Function and inhibition of the lipopolysaccharide floppase MsbA

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

Function and inhibition of the lipopolysaccharide floppase MsbA

presented by Francois Antoine Thelot

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Andrew Kruse
Andrew Kruse (Sep 1, 2021 15:35 EDT)

Typed Name: Dr. Andrew Kruse

Signature: Thomas Bernhardt
Thomas Bernhardt (Sep 2, 2021 17:43 EDT)

Typed Name: Dr. Thomas Bernhardt

Signature:

Typed Name: Dr. Rachelle Gaudet

Signature:

Typed Name: Dr. Xiaochen Bai

Date: August 27, 2021

Function and inhibition of the lipopolysaccharide floppase MsbA

A dissertation presented

by

Francois Antoine Thelot

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Chemistry and Molecular Pharmacology

Harvard University

Cambridge, Massachusetts

August 2021

Dissertation Advisor: Maofu Liao Francois Antoine Thelot

Function and inhibition of the lipopolysaccharide floppase MsbA

Abstract

MsbA is an ATP-binding Cassette (ABC) transporter tasked with flipping lipopolysaccharide (LPS) across the inner membrane of Gram-negative bacteria. Interrupting LPS transport by blocking MsbA leads to cell death, making MsbA an attractive target for developing new antibiotics against multidrug-resistant infections. Although high-resolution structural information can often accelerate drug discovery, studies of MsbA have thus far been complicated by the extraordinary conformational heterogeneity of this transporter.

This dissertation exploits recent progress in cryo-electron microscopy (cryo-EM) to understand the MsbA function cycle and possible inhibition mechanisms. Imaging of MsbA in presence of different nucleotides shows that nucleotide binding is sufficient for substrate release and provides information about which conformations are achievable in lipid membranes and the relative prevalence of each state during continuous ATP turnover. Characterization of MsbA in complex with various modulators illustrates distinct mechanisms of inhibition by small molecules. Notably, the structure of MsbA bound to a decoupler of ATPase activity from LPS transport reveals that the substrate-binding pocket can be remodeled by induced-fit, and that it is an important site of allosteric regulation. In this novel inhibited ABC transporter conformation, significant rearrangements in the transmembrane domains lead to unproductive hydrolysis of ATP. In contrast, structures in complex with an ATPase suppressor demonstrates an inhibition mechanism involving symmetrical separation of the nucleotide-binding domains and impairment of the normal MsbA conformational transition. The findings presented here will inform the future development of modulators of MsbA and other ABC transporters.

Table of Contents:

Table of Figures:

Acknowledgements

First, I am grateful for Maofu's patient mentorship, hands-off teaching style, and for the incredible kindness he has shown me over the years. Structural biology is a fiercely competitive field, and my graduate school years have included both thrills from determining new structures and brutal letdowns after being scooped. Through these ups and downs, one constant in my life has been Maofu's undying optimism. I am thankful for Maofu's trust that I can tackle challenging scientific problems, and that he equipped me to do so by providing insightful advice and an outstanding research environment. I am humbled by his continued support throughout my PhD.

I have been fortunate to overlap with Ben Orlando in the Liao Lab from 2016 to 2020. Ben has taught me probably 80% of everything I know about protein purification, operating cryo-electron microscopes, and data processing. I am grateful to have benefitted from his mentorship. I expect great things from the newly-formed Orlando lab at Michigan State.

Wei, Yanyan, Min, Melissa, Ashlee, Dvir, Young-Cheul, Stuti, Al, Claudio, Shuhui, Xuewu, Kenny, thank you for making the Liao Lab a welcoming and stimulating place to work. I will miss the atmosphere of intense collaboration, as well as the occasional late-night chats about science and life.

Thank you to my dissertation advisory committee, Adrian Salic, Bob Farese, and Andrew Kruse for stimulating discussions during my Preliminary Qualifying Exam, and then throughout multiple meetings. I would further like to thank Andrew for giving me the opportunity to gain teaching experience by TA-ing his biochemistry class, as well as for making critical contributions to my research, including providing the initial push that I study the MsbA stimulator TBT1 in *A.*

baumannii instead of stubbornly working on an *E. coli* structure. What seemed like a simple suggestion had an outsize impact on the course of my research.

Thank you to Peng Yin for welcoming me to his lab for a first-year rotation, and for Di Liu for supervising me throughout this rotation. I am fortunate to have collaborated with them over the past two years on a project that combines the expertise of both the Yin and Liao labs.

I feel privileged to have been given the opportunity to pursue my PhD at Harvard. The Harvard community includes some of the world's best scientists, including in its many core facilities. I specifically would like to thank Zongli Li, Melissa Chambers, and Sarah Sterling for investing considerable time training me as an electron microscopist. Spending quiet time in the Polara room has been the height of my grad school experience.

Thank you to my past research mentors at Duke, Bruce and Bernard, for welcoming me into their labs and introducing me to scientific research. Thanks to their guidance, I couldn't have been better prepared for graduate school.

Many thanks to my brilliant scientist and physician friends, Alex, Bryan, Dvir, Orhan, Tong, Max, Hanchen for making my time in grad school a bearable and at times pleasant experience. If a man truly is the average of his closest friends, then I have done reasonably well for myself and can confidently look to the future.

Thank you to my fiancée Elizabeth for sharing the past few years with me. I'm a bit of a creature of habit, and I appreciate your adventurous spirit and willingness to take me out of my comfort zone. Life wouldn't be as much fun without you.

Finally, I owe any success I've had to the unconditional support and love from my parents Antoinette and Franck. Coronavirus travel restrictions have made separation particularly difficult. I look forward to being reunited soon.

Chapter 1: Introduction

Although most material presented here was specifically generated for this dissertation, select sections are published as:

Thelot, F., Orlando, B. J., Li, Y., & Liao, M. (2020). High-resolution views of lipopolysaccharide translocation driven by ABC transporters MsbA and LptB2FGC. *Curr Opin Struct Biol*, 63, 26-33. doi:10.1016/j.sbi.2020.03.005

1.1 Introduction to ABC transporters

Active transport of ions, metabolites and lipids against their concentration gradient requires release of chemical energy, and typically involves either adenosine triphosphate (ATP) hydrolysis (primary transport) or coupling of transport with a separate substrate going down its electrochemical gradient (secondary transport). Proper molecule transport across biological membranes underpins cell homeostasis and growth and is regulated by diverse types of integral membrane proteins, including Major Facilitator Superfamily (MFS) proteins[1,2], Resistance-Nodulation cell Division (RND) transporters[2], P-type ATPases[3] and ATP-binding Cassette (ABC) transporters[4-6].

ABC transporters are one of Nature's ingenious solutions for coupling chemical energy released by ATP hydrolysis to substrate transport across membranes[4-6]. The ABC superfamily is ubiquitous and strikingly conserved throughout life, hinting at an ancient evolutionary origin[7]. In bacteria, ABC transporters mediate nutrient import, toxin secretion and lipid membrane synthesis[8]. In plants, they participate in cell detoxification and hormone transport[9]. In humans, 48 ABC transporters perform diverse functions including lipid flipping and ion conduction[6,10]. Throughout the tree of life, their relative promiscuity with structurally diverse ligands and general affinity for hydrophobic substrates confers ABC transporters with multidrug transport capabilities[11-13]. Unspecific export of cytotoxic compounds by ABC proteins is of critical importance to human health, because it is a primary mechanism by which bacteria become resistant to antibiotics[11,12], and cancer cells insensitive to chemotherapy[13].

How a single molecular framework accommodates such drastically diverse substrates and cellular functions is a fascinating mystery. ABC transporters are universally composed of four necessary subunits: two Transmembrane Domains (TMDs), and two Nucleotide-Binding Domains

(NBDs), encoded by one (e.g., ABCB1[14]) to four (e.g., Lpt[15]) polypeptides and arranged according to an exact or pseudo two-fold symmetry. These minimal domains may be conjugated to accessory subunits, including substrate-binding proteins (e.g., Maltose importer[16]), extra helical bundles (e.g., SUR1[17]), and peptidase domains (e.g., PCAT[18]), which assist the transporter in its function. These protein domains must precisely coordinate substrate capture, ATP binding and hydrolysis, and substrate translocation. Over the past two decades, considerable structural biology efforts have aimed at understanding the ABC transporter function cycle[4-6], as well as discovering pharmacological tools to modulate transporter conformations[19], with potential applications in treating a range of human diseases.

In the next paragraphs, we first describe the general architecture of ABC transporters, before providing an overview of the critical role of ABC transporters in cell membrane synthesis in Gram-negative bacteria, and finally introducing MsbA, the ABC transporter that is the focus of this study.

1.2 Architecture and mechanism of the Nucleotide-Binding Domains

The Nucleotide-Binding Domains (NBDs) are the catalytic core of ABC transporters and are tasked with binding and hydrolyzing ATP. In most ABC proteins – except ABCE[20] and ABCF[21] families –, NBDs provide the motile force for substrate translocation in the transmembrane domains.

The sequence and structure of NBDs includes large similarities across ABC transporters, hinting at the conserved function of NBDs. From amino to carboxy terminal, the conserved NBD motifs (***Figure 1.1a,b***) include[4,5,22]: 1) The A loop, which provides an aromatic sidechain for pi-stacking with the ATP purine ring; 2) The Walker A (GxxGxGKST, with "x" referring to any arbitrary residue), which stabilizes the α and β phosphates of ATP; 3) The Q loop, whose essential

glutamine residue coordinates an Mg^{2+} ion and stabilizes the ATP γ-phosphate; 4) The ABC signature motif (LSGGQ), which stabilizes the ATP molecule and coordinates an Mg^{2+} ion; 5) The Walker B (ϕϕϕϕDE, with "ϕ" referring to a hydrophobic side-chain), including the essential catalytic glutamate that activates the water nucleophile during ATP hydrolysis; 6) The D-loop (SALD), whose function is uncertain[6]; 7) The H loop, which provides a conserved histidine that participates in coordinating γ-phosphate, attacking water and catalytic Walker B glutamate.

Importantly, the nucleotide is positioned between the signature motif of one NBD, and the Walker A motif of the opposite NBD, implying that ATP binding and hydrolysis demands NBD dimerization in a head-to-tails manner (***Figure 1.1b,c***). Structural characterization of stable ATP-NBD complexes requires preventing ATP hydrolysis, and is usually performed in one of two ways: 1) mutation of the catalytic Walker B glutamate to glutamine, which still allows for ATP binding but renders the transporter incompetent for ATP hydrolysis[23], 2) use of a non-hydrolyzable ATP analogues, such as adenylyl-imidodiphosphate[24] (AMP-PNP), ADP-vanadate[25] or ADP-beryllium fluoride[22].

High-resolution crystal structures of full-length transporters in complex with nucleotides, including those of the maltose importer MalFGK[22] (***Figure 1.1***), have greatly enhanced our understanding of how ABC transporters bind and hydrolyze ATP. In the nucleotide-free state, the two NBDs are kept separate from each other (***Figure 1.1c***). ATP binding promotes NBD dimerization, and triggers a power stroke that propagates to the TMDs and is often accompanied by substrate release. ATP hydrolysis subsequently allows for NBD separation and prepares the transporter for a future ATPase cycle (***Figure 1.1c***).

While the general mechanism by which NBDs bind and hydrolyze ATP is well-documented, important details about NBD function remain elusive. For instance, many

transporters include a functional and a degenerate ATP binding site (e.g., ABCC1[26], TmrAB[27], ABCB11[28]) that is able to bind ATP but hydrolysis-incompetent, while other transporters (e.g. ABCG2[29], ABCB4[30], MsbA[25]) have two functional sites. Further adding complexity, some transporters have two functional ATPase sites, yet with different affinities for ATP such that binding must first occur at one specific site (e.g., ABCB1[31,32]). Such diversity of nucleotide-binding pockets raises questions about the mechanism of transporters with two functional sites: can normal ATPase cycling proceed with a single nucleotide bound to the NBDs? If two nucleotides are bound, is ATP hydrolysis simultaneous at both sites, or does it occur as two sequential events? Further biochemistry efforts may be required to better delineate the exact mechanism by which NBDs coordinate substrate transport and ATP hydrolysis.

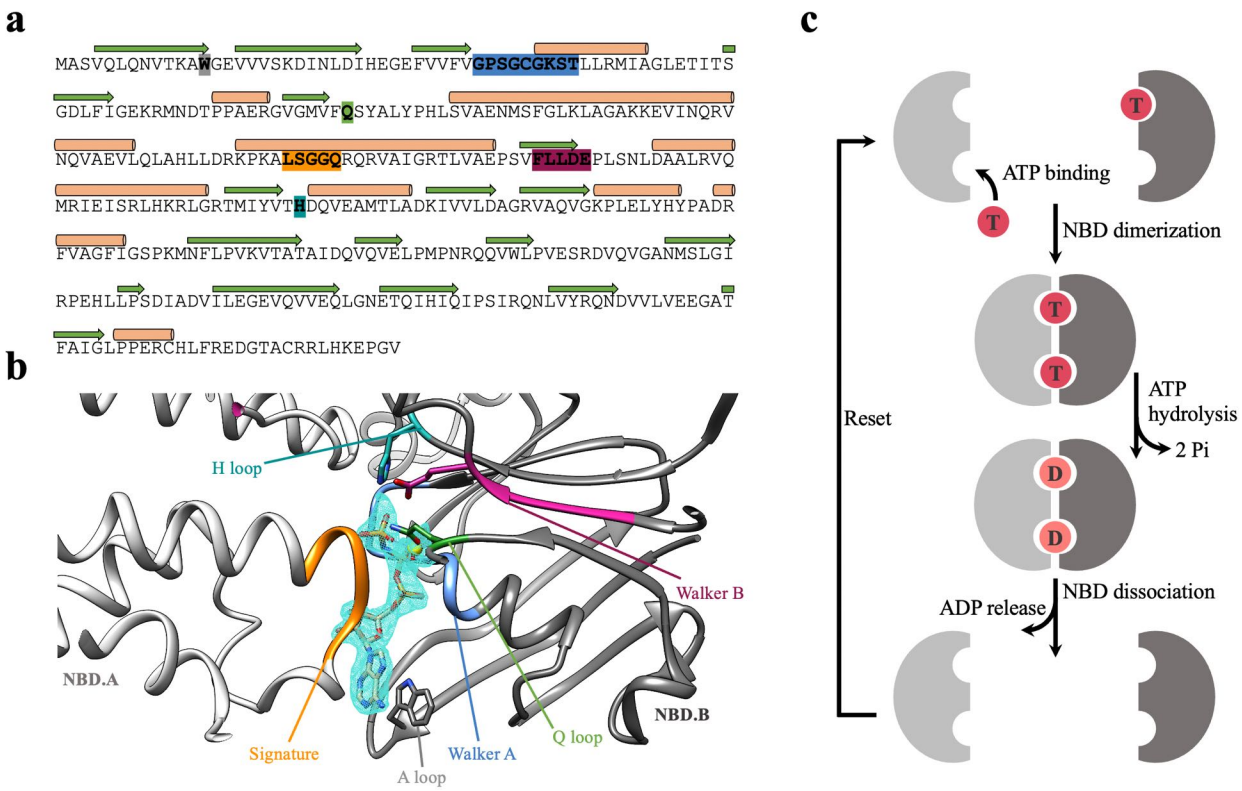


Figure 1.1: Architecture and function of Nucleotide-Binding Domains. a, Sequence and secondary structure for MalK, the Nucleotide-Binding Domain of the maltose importer. Conserved NBD motifs are highlighted with specific coloring (grey: A-loop, blue: Walker A, green: Q-loop, orange: ABC signature motif, maroon: Walker B, turquoise: H-loop). Beta strands (green arrows) and alpha helices (orange cylinders) are labelled above the MalK sequence. **b,** Overview of the nucleotide binding site for MalK in complex with AMP-PNP[22] (PDB: 3RLF). Conserved NBD motifs are labeled. The nucleotide is displayed as zoned density from the deposited electron density map. **c,** Model for NBD dimerization and ATP hydrolysis.

1.3 Structural diversity and classification of transmembrane folds

The transmembrane domains (TMDs) are tasked with recognizing the ABC transporter substrate and promoting its translocation through a lipid membrane. Reflecting the diversity of recognized substrates, TMDs are less structurally conserved than the NBDs and can be used as the basis for ABC transporter classification[33]. ABC transporters have historically been classified based on sequence similarity (e.g., ABC1-3 nomenclature[34]), function (e.g., MDR/MRP classifications[35]), or on TMDs fold as observed in the scarce full-length crystal structures available (e.g., type I vs. type II importer/exporter classification[36]). However, the remarkable proliferation of cryo-EM structures of full-length ABC transporters since ~2015[37,38] has considerably expanded our understanding of existing ABC transporter folds[39] and has provided new opportunities for ABC transporter classification. Under a new nomenclature system, ABC transporters are classified according to TMDs fold in one of seven classes[6,33] (***Figure 1.2***).

Classes I, II and III include TMDs folds that exclusively consist of bacterial and archaeal importers, which are critical for absorption of nutrients such as sugars, amino acids, vitamins and metals. Within importers, classes I and II both require an accessory substrate binding protein, and are distinguished based on their number of transmembrane helices (TMHs), with class I having 6 to 8 TMHs per chain while class II includes 10 TMHs per chain[40]. Class III consists of a peculiar subgroup of ABC importers known as Energy-Coupling Factor (ECF) transporters, which do not require any periplasmic substrate binding protein and include an unusual TMDs fold known as ECF-T, as well as a transmembrane substrate-recognizing protein named S-component[41]. Class IV is the prototypical and most abundant ABC transporter fold and mostly includes exporters, with select exceptions (e.g., the siderophore importer YbtPQ[42]). Type IV transporters can recognize and transport an astoundingly diverse range of substrates, including phospholipids[30],

oligosaccharides[43], and chloride ions[44]. They are composed of 12 TMHs (6 per chain in a dimer, 12 in a monomer) and are characterized by their domain swapped TMHs, which extend to the opposite half of the transporter and couple one TMD with the diametrically opposed NBD. Similar to class IV, class V includes exporters with 12 TMHs but with no domain swapping helices (e.g., ABCG2[29], WzmWzt[45]). Consequently, the two halves of the transporter are kept separate, such that one TMD is only coupled with a single NBD. Type VI transporter includes a single transporter, $LptB_2FGC$[15], which has similar topology to type V exporters except that it is missing amphipathic elbow helices at its N-terminus. Finally, class VII consists of the MacB[46] and LolCDE[47] macrolide and lipoprotein extractors, which only contain 8 TMHs and include large periplasmic domains.

Despite the apparent divergence of TMDs folds, ABC transporters coordinate ATP hydrolysis with substrate binding and transport through a relatively conserved translocation cycle.

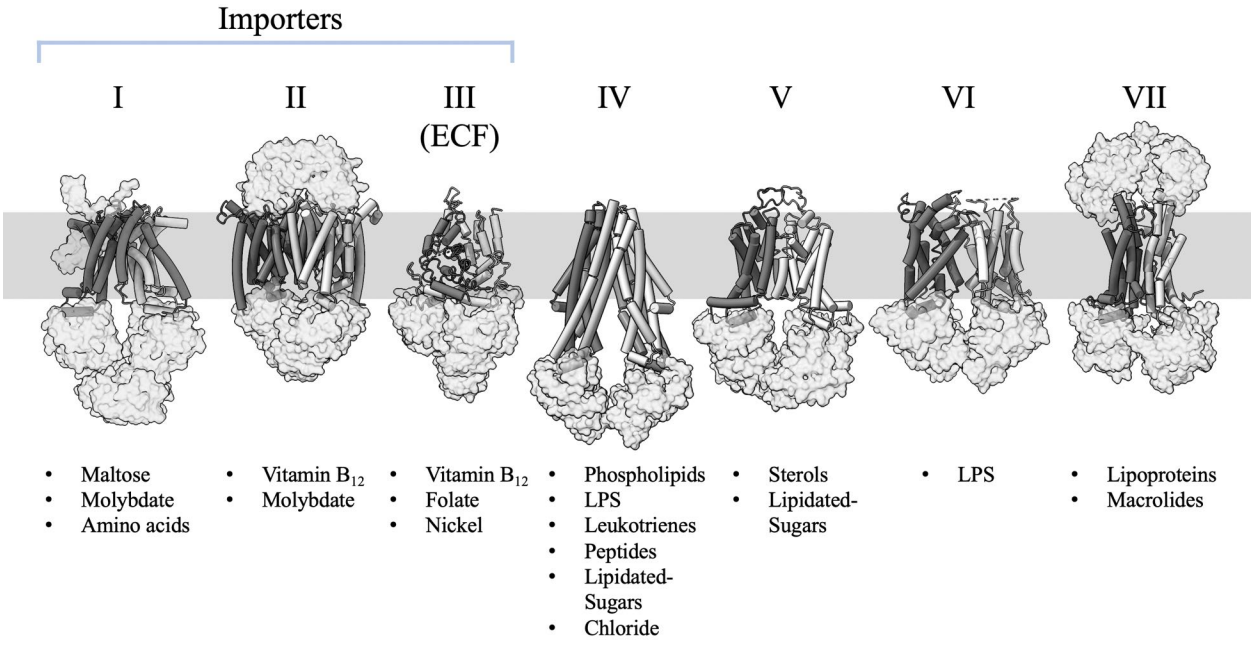


Figure 1.2: Classification of ABC transporter transmembrane domains folds and overview of transported substrates. Overview of the 7 ABC transporter transmembrane domains folds (top). Transmembrane helices are shown as cylinders, while nucleotide-binding domains and accessory domains are displayed as transparent surfaces. The following structures were used for each ABC class: type I: maltose importer (PDB: 3FH6); type II: vitamin B_{12} importer BtuCD (PDB: 2QI9); type III: ECF-CbrT vitamin B_{12} importer (PDB: 6FNP); type IV: MsbA (PDB: 5TV4); type V: ABCG2 (PDB: 6VXF); type VI: $LptB_2FGC$ (PDB: 6MI7); type VII: LolCDE (PDB: 7ARH). Selected substrates for each transporter class are listed (bottom).

1.4 A general model for type IV ABC transporter cycling

ABC transporters are dynamic proteins and adopt distinct conformations throughout the substrate transport cycle. As type IV exporters are the most abundant type of ABC transporter, the major conformations adopted by this transporter fold are outlined below (***Figure 1.3***).

In the absence of bound ligand or nucleotide, the transporter adopts a wide inward-facing conformation, with separate TMDs and inter-NBD spacing up to ~85 Å (i.e., *E. coli* MsbA in dodecyl-maltoside[24]). Substrate entry in the central transporter pocket may occur either through lateral openings enclosed by TMHs 4 and 6 or through the aqueous phase, depending on whether the substrate is hydrophobic and partitions in the membrane (e.g., MsbA[25]) or if it is readily water-soluble (e.g., ABCC1[26]).

Remarkably, substrate binding to the transporter may trigger a large-scale conformation change to a narrow inward-facing conformation, which results in reduction of inter-NBD spacing (***Figure 1.3***, state 2). The impact of substrate binding on transporter conformation has been best characterized for the ABCC1 transporter, which undergoes NBDs tightening from ~35 Å in the ligand-free state to ~22 Å in the presence of its native leukotriene C4 substrate[26]. It seems probable that substrate-induced decrease in inter-NBD distance may favor transporter cycling by increasing the rate of NBD dimerization, and therefore act as a key mechanism for preventing wasteful ATP hydrolysis.

Following substrate binding, two ATP molecules bind the NBDs, promoting a conformation change from narrow inward-facing to outward-open state (***Figure 1.3***, state 2 and 3). Notably, binding of ATP does not require that the NBD be dimerized or even in close proximity to each other, as structures of transporters with separate NBDs yet well-defined bound nucleotides have been determined (e.g., ABCG2 under ATP turnover condition[48]). In the outward-open state, the TMDs rearrange and create an opening between TMHs 1 and 3, allowing for substrate release on the periplasmic or extracellular side of the membrane (***Figure 1.3***, state 3). The outward-open state has so far only been observed in crystallography[24,49] experiments or in rare cryo-EM studies using non-hydrolyzable nucleotides[27]. As all efforts to characterize the outward-open conformation under ATP turnover conditions have been unsuccessful[27,50], it is likely that this conformation is extremely transient, such that the opening between TMHs 1 and 3 only lasts long enough to allow for substrate exit.

Upon substrate release, the TMDs collapse against one another in a conformation referred to as the outward-facing state (***Figure 1.3***, state 4). As the dense TMHs packing precludes substrate binding, the outward-facing state prevents substrate re-entry. Conversion of ATP to ADP is

thought to occur while the transporter is in the outward-facing state and precedes γ-phosphate exit from the transporter through hydrated NBD tunnels[27]. Finally, the NBDs dissociate and the transporter transitions to the wide inward-facing state for a subsequent translocation cycle. Interestingly, the ADP-bound conformation is the most abundant nucleotide-trapped state in ATP turnover experiments[27,50], suggesting that NBD dissociation after ATP hydrolysis may be a rate limiting step of the transporter kinetic cycle.

Successful completion of the substrate translocation cycle allows ABC transporters to perform critical functions in various organisms, including helping bacteria build their complex cell membranes.

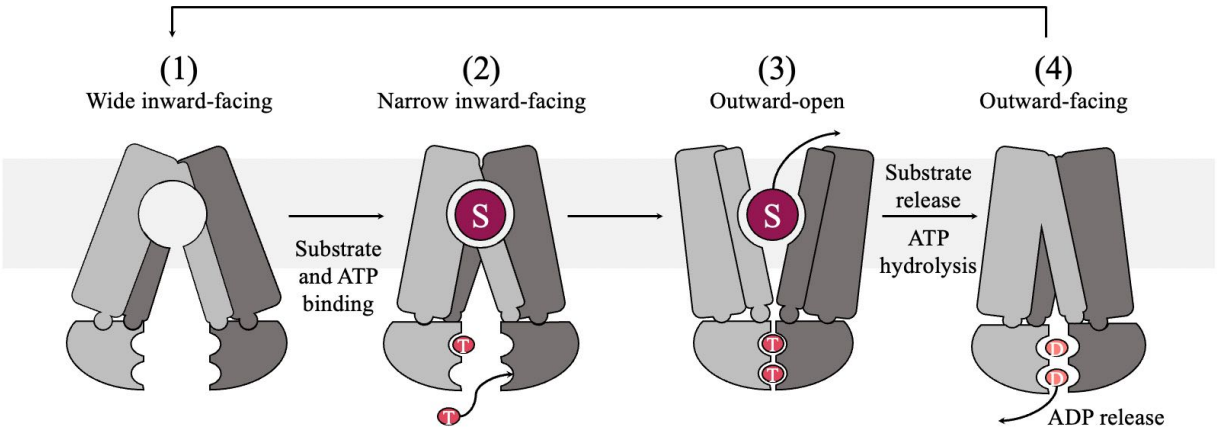


Figure 1.3: Translocation cycle for a type IV ABC exporter. Four major conformations of a type IV exporter throughout the substrate transport cycle. The lipid membrane is shown as a grey rectangle in the figure background. Transported substrate, ATP and ADP are represented as purple circle labeled "S", Red oval labeled "T" and pink oval labeled "D", respectively.

1.5 ABC transporters involved in bacterial cell membrane formation

Gram-negative bacteria are protected from the extracellular environment by an inner membrane (IM) and outer membrane (OM), which are separated by an aqueous periplasmic space about 20 nm wide[51] and a ~5 nm peptidoglycan layer[51] (***Figure 1.4***). The OM and IM have distinct lipid compositions and biophysical properties[52], which render the cell impermeable to antimicrobial compounds. The OM is an asymmetric bilayer, with the outer leaflet consisting of a glycolipid termed lipopolysaccharide (LPS), and the inner leaflet composed of phospholipids, including phosphatidylethanolamine (PE, ~75% of lipids[53]), phosphatidylglycerol (PG, ~20%[53]), and cardiolipin (CL, ~5%[54]). The IM is a symmetric phospholipid membrane with similar composition to the inner leaflet of the OM[52]. The outer leaflet of the IM and both leaflets of the OM also contain lipoproteins, which are membrane-anchored lipid-linked proteins performing various functions in bacterial physiology, including assisting peptidoglycan formation and OM biogenesis[55].

Proper formation of this complex double-membrane cell envelope depends on the efficient transport of (1) phospholipids, (2) lipoproteins, and (3) LPS across the periplasm to their target membrane destination. One common feature of these three classes of molecules is that they are relatively large and amphipathic, and therefore likely substrates to be recognized and translocated by ABC transporters. Indeed, five ABC transporters are involved in cell membrane formation in Gram-negative bacteria: Mla, Lol, WzmWzt, MsbA and Lpt (***Figure 1.4***). Understanding the function of these lipid transporters has exciting applications in perturbing cell membrane formation and developing new antibiotics. Thus, considerable research efforts have recently been dedicated to determining the structure and mechanism of Mla, Lol, WzmWzt, MsbA and Lpt. As of 2017, none of the structures of these five transporters had been solved except for MsbA. As of 2021, multiple structures had been determined for each one of these proteins. The resulting wealth of

structural information has not only advanced our understanding of how these proteins contribute to cell membrane biogenesis, but also clarified the general mechanisms of ABC transporter function. Some of the key findings for each type of transported substrate are described next.

Phospholipid transport pathway: the Maintenance of Lipid Asymmetry (Mla) system is the only ABC transporter known to transport phospholipids in Gram-negative bacteria[52]. It consists of an inner membrane ABC transporter complex (MlaFEDB), a periplasmic component (MlaC) and an Omp-associated MlaA protein. The periplasmic MlaC is presumed to function as a "ferry", shuttling a bound phospholipid between the IM-embedded MlaFEDB and the OM-bound MlaA[52]. Notably, it remains controversial whether the Mla system shuttles lipids from the IM to OM (anterograde transport), OM to IM (retrograde), or bidirectionally[56]. If one assumes anterograde transport, Mla likely functions as a lipid extractor, tasked with extruding its lipid substrate out of the IM and transferring it to the MlaC lipid acceptor. In retrograde transport, Mla would perform the reverse function by receiving lipids from MlaC and facilitating insertion in the IM.

Recent cryo-EM structures have revealed the unusual structure of the MlaFEDB ABC transporter component[56,57], which is not yet classifiable in the TMDs fold nomenclature[33]. Interestingly, the MlaE TMDs are capped with a hexameric MlaD Mammalian Cell Entry (MCE) domain, which is a conserved lipid-binding motif shared with the non-ABC lipid transporter LetB[58]. Furthermore, examination of these high-resolution cryo-EM maps has revealed a pair of stable phospholipids at the interface of the MlaE TMDs and of the MlaD MCE domain[56,57], thereby revealing the structural basis of phospholipid capture by Mla. Despite these achievements, further biochemical efforts will be required to unequivocally determine the directionality of lipid transport by Mla, as well as for characterizing novel Mla inhibitors.

Lipoprotein transport pathway: lipoproteins are recognized by the Sec translocon and twin-arginine system thanks to an N-terminal signal peptide, and subsequently inserted in the IM. Throughout the lipoprotein maturation process, the signal peptide is cleaved off and three acyl chains are covalently conjugated, ensuring that the lipoprotein is retained in the periplasmic leaflet of the IM[47,59]. The Localization of Lipoprotein (Lol) transporter assumes two critical functions in lipoprotein biogenesis: dictating whether triacylated lipoproteins will be kept in the IM or exported to the OM (according to lipoprotein sequence, or "+2 rule" in *E. coli*[59]), and extruding the lipoprotein from the IM.

Similar to Mla, the Lol system is composed of an IM ABC transporter complex (LolCDE), a substrate-binding periplasmic ferry (LolA), and an OM-embedded lipoprotein acceptor (LolB). Yet contrasting with Mla, LolCDE is a type VII ABC transporter with only 4 TMHs per chain and structural homology to the multidrug exporter MacB. Furthermore, each of its TMD is fused to a large periplasmic domain that may transiently bind the lipoprotein substrate and is likely involved in lipoprotein handoff to LolA. The 2021 structures of LolCDE in detergent[47] and nanodisc[59] in several conformations, including apo, nucleotide-bound and LolA-bound[47], have greatly enhanced our understanding of how Lol recognizes lipoproteins and transfers them to the LolA acceptor. The next frontier in Lol-targeting antibiotic development may consist in understanding its mechanism of inhibition by small molecules such as the intriguing lipoprotein transport inhibitor G0507[60].

LPS transport pathway: LPS is a complex glycolipid composed of two major components that are independently synthesized in the cytoplasmic leaflet of the IM: (1) "rough" LPS, which contains a conserved lipid A moiety as well as core oligosaccharide[61-63]; (2) repeating O antigen polysaccharide, which is conjugated to a lipid named undecaprenyl-phosphate (Und-P)[63]. MsbA

and WzmWzt are the corresponding ABC transporter floppases tasked with flipping rough LPS and lipid-linked O antigen across the IM, respectively.

MsbA has been extensively studied over the past two decades due to its excellent stability and biochemical behavior, as well as its structural similarity with human multidrug efflux pumps such as P-glycoprotein[64]. The first set of structures of MsbA in detergent were obtained using X-ray crystallography nearly two decades ago[24,65], though absence of LPS in these early structures made it unclear how MsbA recognizes its lipid substrate and drives its 180º rotation across the IM. More recently, the cryo-EM structure of MsbA in a native-like lipid bilayer clearly resolved a co-purified LPS molecule, revealing how the inner cavity of MsbA forms highly specific contact with the lipid A moiety[25].

Several redundant biochemical pathways are tasked with translocating O antigen across the IM[62], though WzmWzt is the only ABC transporter known to be implicated in flipping Und-P-linked O antigen[66]. Recent crystal and cryo-EM structures of WzmWzt[45,67,68] in apo and nucleotide-bound states have revealed how this type V ABC transporter changes conformations throughout its transport cycle, as well as the architecture of its unique carbohydrate-binding domains (CBDs), which are fused to the NBDs (Wzt) and asymmetrically positioned on one side of the protein[68]. Despite these promising advances, the structure of WzmWzt in complex with lipid-linked O antigen has so far been elusive, and the structural basis for O antigen recognition remains unknown.

Upon flipping of rough LPS and Und-P-linked O antigen across the IM, the two halves of LPS are covalently ligated by the WaaL enzyme[62], then recognized by the Lipopolysaccharide Transport (Lpt) ABC transporter complex. Unlike MsbA and WzmWzt, Lpt is a molecular

extractor rather than a floppase, tasked with extruding LPS from the periplasmic leaflet of the IM and transporting it to the OM via a 7-protein complex ($LptB_2FGCADE$).

Interestingly, despite mediating inter-membrane substrate transport like Mla and Lol, Lpt uses a different solution than the "ferry" system. Instead, the commonly accepted "PEZ" model[69] assumes that polymerization of LptA across the periplasmic space creates a physical bridge, through which LPS molecules are transported. While the existence of the LptA bridge is supported by biochemical evidence[70], structural characterization of a complete Lpt complex has so far been elusive.

The $LptB_2FGC$ IM subunit of Lpt includes interesting structural features, including a unique ABC architecture (type VI TMDs) as well as 3 periplasmic β-jellyroll motifs which contain hydrophobic grooves and can presumably shield the hydrophobic acyl chains of the extracted LPS substrate. Recent crystallography[71,72] and cryo-EM[15,73] works have revealed intricate details of the Lpt transporter, including how LPS is recognized by a ring of basic sidechains similar to that of MsbA[15,64], and have demonstrated that the membrane-embedded LptC paces the conformation cycle of Lpt by inserting itself between the two TMDs. Future efforts in Lpt structural biology may consist in determining the structure of the complete 7-protein complex, as well as of understanding mechanisms of Lpt inhibition.

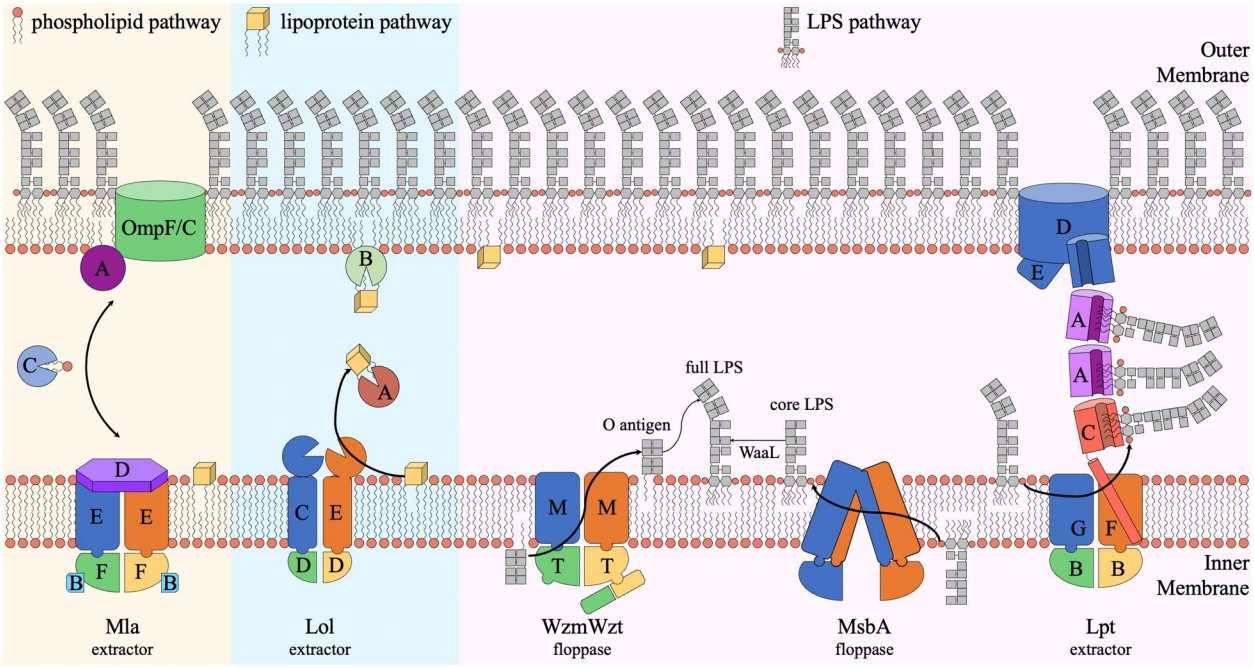


Figure 1.4: Overview of ABC transporters involved in Gram-negative bacteria cell membrane formation. Peptidoglycan layer not shown. Each protein subunit is labeled according to the letter corresponding to the gene encoding it. E.g., "E" on Mla refers to the MlaE protein.

MsbA stands out from Mla, Lol, WzmWzt, and Lpt because it has historically been the most extensively studied of these transporters, both in terms of understanding its functions and mechanisms of inhibition. Furthermore, it is the only transporter in this group to feature a classic type IV ABC transporter topology, suggesting that understanding its function may translate to targeting human type IV transporters such as p-glycoprotein and ABCC1. As MsbA is the focus of the rest of this thesis, we will provide a brief overview of the rich history of MsbA biochemistry and structural biology.

1.6 History of MsbA biochemistry

Identification of the msbA gene in the 1990s resulted from the study of htrB[74,75] (now lpxL, see ***Figure 1.5***), which is an essential gene for proper *E. coli* cell membrane biosynthesis at higher (>32º C) temperatures. Isolation of htrB null clones that could prosper at non-permissive temperatures resulted in the identification of two multicopy suppressor of htrB genes: msbA and msbB[74]. Despite complementing a similar phenotype, MsbA and MsbB belong to different classes of proteins. MsbB (now LpxM, ***Figure 1.5***), similar to HtrB, is an acyltransferase and participates in LPS maturation, such that its overexpression can conceivably overcome htrB deficiency by producing penta-acylated LPS. In contrast, MsbA belongs to the ABC superfamily, which at the time of these early findings was a relatively understudied class of proteins with elusive functions in many organisms[74,75].

While several redundant LPS acylation pathways presumably exist in *E. coli*[76], HtrB is likely the only acyltransferase capable of acylating tetra-acylated LPS at high temperatures. Notably, radioactive phosphate labeling of LPS and membrane fractionation reveal that htrB null mutants accumulate tetra-acylated LPS in the inner membrane when shifted to high temperature[76], suggesting that both htrB and its rescuing gene msbA may be involved in LPS synthesis, and that immature LPS cannot be efficiently transported to its target OM destination[76]. Complementation of htrB null strain by MsbA overexpression further reveals two important pieces of information: 1) MsbA overexpression does not result in LPS acylation[76], indicating that the function of MsbA is distinct from that of HtrB and MsbB; 2) Overexpression leads to transport of tetra-acylated LPS to the OM[76], indicating MsbA is likely an LPS transporter.

In this context, the mechanism by which MsbA complements HtrB deficiency comes into focus. Assuming that MsbA preferentially captures and transports mature LPS (e.g., lipid A,

Figure 1.5) rather than incomplete LPS (e.g., lipid IV_A), it is conceivable that a limited number of MsbA protein copies would be unsuccessful at transporting a large number of tetra-acylated LPS molecules. MsbA overexpression would restore lipid transport because a greater number of MsbA copies would overcome the challenge of transporting a lipid for which MsbA has lower affinity. In a similar fashion, MsbB overexpression restores lipid transport to the OM if one assumes that MsbA has greater affinity for penta-acylated LPS than for tetra-acylated lipids.

Subsequent studies of MsbA function were complicated by the fact that msbA knockouts are lethal and that msbA is located in the same operon as the essential orfE (now lpxK, ***Figure 1.5***) gene[77], preventing the direct study of MsbA-deficient strains. The fortuitous identification[77] of a temperature sensitive MsbA point mutant (A270T) provided an unprecedented opportunity to understand the effects of MsbA depletion. The A270T mutant is similarly active to wild type below 44º C and is rapidly inactivated at non-permissive temperatures[77], although it is unclear how a single mutation at the TMH 5 and 6 junction has such a spectacular influence on MsbA function. To understand the effect of MsbA inactivation, a strain harboring the A270T mutation is first grown under normal conditions at 37º C, such that lipid transport is functional. Temperature increase to 44º C and MsbA impairment then results in a drastic reduction in LPS and PE export to the OM, as assessed by membrane fractionation and thin-layer chromatography[77]. Furthermore, electron microscopy examination of A270T mutant thin sections reveals dramatic twists, bumps and invaginations of the IM[77], consistent with lipid transport blockage. The striking phenotype of MsbA-deficient bacteria highlights the potential of targeting this essential transporter for antibiotic development.

While it became clear in the early 2000s that MsbA is involved with transporting LPS from the IM to the OM, all experiments conducted thus far were compatible with two distinct and

perhaps contradictory models of MsbA function. (1) The flip-flop model[78] suggests that MsbA is a lipid floppase, tasked with capturing LPS in the cytoplasmic leaflet of the IM and flipping it to the periplasmic leaflet, such that MsbA inhibition would result in accumulation of LPS in the cytoplasmic side of the IM. (2) In contrast, the ejection model[78] states that MsbA captures LPS directly from the periplasmic leaflet of the IM and drives the transport of LPS towards the OM, possibly using help from an accessory protein with a similar role to LolA in the Lol system. According to the ejection model, disruption of MsbA function would result in LPS accumulation in the periplasmic leaflet of the IM. Thus, distinguishing between the flip-flop and ejection models hinges on determining LPS localization at membrane leaflet resolution. While it is theoretically possible to address this question *in vitro* using MsbA proteoliposomes and quantifying lipid flip-flop, flipping assays are notoriously challenging and require accurate lipid labeling techniques. Further complicating this experiment, the only known LPS detection assay, based on limulus amebocyte lysate (extract from horseshoe crab blood), has not yet been successfully optimized for LPS flipping assays across liposome membranes.

Instead of a challenging *in vitro* setup, the first elucidation of the role of MsbA involves an astute *in vivo* experiment[78] exploiting an *E. coli* mutant with a constitutionally active PmrA transcription factor, conferring resistance to a class of LPS-binding compounds named polymyxins. Importantly, PmrA controls the expression of membrane-embedded enzymes that are responsible for LPS modification in the periplasmic space: ArnT and EptA (***Figure 1.5***). Assessment of LPS modification in ArnT/EptA expressing bacteria reveals that LPS is chemically modified when wild type MsbA is present, yet unaltered in cells producing A270T MsbA[78]. Since temperature induced MsbA inactivation precludes LPS modification by ArnT and EptA in the

periplasm, this experiment conclusively establishes MsbA as a lipid floppase rather than an extractor.

