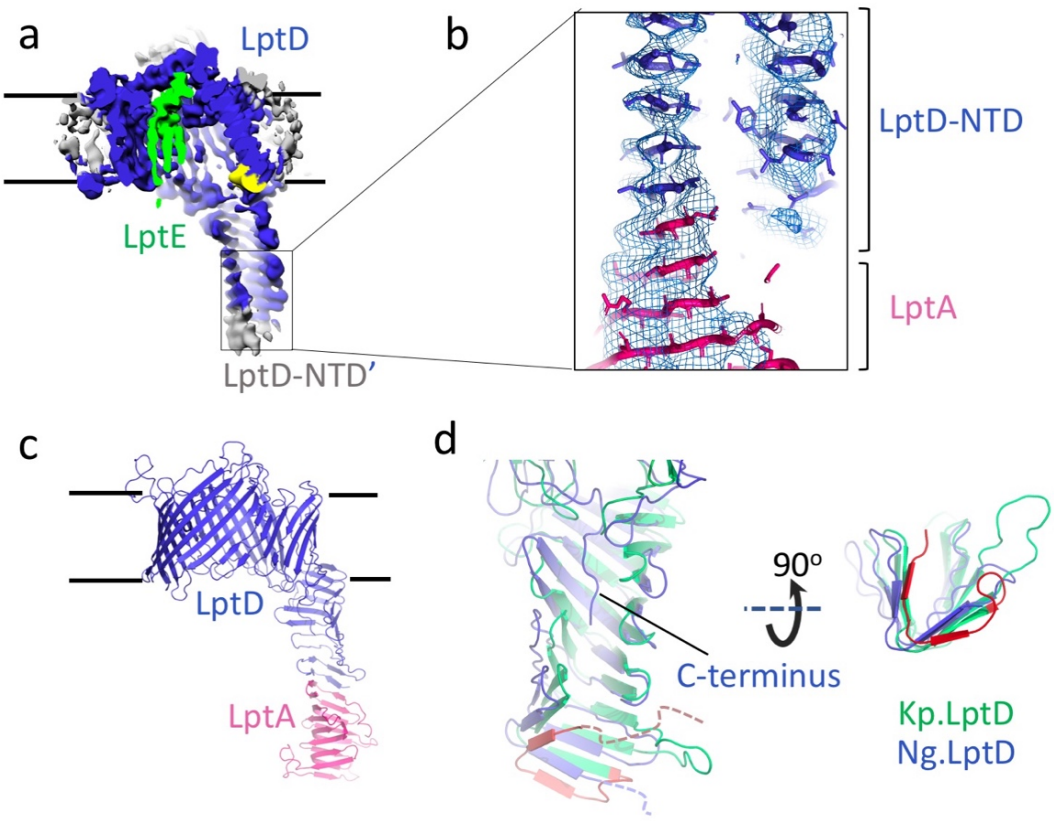


Figure 37 Modeling structure of LptA-lptD complex. (a) Surface representation of the obtained LptDE complex, containing a continuous jellyroll folded glove. (b) The close-up view of the LptD N-terminal jellyroll domain with the homo molecular modeling structure of LptA that fitted into the electron microscopy density. (c) A proposed complex organization for LptD-LptA. (d) Structural comparison of jellyroll region between Ng. LptD and Kp. LptD. The LptD beta-barrel domains have been superimposed. The Ng.LptD N-terminal domain (LPS transporting bridge) is blocked by its C-terminal loop.

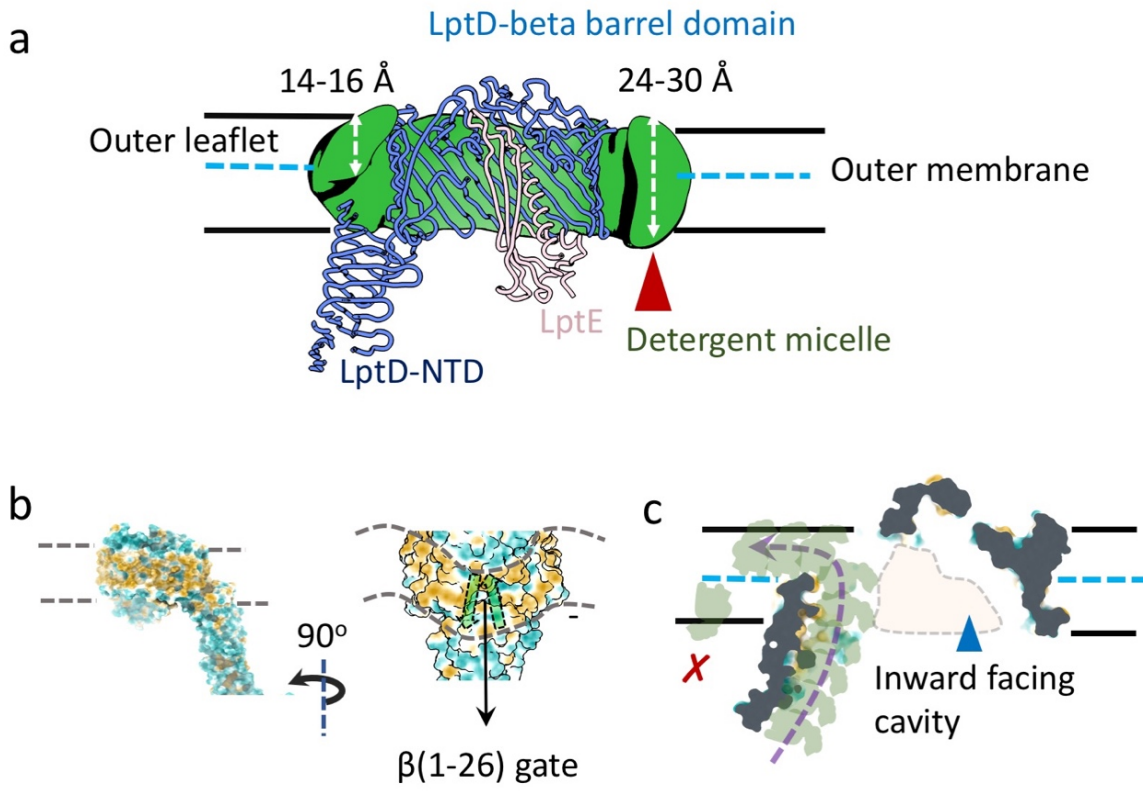


Figure 38 Observation of detergent micelle and the membrane associated surface. (a) The asymmetrical distribution of observed density for the bound detergent micelle. (b) The calculated electrostatic surface potentials suggest a direct membrane access for LptD-NTD (a part of LPS transporting bridge). The hydrophobic surface is colored yellow. (c) A proposed mechanism for LPS assembly and LPS orientation control.

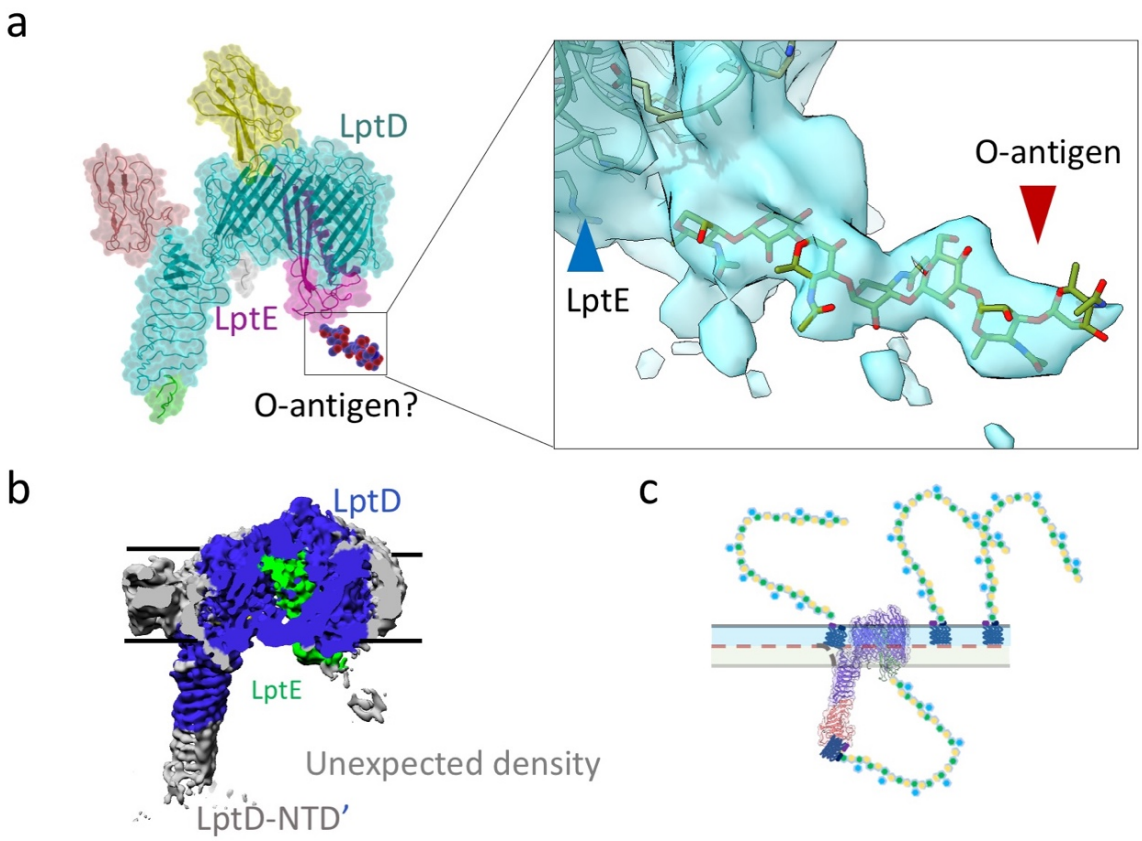


Figure 39 Hypothesis in the context of O-antigen recruitment. (a) Structure representation of LptDE with bound O-antigen. insert: A close-up view of E.coil LPS (O-antigen) docking into the unexpected density that obtained in our study. (b) A representative cryo-EM map with an unexpected additional density at the bottom of LptE, facing towards the periplasm. (c) A proposed model for (LPS) O-antigen recognition.

Table 5 Cryo-EM data collection, refinement and validation statistics.

