Conformational dynamics and substrate specificity in nutrient transporters

Dissertation zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) an der

Universität Hamburg

Fakultät für Mathematik, Informatik und Naturwissenschaft

Fachbereich Chemie

Institut für Biochemie und Molekularbiologie

Vorgelegt von

Kim Bartels

aus Bonn

Hamburg, September 2020

Chairman of the evaluation committee:

Prof. Andrew Torda

Professor at the Institute of Biochemistry and Molecular Biology (University of Hamburg)

Members of the evaluation committee:

Prof. Daniel Wilson

Professor at the Institute of Biochemistry and Molecular Biology (University of Hamburg)

Dr. Christian Löw

Group leader at the European Molecular Biology Laboratory (EMBL) and the Centre of Structural

Systems Biology (CSSB)

Thesis evaluator:

Dr. Christian Löw

Group leader at the European Molecular Biology Laboratory (EMBL) and the Centre of Structural

Systems Biology (CSSB)

Thesis evaluator:

Prof. Dr. Henning Tidow

Professor at the Institute of Biochemistry and Molecular Biology (University of Hamburg)

Date of defense: 11.12.2020

Print permission: 11.12.2020

The work presented in this dissertation was performed between June 2017 and September 2020 in the group of Dr. Christian Löw at the European Molecular Biology Laboratory (EMBL) in Hamburg, Germany.

Table of Contents

I.   List of publications

1.   Membrane Chemistry Tunes the Structure of a Peptide Transporter. Lasitza-Male T, **Bartels K**, Jungwirth J, Wiggers F, Rosenblum G, Hofmann H, Löw C. Angew Chem Int Ed Engl. 2020. doi: 10.1002/anie.202008226. Epub ahead of print. (Shared first author with Lasitza-Male T)

2.   Structure of Prototypic Peptide Transporter DtpA from *E. coli* in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1. Ural-Blimke Y, Flayhan A, Strauss J, Rantos V, **Bartels K**, Nielsen R, Pardon E, Steyart J, Kosinski J, Quistgaard EM, Löw C. J Am Chem Soc. 2019;141(6):2404-2412. doi:10.1021/jacs.8b11343

3.   High-throughput stability screening for detergent-solubilized membrane proteins. Kotov V, **Bartels K**, Veith K, Josts I, Subhramanyam U, Günther C, Labahn J, Marlovits TC, Moraes I, Tidow H, Löw C, Garcia-Alai MM. Sci Rep. 2019;9(1):10379. doi:10.1038/s41598-019-46686-8

II. List of abbreviations

°	Degree
°C	Degree celsius
μg	Microgram
μL	Microliter
μM	Micromolar
μs	Microseconds
μW	Microwatt
2D	Two-dimensional
3D	Three-dimensional
A	Acceptor
Å	Ångström
AB	Antibody
ABC transporter	ATP binding cassette transporter
ACS	Anion-cation symporter
AK-AMCA	β-Ala-Lys-N-7-amino-4-methylcoumarin-3-acetic acid
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BL	Brain total lipid extract
CCCP	Carbonyl cyanide m-chlorophenyl hydrazine
CMC	Critical micelle concentration
Cryo-EM	Cryogenic electron microscopy
cv	Column volume
D	Donor
dCTP	Deoxycytidine triphosphate
DD	D111C/D449C
ddH_2O	Double-distilled water
DDM	n-Dodecyl-β-D-maltopyranoside
Dgal	D-galactonate
dGTP	Deoxyguanosine triphosphate
DM	n-Decyl-β-D-maltopyranoside
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
Dpp	Dipeptide permeases
DTT	1,4-dithiothreitol
E	FRET efficiency
E. coli	*Escherichia coli*

fL	Femtoliter
FRET	Förster resonance energy transfer
g	Gram
GPCR	G-protein-coupled receptor
h	Hours
H^+	Proton
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICH	Intracellular helices domain
IMAC	Immobilized metal affinity chromatography
IMP	Integral membrane protein
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ISSD	Infantile free sialic acid storage disease
K^+	Potassium ion
K_D	Dissociation constant
kDa	Kilodalton
L	Liter
LB	Lysogeny broth medium
LCP	Lipidic cubic phase
LDAO	n-Dodecyl-N,N-dimethylamine-N-oxide
LIC	Ligation independent cloning
LMNG	Lauryl maltose neopentyl glycol
mAU	Milli absorption units
MES	2-(N-Morpholino)ethansulfonsäure
MFS	Major facilitator superfamily
mg	Milligram
min	Minutes
mL	Milliliter
mM	Millimolar
ms	Milliseconds
mV	Millivolt
MWCO	Molecular weight cut-off
N00	Nanobody 00
nanoDSF	Nano differential scanning fluorometry
NG	n-Nonyl β-D-glucopyranoside
ng	Nanogram
Ni-NTA	Nickel nitrilotriacetic acid
nL	Nanoliter
nM	Nanomolar

nm	Nanometer
NM	n-Nonyl-β-D-maltopyranoside
No.	Number
NO_3^{2-}	Nitrate ion
NPT	Type I phosphate transporters
ns	Nanoseconds
nsFCS	Nanosecond fluorescence correlation spectroscopy
OD600 nm	Optical density at 600 nm
Opp	Oligopeptide permeases
PA	Phosphatidic acid
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein data bank
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PEG-maleimide	Methoxypolyethylene glycol maleimide 5000
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PIE	Pulsed interleaved excitation
pM	Picomolar
pNIC	pNIC-CTHF vector
PO_4^{3-}	Phosphate ion
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
POT/PTR	Proton-dependent oligopeptide transporter/ peptide transport family
PS	Phosphatidylserine
psi	Pound-force per square inch
pTH	pTH24 vector
r	Radius
R_0	Förster radius
RASP	Recurrence analysis of single particles
RIS	Rotational isomeric state
RT	Room temperature
s	Seconds
S	Stoichiometry
SapNP	Saposin-derived lipid nanoparticle

SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLC15	Solute carrier 15 family
SLC17	Solute carrier 17 family
smFRET	Single-molecule FRET
SO_4^{2-}	Sulfate ion
TAE	Tris-acetate-EDTA
TB	Terrific broth medium
TBE	Tris-borate-EDTA
TBST	Tris-buffered saline with Tween20
TCEP	Tris(2-carboxyethyl)phosphin
TES	Tris-EDTA-sucrose
TM	Transmembrane helix
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
V	Volt
Val	Valinomycin
VGLUT	Vesicular glutamate transporter
VNUT	Vesicular nucleotide transporter
w/v	Weight per volume
w/w	Weight per weight
WQ	W203C/Q487C
wt	Wildtype
xg	Times gravity
$\Delta\mu H^+$	Electrochemical H^+ gradient
ΔTm	Melting temperature difference
λ	Wavelength

III.   List of tables

IV.   List of figures

1. Zusammenfassung

Proton-abhängige Oligopeptid Transporter (POTs) gehören zur Major-Facilitator-Superfamilie (MFS), einer der größten Familie von Transportern. Angetrieben durch einen Protonengradienten, transportieren POTs Nährstoffe und Peptidmimetika wie β-Lactam-Antibiotika, über die Zellmembran. Dies macht sie zu wichtigen Zielobjekten für die Arzneimittelaufnahme und Medikamentenentwicklung. DtpA, die Dipeptid und Tripeptid Permease aus *Escherichia coli*, ist ein prototypischer prokaryotischer POT, welcher eine bemerkenswerte funktionelle Ähnlichkeit zum humanen Protein PepT1 der Solute Carrier Familie 15 (SLC15) aufweist. Trotz der wichtigen Rolle die POTs spielen, bleibt der Transportzyklus sowie die Dynamik dieser Transporter größtenteils unerforscht. Mehrere Studien legen ein mechanistisches Model nahe, wonach der Transport, als Folge von Substratbindung, auf einem allosterischen Wechsel zwischen innen-offenem, außen-offenem und verschlossenem Zustand beruht. Allerdings gründet dieses Model bis heute hauptsächlich auf Röntgenkristallstrukturen, welche nur für den innen-offenen und verschlossenen Zustand, nicht aber für den außen-offenem Zustand verfügbar sind. In dieser Arbeit wurde Einzelmolekül Förster Resonanz Energie Transfer genutzt, um die Dynamik von DtpA in Detergenslösung sowie rekonstituiert in eine Lipid-Umgebung, zu untersuchen. In Detergenslösung liegt der Großteil der DtpA Moleküle im innen-offenen Zustand vor, während nur ein kleiner Teil im verschlossenen Zustand vorliegt. Hierbei ist der innen-offene Zustand kein statischer Zustand, wie in der Röntgenkristallstruktur suggeriert, sondern ein Ensemble verschiedener innen-offener Zustände, welche sich im Grad der Öffnung auf der zytoplasmatischen Seite unterscheiden. Das Verhältnis der verschiedenen Populationen zueinander konnte weder durch eine Änderung des Detergens, des pH-Wertes noch durch Substratzugabe beeinflusst werden. Im Gegensatz dazu, wenn der Transporter in ‚saposin-derived lipid nanoparticles' (SapNPs) eingebettet ist, ist das Verhältnis der beobachteten Populationen während der smFRET Messungen sehr empfindlich von der Chemie der Lipidkopfgruppe anhängig, welche zur Bildung der SapNPs genutzt wurden. In diesem Fall, wurde der außen-offene Zustand beobachtet, was darauf hindeutet, dass DtpA in einer Lipidumgebung in der Lage ist alle funktional relevanten Konformationszustände anzunehmen.

Wie POTs, so gehören auch Anion-Kation Symporter (ACS) zu den Nährstofftransportern innerhalb der MFS Familie. Die ACS Transporter aus *Escherichia coli* (DgoT, LgoT, GarP, GudP, ExuT und RhmT) transportieren Zuckersäuren, welche als Kohlenstoffquelle dienen. Darüber hinaus werden die Zuckersäuren auch mit der Wirtsanpassung in Zusammenhang gebracht. In dieser Arbeit konnten vier Transporter erfolgreich aufgereinigt und funktionell charakterisiert werden (DgoT, LgoT, GarP und GudP). Die Charakterisierung umfasste einen *In-vivo*-Wachstumstest von Knockout-Mutanten für die jeweiligen Transporter auf verschiedenen Kohlenstoffquellen, einen *In-vitro*-Ligandenbindungstest und einen Liposomen-basierten Aufnahmetest zur Unterscheidung zwischen Liganden und Substraten. Somit konnte D-Galactonat als Substrat von DgoT zugeordnet werden, L-Galactonat von LgoT, Galactarate von GarP und Glucarat und Galactarat als Substrate von GudP. Zusätzlich konnte gezeigt werden, dass D-Galactonat ein Ligand, aber kein Substrat für GarP ist.

Zusammenfassend lässt sich sagen, dass die Konformationszustände, die DtpA annehmen kann, stark von seiner Umgebung abhängen. Während in Detergenslösung nur innen-offene und verschlossene Zustände beobachtet werden, sind in einer Lipid-Umgebung alle funktionsrelevanten Zustände des Transportzyklus vorhanden. Darüber hinaus beeinflussen die Lipidkopfgruppen das Verhältnis der beobachteten Konformationszustände.

Zusätzlich konnten in der ACS-Studie vier der sechs ACS-Transporter aus *Escherichia coli*, DgoT, LgoT, GarP und GudP erfolgreich funktionell charakterisiert werden. Durch eine Kombination von *In-vivo* und *In-vitro* Studien konnten Liganden und Substrate für jeden Transporter bestimmen werden.

2. Abstract

Proton-dependent oligopeptide transporters (POTs) are members of the major facilitator superfamily (MFS), one of the largest families of transporters. Powered by a proton-gradient, POTs transport nutrients and peptide-mimetic drugs such as β-lactam antibiotics across the cell membrane thus they are important targets for drug delivery and design. DtpA, a dipeptide and tripeptide permease from *Escherichia coli*, is a prototypic prokaryotic POT that shows a remarkable functional similarity to human PepT1 of the solute carrier family 15 (SLC15). Despite the fundamental role of POTs, the full transport cycle and dynamics of these transporters remain largely unexplored. Several studies suggest mechanistic models, which rely on an allosteric change between inward-open, outward-open and occluded states, following substrate binding. But up to date, these models are supported mainly by crystal structures which are only available for the inward-open and occluded states but not the outward-open state. In this study, single-molecule Förster resonance energy transfer (smFRET) is used to investigate the dynamics of DtpA when solubilized in detergent and reconstituted into a lipidic environment. In detergent solution, the majority of DtpA molecules adopt the inward-open state and only a small minority the occluded state. This inward-open state is not one defined state as suggested by the crystal structure, but it is an ensemble of several inward-open states which differ in the degree of cytoplasmic opening. The ratio of populations could not be shifted by changing the detergent, substrate addition or pH change. In contrast, when reconstituted into saposin-derived lipid nanoparticles (SapNPs) the ratio of observed populations in smFRET measurements sensitively depends on the chemistry of the headgroup of the lipids used to form the SapNPs. Here, the outward-open conformation was observed, suggesting that in a lipid environment DtpA can adopt all functionally relevant states.

Like POTs, anion-cation symporter (ACS) are another nutrient transporter family among the MFS family. The ACS transporters from *Escherichia coli* (DgoT, LgoT, GarP, GudP, ExuT and RhmT) transport sugar acids which are utilized as an energy source by the bacteria and are implicated in host adaption. Here, four transporters could successfully be purified and functionally characterized (DgoT, LgoT, GarP and GudP). The characterization involved an *in vivo* growth assay of knockout mutants for the respective transporters on various carbon sources, an *in vitro* ligand binding assay and a liposome-based uptake assay to differentiate between ligands and substrates. Thus, D-galactonate could be assigned as the substrate of DgoT, L-galactonate of LgoT, galactarate of GarP and glucarate and galactarate are both substrates of GudP. In addition, D-galactonate was found to be a ligand, but not a substrate, for GarP.

In summary, the conformational states DtpA is able to adopt strongly depend on its environment. Whereas in detergent solution, only inward-open and occluded states are observed, in a lipidic environment all functionally relevant states of the transport cycle are present. Furthermore, the lipid headgroups tune the ratio of the observed conformational states.

Additionally, in the ACS study, four of the six ACS transporters from *Escherichia coli*, DgoT, LgoT, GarP and GudP, could successfully be functionally characterized. A combination of *in vivo* and *in vitro* studies was used to determine ligands and substrates for each transporter.

3. Introduction

3.1. Biological membranes

Biological membranes play an important role for living organisms because they separate the cell from its environment and can divide the cell into specialized compartments. They consist of lipids and proteins, both of which carbohydrates may be attached to. In a biological membrane, the lipid molecules are ordered in a bilayer. The hydrophilic head groups of the lipids are sitting on the outside and the hydrophobic tails are inside the membrane. The hydrophilic headgroups are thus facing the aqueous environment, shielding the hydrophobic membrane core. Proteins attached to or inserted in the membrane confer function such as transport, signal transduction, enzymatic reactions, or act as an anchor to the extracellular matrix and cytoskeleton.

Within the membrane, lipids and proteins are not covalently bound. Thus, in the lateral plane of one leaflet, lipids and proteins can diffuse easily. But the event of lipids crossing from one leaflet to the other, transversal diffusion or 'Flip-Flop', is rare as the hydrophobic headgroup would need to traverse the hydrophobic core region of the bilayer. However, there are certain proteins able to catalyze the movement of lipids between the two leaflets [1]. Hence, the bilayer itself is asymmetric, meaning that both leaflets do not have the same composition [1].

The most prominent example of a biological membrane is the cellular membrane which encloses a cell and thus creates an outside, the extracellular environment, and an inside, the cytoplasm. The membrane barrier prevents leakage of components from within the cell to the outside and, *vice versa*, the intrusion of external molecules [1]. Thus, the cell membrane is essential to establish and maintain an electrochemical gradient [1]. These electrochemical gradients are vital, for example in the generation of action potentials in neurons or in the ATP-synthesis in mitochondria [2]. At the same time however, the membrane cannot completely seal off the cell. Rather a cell needs to perceive and respond to external stimuli as well as take up compounds from or release them into the environment. These functions are performed by proteins embedded in the lipid bilayer.

Some species have more than one cellular membrane, for example gram-negative bacteria having an outer and inner membrane enclosing the cell [1]. Besides the cell membrane, other biological membranes are found forming organelles with specific function in eukaryotic cells such as mitochondria, chloroplasts or lysosomes [1].

3.1.1. Three main lipid types in biological membranes
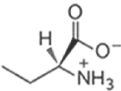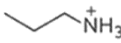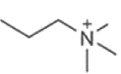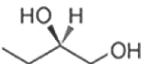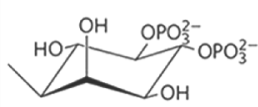
Lipids are the main component of biological membranes. Depending on their structure they are sorted into three main types: (i) phospholipids, (ii) glycolipids and (iii) steroids [1].

Most of the lipids in a membrane are phospholipids [2]. They consist of two hydrophobic fatty acid tails and a hydrophilic headgroup, which involves a phosphate group. The most common fatty acid tails have an acyl chain length between 10 and 24 carbon atoms [1]. If the acyl chains contain only C-C single bonds they are called saturated, if they contain one or more C=C double bonds, they are called unsaturated [1]. The saturation level and chain length influence the melting temperature of a fatty acid. The melting temperature is lower for fatty acids with shorter chain length or unsaturated fatty acids [2]. The lower melting temperature in turn increases the fluidity of the membrane [2].

Phospholipids can be further subdivided into glycerophospholipids and phosphosphingolipids. These two groups differ in the way the headgroup is connected to the fatty acid tails. For glycerophospholipids, glycerol is bound to two fatty acids and attached to the headgroup via a phosphate group. In contrast, phosphosphingolipids do not build on glycerol but instead utilize sphingosine. Sphingosine is an 18-carbon long amino alcohol with a hydrophobic carbon tail that includes a C=C bond and is thus unsaturated. The sphingosine is bound to a fatty acid as well as the headgroup group via the phosphate group (Figure 1).

Common headgroups of glycerophospholipids are serine, ethanolamine, choline, glycerol and inositol. The corresponding phospholipid are called phosphatidylserine (PS), phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylglycerol (PG) and phosphatidylinositol (PI). If no additional group is attached to the phosphate group the phospholipid is called phosphatidic acid (PA) (Table 1). For phosphosphingolipids common headgroups include ethanolamine and choline, those lipids are called sphingomyelin (Figure 1).

Table 1: Glycerophospholipids. Common glycerophospholipids of biological membranes.

Glycerophospholipids	Abbreviation	Headgroup	Formula of headgroup	Net charge at pH 7
Phosphatidic acid	PA	-		-2
Phosphatidylserine	PS	Serine		-1
Phosphatidylethanolamine	PE	Ethanolamine		0
Phosphatidylcholine	PC	Choline		0
Phosphatidylglycerol	PG	Glycerol		-1
Phosphatidylinositol	PI	Inositol		-4

The second major group of membrane lipids is the glycolipids. Their headgroups can be mono- or oligosaccharides. The headgroup can be attached to a phosphate-glycerol backbone with two fatty acids, as in glycerophospholipids, or to a sphingosine backbone with one additional fatty acid as in phosphosphingolipids (Figure 1). Glycolipids are found exclusively in the extracellular leaflet of membranes [1].

Steroids have a very different structure compared to phospholipids or glycolipids. The backbone is formed by a tetracyclic ring consisting of three cyclohexane rings and one cyclopentane ring. At one end of this backbone a hydroxyl group is bound and at the other end a variable side group. Steroids are found in the membranes of animals (cholesterol), plants (stigmasterol) and fungi (ergosterol) [1]. When present in the membrane, steroids lower the fluidity [1]. They are inserted

into the lipid bilayer parallel to the other lipids [1]. The hydroxyl group interacts with the hydrophilic head groups of the other lipids and the aqueous environment outside the membrane, whereas the hydrophobic side group and the tetracyclic ring interact with the hydrophobic core.

In bacterial cell membranes hopanoids are found instead of sterols [3]. The structure of hopanoids are similar to those of steroids but they possess one more cyclohexane ring. Analogous to sterols in eukaryotes, hopanoids modulate the membrane fluidity in bacteria [3].
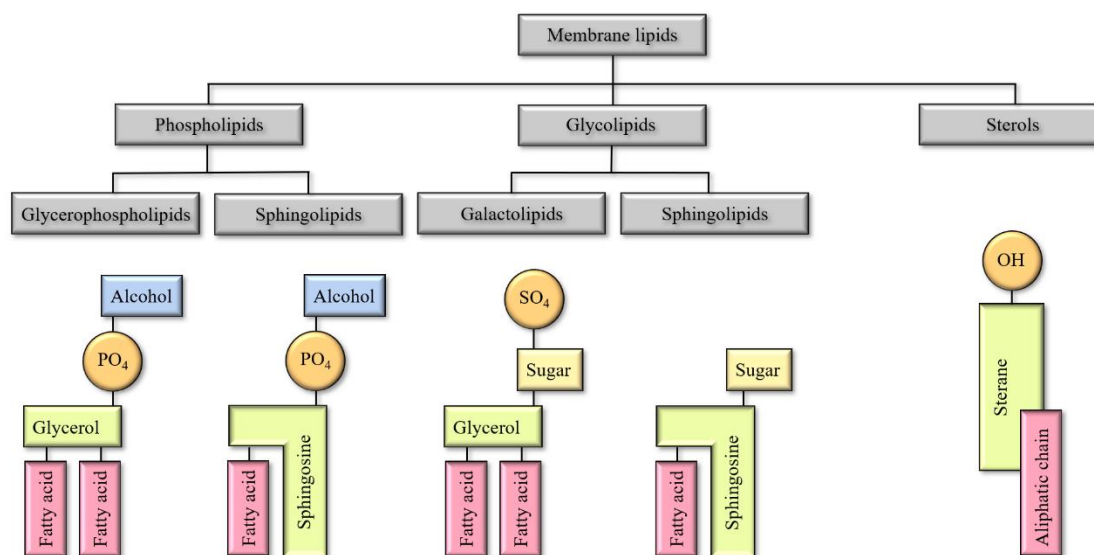
Figure 1: Membrane lipid classification and structure. Classification of membrane lipids based on their building blocks.

3.1.2. Membrane proteins

Besides lipids, the other important component of biological membranes are proteins. Those proteins can be divided into three groups, peripheral membrane proteins, integral membrane proteins (IMPs), and lipid-anchored membrane proteins (Figure 2a). The difference between those groups is that while IMPs have at least one segment of the protein embedded into the lipid bilayer, peripheral proteins do not interact with the hydrophobic core of the membrane. Instead they are loosely associated to the membrane. For instance, they can interact with the membrane lipids or IMPs either by electrostatic interactions or hydrogen bonds [4]. Lipid-anchored membrane proteins on the other hand are covalently bound to a lipid anchor in the membrane. Those lipid anchors can be different lipids, for example glycosylphosphatidylinositol in GPI-anchors [1].

As the proteins studied in this thesis are IMPs, only this group will be introduced in more detail.

IMPs are amphipathic molecules, meaning they have both hydrophilic and hydrophobic regions. Through the hydrophobic region they can interact with the hydrophobic core of the bilayer, thus being embedded in the membrane. The hydrophobic region can be formed by either α-helices or multiple β-sheets [1]. In both cases the side chains of hydrophobic amino acids are pointing towards the hydrophobic membrane core with the hydrophilic backbone shielded.

There are several different topologies possible for IMPs. Not all IMPs span the whole membrane, some insert only into one leaflet and are called monotopic IMPs [1]. Other IMPs span the entire membrane and are called transmembrane proteins. Bitopic IMPs which possess a single transmembrane helix (TM), span the membrane once. Therefore, the N- and C-terminus of the protein are located on opposite sides of the membrane. Depending on whether the N-terminus is on the outside or inside the cytoplasm, they are classified as Type I or Type II respectively.

IMPs that span the membrane several times are called polytopic IMPs or Type III. Here, the TMs belong to one polypeptide chain and are connected by loops. Multimeric IMPs or Type IV, on the other hand, are oligomeric proteins consisting of several polypeptide chains which all possess at least one TM.

Although, IMPs made from α-helices are more common, IMPs made from β-barrels are found for example in the outer membrane of gram-negative bacteria, chloroplasts, or mitochondria, where they form porins [1]. The β-barrel is formed by several β-sheets arranged in a circle. While the side chains of the amino acids pointing towards the membrane are hydrophobic, hydrophilic side chains point inside of the β-barrel, creating a hydrophilic path through the membrane [1] (Figure 2b).

Membrane proteins serve different purposes in the membrane. They can be enzymes catalyzing chemical reactions or receptors for signal transduction, they can serve as markers in cell-cell recognition or as attachment point for the cytoskeleton and extracellular matrix, join cells together in junctions or enable the transport of molecules through the membrane.