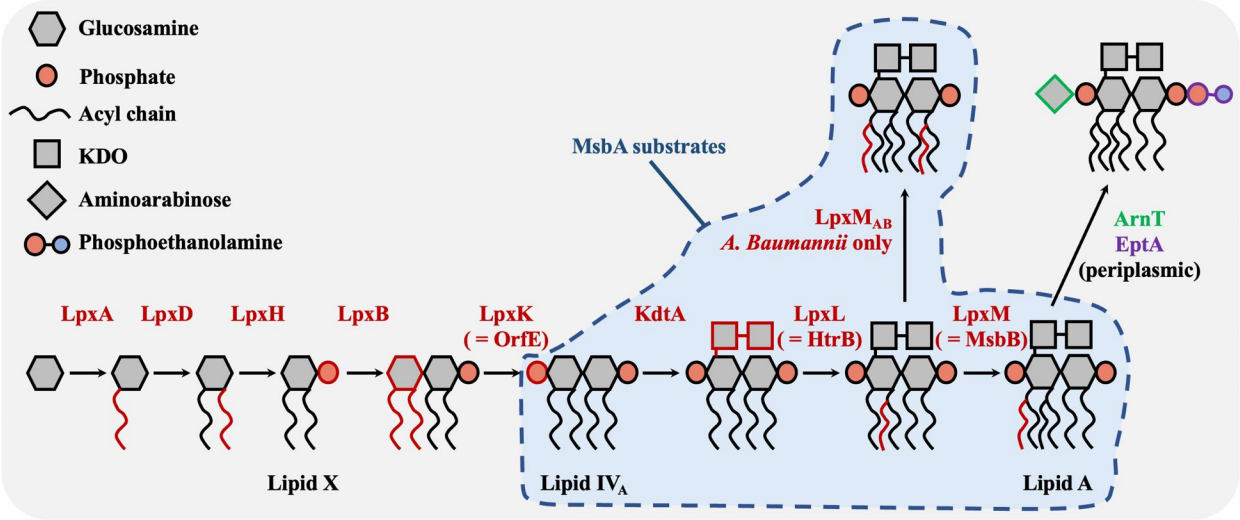


Figure 1.5: Non-exhaustive overview of LPS synthesis pathway and MsbA substrates. LPS-modifying membrane-embedded enzymes are listed in colored text (crimson if in cytoplasmic leaflet, green or purple if periplasmic leaflet). The chemical group just added in a reaction step is colored the same as the enzyme involved in the conjugation. Lipids that can be transported by MsbA are indicated, although completed lipid A is recognized and translocated with greater efficiency than incomplete intermediates such as lipid IV$_A$.

MsbA specificity for various lipids: Although MsbA has a clearly defined role in bacterial physiology, it is nevertheless a relatively promiscuous protein and is in this respect akin to other ABC transporters. Its preferred LPS substrate is hexa-acylated (in *E. coli*), yet MsbA is capable of flipping under-acylated LPS with as few as 4 acyl chains, as indicated by the HtrB rescue experiments[74-76]. In *A. baumannii*, MsbA must accommodate hepta-acylated LPS, due to a LpxM acyltransferase unique to this genus[79] (***Figure 1.5***). There is also evidence that MsbA is active across species, since *A. baumannii* MsbA can rescue *E. coli* cells in which MsbA has been disabled[80]. Thus, MsbA is relatively unspecific in terms of LPS recognition and can flip LPS with anywhere between 4 and 7 acyl chains with variable efficiency. However, different LPS intermediates have contrasting effects on the MsbA activity, such that the immature lipid X and

lipid IV_A (***Figure 1.5***) have no influence on MsbA ATPase activity whereas complete KDO-lipid A stimulates MsbA activity severalfold[81]. Therefore, substrate-induced stimulation of MsbA ensures that most LPS molecules transported to the cell surface are fully formed.

MsbA's ability to flip structurally distinct LPS intermediates raises the question of whether it can act as a general lipid floppase capable of transporting phospholipids across the IM, possibly favoring anterograde transport to the OM. This theory is supported by the observation that disabling MsbA via the A270T temperature-sensitive mutant leads to accumulation of PE in the IM, suggesting that lipid homeostasis between IM and OM is imbalanced[77]. Yet contradicting this claim, inactivation of MsbA in *N. Meningitidis* – which can survive without MsbA provided the LPS biosynthesis pathway is disabled – does not prevent proper shuttling of phospholipids to the OM[82]. Notably, complete elucidation of the role of MsbA in phospholipid flipping is complicated by the fact that disabling MsbA dramatically alters IM lipid composition and morphology[77], possibly destabilizing other phospholipid export systems such as the Mla complex.

Specificity for small molecules: MsbA's sequence and structure similarity with other ABC transporters raises important questions about its specificity for various small molecules: Is MsbA overexpression a viable mechanism for bacterial drug resistance? Does its substrate spectrum overlap with that of homologous transporters such as ABCB1? Early MsbA studies showed that MsbA overexpression failed to confer resistance to common antibiotics such as streptomycin and novobiocin[81], although different groups report contradictory results about whether it may provide erythromycin resistance[81,83]. Similarly, the ABCB1 substrates verapamil and doxorubicin are reported to have no effect on MsbA ATPase activity[81], yet separate studies disagree on whether daunomycin and vinblastine are MsbA modulators[81,83,84]. Finally, both *in vivo* and *in vitro* experiments suggest that MsbA is capable of exporting DNA intercalating dyes such as ethidium

and Hoechst 33342[83,84], illustrating its ability to recognize a diversity of molecular scaffolds. The general lack of consensus about non-LPS MsbA substrates nevertheless makes this field particularly confusing. As a result, this document focuses exclusively on the role of MsbA as an LPS transporter.

Biochemical investigation of MsbA over the past three decades has clarified its critical role in LPS selection and transport across the bacterial IM. Our understanding of MsbA function is further complemented by numerous structures of MsbA in different conformational states.

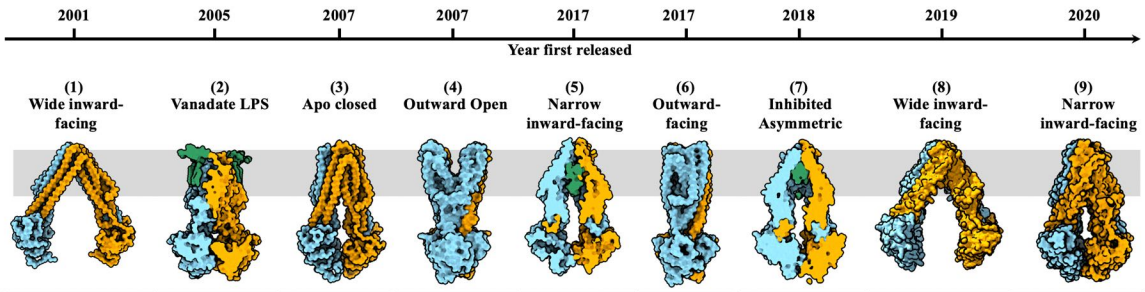
1.7 Structural characterization of MsbA over the years

The history of MsbA structural biology has been quite eventful since the early 2000s and can be broken down into different segments: 1) initial excitement with early X-Ray structures (2001 – 2005); 2) paper retractions, corrections, and MsbA structure draught (2006 – 2016); 3) field renaissance, coinciding with the cryo-EM resolution revolution (2017 – present). Some of the major events are outlined below.

The first structure of MsbA was determined by the Chang group[65], and was a breakthrough when it was published in 2001 because it was the first structure of a full-length ABC transporter. Its structure is significantly different from that of the few membrane proteins crystallized at the time and features an intriguing wide-open conformation with inter-NBD spacing of about 85 Å (***Figure 1.6***, state 1). Notably, the central MsbA pocket is gigantic (~ 850 $Å^2$ of exposed protein surface, large enough to accommodate lipid A and is lined with lysines and arginines, contrasting with the hydrophobic membrane environment. Four years later, the same group published a second MsbA structure in complex with the non-hydrolyzable nucleotide ADP-vanadate and LPS[85] (***Figure 1.6***, state 2). This structure is particularly interesting not only because the nucleotide forced an MsbA conformation that is considerably tighter than the drug-free state, but also because

LPS is bound outside the protein instead of its expected location in the central pocket. The surprising positioning of LPS as if it were embedded in the periplasmic bilayer leaflet (***Figure 1.6***, state 2) then fueled speculation that LPS flipping occurs according to the "credit-card" model[86], in which 180º lipid flipping occurs within the bilayer, with MsbA only providing shielding for the hydrophilic LPS sugar groups. While such mechanism is consistent with that of other known lipid-floppases such as TMEM16 scramblases and P-type ATPases, it doesn't explain the unusual features of the sizable MsbA central pocket. Furthermore, LPS binding outside the TMDs may simply participate in stabilizing crystal contacts[87] rather than correspond to a biologically relevant localization of LPS.

Although the early apo and nucleotide-bound MsbA structures were of great interest to the ABC transporter community, they nevertheless generated some skepticism from scientists who noted that the structure of MsbA was incompatible with previously solved structures of NBDs in isolation. Notably, the subsequent 2006 structure of the bacterial transporter Sav1866 by the Locher group[88] had inverted handedness relative to the published MsbA architecture, which prompted a series of MsbA paper retractions[89-91] once it was discovered that the authors had accidentally inverted the handedness of their electron density maps and proceeded to build right-handed helices into flipped left-handed protein density. The incorrect MsbA structures were later rebuilt and published alongside novel conformations (apo closed and outward open, ***Figure 1.6***, states 3 and 4) in a paper summarizing all the insights gained from the Chang group studies[24]. The backlash resulting from the paper retractions may have negatively impacted the advancement of MsbA structural biology, because no new MsbA structures were determined in the following decade.



Access code	3B5W	1Z2R (retracted)	3B5X	3B60/3B5Y/ 3B5Z	5TV4	5TTP	6BPL / 6BPP	6BL6 / 6O30	6UZ2 / 6UZL
Group	Chang	Chang	Chang	Chang	Liao	Liao	Genentech	Zhang	Walz
Method	X-Ray	X-Ray	X-Ray	X-Ray	Cryo-EM	Cryo-EM	X-Ray	X-Ray	Cryo-EM
Environment	DDM	UDM	DDM	UDM	Nanodisc	Nanodisc	FA-3	FA-3	Peptidisc
Species	*E. coli*	*S. Typhimurium*	*V. Cholerae*	*S. Typhimurium*	*E. coli*	*E. coli*	*E. coli*	*S. Typhimurium*	*E. coli*
LPS?	No	Yes	No	No	Yes	No	Yes	Yes	Yes

Figure 1.6: Timeline of MsbA structures to date. All major structures of full-length MsbA are displayed as surface representation with each monomer colored in either blue or orange, and with surface clipping to highlight LPS density in select structures. LPS is displayed as a dark green surface when modeled in the coordinate file. A table located below the timeline summarizes important information about each structure.

By the mid 2010s, improvements in direct-electron detectors (e.g., Gatan K2 camera[92]) and image processing algorithms (MotionCor2[93], Relion[94]) allowed for the first high-resolution cryo-EM structures of membrane proteins[92,95]. While ABC transporters are particularly challenging targets for cryo-EM studies due to their small size (usually <150 kDa) and flexibility, the early structures of TmrAB[37] and TAP[38] indicated that MsbA may be a suitable target for this new technique. In 2017, the cryo-EM structure of MsbA was determined in lipid nanodiscs[25] – a membrane-mimetic environment – and introduced several new surprising findings: 1) LPS is readily co-purified along with the transporter, revealing the first insights into the structural basis of LPS recognition; 2) *E. coli* MsbA adopts a narrow inward-facing conformation when bound to LPS instead of wide-open (***Figure 1.6***, states 5 and 9); 3) Non-hydrolyzable nucleotides cause a closed outward-facing conformation (***Figure 1.6***, state 6), instead of the previous outward-open conformation. This landmark publication was followed less than one year later by a Genentech study in which a high-throughput chemical screen resulted in identification of a molecular scaffold capable of inhibiting MsbA ATPase activity and thereby its function as an LPS transporter[96]. The

X-Ray structures of MsbA in complex with two of these compounds reveal an unprecedented inhibited MsbA state in which the NBDs surprisingly come together asymmetrically, hinting at a novel inhibition mechanism[96] (***Figure 1.6***, state 7). This structure remains the only small molecule-inhibited MsbA conformation characterized to date.

Structural basis of LPS recognition: All inward-facing MsbA structures solved in recent years (***Figure 1.6***, states 5-9) include LPS density, possibly due to the use of either nanodiscs, peptidiscs, or soft detergents such as FA-3 instead of DDM and UDM. These structures provide unprecedented insights into how MsbA selects LPS among other inner membrane lipids, as well as into the mechanism of LPS flipping (***Figure 1.7a,b***).

It is apparent from the 2017 cryo-EM structure[25] that a ring of charged amino acids including Arg78, Arg148, Gln256, Arg296 and Lys299 forms a tight belt around the LPS glucosamine phosphates and stabilizes the ligand through electrostatic contacts (***Figure 1.7c***). Consistent with this model, an R78A/R148A/K299A mutation abolishes LPS-induced stimulation of ATPase activity as well as LPS co-purification in MsbA pulldowns in mass spectrometry experiments[25], thus confirming the importance of these basic sidechains in LPS recognition. In this context, LPS may serve as a tether bringing together the two MsbA TMDs into a narrow inward-facing conformation. Interestingly, the distance between the glucosamine phosphates and the top of the hydrophobic pocket is just long enough to accommodate the 12/14 carbon length of LPS acyl chains, as opposed to a longer phospholipid carbon chain, suggesting that MsbA acts as a "hydrocarbon ruler"[96]. Thus, the MsbA central pocket is perfectly complementary to its LPS substrate.

Prior models of LPS flipping[24,65] assumed that LPS would undergo a 180° rotation inside the MsbA central pocket, before being released in the periplasmic membrane leaflet. However, the

structure of LPS-bound MsbA[25] captures a previously unrecognized functional state, in which LPS acyl chains are located at the level of the periplasmic membrane leaflet (***Figure 1.7a***). This observation led to the "trap and flip" model[25], which emphasizes that captured LPS has already accomplished much of the translocation across the IM without the upside-down rotation, and is likely followed by simultaneous release of LPS into the periplasmic leaflet and rotation due to hydrophobic preference for the membrane environment. The "trap and flip" model of translation followed by rotation has not been observed for other lipid floppases yet may be the first example of a general mechanism of lipid flipping.

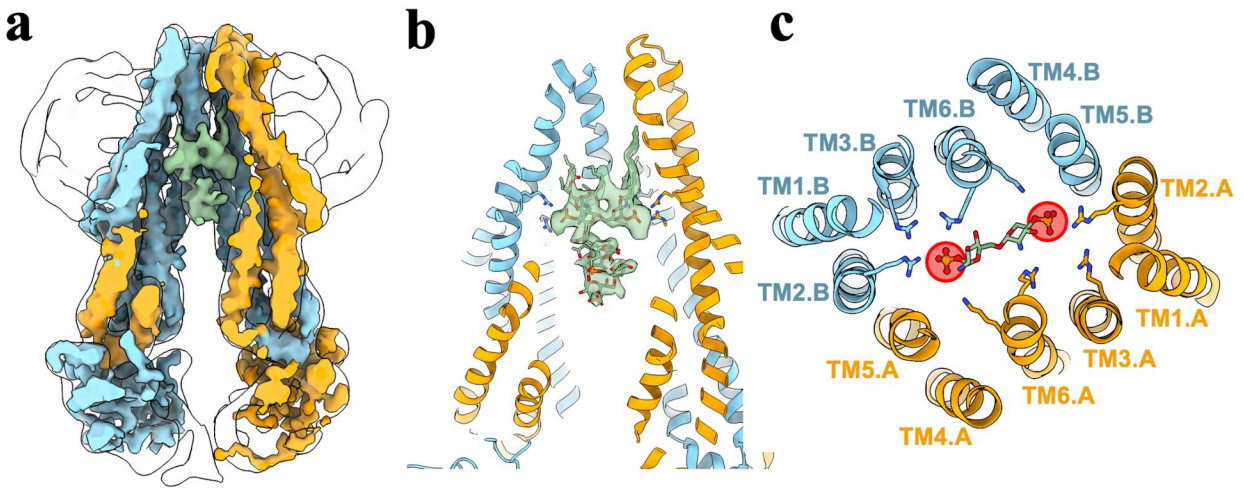


Figure 1.7: Structural basis of LPS recognition by MsbA. a, Cryo-EM structure of *E. coli* MsbA in lipid nanodiscs (EMD-8469), with LPS shown in green. The density is low-pass filtered to 8-Å and displayed as outline to highlight the nanodisc density. **b,** View of the LPS binding pocket, with MsbA model shown and LPS represented as atoms and zoned green density. **c,** Top-down view through the transmembrane domains. LPS glucosamine phosphates are highlighted as red circles, and LPS-interacting basic sidechains are shown as blue and orange.

1.8 Open questions and scope of the thesis

Although decades of basic research have considerably expanded our knowledge of how MsbA captures and translocates LPS across the inner membrane, several important questions pertaining to MsbA function and inhibition remain unexplored.

With regards to MsbA function, the protein structure is known in several conformations and the mechanism of LPS recognition is well-documented. Yet several inconsistencies cloud our understanding of how the states characterized so far fit together in a unified transport cycle: why have drug-free experiments resulted in both wide-open (***Figure 1.6***, states 1 and 8) and narrow inward-facing (***Figure 1.6***, states 5 and 9) conformations? Why have different nucleotides caused either outward open (***Figure 1.6***, state 4) or outward-facing closed states (***Figure 1.6***, state 6)? It is particularly challenging to untangle variations in sample preparation and structural technique with true physiological differences, which warrants further investigation using a consistent system.

Concerning MsbA inhibition, G compounds from Genentech Inc. are the only small molecules known to act as MsbA-specific ATPase inhibitors[96]. However, it is puzzling how binding in of G compounds in symmetrically-related pockets leads to an asymmetric conformation change, and how asymmetric closed NBDs (***Figure 1.6***, state 7) impair the MsbA conformation cycle. Furthermore, the recent identification of TBT1 as a decoupler of ATPase activity from LPS transport[80] raises the question of whether new mechanisms of ABC transporter inhibition remain to be uncovered.

Chapter 2 relates to the normal function of MsbA throughout the ATP turnover and LPS transport cycle. Single-particle cryo-EM imaging under ATP turnover conditions reveals that wide inward-open MsbA can exist in lipid bilayers, and that the ADP-bound state is the dominant closed conformation, indicating that NBD dissociation is a rate-limiting step in the LPS transport cycle.

Comparison with a high-resolution structure of MsbA in a pre-hydrolysis state shows that ATP binding is sufficient for substrate release and suggests that nucleotide γ-phosphate removal destabilizes the lower section of the NBDs.

It is intuitive that small molecule inhibitors may bind and block transporters and thereby arrest their function cycle. More intriguing are the rare compounds that instead increase the pace of a protein machine and disrupt function by overactivation. Chapter 3 reveals the unexpected mechanism by which TBT1 replaces LPS in the central pocket and rearranges the TMDs and NBDs in an occluded and hyperactive MsbA conformation. Cryo-EM structures show that TBT1 binding causes complete disordering of the MsbA domain-swapping helices, and virtual screening against the TBT1 induced-fit pocket leads to identification of novel ATPase modulators.

Chapter 4 is a study of how G compounds[96] block the MsbA conformation cycle not by inducing a closed asymmetric conformation, but by opening the protein symmetrically. Structures in detergent and nanodisc are compared with the prior drug-free structure, and 3D variability analysis and negative-stain EM data indicate that G compounds prevent NBD dimerization by increasing inter-NBD distance.

Insights gained from studying the MsbA function cycle and inhibition mechanisms are briefly summarized in Chapter 5. Future steps and key outstanding questions are discussed.

1.9 References

1. Yan, N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev Biophys* **44**, 257-83 (2015).

2. Piddock, L.J. Multidrug-resistance efflux pumps - not just for resistance. *Nat Rev Microbiol* **4**, 629-36 (2006).

3. Dyla, M., Kjaergaard, M., Poulsen, H. & Nissen, P. Structure and Mechanism of P-Type ATPase Ion Pumps. *Annu Rev Biochem* **89**, 583-603 (2020).

4. Rees, D.C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218-27 (2009).

5. Locher, K.P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487-93 (2016).

6. Thomas, C. & Tampe, R. Structural and Mechanistic Principles of ABC Transporters. *Annu Rev Biochem* **89**, 605-636 (2020).

7. Srikant, S. Evolutionary history of ATP-binding cassette proteins. *FEBS Lett* **594**, 3882-3897 (2020).

8. Davidson, A.L. & Chen, J. ATP-binding cassette transporters in bacteria. *Annu Rev Biochem* **73**, 241-68 (2004).

9. Hwang, J.U. et al. Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle. *Mol Plant* **9**, 338-355 (2016).

10. Plummer, A.M., Culbertson, A.T. & Liao, M. The ABCs of Sterol Transport. *Annu Rev Physiol* (2020).

11. Lubelski, J., Konings, W.N. & Driessen, A.J. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol Mol Biol Rev* **71**, 463-76 (2007).

12. Orelle, C., Mathieu, K. & Jault, J.M. Multidrug ABC transporters in bacteria. *Res Microbiol* **170**, 381-391 (2019).

13. Robey, R.W. et al. Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat Rev Cancer* **18**, 452-464 (2018).

14. Aller, S.G. et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **323**, 1718-22 (2009).

15. Li, Y., Orlando, B.J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. *Nature* **567**, 486-490 (2019).

16. Oldham, M.L., Khare, D., Quiocho, F.A., Davidson, A.L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515-21 (2007).

17. Li, N. et al. Structure of a Pancreatic ATP-Sensitive Potassium Channel. *Cell* **168**, 101-110 e10 (2017).

18. Lin, D.Y., Huang, S. & Chen, J. Crystal structures of a polypeptide processing and secretion transporter. *Nature* **523**, 425-30 (2015).

19. Srikant, S. & Gaudet, R. Mechanics and pharmacology of substrate selection and transport by eukaryotic ABC exporters. *Nat Struct Mol Biol* **26**, 792-801 (2019).

20. Navarro-Quiles, C., Mateo-Bonmati, E. & Micol, J.L. ABCE Proteins: From Molecules to Development. *Front Plant Sci* **9**, 1125 (2018).

21. Murina, V. et al. ABCF ATPases Involved in Protein Synthesis, Ribosome Assembly and Antibiotic Resistance: Structural and Functional Diversification across the Tree of Life. *J Mol Biol* **431**, 3568-3590 (2019).

22. Oldham, M.L. & Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc Natl Acad Sci U S A* **108**, 15152-6 (2011).

23. Johnson, Z.L. & Chen, J. ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. *Cell* **172**, 81-89 e10 (2018).

24. Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-10 (2007).

25. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

26. Johnson, Z.L. & Chen, J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. *Cell* **168**, 1075-1085 e9 (2017).

27. Hofmann, S. et al. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580-583 (2019).

28. Wang, L. et al. Cryo-EM structure of human bile salts exporter ABCB11. *Cell Res* **30**, 623-625 (2020).

29. Taylor, N.M.I. et al. Structure of the human multidrug transporter ABCG2. *Nature* **546**, 504-509 (2017).

30. Olsen, J.A., Alam, A., Kowal, J., Stieger, B. & Locher, K.P. Structure of the human lipid exporter ABCB4 in a lipid environment. *Nat Struct Mol Biol* **27**, 62-70 (2020).

31. Carrier, I., Julien, M. & Gros, P. Analysis of catalytic carboxylate mutants E552Q and E1197Q suggests asymmetric ATP hydrolysis by the two nucleotide-binding domains of P-glycoprotein. *Biochemistry* **42**, 12875-85 (2003).

32. Verhalen, B. et al. Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature* **543**, 738-741 (2017).

33. Thomas, C. et al. Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett* **594**, 3767-3775 (2020).

34. Wang, B., Dukarevich, M., Sun, E.I., Yen, M.R. & Saier, M.H., Jr. Membrane porters of ATP-binding cassette transport systems are polyphyletic. *J Membr Biol* **231**, 1-10 (2009).

35. Borst, P., Evers, R., Kool, M. & Wijnholds, J. A family of drug transporters: the multidrug resistance-associated proteins. *J Natl Cancer Inst* **92**, 1295-302 (2000).

36. Lee, J.Y. et al. Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **533**, 561-4 (2016).

37. Kim, J. et al. Subnanometre-resolution electron cryomicroscopy structure of a heterodimeric ABC exporter. *Nature* **517**, 396-400 (2015).

38. Oldham, M.L. et al. A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. *Nature* **529**, 537-40 (2016).

39. Januliene, D. & Moeller, A. Cryo-EM of ABC transporters: an ice-cold solution to everything? *FEBS Lett* **594**, 3776-3789 (2020).

40. ter Beek, J., Guskov, A. & Slotboom, D.J. Structural diversity of ABC transporters. *J Gen Physiol* **143**, 419-35 (2014).

41. Rempel, S., Stanek, W.K. & Slotboom, D.J. ECF-Type ATP-Binding Cassette Transporters. *Annu Rev Biochem* **88**, 551-576 (2019).

42. Wang, Z., Hu, W. & Zheng, H. Pathogenic siderophore ABC importer YbtPQ adopts a surprising fold of exporter. *Sci Adv* **6**, eaay7997 (2020).

43. Perez, C. et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433-8 (2015).

44. Zhang, Z. & Chen, J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. *Cell* **167**, 1586-1597 e9 (2016).

45. Bi, Y., Mann, E., Whitfield, C. & Zimmer, J. Architecture of a channel-forming O-antigen polysaccharide ABC transporter. *Nature* **553**, 361-365 (2018).

46. Okada, U. et al. Crystal structure of tripartite-type ABC transporter MacB from Acinetobacter baumannii. *Nat Commun* **8**, 1336 (2017).

47. Tang, X. et al. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nat Struct Mol Biol* **28**, 347-355 (2021).

48. Yu, Q. et al. Structures of ABCG2 under turnover conditions reveal a key step in drug transport mechanism. *bioRxiv*, 2021.03.03.433600 (2021).

49. Hutter, C.A.J. et al. The extracellular gate shapes the energy profile of an ABC exporter. *Nat Commun* **10**, 2260 (2019).

50. Wang, L. et al. Characterization of the kinetic cycle of an ABC transporter by single-molecule and cryo-EM analyses. *Elife* **9**(2020).

51. Vollmer, W. & Seligman, S.J. Architecture of peptidoglycan: more data and more models. *Trends Microbiol* **18**, 59-66 (2010).

52. Henderson, J.C. et al. The Power of Asymmetry: Architecture and Assembly of the Gram-Negative Outer Membrane Lipid Bilayer. *Annu Rev Microbiol* **70**, 255-78 (2016).

53. Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol Rev* **40**, 133-59 (2016).

54. Renner, L.D. & Weibel, D.B. Cardiolipin microdomains localize to negatively curved regions of Escherichia coli membranes. *Proc Natl Acad Sci U S A* **108**, 6264-9 (2011).

55. Nakayama, H., Kurokawa, K. & Lee, B.L. Lipoproteins in bacteria: structures and biosynthetic pathways. *FEBS J* **279**, 4247-68 (2012).

56. Chi, X. et al. Structural mechanism of phospholipids translocation by MlaFEDB complex. *Cell Res* **30**, 1127-1135 (2020).

57. Coudray, N. et al. Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *Elife* **9**(2020).

58. Isom, G.L. et al. LetB Structure Reveals a Tunnel for Lipid Transport across the Bacterial Envelope. *Cell* **181**, 653-664 e19 (2020).

59. Sharma, S. et al. Mechanism of LolCDE as a molecular extruder of bacterial triacylated lipoproteins. *bioRxiv*, 2021.04.06.438740 (2021).

60. Nickerson, N.N. et al. A Novel Inhibitor of the LolCDE ABC Transporter Essential for Lipoprotein Trafficking in Gram-Negative Bacteria. *Antimicrob Agents Chemother* **62**(2018).

61. Raetz, C.R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635-700 (2002).

62. Bertani, B. & Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **8**(2018).

63. Sperandeo, P., Martorana, A.M. & Polissi, A. Lipopolysaccharide biogenesis and transport at the outer membrane of Gram-negative bacteria. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1451-1460 (2017).

64. Thelot, F., Orlando, B.J., Li, Y. & Liao, M. High-resolution views of lipopolysaccharide translocation driven by ABC transporters MsbA and LptB2FGC. *Curr Opin Struct Biol* **63**, 26-33 (2020).

65. Chang, G. & Roth, C.B. Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**, 1793-800 (2001).

66. Cuthbertson, L., Powers, J. & Whitfield, C. The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a. *J Biol Chem* **280**, 30310-9 (2005).

67. Caffalette, C.A., Corey, R.A., Sansom, M.S.P., Stansfeld, P.J. & Zimmer, J. A lipid gating mechanism for the channel-forming O antigen ABC transporter. *Nat Commun* **10**, 824 (2019).

68. Caffalette, C.A. & Zimmer, J. Cryo-EM structure of the full-length WzmWzt ABC transporter required for lipid-linked O antigen transport. *Proc Natl Acad Sci U S A* **118**(2021).

69. Okuda, S., Sherman, D.J., Silhavy, T.J., Ruiz, N. & Kahne, D. Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nat Rev Microbiol* **14**, 337-45 (2016).

70. Sherman, D.J. et al. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science* **359**, 798-801 (2018).

71. Luo, Q. et al. Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nat Struct Mol Biol* **24**, 469-474 (2017).

72. Owens, T.W. et al. Structural basis of unidirectional export of lipopolysaccharide to the cell surface. *Nature* **567**, 550-553 (2019).

73. Tang, X. et al. Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism. *Nat Commun* **10**, 4175 (2019).

74. Karow, M. & Georgopoulos, C. The essential Escherichia coli msbA gene, a multicopy suppressor of null mutations in the htrB gene, is related to the universally conserved family of ATP-dependent translocators. *Mol Microbiol* **7**, 69-79 (1993).

75. Polissi, A. & Georgopoulos, C. Mutational analysis and properties of the msbA gene of Escherichia coli, coding for an essential ABC family transporter. *Mol Microbiol* **20**, 1221-33 (1996).

76. Zhou, Z., White, K.A., Polissi, A., Georgopoulos, C. & Raetz, C.R. Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J Biol Chem* **273**, 12466-75 (1998).

77. Doerrler, W.T., Reedy, M.C. & Raetz, C.R. An Escherichia coli mutant defective in lipid export. *J Biol Chem* **276**, 11461-4 (2001).

78. Doerrler, W.T., Gibbons, H.S. & Raetz, C.R. MsbA-dependent translocation of lipids across the inner membrane of Escherichia coli. *J Biol Chem* **279**, 45102-9 (2004).

79. Boll, J.M. et al. Reinforcing Lipid A Acylation on the Cell Surface of Acinetobacter baumannii Promotes Cationic Antimicrobial Peptide Resistance and Desiccation Survival. *mBio* **6**, e00478-15 (2015).

80. Zhang, G. et al. Cell-based screen for discovering lipopolysaccharide biogenesis inhibitors. *Proc Natl Acad Sci U S A* **115**, 6834-6839 (2018).

81. Doerrler, W.T. & Raetz, C.R. ATPase activity of the MsbA lipid flippase of Escherichia coli. *J Biol Chem* **277**, 36697-705 (2002).

82. Tefsen, B., Bos, M.P., Beckers, F., Tommassen, J. & de Cock, H. MsbA is not required for phospholipid transport in Neisseria meningitidis. *J Biol Chem* **280**, 35961-6 (2005).

83. Woebking, B. et al. Drug-lipid A interactions on the Escherichia coli ABC transporter MsbA. *J Bacteriol* **187**, 6363-9 (2005).

84. Reuter, G. et al. The ATP binding cassette multidrug transporter LmrA and lipid transporter MsbA have overlapping substrate specificities. *J Biol Chem* **278**, 35193-8 (2003).

85. Reyes, C.L. & Chang, G. Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science* **308**, 1028-31 (2005).

86. Ballesteros, A. & Swartz, K.J. Lipids surf the groove in scramblases. *Proc Natl Acad Sci U S A* **115**, 7648-7650 (2018).

87. Reyes, C.L. & Chang, G. Lipopolysaccharide stabilizes the crystal packing of the ABC transporter MsbA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 655-8 (2005).

88. Dawson, R.J. & Locher, K.P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180-5 (2006).

89. Miller, G. Scientific publishing. A scientist's nightmare: software problem leads to five retractions. *Science* **314**, 1856-7 (2006).

90. Matthews, B.W. Five retracted structure reports: inverted or incorrect? *Protein Sci* **16**, 1013-6 (2007).

91. Jeffrey, P.D. Analysis of errors in the structure determination of MsbA. *Acta Crystallogr D Biol Crystallogr* **65**, 193-9 (2009).

92. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-12 (2013).

93. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

94. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).

95. Bai, X.C. et al. An atomic structure of human gamma-secretase. *Nature* **525**, 212-217 (2015).

96. Ho, H. et al. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196-201 (2018).

Chapter 2: Dissection of the MsbA function cycle by single-particle cryo-EM

The data presented here is unpublished.

The methods section of this chapter contains excerpts from a methods paper:

Thelot, F. & Liao, M. (2022), Cryo-EM analysis of the lipopolysaccharide flippase MsbA. Lipopolysaccharide transport: Methods and Protocols, *Methods in Molecular Biology* (in press).

Contribution: F. Thelot performed all experiments listed in this chapter, under supervision from M. Liao.

2.1 Introduction

Cryo-EM imaging followed by 3D classification[1,2] allows for characterization of distinct protein conformations within a dataset and provides information about particle populations throughout the function cycle of dynamic proteins. ABC transporters are particularly well-suited to such analysis because they undergo large-scale conformational changes while hydrolyzing ATP, yet the study of these states has so far been restricted to the realm of spectroscopy and intractable by crystallography. Several studies have aimed at characterizing ABC transporter conformations in presence of ATP, including a 2015 negative-stain analysis of MsbA and ABCB1[3], as well as more recent and higher-resolution work on TmrAB[4] and ABCC1[5]. These pioneering projects led to the identification of predominant functional states and demonstrate the power of cryo-EM to investigate ABC transporter dynamics.

ABC transporters are a diverse protein family with significant differences in transmembrane domain folds, ATPase sites and transported substrates. For example, MsbA differs from previously studied transporters because it is a homodimer with two identical functional ATPase sites, whereas TmrAB and ABCC1 include a degenerate ATP-binding pocket. Furthermore, no TmrAB or ABCC1 structure matches the degree of NBD separation of wide-open MsbA as characterized in crystallographic studies[6,7], hinting at possible divergence in transport cycle. Thus, imaging MsbA while it is continuously hydrolyzing ATP would provide new insights into its unique conformational dynamics.

Structural investigation of MsbA over the past two decades has resulted in inconsistencies that blur our understanding of how conformations fit together in a unified model of LPS transport. In apo state, inter-NBD distance has been observed to be 85Å[6], 75Å[7], or 35Å[6,8,9] depending on protein environment (detergent, nanodisc or peptidisc) and technique. A paradigm is emerging that

the wide-open MsbA apo state is a non-physiological artifact of detergent environments[8-10], in support of a model where MsbA adopts a narrow inward-facing conformation in lipid bilayers. Opposing these results, spectroscopic[11,12] and SAXS[13] studies have suggested inward-facing MsbA is remarkably flexible, adopting a continuum of conformations ranging from 40Å to 90Å inter-NBD separation. Whether MsbA adopts discrete or continuous conformations in a lipid bilayer, and the functional role of these states remains to be determined. In the nucleotide-bound state, MsbA has been captured in both outward-open state using ATP analogue AMP-PNP[6], and in a fully closed conformation with ADP-vanadate[8]. While AMP-PNP and ADP-vanadate represent different intermediates of the ATP hydrolysis reaction[14], it also seems possible that the observed conformational differences may arise from divergence in species (*E. coli* vs. *S. Typhimurium*), experimental conditions (b-UDM vs. nanodisc) or to technique (crystallography vs. cryo-EM). Systematic study of MsbA under ATP turnover condition, using a consistent lipid bilayer environment and a single technique may help clarify the relationship between known MsbA states.

2.2 Results

2.2.1 MsbA undergoes dramatic conformational changes while hydrolyzing ATP

E. coli MsbA was purified and reconstituted in palmitoyl-oleoyl-phosphatidylglycerol (POPG) nanodiscs, then exposed to 10 mM ATP/Mg^{2+} at room temperature and vitrified within 10 seconds, such that less than 10% of nucleotides are expected to have been hydrolyzed to ADP upon freezing. Cryo-EM imaging and 2D classification (***Figure 2.1a,b***) reveal 3 major MsbA conformations with varying inter-NBD spacing: wide-open, narrow inward-facing (or "apo-like") and closed. Extensive 3D classification with separate initial models (***Figure 2.1c***) allowed for cryo-EM

reconstructions of the wide-open, apo-like and closed states at 6.8 Å, 5.6 Å and 3.9 Å, respectively. To quantify relative particle populations in each state, particles are arbitrarily referred to as "clean" particles after the final round of global angle search 3D classification is complete. According to this criterion, wide-open, apo-like and closed states include ~35,000, 78,000 and 72,000 particles, respectively (***Figure 2.1c,*** red asterisks).

The wide-open state accounts for ~19% of clean particles (***Figure 2.2a***) and resembles the conformation observed in prior crystallographic studies of apo-state MsbA[6,7]. Existence of this conformation in a lipid bilayer contradicts the hypothesis that the wide-open state is a detergent-induced artifact[8-10], and instead suggests this conformation may play a functional role in LPS transport. Furthermore, the fact that the wide-open state is observed in this ATP turnover experiment but not in prior apo datasets[8] indicates a possible relationship between LPS removal and protein opening. Accordingly, no LPS-like density is present in the TMD (***Figure 2.2b***) consistent with the notion that most LPS substrate was successfully ejected by ATPase cycling.

About 42% of particles adopt an apo-like conformation (***Figure 2.2a***), which is similar to previously characterized apo MsbA in nanodisc[8] and peptidisc[9] including conformational breathing in the NBDs. Two weak densities resembling LPS glucosamine phosphates are visible in the central TMD pocket (***Figure 2.2b***), suggesting that the apo-like particles may contain a mixture of bound and unbound LPS. Comparison of the LPS-free wide-open state and LPS-containing apo-like conformation raises the possibility that the wide-open state corresponds to conformation prior LPS capture, and that substrate binding triggers NBD tightening, similar to how ABCC1 undergoes a conformational change upon LTC4[15] capture.

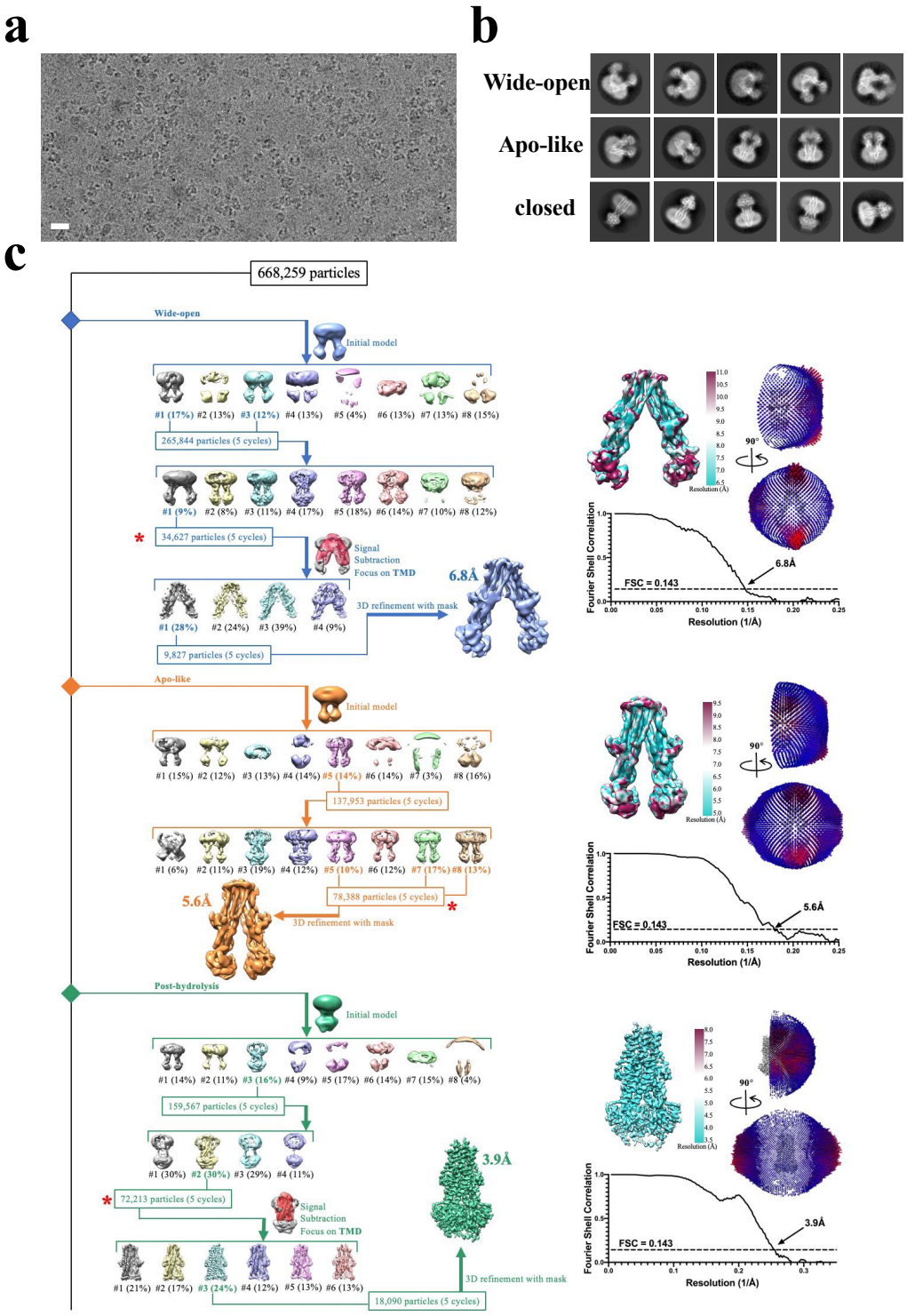


Figure 2.1: Cryo-EM imaging of MsbA under ATP turnover conditions. a, Sample micrograph of MsbA with excess ATP. The scale bar represents 20 nm. **b,** 2D classification reveal three major protein conformations: wide-open, apo-like and closed. Representative averages for each conformation are shown. Box size is 224 Å. **c,** Processing flowchart (left) and validation for each major MsbA conformation. In the processing flowchart, 3 separate initial models were used to

Figure 2.1 (continued)

recover as many particles as possible for each state. Red star symbols refer to "clean" particles, which are the remaining particles after the final round of global angle search 3D classification. "5cyc" refers to the method of selecting any particle being assigned to a class within the final 5 cycles of 3D classification. Structure validation, including ResMap[16] local resolution, angle distribution and gold standard FSC (0.143 cutoff, from independent half maps) are displayed on the right-hand side.

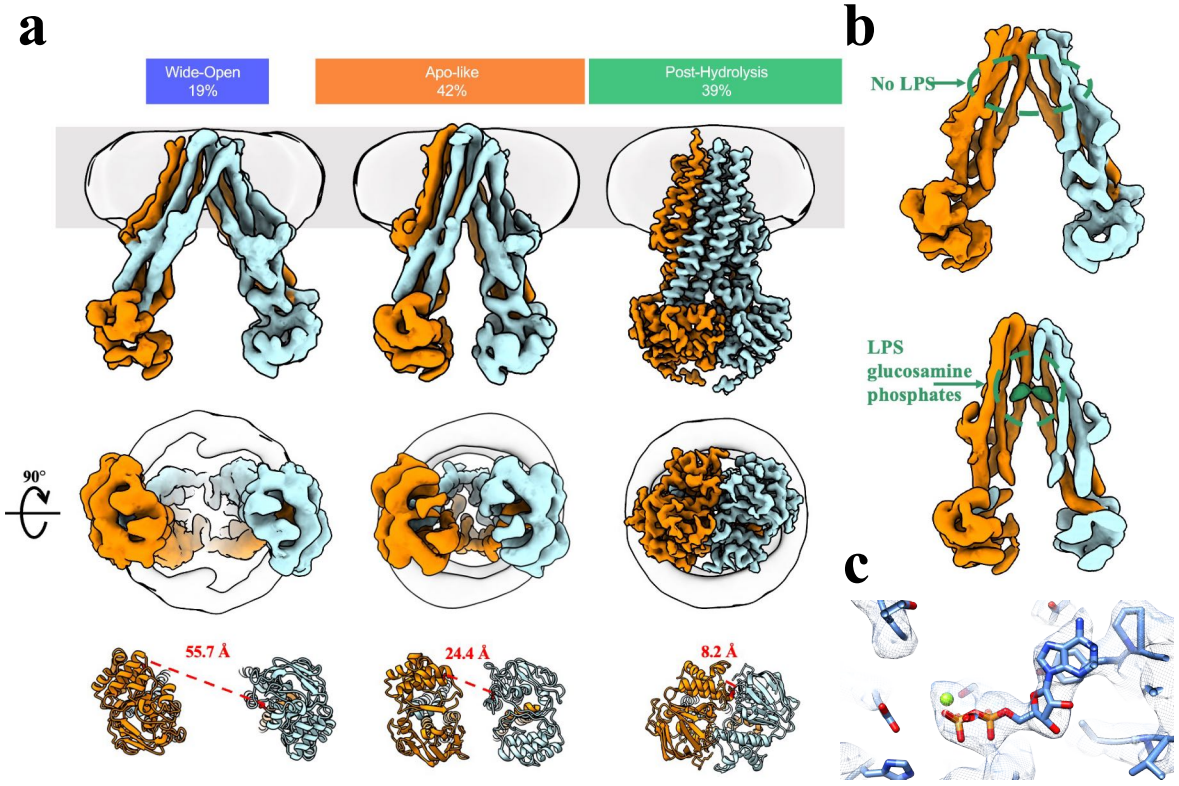


Figure 2.2: MsbA adopts 3 distinct conformations during continuous ATP hydrolysis. a, Cryo-EM densities for the 3 major conformations. Each protein chain is colored either blue or orange, and the nanodisc density is shown as transparent gray. Particle populations are indicated above each state and refer to "clean" particles as defined in Figure 2.1. Inter-NBD distances (red dashed lines) are measured between signature motif S378 and Walker A S482 at alpha carbons. **b,** Examination of substrate densities in wide-open and apo-like MsbA conformations. Only the apo-like, narrow inward-facing state seems to accommodate LPS. **c,** The closed conformation includes a nucleotide density, which is most consistent with ADP. Cryo-EM density and unprocessed, non-zoned cryo-EM map are shown. Since ADP is bound, the closed conformation corresponds to a post-hydrolysis state.