Supplementary Table 1. Cryo-EM data collection and processing				
	NgLptDE-PMb21-PMb51		Apo NgLptDE	
			Monomeric fraction	Dimeric fraction
Microscope	FEI Titan Krios		FEI Titan Krios	FEI Titan Krios
Voltage (keV)	300		300	300
Camera	Gatan K2-Summit		Gatan K2Summit	Gatan K2Summit
Electron exposure (e–/Å^2/frame)	1.42		1.33	1.33
Energy filter slit width (eV)	20 (Gatan Quantum-LS (GIF))		20 (Gatan Quantum-LS (GIF))	20 (Gatan Quantum-LS (GIF))
Pixel size (Å)	0.82		0.64	0.64
Defocus range (μm)	(−0.8) - (−2.8)		(−0.8) - (−2.8)	(−0.8) - (−2.8)
Magnification (nominal)	60’975x (165kx)		78’125x (215kx)	78’125x (215kx)
Number of frames per movie	35		45	45
Number of good micrographs	12’060		3’964	1’578
Refinement procedure	As monomer	As dimer		As dimer

	Overall	Focused on C-terminal	Open conformation	Overall	Focused on dimerization interface	As monomer	
Initial particles (Before/After 2D cleaning)	815’057/490’743			534’661/175’411		1’395’292/196’182	112’358/32’079
Final particles	184’206	184’206 (softmask)	93’151	80’140	80’140 (softmask)	119’115	32’079
Map resolution (Å)	3.40	3.43	4.72	5.27	3.93	4.60	7.46
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	20 - 2.60	20 - 2.90	20 - 4.10	20 – 4.30	20 – 3.40	20 - 4.0	20 – 6.50

5 Chapter V: Structural characterization of SARS-CoV2 Spike-ACE2 complex

Running Title:

Structural investigation of ACE2 dependent disassembly of the trimeric SARS-CoV-2 Spike glycoprotein

Dongchun Ni[1], Kelvin Lau[2], Frank Lehmann[1], Andri Fränkl[1], David Hacker[2], Florence Pojer[2] and Henning Stahlberg[1]

1 Center for Cellular Imaging and NanoAnalytics, Biozentrum, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland

2 Protein Production and Structure Core Facility, SV PTECH PTPSP, Station 19, EPFL, CH-1015 Lausanne, Switzerland

* Corresponding author:
Henning Stahlberg
Center for Cellular Imaging and NanoAnalytics (C-CINA)
Biozentrum, University of Basel
WRO-1058, Mattenstrasse 26
CH-4058 Basel, Switzerland
Phone: +41 61 387 32 62
E-mail: henning.stahlberg@unibas.ch

Running title: Cryo-EM of SARS-CoV-2 S1 and hACE2

Keywords: SARS-CoV-2, COVID-19, Spike, S1/S2, cleavage site, cryo-EM, single particle, pre-fusion, post-fusion, virus entry, RBDs, receptor binding domains, soluble angiotensin-converting enzyme 2

This section introduces a side project of the structural characterization for SARS-CoV2 spike bound to ACE2. Motivated by the impact of the on-going global pandemic of COVID-19, we decided to make a contribution related to this novel coronavirus. The target of our study, the spike protein located on the surface of the virus, is responsible for host cell

invasion. We studied its interactions with ACE2 (a human single-pass transmembrane protein), the key molecule responsible for binding of the spike protein of host cell surface. In this study, we observed a phenomenon suggesting the effect of ACE2-dependent dissociation of the SARS-CoV2 Spike trimer and structural rearrangement of the Spike S1 domain to form a stable, monomeric complex with ACE2. Our research potentially provides a mechanism for the therapeutic role of recombinant soluble ACE2 for treatment of COVID-19.

My contribution: Designing the experiments; sample preparation and dataset acquisition; image processing; structures interpreting; manuscript writing.

This chapter is a submitted manuscript, it is also available on bioRxiv 2020.10.12.336016. (doi: https://doi.org/10.1101/2020.10.12.336016)

5.1 Abstract

The human membrane protein Angiotensin-converting enzyme 2 (hACE2) is known to be the main receptor for host cells invasion of the new coronavirus SARS-CoV-2. The viral surface glycoprotein Spike binds to ACE2, which triggers virus entry into cells. As of today, the role of ACE2 for virus fusion is not well understood. Blocking the transition of Spike from its prefusion to post-fusion state might be a strategy to prevent or treat COVID-19. Here we report a single particle cryo-electron microscopy analysis of SARS-CoV-2 trimeric Spike in presence of the human ACE2 ectodomain. The binding of purified hACE2 ectodomain to Spike induces the disassembly of Spike trimeric form and a structural rearrangement of the its S1 domain to form a stable, monomeric complex with ACE2. This observed ACE2 dependent dissociation of the Spike trimer suggests a mechanism for the therapeutic role of recombinant soluble ACE2 for treatment of COVID-19.

5.2 Introduction

Coronavirus is a family of single-stranded RNA viruses, many of which can infect animals and humans (MacLachlan and Dubovi, 2017; Monto, 1984). The symptoms of coronavirus-related diseases can be mild and mainly occur in respiratory tract. For example, roughly 15%-30% cases of the common cold are caused by human coronaviruses (Mesel-Lemoine et al., 2012). A coronavirus infection sometimes can develop serious illnesses, such as SARS (severe acute respiratory syndrome), MERS (Middle East respiratory syndrome) and also the current pandemic COVID-19 (coronavirus disease 2019) (Tang et al., 2020). SARS-CoV is a beta-coronavirus that caused a pandemic in 2002. SARS-CoV-2 is a novel coronavirus that is genetically similar to the previous SARS-CoV. SARS-CoV-2 causes the ongoing pandemic COVID-19 and has been spreading globally since the first quarter of this year (Ciotti et al., 2020). The symptoms of COVID-19 vary from person to person. In some cases, the illness is very serious, in particular for the elderly (Pascarella et al., 2020). As of today, no specific antiviral drugs were approved for use against COVID-19 and vaccine development is still at the phase of clinical testing.

Cryogenic electron microscopy (Cryo-EM) is a technique for structure determination of biomacromolecules, which has been particularly successful for studying high molecular-weight proteins. Cryo-EM does not require crystallization of the target protein. COVID-19 related protein structures have been widely investigated since the spread of SARS-CoV-2 in this year, using cryo-EM single particle analysis (SPA). The protein nicknamed Spike is with its 180kDa monomeric molecular weight the largest viral surface protein of SARS-CoV-2. It consists of two domains S1 and S2 that are connected by a short linker. Spike forms stable trimers on the virus surface that are attached to the virus membrane. This Spike trimer is the key molecule for host cells receptor binding and invasion of the host cells. The cryo-EM structure of the entire Spike homotrimer was determined recently, showing a mushroom sharped overall architecture (Walls et al., 2020; Wrapp et al., 2020a). As also for the SARS-CoV, the viral fusion bridge from SARS-CoV-2 to the host cell is formed by Spike and the ectodomain of the human Angiotensin-converting enzyme 2 (hACE2), which is the virus receptor on the host cell that triggers virus entry. In vitro studies have shown that the Spike receptor binding domains (RBDs) from SARS-CoV as well as SARS-CoV-2 can both bind to the ectodomain of hACE2 with comparable binding affinities in low nanomolar levels (Lan et al., 2020a). However, the new SARS-CoV-2 exhibits a more

potent capacity of host cells adhesion, as well as a larger virus-entry efficiency than other beta-coronaviruses (Shang et al., 2020).

The membrane-attached hACE2 is known to be the key molecule for the infection by several viruses, including SARS-CoV, Human coronavirus NL63 (HCoV-NL63) and SARS-CoV-2. The infection process primarily involves virus adhesion and fusion (Bao et al., 2020; Fehr and Perlman, 2015; Kuba et al., 2005; Sia et al., 2020). Interestingly, hACE2 may not only serve as a drug target to prevent SARS-Co-2 infection, but hACE2 itself may also be considered as a potential therapeutic drug candidate for the usage against COVID-19 or other beta-coronavirus related diseases. The clinical-grade soluble form of hACE2 has been reported to be a potential novel therapeutic approach for reducing the infection of SARS-CoV-2 (Monteil et al., 2020) by preventing the viral Spike from interacting with other hACE2 present on human cells. Recently, researchers have also characterized the entire architecture of the inactivated authentic virions from SARS-CoV-2 using cryo-electron tomography, observing that post-fusion S2 trimers are distributed on the surface of SARS-CoV-2 virions (Ke et al., 2020; Turonova et al., 2017). The exact role of hACE2 so far is not yet fully understood in terms of its interaction with full-length Spike protein.

In this study, we present a cryo-electron microscopy (cryo-EM) study of the SARS-CoV-2 Spike protein in complex with hACE2. Our analysis reveals a monomeric complex of Spike S1 domain with hACE2, requiring a large structural rearrangement in S1 compared to its isolated structure. Our data show that hACE2 binding induces a conformational change in Spike, leading to Spike trimer dissociation.

5.3 Results

5.3.1 Spike and hACE2 production and its complex assembly.

The prefusion Spike 2P ectodomain was expressed in ExpiCHO cells and affinity purified via its twin Strep-tag. SDS-PAGE analysis showed the presence of pure full-length Spike protein, consisting of both, the S1 and S2 domains at the expected molecular weight of 180 kDa for the Spike monomer. (Figure 40 a). The purified Spike sample in PBS buffer was imaged as negatively stained preparations by transmission electron microscopy (TEM). This revealed the expected trimeric shape, and 2D class averages of selected particles in negative

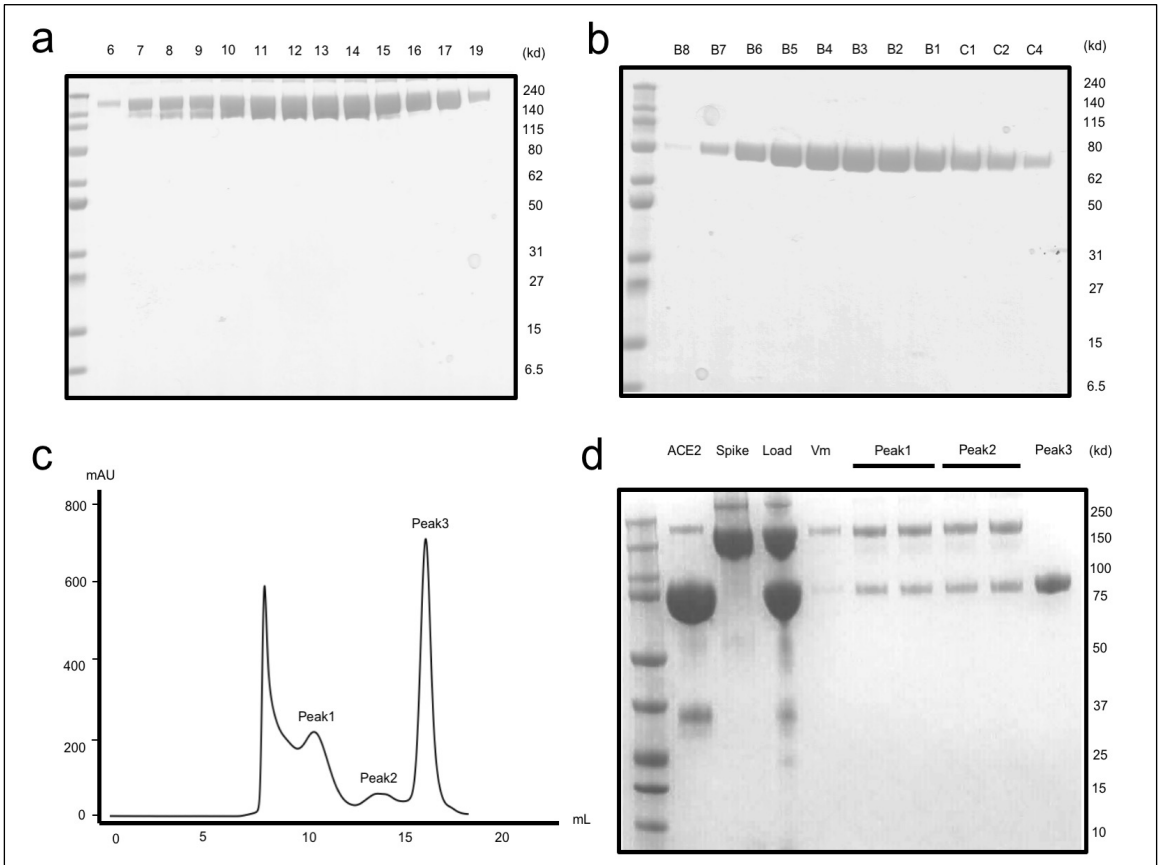


Figure 40 Protein production and purification.

(a) SDS-PAGE of the size exclusion chromatography (SEC) fractions of SARS-CoV-2 Spike (Superose 6 Increase 10/300GL). (b) SDS-PAGE of the anion exchange chromatography fractions from the purified sample of hACE2 ectodomain (HiTrap Q Sepharose HP). (c) A representative size exclusion chromatography curve of Spike-ACE2 complex assembly. 280nm UV absorption is shown. (d) The representative SDS-PAGE of the complex assembly. Lanes ACE2 to Load were after pre-concentration in 100 kDa centrifugal concentrators (Millipore).

stain TEM images showed the typical, mushroom-shaped particles (Figure 41a and c), in accordance with the expected structure of the SARS-CoV-2 Spike in the pre-fusion state.