The transport function is further investigated in this thesis.
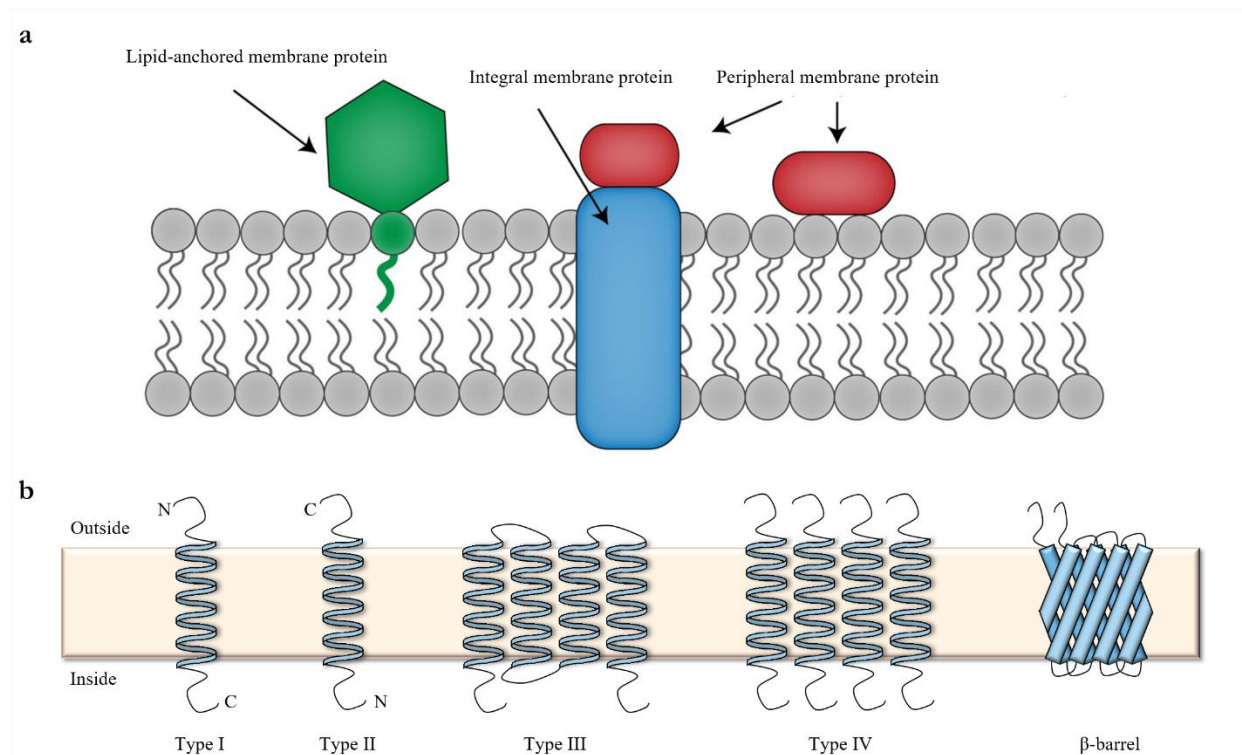
Figure 2: Classification of membrane proteins and topology of integral membrane proteins. (a) Schematic of different groups of membrane proteins. Lipid-anchored membrane proteins are covalently attached to a membrane lipid (green), integral membrane

proteins (IMPs) interact with the hydrophobic core of the membrane (blue), whereas peripheral membrane proteins are attached to the membrane by interaction with either other membrane proteins or the lipid headgroups (red). Figure from [5]. (b) Topology of IMPs. Helical IMPs can span the lipid bilayer one time with either the C-terminus in the cytoplasm (Type I) or the N-terminus on the cytoplasm (Type II). Multipass α-helical IMPs are described as Type III, when the TMs are part of a single polypeptide chain or Type IV in the case of TMs belonging to different polypeptide chains. In addition to α-helical IMPs, β-barrels are IMPs formed from β-sheets.

3.2. Transmembrane transport

For survival, a cell depends on the uptake of nutrients from its environment as well as the release of metabolic byproducts to the environment. To this end, there are several forms of transmembrane transport systems a cell can utilize, depending on the chemical nature and concentration of the substrates. This transmembrane transport can occur in different ways, through passive or active transport [1] (Figure 3).

In passive transport, the molecules move along their concentration gradient which does not require additional energy. There are several ways passive transport can occur. Firstly, small uncharged molecules, for example O_2 or CO_2, can cross through the membrane by simple diffusion. Secondly, charged or large polar molecules can be transported with the aid of carriers in a process called facilitated diffusion [1]. This is utilized by ionophores, channels and transporters. For example, ionophores can aid ions to cross the membrane. There are two classes of ionophores. The carrier ionophores are small molecules that bind ions and thus neutralize their charge. The resulting complex can pass through the membrane and release the ion on the other side [1]. Channel-forming ionophores, on the other side, are peptides that insert into the membrane, creating a hydrophilic pore through which ions can diffuse [6]. Both types of ionophores abolish an established ion-concentration gradient across a membrane and are thus often toxic to cells. Their application ranges from being used as antibiotics, e.g. gramicidin, to their usage in transport assays, e.g. valinomycin [6,7].

In both cases, ion channels and transporters, an IMP provides a passage-way through the membrane. Whereas channels can be open to both sides of the membrane at once allowing rapid movement of molecules through the membrane, transporters only open to one side of the membrane at a time [1].

So far, all introduced transport mechanisms move molecules along their concentration gradient. In a cell, essential nutrients might be scarce and cells might need to enrich certain nutrients. To this end, the cell employs transport mechanisms which are able to move molecules against their concentration gradient. This process is called active transport. In contrast to passive transport, active transport requires energy input. This energy can be provided by ATP hydrolysis in the case of primary active transport, or by the co-transport of a second molecule along its concentration gradient in so-called secondary active transport. Secondary-active transporters harness the energy to move their substrates against their concentration gradient from the ion gradients generated by primary-active transporters [1].
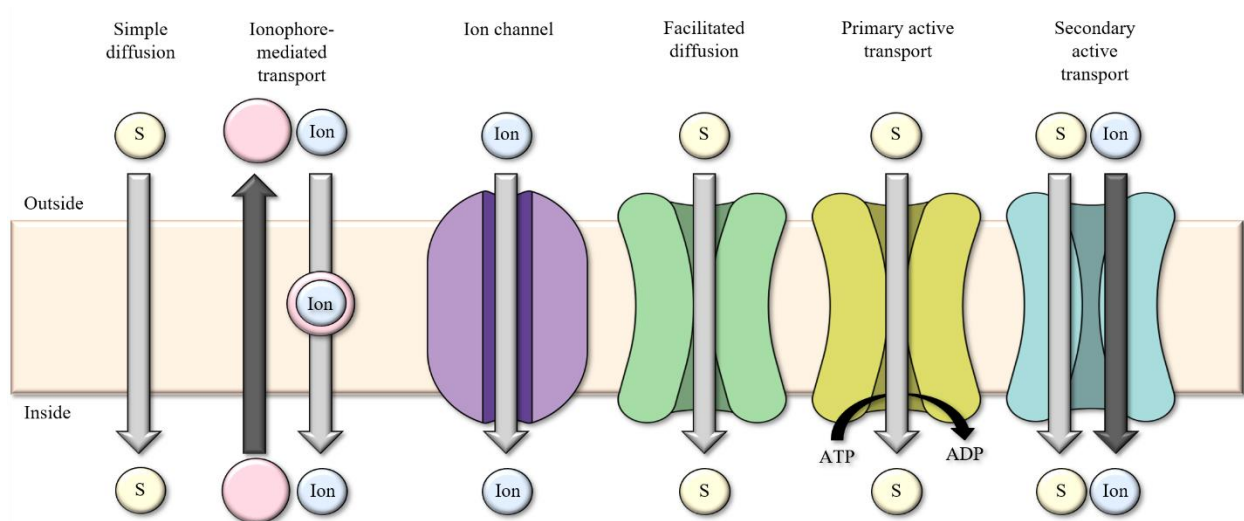
Figure 3: Transport mechanisms across biological membranes. Molecules can cross the lipid bilayer in a number of ways depending on their chemical properties and concentration gradient. Small, nonpolar molecules can diffuse through the membrane along their concentration gradient, this process is called simple diffusion. All other molecules need some variation of a transport system. Ions can be transported through the membrane with the help of ionophores. These are molecules binding ions and moving them through the membrane. In addition, ions can also move through channel proteins to cross the lipid bilayer. Polar molecules or large ions can cross the membrane along their concentration gradient with the help of a carrier protein. This process is called facilitated diffusion. All transport mechanisms described so far are passive, meaning they do not require additional energy input but the molecules are only able to move along their concentration gradient. To move molecules against their concentration gradient, an energy source is needed. This mode of transport is called active transport. Depending on the source of energy that powers the transport, one distinguished primary active transport and secondary active transport. In primary active transport, ATP-hydrolysis is used as energy source and in secondary active transport a second molecule, often an ion, is moved along its concentration gradient supplying the energy to move the main substrate against its concentration gradient.

3.2.1.   Major facilitator superfamily (MFS)

One of the largest transporter families identified to date is the major facilitator superfamily (MFS), which is found in all kingdoms of life [8]. MFS transporters were originally thought to transport carbohydrates but subsequently several subfamilies were characterized which transport a wide range of substrates such as peptides, amino acids, ions, lipids, vitamins and nucleosides [9]. In addition, several MFS proteins can transport drugs or prodrugs, for example, EmrD extrudes various antibiotics from *Escherichia coli* (*E. coli*) conferring multidrug resistance to the bacteria [10,11]. In the case of human PepT1, the protein is able to take up several different medically active compounds like antiviral drugs or beta-lactam antibiotics [12]. Furthermore, mutations or misregulation in several members of the MFS are implicated in diseases such as cancer, schizophrenia, gout, epileptic seizures, diabetes type 2 and Alzheimer's disease [13–15]. Both, the potential of MFS proteins as drug targets as well as for drug delivery constitute their great pharmacological relevance.

Between different MFS subfamilies, the core TMs are more conserved than extracellular and cytoplasmic regions of the proteins, though the overall sequence conservation is quite low, for example, the two sugar-proton symporters from *E. coli*, LacY and FucP are both MFS transporters but their sequence identity is only around 10% [16]. Despite the relatively low

sequence conservation across the MFS they exhibit similar characteristics in (i) the transport mode, (ii) the structural fold and (iii) the transport cycle [9,17].

The transport modes utilized by MFS transporters can be sorted into three groups: uniporter, symporter and antiporter [18] (Figure 4a). Uniporters are classified as facilitators. Here, only one substrate is transported and the transport follows the concentration gradient of this substrate.

In contrast, both symporters and antiporters are secondary active transporters. Here, the transporter either moves two substrates in the same direction, symporters, or in opposite directions, antiporters [9]. In both cases the main substrate of the transporter is moved against its concentration gradient. The energy for this transport is provided by the transport of a second substrate along its concentration gradient [9]. The gradients that are utilized by a secondary active transporter are predominantly sodium or proton gradients, which are usually established by ATP-driven primary active transporters [1]. Prokaryotes use mostly a proton gradient, whereas in eukaryotes a sodium gradient is more frequent [19]. Proton-coupled transporters are found in eukaryotes as well, but mainly in an acidic environment, such as lysosomes or the thylakoid lumen [20,21]. Such acidic environments are less prevalent in eukaryotes compared to prokaryotes which may explain their preference for sodium-coupling [22].

The first published crystal structures of MFS transporters were that of lactose permease LacY and the Glycerol-3-phosphate transporter GlpT both from *E. coli* [23,24]. The first eukaryotic structure was of the phosphate transporter PiPT from the fungus *Piriformospora indica* and the first structure of a human MFS transporter of the facilitated glucose transporter GLUT1 [25,26]. These structures, as well as all known structures of MFS members to date, show a common fold, the so-called MFS fold (Figure 4b).

The MFS fold is comprised of 12 TMs that are arranged into two helical bundles [9]. The first six helices from the N-terminus form the N-terminal domain and the first six helices from the C-terminus form the C-terminal domain (Figure 4b). Both domains are related by a 2-fold pseudosymmetry with the symmetry axis perpendicular to the membrane [9]. Within each domain, the six helices can be divided into subgroups of two three-helical bundles which are called inverted repeats. The inverted repeats making up one domain are related by a 2-fold symmetry axis parallel to the membrane [9] (Figure 4c). The corresponding helices between the four inverted repeats of an MFS transporter have similar structural and functional roles [9]. The first TM of each repeat, TM1, TM4, TM7 and TM10, line the substrate path, while the second TM of each repeats, TM2, TM5, TM8 and TM11 mediate the interface between both domains and the third TM of each repeat, TM3, TM6, TM9 and TM12, give structural integrity.

Both domains can be connected by a loop or additional helices, as is the case for the bacterial members of the proton-dependent oligopeptide transporter (POT) family [27]. There, in addition to the 12 canonical MFS helices, they possess two additional helices. Additional helices are also found in other MFS transporters, though to date no evidence was found for an MFS transporter with fewer than 12 TMs [27,28].

The substrate binding site is located in the middle of the protein at the interface of the N- and C-terminal domain and residues from both domains participate in the substrate coordination [27,29].

How substrates are generally transported by members of the MFS family will be described in the next paragraph.
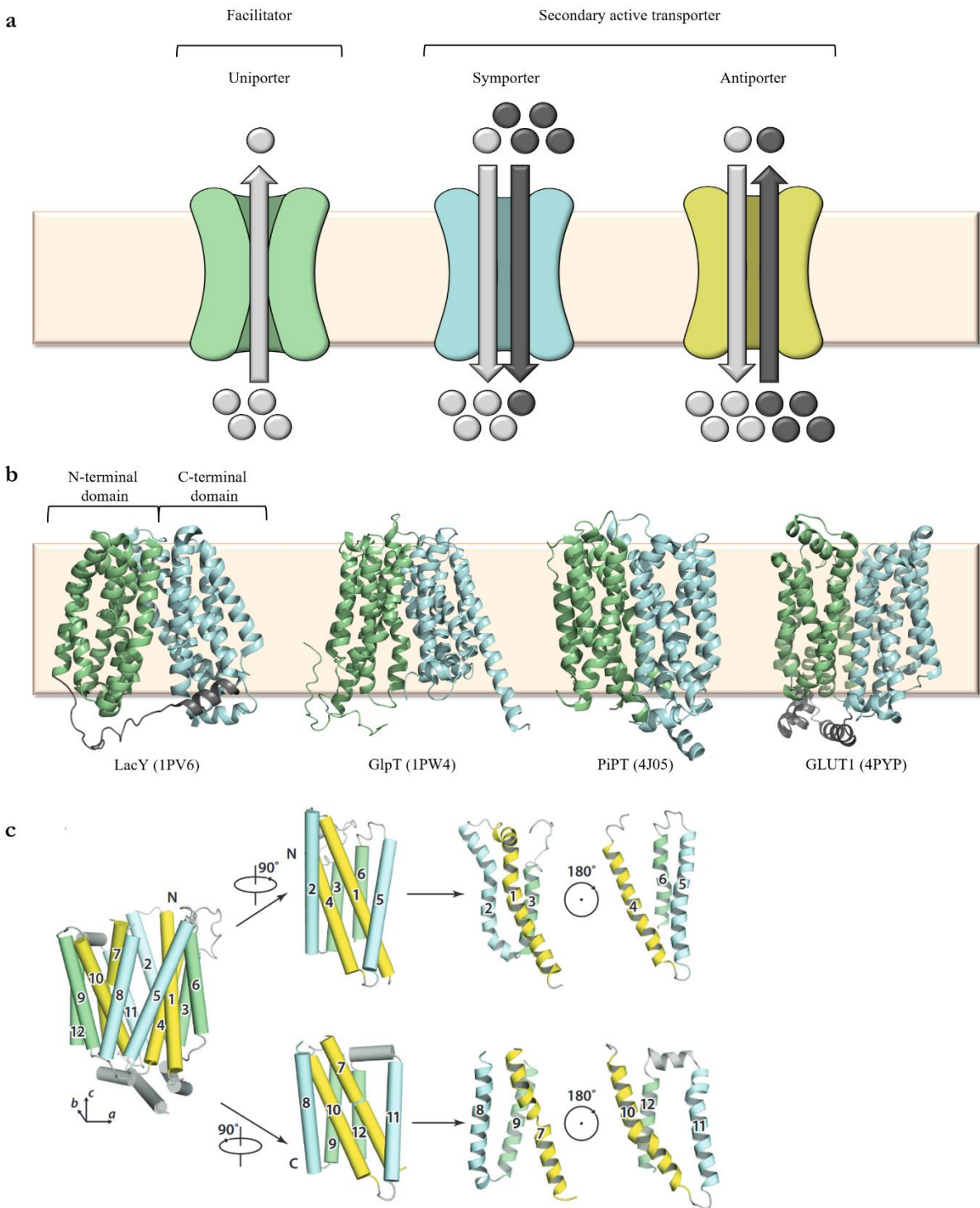
Figure 4: Transport modes and structural features found among MFS transporters. (a) Among MFS transporters three different modes of transport are observed. They can be either uniporters, symporters or antiporters. Uniporter use facilitated diffusion to transport their substrate (light gray sphere) down its concentration gradient. Symporters and antiporters are secondary active transporters that can move their substrate against their concentration gradient. The energy to move the substrate against its concentration gradient is gained from the transport of a second molecule, the co-substrate (dark gray sphere), along its concentration gradient. Usually this co-

substrate is an ion. If the co-substrate is transported in the same direction as the substrate or in the opposite direction, these transporters are called symporters or antiporters, respectively. Figure adapted from [9]. (b) Crystal structures of MFS transporters. LacY and GlpT from *E. coli* were the first MFS transporters for which a high-resolution structure was available. The first eukaryotic MFS transporter structure was PiPT from the fungus *Piriformospora indica* and the first from human was GLUT1. All transporters exhibit the typical MFS fold with 12 TMs arranged in 2 bundles of six TMs. The first six TMs from the N-terminus form the N-terminal domain (green) and the six TMs from the C-terminus form the C-terminal domain (blue). PDB-IDs: 1PV6 (LacY), 1PW4 (GlpT), 4J05 (PiPT) and 4PYP (GLUT1). (c) The N- and C-terminal domains themselves are pseudosymmetric towards one another with the symmetry axis perpendicular to the membrane. Each six-helix-bundle is made from two three-helix-bundles from consecutive TMs. These are called structural repeats. The two structural repeats from one domain are related by a 2-fold symmetry with the symmetry axis being parallel to the membrane. The function of the first TMs of the structural repeats (TM1, TM4, TM7, TM10) (yellow), the second TMs (TM2, TM5, TM8, TM11) (blue) and the third TMs (TM3, TM6, TM9, TM12) (green), is supposed to be equivalent among MFS transporters. Figure from [9].

3.2.1.1.   The transport cycle of MFS transporters

To explain substrate movement by transmembrane transporters, the alternating access model was first proposed by Jardetzky in 1966 [30]. In the alternating access model, the transporter opens its substrate binding site to either side of the membrane but never to both sides at the same time. This distinguishes transporters from channels and requires the transporter to adopt at least two different conformational states: an inward-open state with the substrate binding site exposed to the cytoplasm and an outward-open state with the substrate binding site exposed to the extracellular space [30]. It was suggested that the transition between those states would be accomplished by a rocker-switch motion of the N- and C-terminal domain [31]. Both domains would rotate as rigid bundles with respect to one another, exposing the binding site alternatingly to either side of the membrane [31]. The rotation axis for this rocker-switch model would be parallel to the membrane running through the substrate binding site [31].

In accordance with the rocker-switch model, published crystal structures of MFS transporters show those two proposed states, the inward-open and the outward-open state [23,24,32,33]. In addition, occluded or semi-occluded states have also been determined which are in accordance with the alternative access model but not predicted by the rocker-switch model [30] (Figure 5). Based on this data, the transport model was modified [34].

In this modified clamp-and-switch model, the occluded conformation is achieved by the bending of the tips of the first TMs in an inverted repeat, TM1, TM4, TM7 or TM10 [34]. The bending of the cytoplasmic tips of TM4 and TM10 closes the cytoplasmic side, whereas bending of the periplasmic tips of TM 1 and TM7 closes the periplasmic side. Sometimes the second TM in a structural repeat is also bend, for example in the inward occluded state of PiPT [25]. The bending of the TM tips to form an occluded state is termed the clamping step, the subsequent rotation of the N- and C-terminal domain to move from inward to outward-open or *vice versa* is the switching step.

The residues which are involved in the interaction between N- and C-terminal domain in some but not all conformational states are called gating residues [34]. Depending on the conformational state of the transporter those interactions are either formed when the residues are in close proximity, or broken, when they are moved apart by conformational change. For some MFS transporters, such as the POT family, it has been proposed that these interactions are formed by salt bridges

[35,36] but in other members of the MFS such salt bridges are absent at least on one side of the protein, as is the case for LacY from *E. coli* [37]. Therefore, these gating residues do not necessarily seem to interact via salt bridges.

In summary, for a complete transport cycle, the transporter has to adopt at least three different conformational states which include the outward-open state, inward-open state and occluded state. The transporters can adopt any of these conformational states in the apo or substrate bound form. The conformational change from one state to the next involves the formation and breaking of interactions between so-called gating residues.

The mechanisms of how the transporter transitions from one conformational state to the next was investigated for proton-coupled symporters, especially the intensely studied prototypical MFS transporter LacY from *E. coli* [38]. LacY imports galactosides against their concentration gradient into the cell, coupled to the transport of protons via the electrochemical proton gradient across the membrane [38]. First in the inward-open state, E325 of LacY is protonated. Only the protonated LacY can subsequently bind galactoside [38]. Thus, for symport, proton and galactoside have to be bound to LacY. The binding of galactoside is sufficient to induce the conformational change from the inward-open to the occluded and eventually the outward-open state via induced fit [38].

The induced fit model was described analogous to that of enzyme catalysis [38]. In contrast, the substrate is not chemically altered by the catalyzed reaction, but upon substrate binding to the transporters, the transporter itself changes conformation. In this system, the affinity of the transporter for its substrate is low in the inward-open state and outward-open state compared to the occluded state [38].

In the outward-open state, first the sugar dissociates from LacY, then the protein is deprotonated. The apo protein undergoes conformational changes from the inward-open via occluded states to the outward-open state so another transport cycle can start.

It is noteworthy that the alternate access of LacY is driven by galactoside binding and dissociation and not by the electrochemical H^+ gradient ($\Delta\mu H^+$). $\Delta\mu H^+$ functions as the driving force for the deprotonation of the transporter. This deprotonation step is the rate-limiting step in the transport cycle in the absence of $\Delta\mu H^+$ [38]. $\Delta\mu H^+$ accelerates this step, thus making the transport cycle more efficient [17].

The comparison to other MFS transporters such as XylE, FucP, MelB from *E. coli* and GlcP$_{Se}$ from *Staphylococcus epidermidis* suggests that the mechanism found for LacY might be applicable to other MFS families as well [39].
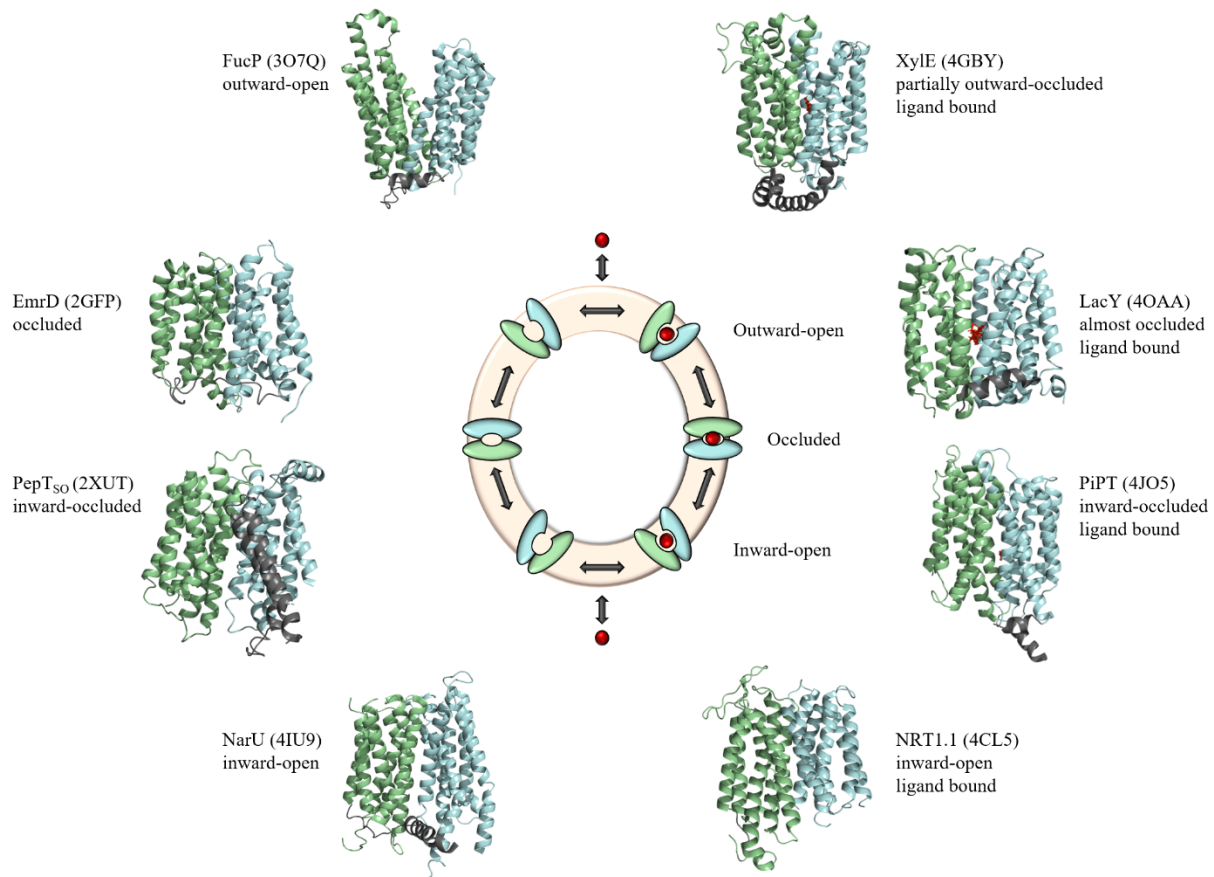
Figure 5: Transport cycle of MFS transporters. In the outward-open state the proteins substrate binding site is exposed to the extracellular site. Upon substrate binding the protein changes conformation and adopts an occluded state where the substrate binding site is closed to both sides of the membrane. In the subsequent inward-open state the substrate can be release and the apo form of the transporter can switch back through the occluded conformation to the outward-open state. High-resolution structures for ligand bound and apo transporters are available for outward-open, inward-open and occluded conformations. However, these structures are from different MFS subfamilies. All structures are colored to highlight the MFS fold with the N-terminal domain in green, the C-terminal domain in blue, additional domains in gray and the ligand in red. PDB-IDs: FucP 3O7Q, XylE 4GBY, LacY 4OAA, PiPT 4JO5, NRT1.1 4CL5, NarU 4IU9, PepTSO 2XUT, EmrD 2GFP. Figure adapted form [9].

3.2.1.2. Nutrient transport by MFS transporters

MFS proteins transport a wide range of substrates. Among these are vital nutrients such as carbohydrates and peptides. In this dissertation, two different MFS subfamilies are discussed, the POT family and the Anion-Cation symporter (ACS) family. Both families are described in detail hereafter.

3.2.1.2.1. Proton-dependent oligopeptide transporter (POT)

One of the many MFS subfamilies involved in nutrient uptake is the POT family, also referred to as the peptide transport (PTR) family [8,40]. As secondary active transporters, POTs utilize an electrochemical gradient to transport its substrates across the membrane. This gradient was shown to be a proton and not a sodium gradient [41]. The ionophore carbonyl cyanide m-chlorophenyl hydrazine (CCCP) allows protons to cross the membrane. Addition of CCCP abolished transport in POTs but the substitution of choline for sodium had no effect on transport [41].

Members of the POT family are found in all kingdoms of life except archaea [19]. Their role is the assimilation of nitrogen for metabolism and growth through the uptake of di- and tripeptides into the cells [19,40]. In plants, these transporters are sorted in the NTR1/PTR family [42]. Here, some transporters evolved to additionally transport nitrate, phytohormones or defense compounds [43–46].