As in TmrAB ATP turnover[4], no outward-open conformation resembling the AMP-PNP crystal structure[6] is observed in our dataset, indicating that the outward-open state, if it exists at all, is short-lived. Instead, the remaining ~39% of particles adopt a closed conformation with collapsed transmembrane helices, similar to the previously characterized 5Å ADP-vanadate structure[8] (***Figure 2.2a***). Nucleotide density consistent with ADP is clearly identifiable at the inter-NBD interface (***Figure 2.2c***), although surrounding sidechains including Q-loop Q424 and H-loop H537 appear completely disordered. The observation that ADP, not ATP, is bound in the NBDs is consistent with prior experiments with TmrAB[4] and ABCC1[5] ATP turnover, and suggests that ATP hydrolysis occurs rapidly while NBD dissociation is a rate-limiting step of the ABC transporter conformation cycle. Thus, this closed conformation from continuous ATP turnover corresponds to the predominant post-hydrolysis state.

Interestingly, the ADP-bound closed conformation described here differs from a prior experiment in which MsbA was exposed to ADP yet adopted an inward-facing conformation[8]. This discrepancy underscores that nucleotide γ-phosphate is required for NBD dimerization and demonstrates that the post-hydrolysis conformation cannot originate from a backwards reaction of inward-facing MsbA binding to ADP.

The closed MsbA structure in complex with ADP (post-hydrolysis in ATP turnover) adopts a similar conformation to that of MsbA bound to ADP-vanadate (hydrolysis intermediate state[14]). Yet, this closed state differs from the outward-open conformation observed when AMP-PNP is used as a nucleotide (pre-hydrolysis state) in crystallography experiments[6]. This observation has led to the hypothesis that ATP hydrolysis causes a collapse of the transmembrane helices into the closed MsbA conformation[8]. As it is challenging to determine whether this discrepancy originates

from differences in experimental settings or represents distinct functional states, a structure of MsbA in nanodiscs in complex with AMP-PNP would provide further insight into MsbA function.

2.2.2 MsbA adopts a closed conformation with collapsed transmembrane helices in the pre-hydrolysis state

MsbA was purified and incorporated into nanodiscs using an identical protocol as in the previous section, incubated with 2 mM of AMP-PNP/Mg^{2+} on ice for ~2 hours, then vitrified. Cryo-EM imaging and processing resulted in a ~3.3 Å reconstruction of MsbA in the closed conformation, with identical transmembrane helices positioning to the post-hydrolysis state (***Figure 2.3a-d***). Notably, the AMP-PNP molecule includes extra density corresponding to γ-phosphate, and the surrounding sidechains are better ordered than in the cryo-EM map of closed MsbA from ATP turnover (***Figure 2.3e***).

While the AMP-PNP conformation shows similarities with the ADP-vanadate[8] and post-hydrolysis states, it formally disproves the prior hypothesis that ATP hydrolysis is required to provoke a collapse of the transmembrane helices from outward-facing to closed conformation[8]. Thus, nucleotide binding is sufficient for both substrate release and induction of the closed state.

In TmrAB, addition of a non-hydrolyzable nucleotide generates an equilibrium of outward-open and closed conformations[4]. Yet despite extensive classification, the outward-open state seems absent from our MsbA dataset. It is possible that a longer incubation with nucleotide (2 hours on ice vs. 5 min at 45ºC in the TmrAB study) resulted in depletion of the outward-open state. Alternatively, the outward-open conformation may be more transient for MsbA than for TmrAB.

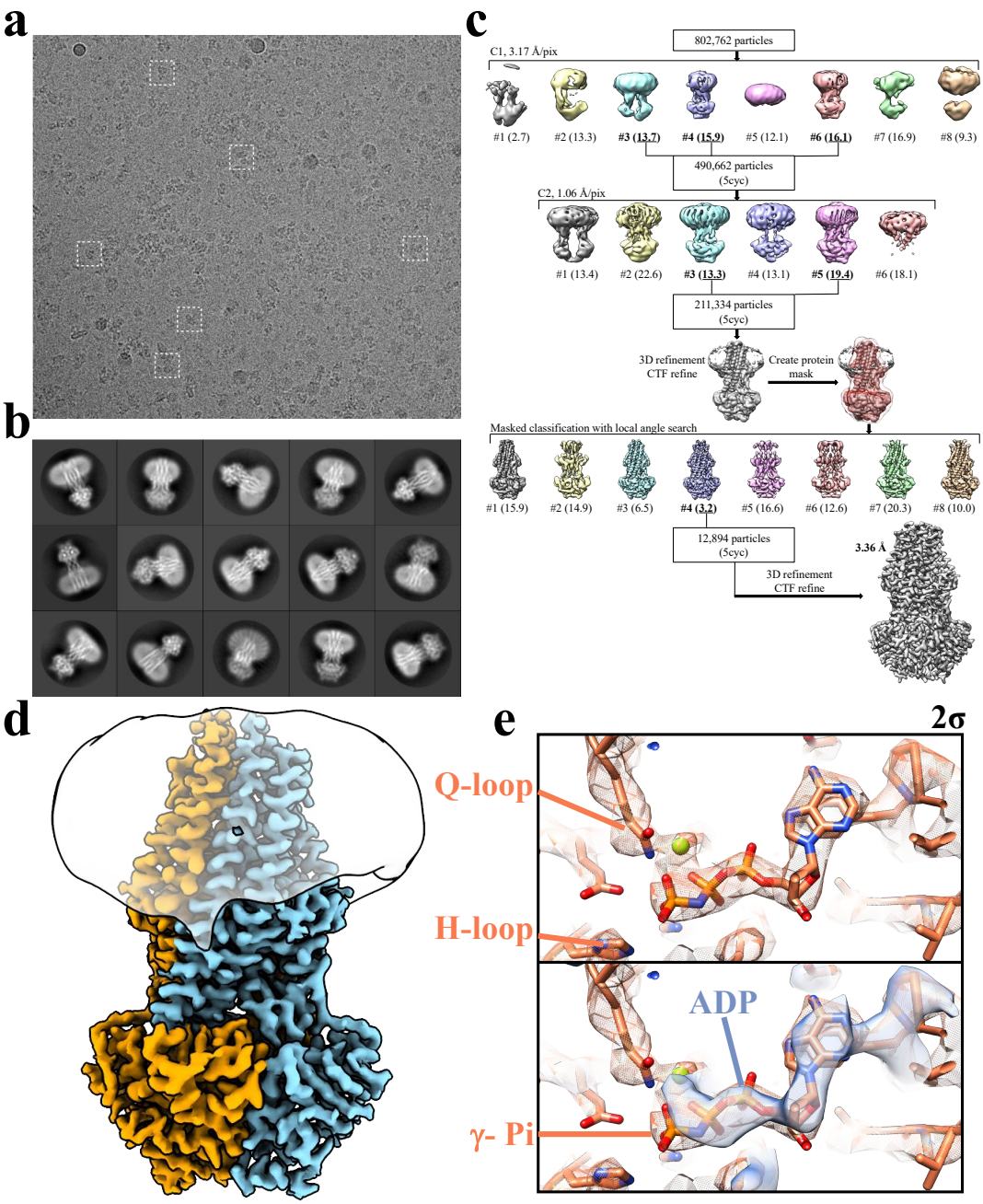


Figure 2.3: Cryo-EM imaging of MsbA in complex with AMP-PNP. a, sample micrograph of MsbA with non-hydrolyzable nucleotide AMP-PNP. Some particles are highlighted with a dashed box. **b,** Class averages for MsbA with AMP-PNP. Box size is 203 Å. **c,** Processing flowchart for the AMP-PNP dataset. "5cyc" refers to the method of selecting any particle being assigned to a class within the final 5 cycles of 3D classification. **d,** Cryo-EM reconstruction of MsbA bound to AMP-PNP, and corresponding to a pre-hydrolysis state. Each protein chain is colored either blue or orange, and the nanodisc density is shown as transparent gray. **e,** Density corresponding to AMP-PNP (top, orange), and comparison with ADP density from the ATP turnover experiment (bottom, blue). Both maps are shown at 2σ contour level. Notably, the Q-loop glutamine (Q424) is well-ordered in the AMP-PNP map yet disordered in the ATP turnover dataset.

2.2.3 Comparison of the pre- and post-hydrolysis MsbA conformations

The AMP-PNP and post-hydrolysis MsbA conformations appear largely identical. However, a finer examination reveals subtle differences, which seem to provide insights into the mechanism of NBD dissociation. Notably, transmembrane helices are of comparable quality when displayed at identical contour levels in both the AMP-PNP and post-hydrolysis reconstructions, yet NBD features are much more poorly defined in the ADP-bound map (***Figure 2.4a***).

In the post-hydrolysis ADP-bound map, residues that are in immediate proximity to the nucleotide, such as the Q-loop Q424 and H-loop H537 appear completely disordered, suggesting increased conformational heterogeneity. Relative NBD destabilization in the post-hydrolysis state is supported by several observations: 1 – Density map RMSD is similar in the TMDs (~2.3 RMSD) and NBDs (~2.2) of the AMP-PNP reconstruction, yet significantly different for the post-hydrolysis state, with TMDs (~1.9) significantly better ordered than the NBDs (~1.45) (***Figure 2.4b***). Furthermore, nucleotide-interacting motifs such as Q-loop and H-loop are more ordered than NBD average (~2.6 RMSD) in the AMP-PNP map, yet worse than average NBD density in the post-hydrolysis reconstruction (~0.8 RMSD). 2 – Model b-factor, which reflect modeling uncertainty during real space refinement, is largest in the extracellular loops of the AMP-PNP model and in the NBDs of the post-hydrolysis model (***Figure 2.4c***). Thus, different protein domains are destabilized in the pre- and post-hydrolysis state. 3 – Modeling discrepancy, as measured by backbone alpha carbon RMSD between AMP-PNP and post-hydrolysis models, is greatest in the NBDs (***Figure 2.4d***).

Although the NBDs appear more disordered in the post-hydrolysis state, NBD positioning is nearly identical, with only a ~0.5-1.0 Å spreading around the ADP molecule (***Figure 2.4e***). This relaxation of the NBDs is reminiscent of the loosening observed when comparing closed TmrAB

with its post-hydrolysis state[4], although the change described for MsbA is less significant and could conceivably be attributed to modeling error rather than relevant conformational change. Finally, prior structural analysis and molecular dynamics simulation in TmrAB resulted in the hypothesis that a pathway links the nucleotide pocket to the solvent, allowing for γ-phosphate release after hydrolysis[4]. In the MsbA AMP-PNP and post-hydrolysis models, the exit tunnel is partly gated by the H-loop histidine and appears closed in both states (***Figure 2.4f***). It is possible that increased flexibility in the NBDs would facilitate phosphate release through this exit pathway.

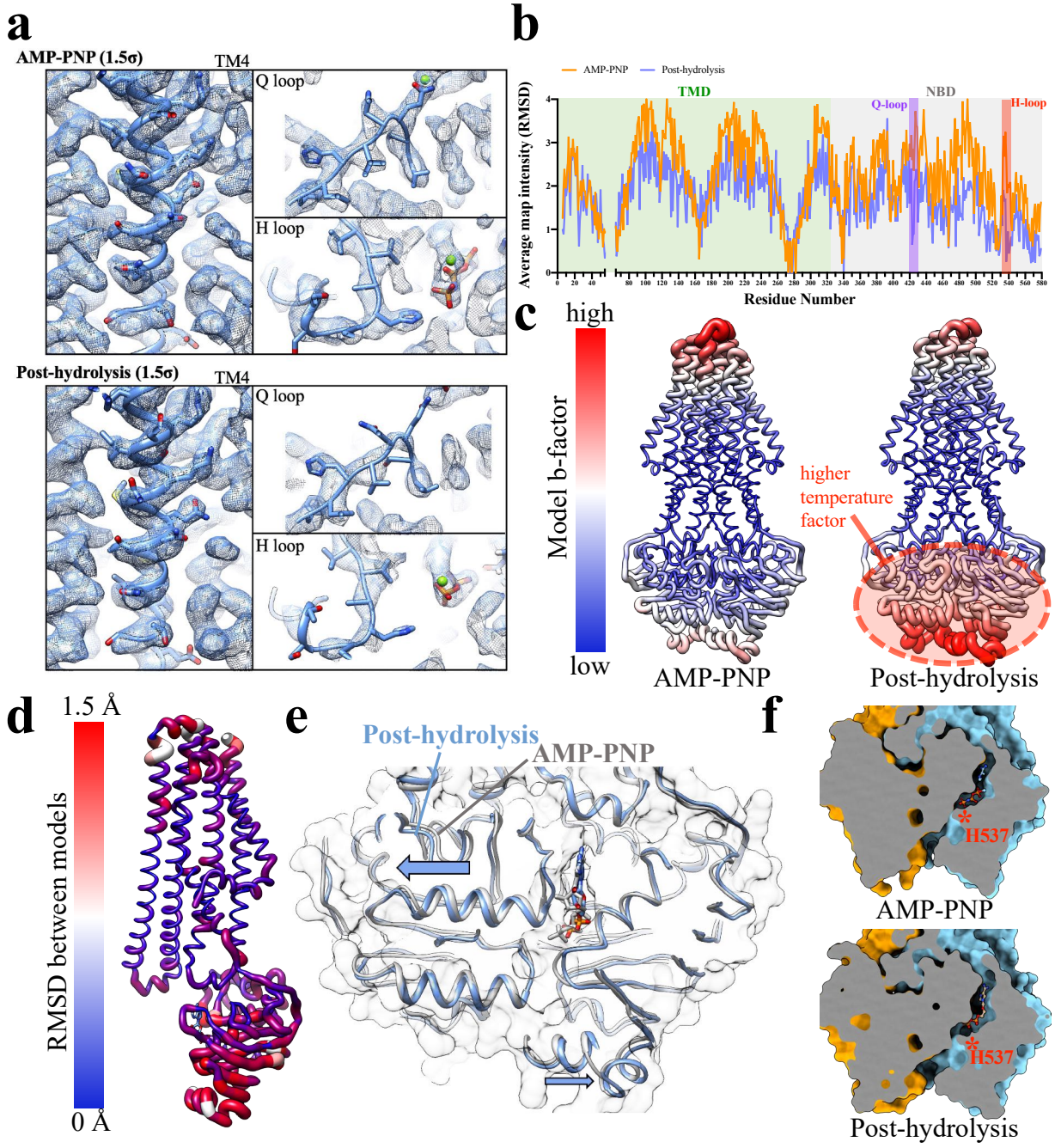


Figure 2.4: Comparison of closed MsbA conformations. a, Side-by-side comparison of AMP-PNP and post-hydrolysis (ATP turnover) map, displayed at the same 1.5σ contour level. While TMDs are well-ordered in both maps, NBD density is considerably weaker in the post-hydrolysis state. **b,** Average map intensity for each residue. NBD density is weaker for the post-hydrolysis than for AMP-PNP, while intensity is more comparable in the TMDs. Regions in contact with the nucleotide, including H-loop and Q-loop, are well-ordered in the AMP-PNP map yet particularly poor in the ADP-bound map. **c,** Model b-factor for each structure, plotted with worm representation (higher radius is higher b-factor). The scale and coloring are calculated independently for each structure. The post-hydrolysis state has comparatively higher b-factor

Figure 2.4 (continued)

relative to the rest of the protein, compared with the AMP-PNP structure. **d,** Mapping of RMSD difference between AMP-PNP and post-hydrolysis protein models, measured at alpha carbon for each residue. The models most differ in the NBDs, suggesting greater conformational differences in this region. **e,** Small-scale rearrangements in the NBDs in the post-hydrolysis state. NBDs seem slightly more spaced-out around the nucleotide in the ADP-bound map. **f,** Cross-section through NBD surface representation, showing possible phosphate release pathway speculated in previous studies[4]. The putative release pathway seems closed in each reconstruction. The positioning of the gating H-loop H537 is shown in red.

2.3 Discussion

Analysis of MsbA under ATP turnover conditions and in complex with AMP-PNP provides new information about the MsbA conformation cycle (***Figure 2.5***).

It is surprising that the wide-open state is readily accessible in nanodisc and accounts for ~20% of particles, as its existence in lipid bilayers had been questioned by the community[8-10]. In contrast with the narrow inward-facing conformation, the wide-open state is lacking density for LPS, suggesting that substrate capture may promote conformational change to the apo-like conformation. Notably, this mechanism of substrate-induced tightening is opposite to TmrAB, in which the substrate can only bind the wide inward-facing conformation[4].

Characterization of the outward-open state resisted all our efforts using both ATP and AMP-PNP, suggesting that this conformation is highly transient and may exist only long-enough to allow for LPS exit. It is unclear why the outward-open conformation is observable for TmrAB when using a non-hydrolyzable nucleotide, although its characterization may be facilitated by the innate stability of this transporter from a thermophilic organism. It remains to be determined whether such degree of outward-opening normally occurs during the ATP turnover cycle[17].

About 40% of particles imaged in presence of excess ATP adopted a closed, post-hydrolysis conformation. The observation that the post-hydrolysis state prevails over ATP-bound

is in line with prior work on TmrAB and ABCC1, which both include one ADP molecule in their functional ATPase site. Similarly, NBD dissociation is likely a rate-limiting in the LPS transport cycle. Comparison of the post-hydrolysis reconstruction with the AMP-PNP state suggests that ADP-bound NBDs are relatively destabilized. While this has not been reported in other transporters, it is conceivable that a homodimer such as MsbA would be more easily subject to destabilization than proteins with one non-functional ATPase site. Thus, NBD disordering upon γ-phosphate release may favor dissociation and promote transporter cycling.

In summary, the MsbA function cycle includes notable differences from that of TmrAB[4] and ABCC1[5], suggesting that analysis of other ABC transporters under ATP turnover conditions may provide further insights into this protein family. Yet despite these advances, the present study does not provide new information about the precise mechanism of LPS flipping and direct validity of the "trap and flip" model[8]. Catching MsbA in the act of flipping LPS may thus be intractable with mere image analysis and will likely require additional biochemical, biophysical and spectroscopy experiments and involve crosslinking of LPS to MsbA.

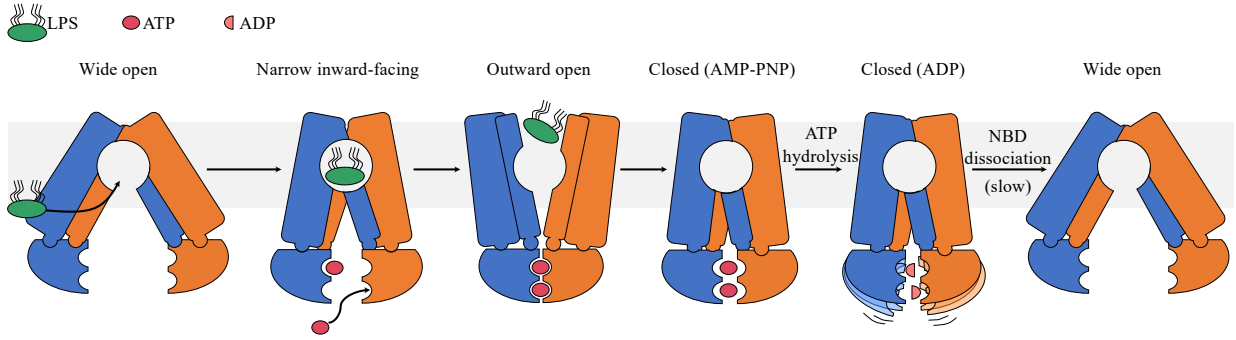


Figure 2.5: Updated model of LPS transport by MsbA. All states except outward-open are characterized in the present study. The pre- and post-hydrolysis closed conformations are displayed separately, with labelling corresponding to the nucleotide present in the cryo-EM maps (AMP-PNP and ADP). In the post-hydrolysis closed conformation, duplicated NBDs and curved lines illustrate that NBDs seem destabilized relative to the pre-hydrolysis state.

	ATP turnover Wide-open	ATP turnover Apo-like	ATP turnover Post-hydrolysis	AMP-PNP
	EMDB: XXXX	EMDB: XXXX	PDBid: XXXX EMDB: XXXX	PDBid: XXXX EMDB: XXXX
Data collection and processing				
Microscope	UMass Talos	UMass Talos	UMass Talos	UMass Titan Krios
Detector	K2	K2	K2	K2
Magnification	36,000	36,000	36,000	130,000
Voltage (kV)	200	200	200	300
Electron Exposure ($e^-/Å^2$)	48	48	48	61.5
Defocus Range (μm)	1.0 – 2.5	1.0 – 2.5	1.0 – 2.5	1.0 – 2.5
Pixel Size (Å)	1.17	1.17	1.17	1.06
Final symmetry imposed	C2	C2	C2	C2
Initial particle images (#)	668,259	668,259	668,259	802,762
Final particle images (#)	9,827	78,388	18,090	12,894
Map resolution (Å)	6.8	5.6	3.9	3.35
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	6.0 – 10.0	5.0 – 10.0	3.5 – 6.0	2.5 – 4.0
Refinement				
Initial model used (PDBid)			5TTP	5TTP
Model resolution (Å)			3.7	3.4
FSC threshold			0.5	0.5
Map sharpening B factor ($Å^2$)			N/A	N/A
Model composition				
Non-hydrogen atoms			8852	8838
Protein residues			1132	1132
Ligand			2	2
B factor ($Å^2$)				
Protein			64.74	31.38
Ligand			67.27	23.48
Model vs. Map				
Overall CC			0.80	0.84
R.m.s. deviations				
Bond lengths (Å)			0.003	0.006
Bond angles (°)			0.632	0.675
Validation				
MolProbity score			2.41	1.76
Clashscore			35.35	8.27
Poor rotamers (%)			0.1	0
Ramachandran plot				
Favored (%)			94.40	95.46
Allowed (%)			5.60	4.36
Disallowed (%)			0	0.18

Figure 2.6: Data collection table and model refinement statistics

2.4 Methods

Protein expression and purification

The MsbA construct used in this study is a donation from the G. Chang lab[6] and consists of *E. coli* MsbA in pET19b with an N-terminal poly-histidine tag. BL21 cells were transformed with the MsbA construct and cultured in LB with 100 µg/mL ampicillin. Protein production was induced by addition of 1mM IPTG when OD_{600} reached ~1.0. Incubator temperature was subsequently lowered to 17°C, and the cells were grown for 12 hours overnight. Upon completion, the cells were harvested by centrifugation for 10 min at 4,000 x g, washed in phosphate-buffered saline and flash frozen in liquid nitrogen before storage at -80°C. The pellet was later thawed and resuspended in ice cold lysis buffer (50 mM Tris base, pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP, 15mL buffer per 1L of cell culture), and vortexed to homogeneity. The cells were lysed using a probe sonicator with a 20% work cycle, at 60% of maximum sonicating power for 20 min (~4 min active sonication). The cell lysate was then centrifugated at 185,000 x g for 1.5 hours at 4°C to harvest cell membranes, which were collected and flash frozen before storage at 80°C. On protein purification day, the membranes were vigorously homogenized in solubilization buffer (lysis buffer containing 1% DDM, 20 mL buffer per 1 L of cell culture) using a Dounce homogenizer kept in ice. The mixture was transferred to a beaker equipped with a stir bar and stirred at 4°C for one hour. Next, the membrane suspension was transferred to ultracentrifuge tubes, centrifugated at 3,000 x g for 15 min to remove large insoluble debris, followed by an additional 185,000 x g spin to pellet thinner insoluble materials. The supernatant was transferred to a column containing 1 mL detergent-equilibrated cobalt resin, washed with 25 mL wash buffer (lysis buffer containing 0.1% DDM and 10 mM imidazole), before protein elution with 5 mL elution buffer (lysis buffer with 0.1% DDM and 250 mM imidazole). The eluted protein was concentrated to 500 µL using a

100 kDa MWCO concentrator, then injected on a Superdex 200 column equilibrated in gel filtration buffer (25 mM Tris pH 8.0, 150 mM NaCl, 5% (v/v) glycerol, 0.5 mM TCEP, 0.05% DDM). Peak fractions (~11 mL elution volume) were collected and concentrated to ~10 mg/mL for subsequent nanodisc reconstitution.

Nanodisc reconstitution

MsbA adopts different conformational states in detergents and nanodiscs, highlighting the importance of studying this dynamic transporter in native-like membrane environments[8]. As the most common MSP1D1 expressing plasmid and the MsbA construct discussed here both produce His-tagged proteins, we describe how to reconstitute 1-to-1 MsbA to nanodisc ratio[18]. If MsbA is fused with a different affinity tag, it is possible to create excess nanodiscs and separate the empty discs by affinity pulldown.

Reconstitution of MsbA into nanodiscs is a process involving simultaneous detergent removal and formation of lipid nanoparticles around the transporter (***Figure 2.7a***). We find the following key factors particularly important for efficient MsbA incorporation: 1) proceed with >1 mg purified protein, 2) ensure that lipid concentration in the reaction is greater than 2 mM, 3) try different lipid to MSP ratios and optimize for a monodisperse gel filtration profile (***Figure 2.7b***). For illustrative purposes, we describe how to set up a 300 µL nanodisc reaction with 1:1:60 MsbA to MSP1D1 to POPG stoichiometric ratio. In mass, this sample reaction includes 1 mg of MsbA, 0.383 mg of MSP1D1 and 0.698 mg of POPG. This protocol needs to be adapted for every nanodisc reconstitution based on MsbA yield in the protein purification step. For the sample reaction described here, 172 µL of cholate buffer (25 mM Tris, 150 mM NaCl, 250 mM sodium cholate) is used to resuspend 5 mg of dried POPG aliquot, with sonication for ~5 min or until no undissolved POPG residues are visible. Next, 100 µL of 10 mg/mL MsbA, 38.3 µL of 10mg/mL

MSP1D1, 137.7 µL of nanodisc buffer (25 mM Tris, 150 mM NaCl), and 24 µL of cholate/POPG solution are mixed and incubated on ice for one hour, before addition of 180 mg of SM-2 biobeads. The sample is subsequently incubated on a nutator overnight at 4ºC. The following day, the sample is filtered and injected on a Superdex 200 column equilibrated in nanodisc buffer (***Figure 2.7b***), then concentrated to ~1.5 mg/mL for further cryo-EM experiments. Prior to cryo-EM sample preparation, we recommend assessing sample quality by negative-stain EM. The sample is first diluted to ~0.02 mg/mL before negative-stain EM grids preparation [19]. MsbA particles must be well-separated and the two nucleotide-binding domains (NBDs) should be clearly distinguishable in individual particle images (***Figure 2.7c***) and 2D class averages (***Figure 2.7d***). Inter-NBD spacing should be decreased compared to MsbA characterized in detergent [3,8].

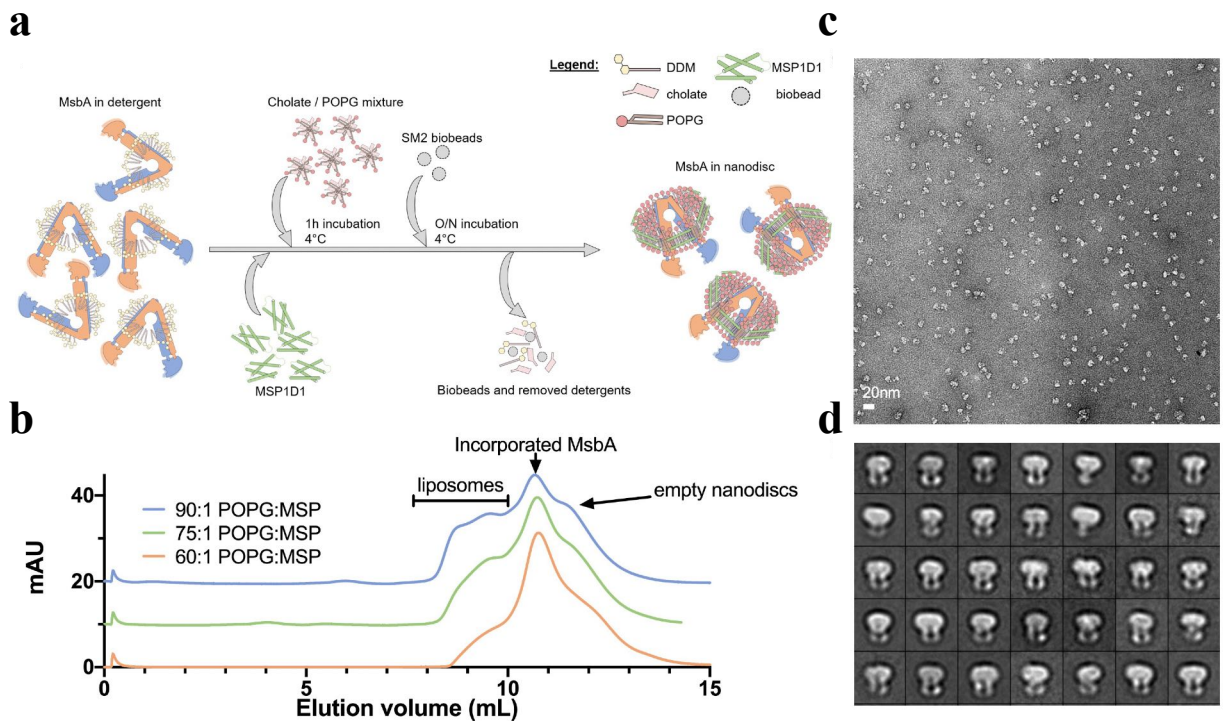


Figure 2.7: MsbA nanodisc incorporation and quality control. a, Cartoon representation of the nanodisc assembly process. DDM-solubilized MsbA (left) is incubated with membrane scaffold protein and lipids. Detergent removal results in the formation of lipid nanodiscs. **b,** Gel filtration profiles of nanodisc reconstitutions with different lipid to MSP1D1 ratio. Additional POPG results in larger amounts of liposomes yet potentially more homogeneous nanodisc reconstitution. The

Figure 2.7 (continued)

profiles are shown as offsets for display purposes. **c,** Sample negative-stain micrograph of MsbA in nanodisc, acquired on Tecnai T12 electron microscope operated at 120 kV at 67,000 x magnification. **d,** Sample negative-stain 2D averages of MsbA in nanodiscs. Box dimension is 215 Å. The nanodisc and nucleotide-binding domains are clearly identifiable in each class average.

Cryo-EM sample preparation

High-resolution imaging of MsbA requires large cryo-EM datasets composed of hundreds of thousands of protein images. It is thus important that each micrograph captures several hundreds of well-defined MsbA particles. MsbA particle distribution is unfortunately highly sensitive to ice thickness, such that few particles localize to thin ice areas. We recommend preparing cryo-EM grids using a Vitrobot plunge freezer, because Vitrobot-frozen grids typically include an ice thickness gradient (***Figure 2.8a***) where it is usually possible to find meshes with appropriate thickness for data collection. Other plunge freezers (e.g., Gatan Cryoplunge 3) may also be used but sample preparation will require careful optimization. On data collection day, we first acquire a low-magnification montage of the grid of interest (***Figure 2.8a***). A Vitrobot-frozen grid will typically include an the ice thickness gradient with regions of thicker ice (smaller mesh size, ***Figure 2.8d***) and regions with thinner ice (larger mesh size, ***Figure 2.8b).*** While the grid square meshes that are far from the ice thickness gradient may appear larger, the ice is typically too thin and MsbA particles are squeezed out of the holes (***Figure 2.8c***). In contrast, regions with thicker ice typically contain more particles in each hole (***Figure 2.8e***). These thicker regions are imaged using a direct electron detector set in counting mode (super-resolution mode is not required), with total exposure of 50-55 e/Å^2. We advise using an automatic data collection program (e.g., SerialEM) and exploiting multi-shot capabilities to increase the number of micrographs acquired in one session [20,21].

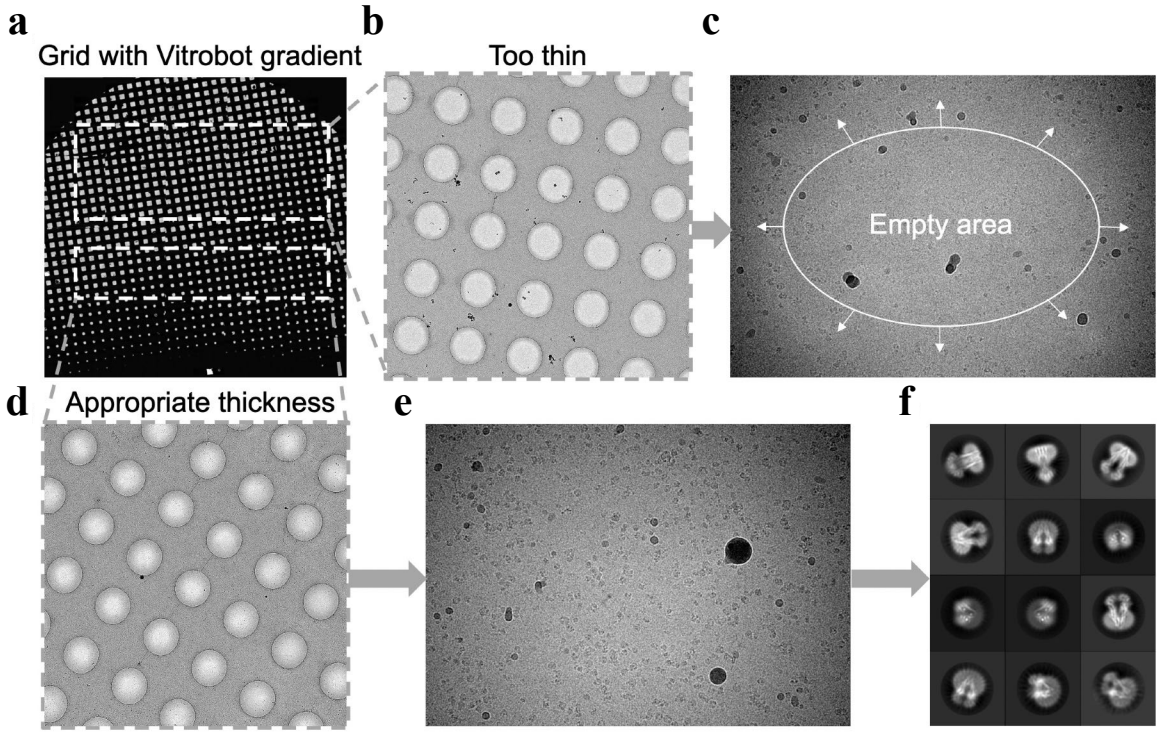


Figure 2.8: Cryo-EM sample preparation and MsbA distribution in ice. a, Low-mag montage of a cryo-EM grid frozen using a Vitrobot plunge freezer. Note that mesh size appears smaller near the thickness gradient (bottom) and increases away from the gradient (top). **b,** Sample mesh in the thinner ice areas. Ice thickness is very shallow near the center of the holes. **c,** Sample micrograph in a thin-ice region. MsbA particles are squeezed out of the center of the holes. **d,** Sample mesh in a region close to the thickness gradient. In each hole, the transition between the center and side of the hole appears smooth. **e,** Sample micrograph in a region near the thickness gradient. Particles are uniformly distributed across the hole. **f,** Sample 2D averages of an MsbA dataset collected in thicker ice. Secondary structure details such as alpha helices are clearly identifiable. Box dimension is 236 Å.

Cryo-EM 3D data processing

Once the dataset is transferred to a workstation, we perform motion correction with MotionCor2[22], fit the contrast transfer function for each micrograph with CTFFIND4[23], and pick particles with SAMUEL[8]. After the particle stack is free of carbon and ice contaminants, 3D classification aims at identifying conformationally homogeneous particles that may be refined to high resolution. 3D processing software have become progressively more sophisticated, such that little user input is now required to reach high-quality reconstructions for some datasets [24]. However, we find that

small and flexible membrane proteins such as MsbA may require expert user input in order to recover the best possible features from 3D processing. Herein, we discuss some techniques that are helpful in improving MsbA cryo-EM reconstructions (see ***Figure 2.9a*** for an overview of the 3D processing pipeline).

The particles are first binned to ~3 Å/pixel and 3D classification is performed on the binned particle set with default classification settings. A published MsbA density (e.g., EMDB:8469) low pass filtered to 35Å may be used as initial model. We recommend running 35 classification cycles and accelerating the task by using the "--fast_subsets" option in Relion. Either C1 or C2 symmetry may be used at this point. At this stage, alpha helices may already be visible as tubes of density. As 3D classification of small flexible proteins tends to be not stable with large number of particles jumping between different 3D classes, we recommend including all particles that were assigned to the best class(es) in the final 5 iterations of classification. We refer to this selection method as "5cyc" selection. Relion auto-refine is run next with default settings. If no secondary structure features are visible, default settings may be overridden by increasing tau2_fudge parameters (see ***Figure 2.9b***), as this setting may favor alignment on higher-resolution features such as alpha helices. A mask of the transmembrane domain (TMD) is then created using UCSF Chimera and fitting a published atomic model (e.g., pdb: 5TV4) of MsbA into the density [25]. We typically focus on the TMD because it is the best ordered region of this transporter and allows for more robust downstream classification. We then follow a signal subtraction protocol to create a density corresponding to nanodisc and NBDs, which will be subtracted from the particle stack (***Figure 2.9c,d***) [2,26]. The resulting particle stack is modified from the original stack, such that 2D projections of the nanodisc and NBDs are subtracted from each particle image (***Figure 2.9d***). Next, 3D classification without alignment is run on the signal subtracted stack ("--skip_align" setting),

including a TMD mask. The class with best features is selected (***Figure 2.9e***). Finally, we run local angle search auto-refine on the selected particle subset, from the original non-signal-subtracted stack. Alternatively, the same job can be launched with either a protein or full particle mask with "--solvent_correct_fsc" argument. Throughout the process, manual inspection is required to avoid picking reconstructions with artifacts from overrefinement, including dust or streaky map appearance.

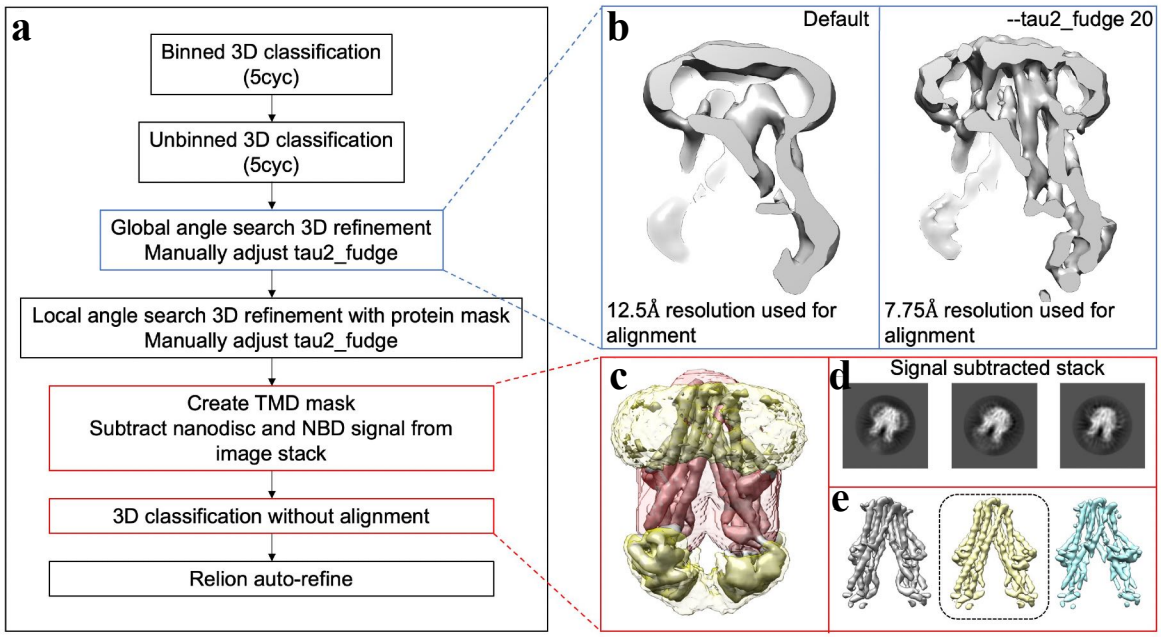


Figure 2.9: Cryo-EM 3D data processing. a, Overview of the 3D data processing pipeline. "5cyc" refers to the 5-cycle particle selection method referred to in the text. Successive rounds of particle classification result in identification of a homogeneous set of particles used in final refinement. **b,** Comparison of default Relion refinement and refinement with adjusted –tau2_fudge value. Increased tau2_fudge value results in improved transmembrane domain features but comes at the expense of possible overfitting. **c,** Mask and maps used in particle stack signal subtraction. 3D refinement map is displayed in gray, and transmembrane domain mask in red. The yellow density corresponds to the projection that is subtracted from the image stack. **d,** 2D averages from the signal-subtracted stack, illustrating that density from the nanodisc and nucleotide-binding domains were successfully removed by the signal subtraction process. Box dimension is 236 Å. **e,** Classification without alignment of the signal subtracted stack. Particles are classified exclusively based on signal from the transmembrane domain. The circled class shows continuous alpha helices and stronger helix pitch than the gray and blue classes and should therefore be selected for final particle refinement.

2.5 References

1. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).

2. Scheres, S.H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-57 (2016).

3. Moeller, A. et al. Distinct conformational spectrum of homologous multidrug ABC transporters. *Structure* **23**, 450-460 (2015).

4. Hofmann, S. et al. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580-583 (2019).

5. Wang, L. et al. Characterization of the kinetic cycle of an ABC transporter by single-molecule and cryo-EM analyses. *Elife* **9**(2020).

6. Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-10 (2007).

7. Padayatti, P.S. et al. Structural Insights into the Lipid A Transport Pathway in MsbA. *Structure* **27**, 1114-1123 e3 (2019).

8. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

9. Angiulli, G. et al. New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins. *Elife* **9**(2020).

10. Wright, J., Muench, S.P., Goldman, A. & Baker, A. Substrate polyspecificity and conformational relevance in ABC transporters: new insights from structural studies. *Biochem Soc Trans* **46**, 1475-1484 (2018).

11. Borbat, P.P. et al. Conformational motion of the ABC transporter MsbA induced by ATP hydrolysis. *PLoS Biol* **5**, e271 (2007).

12. Zou, P., Bortolus, M. & McHaourab, H.S. Conformational cycle of the ABC transporter MsbA in liposomes: detailed analysis using double electron-electron resonance spectroscopy. *J Mol Biol* **393**, 586-97 (2009).

13. Josts, I. et al. Conformational States of ABC Transporter MsbA in a Lipid Environment Investigated by Small-Angle Scattering Using Stealth Carrier Nanodiscs. *Structure* **26**, 1072-1079 e4 (2018).

14. Oldham, M.L. & Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc Natl Acad Sci U S A* **108**, 15152-6 (2011).

15. Johnson, Z.L. & Chen, J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. *Cell* **168**, 1075-1085 e9 (2017).

16. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-5 (2014).