Human ACE2 ectodomain was expressed in HEK293 cells and purified via a poly-histidine immobilized metal affinity chromatography (IMAC) with a Fastback Ni2+ column, followed by another anion exchange column (Figure 40 b). For details, see Methods.

Purified Spike protein was mixed with excess ACE2 (molar ratio Spike:ACE2 of 1:5) and incubated for 12 hours at 4º C. Samples were prepared by negative staining and imaged by TEM. Unexpectedly, the observed particle features were largely different from the typical Spike trimers in shape. A 2D analysis of 5'854 picked negatively stained particles revealed

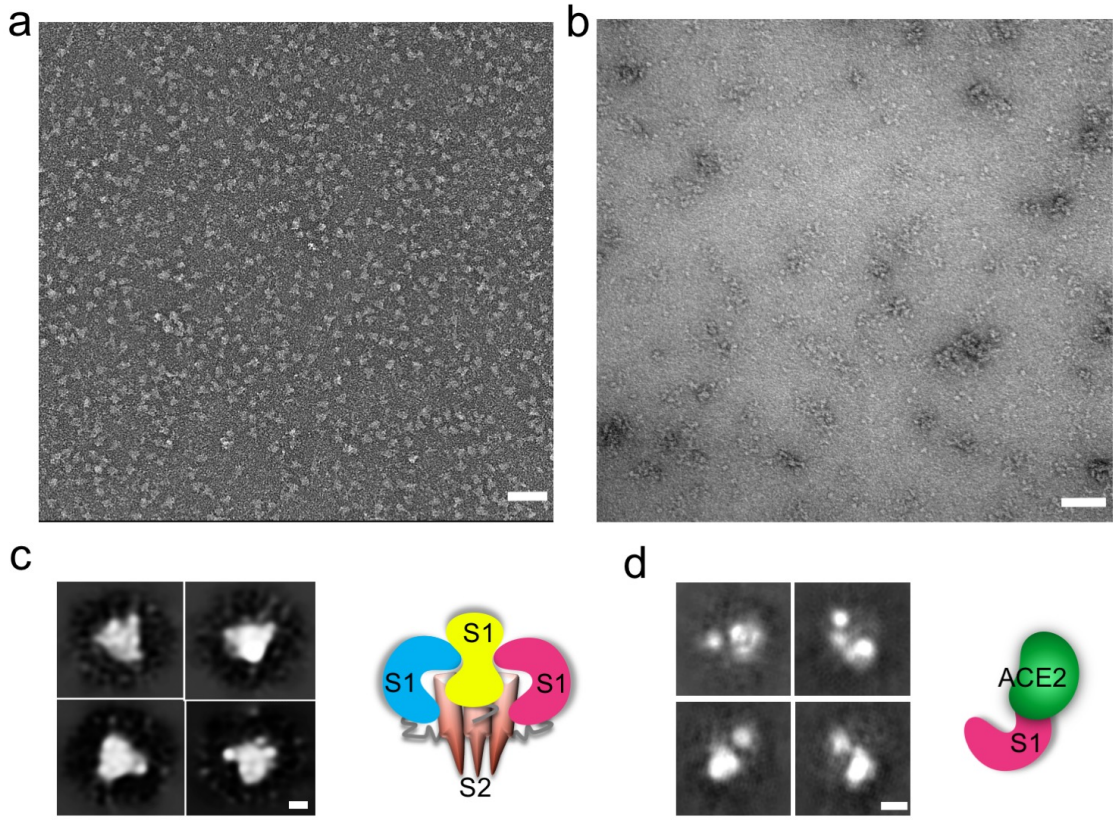


Figure 41 Dissociation of SARS-CoV2 Spike trimer.

(a) Negative stain TEM image of the Spike sample alone (non-incubated sample). (b) Negative stain TEM image of Spike protein incubated with human soluble ACE2 at a molar ratio of 1:5 for 12 hrs at 4ºC, and further purified by SEC. (c) Left: Representative class averages from particles picked from images as in a. Right: A cartoon interpretation of the structure. (d) Left: Representative class averages from particles picked from images as in b. Right: A cartoon interpretation of the structure. Scale bars in a and b are 50nm, in c and d are 3 nm.

in class averages that the majority of the particles were smaller in size and asymmetrical, compared to the non-incubated Spike trimeric samples (Figure 41b and d). This suggests that the prolonged incubation with ACE2 led to Spike trimer dissociation. A similar observation for SARS-CoV-ACE2 complex was recently reported by Song et al (Song et al., 2018). Due to particle heterogeneity, we decided to further purify the complex by size exclusion chromatography (SEC) and indeed the SEC profile showed three distinct peaks, called Peak1, Peak2 and Peak3 (Figure 40 c and d).

By analyzing the 3 peaks by SDS gel and negative stain EM, we could clearly differentiate non-structured aggregates of full-length Spike and hACE2 in Peak1, that did not allow further structural analysis (data not shown), to the excess of unbound hACE2 in Peak3 (Figure 40 c and d). The homogenous Peak2 that contained full-length Spike in complex with hACE2, was further analyzed by cryo-EM.

5.3.2 hACE2 binding can induce disassembly of Spike homotrimer

Peak2 (Spike:ACE2 at molar ratio 1:5 after overnight incubation at 4°C) was vitrified and frozen grids were loaded into a Thermo Fisher Scientific (TFS) Titan Krios cryo-EM instrument, operated at 300kV acceleration voltage, and equipped with a Gatan Quantum-LS energy filter equipped with K2 direct electron detector (Figure 45). 8'927 dose-fractionated images were recorded (Figure 46), from which ~1.7 million particles were extracted and subjected to image processing and 3D reconstruction. The final 3D reconstruction from 72'446 particles at 5.1Å overall resolution showed a density map corresponding to a single, monomeric Spike protein in complex with ACE2 (Figure 42 a). The map allowed docking with available structures for S1 and hACE2 that were taken from the previously reported structures (Spike PDB ID 6VYB and Spike RBD-ACE2 6M0J), revealing a structural rearrangement of the C-terminal domain (CTD) and N-terminal domain (NTD) of S1 compared to a monomer from that Spike structure in the RBDup conformation. The interaction between the S1 RBD and hACE2 is in agreement with several other reported structures of the RBD-ACE2 complex (PDB ID 6M0J, 6VW1 or 6LZG). No additional density for S2 or a fragment of S2 was detected in the reconstruction.

We tested shorter incubation time; by mixing Spike:hACE2 (molar ratio of 1:3) and let it incubate for 3 hours at 4ºC, instead as overnight. No further SEC purification was performed. Subsequently, cryo-EM grids of this sample were prepared and subjected to cryo-EM

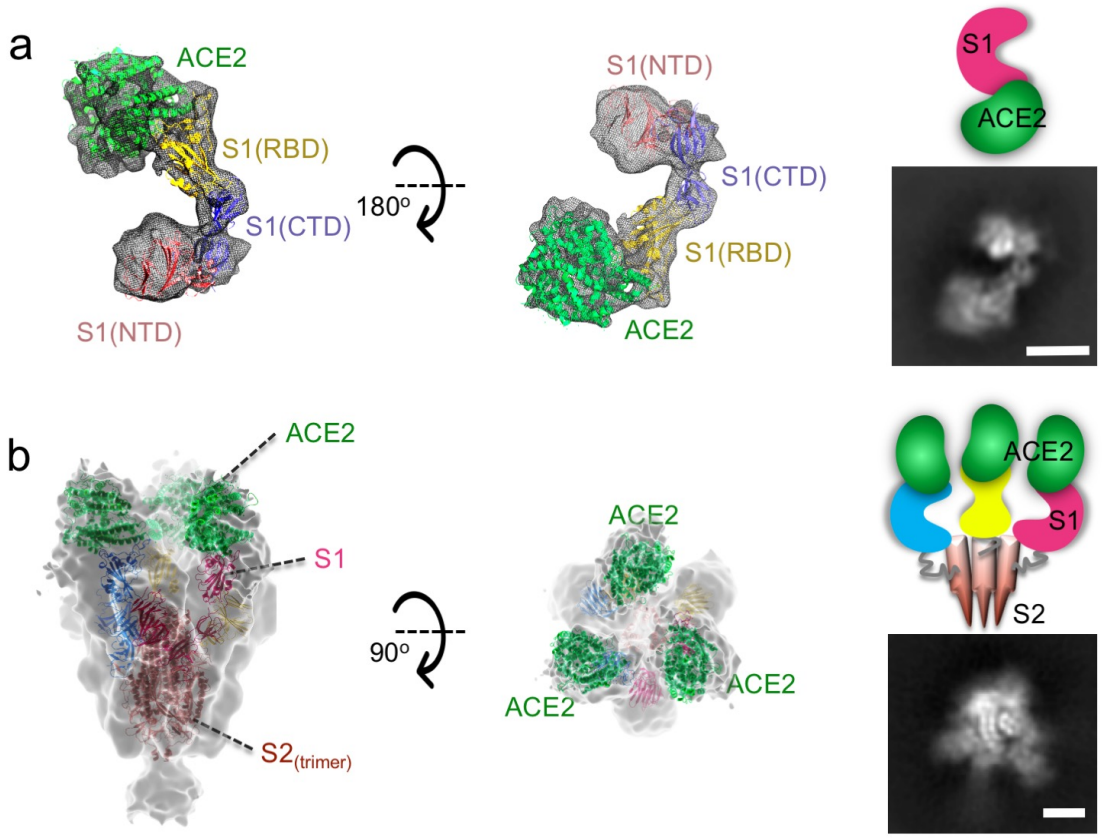


Figure 42 Cryo-EM maps of SARS-CoV-2 Spike-hACE2 complexes and fitted models.

a. The 3D reconstruction of SARS-CoV-2 Spike and human ACE2 (mixed at a molar ratio 1:5) incubated for 12hrs and further purified by SEC shows a structure corresponding to monomeric Spike S1 protein in complex with ACE2. No density for S2 is observed. The N- and C-terminal domains of S1 had to be re-arranged to fit into the map. Left: Side-view, center: 90º rotated view. The structure is colored as follows: ACE2 ectodomain green, Spike S1-RBD yellow, Spike S1-CTD slate blue, Spike S1-NTD salmon. Right: The bottom image shows a representative 2D class average of the S1-ACE2 complex. The upper cartoon is its interpretation. b. 3D reconstruction of Spike and hACE2 (ratio 1:3) incubated for 3 hours shows a trimeric map allowing docking of three ACE2 molecules and three S1 and three S2 molecules, all forming a trimeric complex. The map is low-pass filtered to 9Å resolution for better interpretation. The complex is arranged in a molar ratio of 3:3 (SpikeS1-S2:ACE2). The image on the left indicates the host cell membrane as a cartoon. Center: 90º rotated view. Right: The bottom image shows a representative 2D class average of Spike-ACE2 (3:3) complex, here shown from the top as in the central panel. The upper cartoon is its interpretation (here shown from the side). Scale bar in class averages: 2 nm.

analysis (Figure 47). From 7'045 recorded movies, 615'348 particles were extracted and subjected to classification and 3D analysis. This revealed a small sub-set of 47'901 particles corresponding to the prefusion Spike trimer, which allowed a 3D reconstruction at 4.2Å overall resolution (no symmetry was applied), while some regions of the 3D map showed lower resolution, presumably due to increased flexibility of these areas (Figure 48). To better visualize the ACE2 and further proceed the molecular docking, we computed a resolution-limited map at 9Å resolution (Figure 42 b). This map allowed docking of Spike S1 and S2 and of ACE2, which showed that the complex is composed of Spike and ACE2 in a molar ratio of 3:3 (Spike:ACE2). Three ACE2 molecules were observed to attach to the RBDs of Spike. All three RBDs were in the RBDup conformation and slightly shifted away from the central trimer axis (Figure 42 b). A similar arrangement was also recently observed by Kirchdoerfer et al. (Kirchdoerfer et al., 2018), see also (Zhou et al., 2020).