Although POTs are described as highly promiscuous transporters, in principle accepting all possible 400 combination of dipeptides and 8000 combinations of tripeptides as substrates, they show a certain substrate preference [47]. This preference is based on the physical and chemical properties of the peptides. For example, PepT$_{SO}$ from *Shewanella oneidensis* prefers uncharged to charged peptides and among the charged peptides a negative charge at the C-terminus or a positive charge at the N-terminus of the peptide [48]. In contrast, NmPOT from *Neisseria meningitidis* shows no preference for the position of a negatively charged residue but does not transport peptides with a positive charge at the C-terminus [49]. DtpA from *E. coli* seems to prefer tripeptides over dipeptides whereas the opposite is the case for DtpC from *E. coli*, showing that peptide length is also important for the substrate preference of POTs [50,51].

The observed promiscuity of POTs could be explained by the observation that not only the peptide side chains are coordinated by the transporter but also the peptide main chain. Notably, the main chain coordinating residues are conserved among POTs [41,52].

Single amino acids or tetrapeptides are generally not substrates of POTs, with tetrapeptides deemed to be too large to fit into the transporters binding pockets and single amino acids too small to be sufficiently coordinated in the binding site [48]. An exception is the tetrapeptide Ala-Ala-Ala-Ala, which is small compared to other tetrapeptides, and has been suggested to be transported by the human POT PepT1 and DtpA from *E. coli* [48]. Single amino acids could possibly be transported by the human POTs PHT1 and PHT2 as well as DtpC from *E. coli* [48].

POTs appear to be somewhat stereospecific when it comes to binding substrates. While peptides from L-amino acids are transported, peptides built solely from D-amino acids are not substrates [35]. Although, peptides containing both L- and D-amino acids might be transported by some POTs [53].

In addition to peptides, compounds with a stereochemistry similar to small peptides are transported by POTs as well [12]. This makes them ideal drug targets, as is the case for the POTs from the human pathogens *E. coli* and *Neisseria meningitidis* [49,54]. Here, the transporters could be potentially exploited to import antibiotic compounds into the bacteria [49,54]. In the case of the human POTs PepT1 and PepT2, these transporters are potential drug delivery systems [12]. PepT1 is mainly expressed in the small intestine and PepT2 in the kidney, where these transporters are responsible for intestinal uptake and renal absorption of di- and tripeptides, respectively [55]. In addition, PepT1 was shown to transport various drugs and prodrugs such as beta-lactam antibiotics, antiviral prodrugs, and angiotensin-converting enzyme inhibitors [12]. Since poor bioavailability of medically active compounds is often a hinderance in the development of new drugs, altering medically active compounds to serve as substrates for PepT1 or PepT2, improves their uptake [12,56]. Alterations are effective when

they mimic amino acids or dipeptides, as was the case for the antiviral drug ganciclovir, whose bioavailability was enhanced by the addition of the amino acid valine resulting in valganciclovir, a substrate for PepT1 [12,57].

Besides PepT1 and PepT2, two other POTs are found in humans, PHT1 and PHT2, but compared to PepT1 and PepT2, they are not well studied [55]. These four transporters, as well as the other mammalian POTs, also belong to the solute carrier family 15 (SLC15) [55].

Although the protein sequences of POTs differ across species, they share several characteristics such as (i) the fold, (ii) conserved motifs and (iii) the transport mechanism.
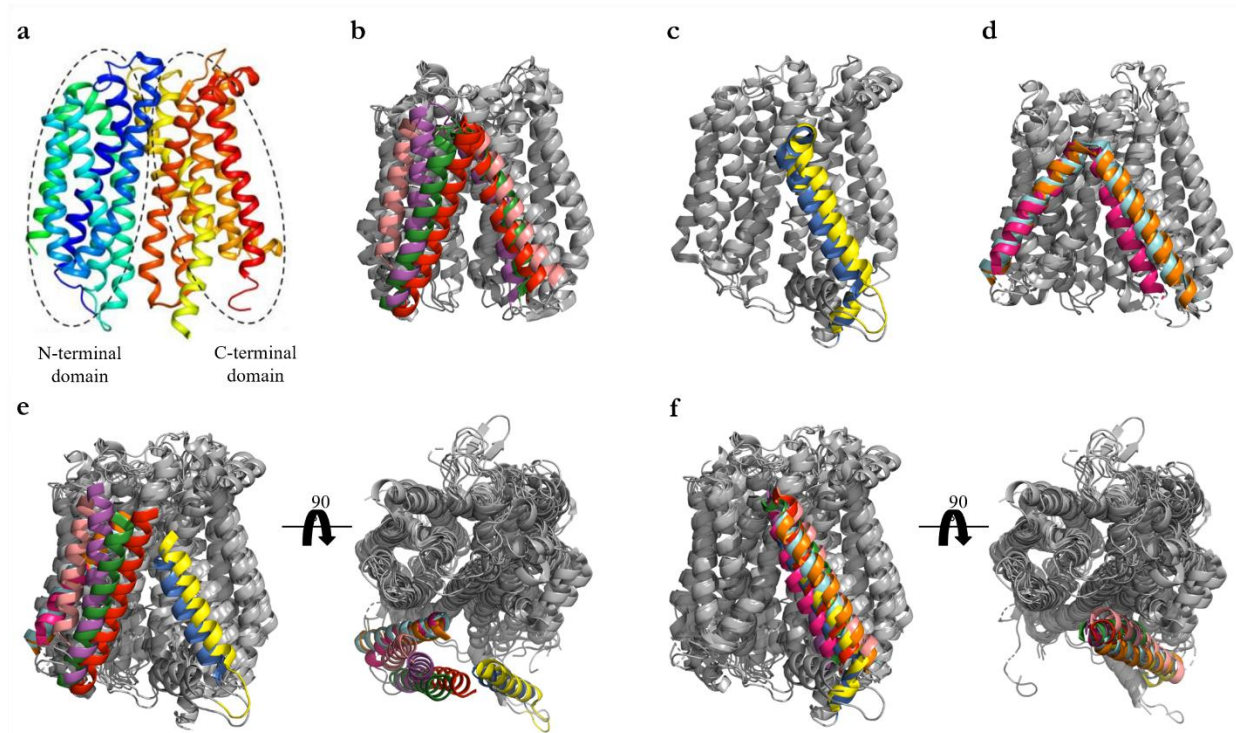
Figure 6: Structure of bacterial POTs. (a) Exemplary structure of a bacterial POT, here $PepT_{So}$ (PDB-ID 2XUT) from *Shewanella oneidensis*. It shows that POTs adopt the typical MFS fold. 12 TMs are arranged in two times six helix bundles. The six TMs from the N-terminus are called the N-terminal domain (blue and green) and the six TMs from the C-terminus are called the C-terminal domain (yellow and red). Here, the structure is in the inward-open state with access to the binding site possible from the cytoplasmic side of the protein but the periplasmic side is closed [58]. (b-d) Conformations of the HaHb domain in crystal structures of bacterial POTs. The 12 core TMs are colored gray and the HaHb domain of (b) $PepT_{St}$ red, GkPOT green, YePEPT salmon, $PepT_{sh}$ purple, (c) $PepT_{So}$ yellow, $PepT_{Xc}$ blue, (d) $PetT_{So2}$ cyan, DtpA orange and DtpD pink. (e-f) Comparison of conformations of the (e) Ha and (f) Hb helices from bacterial POTs. Same color coding as in b-d. PDB-IDs: $PepT_{St}$ 5OXO, GkPOT 4IKV, YePEPT 4W6V, $PepT_{sh}$ 6H7U, $PepT_{So}$ 4UVM, $PepT_{Xc}$ 6EI3, $PetT_{So2}$ 4LEP, DtpA 6GS1 and DtpD 4Q65.

The first crystal structure of a POT family member, $PepT_{SO}$ from *Shewanella oneidensis*, was published in 2011 [59]. Since then, numerous other structures have been published (Table 2). All structures so far show the typical MFS fold. In short, the transporters possess 12 TMs arranged in two continuous six-helix-bundles (Figure 6a). In addition to the 'core' MFS

fold, bacterial POTs possess two additional helices forming the HaHb domain. This domain replaces the intrahelical loop connecting the N- and C-terminal domain in other MFS transporters [60]. Whereas the Hb helix has a similar position in all POT structures published so far, the Ha helix adopts different conformations [60–62] (Figure 6b-f). Between different species, the sequence homology of the HaHb domain is low [61]. Together with the fact that it is only found in bacteria, it has been suggested that the HaHb domain does not play a part in a general conserved transport mechanism [59]. Mutational studies however have suggested that the rigidity of the HaHb domain and the charge of amino acid residues in the loop connecting Ha and Hb are important for the proper function of transporters [60]. Although the function is unknown, it was speculated that the function of the HaHb domain could be sensing of the 'core' movement, improving the protein stability, or that it might be necessary for protein folding [28,60].

The mammalian POTs PepT1 and PepT2 possess an extracellular domain between TM9 and TM10 that is not found in bacterial POTs [63]. This domain is not needed for the transport function but it might serve as interaction platform for other proteins e.g. proteases [63].

Notably, in all available structures POTs are present in either the inward-open or inward occluded state (Table 2). To date, there is no structure of a POT in the outward-open state. Thus, there is no information on all states of the transport cycle, although models of POTs in the outward-open state exist, based on other MFS transporters solved in an outward-open state [32,33]. Similarly, most structures are from bacteria with only one structure solved from a plant transporter, NRT1.1 from *Arabidopsis thaliana* [28,64], and no mammalian or human POT structures are available (Table 2). Nevertheless, because of the conservation of the transport mechanism and important amino acid residues involved in this process, bacterial transporters are a good substitute to provide an understanding of their mammalian relatives.

Among POTs there are three conserved motifs, the ExxERF, PTR2_1 and PTR2_2 motif [19,28] (Figure 7a). The first and third motifs are unique in POTs but the second motif is found across MFS transporters, where it is usually referred to as the A motif [19,28]. The ExxERF motif is located at the periplasmic half of TM1 and has the sequence E-X-X-E-R/K, the second motif is located between TM2 and TM3 and its sequence is G-X$_3$-(D/E)-(R/K)-X-G-[X]-(R/K)-(R/K), the third is located on the cytoplasmic half of TM5 and has the sequence F-Y-X-X-I-N-X-G. Mutations in these motifs lead to a loss of function of the transporter, indicating their importance for the function of the protein [19,28].

POTs are symporters and thus the proposed transport mechanism is similar to that of the MFS symporter LacY (Figure 7b). As a first step, the current transport model suggests the protonation of the transporter, after which the substrate can bind [28]. Substrate binding leads to the conformational change of the transporter from the outward-open to an occluded state by an induced fit mechanism. The proton is subsequently transferred inside the transporter, which adopts the inward-open state where first the substrate is released and then the transporter is deprotonated. The rate-limiting step is proposed to be the reversal of the apo form from the inward-open to the outward-open state. In general, the N-terminal domain seems to be less dynamic than the C-terminal domain [28]. With the N-terminal domain implicated mainly in proton coupling and the C-terminal domain preferably involved in substrate recognition [28].

To restrict or allow access to the binding site, so-called gates are formed by salt bridges that are either broken or formed between TMs. During the transport cycle, the first and second helices of each inverted repeat are participating in the formation of these gates. On the cytoplasmic side those are TM 4, TM5, TM10 and TM11, while the periplasmic gate is formed by TM1, TM2, TM7 and TM8. The formation and breaking of salt bridges is proposed to be controlled by substrate binding and release [28].
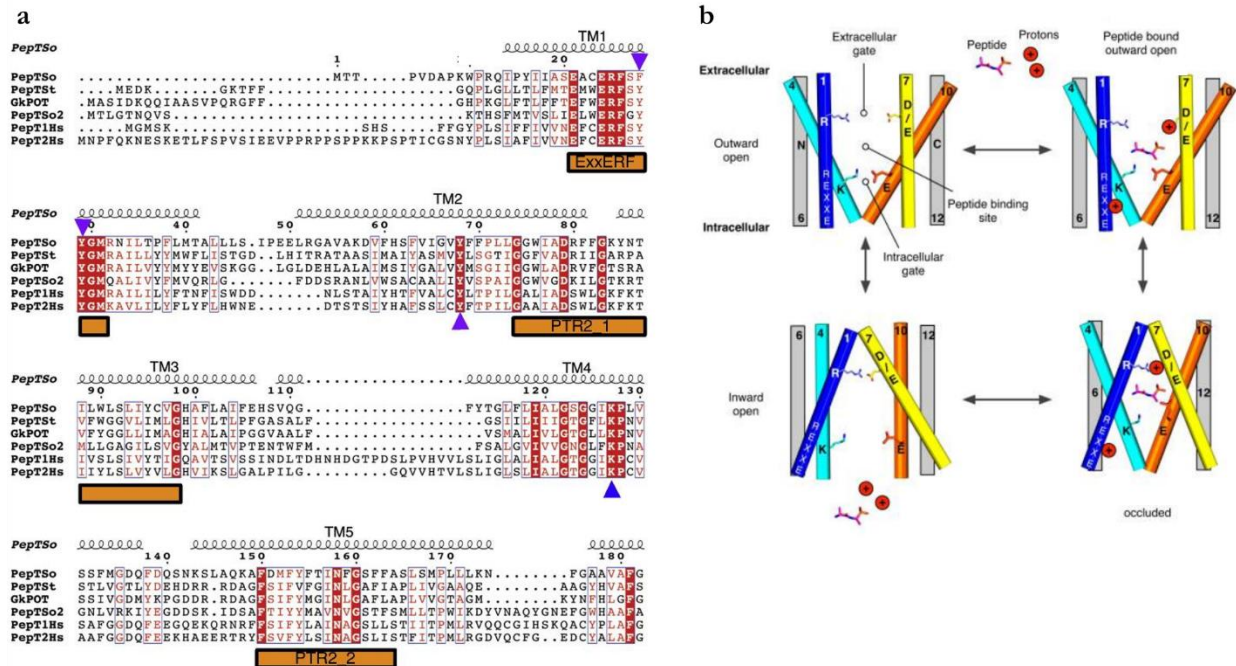
Figure 7: Conserved motifs and transport mechanism of POTs. (a) Sequence alignment of bacterial (PepT$_{So}$, PepT$_{St}$, GkPOT and PepT$_{So2}$) and human (PepT1Hs and PepT2Hs) POTs highlighting the three conserved motifs ExxERF, PTR2_1 and PTR2_2. Figure from [28]. (b) Model for the proton coupled transport in POTs. Conserved residues involved in substrate coordination and proton binding in POTs are shown at their position in the protein as well as the conserved E-X-X-E-R/K motif. TMs involved in closing the substrate binding site (Peptide binding site) on the periplasmic side (extracellular gate) or on the cytoplasmic side (cytoplasmic gate) are shown in color. TMs that act as scaffold for the other TMs are shown in gray. Figure from [58].

Table 2: Available structures of POTs. Published high resolution structures of members of the POT family. For each structure the corresponding PDB-ID is given and it is indicated if the transporters were solved in complex with a ligand and which conformation is adopted in the structure.

Protein	Organism	PDB-ID	Conformation	Ligand	Reference
PepT$_{St}$	*Streptococcus thermophilus*	5OXO	Inward-open	Apo	[62]
		5OXP	(Inward) occluded	PO$_4^{3-}$	[62]
		5OXQ	Inward-open	HEPES	[62]
		5OXK	(Inward) occluded	Ala-Gln	[62]
		5OXM	Inward-open	Ala-Glu	[62]
		5OXL	(Inward) occluded	Ala-Leu	[62]
		5OXN	Inward-open	Phe-Ala	[62]
		6EIA	Inward-open	HEPES	[62]
		6FMR	Inward-open	Apo	[65]
		6FMY	Inward-open	Apo	[65]
		4APS	Inward-open	Apo	[35]

		4XNI	Inward-open	Apo	[66]
		4XNJ	Inward-open	Apo	[66]
		4D2B	Inward-open	Apo	[67]
		4D2C	Inward-open	Ala-Phe	[67]
		4D2D	Inward-open	Ala-Ala-Ala	[67]
		5D58	Inward-open	Ala-Phe	[68]
		5D59	Inward-open	Ala-Phe	[68]
		5D6K	(Inward) occluded	Apo	[69]
		5MMT	Inward-open	Apo	[70]
		6GHJ	Inward-open	Phe-Ala-Gln	[71]
PepT$_{So}$	*Shewanella oneidensis*	4UVM	Inward-open	Apo	[36]
		2XUT	(Inward) occluded	Apo	[59]
PepT$_{So2}$	*Shewanella oneidensis*	4LEP	Inward-open	Alafosfalin	[61]
		4TPJ	Inward-open	Ala-Ala-Ala	[52]
		4TPH	Inward-open	Ala-Tyr	[52]
		4TPG	Inward-open	Ala-Tyr-Ala	[52]
		6JI1	Inward-open	Apo	[72]
		6JKC	Inward-open	Apo	[72]
		6JKD	Inward-open	Apo	[72]
GkPOT	*Geobacillus kaustophilus*	4IKV	Inward-open	Apo	[73]
		4IKW	Inward-open	SO_4^{2-}	[73]
		4IKX	Inward-open	SO_4^{2-}	[73]
		4IKY	Inward-open	Apo	[73]
		4IKZ	Inward-open	Alafosfalin	[73]
DtpA	*Escherichia coli*	6GS1	Inward-open	Apo	[50]
		6GS4	Inward-open	Valganciclovir	[50]
		6GS7	Inward-open	Apo	[50]
DtpD	*Escherichia coli*	4Q65	Inward-open	Apo	[60]
NRT1.1	*Arabidopsis thaliana*	5A2O	Inward-open	NO_3^{2-}	[74]
		5A2N	Inward-open	Apo	[74]
		4OH3	Inward-open	NO_3^{2-}	[64]
YePEPT	*Yersinia enterocolitica*	4W6V	Inward-open	Apo	[41]
PepT$_{Xc}$	*Xanthomonas campestris*	6EI3	Inward-open	Apo	[75]
PepT$_{Sh}$	*Staphylococcus hominis*	6H7U	Inward-open	5-aminolevulinic acid	[76]
		6HZP	Inward-open	5-aminolevulinic acid	[76]
		6EXS	Inward-open	Cys-Gly-3M3SH	[77]
		6GZ9	Inward-open	Valaciclovir	[76]

3.2.1.2.1.1. DtpA

In *E. coli* different systems for peptide uptake evolved, the major ones belonging to the MFS and ATP binding cassette (ABC) transporters [40,78,79]. While the dipeptide permeases (Dpp) and oligopeptide permeases (Opp) are primary active ABC transporters, POTs are secondary active MFS transporters [40,78,79]. In *E. coli*, four POTs are found which are named DtpA (YdgR), DtpB (YhiP), DtpC (YjdL) and DtpD (YbgH), respectively. They can be sorted into two subgroups which share high sequence identity, DtpA and DtpB with 51% and DtpC and DtpD with 56% sequence identity. Between the two subgroups the sequence identity is lower, ranging between 26-28% [80].

DtpA is proposed to be a good substitute to understand the pharmacologically interesting human transporter PepT1 [53]. While the bacterial transporter PepT$_{SO}$ shares a higher sequence identity to PepT1 than DtpA, DtpA has a more similar substrate specificity to PepT1 [48]. For example, PepT$_{SO}$ prefers a positively charged amino acid at the N-terminus and a negatively charged amino acid at the C-terminus of its substrates [48]. In case of PepT1 and DtpA, the opposite is the case. Additionally, DtpA prefers tri- over dipeptides, similar to PepT1 [48,50,81]. Most importantly, DtpA was also shown to interact with a similar set of drugs and prodrugs as PepT1 [12,53]. As an analogue to PepT1, DtpA binds the antiviral drugs valaciclovir and valganciclovir, the beta-lactam antibiotics cefadroxil, cefalexin and cephradine and it shows a low affinity for angiotensin converting enzyme inhibitors captopril and enalapril [12,53]. As mammalian membrane proteins are often difficult to work with due to low stability and expression yield, the fact that DtpA recognizes almost the same set of substrates as PepT1 makes it a suitable proxy for the investigation of the transport cycle dynamics in this dissertation. This is further supported by the common assumption that the members of the POT family share a similar transport mechanism.

3.2.1.2.2. Anion-cation symporter (ACS)

In addition to POTs, a different subfamily of MFS transporters was studied as well in this dissertation, the anion-cation symporter (ACS) family. Members of this family have been annotated in the genomes of bacteria and eukaryotes [82]. Hereafter the ACS transporters found in humans and *E. coli* will be described in more detail.

3.2.1.2.2.1. Human ACS

The mammalian members of the ACS family are also known as the solute carrier family 17 (SLC17) [82]. Originally, they were thought to transport phosphate ions but their substrate range also includes organic anions [82]. They can be sorted into four subgroups of different location and function, (i) type I phosphate transporters (NPT) are expressed at the plasma membrane and implicated in urea metabolism, (ii) Sialin is expressed in lysozymes and part of the glycoprotein and glycolipid metabolism, (iii) vesicular glutamate transporter (VGLUT) and (iv) vesicular nucleotide transporter (VNUT) both are expressed at synaptic vesicles and function in the synaptic storage of neurotransmitter [82] (Table 3).

The proteins NPT1, NPT3, NPT4 and NPT5 belong to the type I phosphate transporter subgroup [82]. All NPTs are expressed in the kidney, whereas NPT1, NPT3 and NPT5 are also expressed in the liver and NPT5 in the stomach and intestine as well [83–86]. Although they transport phosphate, it is suggested that this is not their natural function, because compared to other phosphate transporters expressed at the plasma membrane, they have a low affinity for phosphate [87]. Instead it was proposed that they transport urates, in the hepatic and renal clearance of organic anions [82,86]. In addition,

NPTs secrete drugs like aspirin or beta-lactam antibiotics from cells and thus a reduction of their activity was suggested to improve the bioavailability of those drugs [86,88].

The second subgroup are the VGLUTs. Three proteins belong to the VGLUT subgroup, VGLUT1, VGLUT2 and VGLUT3. They all transport glutamate, but not aspartate into synaptic vesicles [89]. While VGLUT1 and VGLUT2 are found exclusively in glutamatergic neuron, VGLUT3 is also found in other neurons [90]. Although knockout mice for the VGLUTs show severe phenotypes such as premature death, seizures and deafness, they have not been associated to human diseases [91–94]. An exception is VGLUT3 which is linked to non-syndromic deafness [93].

The only protein belonging to the third subgroup is VNUT. VNUT is expressed in synaptic vesicles as VGLUTs but due to its transport of ATP and not glutamate it was sorted into a subgroup on its own [82].

The only protein belonging to the last subgroup is Sialin. Sialin is expressed in lysosomes across all tissues [21,82]. It mediates the proton-coupled efflux of sialic acid and glucuronic acid from degraded glycoproteins and glycolipids [95]. It has also been shown that Sialin recognizes other acidic sugars like lactate and gluconate, but not uncharged sugars such as galactose [95,96]. Mutations of this protein have been linked to severe Mendelian lysosomal storage diseases [96]. The milder version, called Salla disease, is characterized by a reduced transport rate of the protein leading to developmental delays, a low IQ, ataxia, cognitive impairment but patients have a normal life expectancy [97,98]. In contrast, patients with infantile free sialic acid storage disease (ISSD) have more severe neurological symptoms and a shortened life expectancy [95]. In Sialin mutants linked to ISSD, the protein cannot fulfill its transport function anymore [95]. This leads to a buildup of sialic acid in the lysozymes preventing it to be recycled into new glycoproteins and glycolipids [99]. This could explain the changes in myelination of neurons observed in patients [99]. To build up the myelin sheet around the axons of neurons, gangliosides are needed which contain sialic acid [100].

Although SLC17 transporters are implicated in diseases such as gout, hyperuricemia and different neurodegenerative diseases, there are currently no drugs targeting this transporter family [82]. This might also be due to the lack of structural information on ACS transporters. To date, only two high-resolution structures are published, DgoT from *E. coli* and recently VGLUT2 from *Rattus norvegicus* [101,102].

3.2.1.2.2.2. ACS in *E. coli*

The ACS homologues found in bacteria are predicted to transport anions, mostly anionic sugars. Many of the bacterial ACS are not well characterized and their substrates were only proposed by their presence in operons of metabolic enzymes [103]. In *E. coli*, there are six ACS transporters DgoT, LgoT, GudP, GarP, ExuT and RhmT, all transporting anionic sugar coupled to proton symport [101,103–106]. *E. coli* is flexible with respect to its carbon source, it can metabolize different sugar molecules both from mucus and dietary compounds of its host. For example, galactonate is a product of the galactose metabolism in humans [107–110]. It was found that besides glucose, *E. coli* can grow with the sugar acids D-galactonate, L-galactonate, glucarate, glucuronate, galacturonate, and gluconate as sole carbon and energy source [106,111–113]. *E. coli* breaks these sugar acids down to, amongst others, pyruvate which can enter the citric cycle [104,106,112–114]. The utilization of this sugar acid metabolism is implicated in the mammalian gut colonization of *E. coli*. [115–117].

The proposed substrates for the *E. coli* ACS transporters are summarized in Table 3. Functional studies showed that DgoT is highly specific for D-galactonate [101]. In contrast, substrates for LgoT, GudP, GarP, ExuT and RhmT were proposed

based on their respective operons [103–106]. The substrates of all bacterial ACS transporters also serve as inducer for their expression [104–106,118,119].

Bacterial ACS transporters show a high sequence and functional similarity to their human relatives [101]. For example, DgoT serves as a model for Sialin [101]. As this is the only ACS member with high-resolution structures available in two conformational states, the transport mechanism of ACS transporters will be described in more detail based on DgoT.

Table 3: Members of the anion-cation symporter family in *E. coli* and human. The so far annotated members of the ACS family in *E. coli* and humans with their proposed substrates. Where a high-resolution structure is available, the corresponding PDB-ID is given.