17. Gu, R.X. et al. Conformational Changes of the Antibacterial Peptide ATP Binding Cassette Transporter McjD Revealed by Molecular Dynamics Simulations. *Biochemistry* **54**, 5989-98 (2015).

18. Denisov, I.G., Grinkova, Y.V., Lazarides, A.A. & Sligar, S.G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **126**, 3477-87 (2004).

19. Liu, J. et al. Optimized Negative-Staining Protocol for Lipid-Protein Interactions Investigated by Electron Microscopy. *Methods Mol Biol* **2003**, 163-173 (2019).

20. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).

21. Cash, J.N., Kearns, S., Li, Y. & Cianfrocco, M.A. High-resolution cryo-EM using beam-image shift at 200 keV. *IUCrJ* **7**, 1179-1187 (2020).

22. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

23. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).

24. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).

25. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

26. Bai, X.C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S.H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**(2015).

Chapter 3: TBT1 binds in the central MsbA pocket and promotes wasteful ATP hydrolysis by inducing a collapsed inward-facing conformation.

The data presented here is part of a preprint:

Thélot, F. A., Zhang, W., Song, K., Xu, C., Huang, J., & Liao, M. (2021). Distinct allosteric mechanisms of first-generation MsbA inhibitors. bioRxiv, 2021.2005.2025.445681. doi:10.1101/2021.05.25.445681

Contribution: F. Thelot performed molecular cloning, protein purification, cryo-EM sample preparation, cryo-EM data processing, model building and biochemistry experiments. M. Liao performed cryo-EM data processing and supervised the project. W. Zhang and J. Huang conducted in silico docking and MD simulations. K. Song and C. Xu helped with cryo-EM data acquisition.

3.1 Introduction

ATP-binding cassette (ABC) transporters perform diverse functions including importing and exporting small molecules[1-3], mediating ion channel opening[4,5], and translocating lipids across cell membranes[6-8]. Most ABC transporters have two transmembrane domains (TMDs), which interact with and transport substrates, and two nucleotide-binding domains (NBDs), which bind and hydrolyze ATP. Decades of investigation have led to various models to describe ABC transporter mechanisms. In the alternating access model[9-13], the substrate is first recognized by the inward-facing transporter, subsequent ATP binding in the NBDs promotes a transition to the outward-facing state and substrate release, and finally ATP hydrolysis resets the transporter in the inward-facing state. Interrupting such conformational transition cycle has direct applications in treating cancer[14], regulating cholesterol homeostasis[15], and developing new antibiotics[16].

Many small molecule inhibitors were developed against human multidrug ABC transporters, including the ABCB1 inhibitors zosuquidar[3,17,18], tariquidar[3] and elacridar[3], as well as the ABCG2 inhibitors MZ29[19] and MB136[19]. Structural studies have revealed how these compounds bind to the TMDs, interrupting conformational transition and consequently decreasing ATPase activity[3,19,20]. In contrast, there is little structural insight into small-molecule modulation of most other ABC transporters, which demonstrate much narrower substrate spectra. Consequently, we do not know if any modes of small-molecule inhibition are shared across the ABC superfamily, or if there are general druggable conformations or pockets. Such gap of knowledge limits our ability to rationally design drugs targeting many ABC transporters.

MsbA is an ideal model system to understand ABC transporter inhibition because its function cycle is particularly well-understood[6,21,22]. Furthermore, two families of MsbA-specific inhibitors with distinct properties have been reported: one has a tetrahydrobenzothiophene (TBT)

scaffold, named herein as TBT1, and stimulates MsbA ATPase activity despite abolishing LPS transport[23]; the other, here referred to as G compounds, blocks both ATP hydrolysis and LPS transport[24,25]. In this context, we sought to understand how TBT1 and G compounds, while both blocking LPS transport, exert opposite allosteric effects on ATP hydrolysis.

TBT1 was discovered by the Kahne group, using a cell-based high-throughput screening platform designed to identify inhibitors specific to the LPS biogenesis and transport pathway[23]. The screen exploits the observation that LPS biogenesis can be safely knocked out in some Gram-negative strains such as *A. baumannii*, provided the gene knockout involves one of the earlier genes of LPS synthesis (e.g., lpxA, lpxC, lpxD). In this case, the cell is non-committed to LPS synthesis and will not accumulate toxic LPS intermediates as if a downstream protein (e.g., MsbA) had been disabled. The authors then created two *A. baumannii* strains: a strain with functional LPS biosynthesis pathway, and a lpxA knockout that does not produce LPS. Each strain is exposed to a chemical compound library, and the minimal inhibitory concentration (MIC) for each strain is compared. Notably, a bactericidal compound that doesn't target the LPS pathway will have higher MIC in the LPS-containing cell than in the LPS-deficient strain, as LPS-deficient bacteria lacks a proper impermeability barrier. However, a selective inhibitor for LPS synthesis will kill LPS-producing cells much more effectively than the lpxA knockout, because accumulation of LPS intermediate will result in cell death. Thus, this phenotypic screen allows for the efficient identification of inhibitors for the LPS biogenesis pathway.

Screening of 150,000 compounds resulted in identification of an LPS biogenesis inhibitor with TBT scaffold[23]. Interestingly, all mutants that eventually acquired resistance to TBT1 had mutations in the msbA gene, with mutations mapping to the MsbA transmembrane domains. Intriguingly, TBT1 caused dose-dependent increase in MsbA ATPase activity, yet abolished LPS

transport as assessed by electron microscopy tissue sections[23]. Based on this information, Zhang and colleagues hypothesized that TBT1 is an MsbA decoupler of ATPase activity from LPS transport[23].

MsbA inhibition by TBT1 is particularly puzzling because most known inhibitors suppress ATP hydrolysis of ABC transporters in membrane environment[12]. We reasoned that structural information would shine light onto this unusual ABC transporter inhibition mechanism.

3.2 Results

3.2.1 TBT1-bound MsbA adopts an asymmetric, collapsed inward-facing conformation

TBT1 was identified as an LPS transport inhibitor and MsbA ATPase stimulator in strains from the *Acinetobacter* genus[23]. We therefore sought to explore the mechanism of TBT1 inhibition in *Acinetobacter baumannii*, an ESKAPE[26] pathogen responsible for antibiotic-resistant infections in patients. *A. baumannii* MsbA was expressed in *Escherichia coli* cells, purified in dodecyl maltoside (DDM), and reconstituted in palmitoyl-oleoyl-phosphatidylglycerol (POPG) nanodiscs[27,28] (***Figure 3.1a-g***). Basal ATPase activity of *A. baumannii* MsbA in nanodiscs was ~1 μmol ATP/min/mg MsbA (***Figure 3.2a***), which is ~4-fold lower than *E. coli* MsbA[6]. Exposure of MsbA to TBT1 resulted in 4-7x fold stimulation in ATPase activity (***Figure 3.1d,g***), as well as dose-dependent increase in reaction rate in Michaelis-Menten kinetics (***Figure 3.2a***). Notably, TBT1-induced ATPase stimulation was much more pronounced in nanodiscs than in DDM (***Figure 3.1g***), suggesting that nanodiscs are a more suitable system than detergent for structural studies. We thus used single-particle cryo-EM to determine the structure of TBT1-bound MsbA in nanodiscs at an overall resolution of 4.3 Å, with the TMDs at 4.0-Å resolution showing sufficient side-chain densities for model building (***Figure 3.2b***, ***Figure 3.3*** and ***Figure 3.4d***).

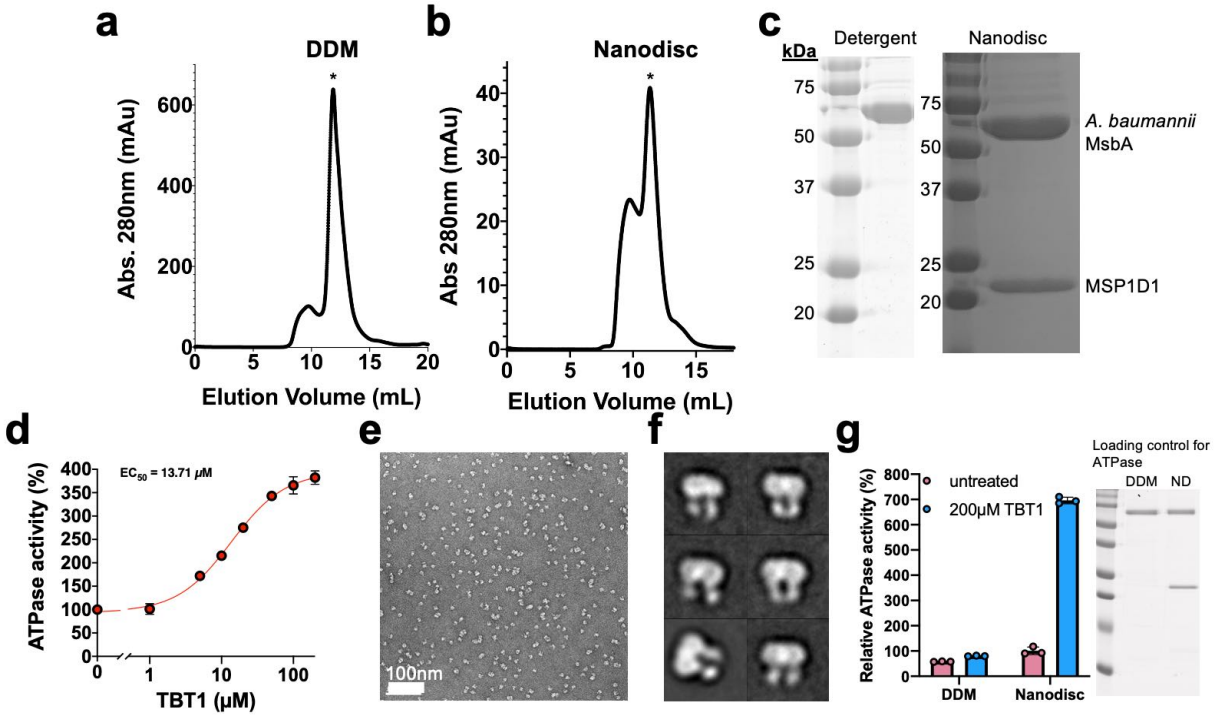


Figure 3.1: Characterization of A. baumannii MsbA. a, Gel filtration profile of purified A. baumannii MsbA in DDM, ran on a Superdex 200 column. The protein elution peak is marked with a star symbol. **b,** Gel filtration profile of A. baumannii MsbA in MSP1D1 nanodiscs. Incorporated MsbA is marked with a star symbol. **c,** SDS-PAGE gels of A. baumannii MsbA in DDM (left) and incorporated in nanodiscs (right). **d,** ATPase activity of A. baumannii MsbA when exposed to increasing TBT1 concentrations. Each point represents mean ± SD (n=3). **e,** Representative negative-stain EM image of A. baumannii MsbA in nanodiscs. **f,** Representative negative-stain 2D class averages of A. baumannii MsbA in nanodiscs. Box size is 215 Å. **g,** Comparison of ATPase activity of A. baumannii MsbA in DDM and nanodiscs (ND) and sensitivity to TBT1-induced stimulation. The data is normalized relative to the ATPase activity of drug-free MsbA in nanodisc. Error bars represent mean ± SD (n=3). The SDS-PAGE gel shown on the right-hand side was used to confirm that a similar amount of protein was loaded in each ATPase reaction.

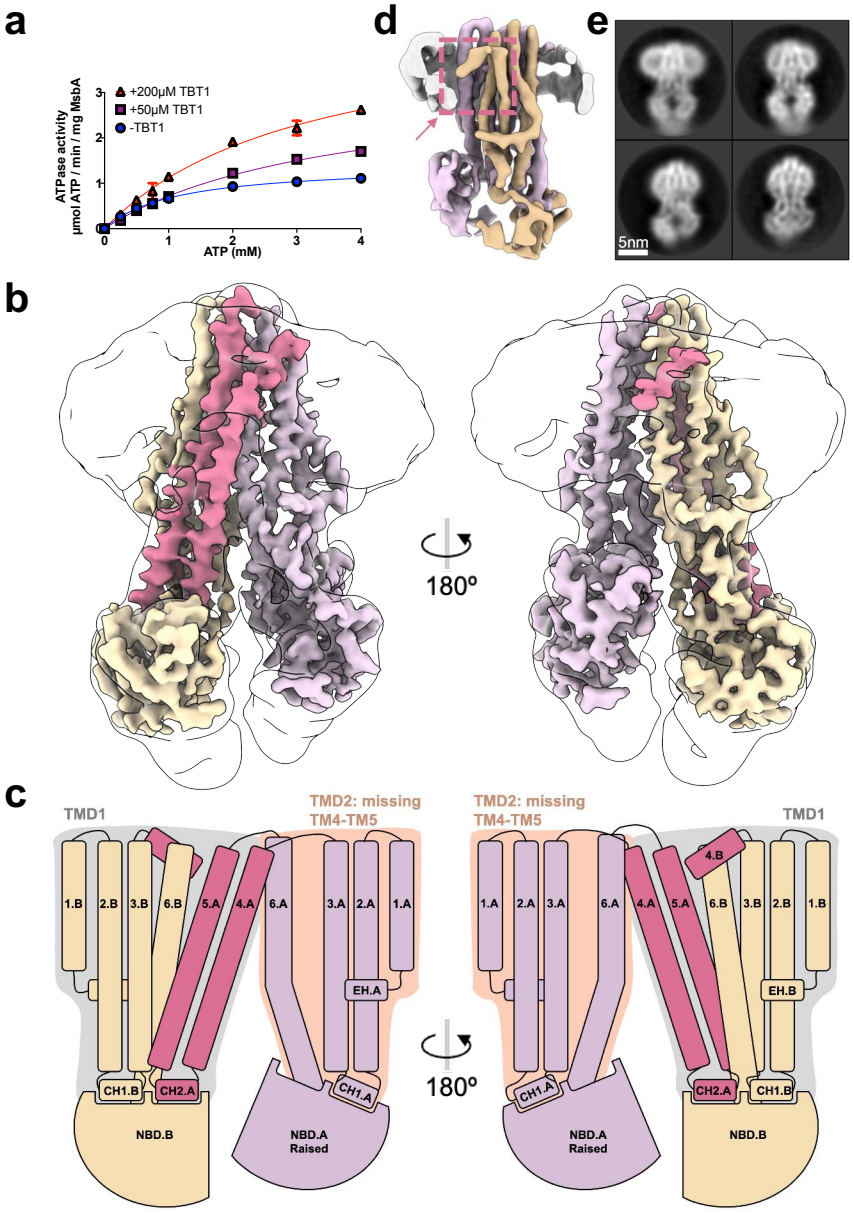


Figure 3.2: TBT1 binding induces a collapsed inward-facing conformation of MsbA. a, ATPase activity of *A. baumannii* MsbA in nanodiscs, measured at varying ATP and TBT1 concentrations. TBT1 increases the rate of ATP hydrolysis in a dose-dependent manner. Each point represents mean ± SD (n=3). **b,** 4.0-Å resolution cryo-EM map of *A. baumannii* MsbA in complex with TBT1, with domain-swapping TM4-TM5 colored fuchsia. The unsharpened map filtered at 10-Å resolution is displayed as outline to show the nanodisc and tightened NBDs positioning. **c,** Cartoon of TBT1-bound MsbA with structural elements indicated, illustrating the structural asymmetry of the two MsbA chains. EH, elbow helix; CH, coupling helix. **d,** Cryo-EM reconstruction low-pass filtered at 6-Å resolution. The dashed box points to the N-terminal end of TM4.B, where TM4-TM5.B becomes disordered and absent from the reconstruction. **e,** Representative 2D class averages of TBT1-bound MsbA, showing constricted TMDs yet separate NBDs. Box size is 203 Å.

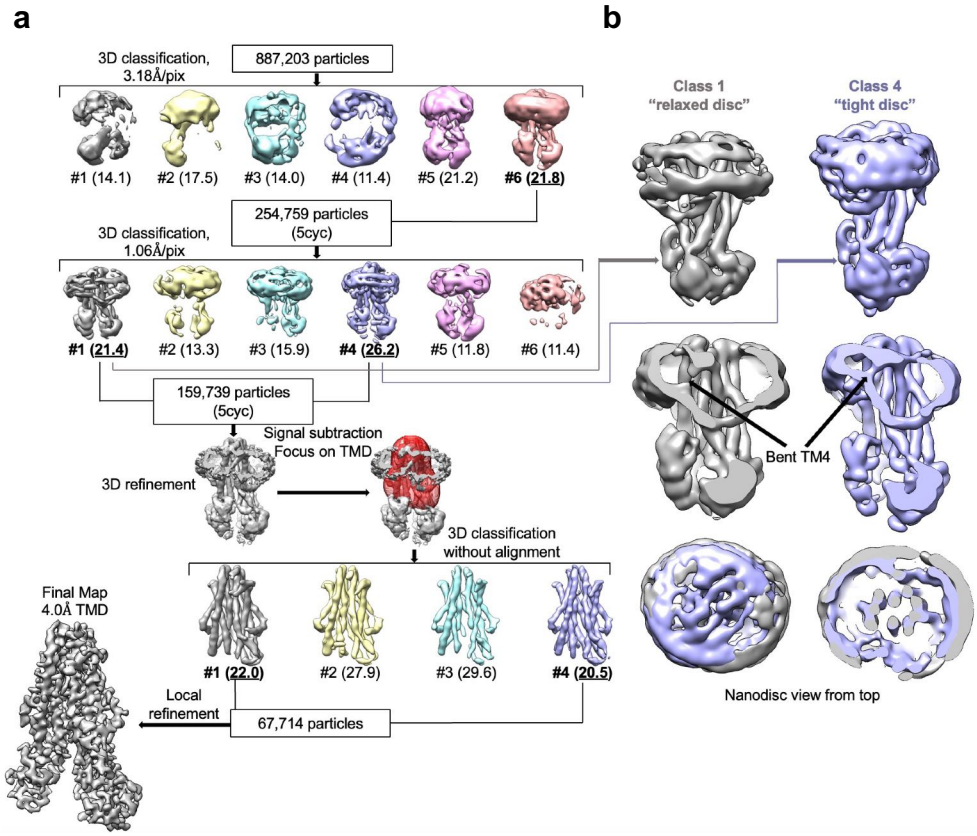


Figure 3.3: Processing flowchart for *A. baumannii* MsbA in complex with TBT1. a, Flowchart showing all 3D classification and refinement steps used to determine the structure of the collapsed inward-facing conformation. "5cyc" refers to the method of including any particle being assigned to a selected class within the final 5 cycles in 3D classification. **b,** Comparison of two classes from the second round of 3D classification, showing that the collapsed inward-facing conformation is compatible with different sizes and shapes of nanodisc.

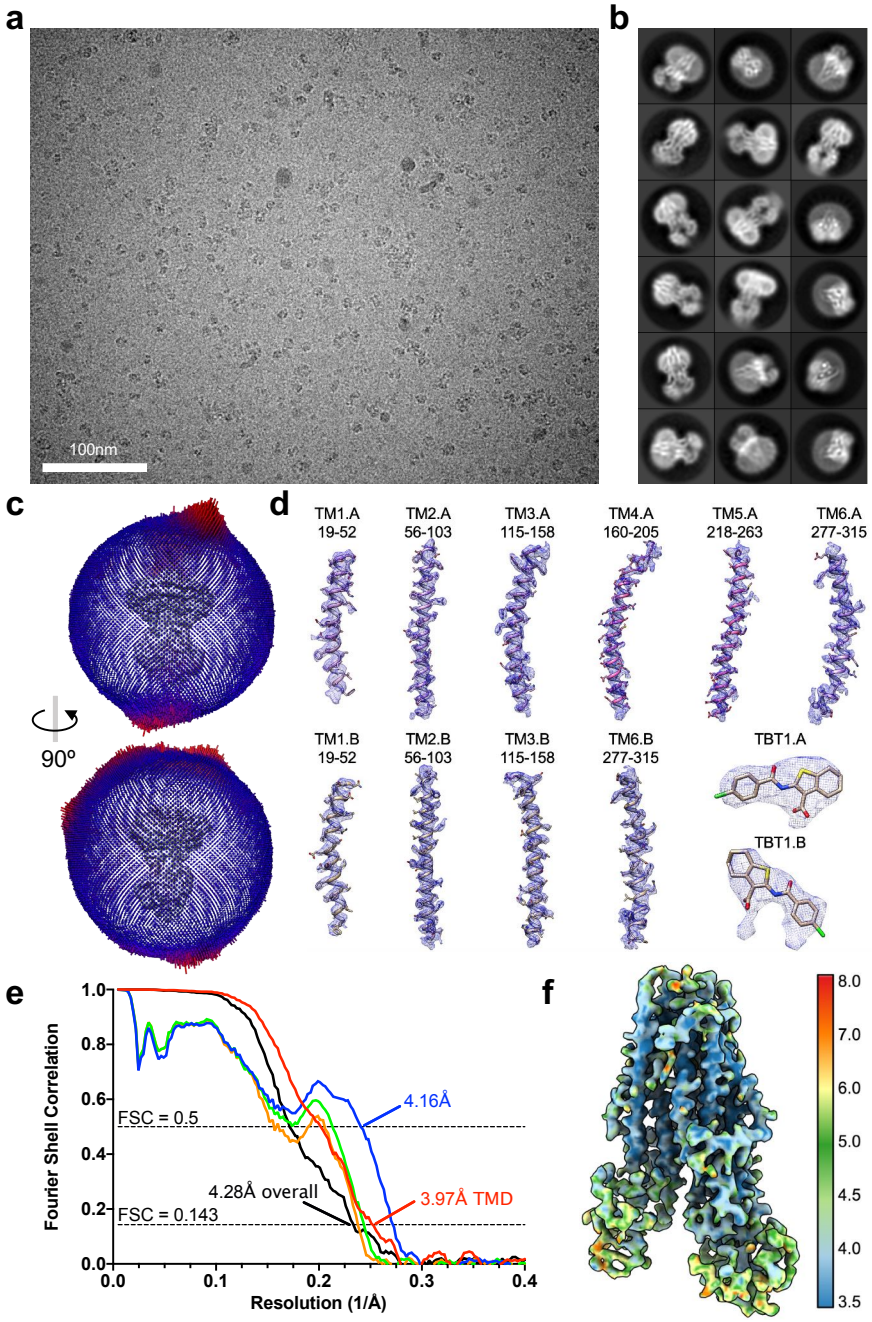


Figure 3.4: Cryo-EM imaging and validation for *A. baumannii* MsbA in complex with TBT1. **a,** Representative cryo-EM image of TBT1-bound MsbA. **b,** 2D class averages of *A. baumannii* MsbA bound to TBT1. The asymmetric positioning of NBDs is readily discernable. Box size is 203 Å. **c,** Angle distribution of the cryo-EM particles included for final 3D reconstruction. **d,** Close-up view of all ordered transmembrane helices and TBT1 densities. **e,** Fourier Shell Correlation (FSC) curves of the final reconstruction. Half map #1 vs. half map #2 for the entire MsbA molecule is shown in black. The remaining FSC curves were calculated for TMDs only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange). **f,** Local resolution of the final cryo-EM map.

The conformation of TBT1-bound MsbA is asymmetric and distinct from all previously determined structures of type IV ABC transporters[29] that are characterized by their domain-swapped TMs. Notably, the domain-swapping TM4-TM5 bundle is well resolved in chain A yet disordered in chain B (***Figure 3.2b,c***), which is evident also in unmasked and unsharpened 3D reconstructions (***Figure 3.5b***). This chain mismatch results in asymmetrical TMDs, with TMD1 including 6 helices (TM4,5.A, TM1,2,3,6.B) and TMD2 including only 4 helices (TM1,2,3,6.A). While TM4.B is mostly unresolved, its N-terminal end is clearly discernable and forms a 61º angle with the C-terminal end of TM3.B, which is ~15º greater than the corresponding angle between the well resolved TM4.A and TM3.A (***Figure 3.5c***). Accordingly, TM4.B kinks outwards, pushing the nanodisc density to bulge upwards from the membrane plane (***Figure 3.2d*** and ***Figure 3.5c***, left). Interestingly, while the consensus map features a bent nanodisc, a subset of particles was incorporated in comparatively flatter and larger nanodiscs yet still adopt the asymmetric TBT1-bound conformation, suggesting that TM4-TM5 bundle destabilization is not dependent on nanodisc tightness (***Figure 3.3b***). The domain-swapping TM4-TM5 bundle is of critical importance because the coupling helix (CH) at the TM4-TM5 junction (CH2), together with CH1 at the TM2-TM3 junction, is responsible for transmitting movement from the catalytically active NBDs to the TMDs (***Figure 3.2c***). Due to disordered TM4-TM5.B, CH2.B is absent for interaction with the NBD in chain A (NBD.A). Accordingly, NBD.A tilts up, moving towards the NBD dimerization interface, and becomes more mobile as its density is considerably weakened compared to that of the opposing NBD.B (***Figure 3.2b,c***). Notably, CH1.A also appears destabilized, likely because it cannot form proper interaction with the drastically shifted NBD.A (***Figure 3.5d***). In summary, the raised NBD.A is disengaged from its two coupling helices and seems decoupled from the TMDs (***Figure 3.5d***).

At first sight, 2D class averages and 3D reconstruction of the cryo-EM images of TBT1-bound MsbA (***Figure 3.2e*** and ***Figure 3.5e***) resemble those of nucleotide-bound MsbA[6], both exhibiting constricted TMDs and reduced inter-NBD distance compared to nucleotide-free MsbA. However, the TBT1-bound MsbA conformation is still inward-facing because no nucleotide is present in the cryo-EM sample and the NBDs remain fully separate. Accordingly, TMD1 is well aligned to the structure of nucleotide-free *E. coli* MsbA (***Figure 3.5f***, root-mean-square-deviation (RMSD) of 5.5 Å over 314 Cα atoms), leaving significant structural mismatch from TMD2. The most noteworthy structural rearrangement in TMD2 is the 13-Å displacement of TM6.A towards the central cavity (***Figure 3.5g***), which seems to pull NBD.A upwards and to close proximity of NBD.B. Thus, the TBT1-bound MsbA structure described here is an unusual "collapsed inward-facing" conformation characterized by two unprecedented features: 1) complete destabilization of one domain-swapping TM4-TM5 helix bundle; 2) highly asymmetric positioning of the NBDs.

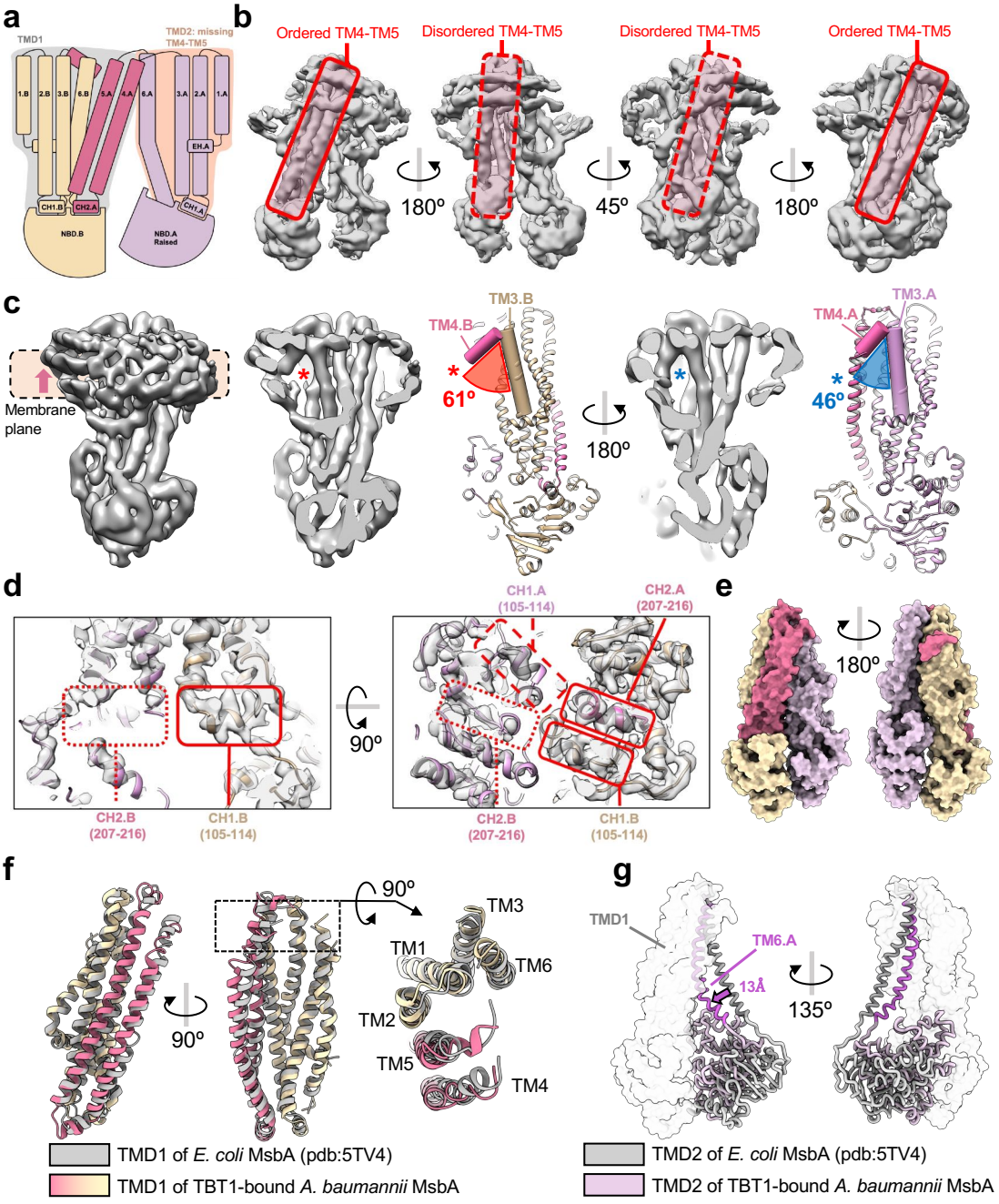


Figure 3.5: Structural analysis of TBT1-inhibited MsbA. a, Reproduction of TBT1-bound MsbA cartoon from Fig. 1c. **b,** Rotated side views of the unsharpened, unmasked cryo-EM map of TBT1-bound MsbA. The location of the domain-swapping TM4-TM5 bundles is highlighted in red box, with solid line when ordered and dashed line when disordered. **c,** Angle comparison between TM3 and the N-terminal end of TM4 for each MsbA chain. In chain B (left, red asterisk), TM4 is kinked outward and pushes the nanodisc scaffold upwards and away from the membrane plane. In chain A (right, blue asterisk), the separation between TM3 and TM4 is narrower. For angle measurements, axes were defined using the "define axis" function in UCSF chimera and

Figure 3.5 (continued)

selecting residues 133-157 for TM3 and 158-165 for TM4. **d,** Map and model highlighting the location of the coupling helices (CH). Coupling helices are boxed with a solid line when present, dashed line when destabilized, and dotted line when disordered. **e,** Surface representation of the model of TBT1-bound MsbA, highlighting the constricted TMDs. **f,** Alignment of the TMD1 bundle of TBT1-bound *A. baumannii* MsbA (colored) with drug-free *E. coli* MsbA (PDB: 5TV4, gray). **g,** Different positioning of TM6.A between TBT1-bound *A. baumannii* MsbA (pink) and drug-free *E. coli* MsbA (PDB: 5TV4, gray). The structurally similar TMD1 is shown as transparent surface, while TM6.A and NBD.A are displayed using worm representation. TM6.A is colored in darker pink or gray to highlight its location. The 13 Å shift of TM6 is measured using Ala308 as reference.

3.2.2 TBT1 binding drives the closure of a wide open MsbA

We then sought to characterize the structure of drug-free *A. baumannii* MsbA to understand how TBT1 binding affects MsbA conformation. To this end, we imaged *A. baumannii* MsbA in identical conditions as described above, but without addition of TBT1 (***Figure 3.6***). The resulting cryo-EM reconstruction at an overall resolution of 5.2 Å enabled unambiguous tracing of each TM helix, revealing a wide inward-facing conformation (***Figure 3.6***). Along with the unpublished structure of wide-open *E. coli* MsbA in nanodisc shown in chapter 2, this is the first example of wide inward-facing MsbA in a lipid bilayer. Such conformation had only been characterized by X-ray crystallography using MsbA in detergents[30,31], and its physiological relevance has been heavily debated[6,32,33]. Our result demonstrates that the wide inward-facing conformation of MsbA can exist in a lipid membrane and may not be an artifact caused by deprivation of membrane environment and usage of detergents. Furthermore, unlike the conformation presented in chapter 2, the wide opening of MsbA seems not due to lack of bound substrate because a clear LPS-like density is present in the inner cavity (***Figure 3.6g***).

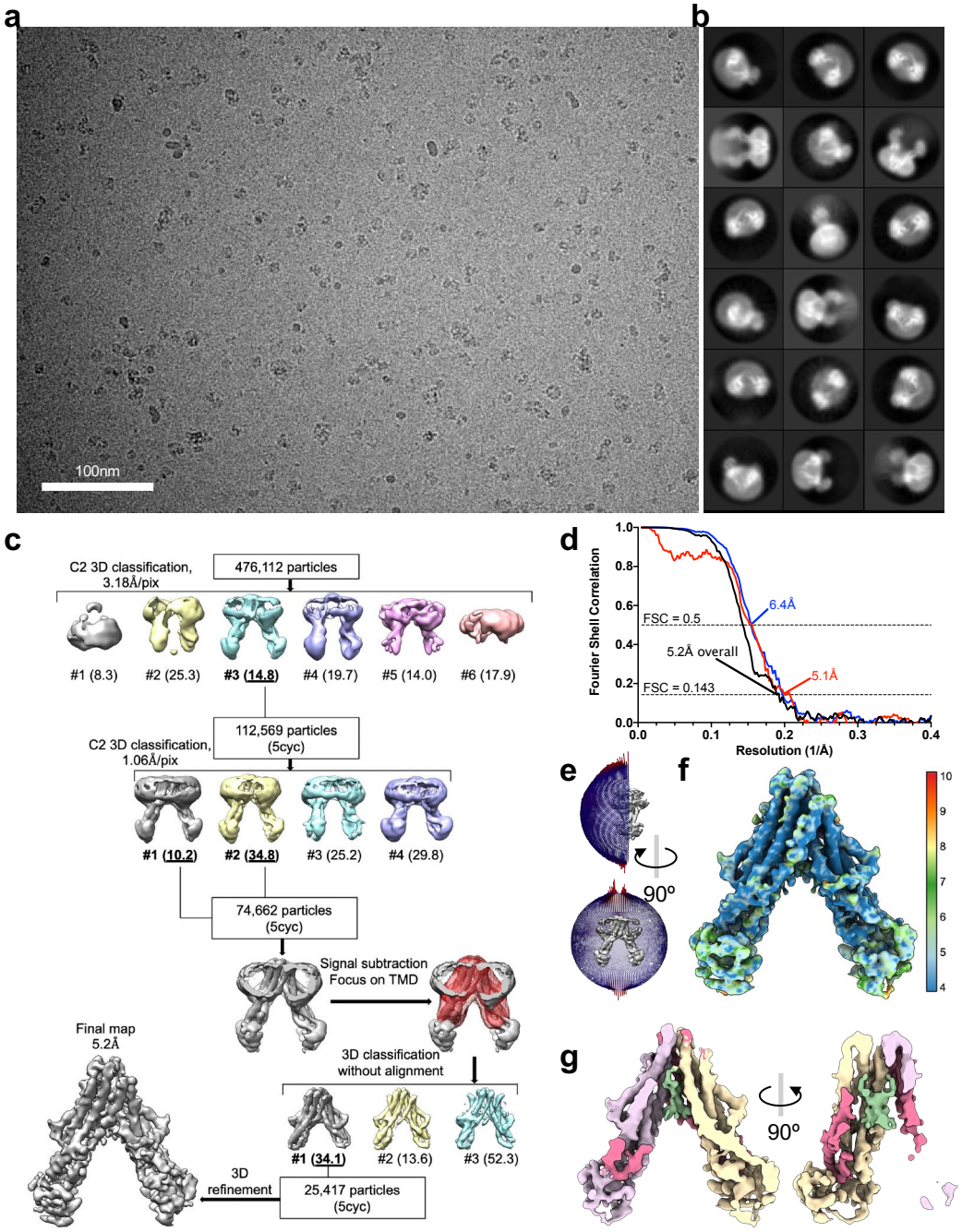


Figure 3.6: Cryo-EM imaging and validation for drug-free *A. baumannii* MsbA. a, Representative cryo-EM image of MsbA in nanodiscs in the absence of compound. **b,** 2D class averages of *A. baumannii* MsbA. Box size is 203 Å. **c,** Processing flowchart for the cryo-EM dataset. **d,** FSC curves of the final reconstruction. Half map #1 vs. half map #2 for the entire protein is shown in black. The remaining curves were calculated for TMDs only: half map #1 vs. half map #2 (red), and model vs. refined map (blue). **e,** Angle distribution of the cryo-EM particles included for final 3D reconstruction. **f,** Local resolution for the final cryo-EM map. **g,** A C2-symmetrized LPS-like density, colored dark green, is present in the central substrate-binding pocket.

Since drug-free *A. baumannii* MsbA has well defined domain-swapping TM4-TM5 bundle in both TMDs and distanced NBDs, TBT1 binding is solely responsible for converting the wide open MsbA to the collapsed inward-facing conformation. To gain insights of TBT1 stimulated ATP turnover of MsbA, we compared three cryo-EM structures: TBT1-bound or drug-free *A. baumannii* MsbA, and *E. coli* MsbA (***Figure 3.7***). Notably, the central substrate-binding pocket of TBT1-bound MsbA is much more constricted than that in the other two structures (***Figure 3.7b,c***), consistent with the notion that the TMDs in TBT1-bound state present similarities with an outward-facing transporter. Upon TBT1 binding, the inter-NBD distance is drastically decreased from ~47 Å to ~20 Å and even slightly shorter than the NBD distance in LPS-bound *E. coli* MsbA (***Figure 3.7d,e***). The two NBDs of TBT1-bound MsbA are positioned asymmetrically, such that one ATP binding site is tightened compared to the opposite ATP site (~19 Å vs. ~22 Å, ***Figure 3.7e***). The structural analysis of TBT1-bound MsbA provides an intuitive explanation for ATPase activity stimulation: removal of TM4-TM5.B bundle and sliding of TM6.A into the central cavity greatly reduce inter-NBD distance, thus increasing the speed of NBD dimerization and ATP hydrolysis.

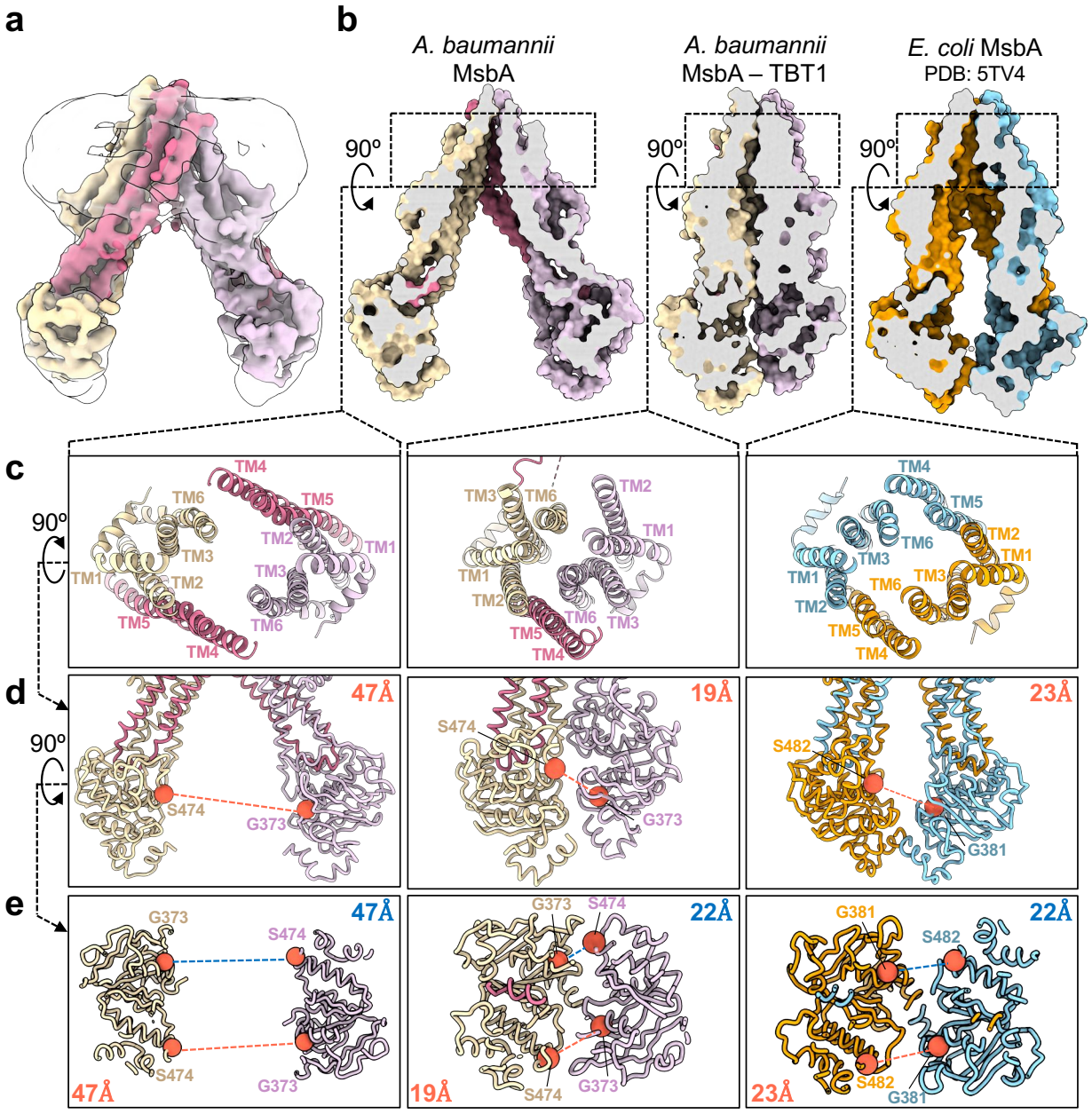


Figure 3.7: Drug-free *A. baumannii* MsbA in nanodiscs adopts a wide inward-facing conformation. a, 5-Å resolution cryo-EM reconstruction of *A. baumannii* MsbA in the absence of ligand. The unsharpened map filtered at 8-Å resolution is displayed as outline to show the nanodisc. **b,** Section through surface representation and comparison for three structures: *A. baumannii* MsbA, TBT1-bound *A. baumannii* MsbA, and *E. coli* MsbA (PDB: 5TV4). **c,** Top-down view of the substrate binding pockets. **d,** Side views of the NBDs. Cα distances between ABC signature motif serine and opposing Walker A glycine residues are indicated as dashed lines, with values rounded to the nearest integer shown in each panel. **e,** Top view of the NBDs, seen from the transmembrane domains. Cα distances between NBDs are indicated similarly to panel d.

3.2.3 TBT1 hijacks the LPS binding site to modulate MsbA

Two densities, each consistent with a TBT1 molecule, are present in the upper region of the TMDs (***Figure 3.8a,b*** and ***Figure 3.9a,b***). Interestingly, these two TBT1 ligands are positioned asymmetrically in the central substrate pocket of MsbA and related by a ~60º rotation angle, with the carboxyl group pointing downwards to the cytoplasm. The general binding mode of TBT1 is reminiscent of how *E. coli* MsbA binds the amphiphilic LPS using a large hydrophobic pocket, which accommodates the lipid acyl chains in LPS, and a ring of basic residues (Arg78, Arg148 and Lys299), which stabilize the phosphate groups on glucosamines[6]. Similarly, the two TBT1 molecules are located right above the ring of basic residues, with the aromatic ring structures of TBT1 positioned in the hydrophobic pocket (***Figure 3.8d***).

Despite being located in asymmetric binding pockets (***Figure 3.9b***), each TBT1 molecule is stabilized by an analogous set of hydrogen-bonding and electrostatic interactions between TBT1 carboxyl groups and neighboring basic sidechains. TBT1.A is oriented parallel to the membrane plane, with its carboxyl group within hydrogen-bonding distance of Lys290.A and forming long-range electrostatic interactions with Lys290.B and Arg72.A, and TBT1.B is rotated relative to TBT1.A, forming a salt bridge with Arg72.B (***Figure 3.8b***). Strikingly, the TBT1 densities and neighboring residues have better-defined features than the rest of the TMDs, suggesting that TBT1 binding stabilizes the surrounding pocket (***Figure 3.8b*** and ***Figure 3.9a***) and allowing for unambiguous orientation of TBT1 into the density map (***Figure 3.9c***).