Structural comparison of Spike-hACE2 complexes .

A detailed analysis of both obtained structures shows significant structural rearrangements in different forms of Spike-ACE2 complex. The resolution of our reconstructed EM maps was not sufficient for building atomic models, but allowed clear placement of available atomic structures for Spike (PDB ID: 6VYB) and ACE2-RBD (PDB ID: 6M0J) (Lan et al., 2020b; Walls et al., 2020). Comparison of the docked and fitted monomeric S1-ACE2 model to that of the trimeric Spike-ACE2 (3:3) showed that a ~30º rotation of the C-terminal and N-terminal sub-domains of Spike S1 were required to bring the Spike S1 protein into the monomeric arrangement with ACE2. After such re-arrangement, the domains of ACE2 and the RBDs of the S1 protein are in good agreement with a reported crystal structure (REF Lan et al.) (Figure 43 b).

When comparing the docked model of the monomeric S1-ACE2 complex with that of the trimeric Spike-ACE2 complex (Figure 43 e), a considerable number of stearic clashes at the interface between Spike S1 (CTD) and its neighboring region from the S2 polypeptide chain was obvious. The docked monomeric S1-ACE2 complex is further structurally incompatible with the observed trimeric arrangement.

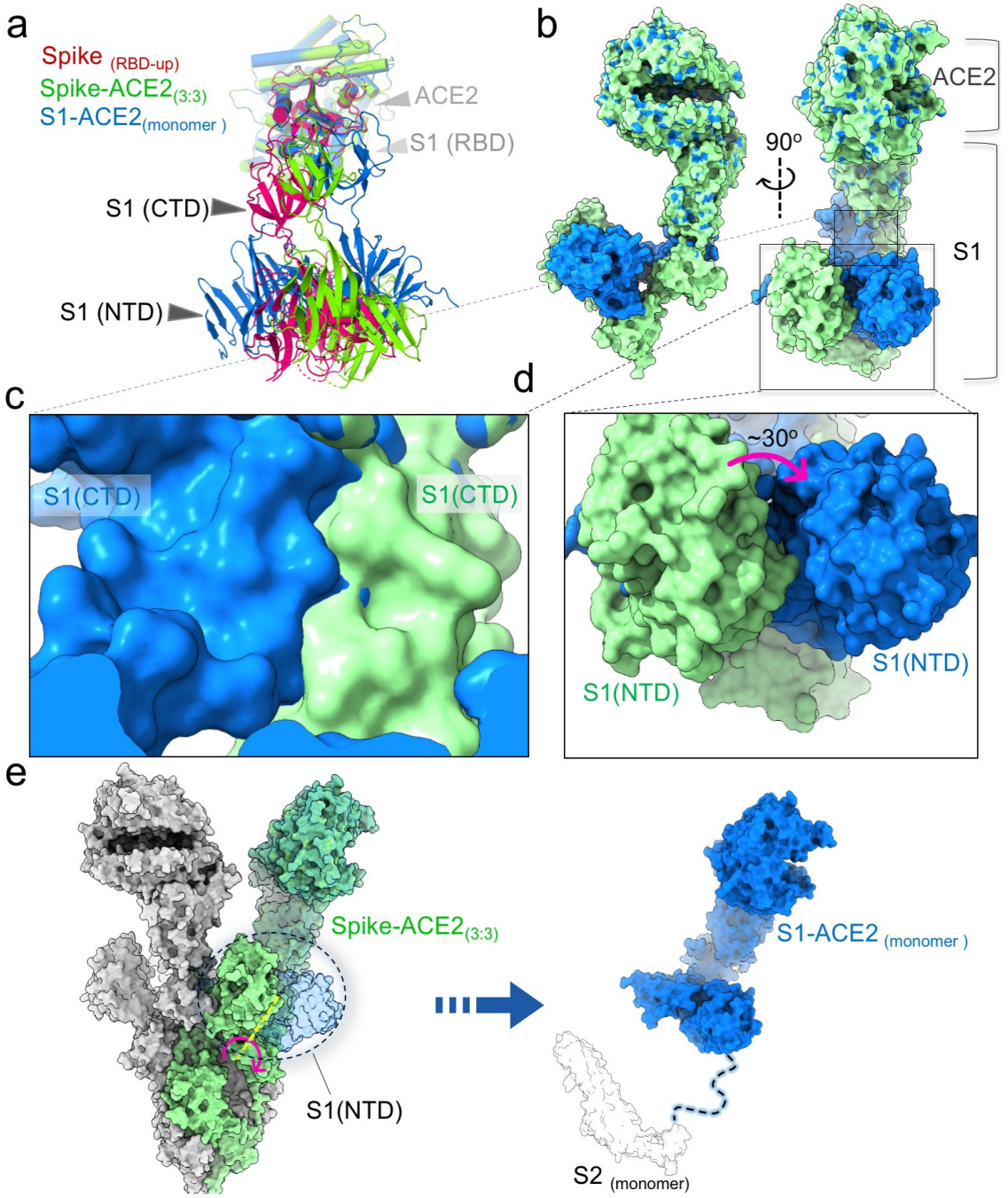


Figure 43 Conformational rearrangement of Spike-hACE2.

(a) Superposition of S1-ACE2 complex structures derived from different conformations. Structures are shown as a cartoon and colored as follows: monomeric Spike-ACE2 complex, blue; trimeric form of Spike-ACE2 complex, green; Spike one RBD-up (PDB ID: 6VYB), red. The RBDs have been superimposed. (b) Structural comparison of S1-ACE2 regions. The movement is presented as two colors: SARS-CoV2 trimeric Spike-ACE2 complex (green) and S1-ACE2 (blue) (c, d) The close-up views show the proposed local structural rearrangements. e. Superposition of S1-ACE2(1:1) with the trimeric form of Spike-ACE2 complex. The RBDs have been superimposed. The rearranged structure of Spike-ACE2 is no longer compatible with formation of trimers so that it dissociates. The right panel is a predicted structure of the dissociated Spike-ACE2 monomer.

5.4 Discussion

The stoichiometric ratio of the complex of Spike:ACE2 on the host cell upon virus entry is not well established. Nevertheless, one ACE2 molecule per Spike trimer is likely sufficient for binding and initializing virus fusion with the host cell (Song et al., 2018). Even though the structure of a post-fusion S2 trimer has recently been determined (Cai et al., 2020), it is not clear how membrane fusion during virus entry is coordinated upon release of the S1-ACE2 caps (Figure 44). We here report the cryo-EM structure of a stable monomeric S1 Spike-hACE2 (1:1) complex. Even though, size-wise it would have been detectable, our particle classification did not reveal any particle class corresponding to an isolated S1 fragment in addition to the observed S1 Spike-ACE2 (1:1) particles (Figure 45c, Figure 46). This is remarkable, since knowledge of the mechanism how the S1 fragment might be detached from the ACE2 receptors after virus entry would be relevant for understanding its mechanism of infection and pathogenicity. The absence of free S1 fragments suggests that the S1-ACE2 complex is rather stable, at least in our in vitro conditions.

Secondly, the S2 domain was not detected in the obtained structure of the Spike-ACE2 monomeric complex (1:1), even though the SDS-PAGE analysis showed that S2 was present as full length in the sample (Figure 40 d). The S2 domain is expected to be connected to the S1 domain via a short loop between S1 and S2, where a Furin protease cleavage site is expected(Belouzard et al., 2009; Haan et al., 2004; Hoffmann et al., 2020).. However, in the absence of stable trimers, the loop between S1 and S2 is likely very flexible, making the S2 domain undetectable by cryo-EM maps. Our cryo-EM analysis either failed to align the S2 domains due to their flexibility, or the S2 domain of the protein was denatured during sample preparation.

An early study presented a potential dose-dependent inhibition of SARS-CoV-2 infection by a recombinant soluble form of hACE2 (Monteil et al., 2020). The mechanism, how the soluble forms of ACE2 would be able to neutralize the virus, is not known. One possible mechanism could be a direct competition between the soluble ACE2 and the host cell hACE2 receptor, so that Spike proteins saturated with soluble ACE2 domains render them unable to interact with host cell ACE2. Here, however, we report that the soluble forms of hACE2 induce the opening and disassembly of the trimeric Spike structure to create the stable Spike S1-ACE2 complex (Figure 44). We propose a mechanism by which the formation of the Spike-ACE2 (3:3) complex induces a high structural flexibility in the Spike trimer, allowing a conformational re-arrangement of the S1 C- and N-terminal domains

when interacting with ACE2. In consequence, the new S1-ACE2 complex is incompatible with a trimeric arrangement, causing the dissociation of the trimeric complex (Figure 44). This hypothesis is supported by the recent manuscript deposited in bioRxiv.org, which describes a similar effect triggered by engineered DARPin molecules (Walser et al., 2020). Therefore, we suppose that the soluble forms of hACE2 may not only block the infection and replication of SARS-CoV-2, but also destroy the trimeric Spike adaptors that are

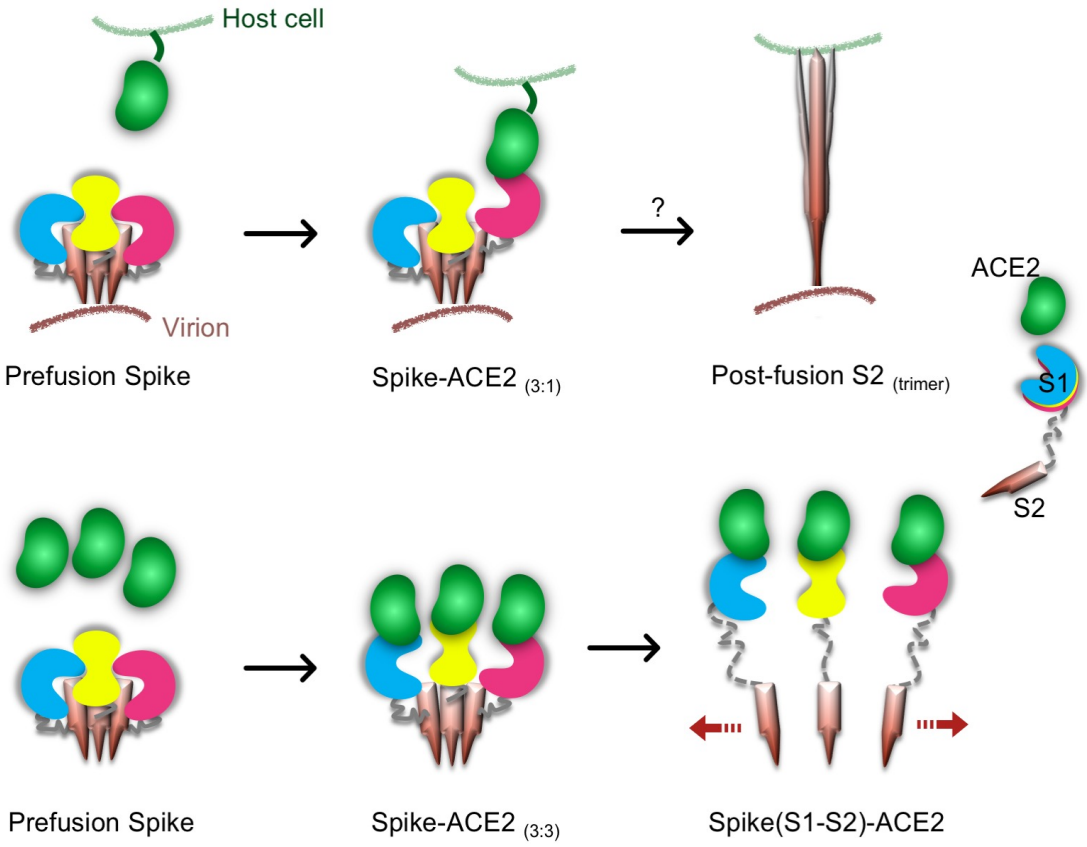


Figure 44 Proposed models for Spike-hACE2 complex formation and structural rearrangement. The upper row shows a possible pathway leading to a conformational change of the trimeric SARS-CoV2 Spike. In this model, one ACE2 molecule binds to one Spike S1 monomer and induces the conformational changes in the trimeric Spike. Subsequently, a post-fusion S2 trimer is formed. The lower row shows a novel proposed pathway leading to Spike trimer disassembly by ACE2. In presence of a high concentration of ACE2 molecules, a Spike-ACE2 (3:3) complex is formed. Structural clashes between the three Spike-ACE2 elements lead to their dissociation. This induces the formation of monomeric Spike-ACE2 complexes.

responsible for viral host membrane fusion. This mechanism suggests a novel therapeutic strategy for the treatment of COVID-19, by adding soluble hACE2 to dissociate the Spike trimer of approaching viruses.