Protein	Organism	PDB-ID	Proposed substrates	Reference
DgoT	*Escherichia coli*	6E9N, 6E9O	D-galactonate	[101]
LgoT	*Escherichia coli*		L-galactonate	[105]
GarP	*Escherichia coli*		D-galactarate, (potentially D-glucarate)	[114,119]
GudP	*Escherichia coli*		D-glucarate, (potentially D-galactarate)	[103,114]
ExuT	*Escherichia coli*		D-glucuronate, D-galacturonate	[106]
RhmT	*Escherichia coli*		L-Rhamnonate	[104]
NPT1 (SLC17A1)	*Homo sapiens*		Phosphate, organic anions, chloride	[82]
NPT3 (SLC17A2)	*Homo sapiens*		unknown	[82]
NPT4 (SLC17A3)	*Homo sapiens*		Organic anions	[82]
NPT5 (SLC17A4)	*Homo sapiens*		unknown	[82]
Sialin (SLC17A5)	*Homo sapiens*		Sialic acid, glucuronic acid, acidic sugars, aspartate, glutamate	[82,120]
VGLUT1 (SLC17A6)	*Homo sapiens*		glutamate	[82]
VGLUT2 (SLC17A7)	*Homo sapiens*	6V4D	glutamate	[102]
VGLUT3 (SLC17A8)	*Homo sapiens*		glutamate	[82]
VNUT (SLC17A9)	*Homo sapiens*		Purine nucleotides	[82]

3.2.1.2.2.3. Sugar acids

Sugar acids are monosaccharides that have one or more carboxyl groups. They can be sorted into different groups depending on the position of the carboxyl groups. Here, three groups of sugar acids will be introduced, aldonic acids, uronic acids and aldaric acids [121].

In aldonic acids, the aldehyde group of an aldose is oxidized to a carboxyl group [121]. The resulting sugar acids are given the suffix -onic acids. For example, oxidation of the aldehyde group in glucose results in the sugar acid gluconic acid, of galactose in galactonic acid and rhamnose in rhamnonic acid (Figure 8). The corresponding anions are called gluconate, galactonate and rhmanonate [121].

Oxidation of the primary hydroxyl group to a carboxyl group gives an uronic acid. Here, the suffix -uronic acid is used. Oxidation of the hydroxyl group of the C6 atom of glucose thus gives rise to glucuronic acid and of galactose to galacturonic acid (Figure 8). Here, the anions are called glucuronate and galacturonate [121].

If both the aldehyde group and the primary hydroxyl group are oxidized to carboxyl groups, these sugar acids are called aldaric acids. The suffix used in this case is -aric acid [121]. In case of oxidation of glucose, the corresponding sugar acid is glucaric acid with the anion named glucarate and in case of galactose galactaric acid and galactarate [121] (Figure 8).

Figure 8: Structures of selected sugar acids based on glucose, galactose and rhamnose. Uronic, aldonic and aldaric acids of the sugars glucose, galactose and rhamnose.

3.2.1.2.2.4. The ACS transport mechanism based on DgoT

The *dgo* operon encodes for a transporter (DgoT), a kinase (DgoK), a dehydratase (DgoD), an aldolase (DgoA) and a regulator (DgoR) [107,112]. DgoK, DgoA and DgoD are needed to break down D-galactonate , which is transported into the cell by DgoT, to pyruvate [107,112] (Figure 9). DgoR binds to a DNA stretch overlapping the *dgo* operon promotor region, thus inhibiting translation [118]. D-galactonate in turn can bind to DgoR which leads to a conformational change of DgoR [118]. DgoR then dissociates from the DNA and the operon can be transcribed [118].
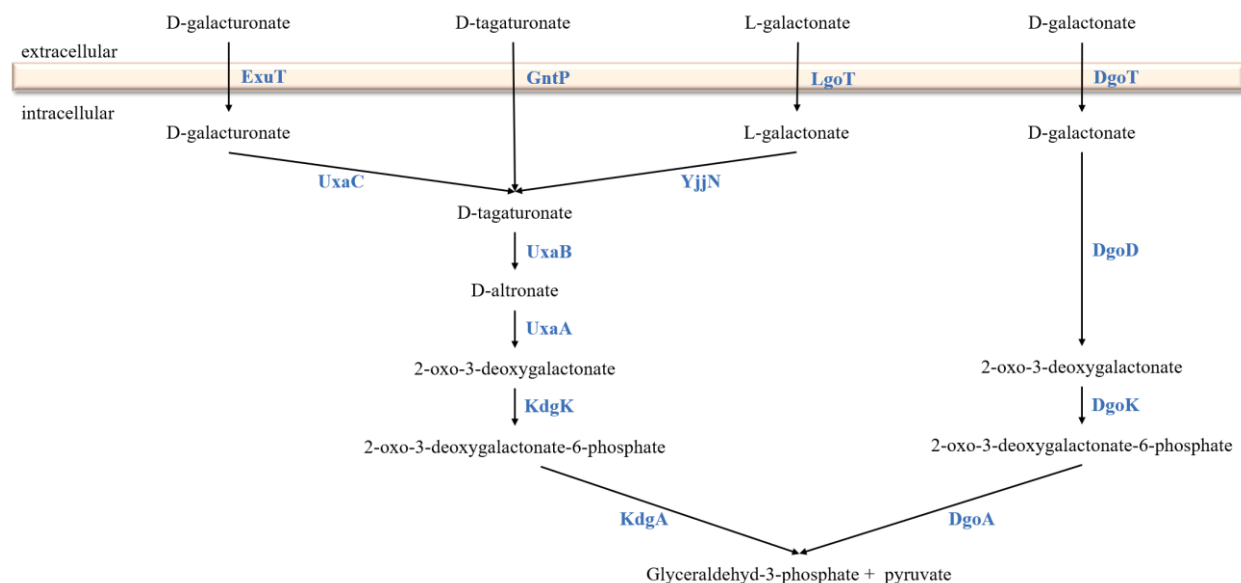
Figure 9: Comparison of the metabolic pathway for L-galactonate and D-galactonate in *E. coli.* Metabolic pathway for the utilization of the sugar acids D-galacturonate, D-tagaturonate, L-galactonate and D-galactonate. Figure adapted from [107,122].

Structures for DgoT are available in the inward-open and outward-open state [101] (Figure 10a-b). The structure for DgoT shows that the protein has 12 TMs arranged in the typical MFS fold with two six-helix-bundles, the N-and C-terminal domain. In addition, two intracellular helices (ICH) domains are located on the cytoplasmic side of the protein [101] (Figure 10a-b). This has also been observed for sugar transporters of other MFS subfamilies. Furthermore, in agreement with what was observed for other MFS families, TM1 and TM7 close the periplasmic gate while T4 and TM10 form the cytoplasmic gate [101].

The substrate binding site is composed of residues from the N- and C-terminal domain. A closer inspection of the binding site could explain the particular substrate specificity of DgoT. The carboxyl group of D-galactonate is coordinated by R47, Y79 and Y44. These residues are conserved among almost all ACS transporters [101]. VGLUTs have a phenylalanine in the position equivalent to Y44 of DgoT that coordinates the carboxyl group of the substrate, and in NPT1 and NPT3 R47 is not conserved [102]. A reason could be that their substrates do not have a carboxyl group and therefore need a different way of coordination. The hydroxyl groups of D-galactonate are coordinated by Q164, Q264, S370 and N393 [101] (Figure 10c). Thus, all hydroxyl groups of D-galactonate are coordinated by DgoT. The substrate binding site is thus highly tailored to D-galactonate. Even similar compounds cannot be ideally coordinated when the hydroxyl groups have different orientations compared to D-galactonate. This is the case for gluconate, which differs from D-galactonate in the position of one hydroxyl group. This hydroxyl group in D-galactonate is coordinated by N393 and Q264 of DgoT [101]. For gluconate this coordination is not possible, explaining why it could not be shown to be a substrate for DgoT [101].

Apart from the binding site, there are several other conserved residues found across the ACS family. Among these, R126 and E133 are implicated in proton translocation during transport in DgoT [101]. E133 is one of two reversible protonation sites on the periplasmic side of the protein [101]. In the outward-open state, E133 forms an interaction with R47 of the substrate binding site. Protonation of E133 interrupts this interaction and the substrate can enter the substrate binding

33

site where it is coordinated by R47. This substrate binding stabilizes E133 in its protonated state. With both E133 and D46 protonated and the substrate bound, the protein switches conformation from the outward-open to the inward-open state through movement of both the N- and C-terminal domain. In the inward-open state, the substrate is released. Upon deprotonation of E133, the interaction to R47 can reform and the transporter in its apo form can adopt the outward-open state again [101] (Figure 10d).

Notably, the VGLUTs have only one protonation site equivalent to E133, and are missing one equivalent to D46 [102]. Compared to DgoT the ratio of protons transporter per substrate might be different for those transporters [102]. Interestingly, NPTs do not have either of those protonation sites conserved, which might be explained by them being sodium-coupled instead of proton-coupled [82,102].
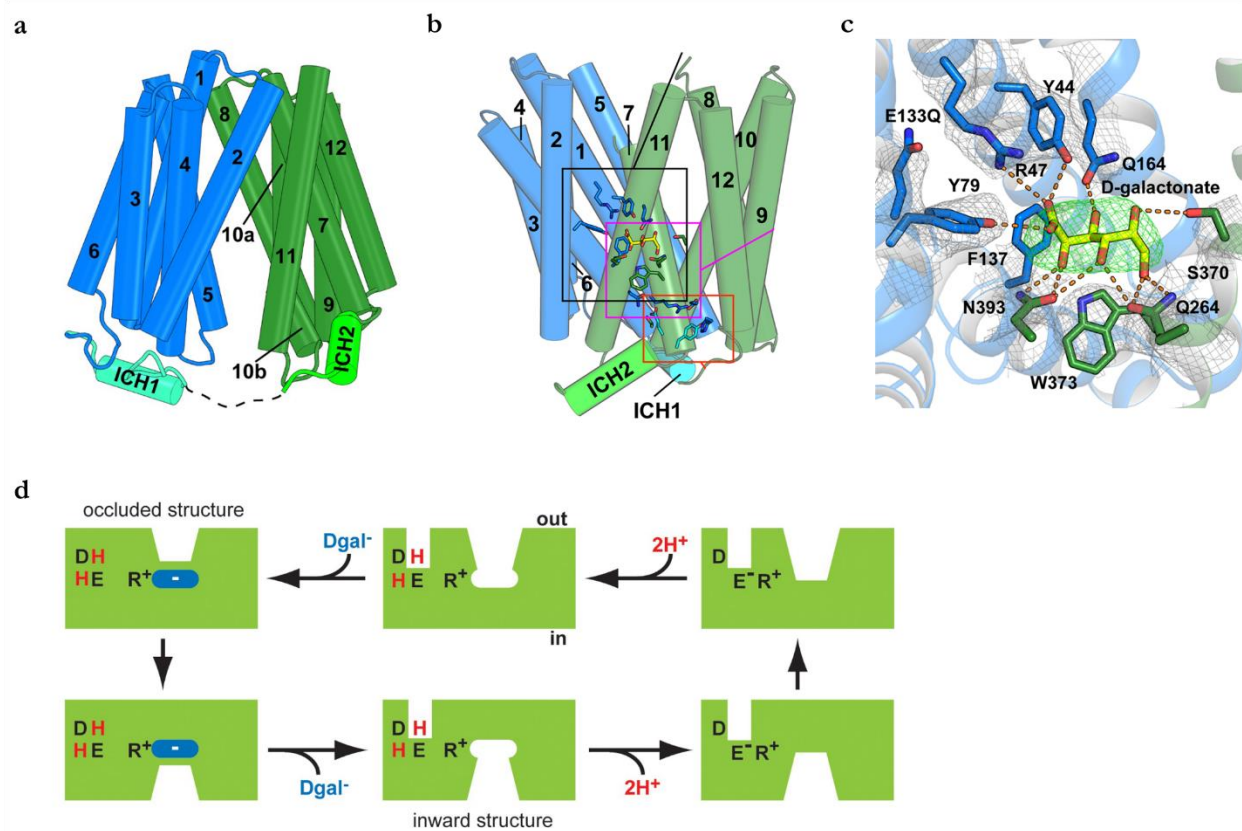
Figure 10: Structure and proposed transport mechanism for DgoT. (a-b) The crystals structures of DgoT from *E. coli* in (a) the inward-open and (b) outward-open state show the typically MFS fold. The 12 TMs of the transporter are arranged in two bundles of six consecutive helices each. The first six helices from the N-terminus form the N-terminal domain (blue) and the six helices from the C-terminus form the C-terminal domain (green). Analogously to what was observed for sugar transporters of other MFS families, DgoT exhibits two intracellular helices (ICH) domains. (c) Coordination of D-galactonate in the substrate binding site. (d) Proposed mechanism for substrate transport by DgoT. In the substrate unbound outward-open conformation the charged side chain of the residues E133 (E⁻) and R47 (R⁺) interact with each other. Protonation of E133 (E) and D46 (D) breaks this interaction, leaving R47 free to bind the substrate (Dgal). Substrate binding leads to the formation of an occluded state. Through conformational changes the transporter then adopts the inward-open state, where the substrate is released. Subsequent deprotonation of E133 and D46 allow for

the reformation of the interaction between E133 and R47. This apo form of the transporter is able to switch back to the outward-open conformation. Figure from [101].

3.2.1.2.2.5. LgoT

Compared to DgoT, little is known for LgoT. The *lgo* operon consists of three genes, LgoT (YjjL), YjjM and YjjN. All three genes were found to be upregulated when *E. coli* was grown on L-galactonate as its only carbon source [105]. YjjM was found to be the regulator for the operon and YjjN to be an oxidoreductase [105]. This enzyme catalyzes the reaction of L-galactonate to D-tagaturonate [123]. Therefore, LgoT was proposed to be the transporter for L-galactonate [105]. Curiously, the metabolic pathway for L-galactonate differs from that of D-galactonate. D-galactonate has its own metabolic pathway with unique enzymes [113]. For L-galactonate only the first step, the conversion to D-tagaturonate is unique. D-tagaturonate is then further broken down in the same pathway that is also utilized for D-galacturonate [113] (Figure 9).

3.3. Methods to study membrane proteins

The involvement in various cellular processes, the potential to be drug targets and their role in disease highlights the importance of studying membrane proteins. For this purpose, high-resolution 3D-structures of membrane proteins are a valuable tool to gain insight into the biological processes performed by membrane proteins [124,125]. High-resolution structures can help to understand the biological function and mechanism as well as show interactions of the protein with substrates, effectors, cofactors, ions, metabolites or other proteins [126]. Although, membrane proteins account for about 60 % of drug targets, only a fraction of the structures deposited in the Protein data bank (PDB) correspond to membrane proteins [127–129].

Despite considerable improvements made in recent years, the major challenge in studying membrane proteins is still the difficulties to express and purify well folded, functional protein [128]. In contrast to soluble proteins, membrane proteins have to be targeted to and inserted into the membrane. The capacity of the translocation machinery and the availability of space in the membrane often limits the copy number a cell is able to produce of a membrane protein in a well folded and functional manner [130]. Furthermore, for most functional and structural studies, membrane proteins are extracted from their natural membranes. Often, detergents are used to keep the protein in solution by shielding the hydrophobic region from the aqueous buffer environment. This extraction and transfer into artificial environments can result in unfolding or loss of function of the protein [131,132].

Taken together, the low yield of membrane proteins from the expression systems and the challenges associated with purification are major limiting factors when studying membrane proteins.

3.3.1. X-ray crystallography of membrane proteins

To obtain near atomic resolution structures, X-ray crystallography is the most common method. Therefore, the membrane protein needs to be first overexpressed. Next, the protein is solubilized using detergents that substitute for the lipid bilayer, followed by purification, crystallization, data collection and structure determination.

Membrane proteins can form two types of crystals, Type I and Type II [133] (Figure 11). Type I crystals are 2D crystals that are stacked in layers on top of each other and are typically obtained through lipidic cubic phase (LCP) crystallization [133,134]. However, type II crystals are more common than type I crystals [135]. In type II crystals the protein is still surrounded by its detergent micelle in the crystal thus, there are only few crystal contacts possible, which are formed by the hydrophilic parts of the proteins [133]. Additionally, the size of the micelles prevents a tight packing of the proteins in the crystals. This typically leads to less well diffracting crystals [135,136]. For type II crystals it is therefore critical to optimize the choice of detergent [135]. Detergents with small micelles are favorable for crystallization, as they facilitate tighter crystal packing but at the same time they can lead to aggregation if the micelle is too small to fully shield the hydrophobic part of the protein. [126]. Because the best detergent for solubilization of the protein might not lead to well-diffracting crystals, a change of detergent during the purification process might be beneficial [126,137]. Even the use of mixed detergent micelles from several detergents and/or lipids can be tried [138].
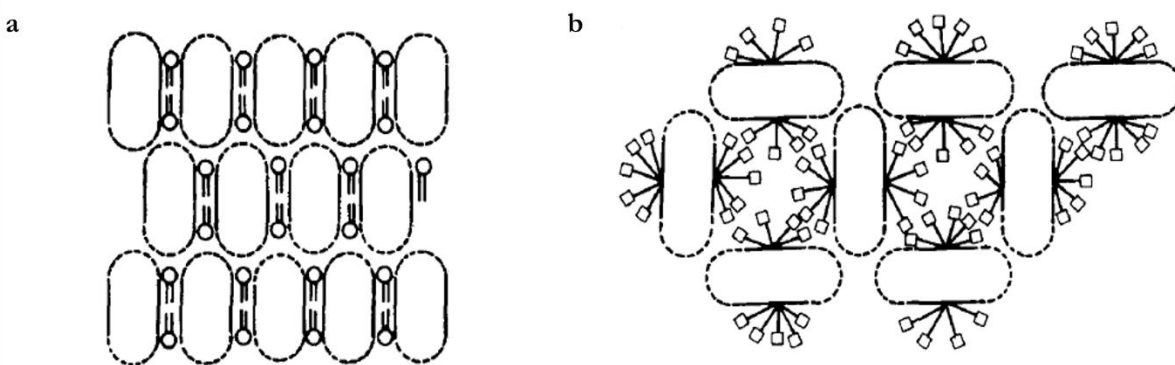
Figure 11: Types of crystals formed by membrane proteins. (a) Type I crystal are 2D crystals which are stacked on top of each other. This crystal type is typical for crystals obtained by the lipidic cubic phase method. (b) Type II crystals are crystallized with the detergent micelle present, leading to weaker crystal contacts. Dotted lines represent the hydrophilic surface of the protein, spheres represent lipids, squares represent detergent. Figure from [135].

Due to several reasons such as the low expression level of membrane proteins, limited success of extracting proteins from the membrane, low purification yields and a lack of well-ordered 3D crystals, obtaining high-resolution structures of membrane proteins is still a challenge [128]. However, several improvements in the fields of expression hosts, development of new detergents and crystal optimization strategies such as the use of crystallization chaperones, like monoclonal antibodies (AB) or heavy-chain only ABs from camelids (nanobodies), or the LCP method were made [128].

In LCP crystallization, the lipidic cubic phase is formed spontaneously upon mixing of lipids with the protein solution [134]. The lipids arrange in a curved continuous bilayer into which the protein can insert [139]. Inside the bilayer the proteins are able to migrate and form type I crystals with the crystal contacts made by both the hydrophilic and hydrophobic part of the protein [136,140]. Typically, monoolein is used as the lipid to form the bilayer [141]. Depending on the temperature and ratio of lipid to protein solution the system can adopt different phases (Figure 12a) [140]. In the case of monoolein, the lipid cubic phase (Pn3m) is formed at room temperature when three parts monoolein are mixed with two parts protein solution

[140]. A phase separation is induced by the addition of precipitant, and if the protein is thereby enriched in one of the phases, nucleation and subsequently crystal growth can occur [140] (Figure 12b).
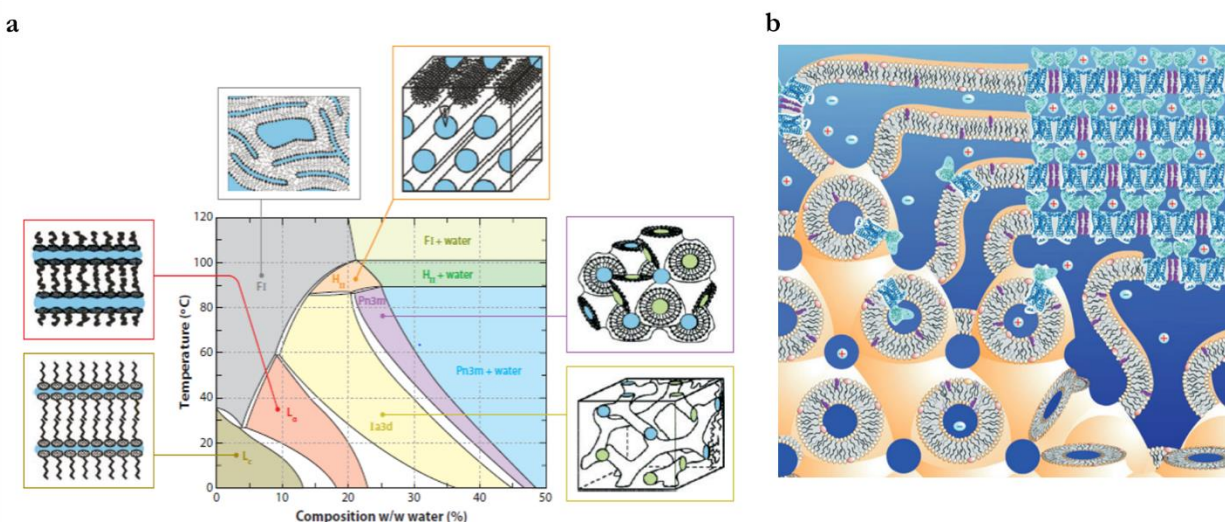
Figure 12: Lipidic cubic phase method. (a) Temperature vs. composition phase diagram for the monoolein/water system with schematic illustration of the different phases that can be adopted. Fl – fluid isotropic phase, HII – inverted hexagonal phase, Lα – lamellar liquid crystalline phase, Lc – lamellar crystal phase, Pn3m – lipid cubic phase, Ia3d – cubic phase. (b) Schematic representation of nucleation and crystal formation of membrane proteins in LCP. Figure from [140].

3.3.2. Förster resonance energy transfer

Whereas X-ray crystallography can provide high resolution structures of macromolecules, the resulting structures only represent static snapshots showing one possible conformation of the molecule. In addition, crystallization may favor a conformational state or the crystallization conditions may induce a conformation that is otherwise not adopted by the protein *in vivo* [142,143]. For cryogenic electron microscopy (cryo-EM), several conformations of a protein can be present in the same sample [144,145]. Analogously to crystallography, here the dynamic information about the transition between different conformational states is missing as well. To fully understand the function of dynamic systems, e.g. the transport cycle of membrane transporters, information about the interconversion time between different conformational states and how often these transitions take place are indispensable.

This information can be obtained by a process called Förster resonance energy transfer (FRET). This mechanism describes the non-radiative transfer of energy via dipole-dipole coupling from one excited fluorophore to another fluorophore which is in close proximity [146].

The two fluorophores used in FRET experiments are referred to as donor (D) and acceptor (A). The prerequisites for FRET to occur are that firstly, the emission spectra of the donor and the absorption spectra of the acceptor are overlapping. Secondly, the absorption spectra of the donor and acceptor have to be separate to archive specific excitation. The same applies for the emission spectra to record specific emission signals. Thirdly, the dipoles of both fluorophores must align and fourthly, the fluorophores must be in the right distance range with respect to each other [146] (Figure 13a-

b). If the donor fluorophore is excited, it can reverse to the ground state through three different processes: emission of a photon, a chemical reaction, or the radiation free transfer of energy to an acceptor. If the latter is the case, the acceptor is excited and can in turn reach its ground state by emission of a photon [146]. The wavelength of the photon that is emitted by the acceptor is of a longer wavelength than the photon that would be emitted by the donor. This change in wavelength of the emitted photons can be used as a measure for the efficiency of FRET. The efficiency of FRET (E) is strongly dependent on the distance between the fluorophores (r) as illustrated by the following equation [147]:

$$E(r) = \frac{R_0{}^6}{R_0{}^6 + r^6}$$

Here, R_0 is the Förster radius, the distance between the two fluorophores where the FRET efficiency is 50% [146]. This is a characteristic of a FRET fluorophore pair and usually ranges between $2 - 6$ nm [148]. For fluorophore distances in the vicinity of R_0, the small changes in distance result in large changes of FRET efficiency (Figure 13c). Thus, the measurements are very sensitive in this region. The more the distance between the fluorophores deviates from R_0, the less sensitive the FRET measurements becomes. As a rule of thumb, FRET efficiencies can be accurately measured when the distance between the donor and acceptor is between $0.5\ R_0$ to $1.5\ R_0$ [149] (Figure 13c). Hence, FRET measurements are suitable to report distances of 1 to 10 nm, depending on the fluorophore pair used [150]. In general, they can detect distance changes up to a resolution of 3 Å [146]. This size and resolution range makes FRET a well-suited technique to study biological molecules. Systems that have been studied using FRET include but are not limited to dynamics of motor proteins, enzymatic reactions, structural transitions in DNA nanomachines, conformational changes of proteins, complex formation and protein folding [151–155]. The observation time of such measurements ranges from ms to minutes, depending on the chosen set-up [156]. Two specific set-ups will be discussed in more detail in chapter 3.3.2.2, FRET measurements of freely diffusing sample and of surface-immobilized samples.
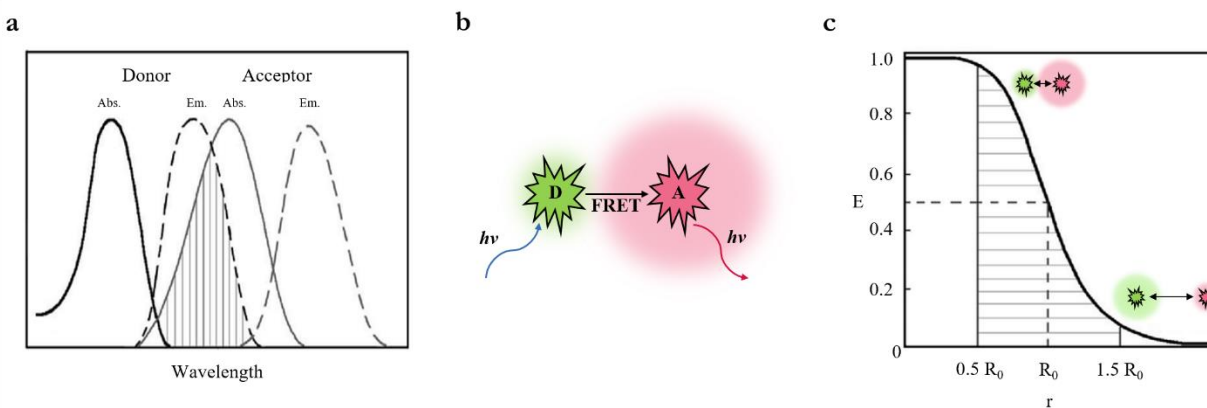
Figure 13: Basic principles of Förster resonance energy transfer (FRET). (a) Exemplary absorption and emission spectra overlay of donor and acceptor fluorophores. The grey shaded area is the overlap of the donor emission with the acceptor absorption spectra which is a prerequisite for FRET to occur. (b) Schematic of energy transfer in FRET. A donor molecule (green star) is excited with light of the appropriate wavelength (blue arrow) but instead of emitting light itself it passes on the energy to an acceptor molecule (red star) in the proximity (black arrow). This acceptor molecule then emits light of a specific wavelength (red arrow) which differs from the wavelength of the light which would have been emitted by the donor. (c) Schematic of the distance dependency of the FRET efficiency.