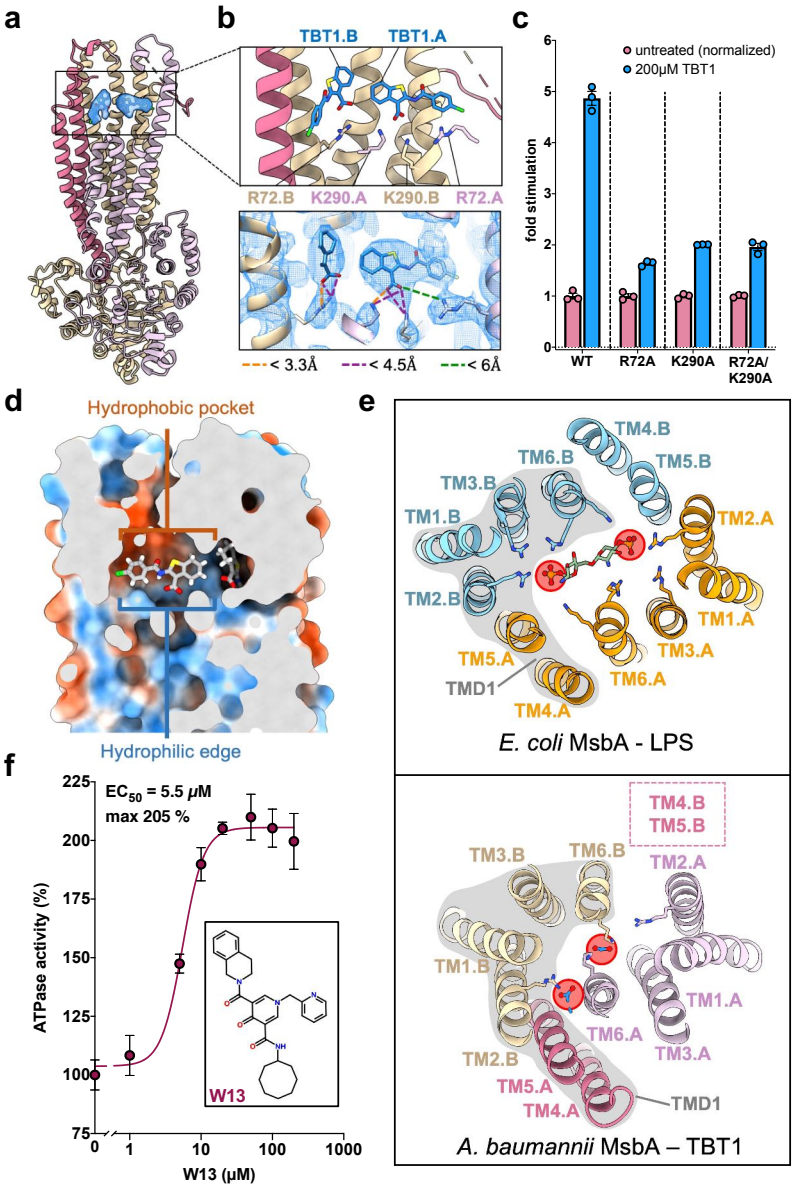


Figure 3.8: TBT1 binding to MsbA. a, Cartoon representation of TBT1-bound *A. baumannii* MsbA colored as in Fig. 1b, with TBT1 colored blue. **b,** Close-up view of the TBT1 binding site (top) and superimposition of model and cryo-EM density (bottom). Distances between TBT1 carboxyl group and neighboring lysine and arginine residues are shown as colored dashed lines. **c,** TBT1-induced ATPase stimulation of wild-type and mutant MsbA. The activity of each protein was normalized to its basal activity without TBT1. Error bars correspond to mean ± SD (n=3). **d,** Surface representation of the TBT1 binding pocket, with hydrophilic and hydrophobic surfaces colored blue and orange, respectively. **e,** Top-down view of *E. coli* MsbA bound to LPS (PDB: 5TV4) and *A. baumannii* MsbA bound to TBT1. TMD1 is conformationally similar in both structures and indicated with a gray background. The phosphate groups on LPS glucosamines and the carboxyl groups of TBT1 are marked with red circles. **f,** ATPase activity of *A. baumannii* MsbA with W13 at increasing concentrations. Each point represents mean ± SD (n=3). The molecular structure of W13 is shown in inset.

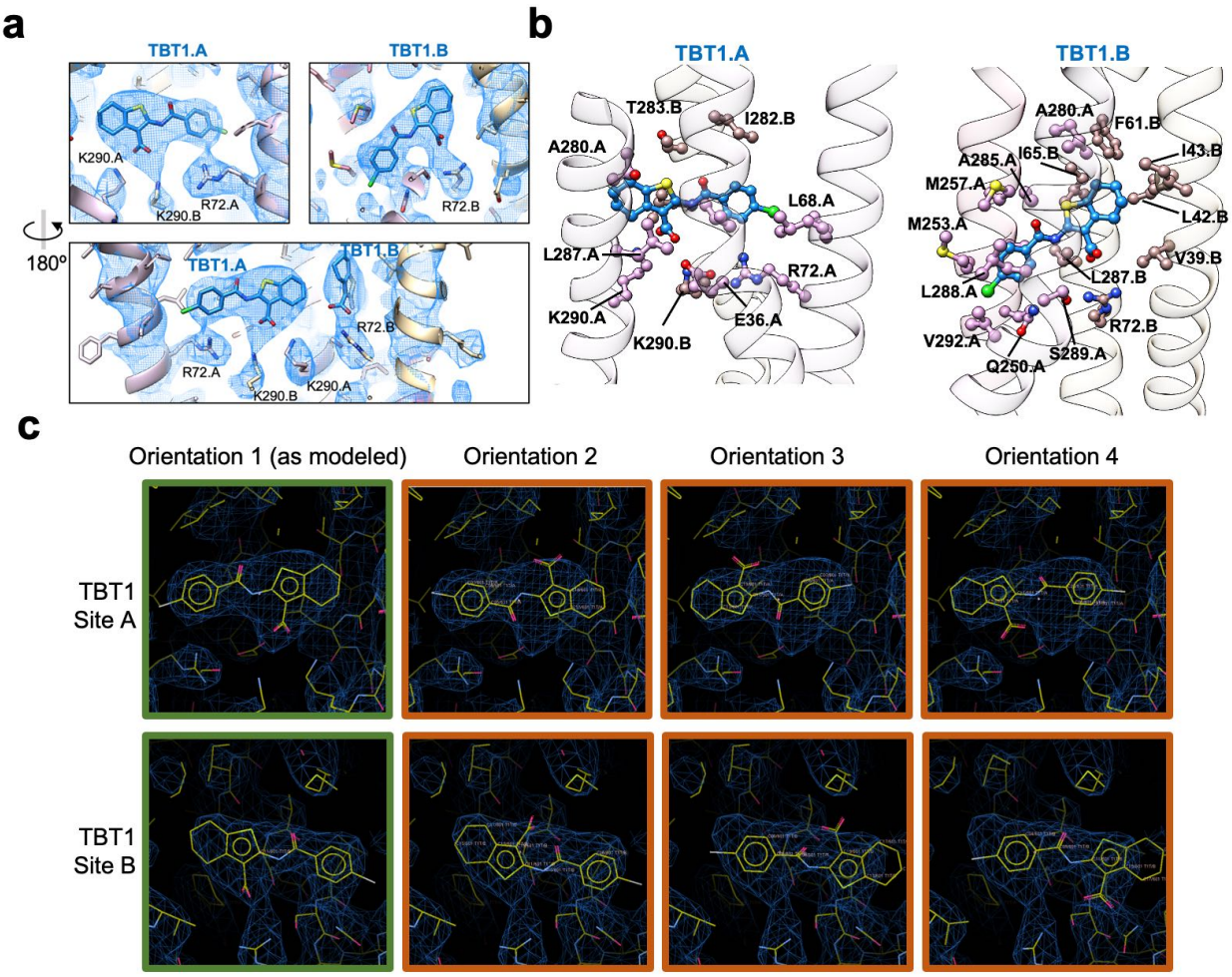


Figure 3.9: Molecular details of TBT1 binding to *A. baumannii* MsbA. a, Superimposition of cryo-EM map and model of the TBT1 binding site. **b,** Binding pocket for each asymmetric TBT1 ligand. **c,** Several TBT1 orientations were fit into each density using the real space refine and regularize functions in Coot. The orientations that were used in the final model are shown with a green outline and are most compatible with the cryo-EM densities.

Our results suggest that the carboxyl group of TBT1 is analogous to the phosphate groups on LPS glucosamines, both forming electrostatic interaction with MsbA. We thus sought to mutate Arg72 and Lys290 to determine whether MsbA retains sensitivity to TBT1-induced ATPase stimulation. *A. baumannii* MsbA with R72A, K290A, or R72A/K290A mutation was purified as wild-type MsbA and reconstituted in nanodiscs (***Figure 3.10a***). All MsbA mutants demonstrated largely diminished TBT1-induced stimulation compared to the wild-type protein, highlighting the importance of electrostatic interactions in TBT1 binding (***Figure 3.8c*** and ***Figure 3.10b,c***). Consistently, chemical modification of the carboxyl group of TBT1 renders the compound incompetent for MsbA inhibition[23]. Although the electrostatic interactions described herein may be the main drivers of TBT1 binding, hydrophobic contacts through the ring structures of TBT1 also participate in stabilizing the inhibitor. For instance, L68F and L150V mutants, which were previously shown to confer resistance to TBT1[23], are in the vicinity of the TBT1 binding sites (***Figure 3.10d***). While not directly adjacent to TBT1, it is conceivable that these mutations allosterically alter the pocket and prevent efficient inhibitor binding.

Despite striking resemblance in LPS and TBT1 recognition by MsbA, differences in TM positioning lead to divergence in ligand binding (***Figure 3.8e***). LPS and TBT1 stabilization both involve TM2 (Arg72) and TM6 (Lys290), although unlike LPS, TBT1 does not clearly interact with TM3. Notably, the two TBT1 copies are a distorted mimic of LPS, with decreased distance between TBT1 carboxyl groups compared to LPS glucosamine phosphates (***Figure 3.8e***). Upon TBT1 binding, TMD2 (TM1,2,3,6.A) undergoes significant changes: TM6.A moves into the center of the inner cavity, presenting Lys290.A for interaction with the hydroxyl group of TBT1.A and dragging the connected NBD.A towards the central axis (***Figure 3.8e***, bottom). TMD2 then collapses around TM6.A, with TM2.A moving closer to TM6.B. It is conceivable that the

convergence between TM2.A and TM6.B would then push TM4-TM5.B laterally. The unsolvable contradiction between NBD.A moving towards the central axis and TM4-TM5.B being pinched outwards results in the dissociation of the CH2.B coupling helix from NBD.A and to the unusually disordered domain-swapped helices of the collapsed inward-facing MsbA conformation.

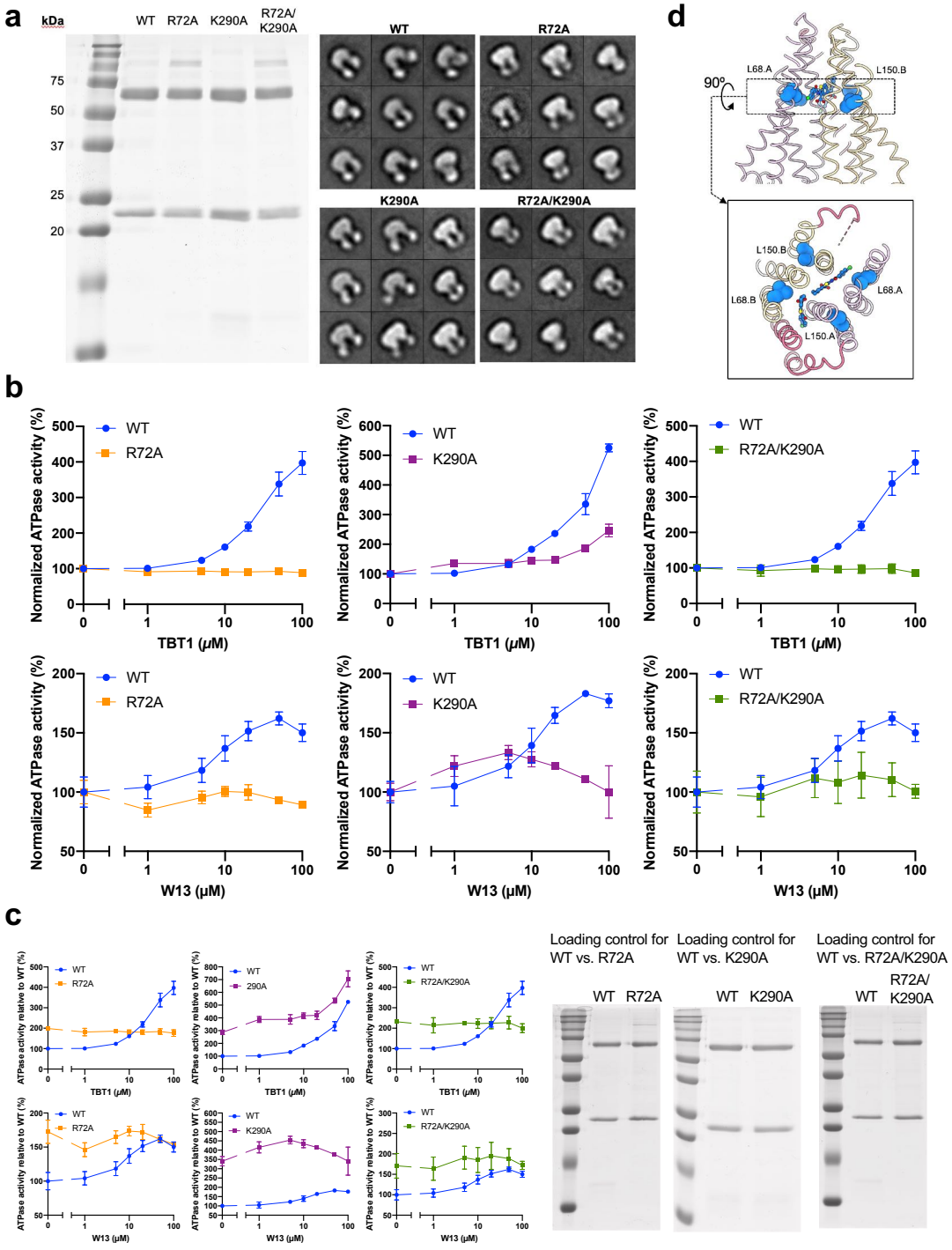


Figure 3.10: Analysis of *A. baumannii* MsbA mutants and effect on TBT1 and W13-induced ATPase stimulation. a, SDS-PAGE gel (left) and negative-stain EM (right) of purified wild-type and mutant MsbA in nanodiscs. *A. baumannii* MsbA runs at ~65 kDa and MSP1D1 at ~22.5 kDa. Box size for the negative-stain averages is 215 Å. **b,** ATPase activity of wild-type and mutant MsbA at different TBT1 (top row) and W13 (bottom row) concentrations. The activity of each

Figure 3.10 (continued)

protein sample was normalized relative to its basal activity in the absence of compound. Error bars correspond to mean ± SD (n=3). The data points for wild-type MsbA activity in the R72A and R72A/K290A panels correspond to the same measurements and are therefore identical in the leftmost and rightmost panels. **c,** Same data as in panel d, normalized relative to the activity of wild-type MsbA in the absence of compounds (left). Notably, mutants seem to have higher basal activity than wild-type, and desensitization to TBT1 or W13 is better seen when normalizing to each mutant as in panel b. Based on a standard phosphate curve, we estimate that less than 10% of ATP molecules were converted to ADP when the reactions were stopped. SDS-PAGE gels used to confirm that a similar amount of protein was loaded in each ATPase reaction are shown on the right-hand side. **d,** Location of mutations (blue spheres) previously found to prevent TBT1-induced ATPase stimulation and LPS transport blockage.

To our knowledge, inhibition by TBT1 represents the first example of a small molecule inhibitor targeting the central binding site of a non-multidrug ABC transporter. Because most ABC transporters have relatively narrow substrate spectra, our findings suggest that mimicking substrate binding can be a generally applicable strategy for developing small molecule modulators for all ABC transporters. To test if the central pocket is useful for structure-based compound discovery, we performed an *in silico* chemical library screen against the TBT1 induced-fit pocket. We screened ~800,000 compounds and determined that one molecule, W13, stimulated ATPase activity with an EC_{50} lower than TBT1 (~5.5 μM vs 13 μM, ***Figure 3.8f*** and ***Figure 3.11b***). Notably, the same mutants that were insensitive to TBT1 also experience lower W13-induced stimulation than wild-type protein (***Figure 3.10b,c***). W13 is structurally dissimilar to TBT1 and considerably larger in size (499 Da vs. 336 Da), and a single W13 molecule is expected to occupy the central MsbA pocket (***Figure 3.11c,d***). Molecular dynamics simulation of W13-bound MsbA in a POPC lipid bilayer suggests a reminiscent binding mode to TBT1, including hydrophobic groups of W13 extending upward in the hydrophobic pocket and carbonyl groups of W13 near TBT1-recognizing arginine and lysine residues (***Figure 3.11d***).

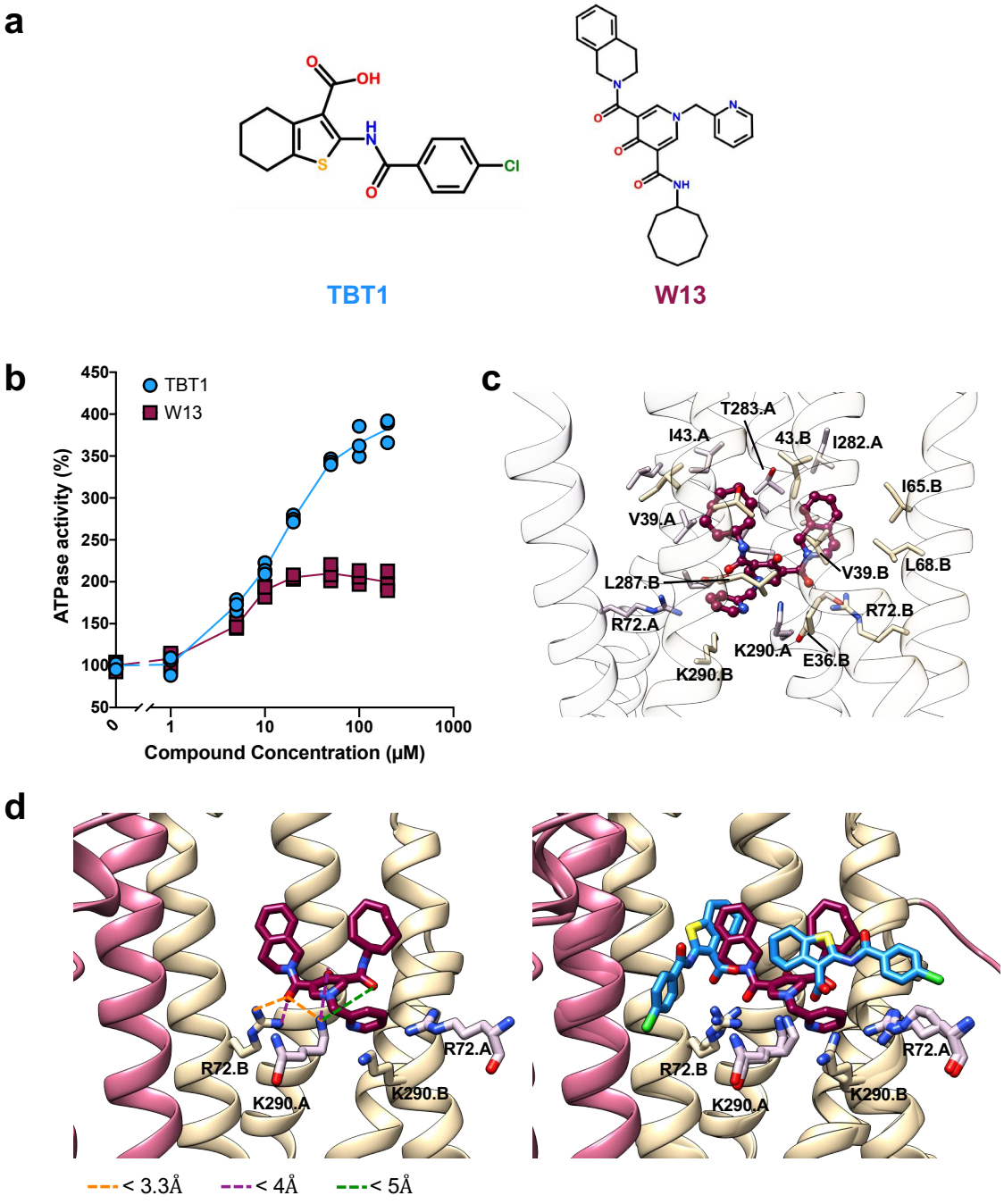


Figure 3.11: Predicted binding of W13 to *A. baumannii* MsbA. a, Molecular structures of TBT1 and W13. **b,** Effect of TBT1 and W13 on MsbA ATPase activity at different compound concentrations. **c,** W13 binding pocket in the central MsbA cavity, as predicted by molecular dynamics simulations. W13 is shown maroon using ball-and-stick representation. Neighboring residues are displayed as sticks with coloring matching their chain assignment. **d,** Predicted W13 fit, with neighboring lysine and arginine residues and distances between sidechains and compound (left). Superimposition of TBT1 (blue) and W13 (maroon) binding pockets (right).

3.2.4 Search for other compounds targeting the TBT1 pocket

W13 is the only ATPase stimulator identified in our *in silico* and ATPase screens. Yet, this screening strategy resulted in identification of several more interesting MsbA ATPase modulators. Our efforts at identifying MsbA inhibitors are summarized next:

Since two copies of TBT1 are asymmetrically bound in the MsbA central pocket, we reasoned that one strategy would consist of covalently tethering two TBT1 molecules, which could increase local TBT1 concentration once the linked molecule enters the MsbA central pocket, possibly lowering TBT1 EC_{50}. Lacking experience in chemical synthesis, we were limited to ordering from commercial compound libraries and could only find one compound resembling linked TBT1 (L-TBT1, see ***Figure 3.12a***). L-TBT1 had no effect on MsbA ATPase activity, even at very high concentrations (***Figure 3.12a***, right). Interestingly, L-TBT1 does not have a complete carboxylic acid tail as in TBT1, and thus (similarly to TBT2 in Zhang et al.[23]) cannot properly engage the central pocket arginines and lysines. Although or attempt at creating a functional tethered TBT1 has been unsuccessful, in-house development of linked molecules with intact carboxylic acid tails may prove to be a fruitful strategy in the future.

Unlike W13, most ATPase modulators identified in our screen were ATPase suppressors rather than stimulators (see W4, W26, W34 in ***Figure 3.12b***). The ATPase inhibitors described here have very different sizes and molecular scaffolds, which may be expected considering that ABC transporters can recognize diverse small molecules. It is particularly interesting that the same central pocket can be targeted to both stimulate (e.g., W13) and inhibit (e.g., W4) MsbA ATPase activity. Further efforts are required to understand how MsbA discriminates between transported substrates (e.g., LPS), non-transported ATPase stimulators (e.g., TBT1) and ATPase inhibitors (e.g., W4).

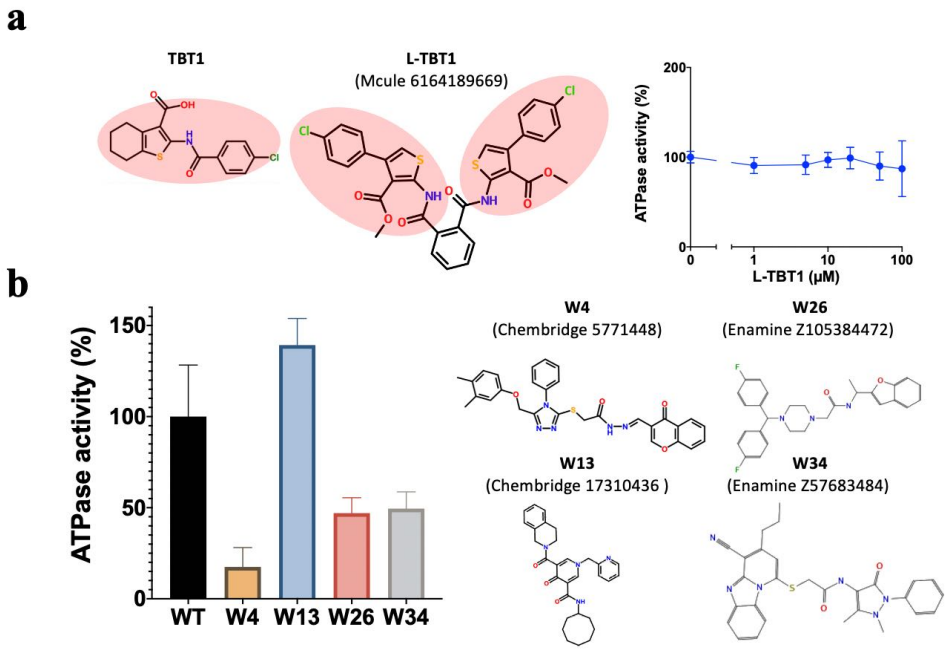


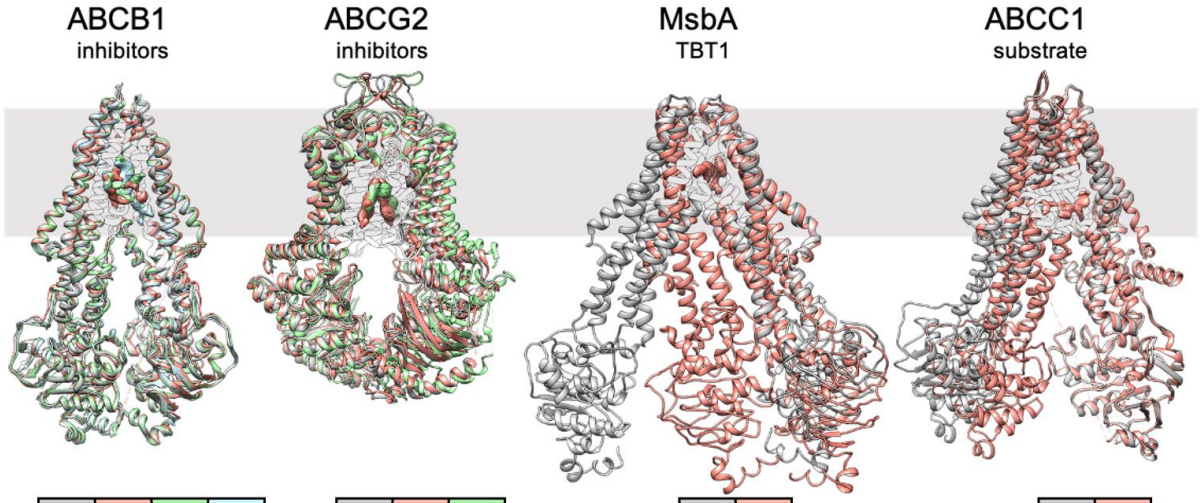
Figure 3.12: ATPase activity modulation by small molecules targeting the central pocket. a, Left, molecular structures of TBT1 and L-TBT1. Right, MsbA ATPase activity at increasing L-TBT1 concentrations. **b,** Left, MsbA activity when exposed to 10 μM of W4, W13, W26 and W34. Right, molecular structures of compounds tested.

3.3 Discussion

The peculiar effect on MsbA conformation and the inhibition mechanism of TBT1 are drastically different from those of small molecule modulators for other ABC transporters, which typically induce more modest conformational changes (e.g., ABCB1 and ABCG2 in ***Figure 3.13***). Interestingly, the narrowed inter-NBD spacing of TBT1-bound MsbA is more akin to the NBD tightening observed when the human multidrug transporter ABCC1 is exposed to its endogenous leukotriene substrate[1] (***Figure 3.13***), further suggesting that TBT1 triggers substrate recognition-like behavior in MsbA by mimicking the much larger LPS substrate. Yet unlike the authentic ligand, TBT1 breaks the symmetry of the MsbA homodimer by destabilizing one domain-swapping helix bundle, in an unprecedented mechanism in ABC transporter family. It is noteworthy that uncoupling of ATPase function and substrate transport by TBT1 occurs through the domain-swapped coupling helix, which is a landmark feature of the type IV exporter fold[29]. Thus, small molecule decouplers acting like TBT1 may be identified for many other type IV exporters such as ABCB1 and ABCC1.

TBT1 face severe drawbacks preventing its application in clinical settings, including low binding affinity to MsbA[23]. Yet compound-bound ABC transporter structures are essential tools for discovering new inhibitors, because molecules such as TBT1 reveal unexpected binding pockets by induced-fit. For example, convenient identification of W13 as an ATPase stimulator is very unlikely if virtual screening was performed against the structure of wide inward-facing, drug-free *A. baumannii* MsbA instead of TBT1-bound MsbA. We therefore hope that these ligand-bound MsbA structures will guide the development of novel classes of antibiotics targeting MsbA. One promising direction for compound optimization may consist of tethering together two TBT1-like molecules, such that inhibitor local concentration is increased upon entering the central pocket.

Another worthwhile undertaking may consist in generalizing the LPS-mimicking inhibition strategy to other Gram-negative strains including antibiotic resistant pathogens. While TBT1 currently only works on *A. baumannii* MsbA, the inhibition mechanism described here should in principle extend to MsbA in other Gram-negative bacteria that presumably use a similar mechanism of LPS binding.



	Apo 7A65	Zos. 7A6F	Tar. 7A6E	Ela. 7A6C
Site 1	16.0Å	16.0Å	16.5Å	16.5Å
Site 2	21.9Å	21.2Å	21.3Å	20.5Å

	Apo 6VXF	MZ29 6FFC	Ima. 6VXH
Site 1	20.4Å	21.4Å	24.5Å
Site 2	20.4Å	21.6Å	24.3Å

	Apo	TBT1
Site 1	46.8Å	19.2Å
Site 2	46.6Å	21.8Å

	Apo 5UJ9	LTC4 5UJA
Site 1	34.8Å	24Å
Site 2	42.9Å	27.8Å

Figure 3.13: Comparison of various cryo-EM structures of apo and compound-bound ABC transporters. Top: alignment of apo structures with inhibited and substrate-bound conformations. Ligands are shown as zoned densities from deposited cryo-EM maps. The multidrug efflux pump ABCB1 adopts a similar conformation whether in apo state (PDB: 7A65) or inhibited by zosuquidar ("zos.", PDB: 7A6F), tariquidar ("tar.", PDB: 7A6E) and elacridar ("ela.", PDB: 7A6C). ABCG2 (PDB: 6VXF) undergoes modest conformational changes in complex with its inhibitors MZ29 (PDB: 6FFC) and imatinib ("ima.", PDB: 6VXH). In contrast, TBT1-bound MsbA undergoes a large-scale conformational change, which is reminiscent of NBD tightening when ABCC1 (PDB: 5UJ9) is exposed to its LTC4 endogenous substrate (PDB: 5UJA). Bottom: Distance between Cα signature motif serine and Walker A glycine for each ATP binding site. For ABCB1, site 1 is S532/G1075 and site 2 is G432/S1177. For ABCG2, S187/G85 is used. For MsbA, S474/G373. For ABCC1, site 1 is S769/G1329 and site 2 is S1430/G681.

	A. baumannii TBT1 In nanodisc	*A. baumannii* Apo In nanodisc
	PDBid: 7MET EMDB: 23803	PDBid: 7RIT EMDB: 23804
Data collection and processing		
Microscope	UMass Titan Krios	UMass Titan Krios
Detector	K3	K3
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron Exposure ($e^-/Å^2$)	52	52
Defocus Range (μm)	1.0 – 2.5	1.0 – 2.5
Pixel Size (Å)	1.06	1.06
Final symmetry imposed	C1	C2
Initial particle images (#)	887,203	476,112
Final particle images (#)	67,714	25,417
Map resolution (Å)	3.97 TMD 4.28 overall	5.1 TMD 5.2 overall
FSC threshold	0.143	0.143
Map resolution range (Å)	3.5 – 8.0	4.5 – 10.0
Refinement		
Initial model used (PDBid)	5TV4	5TV4
Model resolution (Å)	4.16	6.43
FSC threshold	0.5	0.5
Map sharpening B factor ($Å^2$)	N/A	N/A
Model composition		
Non-hydrogen atoms	6557	5616
Protein residues	1019	1138 (all polyA)
Ligand	2	0
B factor ($Å^2$)		
Protein	86.07	223.14
Ligand	35.35	
Model vs. Map		
TMD CC	0.75	0.82
Overall CC	0.69	0.77
R.m.s. deviations		
Bond lengths (Å)	0.003	0.002
Bond angles (°)	0.580	0.472
Validation		
MolProbity score	1.77	1.58
Clashscore	8.08	3.15
Poor rotamers (%)	0	
Ramachandran plot		
Favored (%)	95.26	92.24
Allowed (%)	4.24	7.41
Disallowed (%)	0.49	0.35

Figure 3.14: Data collection table and model refinement statistics

3.4 Methods

Cloning, expression, and purification of MsbA

The gene encoding for *A. baumannii* MsbA was amplified from genomic DNA (ATCC strain 2208, Ref 19606) and cloned into pET-28a vector with an additional N-terminal His tag. *A. baumannii* MsbA alanine mutants were introduced by Quikchange (Agilent) mutagenesis. The MsbA construct was transformed in *E. coli* BL21 DE3 and grown in Terrific Broth medium to OD_{600} ~1.0 at 37°C, before induction with 1mM β-D-1-thiogalactopyranoside (IPTG) and overnight growth at 17°C. Cells were subsequently harvested by centrifugation, resuspended in Buffer A (50 mM Tris pH 7.8, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP), and lysed with a probe sonicator or with a microfluidizer. Lipid membranes were harvested by 75min ultracentrifugation at 185,000 xg at 4°C, frozen in liquid nitrogen and stored at -80°C for future use. Membranes were thawed in buffer A + 1% (w/v) n-dodecyl-β-D-maltopyranoside (DDM), resuspended with a Dounce homogenizer and stirred for ~1 hour at 4°C. Insoluble debris were harvested by centrifugation at ~10,000 xg for 15min, followed by 185,000 xg for 30min. The resulting cell lysate was used for affinity purification and flowed over a TALON cobalt affinity resin (Takara Bio). The cobalt resin was washed with ~20 column volumes (CV) of buffer A + 0.1% DDM + 10 mM imidazole, and the protein was subsequently eluted with 4CV of buffer A + 0.1% DDM + 250 mM imidazole. The MsbA sample was then concentrated on a 100 kDa cutoff concentrator and subjected to size-exclusion chromatography on a Superdex 200 column in a 1:1 buffer A water dilution + 0.05% DDM. Peak fractions were pooled and concentrated to ~10 mg/mL for nanodisc reconstitution. MsbA was incubated with membrane scaffold protein MSP1D1 and palmitoyl-oleoyl-phosphatidylglycerol (POPG) lipids at 1:1:60 stoichiometry in size-exclusion buffer + 20 mM sodium cholate. Detergent was removed using ~65% (w/v) biobeads and the sample was again

subjected to size-exclusion in 25 mM Tris, 150 mM NaCl, 0.5 mM TCEP. The MsbA nanodisc samples were then concentrated and used for ATPase and cryo-EM experiments.

ATPase activity assay

MsbA activity was assessed as previously described[6,20,34]. Briefly, TBT1 (DC chemicals) and W13 (ChemBridge corporation ID 17310436) were solubilized and DMSO and used at final DMSO concentration <2%. 1 µg MsbA was pre-incubated in 4°C nanodisc buffer with varying (TBT1, W13) inhibitor concentration, then mixed with 2 mM ATP + 2 mM $MgCl_2$. The ATPase reaction was stopped after ~20min at 37°C using 12% sodium dodecyl-sulfate (SDS). γ-phosphate dissociation was assessed in a colorimetric assay by addition of a 6% ascorbic acid (w/v), 1% (w/v) ammonium molybdate in 1N HCl, succeeded by mixing with 2% sodium citrate (w/v), 2% (w/v) sodium meta-arsenite, 2% acetic acid (v/v) in 1N HCl. Absorbance at 850 nm was determined using a Flexstation III Molecular Devices spectrophotometer and data was analyzed on GraphPad Prism 8.

Negative-stain EM

Protein samples were assessed for homogeneity by negative-stain EM. All samples were stained using a 1.5% (w/v) uranyl formate solution and imaged on a Tecnai T12 electron microscope (FEI), operated at 120kV and at 67,000x magnification.

Cryo-EM sample preparation and data collection

The *A. baumannii* MsbA nanodisc samples were concentrated to ~1.5 mg/mL. Inhibitor-bound complexes were pre-incubated with 400 µM TBT1 for 2 hours at 4°C, whereas apo *A. baumannii* was directly used for sample preparation. 3 µl sample was applied to glow-discharged Quantifoil (R1.2/1.3, 400 mesh) holey carbon grids and blotted with a Thermo Fisher Vitrobot Mark IV (blot ~7s, force +12) before vitrification in liquid ethane. Apo *A. baumannii* and TBT1-bound *A.*

baumannii MsbA were collected on a Titan Krios equipped with a Gatan K3 camera and a BioQuantum imaging filter. Movie stacks were acquired in super-resolution mode with 0.53 Å pixel size and total exposure of 51.6 e/Å^2. A slit width of 20 eV for energy filter was set during the data collection. SerialEM[35] was used for collection of all datasets. More information about cryo-EM data collection parameters can be found in ***Figure 3.14***.

Cryo-EM data processing

The data was analyzed similarly to previous datasets[6,20,34]. Briefly, movie stacks were binned 2x and motion-corrected with MotionCor2[36], and contrast transfer function was computed with CTFFIND4[37]. All 2D processing was conducted using Simplified Application Managing Utilities of EM Labs (SAMUEL)[6], which relies on the SPIDER[38] image processing system. Briefly, ~2,000 protein images were manually picked to generate initial 2D templates. Auto-picking on 10% of movie stacks were then used to create refined 2D templates, then used to pick particles on the whole dataset. 2D particles were screened using "samtree2dv3.py", which successively runs PCA, *k*-means clustering, and multireference alignment to classify particles. Selected particles after 2D screening were used for 3D processing with RELION3.0[39], including 3D classification, domain masking, signal subtraction[40], and 3D particle refinement. Throughout classification, 3D classes were combined or excluded based on the quality of the transmembrane helices. Overall resolution of cryo-EM map was computed according to the gold standard FSC method, with reconstruction of independent halves of the dataset and determination of resolution using FSC=0.143 criterion, whereas local resolution was estimated with ResMap[41]. All maps were subjected to density modification[42] in phenix, with local sharpening and blurring. No model was provided to the density modification program.

Model building and validation

A homology model of *A. baumannii* MsbA was first created with SWISSMODEL[43] from LPS-bound MsbA (pdb: 5TV4) as input, then manually fit into cryo-EM densities using COOT[44]. TBT1 model and restraints were manually built using ELBOW[45] then fit into the density. The models were then refined against the cryo-EM map using Phenix[46] real space refinement and validated with MolProbity[47]. All figures were generated with UCSF Chimera[48] or ChimeraX[49].

Structure-based virtual screening and molecular dynamics

A two-step virtual screen was performed to identify compounds binding to the TBT1 pocket. First, flexible docking calculations were carried out for the ChemBridge library (~800,000 compounds) with AutoDock Vina[50]. The center of the TBT1 pocket was selected as coordinate origin, and the docking box set to a 25Å size. Other parameters of Vina were set as default. The next step involved re-scoring of the top 1% docked compounds by an in-house protein-ligand interaction similarity protocol. The BioLip libary[51], which curates biologically relevant ligand-protein binding interactions, was used to search for similar template pockets as the TBT1 site with PPS-align (availability: https://zhanglab.ccmb.med.umich.edu/PPS-align/). A weighted Jaccard score combining the pocket similarity and ligand similarity was used to rank docked compounds. The top 40 compounds were selected for further validation, and 17 compounds were ordered for experimental testing according to availability. Once W13 was identified as a hit, its binding pose was verified by molecular dynamics (MD) simulation at 310 K and 1 atm. MsbA were embedded into a lipid bilayer consisted of 215 POPG molecules and then solvated into a water box consisted of 28,395 TIP3P water and 0.15 M NaCl. Independent simulations in POPC lipid bilayer or 75%POPE/25%POPG bilayer were also carried out and consistent results were obtained. The simulation system included 130,779 atoms and had a dimension of 92×92×165 Å. The force field

parameters of W13 ligand were generated by CGenFF[52], and the protein was modeled by the CHARMM36m force field[53]. System setup and analysis were performed with CHARMM[54] and MD simulations were carried out using OpenMM[55].

3.5 References

1. Johnson, Z.L. & Chen, J. Structural Basis of Substrate Recognition by the Multidrug Resistance Protein MRP1. *Cell* **168**, 1075-1085 e9 (2017).

2. Taylor, N.M.I. et al. Structure of the human multidrug transporter ABCG2. *Nature* **546**, 504-509 (2017).

3. Nosol, K. et al. Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter ABCB1. *Proc Natl Acad Sci U S A* **117**, 26245-26253 (2020).

4. Zhang, Z. & Chen, J. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. *Cell* **167**, 1586-1597 e9 (2016).

5. Li, N. et al. Structure of a Pancreatic ATP-Sensitive Potassium Channel. *Cell* **168**, 101-110 e10 (2017).

6. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

7. Qian, H. et al. Structure of the Human Lipid Exporter ABCA1. *Cell* **169**, 1228-1239 e10 (2017).

8. Olsen, J.A., Alam, A., Kowal, J., Stieger, B. & Locher, K.P. Structure of the human lipid exporter ABCB4 in a lipid environment. *Nat Struct Mol Biol* **27**, 62-70 (2020).

9. Rees, D.C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218-27 (2009).

10. Locher, K.P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487-93 (2016).

11. Thomas, C. & Tampe, R. Multifaceted structures and mechanisms of ABC transport systems in health and disease. *Curr Opin Struct Biol* **51**, 116-128 (2018).

12. Srikant, S. & Gaudet, R. Mechanics and pharmacology of substrate selection and transport by eukaryotic ABC exporters. *Nat Struct Mol Biol* **26**, 792-801 (2019).

13. Szollosi, D., Rose-Sperling, D., Hellmich, U.A. & Stockner, T. Comparison of mechanistic transport cycle models of ABC exporters. *Biochim Biophys Acta Biomembr* **1860**, 818-832 (2018).

14. Shukla, S., Ohnuma, S. & Ambudkar, S.V. Improving cancer chemotherapy with modulators of ABC drug transporters. *Curr Drug Targets* **12**, 621-30 (2011).

15. Plummer, A.M., Culbertson, A.T. & Liao, M. The ABCs of Sterol Transport. *Annu Rev Physiol* (2020).

16. Lewis, V.G., Ween, M.P. & McDevitt, C.A. The role of ATP-binding cassette transporters in bacterial pathogenicity. *Protoplasma* **249**, 919-42 (2012).

17. Alam, A., Kowal, J., Broude, E., Roninson, I. & Locher, K.P. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. *Science* **363**, 753-756 (2019).

18. Alam, A. et al. Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1. *Proc Natl Acad Sci U S A* **115**, E1973-E1982 (2018).

19. Jackson, S.M. et al. Structural basis of small-molecule inhibition of human multidrug transporter ABCG2. *Nat Struct Mol Biol* **25**, 333-340 (2018).

20. Orlando, B.J. & Liao, M. ABCG2 transports anticancer drugs via a closed-to-open switch. *Nat Commun* **11**, 2264 (2020).

21. Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-10 (2007).

22. Thelot, F., Orlando, B.J., Li, Y. & Liao, M. High-resolution views of lipopolysaccharide translocation driven by ABC transporters MsbA and LptB2FGC. *Curr Opin Struct Biol* **63**, 26-33 (2020).

23. Zhang, G. et al. Cell-based screen for discovering lipopolysaccharide biogenesis inhibitors. *Proc Natl Acad Sci U S A* **115**, 6834-6839 (2018).

24. Ho, H. et al. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196-201 (2018).

25. Alexander, M.K. et al. Disrupting Gram-Negative Bacterial Outer Membrane Biosynthesis through Inhibition of the Lipopolysaccharide Transporter MsbA. *Antimicrob Agents Chemother* **62**(2018).

26. Boucher, H.W. et al. Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America. *Clinical Infectious Diseases* **48**, 1-12 (2009).