5.5 Methods

5.5.1 Protein production and purification

Spike protein: The gene for the prefusion ectodomain of the SARS-CoV2 Spike 2-P protein (the prefusion stabilized construct (2P) that includes the putative furin cleavage site mutated. Plasmid was a generous gift from Prof. Jason McLellan, University of Texas, Austin;) (Wrapp et al., 2020) was transiently transfected into suspension-adapted ExpiCHO cells (Thermo Fisher) with PEI MAX (Polysciences) in ProCHO5 medium (Lonza). After 1 h, dimethyl sulfoxide (DMSO; AppliChem) was added to 2% (v/v). Incubation with agitation was performed at 31°C and 4.5% CO2 for 5 days. The transparent supernatant was passed over a Strep-Tactin column (IBA Lifesciences) and bound protein was eluted with PBS buffer + 2.5 mM desthiobiotin. The elute was dialyzed against PBS buffer. The average protein yield for Spike 2P was 15 mg/L culture.

hACE2 protein: The gene for the secreted expression of ACE2-Fc-His was a generous gift from Prof. Jason McLellan. The construct was transiently transfected into HEK293 cells (Thermo Fisher) with PEI MAX (Polysciences) in EX-CELL 293 serum-free medium (Sigma), supplemented with 3.75 mM valproic acid. Incubation with agitation was performed at 37°C and 4.5% CO2 for 8 days. The clear supernatant was purified as follows via MabSelect Protein A resin (Cytiva), cleavage of the Fc-His tag by overnight treatment with 3C protease, removal of the cleaved tag and 3C protease by NiNTA followed by an anion exchange column. The purified protein was finally dialyzed into PBS buffer. The average yield for cleaved ACE-2 was 13 mg/L culture.

5.5.2 Negative stain TEM imaging

For all protein samples, 3ul of sample solution was applied to glow-discharged carbon-coated Cu-grids and incubated for 1min. The free liquid was removed by blotting with filter paper and the grids were stained for 25s with 2% uranyl acetate solution. The negatively stained preparations were imaged with a Tecnai G2 Spirit TEM, operated at 120kV at a magnification of 135kx. Image were recorded with a Veleta CCD camera (EMSIS GmbH, Münster, Germany). The 2D analysis of the negative stain TEM images was performed with the program cisTEM1.0 (Grant et al., 2018).

5.5.3 Cryo-EM imaging

Two different preparations of cryo-EM grids were performed. Purified Spike and ACE2 proteins were mixed at the molar ratio of 1:5 and incubated for 12 hours at 4ºC. After that,

the sample was subjected to size exclusion chromatography (SEC) with a Superose 6 increase (10/300) column, and the fractions from Peak2 were pooled and concentrated in 100 kDa centrifugal concentrators (Millipore).

Alternatively, purified Spike and ACE2 were mixed at the molar ratio of 1:3 (Spike:ACE2) an incubated for 3 hours at 4ºc, without further purification via sec.

For both samples, the concentration was adjusted to 0.5mg/ml. Cryo-EM grids were prepared with a Vitrobot Mark IV (Thermo Fisher), using a temperature of 4°C and 100% humidity. 4 µL of sample was applied onto glow-discharged Quantifoil holey carbon grids (R2.0/2.0, 200 mesh, copper) and blotted for 2.0-3.0s. The grids flash-frozen in a liquid ethane, cooled by liquid nitrogen.

For both samples, dose-fractionated images (i.e., movies) were recorded with a Titan Krios (Thermo Fisher), operated at 300kV, and equipped with a Gatan Quantum-LS energy filter (20 eV zero-loss energy filtration) followed by a Gatan K2 Summit direct electron detector. The data collection statistics is presented in Table 6. Images were recorded in counting mode, at a magnification yielding a physical pixel size of 0.82Å at the sample level. Images were automatically recorded with the SerialEM program (Mastronarde, 2003) at a defocus range of -0.8 ~ -2.5um. All the movies were gain-normalized, aligned, dose weighted and averaged with the program of MotionCor2 (Zheng et al., 2017) within FOCUS (Biyani et al., 2017), which also was used to sort images and reject images of insufficient quality. The pre-processed micrographs were imported into cryoSPARC V2 (Punjani et al., 2017).

5.5.4 Image processing

Data processing of the monomeric Spike-ACE2 complex was conducted in CryoSPARC V2 (Punjani et al., 2017). The defocus and contrast transfer function (CTF) values were estimated on 8’927 micrographs and 7’927 good images were selected. 1’685’202 particles were auto-picked by blob picking, followed later by another round of template picking. Particle clearing was performed by two rounds of 2D classifications, resulting in a particles stack of 435’410 particles. 3D references were generated using ab-initio reconstruction (CryoSPARC V2) and followed by two rounds of 3D hetero-refinements. The final particle set contained 72’446 particles, leading to a 3D map at 5.1Å overall resolution, as estimated by Fourier Shell Correlation (FSC) using the 0.143 cutoff criterion.

For the trimeric Spike-ACE2 complex, image processing was performed similarly. The final reconstruction produced a 3D map at 4.2Å overall resolution.

5.5.5 Model interpretation

Protein models were generated from reported structures (Spike: PDB ID 6VYB; ACE2-RBD: PDB ID 6M0J) (Lan et al., 2020b; Walls et al., 2020). For the S1-ACE2 structure, the model was manually docked into the EM density with the program Chimera (Pettersen et al., 2004) and further refined using rigid-body fitting in COOT (Emsley et al., 2010). For the Spike-ACE2 trimer, the density corresponding to ACE2 was relatively weak, so that low-pass filtration to 9Å resolution was applied to the map before proceeding with the docking of ACE2 as described above.

5.5.6 Figure preparation.

Figures were created using the software PyMOL (PyMOL Molecular Graphics System, DeLano Scientific), Chimera and ChimeraX (Goddard et al., 2018).

5.5.7 Data availability

The EM maps for the SARS-CoV2 Spike-ACE2 trimer was deposited on EMDB with ID code EMD-11833 and the EMDB-ID for S1-ACE1 was EMD-11832.

Acknowledgements

We thank Prof. Jason McLellan from University of Texas, Austin for the Plasmids.We thank Lubomir Kovacik and Mohamed Chami for the assistance in cryo-electron microscopy. This work was supported by the Swiss National Science Foundation, grant NCCR TransCure.We thank Laurence Durrer and Soraya Quinche from PTPSP, EPFL for work in mammalian cells culture.

Author contributions

D.N. did imaging and image processing and data analysis; D.H, K.L. and F.P. produced and purified the protein samples of hACE2 and Spike; F.L. assembled the complex; A.F. performed negative stain TEM. H.S. supervised the project. All the authors edited the manuscript.

Competing financial interests

The authors declare no competing financial interests.

5.6 References:

Bao L, Deng W, Huang B, Gao H, Liu J, Ren L, Wei Q, Yu P, Xu Y, Qi F, Qu Y, Li F, Lv Q, Wang W, Xue J, Gong S, Liu M, Wang G, Wang S, Song Z, Zhao Linna, Liu P, Zhao Li, Ye F, Wang H, Zhou W, Zhu N, Zhen W, Yu H, Zhang X, Guo L, Chen L, Wang C, Wang Y, Wang X, Xiao Y, Sun Q, Liu H, Zhu F, Ma C, Yan L, Yang M, Han J, Xu W, Tan W, Peng X, Jin Q, Wu G, Qin C. 2020. The pathogenicity of SARS-CoV-2 in hACE2 transgenic mice. Nature 583:830–833. doi:10.1038/s41586-020-2312-y

Belouzard S, Chu VC, Whittaker GR. 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. PNAS 106:5871–5876. doi:10.1073/pnas.0809524106

Biyani N, Righetto RD, McLeod R, Caujolle-Bert D, Castano-Diez D, Goldie KN, Stahlberg H. 2017. Focus: The interface between data collection and data processing in cryo-EM. J Struct Biol 198:124–133. doi:10.1016/j.jsb.2017.03.007

Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, Rawson S, Rits-Volloch S, Chen B. 2020. Distinct conformational states of SARS-CoV-2 spike protein. Science. doi:10.1126/science.abd4251

Ciotti M, Ciccozzi M, Terrinoni A, Jiang W-C, Wang C-B, Bernardini S. 2020. The COVID-19 pandemic. Crit Rev Clin Lab Sci 57:365–388. doi:10.1080/10408363.2020.1783198

de Haan CAM, Stadler K, Godeke G-J, Bosch BJ, Rottier PJM. 2004. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. J Virol 78:6048–6054. doi:10.1128/JVI.78.11.6048-6054.2004

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Cryst D 66:486–501. doi:10.1107/S0907444910007493

Fehr AR, Perlman S. 2015. Coronaviruses: An Overview of Their Replication and Pathogenesis. Coronaviruses 1282:1–23. doi:10.1007/978-1-4939-2438-7_1

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci 27:14–25. doi:10.1002/pro.3235

Grant T, Rohou A, Grigorieff N. 2018. cisTEM, user-friendly software for single-particle image processing. eLife 7:e35383. doi:10.7554/eLife.35383

Hoffmann M, Kleine-Weber H, Pöhlmann S. 2020. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Molecular Cell 78:779-784.e5. doi:10.1016/j.molcel.2020.04.022

Ke Z, Oton J, Qu K, Cortese M, Zila V, McKeane L, Nakane T, Zivanov J, Neufeldt CJ, Cerikan B, Lu JM, Peukes J, Xiong X, Kräusslich H-G, Scheres SHW, Bartenschlager R, Briggs JAG. 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature 1–7. doi:10.1038/s41586-020-2665-2

Kirchdoerfer RN, Wang N, Pallesen J, Wrapp D, Turner HL, Cottrell CA, Corbett KS, Graham BS, McLellan JS, Ward AB. 2018. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. Scientific Reports 8:15701. doi:10.1038/s41598-018-34171-7

Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, Huan Y, Yang P, Zhang Y, Deng W, Bao L, Zhang B, Liu G, Wang Z, Chappell M, Liu Y, Zheng D, Leibbrandt A, Wada T, Slutsky AS, Liu D, Qin C, Jiang C, Penninger JM. 2005. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. Nature Medicine 11:875–879. doi:10.1038/nm1267

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215–220. doi:10.1038/s41586-020-2180-5

MacLachlan NJ, Dubovi EJ, editors. 2017. Chapter 24 - CoronaviridaeFenner's Veterinary Virology (Fifth Edition). Boston: Academic Press. pp. 435–461. doi:10.1016/B978-0-12-800946-8.00024-6

Mastronarde DN. 2003. SerialEM: A Program for Automated Tilt Series Acquisition on Tecnai Microscopes Using Prediction of Specimen Position. Microscopy and Microanalysis 9:1182–1183. doi:10.1017/S1431927603445911

Mesel-Lemoine M, Millet J, Vidalain P-O, Law H, Vabret A, Lorin V, Escriou N, Albert ML, Nal B, Tangy F. 2012. A Human Coronavirus Responsible for the Common Cold Massively Kills Dendritic Cells but Not Monocytes. J Virol 86:7577–7587. doi:10.1128/JVI.00269-12

Monteil V, Kwon H, Prado P, Hagelkrüys A, Wimmer RA, Stahl M, Leopoldi A, Garreta E, Hurtado Del Pozo C, Prosper F, Romero JP, Wirnsberger G, Zhang H, Slutsky AS, Conder R, Montserrat N, Mirazimi A, Penninger JM. 2020. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. Cell 181:905-913.e7. doi:10.1016/j.cell.2020.04.004

Monto AS. 1984. Coronaviruses. Viral Infections of Humans: Epidemiology and Control.

Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, Scarlata S, Agrò FE. 2020. COVID-19 diagnosis and management: a comprehensive review. J Intern Med 288:192–206. doi:10.1111/joim.13091

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi:10.1002/jcc.20084

Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14:290–296. doi:10.1038/nmeth.4169

Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. 2020. Structural basis of receptor recognition by SARS-CoV-2. Nature 581:221–224. doi:10.1038/s41586-020-2179-y

Sia SF, Yan L-M, Chin AWH, Fung K, Choy K-T, Wong AYL, Kaewpreedee P, Perera RAPM, Poon LLM, Nicholls JM, Peiris M, Yen H-L. 2020. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. Nature 583:834–838. doi:10.1038/s41586-020-2342-5

Song W, Gui M, Wang X, Xiang Y. 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLOS Pathogens 14:e1007236. doi:10.1371/journal.ppat.1007236

Tang D, Comish P, Kang R. 2020. The hallmarks of COVID-19 disease. PLoS Pathog 16:e1008536. doi:10.1371/journal.ppat.1008536

Turoňová B, Sikora M, Schürmann C, Hagen WJH, Welsch S, Blanc FEC, Bülow S von, Gecht M, Bagola K, Hörner C, Zandbergen G van, Landry J, Azevedo NTD de, Mosalaganti S, Schwarz A, Covino R, Mühlebach MD, Hummer G, Locker JK, Beck M. 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science. doi:10.1126/science.abd5223

Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181:281-292.e6. doi:10.1016/j.cell.2020.02.058

Walser M, Rothenberger S, Hurdiss DL, Schlegel A, Calabro V, Fontaine S, Villemagne D, Paladino M, Hospodarsch T, Neculcea A, Cornelius A, Schildknecht P, Matzner M, Hänggi M, Franchini M, Kaufmann Y, Schlegel I, Iss C, Loser T, Mangold S, Herzog C, Schiegg D, Reichen C, Radom F, Bosshart A, Lehmann A, Haeuptle MA, Zürcher A, Vagt T, Sigrist G, Straumann M, Proba K, Veitonmäki N, Dawson KM, Zitt C, Mayor J, Ryter S, Lyoo H, Wang C, Li W, Drulyte I, Binz HK, Waal L de, Stittelaar KJ, Lewis S, Steiner D, Kuppeveld FJM van, Engler O, Bosch B-J, Stumpp MT, Amstutz P. 2020. Highly potent anti-SARS-CoV-2 multi-DARPin therapeutic candidates. bioRxiv 2020.08.25.256339. doi:10.1101/2020.08.25.256339

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1260–1263. doi:10.1126/science.abb2507

Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14:331–332. doi:10.1038/nmeth.4193

Zhou T, Tsybovsky Y, Olia AS, Gorman J, Rapp M, Cerutti G, Chuang G-Y, Katsamba PS, Nazzari A, Sampson JM, Schön A, Wang P, Bimela J, Shi W, Teng I-T, Zhang B, Boyington JC, Sastry M, Stephens T, Stuckey J, Wang S, Friesner RA, Ho DD, Mascola JR, Shapiro L, Kwong PD. 2020. Cryo-EM Structures Delineate a pH-Dependent Switch that Mediates Endosomal Positioning of SARS-CoV-2 Spike Receptor-Binding Domains. bioRxiv 2020.07.04.187989. doi:10.1101/2020.07.04.187989

Table 6 Cryo-EM data collection, refinement and validation statistics.

	Spike monomer	Spike trimer
Cryo-electron microscopy data collection and processing		
Microscope	FEI Titan Krios	FEI Titan Krios
Voltage (keV)	300	300
Camera	Gatan K2-Summit	Gatan K2-Summit
Electron exposure (e–/Å^2/frame)	1.5	1.0
Energy filter slit width (eV)	20 (Gatan Quantum-LS)	20 (Gatan Quantum-LS)
Pixel size (Å)	0.82	0.82
Defocus range (μm)	(−0.8) - (−2.5)	(−0.8) - (−2.5)
Magnification (nominal)	60’975x (165kx)	60’975x (165kx)
Number of frames per movie	40	40
Number of good micrographs	8’927	7’045
Initial particles	435’410	201’374
Final particles	72’446	47’901
Symmetry imposed	C1	C1
Map resolution (Å)	5.1	4.2
FSC threshold	0.143	0.143
Map local resolution range (Å)	20 -4.80	20 -3.60
EMDB ID	EMD-11832	EMD-11833

5.7 SUPPLEMENTARY FIGURES

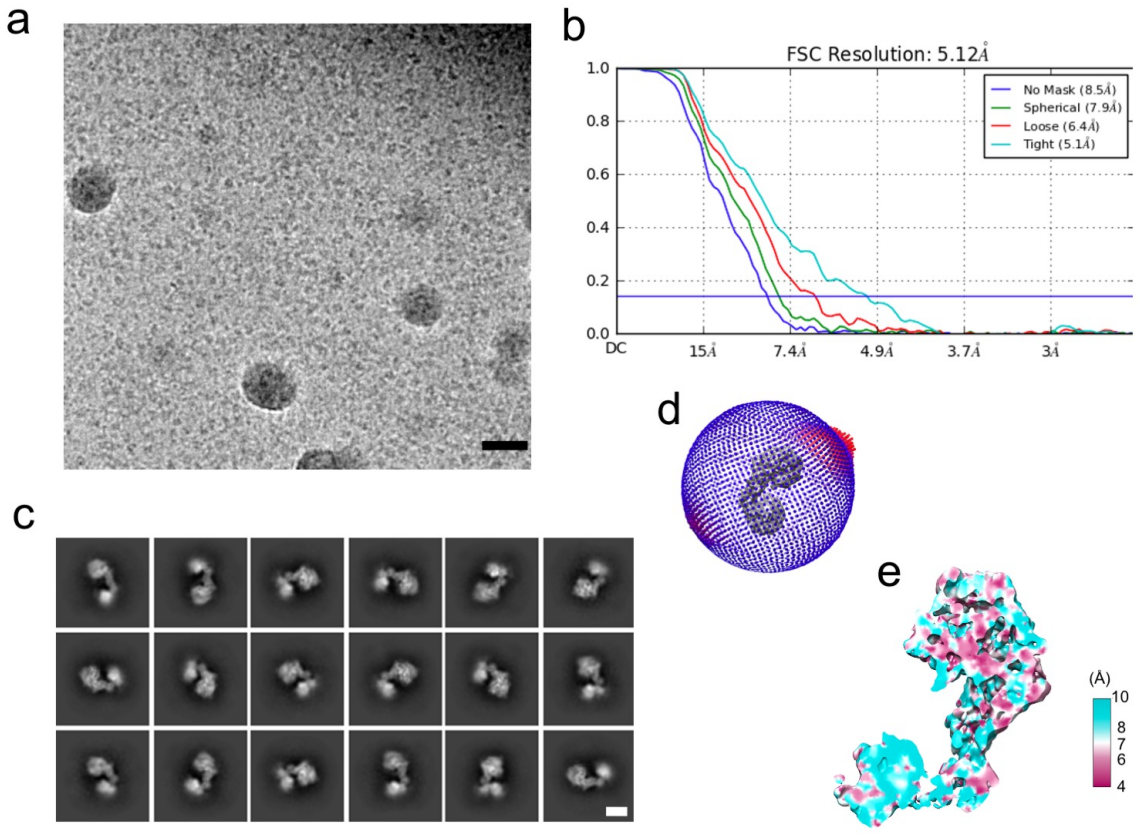


Figure 45 Data quality of Spike-ACE2 monomeric complex (S1-ACE2) sample.

(a) A representative micrograph. (b) Overall Resolution estimation (FSC, 0.143). (c) Representative 2D average classes. (d) Direction distributions. (e) Local Resolution estimation (MonoRes). Scale bars in a and c are 50nm and 3 nm.

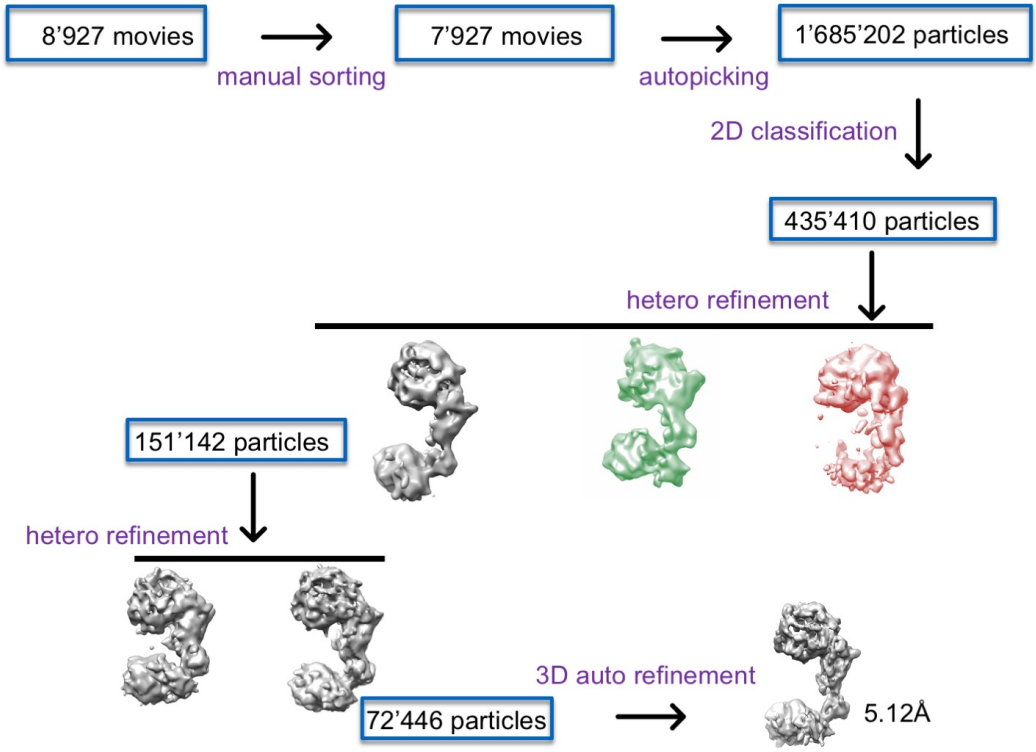


Figure 46 Processing workflow for the Spike-ACE2 monomeric complex (S1-ACE2).