The FRET efficiency (E) changes with the distance (r) between the donor (green star) and acceptor (red star). The distance between the fluorophores at which the FRET efficiency is 0.5 is defined as the Förster radius (R_0). The useful range to measure FRET is between 0.5 times R_0 and 1.5 times R_0 (grey shaded area). Within this range, changes in distance between the fluorophores sensitively translate into changes in FRET efficiency. Figure adapted from [149].

3.3.2.1. Labeling of samples for FRET measurements

Different biological macromolecules can be labeled for FRET measurements. The work described in this dissertation was carried out solely on protein samples and therefore only labeling of those will be described here.

Introducing fluorophores for FRET measurements into the protein sequence can be realized by either exploiting the presence of intrinsic tryptophan residues, the incorporation of unnatural amino acids or through chemical modification of side chains [149]. Among these, the chemical modification of cysteine residues is the most widely used approach [149]. On the one hand, site-directed mutagenesis allows to site-specifically introduce cysteine residues where the labels should be placed. Ideally, the distance between them is in the range of the chosen fluorophore pair's R_0. On the other hand, cysteine residues are unique among the proteinogenic amino acids for their thiol group. This can be exploited by equipping the fluorophore with a maleimide group, resulting in thiol-maleimide coupling [157]. Among the most commonly used fluorophores are cyanine dyes and fluorophores of the Alexa series [158,159]. Both are available with different functional groups for coupling to the protein sample.

Fluorophores that are well suited for FRET studies exhibit certain characteristics: (i) they are stable under high photon flux, (ii) they have a high molecular absorptivity, meaning that they absorb light of a given wavelength well, (iii) they have a high fluorescence quantum yield, which is a high ratio of absorbed to emitted photons, (iv) they show minimal blinking, (v) they are small and (vi) they are soluble in water, as experiments will commonly be carried out in aqueous buffer solutions [160].

Blinking is a process where a fluorophore spontaneously enters a nonfluorescent state [146]. In contrast to photobleaching, the fluorophore can leave this state and become fluorescent again [146]. In photobleaching the fluorophore is irreversibly destroyed through a chemical reaction [160]. This is a limiting factor for the observation time of a molecule. This behavior can be advantageous as well e.g. to test if the signal was originating from a single molecule, where one would expect a single photobleaching event, or from multiple molecules, with multiple events observed [146].

3.3.2.2. Single-molecule FRET of freely diffusing and surface-immobilized samples

Although the theory of FRET has already been developed in the 1940s and the technique was proven to be suitable as a spectroscopic ruler in the 1960s, the first single-molecule FRET (smFRET) measurements became technically possible in the 1990s [147,161,162]. Reasons for this are that the measurement of a single molecule in contrast to ensemble measurements requires high-powered illumination e.g. a laser, the reduction of background fluorescence and especially sensitive detection [147,161,162].

Single-molecule measurements have the advantage that changes in the sample are not averaged out, while this could be the case in ensemble measurements where the changes are happening asynchronously [156]. Both, freely diffusing samples as well as surface-immobilized samples can be measured in a single-molecule regime. Typically, a confocal microscope is

used to observe freely diffusing sample in solution [160]. Here, the low sample concentration, usually 10 to 100 pM, and small excitation volume of around 1 fL, ensure that only one molecule is in the confocal spot at any given time [160]. At the same time, the low concentration also makes it more likely that a molecule that just left the confocal spot returns to it instead of a new molecule entering the confocal spot [163]. This circumstance is exploited further during downstream analysis of the recorded FRET data, using a process called the recurrence analysis of single particles (RASP) [163], which will be explained in more detail in chapter 3.3.2.4.

Each time a molecule enters the confocal spot, the donor molecule is excited by a laser and the fluorescence is recorded in form of a burst signal [160]. The bursts from all molecules that passed through the observation volume during the course of the measurements are later summed up in a FRET histogram[160]. In these histograms the number of molecules that were observed to exhibit a certain FRET are displayed (Figure 14a).

Gaussian functions are fitted to the recorded histograms to determine the number of populations [148]. By analyzing the width, position and number of the populations in a histogram, one can gain information about the dynamics of the studied system [148]. For example, if the conformational changes occurring in the system are much slower than the diffusion time through the confocal spot, the histogram will show two static species [148]. Otherwise, if the conformational changes are much faster than the diffusion time, the data will average out and different populations are no longer distinguishable.

Four processes influence the position and width of peaks, among those are, (i) the intramolecular dynamics of sample and in relation (ii) the observation time. Furthermore, (iii) the time the donor stays excited before going back to ground state (the fluorescence lifetime of the donor) and (iv) the rotational correlation time of the fluorophores [160]. Ideally both fluorophores are able to rotate freely as the occurrence of energy transfer between them depends on their orientation towards each other. The anisotropy of the fluorophores is a measure to judge if the free rotation is possible. Here, a value below 0.2 for the anisotropy is presumed to be good [146]. If the dyes are found to be able to rotate freely and quickly the FRET efficiency can be used to calculate distances directly from the relative fluorescence, meaning the ratio of photons emitted by the donor and acceptor.

Importantly, FRET only reports one-dimensional distance measurements, whereas proteins are three-dimensional entities. To interpret the different populations observed in a FRET histogram, high resolution structures are therefore helpful. Theoretical FRET efficiencies can be calculated from the distances measured in the structure and then compared to the data obtained during FRET measurements. One should keep in mind that the measurement of multiple fluorophore pairs at different positions of the protein might be useful. One fluorophore pair might not change its distance although the protein moves, e.g. in rigid body movement, when they are in the same domain or maybe movements are more complex than a rigid body movement. In this case more than one distance information would be needed to gain a complete picture. With diffusion-based measurements, dynamics between 0.1 ms to 10 ms can be studied [156]. The measurements as a whole can last from minutes up to hours depending, among other things, on the labeling quality of the sample [160]. The higher the fraction of molecules that possess a donor and an acceptor molecule, the shorter the acquisition time. The observation time for one molecule is its diffusion time through the confocal spot, typically in the range of 1 ms [156]. In contrast, when immobilized samples are measured with a confocal microscope, the observation time per molecule can be as high as tens of seconds until that molecule undergoes photobleaching [156,160].

For surface-immobilized samples, it is not the bursts of different molecules diffusing through the confocal spot that are recorded, but the same molecule over time. These time traces can last from seconds to hours depending on the

photostability of the fluorophores [146]. Time traces from many molecules are then subsequently summed up into histograms (Figure 14b). Compared to diffusion measurements, slower dynamics on the timescale from seconds to minutes can be followed with this technique [160].

The data presented in this dissertation was measured for freely diffusion sample, therefore different analysis techniques for data recorded with this set-up will be discussed.
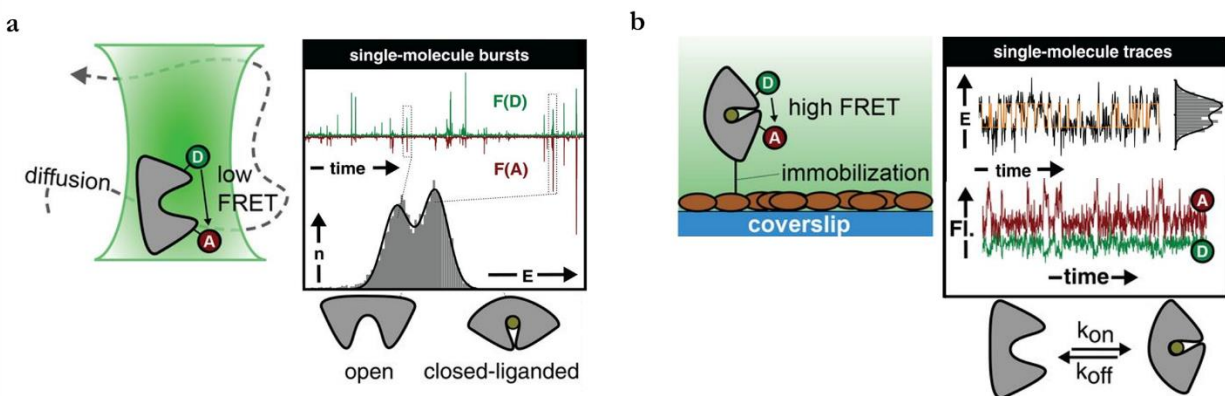
Figure 14: Experimental set-ups to measure single-molecule FRET (smFRET). (a) Diffusion based smFRET measurements. Here, molecules freely diffuse in solution. Due to the very low concentration of the molecules, statistically only one enters the confocal spot at any given time, thus making this set-up a single-molecule experiment. When a molecule enters the confocal spot, bursts of fluorescence from the donor or acceptor molecule are recorded. These single-molecule bursts are recorded over a period of time and then summed up into FRET histograms according to how often (n) a burst of a certain FRET efficiency (E) was observed. The histogram can subsequently be fitted with Gaussians corresponding to different FRET populations. These populations can in turn be assigned, e.g. to different conformational states of the protein according to prior knowledge of the system, for example from high-resolution structures. (b) SmFRET on surface-immobilized sample. Here, the molecules are fixed in place onto a surface. This enables to focus on a single molecule and to record its FRET trace over a period of time. Several traces of different molecules can then be summed up to FRET histograms and analyzed as described for diffusion-based smFRET measurements. Figure taken from [156].

3.3.2.3. FRET analysis using pulsed interleaved excitation

Due to labeling imperfections, photobleaching or blinking, not every protein in a sample will carry one active donor and one active acceptor molecule [148]. Among these, especially the ones which do not have an active acceptor molecule (D-only) interfere with the analysis of the FRET data. This D-only population will show a peak in the FRET histogram around a FRET efficiency of 0. This could make it more difficult to detect FRET populations of very low FRET efficiencies. D-only signal could originate from incomplete labeling, the acceptor being photobleached or the acceptor blinking [148].

Pulsed interleaved excitation (PIE) makes it possible to obtain information on all fluorophore positions on a protein and sort them into high FRET, low FRET, D-only, and acceptor only species [164]. To do this, the sample is excited at the wavelength for the donor and the acceptor, respectively. At the same time, the alternation between those two wavelengths

is faster than the diffusion time of the molecule through the confocal spot. Thus, for each molecule entering the confocal spot, information about its behavior at donor and acceptor excitation wavelengths is obtained [164].

Proteins that carry an active donor and acceptor within the right distance will exhibit FRET upon excitation of the donor, with the FRET efficiency dependent on the distance of the fluorophores (Figure 15ai). Proteins that only carry an active donor but the acceptor molecule is either too far away for FRET to occur or not active due to photobleaching or blinking (Figure 15aii) or the protein was not labeled with both fluorophores (Figure 15aiii) will exhibit only donor emission upon donor excitation. Proteins that do not have an active donor will not exhibit any emission upon donor excitation (Figure 15aiv). *Vice versa*, upon acceptor excitation, proteins with both fluorophores in FRET distance (Figure 15av) as well as out of FRET distance (Figure 15avi), and such molecules that have only an active acceptor (Figure 15aviii) will show acceptor emission. Proteins with only an active donor will show no emission (Figure 15avii). Those species can then be sorted according to the number of different fluorophores detected, the stoichiometry S, and the FRET efficiency E they exhibited (Figure 15b). All molecules with a stoichiometry of 1 and a FRET efficiency of 0 are thus D-only and they can be removed from the raw data.
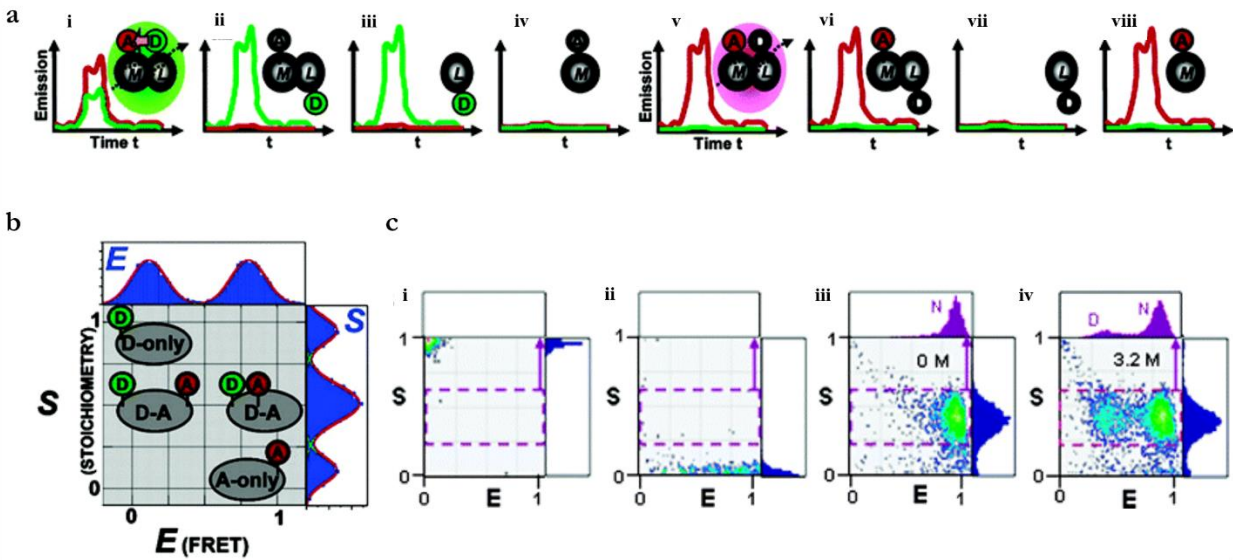
Figure 15: Stoichiometry plots from pulsed interleaved excitation (PIE) experiments. (a) Due to labeling imperfections, photobleaching and blinking several species of labeled molecules can be found in labeled samples for FRET. They can be distinguished due to their different behavior when they are excited at the donor (D) (i-iv) and acceptor (A) (v-viii) wavelength. (i) Molecules that have a donor and acceptor label in the right distance for FRET to occur will exhibit fluorescence for the donor and acceptor molecule when excited at the donor wavelength. (v) The same molecule will show fluorescence for the acceptor if it is excited at the acceptor wavelength. In contrast, a molecule with a donor and an acceptor, but where the acceptor is too far away for FRET, will only show (ii) donor fluorescence when excited at the donor wavelength and (vi) acceptor fluorescence when excited at the acceptor wavelength. Molecules that carry only a donor molecule and no active acceptor due to labeling imperfections, photobleaching or blinking (D-only) will show (iii) donor fluorescence when excited at the donor wavelength but (vii) no fluorescence when excited at the acceptor wavelength. (iv-viii) The inverse is the case for molecules with an active acceptor but no active donor. (b) 2D stoichiometry (S)-FRET efficiency (E) histogram of differently labeled species. These histograms allow us to distinguish between D-only and low FRET species and acceptor-

only and high FRET species. (c) Exemplary S-E histograms of (i) D-only species, (ii) acceptor-only species as well as (iii) high FRET and (iv) a mix from high FRET and low FRET species. Adapted with permission from [148]. Copyright 2020 American Chemical Society.

3.3.2.4. FRET analysis using recurrence analysis of single particles

As mentioned above, at the very low concentrations commonly used in diffusion experiments, it is statistically much more likely for the molecule, having just left the confocal spot, to return to it, rather than a new molecule entering [163]. In consequence, bursts which are recorded shortly after one another are likely originating from the same molecule [163]. Using RASP, it is then possible to determine if the FRET efficiency changed between those bursts from the same molecule, opening the possibility of investigating even faster dynamics from 50 μs to 100 ms [163].

To perform the first step of RASP, all bursts with one FRET efficiency, e.g. the folded population, are picked from the collected data set. From those initial bursts the initial histogram is constructed. Then all bursts that were recorded at a specific time after each initial burst are selected, this is called the recurrence interval. From these bursts, the recurrence histogram is constructed. This process can be repeated while the recurrence interval is increased (Figure 16). From each recurrence histogram the fraction of molecules per population in the histogram, e.g. folded versus unfolded, is calculated and compared to the ratio that would have been expected from a non-dynamic system [163]. The variations then allow to determine the conversion kinetics between different FRET populations.
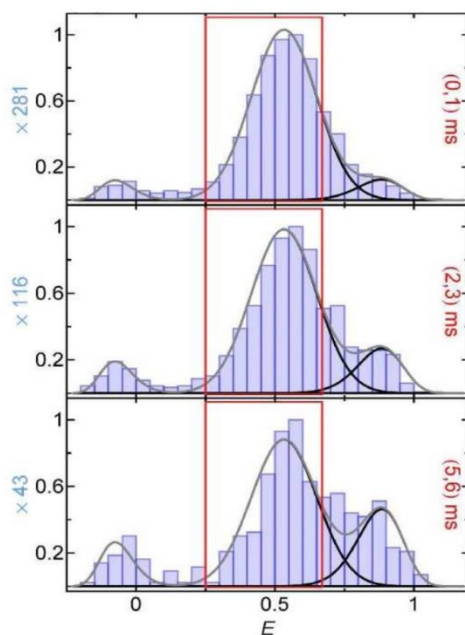
Figure 16: Example of recurrence analysis of single particles (RASP). Exemplary recurrence histograms at different recurrence times. The FRET population around FRET efficiency 0 corresponds to molecules with an inactive acceptor, the population around FRET efficiency 0.5 corresponds to the unfolded state and the FRET efficiency around 0.9 corresponds to the folded state. Initial bursts were taken with FRET efficiencies indicated by the red box. Recurrence time intervals are indicated by the red numbers on the right. Figure from [163].

4. Aim of this work

Although membrane proteins constitute a large portion of both the proteome, 20-30% of all sequenced genomes[127], and of possible drug targets, various aspects about them remain unclear. Here, MFS transporters from two families, the POT and ACS family, are studied to obtain insight into the transport mechanism as well as substrate recognition of these proteins.

POTs are important for the uptake of di- and tripeptides but some members were also found to be able to transport various drugs and prodrugs. Recently published structures of several different POTs with ligands shed light on how these transporters acquire their broad substrate promiscuity. Nevertheless, only a subset of possible conformations that the transporters have to adopt for a full transport cycle are represented by these published structures. Moreover, these structures only provide a static view of the dynamic transport process.

Therefore, in the first part of this work the dynamics of the transport cycle of POTs is studied, using a POT from *E. coli* (DtpA) as a model system. Here, the following questions are addressed:

- Does the transporter adopt different conformational states when in detergent solution?
- Are the conformational states the transporter is able to adopt influenced by its environment e.g. detergent solution or lipid bilayer?
- What are the timescales of interconversion between different conformational states?

In contrast to the broad substrate range commonly found among POTs, especially the *E. coli* members of the ACS family are very substrate-specific. Often accepting only one stereoisomer of a single substrate. Therefore, the following questions are addressed in the second part of the project:

- What are the substrates of each transporter?
- How is the substrate-specificity achieved?
- How is the substrate coordinated in the binding site?

5. Results and discussion

5.1. Structural dynamics of POTs monitored by smFRET

POTs are pharmacologically interesting targets for drug delivery. However, an understanding of their transport activity in atomic detail is lacking. Current high-resolution structures of POTs are mostly crystallized in two conformational states, inward-open and occluded, while missing an outward-open state (Table 2). Moreover, these structures are static snapshots and do not reveal dynamic information on the transport cycle. However, detailed knowledge of the complete transport cycle is required for rational drug design.

To gain insight into the substrate translocation, the transport cycle of POTs is studied here by smFRET. This technique allows to follow dynamic processes by reporting the relative distance of two fluorophores with respect to one another [146]. In the following study, smFRET was used to study a transmembrane protein from the POT family. Members of the POT family exhibit the typical MFS fold. Due to the nature of their function, transporting molecules across a membrane, POTs are supposed to be highly dynamic proteins [58]. Based on the alternative access mechanism and available structures of MFS transporters, at least three conformational states are required for a full transport cycle, the inward-open, occluded and outward-open state [9,34]. The conformational changes between those states are described as rocker-switch motions [165]. The two extreme conformational states of this transport cycle are the inward-open, where the substrate binding cavity is accessible from the cytoplasm, and the outward-open state, where the substrate binding cavity is accessible from the extracellular environment (the periplasm in the case of gram-negative bacteria or lumen for endosomal and lysosomal transporters). When the transporter is labeled with a fluorophore in each domain, the different conformational states can be distinguished by smFRET. In each conformational state, the fluorophores are at a certain distance leading to specific FRET efficiencies which can be assigned to distinguished states. The number of FRET populations thus indicate the presence of different conformational states. In addition, the interconversion of those FRET population gives information about the timescales associated with switching between the different conformational states.

SmFRET was already successfully used to study conformational changes in other membrane proteins such as ABC transporters, G protein-coupled receptors (GPCRs), ion channels or neurotransmitter transporters [154,166–169]. Here, it is extended to POTs. The workflow can be roughly summarized as follows:

i. Selection of a model system
ii. Selection of a fluorophore pair
iii. Generation of mutants that can be site-specifically labeled with a fluorophore (single mutants)
iv. Combination of these single mutations to generate mutants that can be labeled at two positions (FRET mutants)
v. SmFRET measurements of FRET mutants in detergent solution and lipidic environment

5.1.1. Sample preparation for smFRET experiments

In this study, DtpA, a POT from *E. coli*, was chosen as model protein. There are several reasons for studying DtpA as a model system. Firstly, although it is a bacterial protein, DtpA transports a similar set of di- and tripeptides as well as a subset of peptidomimetic drugs compared to its human homologue PepT1 [53]. Therefore, DtpA is considered a suitable proxy for studies of the transport mechanism and the conformational variety of POTs.

Secondly, DtpA was shown to be a monomer in detergent solution as well as in the membrane-like environment of SapNPs [50,53]. As described previously, smFRET is measured at picomolar concentrations. At this concentration it is assumed that the FRET signal originates from two fluorophores of the same protein. In the case of oligomeric proteins, the signal could also result from proteins in close proximity to each other. Another advantage of DtpA is that protocols for expression and purification were established and most importantly the structure of the inward-open state of DtpA has been published (Figure 17a) [50]. The crystal structure allows us to measure the distances between fluorophore pairs at various positions and calculate the expected FRET efficiency of those fluorophore pairs for the inward-open conformation. Thus, it is possible to assign FRET populations to the different conformational states of DtpA. Moreover, in the crystal structure DtpA is in complex with a nanobody [50]. The nanobody (N00) is a conformational binder [50]. It binds to the periplasmic side of the protein and stabilizes the inward-open conformation. By adding N00 to DtpA, the periplasmic side of the protein will be closed and the conformational variety might be shifted to the inward-open state.
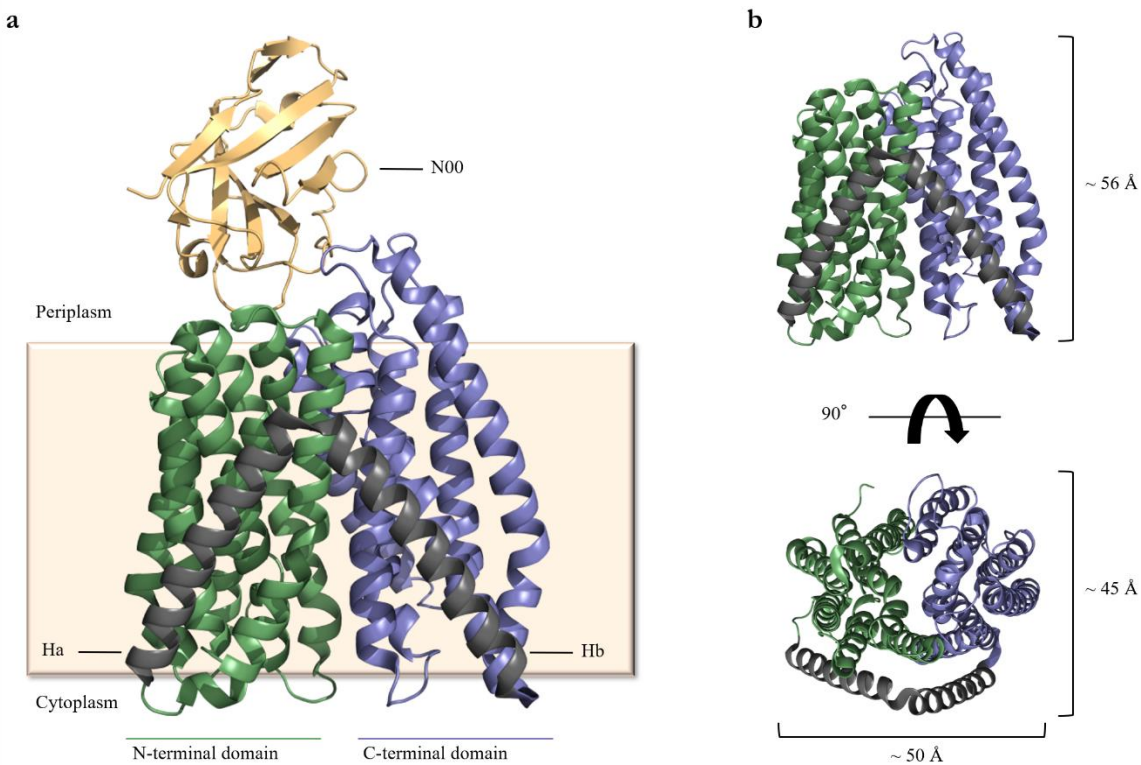
Figure 17: Structure of DtpA. (a) Crystal structure of the DtpA-N00 complex (PDB-ID 6GS7). DtpA is presented in the inward-open state with the nanobody (N00) bound on the periplasmic side. The N-terminal domain of DtpA is shown in green, the C-terminal domain in blue and the HaHb domain in gray. N00 is colored yellow. (b) Dimensions of DtpA.