27. Denisov, I.G. & Sligar, S.G. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem Rev* **117**, 4669-4713 (2017).

28. McLean, M.A., Gregory, M.C. & Sligar, S.G. Nanodiscs: A Controlled Bilayer Surface for the Study of Membrane Proteins. *Annu Rev Biophys* **47**, 107-124 (2018).

29. Thomas, C. et al. Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Lett* **594**, 3767-3775 (2020).

30. Higgins, C.F. & Linton, K.J. Structural biology. The xyz of ABC transporters. *Science* **293**, 1782-4 (2001).

31. Padayatti, P.S. et al. Structural Insights into the Lipid A Transport Pathway in MsbA. *Structure* **27**, 1114-1123 e3 (2019).

32. Angiulli, G. et al. New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins. *Elife* **9**(2020).

33. Wright, J., Muench, S.P., Goldman, A. & Baker, A. Substrate polyspecificity and conformational relevance in ABC transporters: new insights from structural studies. *Biochem Soc Trans* **46**, 1475-1484 (2018).

34. Li, Y., Orlando, B.J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB2FGC complex. *Nature* **567**, 486-490 (2019).

35. Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y. & Mastronarde, D.N. Software tools for automated transmission electron microscopy. *Nat Methods* **16**, 471-477 (2019).

36. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

37. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).

38. Shaikh, T.R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* **3**, 1941-74 (2008).

39. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).

40. Bai, X.C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S.H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**(2015).

41. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-5 (2014).

42. Terwilliger, T.C., Ludtke, S.J., Read, R.J., Adams, P.D. & Afonine, P.V. Improvement of cryo-EM maps by density modification. *Nat Methods* **17**, 923-927 (2020).

43. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).

44. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

45. Moriarty, N.W., Grosse-Kunstleve, R.W. & Adams, P.D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074-80 (2009).

46. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

47. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

48. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

49. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).

50. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-61 (2010).

51. Yang, J., Roy, A. & Zhang, Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res* **41**, D1096-103 (2013).

52. Vanommeslaeghe, K. et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **31**, 671-90 (2010).

53. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71-73 (2017).

54. Brooks, B.R. et al. CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-614 (2009).

55. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).

Chapter 4: G247 acts as symmetrical transmembrane domains wedge and prevents MsbA conformational cycling.

The data presented here is part of a preprint:

Thélot, F. A., Zhang, W., Song, K., Xu, C., Huang, J., & Liao, M. (2021). Distinct allosteric mechanisms of first-generation MsbA inhibitors. bioRxiv, 2021.2005.2025.445681. doi:10.1101/2021.05.25.445681

Contribution: F. Thelot performed molecular cloning, protein purification, cryo-EM sample preparation, cryo-EM data processing, model building and biochemistry experiments. M. Liao supervised the project.

4.1 Introduction

The previous chapter focused on the unusual mechanism by which TBT1 prevents LPS transport yet stimulates MsbA ATPase activity. The mechanism of action of TBT1 differs from that of G compounds[1], which is the only other class of MsbA-specific inhibitors and suppress both LPS transport and ATPase activity. We seek to understand how TBT1 and G compounds differentially affect MsbA activity yet both prevent efficient LPS transport.

G compounds were discovered by a research group at Genentech in a high-throughput chemical screen of ~3 million compounds against *E. coli* MsbA reconstituted in lipid nanodiscs[1]. Selection of MsbA ATPase inhibitors and subsequent in-cell assays resulted in identification of G592, a small molecule with bactericidal properties. Chemical synthesis efforts on G592 led to the discovery of G092, G247 and G907, with improved IC_{50} for MsbA and lower minimum inhibitory concentration (MIC) relative to the original compound. Two crystal structures of *E. coli* MsbA in complex with G907 and G092 demonstrate that G compounds generate by induced-fit a binding pocket enclosed by TMHs 4,5 and 6 and are mostly recognized through hydrophobic contacts. Although two G compounds are symmetrically bound in each MsbA homodimer, they trigger an allosteric conformational change in the NBDs characterized by asymmetric NBD tightening. These crystal structures suggest that the tightened asymmetric NBDs are uncoupled from the transmembrane domains and cannot properly hydrolyze ATP.

Although previous crystal structures provide useful insights into G compound inhibition, we reasoned that structural investigation by cryo-EM would be of interest for several reasons: 1) It is unclear how symmetrical drug binding triggers an asymmetric conformational change in MsbA, and how the asymmetric state results in impaired conformational cycling; 2) The extraordinary conformational flexibility of ABC transporters can result in different states being

captured depending on whether the protein is studied by crystallography[2-4] or cryo-EM[5,6]. Thus, cryo-EM study may reveal a different conformation from prior work; 3) Comparison of a cryo-EM structure with previous drug-free MsbA in nanodisc provides unbiased insight into G compound inhibition, because the protein is studied under identical conditions.

4.2 Results

We purified *E. coli* MsbA in detergent and nanodisc (***Figure 4.1***) and confirmed that G247, the most potent G compound, suppressed MsbA ATPase activity at nanomolar concentrations (***Figure 4.1d***). To ensure that MsbA conformation is consistent in different protein environments, we subsequently acquired datasets of G247-bound *E. coli* MsbA in POPG nanodisc (***Figure 4.2***) and DDM (***Figure 4.3***). G247 binding provided unexpected ordering of the usually flexible transmembrane domains, such that little 3D processing was required to achieve high-resolution structures (***Figure 4.2e*** and ***Figure 4.3a***). Transmembrane domains resolution changes from ~3.9 Å for the nanodisc dataset, down to 3.5 Å for the C2-averaged DDM reconstruction, allowing for unambiguous sidechains and drug density modeling.

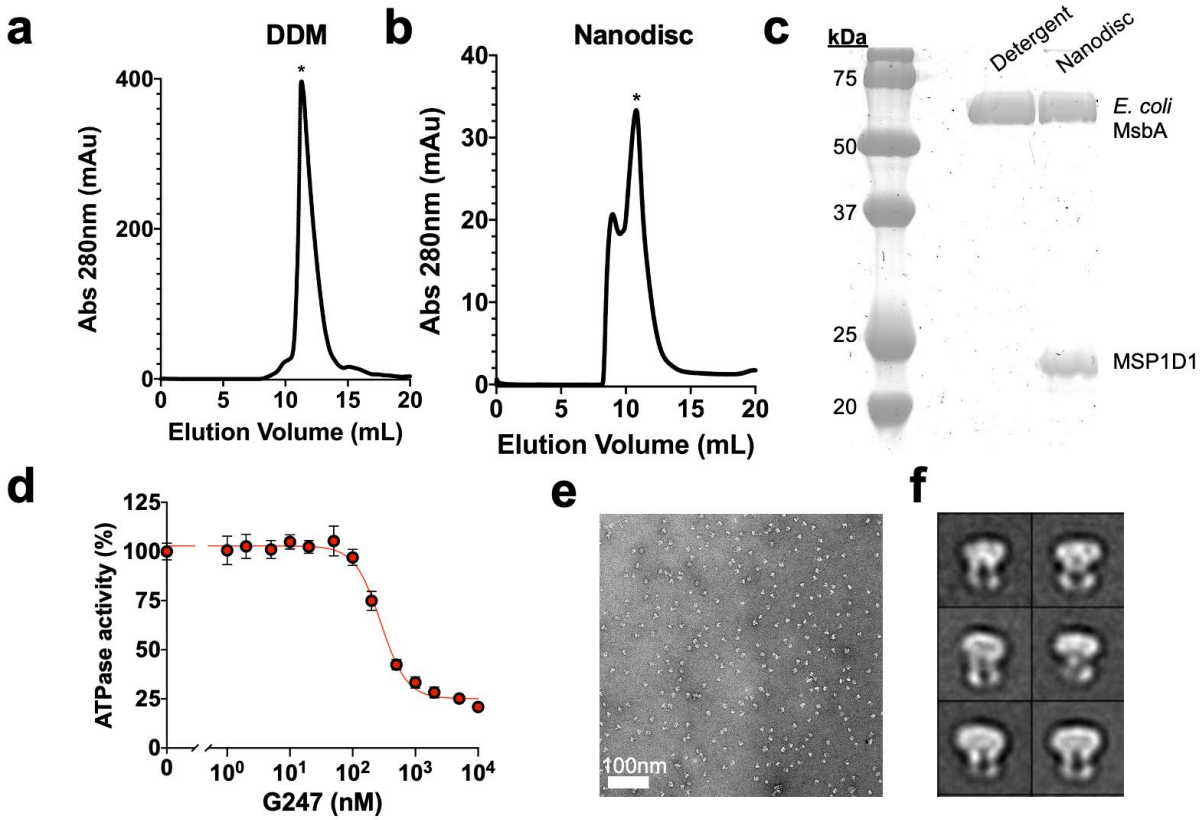


Figure 4.1: Characterization of *E. coli* MsbA. a, Gel filtration profile of *E. coli* MsbA in DDM detergent. **b,** Gel filtration profile of *E. coli* MsbA in nanodiscs. **c,** SDS-PAGE analysis of purified *E. coli* MsbA after gel filtration. **d,** ATPase activity of *E. coli* MsbA with increasing G247 concentration. Each point represents mean ± SD (n=3). **e,** Negative-stain EM image of *E. coli* MsbA. **f,** Negative-stain 2D class averages of *E. coli* MsbA in nanodiscs, with a 215 Å box size.

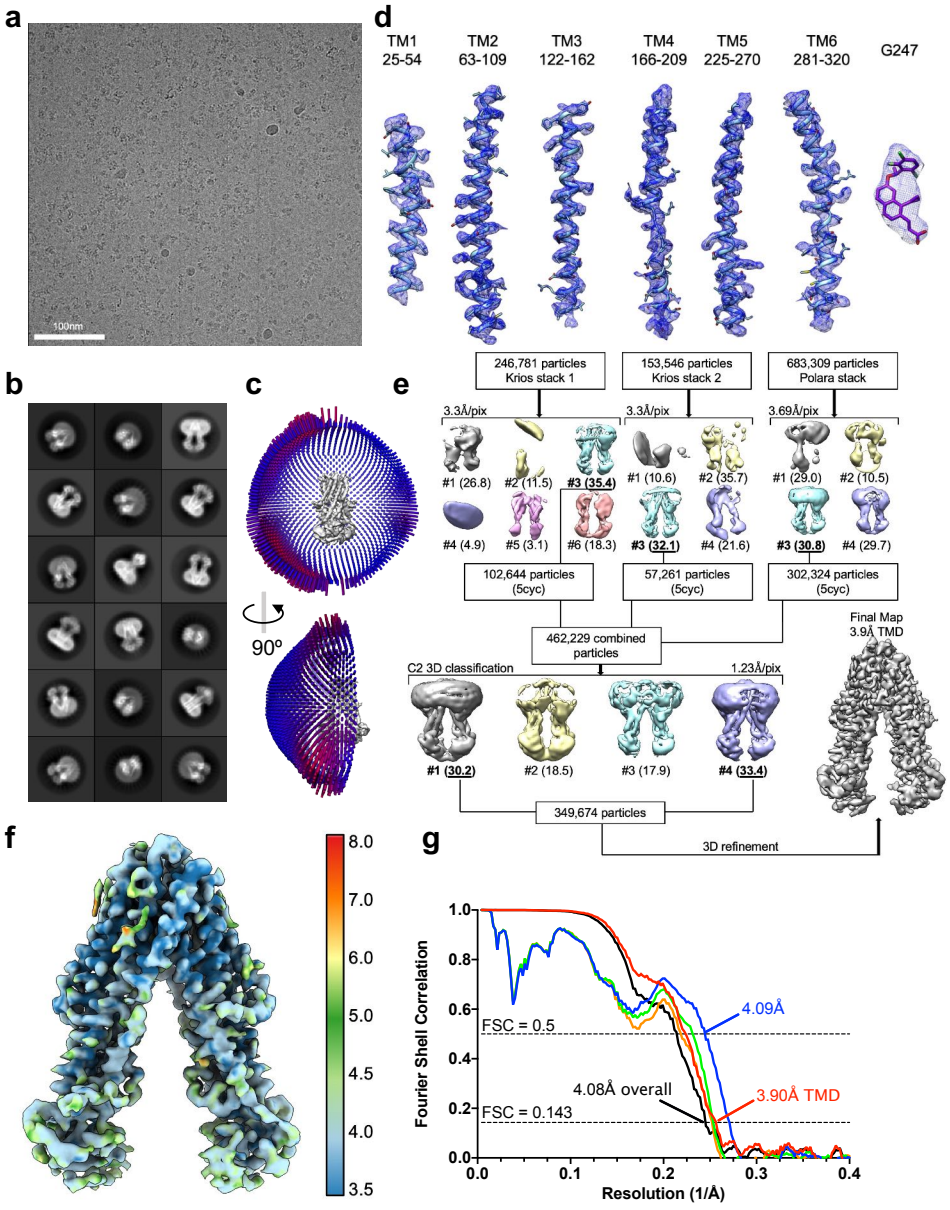


Figure 4.2: Cryo-EM imaging and validation for *E. coli* MsbA in nanodiscs in complex with G247. a, Representative cryo-EM image of *E. coli* MsbA in complex with G247. **b,** 2D class averages. Box size is 236 Å. **c,** Angle distribution of cryo-EM particles included in the final 3D reconstruction. **d,** Superimposition of cryo-EM density and model of TM helices and G247. **e,** Processing flowchart for MsbA in nanodisc in complex with G247. "5cyc" refers to the method of including any particle being assigned to a selected class within the final 5 cycles in 3D classification. The second round of 3D classification and final refinement were performed with C2 symmetry because no apparent asymmetric classes were present in the dataset. **f,** Local resolution of the final cryo-EM map. **g,** Fourier Shell Correlation (FSC) curves of the final reconstruction. Half map #1 vs. half map #2 for the entire MsbA molecule is shown in black. The remaining FSC curves were calculated for TMDs only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange).

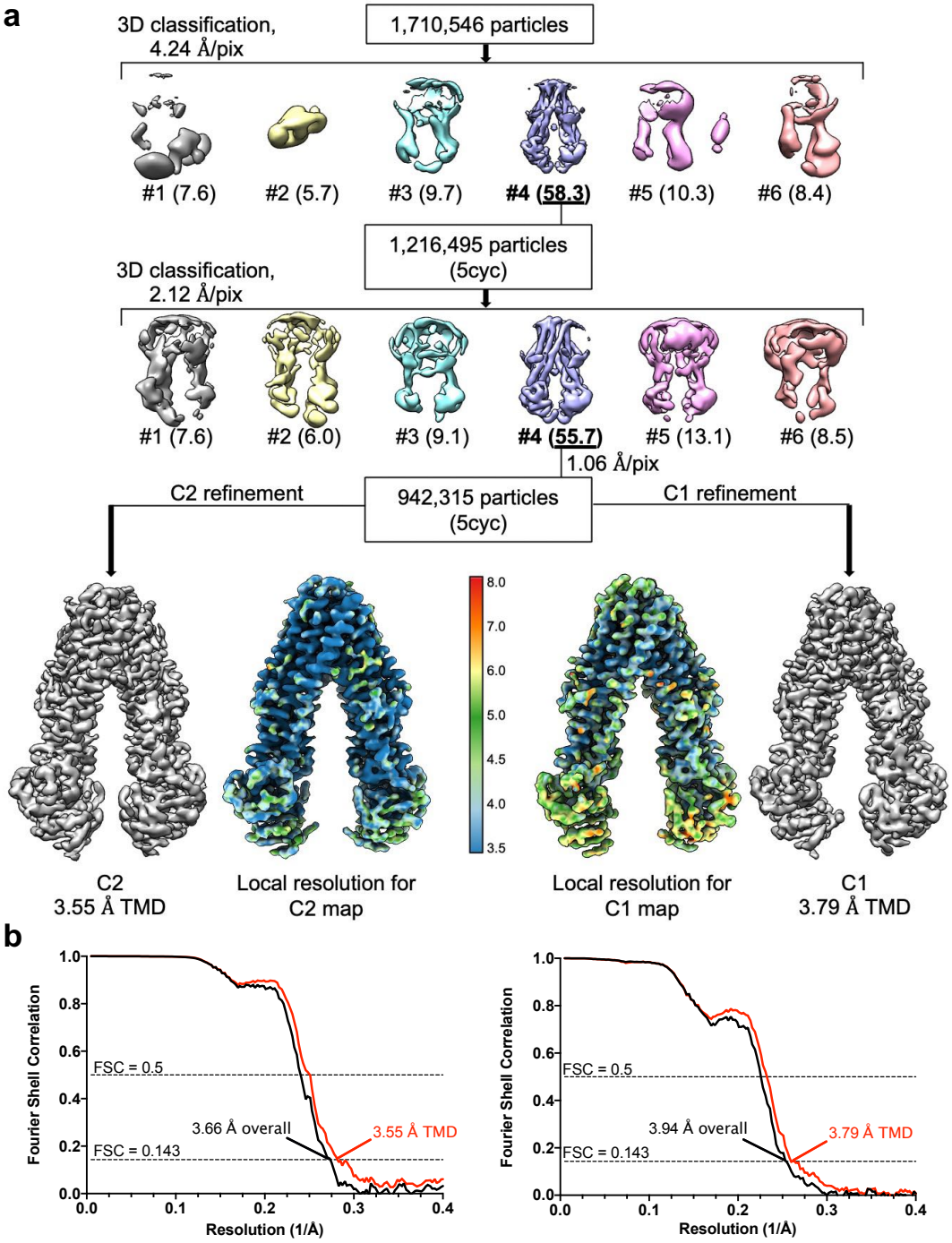


Figure 4.3: Cryo-EM imaging and validation for *E. coli* MsbA in DDM in complex with G247. a, Processing flowchart for MsbA in DDM in complex with G247. "5cyc" refers to the method of including any particle being assigned to a selected class within the final 5 cycles in 3D classification. The final C1 and C2 maps are nearly identical, indicating that G247 induces an open symmetrical MsbA conformation. **b,** Fourier Shell Correlation (FSC) curves of the final C2 (left) and C1 (right) reconstructions. Half map #1 vs. half map #2 for the whole protein and for the TMDs are displayed in black and red, respectively.

Previous crystal structures of G compound-bound MsbA[1] have NBDs that are brought closer together compared to apo state and positioned asymmetrically with one NBD raised relative to the other (***Figure 4.4b***, right). In contrast, our cryo-EM structures in nanodisc and detergent exhibit clear C2 symmetry, even when refined without symmetry constraints (***Figure 4.4b***, right and ***Figure 4.5b,c***). Asymmetry in the crystal structures stems from Arg190 which forms a salt bridge with the acrylic acid tail of the G compounds only in one subunit of MsbA. In cryo-EM structures, Arg190 is best ordered in the higher-resolution DDM structure and appears too distant from G247 to form a salt bridge (***Figure 4.5e***). Furthermore, TM4,5,6 bundle is positioned differently relative to TM1,2,3 in the X-Ray and cryo-EM structures (***Figure 4.5f***), which ultimately impacts the spacing between coupling helices (CH1 and CH2) and NBD positioning. While CH1-CH2 spacing is reduced in X-Ray structures compared to the drug-free conformation, CH1-CH2 distance is symmetrically increased in the G247-bound state (***Figure 4.5g***). Notably, the closed asymmetric conformation of X-Ray structures is not observed in any round of 3D classification (***Figure 4.2e*** and ***Figure 4.3a***), we sought to determine if an asymmetric subset of particles could be identified using cryoSPARC variability analysis[7] (***Figure 4.6a***). Since the closed asymmetric conformation remained elusive despite our best efforts, we posit that the open C2-symmetric state is the predominant conformation of G247-bound MsbA in solution.

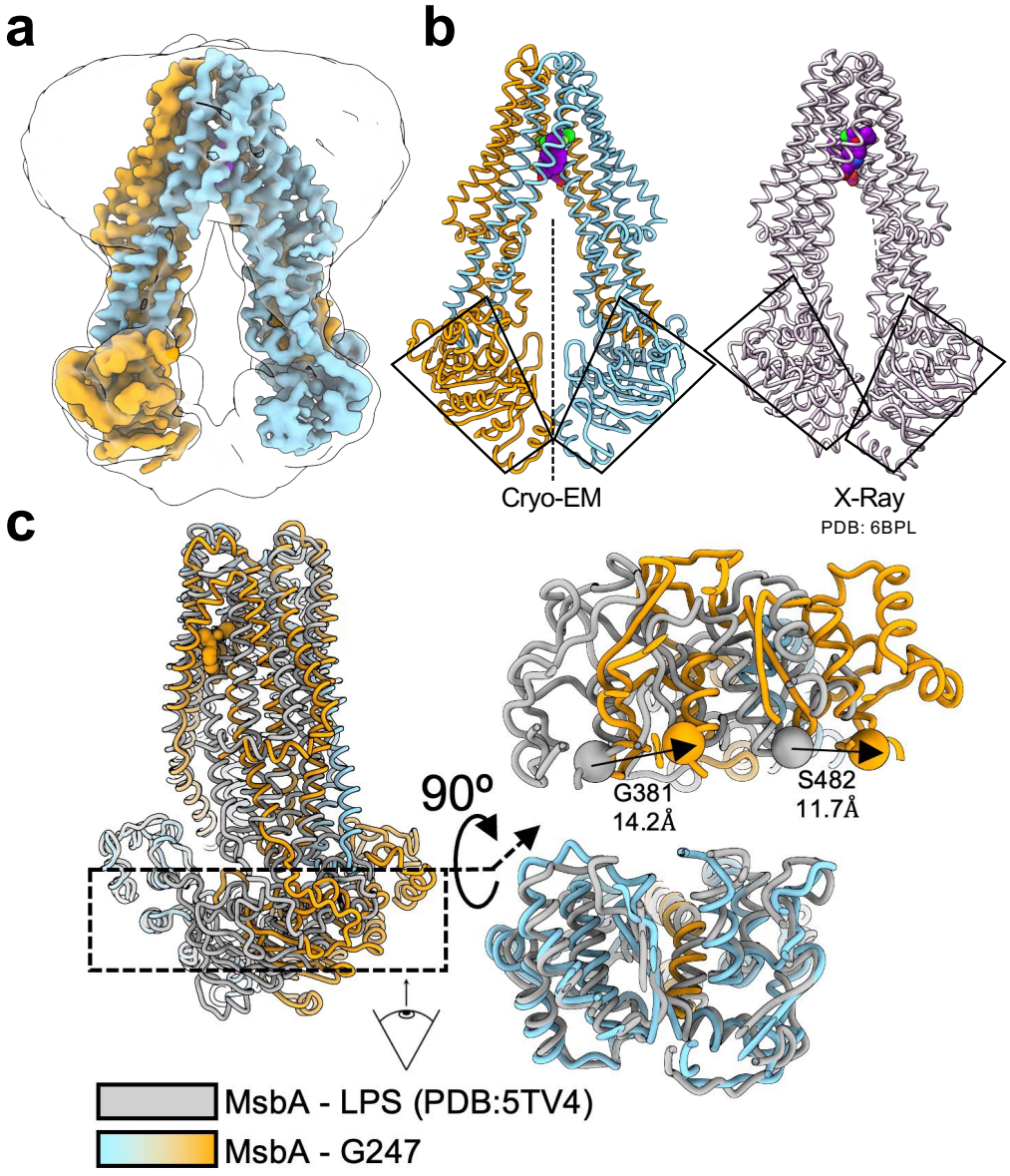


Figure 4.4: G247 symmetrically increases inter-NBD spacing. a, 3.9-Å resolution cryo-EM reconstruction of *E. coli* MsbA in complex with G247. The unsharpened map filtered at 6-Å resolution is displayed as outline to show the nanodisc. **b,** Comparison of MsbA structures determined by cryo-EM or X-ray crystallography (PDB: 6BPL). The cryo-EM structure adopts near-perfect C2 symmetry, while NBDs are asymmetrically positioned in the crystal structure. **c,** Comparison of the cryo-EM structures of drug-free (PDB: 5TV4, gray) and G247-bound (colored) MsbA, both in nanodiscs. G247 pushes NBDs away from each other, compared to the uninhibited state. The NBD shift is apparent when viewing MsbA from the side (left panel), and from the cytoplasmic space (right panel).

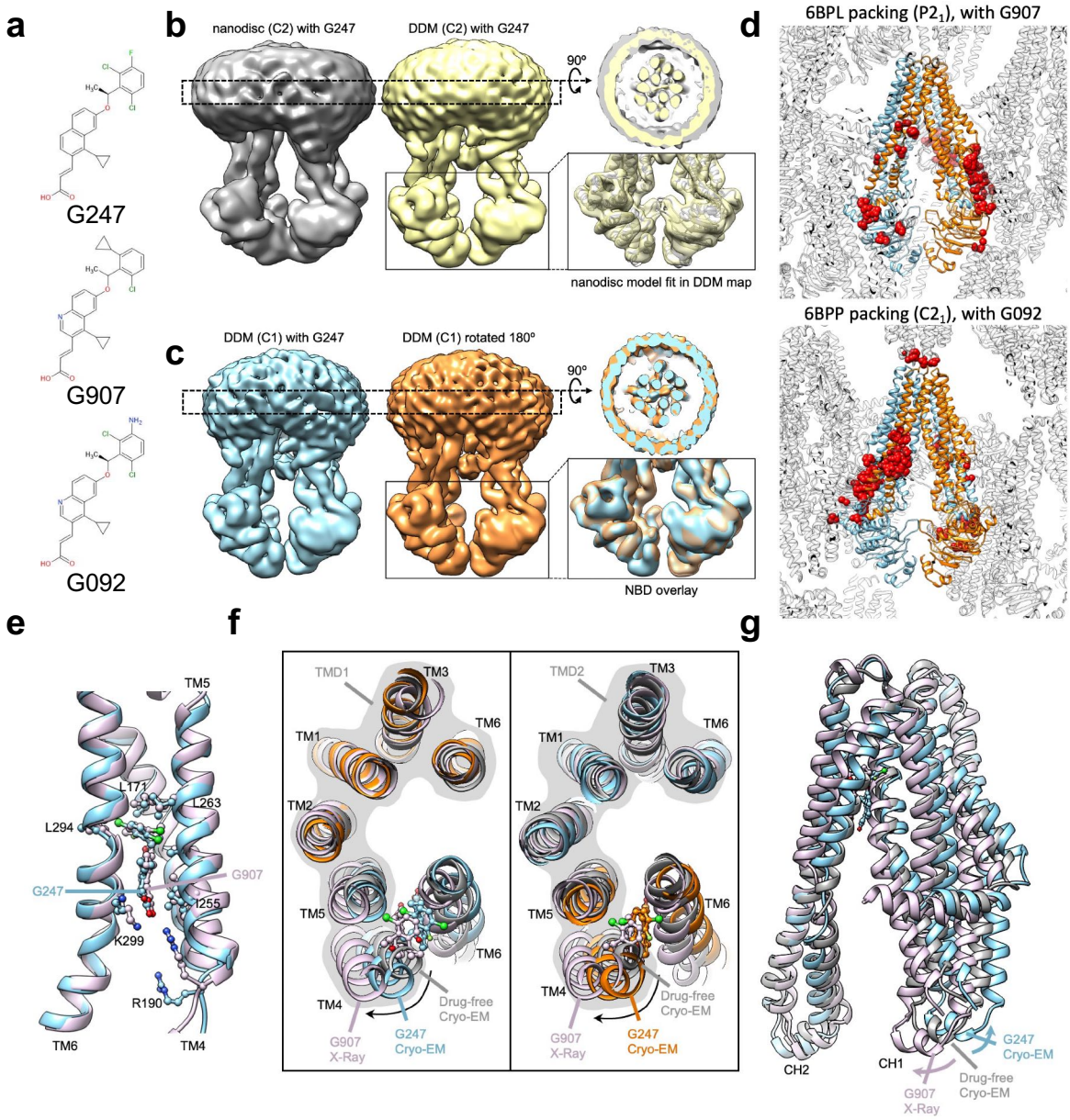


Figure 4.5: Comparison of the cryo-EM G247-bound open symmetrical conformation with prior studies. a, Molecular structure of G247 and of G compounds used in prior crystallization studies. **b,** Comparison of the nanodisc (gray) and DDM (yellow) C2 reconstructions, low-pass filtered to 6-Å. The cross-section through the TMDs (top right panel) shows that although the MSP1D1 nanodisc is wider than the DDM micelle, TMs positioning is identical. The nanodisc model was fit into the DDM reconstruction and matches the density well including in the more flexible NBDs (bottom right inset). **c,** The C1-refined DDM map (blue) is superimposed with the same map after 180° rotation (orange), showing nearly perfect symmetrical structure including in the NBDs. **d,** Cryo-EM G247-bound conformation (blue and orange) fit into the crystal packing from previous work (PDB 6BPL and 6BPP, colored gray). The cryo-EM model was aligned to the crystal structures using one TMD chain (residues 1-325). Steric clashes with surrounding MsbA

Figure 4.5 (continued)

copies are displayed as red spheres using default settings in UCSF chimera. **e,** Comparison of drug-binding pocket for G247 (blue) and G907 (pink, PDB:6BPL), as seen from the LPS binding site. Alignment performed on TM4,5,6. **f,** Helix positioning for drug-free MsbA (gray, PDB:5TV4), G247-bound MsbA (blue and orange), and G907-bound crystal structure (pink, PDB:6BPL). Alignment performed on each TMD, shown with gray background. **g,** Comparison of coupling helices (CH) positioning for drug-free MsbA (gray, PDB:5TV4), G247-bound MsbA (blue), and G907-bound crystal structure (pink, PDB:6BPL). Compared to the reference drug-free cryo-EM structure, the distance between CH1 and CH2 is reduced for the G907-bound crystal structure and increased for G247-bound MsbA. Only one TM chain is shown for clarity. Alignment is performed on TMD1, as shown in panel f.

Comparing the cryo-EM structures of G247-bound and drug-free MsbA[5] reveals the mechanism of G compound inhibition which is different from previously anticipated: instead of pushing one NBD towards the dimerization interface by 10-15 Å[1], G247 binding displaces the NBDs symmetrically and away from each other, increasing inter-NBD distance by ~13 Å (***Figure 4.4c***). This model is consistent with 3D variability analysis (***Figure 4.6a***), in which G247-bound MsbA NBDs move away from the dimerization axis relative to the NBDs of drug-free MsbA. In summary, while drug-free MsbA has NBDs aligned in a head-to-tail manner and thus primed for ATP hydrolysis, the increased inter-NBD distance in G247-bound MsbA is expected to reduce NBD dimerization efficiency and ATPase activity.

To formally test this hypothesis, we exposed MsbA to the non-hydrolyzable nucleotide AMP-PNP, in the presence or absence of G247 (***Figure 4.6b***). When subjected to 1mM AMP-PNP, nearly all MsbA particles showed a closed conformation with dimerized NBDs. In contrast, 5 μM of G247 was sufficient to shift more than 60% of particles to an open conformation with separate NBDs (***Figure 4.6b***), with consistent results using both 2D and 3D classification. Taken together, our EM analyses demonstrate that G compounds act as TMD wedges to symmetrically increase NBD separation, thus inhibiting conformational transition and ATP hydrolysis.

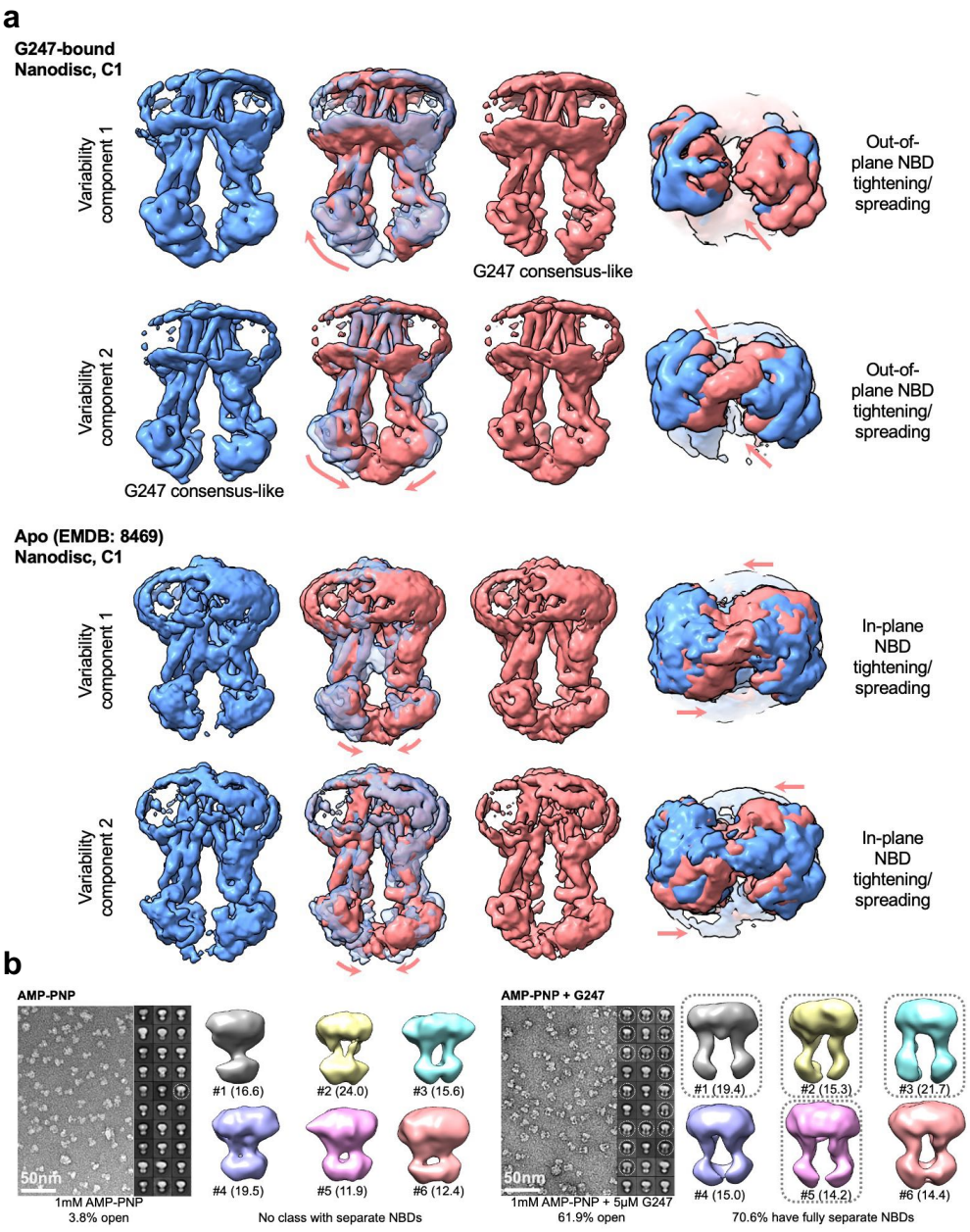


Figure 4.6: Variability and conformation analysis for the G247-bound MsbA conformation. **a,** CryoSPARC variability analysis for *E. coli* MsbA in nanodisc, when bound to G247 (top) and drug-free (bottom, EMDB:8469 from a prior publication[5]). The first two variability components are displayed, with the first and last frame of each volume series displayed blue and red, respectively. While the NBDs are dynamic in each dataset, they tend to be more misaligned in the G247-bound reconstructions than in the drug-free data (rightmost panels). **b,** Representative negative-stain EM images and 2D class averages of *E. coli* MsbA exposed to AMP-PNP, in the absence and presence of G247. G247 prevents NBD closure when the protein is exposed to AMP-PNP, a non-hydrolyzable nucleotide. The computational analysis was performed twice with comparable outcomes.

	***E. coli* G247 In nanodisc**	***E. coli* G247 In DDM**
	PDBid: 7MEW EMDB: 23805	EMDB (C2): 24446 EMDB (C1): 24447
Data collection and processing		
Microscope	UMass Titan Krios / HMS Polara	Harvard Titan Krios
Detector	K2	K3
Magnification	165,000 / 31,000	81,000
Voltage (kV)	300	300
Electron Exposure ($e^-/Å^2$)	63 / 52	52
Defocus Range (μm)	1.0 – 2.5	1.0 – 2.5
Pixel Size (Å)	0.83 / 1.23	1.06
Final symmetry imposed	C2	C2 / C1
Initial particle images (#)	159,905 / 302,324	1,710,546
Final particle images (#)	349,674	942,315
Map resolution (Å)	3.90 TMD 4.08 overall	3.55 / 3.79 TMD 3.66 / 3.94 overall
FSC threshold	0.143	0.143
Map resolution range (Å)	3.5 – 8.0	3.0 – 8.0
Refinement		
Initial model used (PDBid)	5TV4	5TV4
Model resolution (Å)	4.09	
FSC threshold	0.5	
Map sharpening B factor ($Å^2$)	N/A	N/A
Model composition		
Non-hydrogen atoms	7376	
Protein residues	1144	
Ligand	2	
B factor ($Å^2$)		
Protein	96.94	
Ligand	46.37	
Model vs. Map		
TMD CC	0.77	
Overall CC	0.72	
R.m.s. deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.624	
Validation		
MolProbity score	1.61	
Clashscore	9.61	
Poor rotamers (%)	0	
Ramachandran plot		
Favored (%)	97.46	
Allowed (%)	2.54	
Disallowed (%)	0	

Figure 4.7:Data collection table and model refinement statistics

4.3 Discussion

We have shown that G247 acts as a TMD wedge, symmetrically increasing inter-NBD spacing and preventing MsbA conformational transition. This mechanism is different from that hypothesized from crystal structures, involving asymmetric tightening of the NBDs and dissociation of coupling helices. The wedge mechanism described here shows similarities with other ABC transporter inhibitors including zosuquidar, which also reduces the activity of the homologous ABCB1 transporter by pushing its NBDs away from each other[8]. The symmetrical TMD wedge mechanism is significantly different from how TBT1 inhibits MsbA function. In each case, inter-NBD distance is correlated with stimulation or inhibition of ATPase activity, suggesting that inter-NBD spacing provides a heuristic for classifying ATPase inhibitors from stimulators. Since most ABC transporter inhibitors identified so far are ATPase suppressors, it is likely that more compounds act similarly to G247 than to TBT1.

It is particularly interesting that the closed asymmetric state remained elusive in both nanodisc and detergent, despite extensive classification and 3D variability analysis. Although it is possible that this conformation may transiently occur, it is likely stabilized by crystal contacts rather than a dominant state in solution. In this respect, the closed asymmetric crystal structure is comparable to the outward-open MsbA state, which has so far resisted all our cryo-EM efforts as outlined in chapter 2. Discrepancies between cryo-EM and X-Ray crystallography are rare, but previous examples can be found in the literature for flexible membrane proteins. For example, drug-free ABCB1 has straight TM4 and TM10 in crystallography studies[9] yet kinked helices in single-particle EM[10], providing different insights into mechanisms of substrate selection. Thus, although crystallography is a useful technique to trap transient ABC transporter states, single-

particle EM may be a more reliable method to determine the dominant conformation adopted by membrane proteins.

G compounds are relatively hydrophobic, such that their partitioning in lipid membranes facilitates binding to MsbA even at low concentrations. The downside of G compound hydrophobicity is their tendency to bind plasma proteins[11] and to fail at reaching their bacterial target in practice. Although this specific molecular scaffold did not result in a potent antibiotic, the induced-fit pocket it creates upon MsbA binding may be amenable to targeting by other classes of small molecules. Future research efforts may employ computational docking, as in chapter 3, to identify new MsbA TMD wedges with improved deliverability.

4.4 Methods

Cloning, expression, and purification of MsbA

The *E.coli* MsbA construct used in this study was originally obtained from the G. Chang lab[3,5] and consists of N-terminally His tagged MsbA in pET-19b. As in previous chapters, MsbA was transformed in *E. coli* BL21 DE3 and grown in Terrific Broth medium to OD_{600} ~1.0 at 37°C, with 1mM IPTG induction and growth at 17°C overnight. The cells were harvested by centrifugation at 4,000 xg, washed in Phosphate-buffered saline, then resuspended in Buffer A (50 mM Tris pH 7.8, 300 mM NaCl, 10% glycerol, 0.5 mM TCEP). Cell lysis was performed either with a probe sonicator (60% power, 20% "on" time for ~5 min) or with a microfluidizer operated at 800 bar. Lipid membranes were harvested by 1 hour ultracentrifugation at 185,000 xg at 4°C, followed by flash-freezing in liquid nitrogen and storage at -80°C. Membranes were subsequently resuspended in buffer A + 1% (w/v) DDM, homogenized, and stirred for 1-2 hours at 4°C. Cell debris were removed by centrifugation at ~10,000 xg for 15min, followed by 100,000 xg for 30min. The resulting solubilized membranes were flowed over a TALON cobalt resin (Takara Bio), washed with ~20 column volumes (CV) of buffer A + 0.1% DDM + 10 mM imidazole, and eluted with 4CV of buffer A + 0.1% DDM + 250 mM imidazole. The protein was concentrated on a 100 kDa cutoff concentrator and further purified by gel filtration over a Superdex 200 column in a 1:1 buffer A water dilution with 0.05% DDM. Peak fractions were combined and concentrated to ~10 mg/mL for nanodisc incorporation. The protein was incubated with MSP1D1 and POPG at 1:1:60 stoichiometric ratios, in 20 mM cholate buffer. The DDM was removed using ~65% (w/v) SM2 biobeads before size-exclusion chromatography in 25 mM Tris and 150 mM NaCl. MsbA incorporated in nanodiscs was then concentrated for cryo-EM analysis.

Negative-stain EM

Protein samples were assessed for homogeneity by negative-stain EM. All samples were stained using a 1.5% (w/v) uranyl formate solution and imaged on a Tecnai T12 electron microscope (FEI), operated at 120kV and at 67,000x magnification. For quantification of MsbA conformations, only clearly inward-facing classes with 2 visible and separate NBDs were considered. The 1mM AMP-PNP/Mg^{2+} sample contained 1 inward-facing class in 30 classes (405 particles in 10,688) and the 1mM AMP-PNP/Mg^{2+} + 5μM G247 contained 18 inward-facing classes in 30 classes (10,784 particles in 17,427).

Cryo-EM sample preparation and data collection

E. coli MsbA in nanodisc was concentrated to ~1.5 mg/mL and incubated with 100 μM G247 for 2 hours at 4°C, whereas the DDM dataset was concentrated to 16 mg/mL and incubated with 500 μM G247 to account for the increased protein concentration. 3 μl of sample were applied to Quantifoil (R1.2/1.3, 400 mesh) grids and blotted with either a Gatan Cryoplunge 3 (blot ~3s) or Thermo Fisher Vitrobot Mark IV (blot ~7s, force +12) and vitrified in liquid ethane. The G247-bound *E. Coli* MsbA nanodisc sample combines data from two microscopes: #1 Titan Krios equipped with a K2 camera (0.42 Å super-resolution pixel size, 62.7 e/$Å^2$ exposure) and a BioQuantum imaging filter (with a slit width set to 20 eV); #2 FEI T30 Polara with a K2 camera (0.62 Å super-resolution pixel size, 52 e/$Å^2$ exposure). The G247-bound DDM sample was acquired on the Harvard Medical School Titan Krios 1, equipped with a K3 camera (1.06 Å pixel size in counting mode, 52 e/$Å^2$ exposure). SerialEM[12] was used for collection of all datasets. More information about cryo-EM data collection parameters can be found in ***Figure 4.7***.

Cryo-EM data processing

Movie stacks were binned 2x or left un-binned depending on whether the data was acquired in super-resolution mode, and motion-corrected using MotionCor2[13]. The micrograph CTF fits were estimated with CTFFIND4[14]. All 2D processing was conducted using Simplified Application Managing Utilities of EM Labs (SAMUEL)[5]. 2,000 protein images were manually picked to generate initial 2D templates. Auto-picking on 10% of movie stacks were then used to create refined 2D templates, then used to pick particles on the whole dataset. 2D particles were screened using "samtree2dv3.py. All subsequent 3D processing was performed with RELION3.0[15], including 3D classification, domain masking, signal subtraction[16], and 3D particle refinement. Overall resolution of cryo-EM map was computed according to the gold standard FSC method, with reconstruction of independent halves of the dataset and determination of resolution using FSC=0.143 criterion, whereas local resolution was estimated with ResMap[17]. All maps were subjected to density modification[18] in phenix, with local sharpening and blurring.

Model building and validation

G247-MsbA was fit into the cryo-EM map using 5TV4.pdb as reference, then adjusted with COOT. G247 model and restraints were manually built using ELBOW[19] then fit into the density. The models were then refined against the cryo-EM map using Phenix[20] real space refinement and validated with MolProbity[21]. All figures were generated with UCSF Chimera[22] or ChimeraX[23].