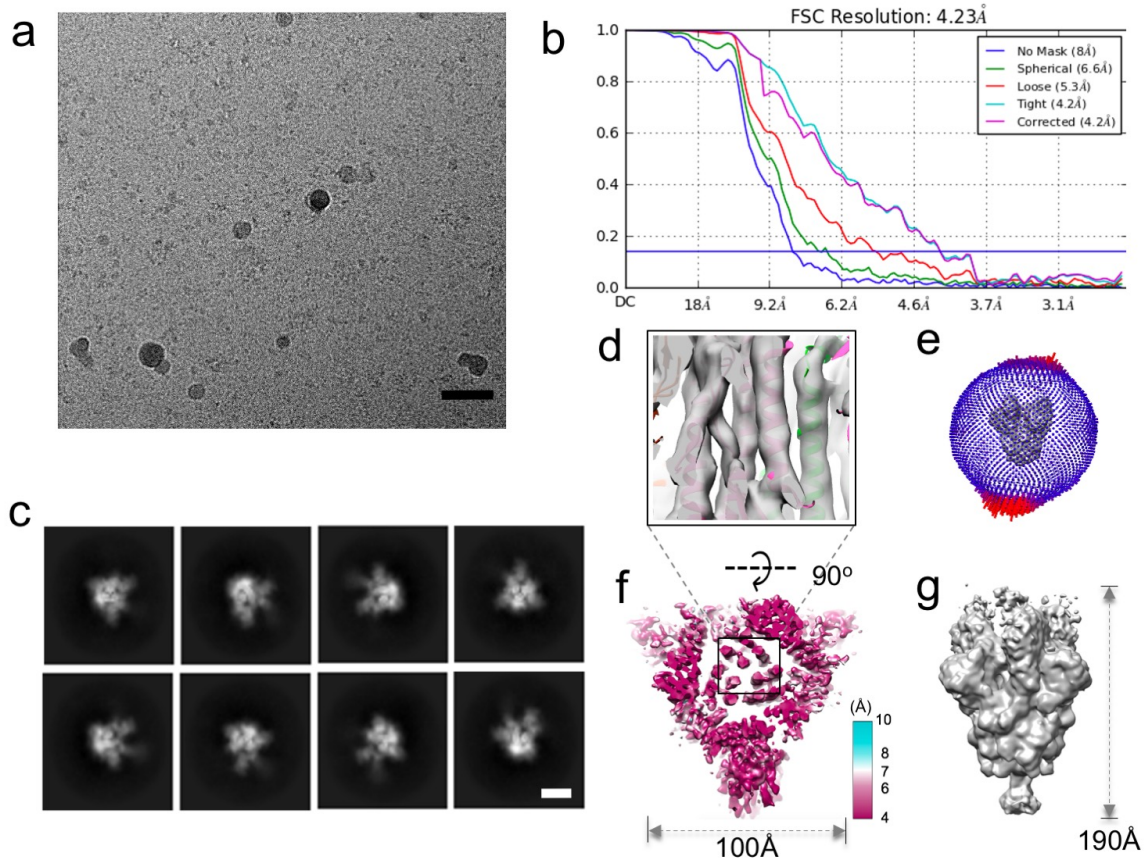


Figure 47 Data quality of the sample of Spike-ACE2 trimeric complex.

(a) A representative micrograph. (b) Overall Resolution estimation (FSC, 0.143). (c) Representative 2D average classes. (d) Model fitted into the S2 trimeric core (The map was low-pass filtered). (e) Distribution of particle orientations. (f) Local resolution level at the best resolved regions of the trimeric form of Spike-ACE2 complex (bottom view, MonoRes). (g) The low-pass filtered EM map at 9Å resolution (For model generation). Scale bars in a and c are 50nm and 3 nm.

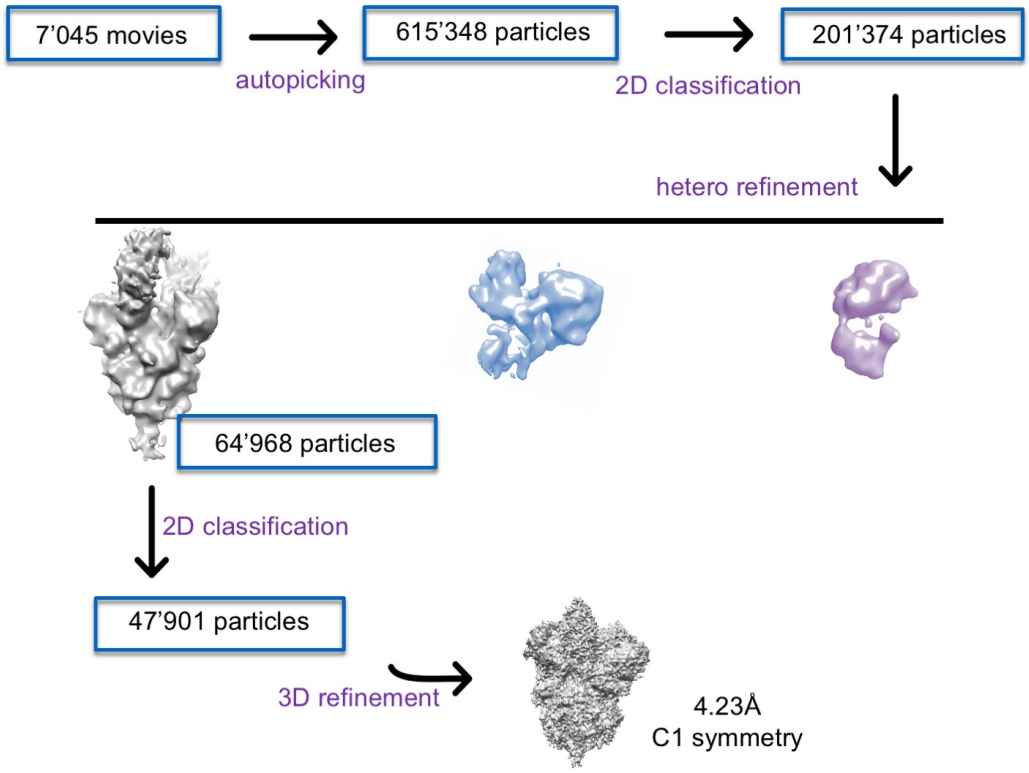


Figure 48 Processing workflows for the trimeric Spike-ACE2 complex.

6 Chapter VI: Conclusion and Outlook:

Cryo-EM has undergone a resolution revolution in 2013. The amazing progress of cryo-EM allowed it to become a major player in the field of drug discovery. For membrane proteins, it is still quite challenging to determine structures with clearly defined side-chains using cryo-EM SPA. Several issues can be limiting during both sample preparation and data processing, such as protein expression and isolation, protein stability, protein denaturation, preferred orientation and protein flexibility. A variety of methods have been developed in order to improve the resolution and prevent such problems. For example, protein production is one of the main issues for the membrane protein cryo-EM studies. For many membrane protein samples, to purify a sufficiently concentrated sample for cryo-EM grid preparation is not trivial. This is especially true for human membrane proteins as well as for protein complexes that have several subunits. In addition, protein folding and membrane insertion is often incorrect in artificial expression systems, where the complexity of multiple steps of protein synthesis and protein quality control is missing.

To give an idea how reliable and efficient this approach is, I have presented in this thesis three studies regarding structure-based drug discovery for membrane protein targets. These projects have provided the structural basis for drug development.

What have we learned from the studies presented in this thesis?

The results from Chapter III demonstrate that it is now feasible to visualize interactions of membrane proteins with small molecule ligands using cryo-EM. Obtaining good ligand densities is still problematic for some special situations like ABCG2, where the binding site is located exactly at the symmetry axis. This leads to the misalignment of the pseudo-symmetric ligand densities during data processing. In case of crystallographic approaches, capturing such a ligand molecule in the specific cavity requires relatively high binding affinities. For this reason, the protein-drug incubation should be done properly in order to increase ligand`s occupancy in 3D crystals. On the other hand, cryo-EM SPA requires much lower binding affinities, significantly increasing the chances of capturing the drugs.

For membrane protein cryo-EM, a good choice of the detergent is important in order to be able to reach high-resolution. A suitable detergent will influence the sample behavior

and the data quality, which is major issue during the sample preparation. For example, DDM-reconstituted membrane protein samples require extremely high concentration for cryo-EM grids preparation, approximately 4 to 8 mg/ml, which then reaches the concentration ranges required for X-ray crystallography. Despite such high protein concentration, the distribution of particles in images is usually still too low with limited views of proteins in the vitreous ice, resulting in unsuccessful 3D reconstruction and limited high-resolution information. However, with other types of detergents such as LMNG or lipid-nanodiscs, which were applied for the here presented studies, the protein concentration can be 5-10 times lower compared to DDM. The lipid-nanodisc reconstituted samples are usually oriented nicely randomly with excellent particle distributions.

Finally, cryo-EM delivers insights into conformation changes of the protein that required for accomplishing its specific biological functions. Cryo-EM-determined protein structures are determined in a more natural environment compared to other methods. Cryo-EM does not require forming crystalline lattices with the proteins, and usually does not require protein labeling. Nowadays, with the new generation of image processing software, it is possible to capture multiple distinct protein conformations from one dataset. Furthermore, it is also possible now to isolate the high free-energy intermediate conformational fractions from cryo-EM data, which potentially reveals the biological mechanism with the meaningful structures. In Chapter IV describing the LptD/E work, we have captured such an intermediate conformation, providing direct evidence for LPS outer membrane assembly lateral opening hypothesis, which suggests a direction for antibiotic discovery in the near future.

An interesting topic that comes to mind regarding two of the projects presented in this thesis is the source of energy of translocation machineries. For ABC transporters, it is relatively well understood. It is clear that the ATP binding drives the NBDs dimerization and conformation swapping, consequently resulting in the translocation and further release of the substrate. As long as the ATP molecules are hydrolyzed and the Pi and ADP molecules get released from the ATP binding site, the transporters are returned to the initial state and prepared for the next transporting cycle. In contrast to the ATP coupling

protein machineries, the situation for outer membrane transporters is complicated as there is no ATP resource in this specific area. Thus, the "energy" is stored in the environment or by the protein molecules themselves. For the LptD/E complex, it seems that the beta barrel lateral gate is switched from the closed to the open conformation spontaneously and the energy for this switch may be provided by Brownian motion within the fluidity of the lipid bilayer. At the same time, protein-protein interactions may induce the conformation changes as well, such as the LptA-D regulation in the context of the formation of the LPS transporting bridge, or some other outer membrane transporters, like the type I pili secretion.

To sum up, as a follow up for my work, I would suggest to perform the cell-based cryo-EM analysis for membrane protein drug targets in presence and absence of their ligands. Here, the cell-based cryo-EM structure determination means that the cryo-EM images will be recorded directly on the frozen cells, which has a much closer physiological relevance. In a way, this will be like conventional cellular imaging by electron microscopy, but in the future will be possible at a much higher resolution. Cryo-electron tomography (cryo-ET) and sub-volume analysis (SVA) can now routinely reach 1 nm resolution, and a few first cases exist where the combination of SVA with single particle image processing allowed reaching side-chain resolution (3Å) for proteins that were captured within the cellular context (Bouvette et al., 2020) [1]. In my view, this kind of approach will reach high resolution also for more challenging proteins in the coming years. However, this would still be a challenging task.

Alternatively, for each of the projects presented in my PhD thesis, I would like to propose relatively easier strategies to continue research:

For Chapter III, we have performed structural studies of human ABCG2 bound to anti-cancer drugs. What can be done next? We have preliminarily obtained inhibitory nanobodies against human ABCG2 and the structures were resolved by cryo-EM (data not

[1] Bouvette J, Liu H-F, Du X, Zhou Y, Sikkema AP, Da Fonseca Rezende E Mello J, Klemm B, Huang R, Schaaper RM, Borgnia MJ, Bartesaghi A. 2020. Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. bioRxiv 2020.09.24.294983. doi:10.1101/2020.09.24.294983

shown [2]). These nanobodies bind to ABCG2 at the intracellular side, which therefore makes it not easy to deliver them inside the cells to block the transporter, making them bad drug candidates. In addition, another inhibitory 5D3-Fab that we used in this study was also not suitable for use as a drug due to its brain toxicity and low binding efficiency. In this context, it should be interesting as future research to proceed with sybody-based drugs for ABCG2, similar to the approach in Chapter IV. This could result in cell surface bound nanobodies with a higher binding affinity. The final goals would be to determine cryo-EM structures of these sybodies bound to ABCG2 and to test the inhibition in vivo if possible. Another direction would be to search new small-molecule drug candidates with higher anti-cancer activity or modify the existing drugs in order to abolish the binding affinity for ABCG2 through a structure-based design strategy, such as the mitoxantrone-like molecules.