After a suitable model system for smFRET was found, compatible fluorophores needed to be selected. As described previously, in FRET an excited fluorophore (donor) transfers its energy onto another fluorophore (acceptor) instead of emitting light. The acceptor in turn emits light of a different wavelength. For this to happen, the emission spectrum of the

donor and the absorption spectrum of the acceptor have to overlap. Additionally, for a maximal change of FRET efficiency in respect to the distance of the fluorophores, the Förster-distance should be within the expected range of movement.

Here, the aim is to observe DtpA switching between different conformational states, namely the inward-open state, occluded state and outward-open state. By placing one fluorophore in each of the two domains, the conformational changes can be followed using smFRET. The fluorophore pair with an appropriate Förster-distance depends on the dimensions of DtpA, meaning the distance between residues in the N-terminal and C-terminal domain. As the width and height of DtpA are roughly 50 Å (Figure 17b), so should the Förster-distance of the chosen fluorophore pair be. A fluorophore pair fulfilling these criteria is Alexa Fluor 488 and Alexa Fluor 594 with a Förster-distance of 54 Å and sufficient spectral overlap (Figure 18) [170].
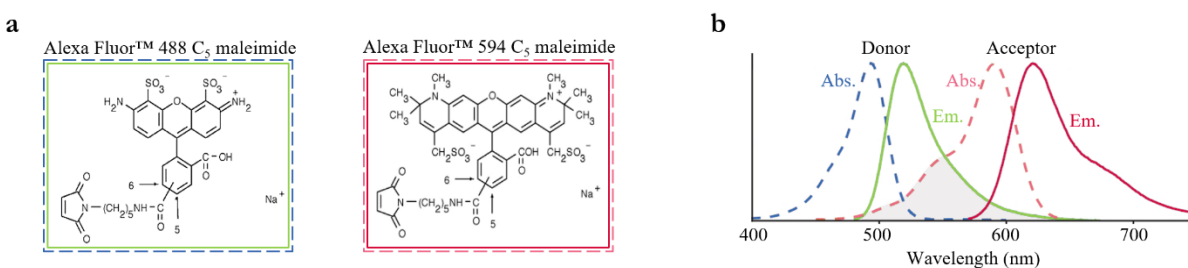
Figure 18: Alexa fluorophores used for smFRET measurements. (a) Chemical structure of Alexa Fluor™ 488 C$_5$ maleimide and Alexa Fluor™ 594 C$_5$ maleimide. Figure from [171,172]. (b) Excitation (Abs.) and emission (Em.) spectra of Alexa Fluor™ 488 C$_5$ maleimide (Donor) and Alexa Fluor™ 594 C$_5$ maleimide (Acceptor). Figure adapted from [173]. The gray area shows the overlap for the emission spectra of Alexa Fluor™ 488 C$_5$ maleimide and the excitation spectra for Alexa Fluor™ 594 C$_5$ maleimide. This overlap is a requirement for two fluorophores to be used for FRET measurements.

5.1.1.1. Generation of mutants for labeling with fluorophores

Fluorophores were attached to the protein via thiol-maleimide click chemistry [157]. Here, the thiol group is found in the cysteine residues of the protein while the maleimide group is supplied as a functional group coupled to the fluorophores. For the labeling strategy to be site-specific, all intrinsic cysteine residues have to be removed from the protein. Wildtype DtpA has three cysteine residues, C140, C200 and C360. All were replaced by serine residues, thereby creating a cysteine free (Cys-free) mutant. This Cys-free mutant served as background for introducing new cysteine residues at the desired labeling positions.

In the first step, one new cysteine residue was introduced at the desired labeling position. The generated mutants are called single mutants as they only have one single cysteine residue in their sequence. After a short quality control ensuring that they displayed wildtype-like stability and functionality, single mutants that passed the quality control experiments were further mutated into double mutants. These mutants are termed FRET mutant hereafter. FRET mutants were to pass the same quality control experiments as single mutants before they were measured with smFRET.

5.1.1.1.1. Generation of single cysteine mutants for labeling with fluorophores

Several points have to be considered when choosing fluorophore positions. (i) The position should be at the solvent exposed surface of the protein. For labeling, the side chain of the cysteine residue needs to be accessible to the fluorophore. This is not the case if the amino acid is, for example, buried inside the protein. Similarly, the transmembrane part of the protein would be inaccessible for fluorophores as it might be shielded either by the detergent micelle or lipids. (ii) The positions should be at the end of transmembrane helices and not inside loop regions. Loops might be flexible on their own without the conformation of the protein changing. In this case, different FRET populations could originate from movement of the loops and not by conformational switching of the protein itself. (iii) Close proximity to tryptophan residues should be avoided as they would quench the FRET signal [174]. Quenching is a term used for processes that reduce the intensity of a fluorophore. As the fluorescence intensity is measured during the FRET experiment, quenching is undesirable. For both used fluorophores, Alexa Fluor 488 and Alexa Fluor 594, tryptophan was found to quench their fluorescence intensity [174]. (iv) In case of labeling via cysteine residues, it is recommended to use labeling positions that are not in close proximity of glutamate or aspartate residues as this would reduce the labeling efficiency. The negatively charged amino acids increase the pKa of the thiol group of the cysteine when aspartate or glutamate are placed adjacent to a cysteine residue. This increase in pKa in turn leads to slower kinetics of the thiol-maleimide click reaction [175].

Based on these criteria, 13 possible labeling positions were identified (Figure 19 and Table 4). All listed point-mutations were introduced via site-directed mutagenesis with a mutagen primer pair and verified by sequencing.
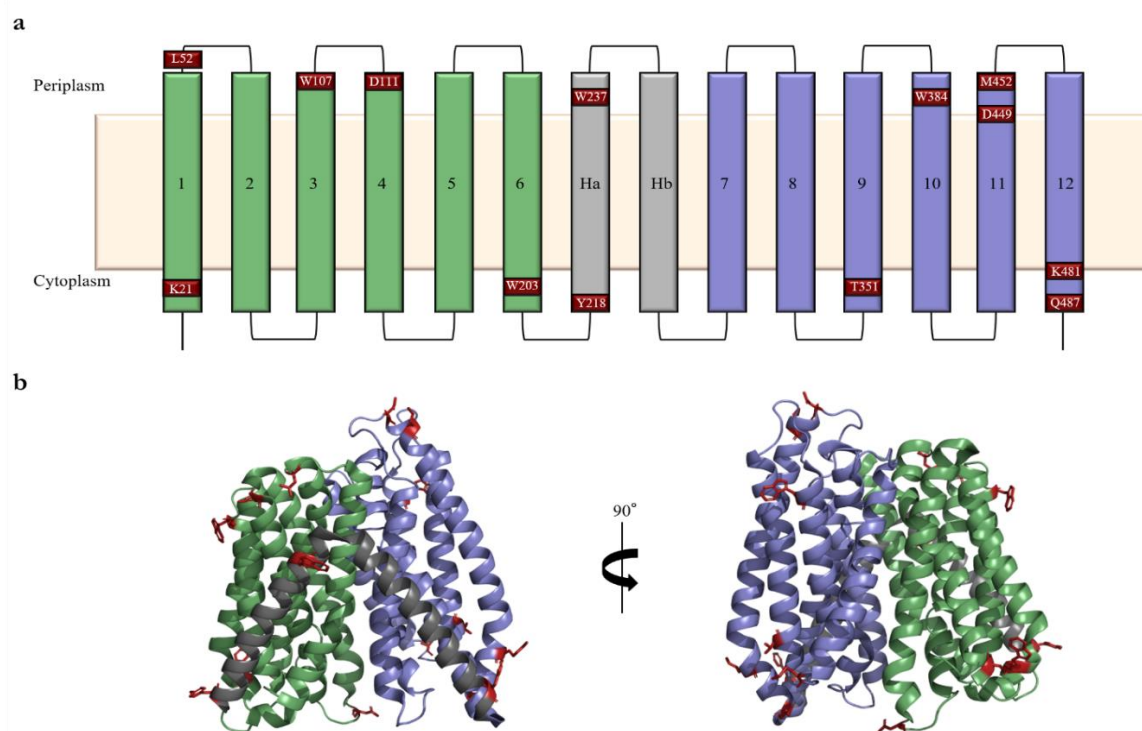
Figure 19: Overview of labeling positions on DtpA from _E. coli_. (a) Schematic overview of the labeling positions with respect to their position on the TMs. N-terminal domain TMs depicted in green, C-terminal domain helices in blue and the HaHb domain in gray.

The positions of amino acids chosen for labeling are shown as red boxes. (b) Labeling positions shown on the crystal structure of DtpA. Same color code as above, amino acids chosen for labeling are shown as red sticks.

Table 4: Point mutations introduced in DtpA for FRET labeling. + means this experiment was successful, - means this experiment was not successful, n.d. means that this experiment was not conducted for the mutant. Successful experiments meant that the mutant could be cloned, expressed, purified, it bound ligands and the cysteine residue was accessible. When this was not the case the experiments were deemed not successful.

Mutation	Domain	Cloned	Expressed	Purified	Ligand binding	Cysteine accessibility
C140S	N-terminal	+	+	+	+	+
C200S	N-terminal	+	+	+	+	+
C360S	C-terminal	+	+	+	+	+
K21C	N-terminal	+	+	+	+	+
L52C	N-terminal	+	+	+	+	+
W107C	N-terminal	+	+	+	+	-
D111C	N-terminal	+	+	+	+	+
W203C	N-terminal	+	+	+	+	+
Y218C	HaHb	+	+	+	-	n.d.
W237C	HaHb	+	+	+	+	-
T351C	C-terminal	+	+	+	+	+
W384C	C-terminal	+	+	+	+	+
D449C	C-terminal	+	+	+	+	+
M452C	C-terminal	+	+	-	n.d.	n.d.
K481C	C-terminal	+	+	+	+	+
Q487C	C-terminal	+	+	+	+	+

5.1.1.2. Purification of proteins for smFRET

As mentioned above, an established protocol for overexpression and purification of DtpA was available [50]. However, the purification protocol had to be adapted. (i) The choice of detergent had to be optimized, (ii) an immobilized metal affinity chromatography (IMAC) step to remove TEV protease (negative IMAC) was added.

In general, point-mutants of DtpA were less thermally stable than the wildtype protein. To compensate for this, the detergent used for purification was switched from DDM to LMNG, a maltose-neopentyl glycol. Maltose-neopentyl glycol detergents often lead to an enhanced structural stability of membrane proteins [176]. This family of detergents is commonly used for extraction of notoriously unstable membrane proteins from their native lipid bilayer as well as for structure determination of those proteins using X-ray crystallography or cryo-EM [131,176]. In LMNG the FRET mutants were sufficiently stable for purification, labeling and smFRET measurements.

Moreover, the original purification protocol did not include a negative IMAC step to remove the His-tagged TEV protease after cleavage of the His-tag from DtpA [50]. This is problematic because TEV protease itself contains several cysteine

residues and thus might be labeled with fluorophores as well [177]. If TEV was not efficiently removed, samples would contain two labeled proteins, TEV and DtpA. In the FRET histograms it would therefore not be possible to decide which signal originates from DtpA and which corresponds to contamination from TEV protease. The reason why it was possible that TEV protease could still be present after purification was the DtpA construct used. Originally, DtpA was overexpressed in a pTH24 vector (pTH) with an N-terminal and a C-terminal His-tag [50]. Of those two His-tags, only the N-terminal was cleavable by TEV protease (Figure 20a). Therefore, it was not possible to perform a negative IMAC step in order to remove TEV protease completely. To not contaminate the FRET measurements, the DtpA variants were re-cloned in a pNIC-CTHF vector (pNIC) [178]. This vector has a C-terminal His-tag that can be cleaved off (Figure 20b). This allows addition of a negative IMAC step to the purification protocol, thus enabling removal of the TEV protease (Figure 20c).
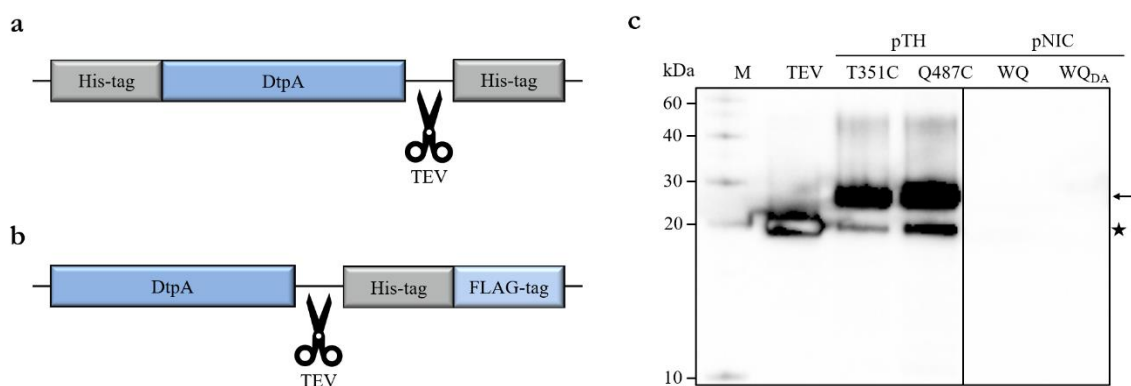
Figure 20: Construct design for DtpA FRET mutants. (a) Schematic of the pTH24 (pTH) construct for FRET mutants. Here, the DtpA variant is expressed with both a N-terminal and a C-terminal His-tag. Only the C-terminal His-tag is cleavable by TEV protease, resulting in a protein sample with one His-tag after purification. For this construct no negative IMAC step was done during purification. (b) Schematic of the pNIC-CTHF (pNIC) construct for FRET mutants. Here, the DtpA variant is expressed with C-terminal His-tag and FLAG-tag. Both tags are cleavable by TEV protease, resulting in a protein sample without His-tag after purification. For this construct a negative IMAC step was done during purification. (c) Western blot analysis to detect TEV contamination in the final purified protein sample. The membrane was stained with anti-His-tag antibodies. In both mutants expressed as pTH construct (T351C and Q487C), the transporter band is observed at 35 kDa, denoted by an arrow. As this construct has a non-cleavable His-tag it is visible on the Western blot in contrast to the mutants from the pNIC construct (unlabeled protein WQ and labeled protein WQ$_{DA}$) where the His-tag is cleaved during purification. Due to the lack of a negative IMAC step for pTH constructs, there is still TEV protease present in the samples after purification. Here seen as band at 20 kDa, denoted by a star. For the pNIC samples TEV is removed due to the addition of a negative IMAC step to the purification protocol.

With this optimized purification protocol, the overexpression and purification of 12 out of the 13 initially designed single mutants was successful (Figure 50 Appendix). Only mutant M452C precipitated heavily during purification and was therefore excluded from further analysis. The remaining 12 single mutants were subjected to a series of quality control experiments.

5.1.1.3. Quality control of single cysteine mutants

In order to label DtpA for FRET measurements several point-mutations were introduced in the protein sequence. The positions of these point mutations were chosen such that they would allow for a maximum sensitivity of change in FRET efficiency. Since the chosen mutation sites might be important for the structural integrity or functionality of the protein, it was important to ensured that the mutants used for smFRET measurements still resemble the wildtype protein as close as possible. For this, quality control experiments were conducted for every mutant. The quality control experiments were performed to test the single mutants for their thermal stability and ability to bind ligands (thermal shift assay) and the accessibility of the introduced cysteine residues for labeling (PEGylation assay).

5.1.1.3.1. *In vitro* ligand binding assay of single cysteine mutants

After the successful purification of 12 single mutants, their thermal stability and ability to bind different ligands was tested. The ability to bind ligands was used as read-out for the retained functionality of the protein. For testing binding, a small subset of three ligands, two dipeptides and a tripeptide, were chosen. These ligands were already shown to bind and stabilize wildtype DtpA [50]. In addition, as smFRET is measured at room temperature (RT), the thermal stability had to be sufficiently high for the protein to not denature.

For the *in vitro* ligand binding assay, nano differential scanning fluorometry (nanoDSF) was used. Here, the fluorescence intensity of tryptophan residues of the protein is measured while the protein is being heated [179]. Upon heating, the protein unfolds and the chemical environment of the tryptophan residues changes, which in turn alters their fluorescence. Thus, by recording the ratio of the fluorescence intensities at 330 nm and 350 nm one can follow the protein's thermal unfolding. The inflection point of this unfolding curve is defined as the melting temperature (Tm). To better visualize the Tm, the first derivative of the unfolding curve is plotted against the temperature. In this case, the peak of the curve corresponds to the Tm of the protein. If ligands are interacting with the transporter, it might result in the stabilization of the protein and in turn heat denaturation might occur at higher temperatures (Figure 21a).

This method has several advantages: it requires only little material (10 μL of protein solution per tested ligand), it can be performed at relatively low concentrations (9 μM in the case of wildtype DtpA and DtpA mutants) and there is no need for labeling compared to traditional thermal shift assays.

In general, single mutants are less stable then the wildtype which is not surprising, as it is expected that the point-mutations introduced in the protein might have a destabilizing effect. In DDM, the Tm of the tested mutants ranged from 35 °C for W384C to 41 °C for D111C compared to 48 °C for wildtype DtpA (Figure 21b). After purification in LMNG, instead of using DDM, the screened single mutants were sufficiently stable to withstand labeling and FRET measurements. In LMNG, the thermal stability increased for wildtype DtpA to 57 °C. The tested single mutants showed increased thermal stability as well, with the Tm's ranging from 44 °C for T351C to 53 °C for K481C (Figure 21c).

In addition, all single mutants still bound the tested set of ligands. On average, in DDM, addition of the dipeptide Ala-Leu (AL) increased the Tm by 7 °C, addition of Ala-Phe (AF) by 11 °C, and of the tripeptide Ala-Phe-Ala (AFA) by 12 °C. (Figure 21b). The same trend was observed in LMNG. Here, addition of AL on average increased the Tm by 4 °C, addition of AF by 6 °C, and addition of AFA by 7 °C (Figure 21c). An exception was the mutant Y218C. Here, no Tm could be

detected. Furthermore, addition of ligand did not alter the unfolding curve, meaning no binding could be shown. Thus, no FRET mutants based on this position were generated for future experiments (Figure 21c).
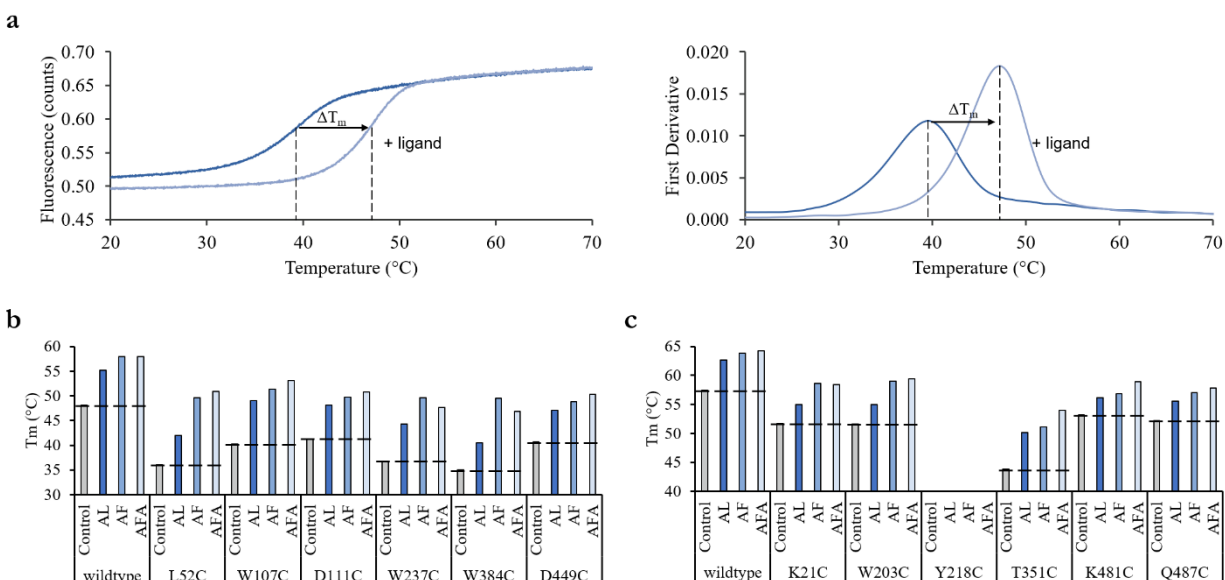
Figure 21: Thermal stability and ligand binding of single mutants and FRET mutants of DtpA. (a) Schematic of thermal stability measurements using nanoDSF. The fluorescence of tryptophan and tyrosine residues is recorded at 350 nm and 330 nm respectively while the samples are heated from 20 °C to 70 °C. The ratio of the fluorescence at 350 nm and 330 nm is plotted against the temperature. As the fluorescence of tryptophan and tyrosine residues depends on the surrounding, a change in surrounding for example due to unfolding of the protein leads to a change in fluorescence. The inflection point of this unfolding curve is defined as the melting temperature (Tm). Binding of ligands that stabilize the protein leads to a change in Tm. To make the Tm more prominent, the first derivative of the unfolding curve is plotted against the temperature. In this case the peak of the curve corresponds to the Tm of the protein. (b-c) Thermal stability and ligand binding of single mutants in (b) DDM and (c) LMNG. All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. For Y218C no Tm could be determined.

5.1.1.3.2. Cysteine accessibility assay of single cysteine mutants

By this point, eleven single mutants were shown to be functional and sufficiently stable for purification and labeling. Next, the accessibility of the cysteine residues for reaction with maleimide coupled fluorophores, was tested. Although care was taken to select labeling positions where the side chains point towards the surface of the protein, it could still be possible that the position is shielded by the detergent micelle and thus not accessible. To test if this is the case, the detergent-solubilized single mutants were incubated with maleimide-coupled polyethylene glycol (PEG). Here, the same maleimide-thiol reaction takes place as it would for the FRET fluorophores. Instead of attaching the fluorophore to the transporter, a PEG-molecule is added. This increases the molecular size of the protein and leads to a change in migration on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 22a-b).

For FRET measurements, it is important to have a well labeled sample. Proteins without an active acceptor will not exhibit FRET. At the same time, proteins which carry only a donor molecule will present as a population at a FRET efficiency of

0. Both cases do not add information to the FRET measurement and increase the measurement time as more data need to be acquired to observe a sufficient number of donor-acceptor labeled molecules.

The tested cysteine positions differed greatly in accessibility to PEGylation (Figure 22c). While some mutants, such as D111C, W203C, W384C, D449C or Q487C, can be almost completely PEGylated, others, such as K21C, L52C and T351C, are less accessible. Nevertheless, all these mutants show sufficient labeling to proceed with them. In the case of K481C, only a very small fraction of the protein can be PEGylated. But the mutant was not excluded from the project at this point as it was still possible to access at least a fraction of the cysteine residues. In contrast, W107C and W237C did not show any PEGylation. It was therefore concluded that those cysteine residues are not accessible and work on both mutants was not continued (Figure 22c).
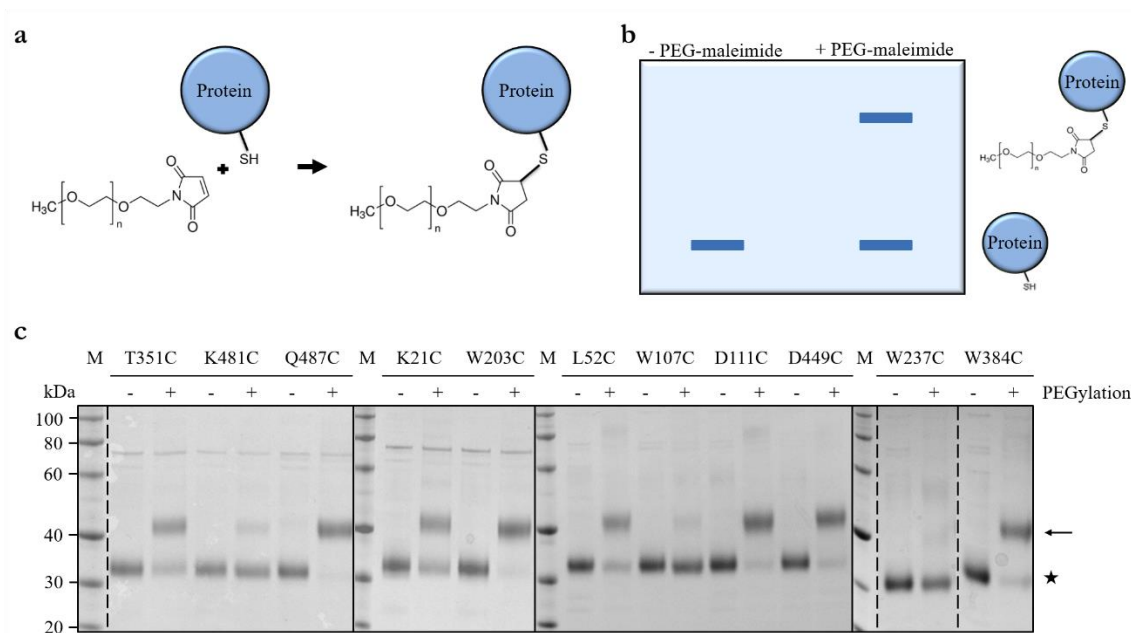
Figure 22: Testing the accessibility of introduced cysteine residues for labeling with FRET fluorophores. (a) Schematic of the thiol-maleimide reaction between the cysteine group of a protein and the maleimide group attached to a PEG molecule. This forms a covalent bond and attaches the PEG molecule to the protein. (b) Schematic of a gel shift assay to judge the PEGylation of a protein. If a PEG molecule is covalently attached to the protein this will increase its molecular weight and cause a band shift on SDS-PAGE to higher molecular weight. (c) Accessibility of cysteine residues in single mutants of DtpA. Mutants were incubated with PEG-maleimide (+) and run on SDS-PAGE to test if a band shift appeared. DtpA mutants run at 35 kDa (black star), addition of one PEG-molecule shifts the band to 45 kDa (black arrow). As a control, sample not treated with PEG (-) was also run.