4.5 References

1. Ho, H. et al. Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196-201 (2018).

2. Higgins, C.F. & Linton, K.J. Structural biology. The xyz of ABC transporters. *Science* **293**, 1782-4 (2001).

3. Ward, A., Reyes, C.L., Yu, J., Roth, C.B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-10 (2007).

4. Padayatti, P.S. et al. Structural Insights into the Lipid A Transport Pathway in MsbA. *Structure* **27**, 1114-1123 e3 (2019).

5. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

6. Angiulli, G. et al. New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins. *Elife* **9**(2020).

7. Punjani, A. & Fleet, D.J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702 (2021).

8. Alam, A., Kowal, J., Broude, E., Roninson, I. & Locher, K.P. Structural insight into substrate and inhibitor discrimination by human P-glycoprotein. *Science* **363**, 753-756 (2019).

9. Ward, A.B. et al. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. *Proc Natl Acad Sci U S A* **110**, 13386-91 (2013).

10. Alam, A. et al. Structure of a zosuquidar and UIC2-bound human-mouse chimeric ABCB1. *Proc Natl Acad Sci U S A* **115**, E1973-E1982 (2018).

11. Alexander, M.K. et al. Disrupting Gram-Negative Bacterial Outer Membrane Biosynthesis through Inhibition of the Lipopolysaccharide Transporter MsbA. *Antimicrob Agents Chemother* **62**(2018).

12. Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y. & Mastronarde, D.N. Software tools for automated transmission electron microscopy. *Nat Methods* **16**, 471-477 (2019).

13. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

14. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).

15. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).

16. Bai, X.C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S.H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**(2015).

17. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-5 (2014).

18. Terwilliger, T.C., Ludtke, S.J., Read, R.J., Adams, P.D. & Afonine, P.V. Improvement of cryo-EM maps by density modification. *Nat Methods* **17**, 923-927 (2020).

19. Moriarty, N.W., Grosse-Kunstleve, R.W. & Adams, P.D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074-80 (2009).

20. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

21. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).

22. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

23. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).

Chapter 5: Summary and Outlook

The findings presented in this thesis broadly fall under two topics: MsbA function (chapter 2) and inhibition (chapters 3 and 4). Major advancements regarding the MsbA function cycle include:

(1) The MsbA wide-open state is accessible in lipid bilayers when the protein is actively hydrolyzing ATP. Furthermore, lack of LPS in this conformation but not in the narrow inward-facing state suggests that LPS binding contributes to a conformational transition to a tighter conformation.

(2) Nucleotide binding is sufficient for both LPS ejection and complete closure of the TMDs. Thus, ATP hydrolysis does not cause a rearrangement of the TMDs from outward-open to closed state, as previously anticipated[1].

(3) Under ATP turnover conditions, the predominant closed conformation is ADP-bound. Therefore, ATP hydrolysis occurs very fast and NBDs dissociation is likely a rate-limiting step in LPS transport.

(4) ADP-bound NBDs seem destabilized relative to AMP-PNP-bound MsbA, possibly due to lack of γ-phosphate. Release of γ-phosphate may promote dissociation of the NBDs and conformation cycling.

The role of LPS in MsbA conformational change needs to be evaluated more accurately. For example, the hypothesis that LPS provokes MsbA closure to the narrow inward-facing state is weakened by the structure of drug-free, wide-open *A. baumannii* MsbA presented in chapter 3. Yet in this structure, there is a species mismatch between protein (*A. baumannii* MsbA) and substrate (*E. coli* LPS). It is conceivable that a protein designed to recognize substrate with 7 acyl chains would not properly change conformations when bound to a substrate with 6 acyl chains. Alternatively, MsbA from different species may assume distinct conformations in the LPS-free

state. Elucidation of this question will likely require purification of MsbA from an organism that does not produce LPS (e.g., *L. lactis,* which is capable of producing MsbA[2]), followed by supplementation of commercial purified LPS at the nanodisc reconstitution step.

One early goal of this dissertation project was to capture intermediate states of LPS flipping, either using AMP-PNP or careful classification of particles from an ATP turnover experiment. Although this endeavor has been unsuccessful, this negative result suggests that future efforts should focus on biochemically crosslinking the incredibly transient flipping intermediate. Several approaches are possible, including using cysteine crosslinking to prevent complete conformational transition to the outward-facing state (e.g., by introducing cysteines at select positions lining helices 1 and 3), or by directly crosslinking MsbA to its substrate by incorporation of unnatural amino acids. Either of these options will require considerable trial and error.

It is my opinion that the primary scientific contribution of this thesis is the description of new mechanisms of MsbA inhibition, and specifically the structure of MsbA bound to the TBT1 decoupler. These key findings are illustrated in ***Figure 5.1*** and are summarized next:

(1) Two copies of TBT1 asymmetrically bind in the central MsbA pocket and mimic the endogenous LPS substrate. As with LPS, TBT1 establishes both hydrophobic and electrostatic interactions with MsbA.

(**2) Binding of TBT1 leads to drastic structural rearrangements predominantly within one half of the transporter**, including the sliding of TM6.A into the central cavity, and destabilization of domain-swapping TM4-TM5.B bundle and coupling helices. These conformational changes result in one NBD moving towards the NBD dimerization interface.

Reduction of inter-NBD distance promotes ATP hydrolysis and accounts for the stimulatory effect of TBT1.

(3) The central substrate pocket, reshaped by TBT1, is a viable target for computational drug discovery. Virtual screening and ATPase measurements result in convenient identification of modulators, including both ATPase stimulators and suppressors.

(4) G247 is a TMDs wedge and symmetrically pushes the NBDs away from each other, as indicated by comparison with the drug-free conformation and 3D variability analysis. This increased inter-NBD spacing prevents NBD closure and conformational transition, and results in inhibition of ATPase activity.

(5) In these inhibitor-bound structures, inter-NBD spacing compared to drug-free state is correlated with ATPase activity modulation. Thus, inter-NBD spacing may be a useful heuristic for classifying ATPase stimulators from suppressors.

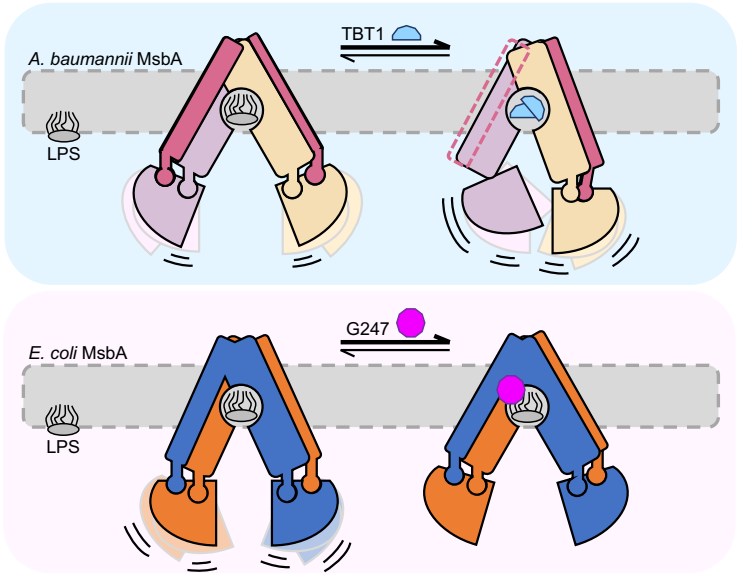


Figure 5.1: Distinct MsbA inhibition mechanisms of positive and negative allosteric ATPase modulators. Proposed models for the divergent action of TBT1 (top) and G247 (bottom). The inner membrane is shown as dashed gray box. For illustrative purposes, the weakly associated CH1.A is not represented in TBT1-bound MsbA. Curved lines and duplicated NBDs illustrate NBD conformational heterogeneity and their ability to dimerize for hydrolyzing ATP, in all conformations except G247-bound MsbA.

Closing thoughts

Complete disordering of domain-swapping helices in the TBT1 conformation is unexpected and differs from any ABC transporter structure characterized to date. While novelty is always exciting, I experience a certain sense of dread at the thought that the TBT1 structure is an interesting artifact or experiment defect, rather than the true MsbA stimulated state. At the core of this interrogation is that the burden of proof for showing physiological relevance of a structure is too high: no structure, whether X-Ray or cryo-EM, is free of potential experiment bias, except possibly when characterized *in situ* by cryo-electron tomography. Perhaps one day technical advancements in tomography and direct imaging of TBT1-treated cells will confirm or invalidate my work.

The state of structural biology is vastly different today from when I entered graduate school in 2016. Due to technical improvements and democratization of cryo-EM, the structures of many – arguably, most – critical ABC transporters in human and bacteria are now known. A very recent and exciting event is the development of stunningly accurate protein structure prediction programs, most notably AlphaFold[3,4]. Although these predictors do not handle oligomers well at the moment, it seems only a matter of time before this technical hurdle is addressed. Thus, for many proteins, the scientific value derived from determining a drug-free structure has now considerably decreased.

Concurrently, substantial advances in virtual screening now make it possible to determine binding energetics for hundreds of millions of ligands[5], in some cases resulting in the identification of compounds with remarkable binding affinity and subtype selectivity[6]. Will the combined power of accurate structure prediction, together with virtual screening, render the structural biology pipeline obsolete? What place is there for structural biologists in this new era?

I find solace in thinking that most conclusions presented in this thesis are out of the immediate reach of existing structural bioinformatics programs, because MsbA is simply too flexible a protein and a single snapshot is not an accurate representation of its behavior in lipid membranes. It is outstanding that the only known MsbA inhibitors, TBT1 and G compounds, both remodel the transmembrane domains and define unpredictable binding pockets by induced-fit mechanisms. Hence for the foreseeable future, development of modulators of MsbA and many other flexible transporters will rely on empirical structure determination before virtual compound screening can be reliably utilized.

After two decades of structural investigation, and despite my best efforts over the past few years, MsbA and ABC transporter pharmacology still escapes our understanding. The inhibition mechanisms of these transporters appear incredibly diverse and complex[7-9], and in many cases may be protein-dependent and not generalizable across the ABC superfamily. Yet each new inhibitor-bound structure, and every new binding pocket revealed by induced-fit, brings us a little closer to the goal of developing new drugs. For the first time, this goal seems almost within reach.

References

1. Mi, W. et al. Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237 (2017).

2. Woebking, B. et al. Drug-lipid A interactions on the Escherichia coli ABC transporter MsbA. *J Bacteriol* **187**, 6363-9 (2005).

3. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* (2021).

4. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* (2021).

5. Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224-229 (2019).

6. Stein, R.M. et al. Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609-614 (2020).

7. Nosol, K. et al. Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter ABCB1. *Proc Natl Acad Sci U S A* **117**, 26245-26253 (2020).

8. Orlando, B.J. & Liao, M. ABCG2 transports anticancer drugs via a closed-to-open switch. *Nat Commun* **11**, 2264 (2020).

9. Nosol, K. et al. Structures of ABCB4 provide insight into phosphatidylcholine translocation. *Proceedings of the National Academy of Sciences* **118**, e2106702118 (2021).

Appendix: Sub-3 Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly

The data presented here is part of a preprint:

Liu, D.‡, **Thelot, F. A.**‡, Piccirilli, J. A., Liao, M., & Yin, P. (2021). Sub-3 Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *bioRxiv*, 2021.2008.2011.455951. doi:10.1101/2021.08.11.455951

‡ *Denotes equal contribution*

Contribution: D. Liu designed the RNA constructs, performed oligomer assembly, conducted biochemical assays, and built atomic models with F. Thelot. F. Thelot prepared the cryo-EM samples, acquired the cryo-EM data, and performed cryo-EM data processing with guidance from M. Liao. D. Liu and F. Thelot wrote the manuscript with contributions from all authors. P. Yin and M. Liao supervised the project.

A.1 Introduction

The appreciation for the versatile and far-reaching roles that RNAs can play has been growing over the past three decades[1]. Besides guiding protein biosynthesis as viewed classically, RNAs are now understood to regulate gene expression and modulate other important biological processes at different levels and by various mechanisms[2,3], such as binding proteins, recognizing metabolites and catalyzing chemical transformations. Evidence shows that more than 85% of human genome is transcribed, but only less than 3% of genome sequence is protein-coding[4], suggesting a tremendous portion of transcribed RNAs with functions unknown. Furthermore, the functional capacity of RNAs has been remarkably expanded with *in vitro* selection and evolution[5-7]. In order to elucidate the mechanisms underlying these functional RNAs, high-resolution structural information is necessary. However, RNA structural biology, traditionally being a primary territory of X-ray crystallography, is challenging: (1) the intrinsic properties of RNAs such as (a) poorly differentiated anionic surface, (b) irregular and elongated shape, and (c) structural flexibility and conformational heterogeneity, make it notoriously difficult to obtain well-diffracting crystals; (2) phase determination of RNA crystals is nontrivial due to the lack of convenient and universal strategies such as selenomethionine substitution[8] for protein crystallography.

Without the need for procuring crystals and solving phase problem, cryo-EM is gaining increasing popularity and its obtained resolution is now beginning to rival that of X-ray crystallography thanks to the on-going advances in instrumentation and software[9]. However, the application of cryo-EM in RNA structural determination has not been well explored. So far, there are only two reported examples of RNA-only structures[10-12] determined by cryo-EM that achieve a resolution of 4.5 Å or better. The first among them is the 4.5 Å structure of Lactococcus lactis group IIA intron[10], a large RNA containing > 600 RNA nucleotides (nt). The other more recent

case is the smaller 119-nt Mycobacterium sp. SAM-IV riboswitch[12]: cryo-EM maps at 3.7 Å and 4.1 Å resolution were reported for the apo and ligand-bound states, but a large dataset (~ two million initial particles) was required for the reconstruction. Even for the best-resolved map at 3.7 Å, the RNA bases are barely separate and their conformations are poorly defined, and therefore, making the model building process heavily dependent on computer modeling[13].

Similar to crystallography, structural flexibility of RNA molecules is a major limiting factor for their high-resolution cryo-EM analysis: (1) double-stranded A-helix—the basic secondary structure element of RNA—has a much shorter persistence length (~64 nm, or ~230 RNA base-pairs[14]) than its protein counterpart, α-helix (~100 nm, or ~670 amino acid residues[15]); (2) folded RNAs normally have less long-range tertiary interactions to stabilize the overall 3D architectures than proteins[16]. Besides flexibility, a lot of structured RNAs are of relatively small size (<100 kDa, or <300 nt), making their cryo-EM determination more challenging. Herein, we present ROCK (RNA oligomerization-enabled cryo-EM via installing kissing-loop), a nanoarchitectural engineering strategy derived from nucleic acid nanotechnology[17-21] to address the challenges in RNA cryo-EM. Kissing-loop (KL) sequences are installed onto the peripheral stems without perturbing the functional domain to mediate self-assembly of the RNA of interest (ROI) into a closed homooligomeric ring (***Figure A.1***). Analogous to the quaternary structure of proteins, the assembled RNA structure has a multiplied molecular weight and each constituent monomeric unit is expected to have decreased structural flexibility due to the geometric restraint. KLs, instead of sticky ends, are chosen as the assembly mediator because their paranemic characteristic[22-25] (i.e. the two interacting loops within the KLs are topologically closed and can separate without the need for strand scission) minimizes the strand breaks (and thereby, the number of unique strands) and dispenses with introducing permutation to the ROI.

Properties	Standalone RNA	Assembled RNA
Molecular weight	Small	Multiplied
Conformation	Highly flexible	Flexibility mitigated
Folding	Strong misfolding tendency, difficult for purification	Alternate folding eliminated by purification
Shape	Irregular	Symmetric, characteristic

Figure A.1: Concept of RNA oligomerization-enabled cryo-EM via installing KL (ROCK). The engineering of a target RNA for the self-assembly of a closed homooligomeric ring, which is more amenable for cryo-EM structural determination compared to the standalone monomer RNA.

A.2 Results

A.2.1 Engineering RNA for homomeric self-assembly

We first used the Tetrahymena group I intron (TetGI) as an ROI to explore ROCK because, as the first discovered catalytic RNA[26], the TetGI is one of the most investigated RNAs and serves as a rewarding model for research of RNA biochemistry and structural biology[27,28]. ***Figure A.2a*** shows its secondary structure derived from previous prediction[29] and our present study., Besides the catalytic core (containing P4-P6, P3-P9 and P1-P10 domains) conserved for all group I introns, the TetGI, as a subgroup IC1 intron, also possesses peripheral insertions, P2-P2.1 and P9.1-P9.2 domains (***Figure A.2b***). Though a number of X-ray structures of partial TetGI (including the P4-P6 domain[30,31] and the core[32,33]) have been determined by crystallography since its discovery

nearly four decades ago, its complete structure has not been determined at high resolution. Nonetheless, the complete TetGI has been modelled computationally based on phylogeny, biochemistry data and a number of distance constraints derived from long-range tertiary interactions[29].

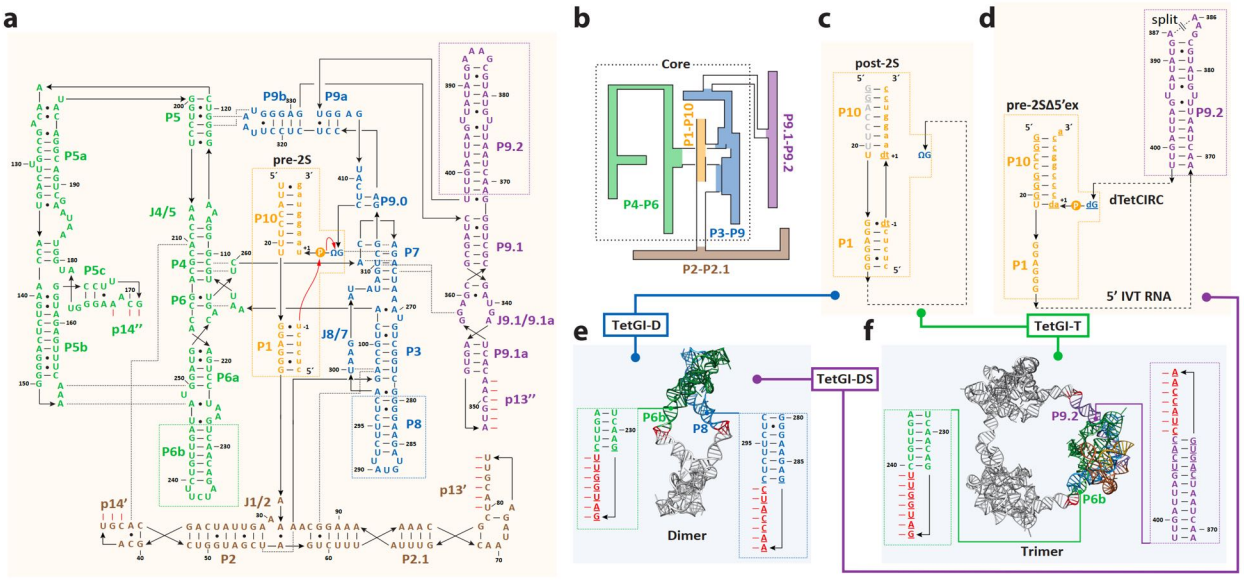


Figure A.2: Construct engineering of the Tetrahymena group I intron (TetGI) for ROCK (RNA oligomerization-enabled cryo-EM via installing kissing-loop). a, Sequence and secondary structure of the wild-type (WT) TetGI in the pre-2S state. Several key tertiary contacts are shown with dashed lines. Short lines (–) indicate the canonical Watson-Crick (WC) base-pairs (bps), and dots (•) indicate the non-WC bps. Dashed boxes mark the locations to be engineered in our current study. **b,** Schematic shows the intron's catalytic core, which encloses the P4-P6, P1-P10 and P3-P9 domains, and the two peripheral P2-P2.1 and P9.1-P9.2 domains. **c, d,** Engineered constructs corresponding to the post-2S (c) and the pre-2SΔ5'ex (pre-2S without 5'-exon bound; d) reaction states of the intron. Lower-case nucleotides are from exons, underlined nucleotides are mutated from the WT, and grey nucleotides (in c) are not present in the final constructs due to the cleavage after U20 during the transcription. In d, the RNA construct has a strand split at the loop of P9.2 (between A386 and A387) and is formed by a 5' RNA fragment synthesized by in vitro transcription (IVT) and a chemically synthesized 3' RNA with two deoxy mutations (dTetCIRC). **e, f,** The dimeric and trimeric assemblies are designed by engineering P6b and P8, and P6b and P9.2, respectively, so that the sequences (red) for forming kissing-loop (KL) motifs are installed to mediated the cohesion of the monomeric units. The design of dimer is based on the previous crystal structure of the TetGI core domain (PDB code: 1x8w), and the design of trimer is based on a computer model by Michel, Westhof and coworkers. In our current study, three TetGI constructs are designed and studied by cryo-EM: TetGI-D (dimeric construct of the post-2S state), TetGI-T (trimeric construct of the post-2S state), and TetGI-DS (dimeric construct of the pre-2SΔ5'ex state).

The TetGI catalyzes two consecutive transesterification reactions, and we chose two of its reaction states in our construct design. ***Figure A.2c*** presents a construct initially designed as the post-2S (the state after the second step of splicing) complex with two deoxy substitutions introduced in the ligated exon mimic (TetLEM; added in trans) to prevent the reverse reaction of the second step of splicing[34]. However, the intron RNA was cleaved co-transcriptionally between U20 and U21 as indicated by dideoxynucleotide sequencing analysis[35], probably due to the formation of a hairpin at the RNA's 5'-end. Therefore, this construct (***Figure A.2c***) can also be regarded as the complex formed by the trans-acting TetGI ribozyme (for the restriction endonucleolytic reaction[36]) and its substrate. ***Figure A.2d*** shows a second construct, termed pre-2SΔ5'ex, corresponding to the pre-2S (the state before the second step of splicing) complex[34] but without the 5'-exon. This construct is formed by two fragments: an in vitro transcribed (IVT) RNA corresponding to the 5' of the TetGI till A386 and a chemically synthesized 37-nt chimeric oligonucleotide (dTetCIRC) enclosing the 3' fragment of the TetGI and the 3'-exon. We introduced mutations to the 5' sequence of the IVT RNA to prevent the possible formation of aforementioned hairpin presumably responsible for the cleavage near the 5' end. Two deoxy substitutions are present in dTetCIRC to inhibit the hydrolysis at 3' splice site[37]. An additional u(+1)a mutation restores the base-pairing at the base of P10 and potentially further improves the construct's rigidity[34]. This pre-2SΔ5'ex complex has two features that might be beneficial for obtaining a higher-resolution structure: (1) the extra covalent linkage between helices P9.0 and P10 is expected rigidify the architecture[34]; and (2) the removal of 5'-exon is intended to eliminate its binding/unbinding equilibrium and thereby to improve the conformational homogeneity.

According to the knowledge gained from the previous structural and functional studies,

stems P6b, P8 and P9.2 are extended away from the catalytic core and are not participating in tertiary interactions; therefore, they were chosen for ROCK engineering. For the choice of KL throughout this work, we use the 7-bp KL having an NMR structure[38] of an angle from ~110° to ~130° (~120° for the average structure). Consequently, a homodimer (***Figure A.2e***) was designed based on the X-ray model of the core[33] by engineering P6b and P8; and a homotrimer (***Figure A.2f***) on the computer model[29] by engineering P6b and P9.2. The lengths of the stems on which the KL sequences are installed need to be optimized to ensure formation of ring instead of other linear or spiral assemblies. This can be readily performed in silico with software such as NanoTiler[39]. In our current study, three TetGI constructs are designed and studied by cryo-EM: two dimeric constructs, TetGI-D and TetGI-DS, designed as the post-2S and pre-2SΔ5'ex states, respectively; and a trimeric construct, TetGI-T, designed as the post-2S state.

A.2.2 Sub-3 Å cryo-EM structure of the TetGI-DS construct

The best-resolved cryo-EM map of the TetGI was obtained from the TetGI-DS construct, so we first present the cryo-EM analysis of this construct. The characteristic shape of the TetGI-DS dimer, clearly discerned from the raw micrographs (***Figure A.3***) and 2D class-averages (***Figure A.4a***), is beneficial for particle-picking and initial alignment. Reconstruction with C_2 symmetry applied to the homodimer only delivers a cryo-EM map of a medium resolution (see ***Figure A.4b*** for the whole C_2 dimer, and ***Figure A.4c*** for the symmetrized monomer, or the C_2 monomer), but symmetry-expansion (SE) allows finer 3D classification of the monomers (***Figure A.3***) and refinement of these SE monomers yields a significantly improved resolution (***Figure A.4d***): as estimated by the Fourier shell correlation (FSC) curves (***Figure A.4e***), the overall resolution for the full map of the SE monomer is 2.98 Å at FSC = 0.143, with the core arriving at 2.85 Å

resolution, superior to the best-resolved group I intron to date—the 3.10 Å crystal structure[34] of the Azoarcus group I intron (AzoGI). This represents the first sub-3 Å cryo-EM map obtained for an RNA-only structure and enables the de novo model building of the complete TetGI (***Figure A.4f,g***). Remarkable details of structural features pertaining to the sugar-phosphate backbone and nucleobases of RNA are resolved (***Figure A.4h***; see ***Figure A.5*** for comparing the features of different X-ray and EM maps). As a proof for the excellent map quality, we show in ***Figure A.4i*** that the density for individual base-pair (bp) could be well separated without breaking the continuity of the backbone density at a wide range of contour levels. The distinct geometries of different types of bps (***Figure A.4j,k***) can be readily recognized. Remarkably, structural features for some exocyclic amino groups could be spotted (blue arrows in ***Figure A.4i-k***). Prominently, we can visualize the strong intensities of ordered Mg^{2+} ions (***Figure A.4l***). This demonstrates the power of cryo-EM in the identification and determination of native metal ions, which have profound roles in RNA folding and sometimes serve as ligands in RNA catalysis[40], dispensing with the complicated screening of nonnative heavy metals, which serve as anomalous scatterers and native metal mimics, in X-ray crystallography.

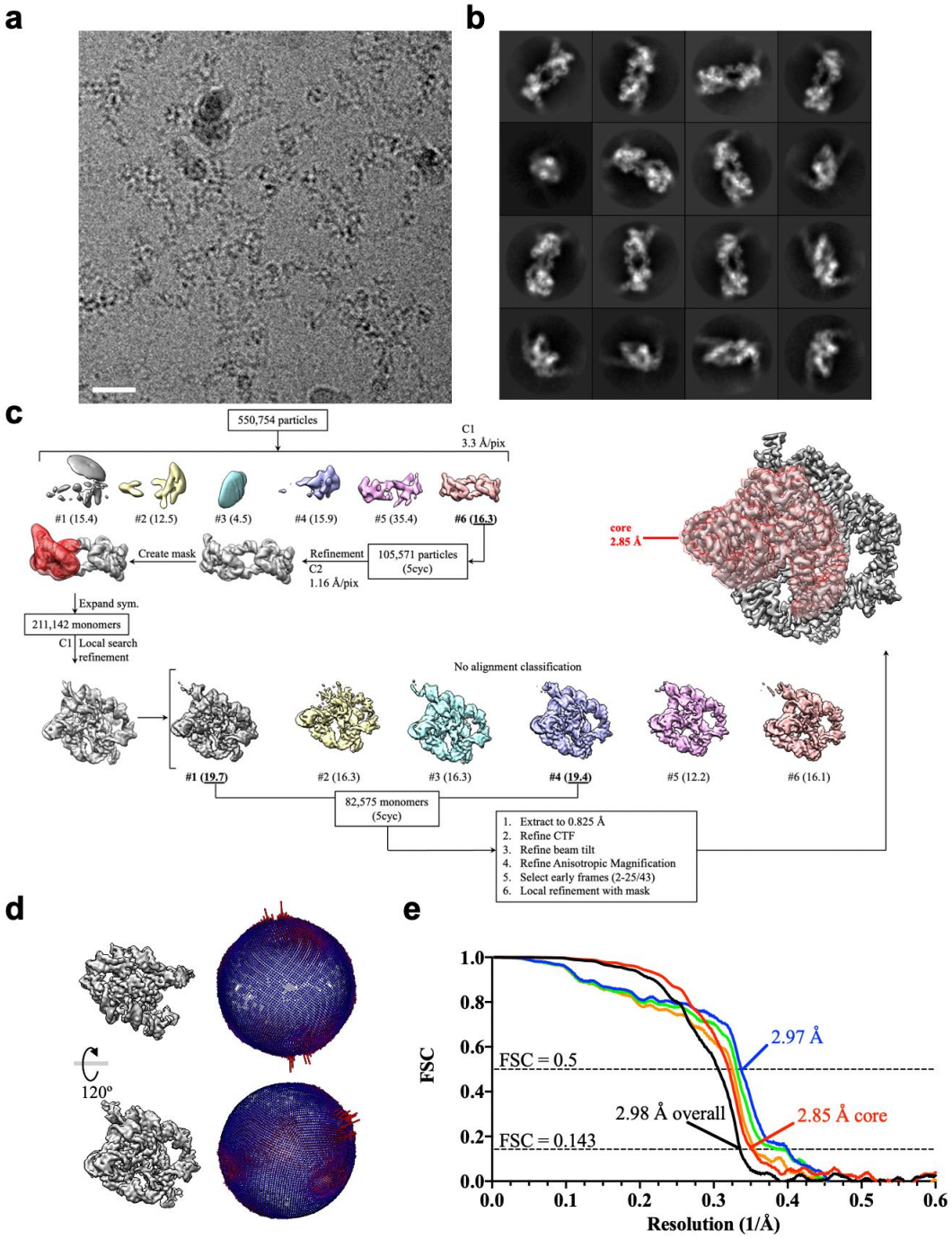


Figure A.3: Cryo-EM imaging, processing and validation for TetGI-DS. a, Representative cryo-EM image of TetGI-DS. The scale bar represents 20 nm. **b,** 2D class averages of TetGI-DS. Box size is 264 Å. **c,** Processing flowchart for the TetGI-DS dataset. **d,** Angle distribution for the particles included in the final 3D reconstruction. **e,** Fourier Shell Correlation (FSC) curves of the final TetGI-DS reconstruction. Half map #1 vs. half map #2 for the entire monomer is shown in black. The remaining FSC curves were calculated for the core domains only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange).

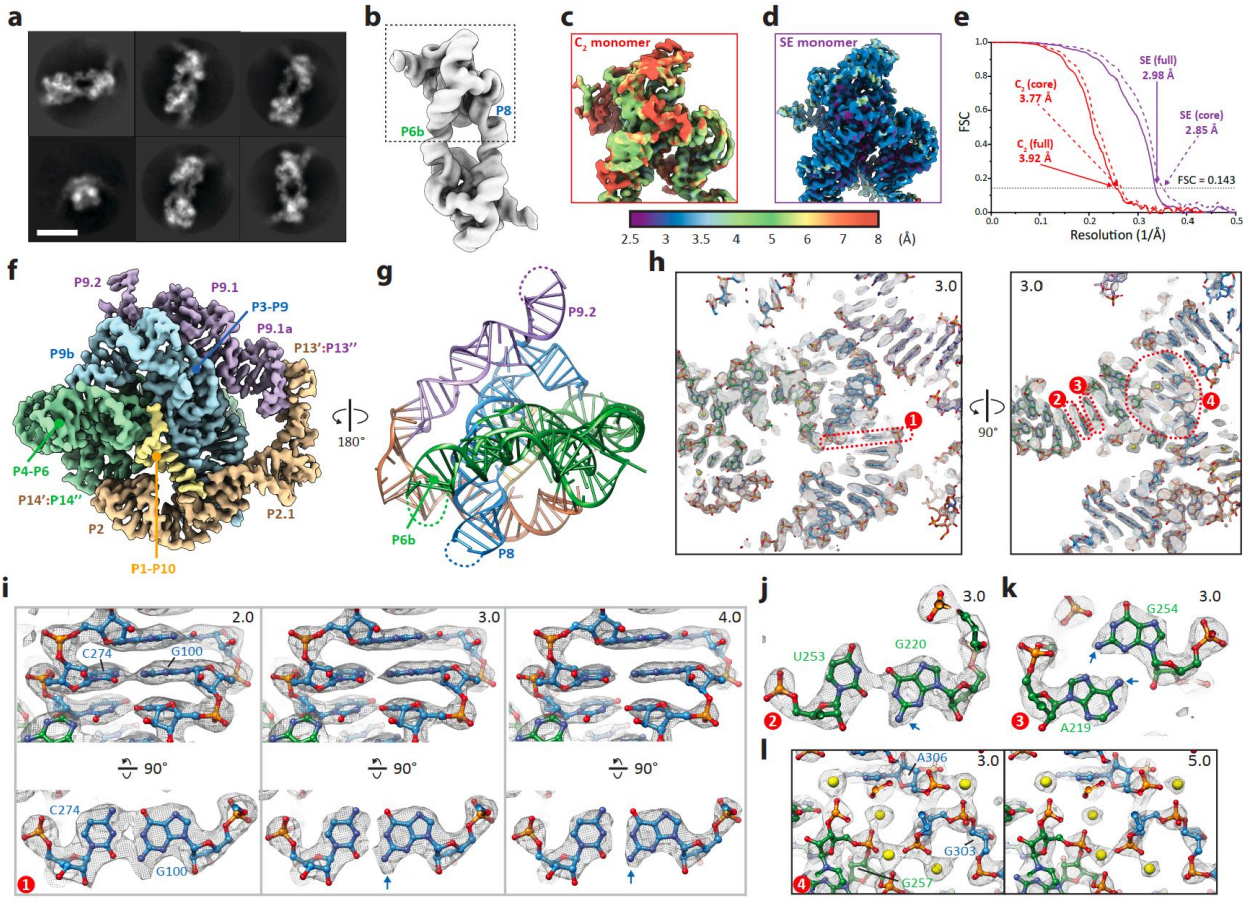


Figure A.4: Sub-3 Å resolution cryo-EM map of TetGI-DS enabled by ROCK. a, Representative 2D class averages of TetGI-DS. Scale bar is 10 nm. **b**, Cryo-EM map of TetGI-DS refined with C_2 symmetry. **c-d**, Reconstruction with symmetry expansion (SE) improves the resolution of the monomer as indicated by comparing the local resolution maps of the monomer from the C_2 dimer (C_2 monomer, **c**) and after SE (SE monomer, **d**), and Fourier shell correlation (FSC) curves (**e**) calculated for the full (solid lines) and the core (dashed lines) of the C_2 monomer (red) and SE monomer (purple). **f, g**, Cryo-EM map (**f**) and the built atomic model (**g**) colored according to the secondary structure in ***Figure A.2***. **h**, Two clipped views of the map (grey mesh) with the atomic model (sticks) illustrating the quality of the high-resolution cryo-EM map. Numbers in the corners of the figure panels (through all figures) indicate the root mean square deviation (RMSD, analogous to σ for X-ray crystallography maps) contour level of the raw EM map (without any zoning or carving applied). Numbered (1 through 4) red dashed box mark the regions in **i-l**. **i-l**, Detailed views of an example WC bp (G100:C274, **i**), a wobble bp (G220:U253, **j**), a non-canonical bp (A219:G254, **k**), and a region enclosing five Mg^{2+} cations (**l**). Blue arrows point to the feature in the map corresponding to exocyclic amino groups. In **i**, two views related by a rotation of 90° are shown with maps rendered at three different contour levels to show the quality of the map. In **l**, the same area is shown with maps rendered at two different contour levels to show the strong map intensities of Mg^{2+} cations.

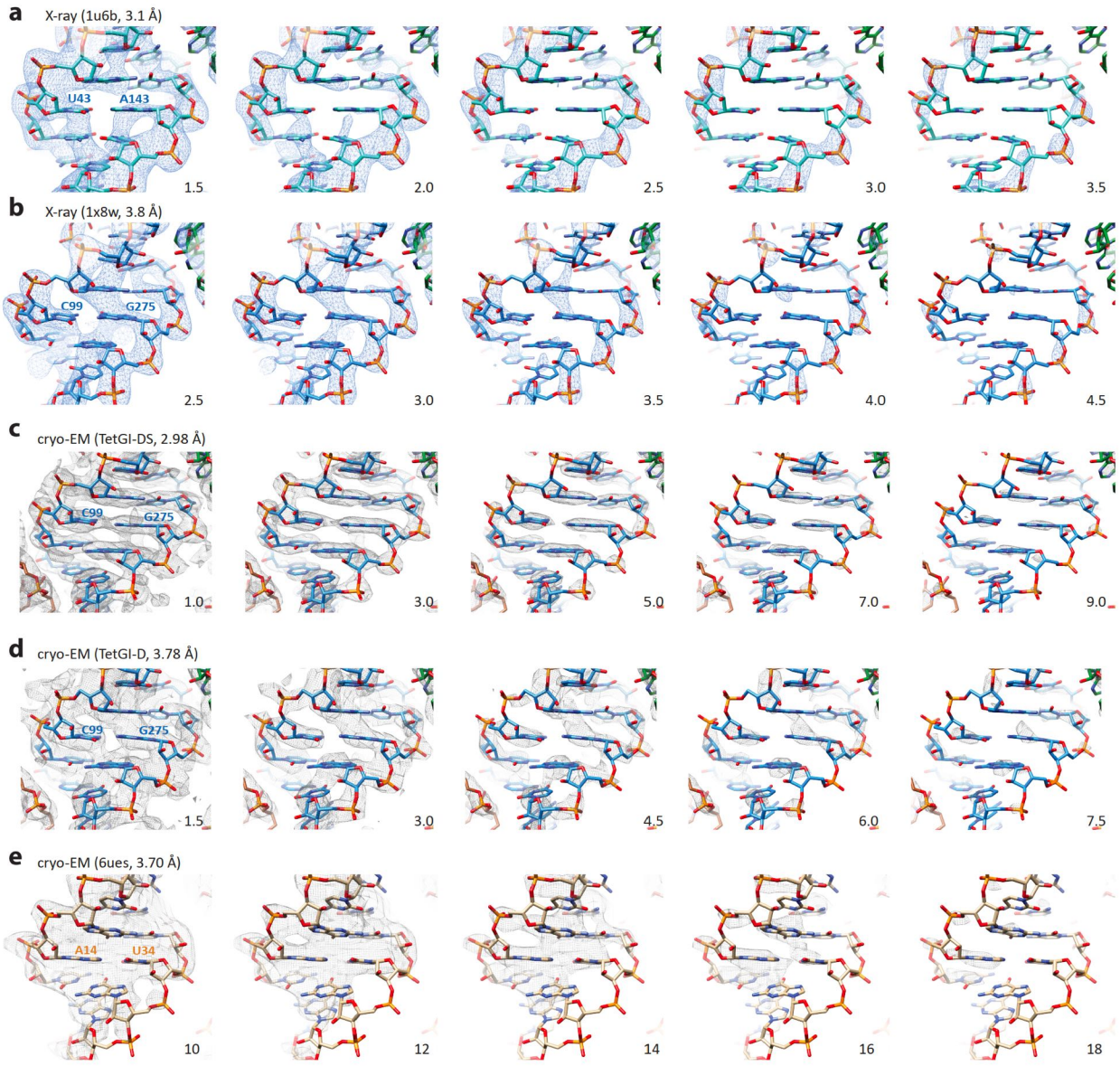


Figure A.5: Comparing features of X-ray and cryo-EM maps at different resolutions. a, b, Structural models and X-ray maps of P3 helices of the AzoGI (**a**) and TetGI (**b**). The maps are generated from structure factors deposited at the PDB (PDB codes and resolutions are given in parentheses). **c, d**, Structural models and EM maps of P3 helices of TetGI-DS (**c**) and TetGI-D (**d**). **e**, Structural model and EM map of the P2a helix of the SAM-IV riboswitch in the apo state. The contour levels (σ for X-ray maps and RMSD for EM maps) are indicated at the bottom right corner. Unlike the X-ray maps, phosphate groups are not necessarily the most prominent intensity for medium- or low-resolution cryo-EM maps of RNA due to resolution-dependent signal loss for phosphate groups in cryo-EM maps[41]. This underscores the difficulty for interpreting medium- or low-resolution cryo-EM maps of RNA.

A.2.3 Assembly, activity and cryo-EM analyses of TetGI-D and -T.

The constructs TetGI-D and -T are both designed in the post-2S state, enabling us to investigate them in parallel to assess the validity of ROCK. Because the cation identities and concentrations can profoundly influence the folding and assembly of RNAs[25], we tested the assembly of TetGI-D and -T (along with monomer control, TetGI-M) in annealing buffers containing different concentrations of Mg^{2+} and Na^{+} and analyzed the assemblies by native polyacrylamide gel electrophoresis (nPAGE) as shown in ***Figure A.6a***. Indeed, different annealing buffers resulted in different assembly patterns of the constructs and the maximum yields of the desired dimer for TetGI-D (lanes 4 and 5) and the desired trimer for TetGI-T (lanes 10 and 11) were obtained in a buffer containing no Na^{+} and 1 or 3 mM Mg^{2+}. We note that the Mg^{2+} concentrations achieving the maximum yields of the desired oligomers are close to the previously determined concentration (2 mM) optimal for the TetGI folding[42], implying that the homomeric self-assembly system can serve as a read-out platform for optimizing conditions of the ROI folding. Consequently, we assembled the TetGI-D dimer and TetGI-T trimer in large scale with a buffer of 3 mM Mg^{2+} (which is also the buffer condition for the assembly of TetGI-DS dimer) and purified them by preparative nPAGE for the subsequent activity assays and cryo-EM analyses.

To ensure the catalytic activity of the engineered homomeric TetGI constructs, we assayed their trans-acting endonucleolytic activity[36] (***Figure A.6b***). The reaction kinetics for TetGI-M, -D and -T are almost identical within the first 15 minutes of reaction when less than 50% of substrate is cleaved, indicating that configuring the TetGI within the homooligomers does not notably affect its activity. Interestingly, TetGI-D and -T are slightly slower in cleaving the rest substrate than TetGI-M, probably due to that the tighter binding of the product to the homooligomers (i.e. the avidity effect) limits the reaction rate[43] more significantly in the later stages of the assayed reaction.

Figure A.6c shows the cryo-EM map of the C_2 dimer of TetGI-D and symmetry expansion of it allows finer 3D classification, revealing two conformations that are different in P1 (***Figure A.6d,e***): ***Figure A.6d*** is the conformation with double-stranded P1 docked between the P4-P6 and P3-P9 domains; ***Figure A.6e*** is the conformation without TetLEM bound, so its P1 (or internal guide sequence, IGS) is single-stranded and undocked. Except for P1, the other parts of the two conformations are almost identical, so the final refinement was conducted by combining these two classes, resulting in a final cryo-EM map with a resolution better than either class (***Figure A.6f***). The cryo-EM reconstruction of TetGI-T (***Figure A.6g,h***) was similarly performed, and the its resolution is slightly lower than that of the TetGI-D (***Figure A.6i***). The overall architectures of the cryo-EM structures of TetGI-D and -T are almost identical and the only differences reside in the helical directions of P6b and P9.2 due to the different geometric restraints applied (***Figure A.6j***). Because TetGI-D has a higher overall resolution than TetGI-T, the map quality of TetGI-D is better than TetGI-T in most parts of structure as indicated by the atom-averaged local resolution (***Figure A.6k***) and map intensity (***Figure A.6l***) calculated for each residue. However, in the peripheral regions of P2.1 and P9.1-P9.2, the map quality of TetGI-T is comparable to or even better than that of TetGI-D, likely reflecting the reduced conformational dynamics of these regions in TetGI-T due to that the geometrically restrained P9.2. This result testifies to our proposal of using oligomerization to mitigate structural flexibility and suggests that different regions might be selectively stabilized by different oligomeric constructs.