For Chapter V, the analysis of the Spike:ACE2 complex, it would be important to confirm that an overloading with soluble ACE2 can indeed disrupt the Spike timer also in vivo. Cryo-ET investigation of SARS-CoV2-infected host cells would be able to provide valuable insight into the physiological stoichiometry of the Spike and ACE2 during and after virus invasion. Still, small molecules might be the most promising drug candidates to target the Spike of SARS-CoV2. Cryo-EM SPA is the most promising tool for such studies.

For Chapter IV, we have discovered the new conformations and new complex arrangement of LptD/E involved in the LPS functional cycle, mainly using cryo-EM SPA. Additionally, with our structures, we were able to make new observations that may be related to antibiotic discovery. First of all, in our structure, the C-terminal tail of LptD is inserted into the LptD-NTD jellyroll domain, which completely blocks the LPS transporting bridge. Thus, this C-terminal sequence might act as inhibitor[3]. This requires further investigation. Secondly, in our dimeric structure, we have suggested a potential organization of the LptA-LptD

[2] Rossitza Irobalieva, Scott M. Jackson, Ioannis Manolaridis, Dongchun Ni, Julia Kowal, Henning Stahlberg#, Kaspar P Locher1# Nanobody screening and the structural basis of nanobody-based inhibition of human multidrug transporter ABCG2 (in preparation)

[3] Mathieu Botte*, Dongchun Ni*, Stephan Schenck1*, Iwan Zimmermann, Mohamed Chami, Nicolas Bocquet, Denis Bucher, Pascal Egloff, Matilde Trabuco, Robert K.Y. Cheng, Janine Brunner, Markus A. Seeger, Henning Stahlberg# and Michael Hennig# Insight into the transport mechanism of N. gonorrhoeae LPTDE provided by the cryo-EM structure in complex with Pro-Macrobodies. (to be submit) Equal first author; Chapter IV on this thesis

complex. With the proposed open/close LPS transporting bridge regulation mode, another inhibitor could be designed based on the interface between LptA and LptD-NTD. In addition, there is an existing antibiotic called murepavadin (LptD targeting drug) with a so far unknown binding site. Therefore, cryo-EM SPA would be a powerful tool to help us to understand how this cyclic peptide works for the inhibition of LptD.

Finally, because our group is highly interested in neurodegenerative diseases, I have proposed a direction relevant for this field. Mitochondrial outer membrane proteins: VDAC-based complex. Since the function of mitochondria is highly related to the neurodegenerative disorder and the VDAC protein is an important molecule that regulates mitochondrial ion homoeostasis, it would be a good target to be further investigated, especially the structures of VDAC-ligand complexes. In addition, the VDAC channel is a good example to further develop the technology specifically for the aim of small membrane protein (below 40KDa) cryo-EM SPA.

What will happen to the field of structural pharmacology in the near future? Here I propose three possible directions, which may become reality in next years.

1) High-resolution structure determination by cryo-EM of cellular or non-purified samples. This approach determines high-resolution structures without purification by directly imaging entire cells. In my opinion, this may be not too hard and will happen probably very soon, for example in 3 years.

2) High-throughput and high-resolution cryo-EM SPA. The current systems are already very fast and highly sensitive. However, they are still not sufficiently fast for high-throughput screening and data collection in the context of drug discovery. Therefore, it may require a new system with better mechanical stability. My prediction is around 5 years.

3) High-resolution cryo-EM structure determination for asymmetrical small membrane proteins, below 40 kDa without any protein binders (Fabs or nanobodies). It sounds not very difficult, but in fact it is. Such protein sample is just like a blob and there are no features useful for particle alignment. In order to obtain structures with side chain resolution, the SNR must be improved dramatically and requires the next generation of electron microscope and new cameras. This will be probably true in around 8 years.

7 Acknowledgments

First of all, I express the great gratitude to my supervisor Henning Stahlberg. Without his help and the supporting from him, I wouldn`t get the fantastic opportunity to work on such excellent projects over my PhD in the past three years. I have learned really a lot and thanks a lot. I appreciate Camilo Perez for being my secondary supervisor and a lot of helps for the PAC meeting. I`m very much thankful to Andrea Ablasser for reviewing my thesis as the external expert and participating in my PhD defense. I thank Thomas Braun for chairing my PAC meeting. Greatly thank Jaroslaw Sedzicki for many discussions over my PhD as well as his help for the comments on my thesis. I thank the whole C-CINA and bioem groups, especially I thank Lubomir Kovacik, Kenneth Goldie and Mohamed Chami for the assistance of the EM operations and so on. I thank Frank Lehmann for offering the assistance of protein production and Karen Bergmann for solving the issues related to my PhD. Thank Kai Schmitz for being of my neighbor in the office and the lab and for the daily chatting. Secondly, I thank Prof. Kaspar Locher and his group members, especially Ioannis Manolaridis, Julia Kowal and Scott Jackson for the wonderful collaboration on ABCG2 project. At the same time, I thank Michael Hennig and Mathieu Botte and many other members from leadxpro for the collaboration on the LptD/E project. Thank Kelvin and Florence Pojer for providing the samples of Spike and ACE2, so that we can investigate the structures related to COVID-19 in this year. I acknowledge the NCCR-TransCure for providing the good network to do the research about the membrane protein drug targets. Finally, I must thank my wife Lihua for the her supporting in the past many years.

8 CV

Dongchun Ni
PhD student
Date of Birth: 11/28/1989
Tel: 078 746 3313
Email: dongchun.ni@unibas.ch

C-CINA, Biozentrum, Uni Basel, Mattenstrasse 26, Room P28, CH-4058 Basel, Switzerland

Research interests:
Single particle Cryo-EM study of ABC transporters, membrane protein drug targets.

I joined Prof. Henning Stahlberg's lab at the University of Basel as a PhD student in December 2017. I used to be a master`s student at Tianjin Medical University as well as the Institute of Biophysics of the Chinese Academy of Sciences, focusing on outer membrane protein structures and functions by X-ray crystallography. After that, I went to Harvard Medical School/Boston Children's Hospital for the Visiting research collaboration, mainly using the crystallography and Cryo-EM. Currently, I got the great opportunity to apply the high-resolution cryo-electron microscopy to unravel the 3D structure of proteins and understand their biological function and involvement in human diseases.

Education:

Dec, 2017 -	Universität Basel, Basel, Switzerland PhD student, in Structural Biology& Cryo-EM
Sep, 2011 - Jul, 2014	Tianjin Medical University, Tianjin, China Master`s student, in Public Health Toxicology.

Jan, 2012 - April, 2014	Institute of Biophysics, Chinese Academy of Sciences, Beijing, China Graduate student, in Structural Biology, Membrane protein crystallography.
Sep, 2007 - June, 2011	South-Central University for Nationalities, Wuhan, China Undergraduate student, in Pharmaceutical Science.

Position:

Jan, 2017- May, 2017	Department of Cardiovascular Diseases, Tianjin XiQing Hospital, Tianjin , China Research assistant
Dec, 2015 - Dec, 2016	Harvard Medical School/ Boston Children`s Hospital, Boston, USA Research scholar
Jul, 2014 - Dec, 2015	Department of Cardiovascular Diseases, Tianjin Xiqing Hospital, Tianjin , China Research assistant
July, 2010- Apr, 2011	Department of Neurobiology, Peking university, Beijing, China Internship

Previous publications (not during PhD studies):

1. **Ni D**, Yang K, Huang Y*. Preliminary X-ray crystallographic studies of the β-barrel domain of BamA, a membrane protein essential for outer membrane protein biogenesis. Acta Crystallographica Section F: Structural Biology Communications. 2014, 70(3) :332-335

2. **Ni D,** Wang Y, Yang X, Zhou H, Hou X, Cao B, Lu Z, Zhao X, Yang K, Huang Y*. Structural and functional analysis of the β-barrel domain of BamA from Escherichia coli. FASEB J.2014 ,28(6):2677-85.

3. Cao B, Zhao Y, Kou Y, Ni D, Zhang XC, Huang Y*. Structure of the nonameric bacterial amyloid secretion channel. Proc Natl Acad Sci. U S A. 2014 Dec 16;111(50):E5439-44

4. **Ni D**, Huang Y*. The expression, purification, and structure determination of BamA. Methods in Molecular Biology. 2015;1329:169-78.

5. Wang, Y., Pannuri, A.A., **Ni, D**., Zhou, H., Cao, X., Lu, X., Romeo, T. and Huang, Y., 2016. Structural Basis for Translocation of a Biofilm-supporting Exopolysaccharide across the Bacterial Outer Membrane. Journal of Biological Chemistry, 2016, 291(19), pp.10046-10057.

6. Han L, Zheng J, Wang Y, Yang X, Liu Y, Sun C, Cao B, Zhou H, **Ni D**, Lou J, Zhao Y, Huang Y*. （2016）Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. Nat Struct Mol Biol. Mar; 23（3）:192-6

7. Chen H, Song G, Zhang Y, **Ni D**, Zhang X, Huang Y*, Lou J. Mechanical unfolding of a β-barrel membrane protein by single-molecule force spectroscopy. Science China Life Sciences. 2020 Jul 28:1-3.

9 Publications (PhD studies):

1. Julia Kowal*#, Dongchun Ni*, Scott M Jackson1, Ioannis Manolaridis1, Henning Stahlberg#, Kaspar P Locher1# Structural basis of drug recognition by multidrug transporter ABCG2. (submitted) Equal first author; Chapter III on this thesis

2. Mathieu Botte*, Dongchun Ni*, Stephan Schenck1*, Iwan Zimmermann, Mohamed Chami, Nicolas Bocquet, Denis Bucher, Pascal Egloff, Matilde Trabuco, Robert K.Y.

Cheng, Janine Brunner, Markus A. Seeger, Henning Stahlberg# and Michael Hennig# Insight into the transport mechanism of N. gonorrhoeae LPTDE provided by the cryo-EM structure in complex with Pro-Macrobodies. (to be submit) Equal first author; Chapter IV on this thesis

3. Dongchun Ni, Kelvin Lau, Frank Lehmann, Andri Fränkl, David Hacker, Florence Pojer and Henning Stahlberg# Structural investigation of ACE2 dependent disassembly of the trimeric SARS-CoV-2 Spike glycoprotein. bioRxiv 2020.10.12.336016; doi: https://doi.org/10.1101/2020.10.12.336016 First author; Chapter V on this thesis

4. Guerrero-Ferreira, Ricardo, Lubomir Kovacik, Dongchun Ni, and Henning Stahlberg. "New insights on the structure of alpha-synuclein fibrils using cryo-electron microscopy." Current Opinion in Neurobiology 61 (2020): 89-95.

5. Dongchun Ni, Ioannis Manolaridis, Scott M Jackson, Julia Kowal, Henning Stahlberg#, Kaspar P Locher# Structure of wild type Human ABCG2 in a Vanadate-trapped post-hydrolytic state and the inhibitory 5D3 Fabs upon closing. (in preparation)

6. Qin Yu. Dongchun Ni et al Structural basis of … (in preparation)

7. Rossitza Irobalieva, Scott M. Jackson, Ioannis Manolaridis, Dongchun Ni, Julia Kowal, Henning Stahlberg#, Kaspar P Locher1# Nanobody screening and the structural basis of nanobody-based inhibition of human multidrug transporter ABCG2 (in preparation)