5.1.1.4.   Combination of single cysteine mutants to generate FRET mutants for labeling with fluorophores

The nine single mutants that passed the quality control were K21C, L52C, D111C, W203C, T351C, W384C, D449C, K481C and Q487C. Those single mutants were then combined to generate double mutants which carry two cysteine residues for labeling. These FRET mutants were sorted into different groups according to the positions of the cysteine residues in respect to each other (Figure 23). Mutants with both labeling positions on the cytoplasmic side but in different

domains are called cytoplasmic mutants, the corresponding mutants labeled on the periplasmic side are called periplasmic mutants. If one labeling position is on the cytoplasmic side and one on the periplasmic side of either the N- or C-terminal domain, they are called N- and C-terminal mutants, respectively. If the labels are opposite to each other in regard to the membrane and in different domains, they are termed across mutants.
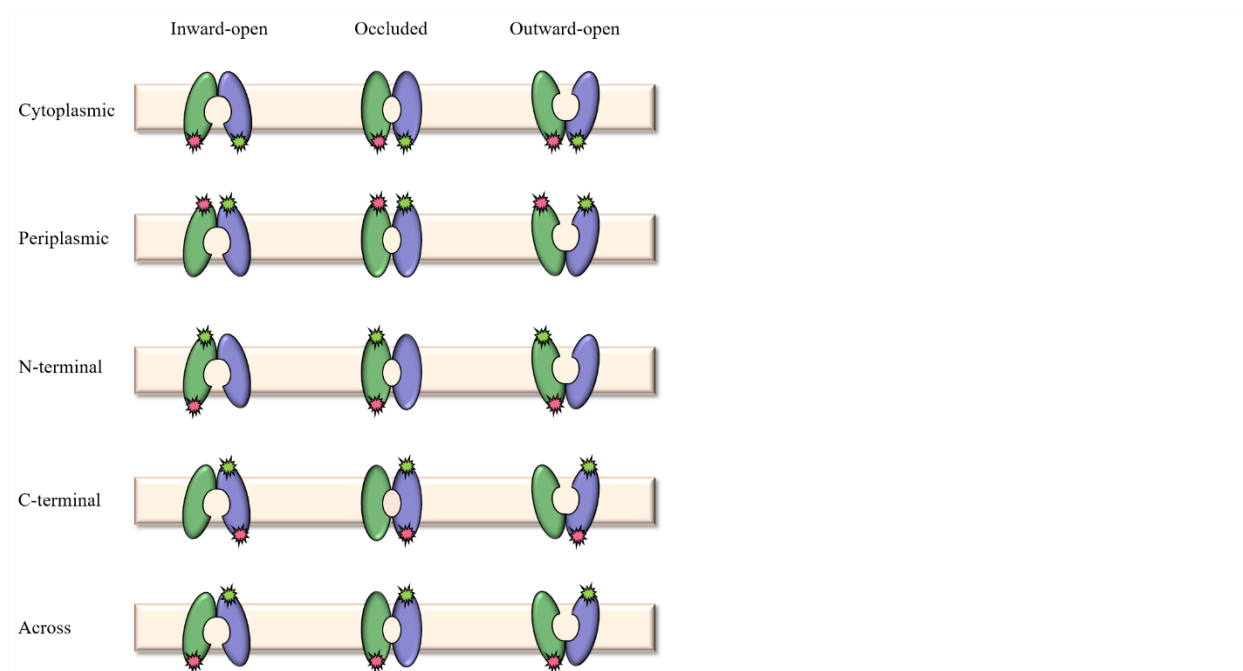
Figure 23: Overview of different positions of the FRET dye pair. FRET mutants are sorted into different groups according to the positions of the FRET dyes towards each other in the labeled protein. In cytoplasmic mutants, both FRET dyes are on the cytoplasmic side of the protein, with one dye in the N-terminal and one in the C-terminal domain. For periplasmic mutants they are both on the periplasmic side. For N-terminal mutants, both dye positions are in the N-terminal domain but one is on the cytoplasmic and one is on the periplasmic side of the protein. The equivalent mutants but with FRET dyes in the C-terminal domain are called C-terminal mutants. If both dyes are on either side of the membrane in opposite domains, the mutants are called across mutants. For each case, the position of the dyes towards each other is depicted for the protein in the inward-open, occluded or outward-open state. The N-terminal domain is depicted in green and the C-terminal domain in blue. The HaHb domain in omitted. The FRET dyes are shown as stars.

The expected FRET efficiencies for different labeling positions were calculated based on the available crystal structure of DtpA [50] (PDB-ID 6GS7) (Table 5). To simulate the overall flexibility of the Alexa dyes, the dyes were modeled as a bulky sphere for the fluorophore moiety and a polyethylene chain of eleven monomers for the linker. The linkers were attached to the Cβ atom of the amino acids chosen as labeling position (Figure 24).

a
Alexa Fluor 488 C₅ Maleimide

b
RIS Linker-Dye Model

$r = 3.5$ Å
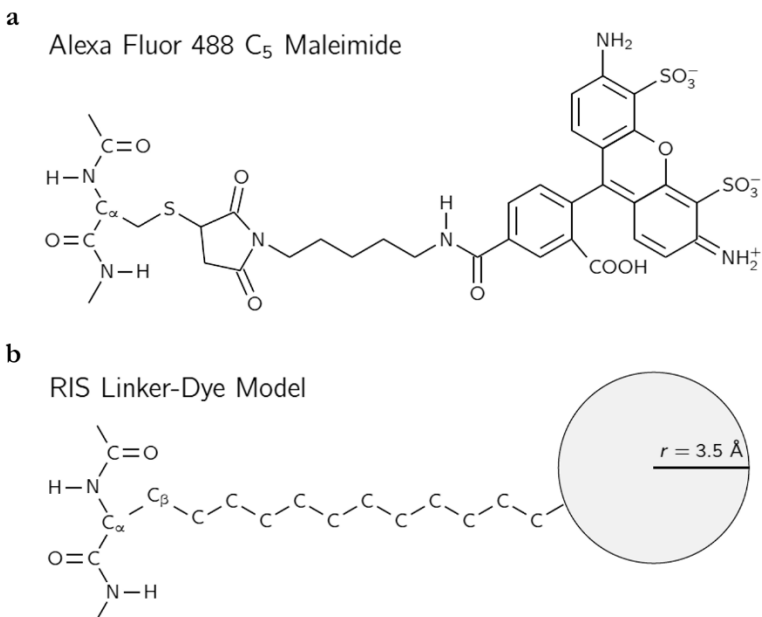
Figure 24: Rotational isomeric state model for Alexa Fluor™ C₅ maleimide dyes. (a) Structural model of Alexa Fluor™ 488 C₅ maleimide (b) Rotational Isomeric State (RIS) model for the Alexa Fluor™ C5 maleimide dyes shown above. The maleimide moiety was kept but the linker from the Alexa Fluor molecule was replaced by a polyethylene chain and the fluorophore moiety was modeled as a sphere with a radius of 3.5 Å. Figure curtesy of Jakub Jungwirth.

Then, possible dye conformations were simulated by rotating each bond of the polyethylene linker to form the anti, gauche- and gauche+ rotamers. Rotamers that clashed with the protein structure were discarded. The FRET efficiency was calculated for all possible pairs of the remaining rotamers. These calculations were performed by Jakub Jungwirth from the Weizmann Institute of Science.

Table 5: FRET mutants of DtpA generated for smFRET measurements. DtpA mutants were generated that could be labeled at two positions to measure smFRET. The expected FRET efficiency for each mutant when in the inward-open state was calculated using the available crystal structure of DtpA. The calculations were done in the presence and absence of the nanobody N00.

Mutant	Label positions	Cα-Cα distance (inward-open state)	Expected FRET (inward-open state)	Expected FRET (inward-open state N00 bound)
K21C/T351C	Cytoplasm	37.1 Å	0.52	0.50
K21C/K481C	Cytoplasm	44.3 Å	0.25	0.24
K21C/Q487C	Cytoplasm	39.3 Å	0.63	0.62
L52C/T351C	Across	44.3 Å	0.33	0.33
L52C/W384C	Periplasm	31.3 Å	0.54	0.53
L52C/D449C	Periplasm	22.5 Å	0.88	0.89
D111C/W203C	N-terminal	42.0 Å	0.56	0.56
D111C/T351C	Across	48.2 Å	0.19	0.19
D111C/W384C	Periplasm	38.2 Å	0.34	0.34
D111C/D449C	Periplasm	31.1 Å	0.64	0.65
W203C/T351C	Cytoplasm	40.5 Å	0.47	0.47
W203C/K481C	Cytoplasm	47.0 Å	0.22	0.22
W203C/Q487C	Cytoplasm	41.5 Å	0.54	0.55
T351C/W384C	C-terminal	26.7 Å	0.92	0.92
D449C/Q487C	C-terminal	55.5 Å	0.17	0.17

5.1.1.4.1.1. Quality control of FRET mutants

Analogue to the single mutants, for the FRET mutants a series of quality control experiments were conducted as well. The quality control experiments tested the mutants for their activity (*in vivo* uptake assay), their thermal stability and ability to bind ligands (thermal shift assay), their structural integrity (complex formation with a conformational nanobody) and the accessibility of the introduced cysteine residues for labeling (PEGylation assay).

5.1.1.4.1.1.1. *In vivo* uptake assay of FRET mutants

The activity of FRET mutants was tested using an *in vivo* uptake assay as previously described but with minor changes [53]. Here, a fluorescent reporter molecule, β-Ala-Lys-N-7-amino-4-methylcoumarin-3-acetic acid (AK-AMCA), is taken up by DtpA. This is possible due to the promiscuity of the transporter that recognizes the dipeptide β-Ala-Lys as substrate even if it is coupled to the fluorophore AMCA [53]. After incubation with AK-AMCA, the cells are washed with AK-AMCA free buffer, to remove the fluorophore from the buffer surrounding the cells, and the residual fluorescence inside the cells is measured with an excitation wavelength of 350 nm and emission wavelength of 450 nm (Figure 25a).

Several FRET mutants show close to or even higher uptake of AK-AMCA than the wildtype protein. Those mutants were W203C/Q487C, L52C/D449C and D111C/D449C. Also, W203C/T351C, D111C/W203C, T351C/W384C,

D449C/Q487C, L52C/T351C and D111C/T351C still showed 62-92% of the wildtype's uptake (Figure 25b). Those mutants were deemed to have a wildtype-like activity and work on them was continued.

Cytoplasmic FRET mutants based on the K21C mutation were on average less active then those based on the point-mutation W203C. Therefore K21C/T351C, K21C/K481C and K21C/Q487C were excluded from further experiments. The FRET mutant W203C/K481C still showed a high AK-AMCA uptake with 77% compared to the wildtype. But considering that the position K481C was previously shown to be poorly accessible for labeling and that two other cytoplasmic FRET mutants showed even higher uptake, this mutant was also not followed up. The FRET mutants L52C/W384C and D111C/W384C showed the poorest uptake of all tested FRET mutants, 27% and 30% respectively. It was concluded that they may no longer be active.

Up to this point, nine FRET mutants L52C/T351C, L52C/D449C, D111C/W203C, D111C/T351C, D111C/D449C, W203C/T351C, W203C/Q487C, T351C/W384C and D449C/Q487C were still considered for smFRET measurements.
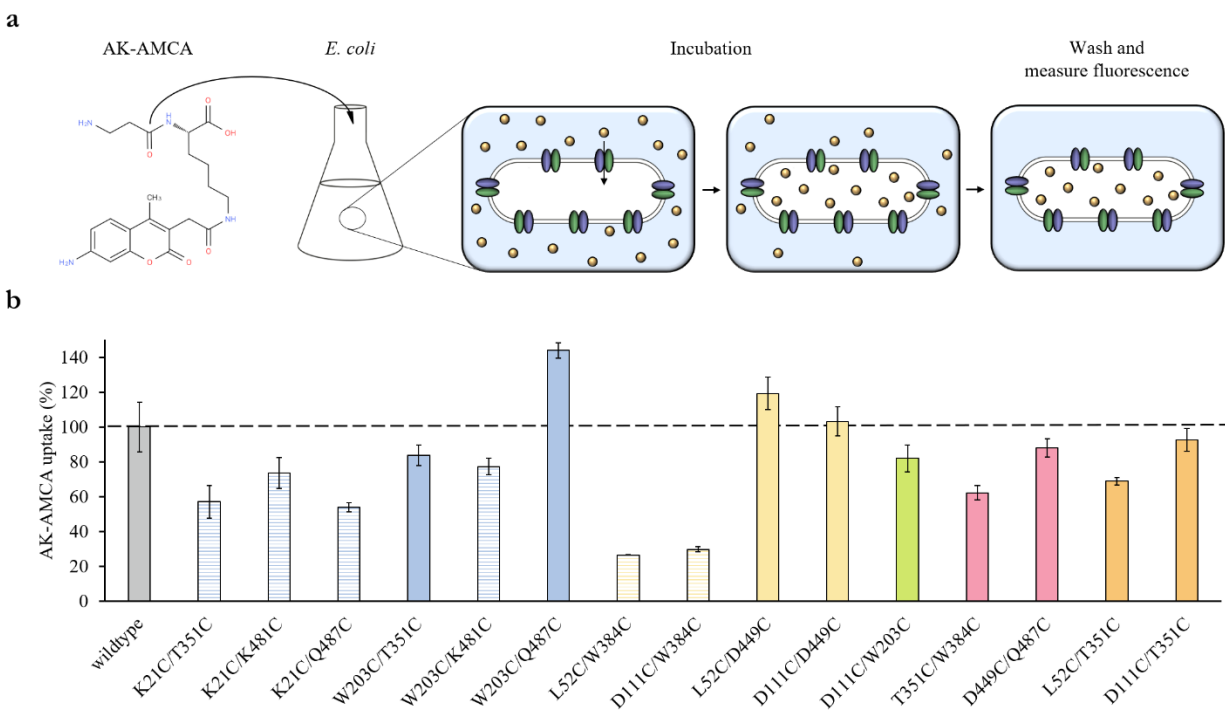
Figure 25: AK-AMCA uptake assay to test the activity of FRET mutants *in vivo*. (a) Schematic of a β-Ala-Lys-N-7-amino-4-methylcoumarin-3-acetic acid (AK-AMCA) *in vivo* uptake assay. Cells expressing the protein are incubated with AK-AMCA. Because the fluorophore AMCA (yellow spheres) is coupled to the dipeptide AK, it can be taken up by DtpA (N-terminal domain green oval, C-terminal domain blue oval). During the incubation time AK-AMCA thus accumulates in the cell. After the reaction is stopped, AK-AMCA is washed off outside the cell and the remaining fluorescence in the cells is measured. (b) Comparison of AK-AMCA uptake of wildtype DtpA to FRET mutants. Wildtype DtpA uptake was set to be 100%. Cytoplasmic mutants are shown in blue, periplasmic mutants are shown in yellow, N-terminal mutants are shown in green, C-terminal mutants are shown in red, across mutants are shown in orange. Mutants that were not used in further experiments are shown with stripes. Measurements were done in triplicates.

5.1.1.4.1.1.2. *In vitro* ligand binding assay of FRET mutants

After establishing that nine FRET mutants were still active, their thermal stability and ability to bind different ligands was tested. The same subset of three ligands as for single mutants was tested.

Similarly to single mutants, FRET mutants were less stable compared to wildtype DtpA. In LMNG, wildtype DtpA exhibits a Tm of 57 °C while the FRET mutants show Tm's between 49 °C for L52C/D449C and 51 °C for W203C/Q487C. But an increase of the Tm for all FRET mutants was observed upon the addition of ligands. Addition of AL peptide increased the Tm on average by 3 °C and addition of AF or AFA by 6 ° C (Figure 26). Therefore, I concluded that all tested FRET mutants were functional.
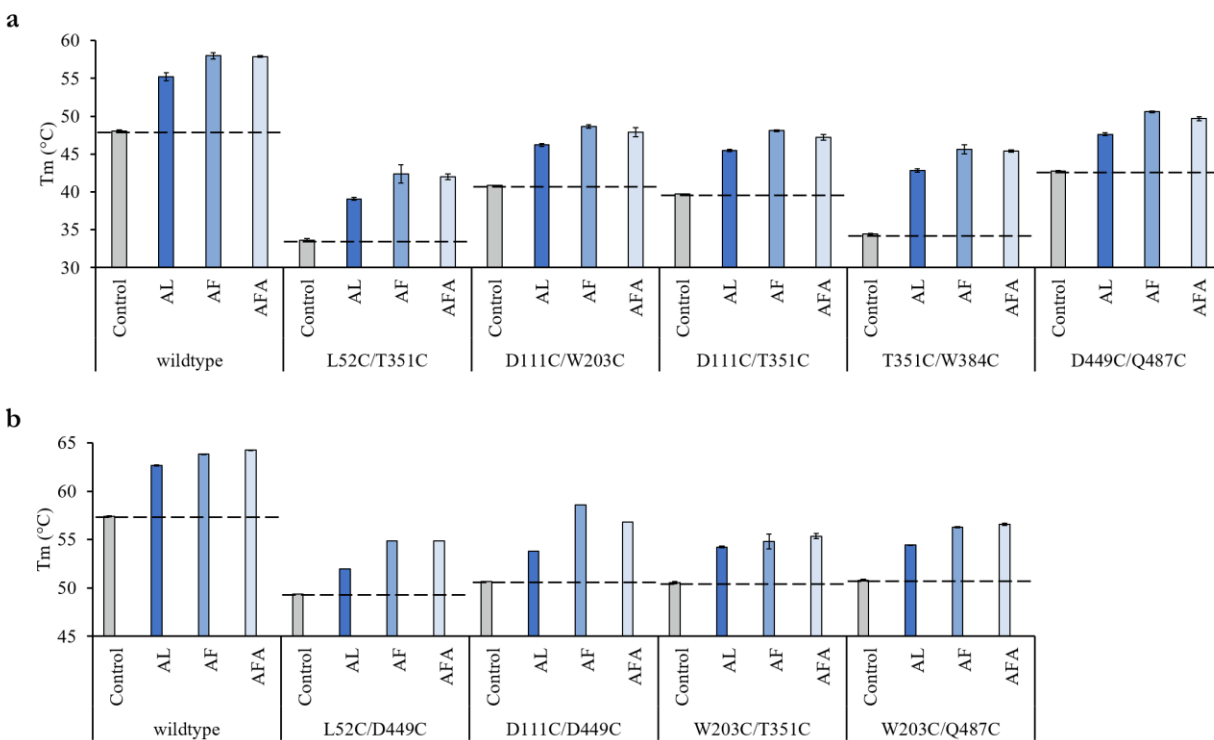
Figure 26: Thermal stability and ligand binding of FRET mutants of DtpA. (a-b) Thermal stability and ligand binding of FRET mutants in (a) DDM. (b) LMNG. All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. All measurements done in triplicates.

5.1.1.4.1.1.3. Cysteine accessibility assay of FRET mutants

As already described for single mutants, for FRET mutants the accessibility of their cysteine residues was tested as well. Again, there are differences between the FRET mutants depending on where the cysteine residues were introduced. Both cysteine residues in W203C/Q487C were fully PEGylated while for all other tested FRET mutants at least half of the sample could be PEGylated at two positions (Figure 27). Therefore, no mutants were excluded at this stage.
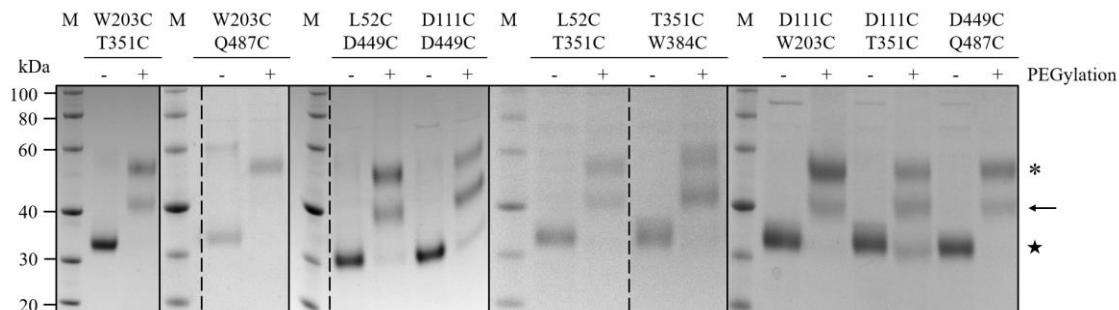
Figure 27: Testing the accessibility of introduced cysteine residues for labeling with FRET fluorophores. Accessibility of cysteine residues in FRET mutants of DtpA. Mutants were incubated with PEG-maleimide (+) and run on SDS-PAGE to test if a band shift appeared. DtpA mutants run at 35 kDa (black star), addition of one PEG-molecule shifts the band to 45 kDa (black arrow), addition of two PEG-molecules to 55 kDa (black asterisk). As a control, sample not treated with PEG (-) was also run.

5.1.1.4.1.1.4. Complex formation of FRET mutants with a conformation-specific nanobody

For FRET mutants that passed the previous steps of quality control, one final experiment was conducted. In the published crystal structure of DtpA, a nanobody was used as crystallization scaffold and was shown to bind to the periplasmic side of the protein [50]. If the nanobody N00, a specific conformational binder, is still able to bind to the FRET mutants, it suggests that their structure is still resembling that of the wildtype at least on the periplasmic side.

To monitor the complex formation, biolayer interferometry was used (Figure 28a). This technique measures the change in optical thickness on a sensor when the protein complex is formed [180]. First, the nanobody is immobilized on a sensor. If the FRET mutants interact with the nanobody, the resulting change in optical thickness is detected (Figure 28b). In addition to qualitatively reporting the complex formation, the dissociation constant (K_D) can also be determined by measuring the association and dissociation of FRET mutants to the nanobody.
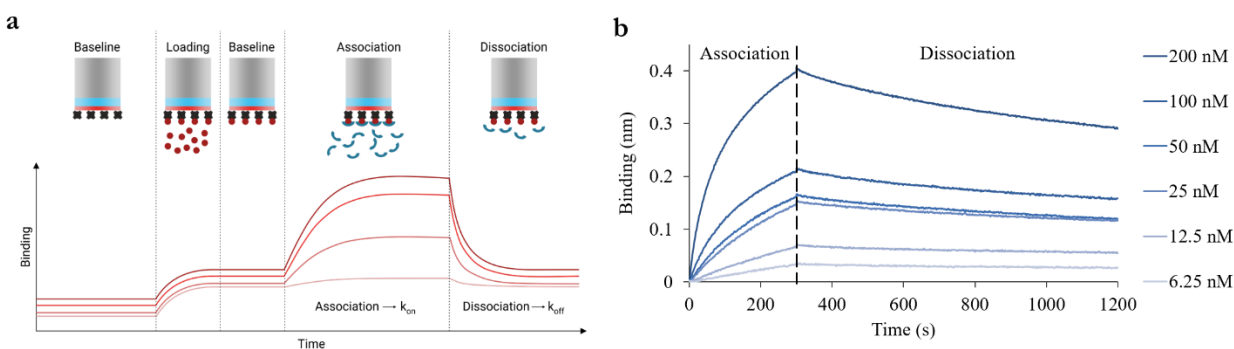
Figure 28: Binding affinity of FRET mutants to nanobody N00. (a) Schematic of a biolayer interferometry experiment using the Octet-RED96 machine (Forte bio). After establishing a baseline in buffer, the first protein is loaded onto biosensors in the loading step. Excess protein is washed off and the sensor dipped in a solution containing the second protein. In this association step a complex between the two proteins is formed. The last step is the dissociation where the sensor with the protein complex is brought back to buffer, where the second protein dissociates from the complex. During the experiment, the thickness increase of the sensor tip is measured. The thickness increases with protein binding to the sensor and complex formation and decreases with complex dissociation.

Figure from [181]. (b) Raw data of the association and dissociation step of FRET mutant W203C/Q487C binding to N00. N00 was immobilized on the sensor. Association of W203C/Q487C to N00 was done using 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM and 6.25 nM of W203C/Q487C.

From the nine tested FRET mutants, eight formed a complex with the nanobody and are thus still structurally intact on the periplasmic side (Table 6). Only for L52C/T351C there was no binding detectable (Table 6). The K_D values for wildtype and FRET mutants, which formed a complex, are all in a similar low nanomolar range, between 27.7 nM to 2.02 nM for the FRET mutants and 7.13 nM for wildtype DtpA (Table 6). To ensure that the labeling of the FRET mutants with the fluorophores does not hamper the complex formation, the complex formation was performed with labeled FRET mutants.

To test whether the labeling led to a change in K_D, for one FRET mutant, W203C/Q487C, the binding affinity was measured for the labeled and unlabeled protein. For both labeled and unlabeled W203C/Q487C, the K_Ds are still in the same range at 5.40 nM and 7.36 nM respectively, showing that labeling does not influence the K_D (Table 6).

At this step only one FRET mutant was excluded from the study, L52C/T351C.

Table 6: K_D of wildtype DtpA and FRET mutants binding to nanobody N00. K_Ds were determined by biolayer interferometry. Biotinylated N00 was immobilized on streptavidin sensors and binding was assessed at 200 nM, 100 nM, 50 nM, 25 nM, 12.5 nM and 6.25 nM of either wildtype DtpA or FRET mutants. The data was processed assuming a 1:1 stoichiometry of the DtpA-N00 complex. Full X^2 and R^2 give information on the quality of the fit.

Mutant	Labeled	K_D	K_D error	Full X^2	Full R^2
wildtype	-	7.13 nM	0.03 nM	0.073	0.9988
L52C/T351C	+	-	-	-	-
L52C/D449C	+	27.7 nM	0.53 nM	0.1481	0.9942
D111C/W203C	+	9.57 nM	0.07 nM	0.0479	0.9986
D111C/T351C	+	2.02 nM	0.17 nM	0.3507	0.9778
D111C/D449C	+	16.5 nM	0.21 nM	0.1844	0.9958
W203C/T351C	+	6.31 nM	0.05 nM	0.267	0.9953
W203C/Q487C	-	7.36 nM	0.05 nM	0.1616	0.9971
W203C/Q487C	+	5.40 nM	0.03 nM	0.1666	0.9983
T351C/W384C	+	2.40 nM	0.21 nM	1.5871	0.7924
D449C/Q487C	+	5.58 nM	0.56 nM	0.5846	0.9584

5.1.1.4.1.2. Labeling of FRET mutants for smFRET

The labeling step was incorporated into the purification protocol. This was necessary to remove unbound fluorophores after labeling. After solubilization from the membrane and a first IMAC step, the protein was labeled with Alexa fluorophores. Labeling was done with excess of fluorophores. To remove all non-bound fluorophores, a second IMAC step was introduced to the purification protocol. After labeling, the protein is immobilized, and the sample is washed extensively. This removes the unbound fluorophores and reduces the background noise in smFRET measurements.