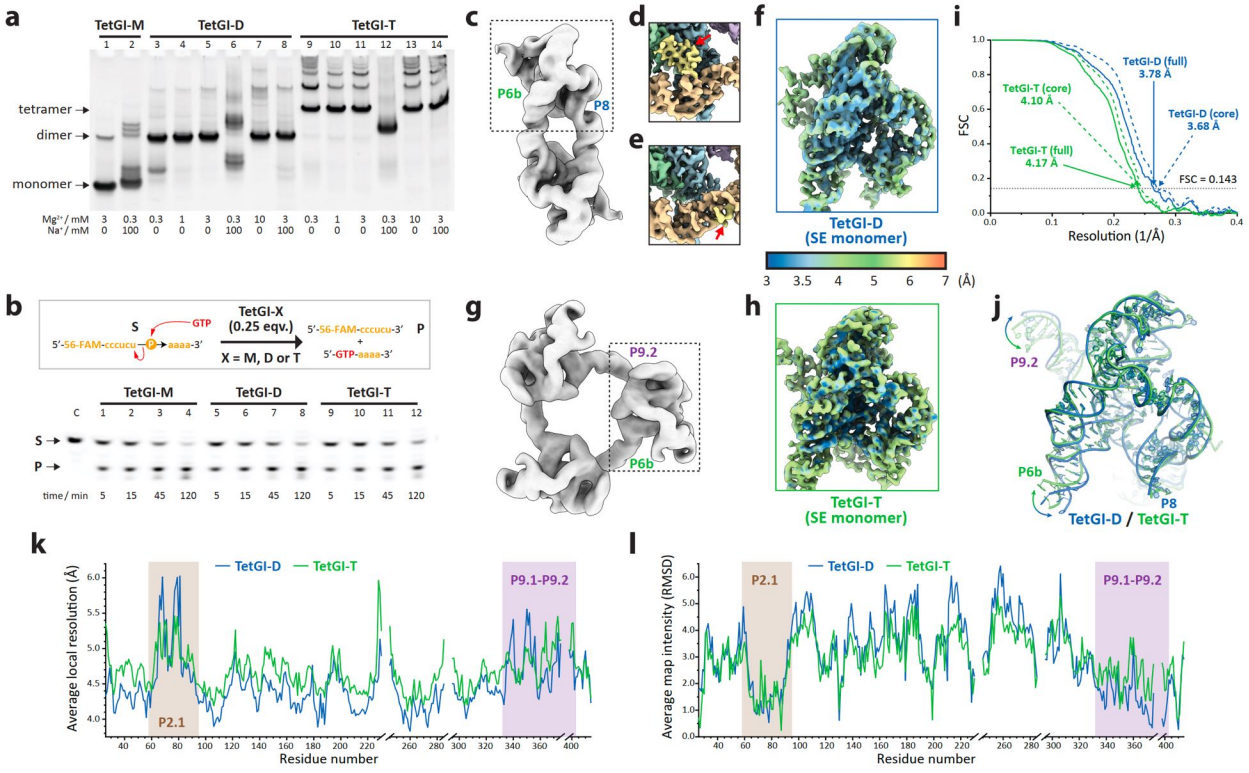


Figure A.6: Assembly, activity and cryo-EM analyses of the dimeric and trimeric post-2S TetGI constructs, TetGI-D and -T. a, Native polyacrylamide gel electrophoresis (nPAGE) analyses of the assembly products under different annealing buffers. TetGI-M is the corresponding monomer control. The bands for the target dimer and trimer, along with the monomer control, are marked on the left of the gel. **b**, The TetGI-D and -T constructs are enzymatically active as determined by the trans-acting endonucleolytic activity. The substrate (S, 2 μM) is a 10-nt single-stranded RNA 5' labeled with 6-fluorescein amidite (56-FAM) and it is cleaved by an exogenic GTP cofactor (1 mM) at 37 °C in the presence of the intron constructs (0.5 μM of monomer units). **c,** Cryo-EM map of TetGI-D refined with C_2 symmetry. **d, e**, 3D classification of the SE monomers of TetGI-D results in two distinct conformations with (**d**) or without (**e**) the binding of the ligated exon mimic (TetLEM). Red arrows point to the P1 region, which is double-stranded in **d** and single-stranded in **e**. **f**, Local resolution map of the SE monomer of TetGI-D refined by combining the two classes in **d** and **e**. **g,** Cryo-EM map of TetGI-T refined with C_3 symmetry. **h**, Local resolution map of the SE monomer of TetGI-T. The shown map is also refined by combining the two classes corresponding to the conformations with double-stranded and single-stranded P1. **i**, FSC curves calculated for the full (solid lines) and the core (dashed lines) of the TetGI-D (blue) and -T (green) SE monomer. **j**, Overlay of the structures of TetGI-D (blue) and -T (green) based on the cryo-EM maps. Double-headed arrows mark the structural differences of P6b and P9.2 of the two structures. **k, l**, Plots of the atom-averaged local resolution (**k**) and map intensity (**l**) calculated for each residue of TetGI-D (blue) and -T (green). Though TetGI-T has an overall lower map quality (marked by higher values in **k** and lower values in **l**) than TetGI-D, it has comparable or better map quality in the regions of P2.1 and P9.1-P9.2 (highlighted by brown and purple shades) because these peripheral domains are geometrically restrained TetGI-T.

A.2.4 Newly visualized interactions involving the TetGI's peripheral domains

Our cryo-EM structure supports the configurations of the TetGI's peripheral domains (P2-P2.1 and P9.1-P9.2) predicted by the decades-old computer model[29] and glimpsed by a recent medium-resolution (6.8 Å) cryo-EM structure[13]. As expected for a subgroup IC1 intron, P2-P2.1 connects P5c and P9.1-P9.2 via the tertiary base-pairings of P14 and P13 and these peripheral elements constitute a pseudo-continuous belt enclosing the core (***Figure A.7a***). Besides the overall structural organization, the present high-resolution cryo-EM structures lend us a clearer view of the structural elements involving the peripheral domains (***Figure A.7b-f***). The tertiary interaction of P14 (***Figure A.7b***) consists of the nucleotides U43, G44 and C45 (p14') pairing with A172, C170 and G169 (p14''). An unexpected feature of this interaction is the unpaired A171, which does not pair with U43 as previously suggested. Also notable is that the bulged A210 from P4, which destabilizes the folding of the isolated P4-P6 domain[31] and was either eliminated[31,44] or mutated[33] in previous structural studies, participates in a long-range base-triple with noncanonical C41:A46 pair from P2 (inset of ***Figure A.7b***), and possibly reinforces P14. The tertiary interaction of P13 (***Figure A.7c***) is a 6-bp duplex formed by the base-pairings of U75 through U80 (p13') and A352 through A347 (p13''), and stacks coaxially between the G73:C81 pair of P2.1 and the G346:C353 pair of P9.1a, bearing a conformational resemblance with some other 6-bp KL complexes[45,46]. A 4-nt bulge consisting of A69 through A72 is present near the tip of P2.1, making a major contribution to the bent shape of the P2.1-P13-P9.1a pseudo-continuous helix.

Joining P9.1-P9.2 and P9a-P9b is a four-way junction (4WJ) with the flanking helices stacked in a left-handed parallel configuration[47] (***Figure A.7d*** and its top-right insert). Though other long-range tertiary interactions may be the major determinant for the configuration of this 4WJ, it is stabilized by the sugar-phosphate interactions of the nucleotides from the exchanging

strands at the crossover site (bottom-right insert of ***Figure A.7d***). Another covalent linkage of the core and peripheral domains is the complex multiway junction connecting P2-P2.1 with P1 and P3-P8 (***Figure A.7e***). Two tertiary interactions involving the sugar edges and centered respectively by A97 and A95 (inserts of ***Figure A.7e***) have been identified to stabilize the juxtaposition of the pseudo-continuous helices of P2-P2.1 and P3-P8.

Another newly visualized long-range tertiary contact between the peripheral and core domains involves the interaction of G358 from J9.1/9.1a and the minor groove of P7 (***Figure A.7f***) and it is likely to have functional implications because (1) a similar contact involving the peripheral P7.2 also occurs in the crystal structure[48] of Twort group I intron (TwoGI, a subgroup IA2 intron), and (2) the purine-rich internal loop where the TetGI's G358 is located is conserved for different subgroup IC1 and IE introns possessing the P2.1-P13-P9.1a pseudo-continuous helix, suggesting the presence of this contact in these introns. Because the active site of the intron is located on the major-groove side of P7, this newly visualized tertiary contact is likely to buttress the active site from the other side. Previous biochemical and chemical probing studies[49,50] indicate that P9.1-P9.2 functions to stabilize the P3-P7 region of the core and its removal affects some of the intron's reactions involving the 3'-exon. However, so far, the major focus of the function of P9.1-P9.2 has been directed to the apical loop of P9.1a (p13'') due to that its participation in the P13 interaction can be readily predicted by sequence complementarity. Here, our high-resolution structure rendered by ROCK delivers an overlooked, alternative structural basis for the role of P9.1-P9.2.

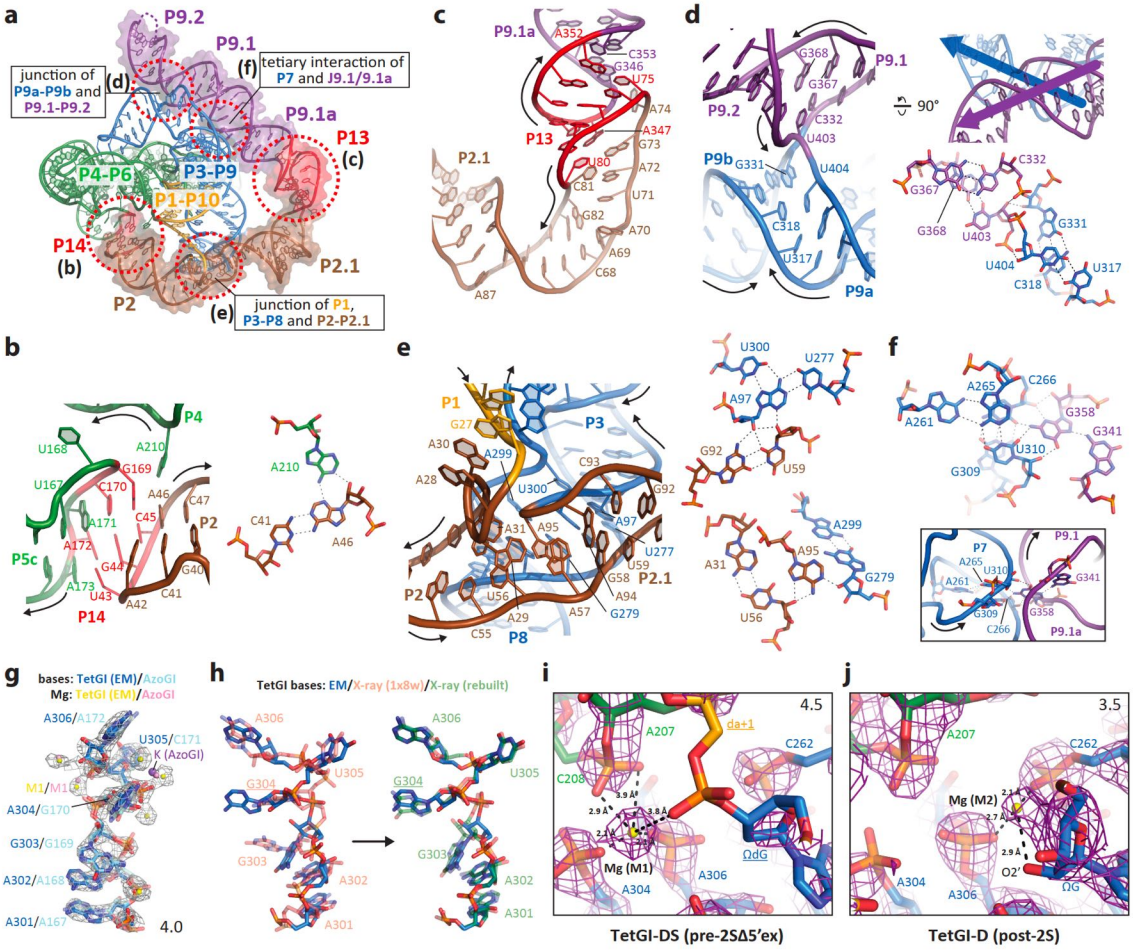


Figure A.7: Structural insights gained from the cryo-EM structures of TetGI. a, An overview of tertiary interactions and junction structures involving the peripheral domains (highlighted by surface rendering) of TetGI. Unless specified otherwise, all the TetGI structural elements are from TetGI-DS. **b, c**, Structural details of the tertiary interactions P14 (**b**) and P13 (**c**). Insert in **b** in shows the base-triple formed by the flipped-out A210 from P4 and C41 and A46 from P2. **d, e**, Structures of the four-way junction (4WJ) of P9a-P9b and P9.1-P9.2 (**d**), and the junctions of P1, P3-P8 and P2-P2.1 (**e**). Inserts of **d** show the strand directions of the continuous strands of the 4WJ (top), and the interaction details at the crossover site (bottom). Inserts of **e** show the details of two tertiary contacts between the P2-P2.1 and P3-P9 domains. **f**, Structural details of the tertiary contact formed by the docking of G358 of J9.1/9.1a into the minor groove of P7. **g**, A comparison of the J8/7 region of our TetGI EM structure and the Azoarcus group I intron (AzoGI) crystal structure (PDB code: 1u6b) reveals close structural resemblance. Gray mesh is the EM map contoured at 4.0 RMSD level and carved within 2.0 Å of the displayed atoms. **h**, The high-resolution cryo-EM structure enables the rebuilding and re-refinement of J8/7 of the previous crystal structure of the TetGI core (PDB code: 1x8w). The structures before (pink) and after (green) rebuilding are overlayed with our EM structure (blue). Underlined nucleotides indicate the mutations introduced for the crystallographic study. **i, j**, Active-site Mg^{2+} cations, M1 (functioning as activating the attacking nucleophile) and M2 (functioning as stabilizing the leaving group), are observed in the EM maps (magenta meshes, contour levels indicted in the top-right corner) of TetGI-DS (**i**) and TetGI-D (**j**), respectively.

A.2.5 J8/7 and active-site magnesium ions of TetGI

In the pre-2SΔ5'ex construct of TetGI-DS, P1 is in the single-stranded state, leading us to compare this construct's J8/7 (functioning as a docking site for P1) with those in the previous crystal structures of AzoGI[34] (with a double-stranded P1-P2; P2 considered as part of P1) and TetGI core[33] (without P1). As shown in ***Figure A.7g***, the configurations of J8/7 in the TetGI-DS cryo-EM structure and the AzoGI crystal structure are nearly identical. Some discrepancies lie in the nearby metal ions: for instance, we spot a Mg^{2+} ion interacting with the phosphates of A301 and A302 (corresponding to A167 and A168 of the AzoGI) that has been shown theoretically to stabilize the stack-exchange junction at P3-P8[51], but not observed in the AzoGI crystal structure. Though our cryo-EM model and the previous X-ray model of TetGI considerably differ in J8/7 (***Figure A.7f,g***), we noticed that the X-ray model was poorly built in this region. After model rebuilding and re-refinement based on our cryo-EM structure, we see a substantial improvement of the map quality and refinement statistics, and the structure differences in this region are eliminated (***Figure A.7h***). These results indicate that J8/7 has been pre-organized for the P1 docking and showcase the potential of high-resolution cryo-EM structures in improving the RNA model building that has been difficult for low-resolution RNA crystals.

In the cryo-EM structure of TetGI-DS (***Figure A.7i***), we observed an active-site Mg^{2+} corresponding to the location of M1 observed in the crystal structure of the AzoGI ribo-ΩG pre-2S complex[52]. M1 functions mainly as a nucleophile activator, and its presence in the 5'-exon-free intron (pre-2SΔ5'ex state) suggests its role of activating a water molecule in the 3' splice site hydrolysis reaction[37]. However, the other metal ligand M2 (functioning as stabilizing the leaving group) in the AzoGI crystal structure is not observed in our TetGI-DS construct, likely due to the deoxy substitution of ΩG (G414). We note that the M2, though present in the deoxy-ΩG pre-2S

AzoGI[34], is a monovalent metal. In our current cryo-EM study, the only metal ion present in the buffer is Mg^{2+}. Nonetheless, the density for M2 could be spotted in the cryo-EM map of TetGI-D, which has a ribo-ΩG (***Figure A.7j***), though we could not unambiguously build M1 due to the limited map quality of TetGI-D.

A.2.6 Extending the application of ROCK to smaller RNA structures

We next set out to assess ROCK for smaller RNA structures and with more readily accessible instruments (Polara microscope with K2 camera, rather than Krios microscope with K3 camera for the TetGI structures). The AzoGI is a subgroup IC3 intron of 206 nt (~70 kDa): compared with the TetGI, it lacks the extensive peripheral domains that function to facilitate RNA folding and to reinforce the overall structure. Both the smaller molecular weight and the presumed increase of structure flexibility make the AzoGI a more challenging ROI for cryo-EM. However, the smaller size and simple fold of the AzoGI make it attractive for crystallographic studies[34,52,53], which provide us the structural basis for ROCK engineering. We designed a construct based on the post-2S complex of AzoGI and engineered its P5a and P8a for the assembly of a homotrimer (***Figure A.8a,b***; referred to as AzoGI-T). Similar to the TetGI-D and -T, two deoxy substitutions are introduced in the in-trans added ligated exon mimic (AzoLEM). Using a similar workflow as the TetGI constructs, we obtained a 4.9 Å cryo-EM map of the SE monomer (***Figure A.8c,d***). This map corresponds to a conformation with P1-P10 tightly docked (***Figure A.8e***), which is the most populated and best resolved conformation from 3D classification and is similar as the post-2S AzoGI crystal structure[53]. Besides, we refined the cryo-EM map corresponding to an alternative open conformation (***Figure A.8f***). Though this map is of a significantly lower resolution (~ 8Å), the relative movement of the P2-P1-P10 and P4-P5 could be clearly discerned (***Figure A.8f***),

demonstrating the power of cryo-EM in the study of RNA conformational dynamics.

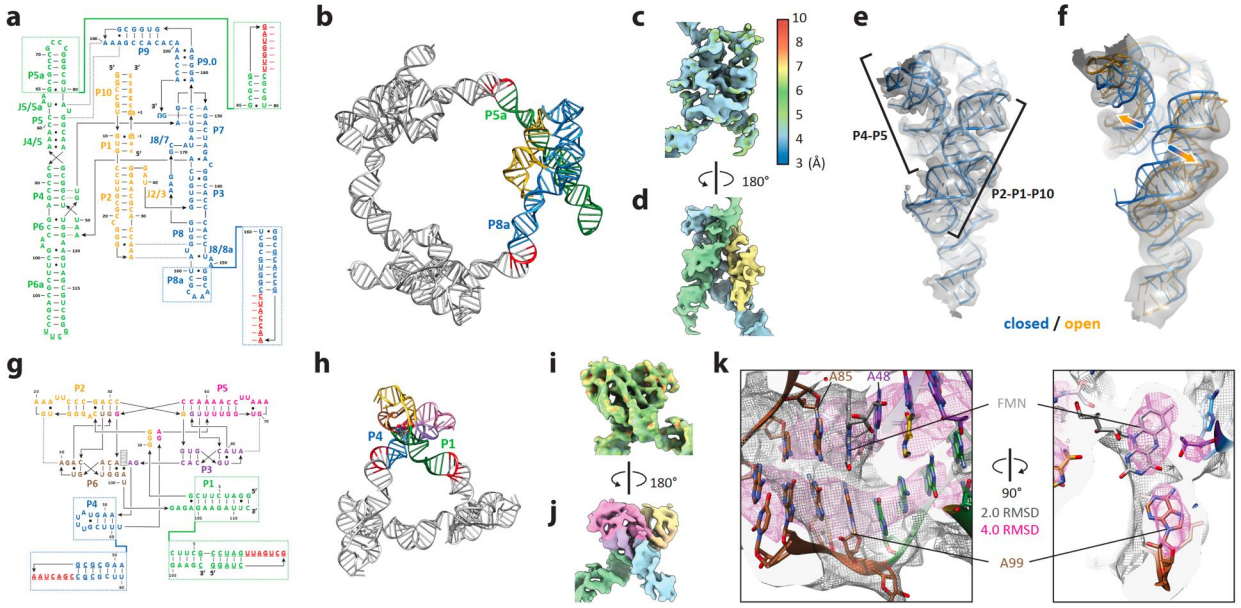


Figure A.8: The homomeric self-assembly strategy applied to two smaller structured RNAs. **a**, Sequence and secondary structure of the Azoarcus group I intron (AzoGI). Dashed boxes mark the engineering of P5a and P8a for the trimeric self-assembly, AzoGI-T. **b**, Computer model of the assembled homotrimer of AzoGI-T. **c, d**, Cryo-EM maps rendered by local resolution (**c**) and coloring (**e**) based on the secondary structure in **a**. **e**, **f,** A comparison of the best-resolved closed conformation (**e**) from 3D classification of the SE monomer and another less-resolved (~8.0 Å) open conformation. In **f**, the fitted atomic model of the open conformation (orange) is overlayed with that of the close conformation (blue) and the arrows mark the structural movement of P2-P1-P10 and P4-P5. **g**, Sequence and secondary structure of the Fusobacterium nucleatum FMN-riboswitch (FMNrsw). Dashed boxes mark the engineering of P1 and P4 for the trimeric self-assembly, FMNrsw-T. **h**, Computer model of the assembled homotrimer of FMNrsw-T. **i, j**, Cryo-EM maps rendered by local resolution (**i**) and coloring (**j**) based on the secondary structure in **g**. The color keys for **i** are the same as **c**. **k**, Clipped views showing the cryo-EM map of the FMN ligand and its binding environment. The cryo-EM map is rendered at two different contour levels: grey mesh at 2.0 RMSD and magenta mesh at 4.0 RMSD.

Lastly, we challenged ROCK to an even smaller ROI—the Fusobacterium nucleatum FMN-riboswitch (FMNrsw) of 112 nt (~35 kDa). Based on its crystal structure[54], we installed the KL sequences onto its P1 and P4 for the assembly of a homotrimer (***Figure A.8g,h***; referred to as FMNrsw-T). Interestingly, this construct assembles into two different homooligomers, dimer or trimer, when annealed in different buffer conditions, and ligand-binding assays show that the RNA within the expected trimer is in its functionally relevant conformation—the dimer therefore must be a kinetic product with distorted structures. Cryo-EM analysis was then conducted on the purified trimer with ligand bound, and due to the small size of this ROI, the final refinement after SE was conducted on the whole trimer particles with one monomer focused. Ultimately, we obtained the 5.9 Å resolution cryo-EM map of the focused monomer within the FMNrsw-T trimer (***Figure A.8i,j***). The presence of the bound ligand along with its binding environment can be visualized in this medium-resolution cryo-EM map thanks to the strong map intensities of the ligand and its vicinity (***Figure A.8k***).

	TetGI-DS	TetGI-D	TetGI-T	AzoGI-T	FMNrw-T
	PDBid: 7R6L EMDB: 24281	PDBid: 7R6M EMDB: 24282	PDBid: 7R6N EMDB: 24283	EMDB: 24284	EMDB: 24285
Data collection and processing					
Microscope	Titan Krios	Titan Krios	Titan Krios	Polara	Polara
Detector	K3	K3	K3	K2 Summit	K2 Summit
Magnification	105,000	105,000	105,000	31,000	31,000
Voltage (kV)	300	300	300	300	300
Electron Exposure ($e^-/Å^2$)	47	47	47	52	52
Defocus Range (μm)	0.8-2.0	0.8-2.0	0.8-2.0	1.0-2.5	1.0-2.5
Acquisition Pixel Size (Å)	0.825	0.825	0.825	1.23	1.23
Final symmetry imposed	C1	C1	C1	C1	C1
Initial particle images (#)	550,754	384,170	124,353	474,349	471,007
Final particle images (#)	82,575 S.E. monomers	113,548 S.E. monomers	85,596 S.E. monomers	486,860 S.E. monomers	266,623 S.E. monomers
Reconstruction Pixel Size (Å)	0.825	1.03	1.24	1.23	1.23
Map resolution (Å)	2.85 core / 2.98 overall	3.68 core / 3.78 overall	4.10 core / 4.17 overall	4.9	5.9
FSC threshold	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.5-6.0	3.5-8.0	3.0-7.0	4.0-8.0	4.0-8.0
Refinement					
Model resolution (Å)	2.97	3.82	4.20		
FSC threshold	0.5	0.5	0.5		
Map sharpening B factor ($Å^2$)	N/A	N/A	N/A	N/A	N/A
Model composition					
Non-hydrogen atoms	7885	7810	7588		
RNA bases	368	365	354		
Ligand	16	12	12		
B factor ($Å^2$)					
RNA bases	70.20	104.18	89.86		
Ligand	36.67	57.50	46.19		
R.m.s. deviations					
Bond lengths (Å)	0.008	0.009	0.008		
Bond angles (°)	0.782	1.109	1.069		
Validation					
MolProbity score	2.42	2.46	2.41		
Clashscore	5.06	5.70	4.91		

Figure A.9: Data collection table and model refinement statistics

A.3 Discussion

In the present work, we have determined the cryo-EM structures of three different ROIs with sizes ranging from ~35 to ~140 kDa and belonging to two important categories of functional RNAs—ribozymes and riboswitches, showcasing the potential universality ROCK. Though strategies for construct engineering are common in RNA crystallography[55], their results are usually unforeseeable, and their success largely depends on serendipity. ROCK represents a rational construct engineering strategy for RNA cryo-EM studies thanks to the programmability of RNAs and the advances of nucleic acid nanotechnology[17-21]. Previous RNA cryo-EM studies aiming for high-resolution structures entail a large number of initial particles[11-13], making the workflow experimentally and computationally demanding. Our strategy greatly improves the usable particles via mitigating the conformational dynamics using geometric restraints and improving sample homogeneity by natively purifying the target homooligomers. Further, the resulting symmetric homomeric assemblies are also preferred subjects for cryo-EM: (1) the characteristic shapes are more convenient for initial alignment of the particles; and (2) special EM image processing procedures such as symmetry expansion and individual subunit-focused classification and refinement can be utilized for achieving unprecedented resolution for RNA-only structures. The principle of ROCK is also envisioned to apply for RNA-protein complexes or even protein-only structures[56,57].

It is also important to realize the limitations of ROCK, but they are not unsolvable. Firstly, prior knowledge of the ROI's structures is required for construct engineering. This can be achieved by exploiting atomic models from solved partial structures (by crystallography or NMR) or computer modelling, both of which have been demonstrated in our current work, or via other methods such as atomic force microscopy[58], small-angle X-ray scattering[59], comparative gel

electrophoresis[60], and preliminary low- or medium-resolution cryo-EM models[13]. Secondly, de novo designed oligomerization is likely to alter the native structure and function. However, this is a universal concern for all structure biology studies, and is not unique to ROCK: activity assays are inevitable to ensure the biological meaningfulness of the structural insights obtained. Thirdly, the ROI should have at least two nonessential helices for ROCK engineering. This concern could be addressed by choosing the RNA possessing more extended peripheral structural elements, or by performing the operation of motif fusion[25] to attach additional helices to the ROI. We believe that ROCK unleashes the largely unexplored potential of cryo-EM in RNA structural studies, opening new opportunities for elucidating the mechanisms of functional RNAs and facilitating the design- and structure-based approaches to the invention of RNA-targeting therapeutics[61].

A.4 Methods

RNA construct design

The software NanoTiler[39] (developed by Bruce Shapiro's group) was used to design the RNA constructs with kissing-loops (KLs) installed for homomerization. Throughout this work, we chose the 7-bp KL that is adapted from the RNA I-RNA II complex of the *E. coli ColE1* plasmid[62] and has been determined by NMR[38] (PDB code: 2bj2) to have an included angle of ~120°. For TetGI-D and TetGI-DS, the design was based on crystal structure[33] of the core domains of the TetGI (PDB code: 1x8w). Before being imported into NanoTiler as structural modules, the PDB files of 1x8w (the TetGI core) and 2bj2 (the 7-bp KL) were first edited so that the helices (P6b and P8 for the TetGI core and the two stems of the 7-bp KL) to be connected in the final construct were trimmed to 3 bps. Then NanoTiler's command "clone" was used to generate another copy of these structural models; commands "genhelixconstraint" and "opthelices" were used to generate ideal A-form RNA helices of certain lengths for connecting the TetGI core and the 7-bp KL; the "start_score" in the output of the "opthelices" for placing the last connecting helices indicates how good is the ring closure and a lower value indicates better ring closure. The processes can be automated with the "NanoScript" of NanoTiler in which "foreach" loops can be incorporated to iterate through connecting helices of different lengths for screening for the optimal lengths for ring closure. Similarly, the trimeric construct TetGI-T was designed based on the computational model[29] of the complete TetGI kindly provided by Prof. Eric Westhof; AzoGI-T based on the crystal structure[34] of the Azoarcus group I intron (AzoGI) (PDB code: 1u6b); and FMNrsw-T based on the crystal structure[54] of the FMN riboswitch (FMNrsw) bound to FMN (PDB code: 3f2q). After installing the KL sequence and connector helices onto the peripheral stem-loops of each target RNA, the secondary structure of each newly formed stem-loop was further checked by Mfold[63] to ensure

that the stem-loop folds correctly.

RNA preparation and nanostructure assembly

All RNA molecules were synthesized by *in vitro* transcription (IVT) using the HiScribe™ T7 High Yield RNA Synthesis Kit from the New England Biolabs (NEB). The corresponding DNA templates for *in vitro* transcription were the PCR products of gene fragments ordered from WuXi Qinglan Biotech. The PCR experiments were conducted using the Q5® Hot Start High-Fidelity DNA Polymerase (NEB) following the recommended protocol provided by NEB. Primers and other modified oligonucleotides were ordered from the Integrated DNA Technologies (IDT). All IVT RNA molecules were purified by denaturing PAGE (containing 7M urea), then ethanol precipitated and suspended in pure water. The RNA concentration was determined by measuring OD_{260}.

Protocols for RNA nanostructure assembly were adapted from a previous publication[25]. RNAs were first denatured at 85 °C for 1 min and snap-cooled on ice. Then, the annealing buffers containing 20 mM of Tris-acetate (pH 8.0) and varied concentrations of Mg^{2+} (using 100 mM Mg^{2+} stock solution containing 110 mM of $MgCl_2$ and 10 mM of EDTA) and Na^+ (using 1 M Na^+ stock solution containing 1 M of NaCl) were added to the denatured RNAs (to a final RNA concentration of ~800 nM). The mixtures were then annealed from 70 °C to 4 °C with the following protocol: 70 °C to 50 °C over 6 min, 50 °C to 37 °C over 20 min, 37 °C to 4 °C over 2 hr. The annealed RNAs were analyzed by native polyacrylamide gel electrophoresis (nPAGE) in 0.5 × TBE buffer supplemented by 3 mM of $MgCl_2$ (so that the final concentration of Mg^{2+} is 2 mM because 1 mM of EDTA is included in 0.5 × TBE). To prepare the samples for cryo-EM experiments, the cation contents in the annealing buffers were chosen based on the nPAGE results: for TetGI-DS, TetGI-D, TetGI-T and AzoGI-T, the annealing buffer contains 1 mM of Mg^{2+} and 1.2 equivalents (equiv.)

of associated synthetic oligonucleotides (dTetCIRC for TetGI-DS; TetLEM for TetGI-D and -T; and AzoLEM for AzoGI-T); for FMNrsw-T, the annealing buffer contains 0.3 mM of Mg^{2+} and 10 mM of Na^{+} supplemented by 1 mM of FMN. The assembled nanostructures were purified by preparative nPAGE, eluted to the annealing buffer of 3 mM Mg^{2+}, and concentrated to ~1 μg/μL using Amicon Ultra centrifugal filters (MWCO 30 kDa).

Cryo-EM sample preparation and data acquisition

Before grid preparation, Mg^{2+} and other components (such as short oligonucleotides or ligand) were added to the purified and concentrated RNA assemblies. For the four ribozyme constructs (TetGI-DS, TetGI-D, TetGI-T and AzoGI-T), Mg^{2+} was added to a final concentration of 30 mM to ensure complete folding[51]. Additional synthetic oligonucleotide was added for each of the ribozyme construct in case of the dissociation due to equilibrium or during the purification processes: for TetGI-DS, 1.2 equiv. of dTetCIRC was added; for TetGI-D and -T, 3 equiv. of TetLEM was added; and for AzoGI-T, 3 equiv. of AzoLEM was added. For FMNrsw-T, Mg^{2+} was added to a final concentration of 10 mM, and FMN to a final concentration of 200 μM. After the addition of Mg^{2+} and other components, the samples were left at room temperature for 15 min before grid preparation.

Each RNA sample was applied to a glow-discharged Quantifoil (R1.2/1.3, 400 mesh) holey carbon grid. The three TetGI samples were vitrified in liquid ethane with a Thermo Fisher Vitrobot Mark IV (blot for ~7s, force +12) and imaged on a Titan Krios microscope equipped with a Gatan K3 camera. The Krios movie stacks were acquired in counting mode, with a physical pixel size of 0.825 Å, total exposure of 47 e/Å^2 and at a defocus ranging from -0.8 to -2.0 μm. The AzoGI-T and FMNrsw-T samples were frozen using a Gatan Cryoplunge 3 (blot ~3s) and imaged on an FEI T30 Polara microscope equipped with a K2 summit detector. The Polara micrographs were

acquired in super-resolution mode, with 0.62 Å super-resolution pixel size and 52 e/Å^2 exposure, with a defocus ranging from -0.8 to -2.0 μm. SerialEM[64] was used for collection of all datasets. More information about cryo-EM data collection can be found in ***Figure A.9***.

Cryo-EM data processing

The overall processing pipeline was similar for all the datasets. Briefly, movie stacks were binned 2× if collected in super-resolution mode, then motion-corrected with MotionCor2[65]. The contrast transfer function was subsequently computed with CTFFIND4[66]. Particle picking and 2D particle curation were performed with Simplified Application Managing Utilities of EM Labs (https://liao.hms.harvard.edu/samuel), a set of protocols built on the SPIDER[67] image processing system. About 2,000 particles were manually picked to generate 2D initial models, which were used to auto-pick 10% of the micrographs and generate refined 2D templates. The refined templates were then used to auto-pick from all micrographs. The particles were curated with "samtree2dv3.py", which runs iterative principal component analysis (PCA), k-means clustering and multireference alignment. Selected particles were then imported to RELION 3.0[68], which was used for all subsequent processing steps. Briefly, the particles were downsampled and sorted with 3D classification, before unbinning and 3D refinement with enforced symmetry. As the constructs include monomers of uneven integrity, we used symmetry expansion[69] to address the pseudo-symmetry of the particles and separate the most stable monomers. Monomers were then classified through 3D classification without particle alignment, and selected classes were refined using local angle search refinement. Monomer resolution was calculated with the FSC = 0.143 criterion on half maps from independent halves of the dataset, while local resolution was determined with ResMap[70]. All maps were density modified with "phenix.resolve_cryo_em"[71] that is recently incorporated in PHENIX[72], and no model information was provided to avoid any possible bias.

Structure modelling

To build the structural model of the TetGI, we began with the core domains by rigid-body fitting the crystallographic model of the core (PDB code: 1x8w) into the TetGI-DS map using UCSF Chimera[73] and performing several rounds of real-space refinements[74] in PHENIX[72]. Though the crystallographic model mostly agreed with our cryo-EM map, two regions of the core show the most significant discrepancy—the region of P9.0 and G414, which were not present in the crystallographic model, and the region near J8/7, which we believe was built incorrectly in the crystallographic model. These two regions were manually built in COOT[75]. To build the peripheral domains of TetGI, ideal A-form helices were generated and fitted to the TetGI-DS map using UCSF Chimera[73] based on the known secondary structure. The rest of structure was manually built in COOT[75]. The most difficult region to build was P2.1 to P13, where the map density was relatively weaker and the secondary structure assignment was ambiguous. For instance, any residue from A87 through A90 could be possibly unpaired, and the exact number of base-pairs in kissing-loops P13 was not known. We manually tested different possibilities of secondary structure assignments for this region to ensure its successful joining to the helices on both sides. The TetGI-T map was referenced to build the region of P2.1 to P13 because the map quality is better for this region than the TetGI-DS map. The complete TetGI model was finally real-space refined against the maps of the three TetGI constructs and manually checked and adjusted in COOT. The statistics of model refinement and validation were tabulated in ***Figure A.9***.

For AzoGI-T, we used the AzoGI crystal structure (PDB code: 1u6b) as the initial model, which was rigid-body fitted to the map to the AzoGI-T map using Chimera, and then the atomic model was iteratively improved by real-space refinements in PHENIX and manual adjustments in COOT. Because the final cryo-EM map for FMNrsw-T is the trimer instead of SE monomer, the

designed NanoTiler model was used as the initial model to fit the map by MDFF[76], and then the atomic model was iteratively improved by real-space refinements in PHENIX and manual adjustments in COOT.

A.5 References

1. Breaker, R.R. & Joyce, G.F. The expanding view of RNA and DNA function. *Chem Biol* **21**, 1059-65 (2014).

2. Mortimer, S.A., Kidwell, M.A. & Doudna, J.A. Insights into RNA structure and function from genome-wide studies. *Nat Rev Genet* **15**, 469-79 (2014).

3. Serganov, A. & Patel, D.J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**, 776-90 (2007).

4. Hangauer, M.J., Vaughn, I.W. & McManus, M.T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet* **9**, e1003569 (2013).

5. Robertson, D.L. & Joyce, G.F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467-8 (1990).

6. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-22 (1990).

7. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).

8. Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal* **9**, 1665-1672 (1990).

9. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat Methods* **13**, 24-7 (2016).

10. Qu, G. et al. Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol* **23**, 549-57 (2016).

11. Li, S. et al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat Struct Mol Biol* **26**, 1094-1105 (2019).

12. Zhang, K. et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 A resolution. *Nat Commun* **10**, 5511 (2019).

13. Kappel, K. et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods* **17**, 699-707 (2020).

14. Abels, J.A., Moreno-Herrero, F., van der Heijden, T., Dekker, C. & Dekker, N.H. Single-molecule measurements of the persistence length of double-stranded RNA. *Biophys J* **88**, 2737-44 (2005).

15. Choe, S. & Sun, S.X. The elasticity of alpha-helices. *J Chem Phys* **122**, 244912 (2005).

16. Kessel, A. & Ben-Tal, N. *Introduction to proteins: structure, function, and motion*, (CRC Press, 2010).

17. Seeman, N.C. Nanomaterials based on DNA. *Annual Review of Biochemistry* **79**, 65-87 (2010).

18. Seeman, N.C. *Structural DNA Nanotechnology*, (Cambridge University Press, 2016).

19. Seeman, N.C. & Sleiman, H.F. DNA nanotechnology. *Nature Reviews Materials* **3**, 17068 (2017).

20. Guo, P. The emerging field of RNA nanotechnology. *Nature Nanotechnol.* **5**, 833-42 (2010).

21. Grabow, W.W. & Jaeger, L. RNA self-assembly and RNA nanotechnology. *Acc Chem Res* **47**, 1871-80 (2014).

22. Zhang, X., Yan, H., Shen, Z. & Seeman, N.C. Paranemic Cohesion of Topologically-Closed DNA Molecules. *Journal of the American Chemical Society* **124**, 12940-12941 (2002).

23. Wang, X. et al. Paranemic Crossover DNA: There and Back Again. *Chem Rev* **119**, 6273-6289 (2018).

24. Geary, C.W. & Andersen, E.S. Design Principles for Single-Stranded RNA Origami Structures. 1-19 (Springer International Publishing, Cham, 2014).

25. Liu, D. et al. Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. *Nature Chemistry* **12**, 249-259 (2020).

26. Kruger, K. et al. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147-157 (1982).

27. Hougland, J.L., Piccirilli, J.A., Forconi, M., Lee, J. & Herschlag, D. How the Group I Intron Works: A Case Study of RNA Structure and Function. in *RNA World* (eds. Gesteland, R.F., Atkins, J.F. & Cech, T.R.) 133-205 (Cold Spring Harbor Laboratory Press, 2006).

28. Golden, B.L. Group I Introns: Biochemical and Crystallographic Characterization of the Active Site Structure. in *Ribozymes and RNA Catalysis* 178-200 (The Royal Society of Chemistry, 2007).

29. Lehnert, V., Jaeger, L., Michele, F. & Westhof, E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chemistry & Biology* **3**, 993-1009 (1996).

30. Cate, J.H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-85 (1996).

31. Juneau, K., Podell, E., Harrington, D.J. & Cech, T.R. Structural Basis of the Enhanced Stability of a Mutant Ribozyme Domain and a Detailed View of RNA–Solvent Interactions. *Structure* **9**, 221-231 (2001).

32. Golden, B.L., Gooding, A.R., Podell, E.R. & Cech, T.R. A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science* **282**, 259-64 (1998).

33. Guo, F., Gooding, A.R. & Cech, T.R. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* **16**, 351-62 (2004).

34. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. & Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50 (2004).

35. Hahn, C.S., Strauss, E.G. & Strauss, J.H. Dideoxy sequencing of RNA using reverse transcriptase. *Methods Enzymol* **180**, 121-30 (1989).

36. Zaug, A.J., Been, M.D. & Cech, T.R. The Tetrahymena ribozyme acts like an RNA restriction endonuclease. *Nature* **324**, 429-33 (1986).

37. Inoue, T., Sullivan, F.X. & Cech, T.R. New reactions of the ribosomal RNA precursor of Tetrahymena and the mechanism of self-splicing. *Journal of Molecular Biology* **189**, 143-165 (1986).

38. Lee, A.J. & Crothers, D.M. The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* **6**, 993-1007 (1998).

39. Bindewald, E., Grunewald, C., Boyle, B., O'Connor, M. & Shapiro, B.A. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. *J Mol Graph Model* **27**, 299-308 (2008).

40. Woodson, S.A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* **9**, 104-9 (2005).

41. Wang, J. & Moore, P.B. On the interpretation of electron microscopic maps of biological macromolecules. *Protein Sci* **26**, 122-129 (2017).

42. Rook, M.S., Treiber, D.K. & Williamson, J.R. An optimal Mg(2+) concentration for kinetic folding of the tetrahymena ribozyme. *Proc Natl Acad Sci U S A* **96**, 12471-6 (1999).

43. Herschlag, D. & Cech, T.R. Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. *Biochemistry* **29**, 10172-80 (1990).

44. Ye, J.D. et al. Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc Natl Acad Sci U S A* **105**, 82-7 (2008).

45. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nature Struct. Biol.* **8**, 1064-8 (2001).

46. Lebars, I. et al. Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res* **36**, 7146-56 (2008).

47. Lilley, D.M. Structures of helical junctions in nucleic acids. *Quarterly Reviews of Biophysics* **33**, 109-159 (2000).

48. Golden, B.L., Kim, H. & Chase, E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* **12**, 82-9 (2005).

49. Barfod, E.T. & Cech, T.R. Deletion of nonconserved helices near the 3' end of the rRNA intron of Tetrahymena thermophila alters self-splicing but not core catalytic activity. *Genes Dev* **2**, 652-63 (1988).

50. Laggerbauer, B., Murphy, F.L. & Cech, T.R. Two major tertiary folding transitions of the Tetrahymena catalytic RNA. *The EMBO Journal* **13**, 2669-2676 (1994).

51. Denesyuk, N.A. & Thirumalai, D. How do metal ions direct ribozyme folding? *Nat Chem* **7**, 793-801 (2015).

52. Stahley, M.R. & Strobel, S.A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587-90 (2005).

53. Lipchock, S.V. & Strobel, S.A. A relaxed active site after exon ligation by the group I intron. *Proc Natl Acad Sci U S A* **105**, 5699-704 (2008).

54. Serganov, A., Huang, L. & Patel, D.J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233-7 (2009).

55. Zhang, J. & Ferre-D'Amare, A.R. New molecular engineering approaches for crystallographic studies of large RNAs. *Curr Opin Struct Biol* **26**, 9-15 (2014).

56. Liu, Y., Gonen, S., Gonen, T. & Yeates, T.O. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc Natl Acad Sci U S A* **115**, 3362-3367 (2018).

57. Liu, Y., Huynh, D.T. & Yeates, T.O. A 3.8 A resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat Commun* **10**, 1864 (2019).

58. Schön, P. Imaging and force probing RNA by atomic force microscopy. *Methods* **103**, 25-33 (2016).

59. Chen, Y. & Pollack, L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip Rev RNA* **7**, 512-26 (2016).

60. Lilley, D.M. Analysis of branched nucleic acid structure using comparative gel electrophoresis. *Q Rev Biophys* **41**, 1-39 (2008).

61. Warner, K.D., Hajdin, C.E. & Weeks, K.M. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **17**, 547-558 (2018).

62. Tomizawa, J.-i. Control of ColE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* **38**, 861-870 (1984).

63. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15 (2003).

64. Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y. & Mastronarde, D.N. Software tools for automated transmission electron microscopy. *Nat Methods* **16**, 471-477 (2019).

65. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).

66. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-21 (2015).

67. Shaikh, T.R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat Protoc* **3**, 1941-74 (2008).

68. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).

69. Scheres, S.H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-57 (2016).

70. Kucukelbir, A., Sigworth, F.J. & Tagare, H.D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-5 (2014).

71. Terwilliger, T.C., Ludtke, S.J., Read, R.J., Adams, P.D. & Afonine, P.V. Improvement of cryo-EM maps by density modification. *Nat Methods* **17**, 923-927 (2020).

72. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).

73. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).

74. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).

75. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).

76. McGreevy, R., Teo, I., Singharoy, A. & Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50-60 (2016).