A size exclusion chromatography (SEC) step was used to ensure homogeneity of the labeled sample. Here, the absorbance of the sample was recorded at different wavelength, 280 nm, 488 nm and 594 nm. The protein absorbance is measured at 280 nm whereas the absorbances of both Alexa dyes were measured at 488 nm and 594 nm, respectively. After labeling, DtpA elutes with a homogeneous peak with a considerable absorbance at the wavelength for both fluorophores, at the same time of the protein elution (Figure 29a-b, Figure 51 and Figure 52 Appendix). These results support a sufficient labeling of the sample.

Importantly, labeling of FRET mutants with the fluorophores did not affect ligand binding as was tested for the two mutants W203C/T351C and W203C/Q487C. In addition, both are still sufficiently stable for FRET measurements with a Tm of 55 °C and 49 °C respectively. Addition of the same ligands as in the quality control experiments before, AL, AF and AFA, increase the Tm of W203C/T351C by 2 °C to 4 °C and of W203C/Q487C by 3 °C to 4 °C (Figure 29c).
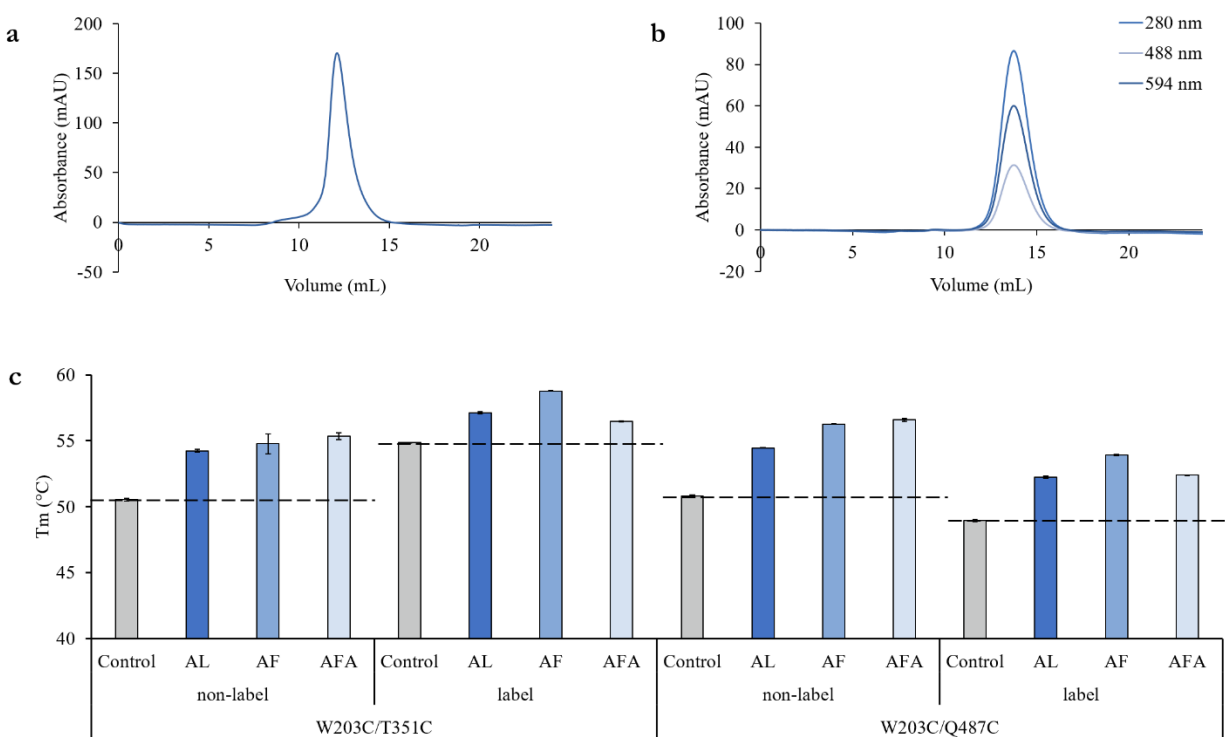
Figure 29: Labeling and quality control of FRET mutants. (a) SEC profile of FRET mutant W203C/Q487C without labeling with Alexa fluorophores. Absorbance was measured at 280 nm for the protein. (b) SEC profile of labeled W203C/Q487C after labeling with Alexa fluorophores. Absorbance was measured at 280 nm for the protein and at 488 nm and 594 nm for the donor and acceptor fluorophore respectively. (c) Thermal stability and ligand binding of non-labeled and labeled FRET mutants W203C/T351C and W203C/Q487C in LMNG. All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. All measurements done in triplicates.

5.1.1.5.   Reconstitution of FRET mutants into a lipidic environment

For several membrane proteins it was previously shown that their activity might be modulated by the surrounding lipid bilayer [182–184]. For this reason, FRET mutants were reconstituted into SapNPs for FRET measurements as well.

5.1.1.5.1. Reconstitution of FRET mutants into SapNPs

Reconstitution of a protein into SapNPs inserts it from a detergent solution back into a lipid bilayer. In the presence of lipids, the lipid-binding protein saposin A (SapA) forms SapNPs [185]. In these disc-like nanoparticles, SapA encloses a patch of lipids like a belt [185]. SapNPs have been successfully utilized to provide a near native environment for membrane proteins [183,186,187].

Here, an already established protocol was followed [185]. The detergent-solubilized FRET mutants were incubated with lipids to replace the detergent molecules and SapA to form the scaffold around the lipids holding the nanoparticle together. To test not only the influence of a lipid bilayer but also different lipid compositions of this bilayer, SapNPs were formed with several lipids: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (POPA) and the lipid mixture brain total lipid extract (BL) (Table 41 Appendix). After assembly of the nanoparticle, a SEC step was used to ensure homogeneity of the sample and remove free SapA as well as empty discs from the reconstituted sample. Here, the absorbance was again measured at three different wavelengths, 280 nm, 488 nm and 594 nm, to detect the protein absorption as well as the donor and acceptor signal, respectively. Again, the protein elutes with a considerable absorbance at the wavelength for both fluorophores, at the same time of the protein elution supporting the notion that labeled protein was incorporated into SapNPs (Figure 30a, Figure 53 and Figure 54 Appendix). The coelution of FRET mutants with SapA in the peak fractions of the SEC run showed successful formation of the SapNPs (Figure 30b).

As was previously shown for detergent-solubilized DtpA[186], small angle X-ray scattering (SAXS) measurements using wildtype DtpA, showed that DtpA is reconstituted into SapNPs as a monomer (Figure 30c-d). Hence, subsequent FRET measurements should give signals that originate from intramolecular and not intermolecular interactions.

After reconstitution into SapNPs, the thermal stability and the ability of the FRET mutants to still bind ligands and the nanobody was tested as well. Under these conditions, the results were similar to the samples tested in detergent. In general, the Tm of the FRET mutant W203C/Q487C in different SapNPs did not vary much from its Tm in LMNG, with 51 °C in LMNG, 51 °C in POPA, 53 °C in POPE, 52 °C in POPS and 50 °C in BL (Figure 30e). But here, one has to consider that LMNG alone leads to a considerable stabilization of the protein. Same as for detergent-solubilized sample, the FRET mutants reconstituted into SapNPs were still able to bind the tested ligands as well as the nanobody. On average the addition of the ligands AL, AF and AFA increased the Tm by 4 °C to 5 °C and the addition of N00 leads to an increase of 16 °C on average (Figure 30e).
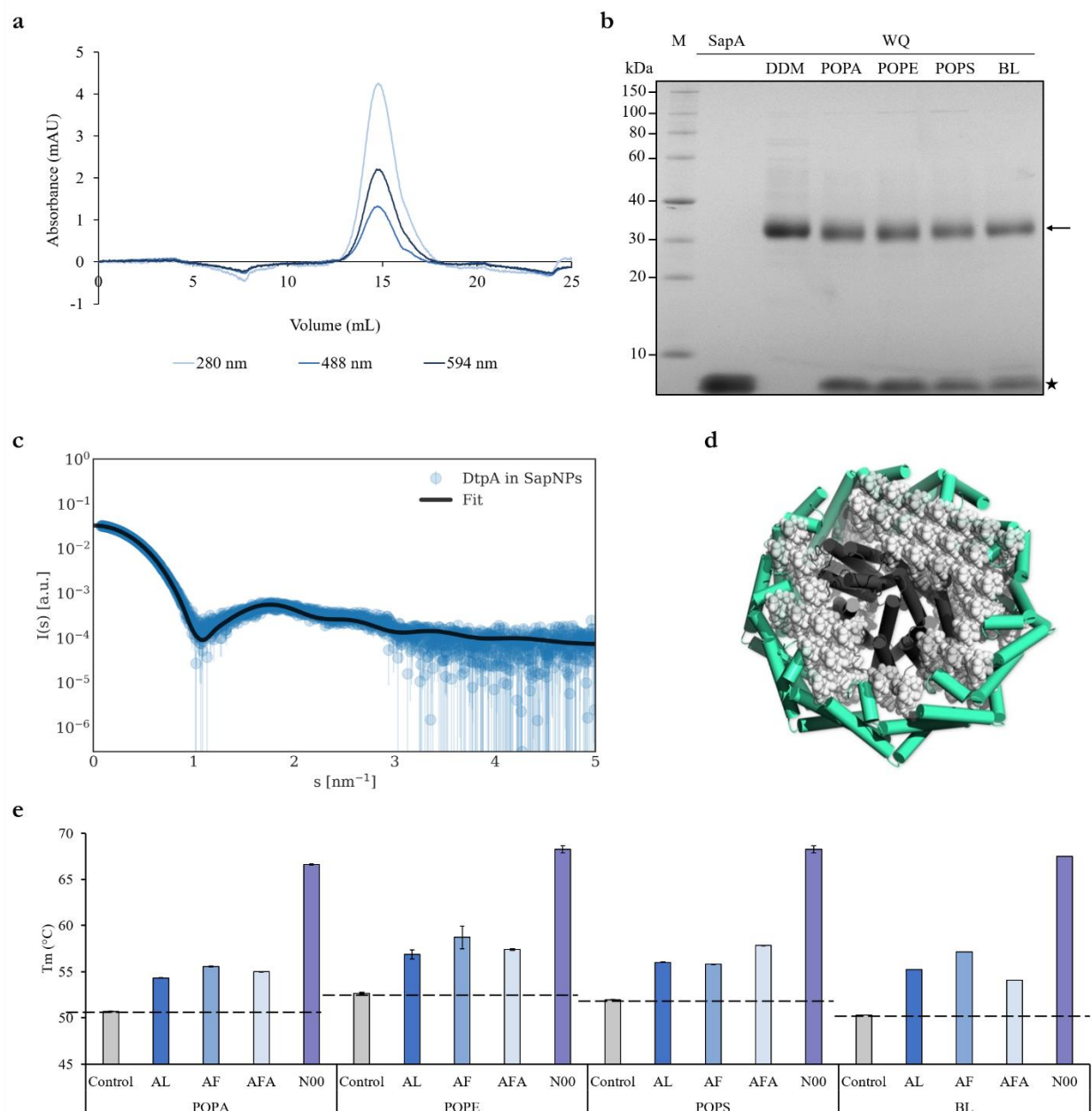
Figure 30: Reconstitution of labeled FRET mutants into SapNPs. (a) SEC profile of W203C/Q487C (WQ) reconstituted into POPE SapNPs. Absorbance was measured at 280nm for the protein and at 488 nm and 594 nm for the donor and acceptor fluorophore respectively. (b) SDS-PAGE of peak fractions from SEC runs after reconstitution of W203C/Q487C into SapNPs made with different lipids. W203C/Q487C run at 35 kDa (black arrow) and saposin A (SapA) at < 10 kDa (black star). After reconstitution, both proteins elute together, showing the formation of the SapNPs with W203C/Q487C reconstituted. (c) SAXS analysis of W203C/Q487C in POPE SapNPs. The calculated scattering curve fits the experimental data. (d) Model of DtpA (dark gray) in a lipid SapNP stabilized by saposin A (green). (e) Thermal stability and ligand binding of FRET mutant W203C/Q487C in SapNPs with different lipid composition. All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. All measurements done in triplicates.

5.1.2. SmFRET in solution

After ensuring the FRET mutants resemble the wildtype as closely as possible, FRET was measured for detergent-solubilized and SapNP reconstituted FRET mutants.

Of the FRET mutants which passed the quality control experiments, W203C/Q487C was by far superior in labeling efficiency which is why most experiments were performed using W203C/Q487C (Figure 27). Based on the available crystal structure of DtpA, the expected FRET efficiency for the inward-open state was calculated to be 0.54 for W203C/Q487C. Therefore, a FRET population at a FRET efficiency around 0.5 would correspond to the inward-open state of the transporter. A FRET population at higher FRET efficiency corresponds to a conformational state where the cytoplasmic side of the protein is closed which would fit to either an occluded or outward-open state. Due to the absence of crystal structures of those two states for DtpA, the expected FRET efficiencies cannot be calculated and those two states cannot be distinguished in this set-up.

All following smFRET data were measured and analyzed in collaboration with Tanya Lasitza Male and Hagen Hofmann from the Weizmann Institute of Science, Rehovot, Israel. All shown histograms are normalized to the bar with the maximum amount of bursts, meaning the height of the largest bar. This was done to better compare the ratios of the observed FRET populations, but typically between 5000 and 15000 bursts were collected per histogram.

5.1.2.1. SmFRET of detergent-solubilized DtpA

First, it was investigated which conformational states DtpA adopts when in detergent solution. Membrane proteins are routinely extracted from membranes using detergents such as DDM or LMNG and most of the available POTs structures to date were determined from such detergent-solubilized samples [41,50,52,59–61,74,75]. While detergent extraction and solubilization are wieldy used methods to enable working with hydrophobic membrane proteins in an aqueous solution, evidence is growing that the protein function might not always be preserved under these conditions [131,132]. Therefore, it is important to test whether all functionally relevant conformations can be sampled by the transporter in this rather artificial environment.
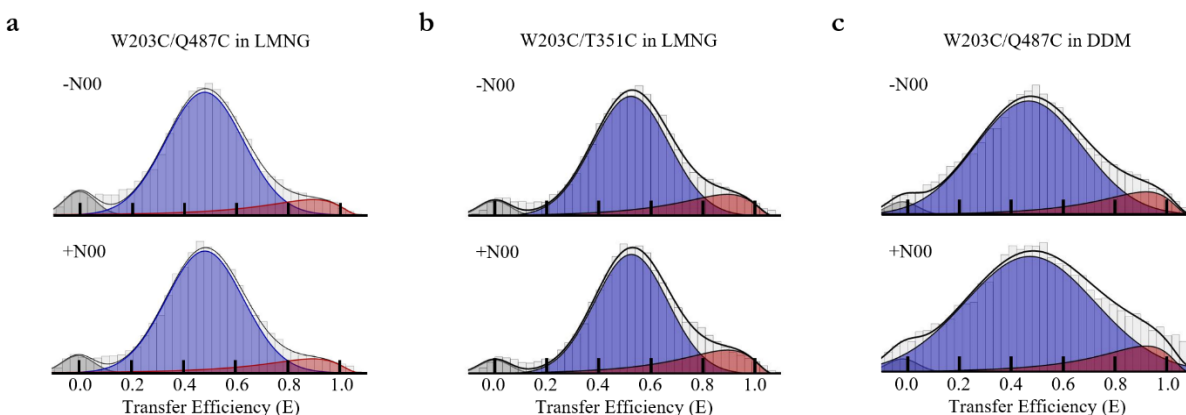
Figure 31: SmFRET of cytoplasmic FRET mutants in DDM and LMNG. FRET histograms of (a) FRET mutant W203C/Q487C in LMNG, (b) FRET mutant W203C/T35C in LMNG and (c) FRET mutants W203C/Q487C in DDM without and with addition of N00 at a final concentration of 8 µM. Two FRET populations are observed. A low FRET population at a FRET efficiency around 0.5

(blue) and a high FRET population at a FRET efficiency of around 0.9 (red). Contribution of molecules without an active acceptor are shown in gray. Figure from [188].

When smFRET for the DtpA variant W203C/Q487C in LMNG was measured, two populations, a major one at medium FRET efficiency (E = 0.48) and a smaller population at high FRET efficiency (E = 0.89) were observed (Figure 31a). The FRET efficiency calculated from the crystal structure of DtpA in the inward-open state (E = 0.54) agrees well with the one observed in the FRET experiments (Figure 31a). Therefore, the major population represents an inward-open state. The same distribution was observed for W203C/Q487C in a different detergent (DDM) as well as for a different DtpA variant in LMNG, namely W203/T351C (Figure 31b-c). This suggests that those two conformations are not an artifact of the detergent used or of the mutations introduced.

To further investigate whether the observed major population is a single defined state or an ensemble of several states with different degrees of cytoplasmic opening, the donor fluorescence lifetime for W203C/Q487C in LMNG was analyzed.

The fluorescence lifetime describes, how long a fluorophore exists in the excited state before relaxing back to the ground state. The donor fluorescence lifetime in the absence and presence of the acceptor can be used to calculate the FRET efficiency. For example, in the case of static distances between the fluorophores, this relationship should be linear [160]. If an ensemble of distances is present, this simple relationship will not fit the data anymore. Here, the normalized donor lifetime is plotted against the transfer efficiency and in case of a single static state, a linear dependence is expected (black solid line Figure 32a). Instead, the data deviates from this linear behavior which suggests the presence of a distribution of inward-open states.
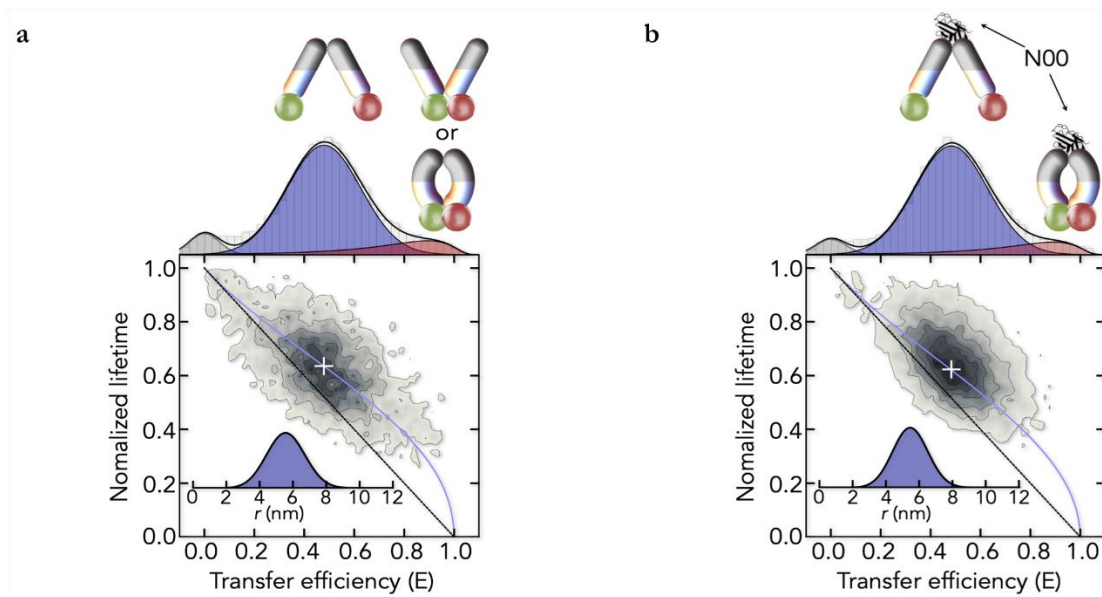
Figure 32: SmFRET of FRET mutant W203C/Q487C in LMNG. FRET mutants W203C/Q487C in DDM (a) without and (b) with addition of N00 at a final concentration of 8 μM. Two FRET populations are observed. A low FRET population at a FRET

efficiency around 0.5 (blue) and a high FRET population at a FRET efficiency of around 0.9 (red). Contribution of molecules without an active acceptor are shown in gray. 2D correlation map between the normalized donor fluorescence lifetime and transfer efficiency are shown at the bottom showing the dependence for a single donor-acceptor distance (black line) and the observed distance distribution (blue line). The inset shows the distance distribution of the donor-acceptor distance. Figure from [188].

These inward-open states differ in the width of the opening of the transporter to the cytoplasmic side. To judge how much they differ from each other, the distance distribution within this ensemble was calculated (insert Figure 32a). The inward-open states sampled by W203C/Q487C differ by 1.1 nm around the mean. This is a considerable difference for a relatively small protein like DtpA (54 kDa).

Thus, in detergent solution DtpA is able to adopt several inward-open states which differ in their cytoplasmic opening. The inward-open state captured in the crystal structure might only represent the most abundant one.

After assigning the major peak of the smFRET histogram of W203C/Q487C in LMNG to the inward-open state, the minor peak observed at E = 0.89 was investigated (Figure 32a). The higher FRET efficiency corresponds to a decreasing distance between the fluorophores, meaning a closed cytoplasmic side. Based on the transport model (Figure 5), two states are predicted to have a closed cytoplasmic side, the outward-open and the occluded state. These two states have a closed cytoplasmic side but differ on the periplasmic side. While in the outward-open state, the transporter is open on the periplasmic side, which in the case of the occluded state is closed. Since the W203C/Q487C variant carries both fluorophores on the cytoplasmic side, information about the conformational state on its periplasmic side is lacking. Thus, the outward-open and occluded states cannot be distinguished based on the observed FRET efficiency in this set-up.

To overcome the obstacle of distinguishing between the outward-open and occluded states, a biochemical approach was used instead. Here, the conformation-specific nanobody (N00) was added to the W203C/Q487C variant in LMNG and smFRET was remeasured. N00 was co-crystalized with DtpA [50] and binds to the wildtype as well as to the FRET mutants with nanomolar affinity (Table 6). In the crystal structure, N00 binds to DtpA on the periplasmic side, locking the protein in an inward-open state. Addition of N00 to W203C/Q487C however did not alter the ratio or FRET efficiency of the populations observed in the smFRET histogram. There remains a major peak at E = 0.48 nm and a minor peak at E = 0.89 nm (Figure 32b). In the case of N00 bound W203C/Q487C, the minor peak at high FRET efficiency corresponds to the occluded state. Upon N00 binding, the periplasmic side is closed and at the same time the FRET efficiency indicates a closed cytoplasmic side.

As a similar high FRET peak for W203C/Q487C without N00 bound is observed, this might also represent the occluded state. However, at the same time the possibility cannot be excluded that the presence of molecules in the outward-open conformation contributes to this peak. N00 binding might not lead to the formation of the outward-open state but the occluded state instead. Thus, molecules in the outward-open and/or occluded state might contribute to the minor peak in the W203C/Q487C sample (Figure 32a) whereas for the minor peak in the N00 bound W203C/Q487C sample (Figure 32b) only molecules in the occluded state would contribute.

To exclude the possibility that the peak at high FRET efficiency results from a misfolded and kinetically trapped state, RASP analysis was performed. This technique is used to test if there is a direct exchange between subpopulations which in this case are the minor and the major peak [163]. For this, the assumption is exploited, that since the experiments are performed at picomolar concentrations, it is much more likely that a molecule which just left the confocal spot returns

instead of a new molecule coming in [163] (Figure 33a). This holds true for several tens of milliseconds after the initial molecule was first recorded [163]. If molecules of different FRET populations interconvert on a millisecond timescale, those molecules are changing conformation between leaving and re-entering the confocal spot. To analyze if this is the case, one selects all bursts from molecules exhibiting a certain FRET efficiency, here all molecules contributing to the minor peak. With these bursts, the so-called initial histogram is constructed, consisting of only one population. Then all bursts are selected that are observed during a specific time delay after the initial burst e.g. 1 ms. With these bursts, a so-called recurrence histogram is constructed. The recurring molecule is presumably the same molecule re-entering the confocal spot within 1 ms. When RASP was performed for the W203C/Q487C data set, with increasing time delay in the recurrence histograms, the minor peak decreased and the major peak increased (Figure 33b).
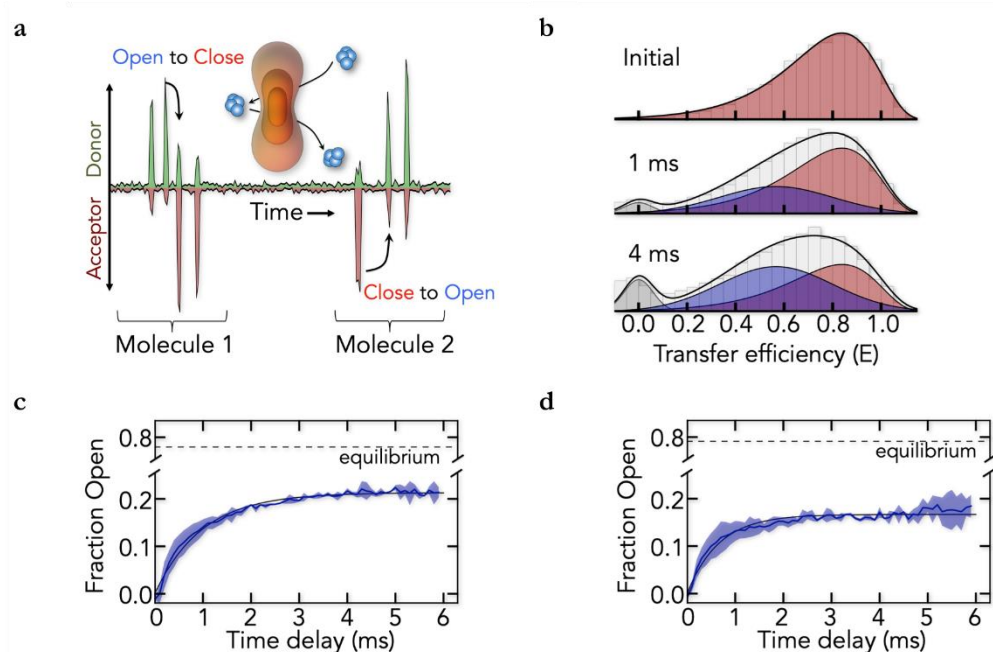
Figure 33: High and low FRET peaks in detergent are in dynamic exchange. (a) Schematics of a RASP experiment. Donor (green) and acceptor (red) signal recorded over time with the reentry of the same molecule several times to the confocal spot (orange). (b) Initial and recurrence histograms at two different recurrence times. Initially the high-FRET population was selected, but the recurrence histograms show the formation of a low-FRET species over-time. (c-d) Formation of the low-FRET population (blue) with a single-exponential fit (black line), (c) without and (d) with a final concentration of 8 μM N00. The dashed line indicates the equilibrium fraction of the low-FRET population. (Figure from [188].

When the fraction of molecules of the major peak (Fraction Open) was plotted against the time delay it was observed that after 1 ms the fraction of the major peak did not increase anymore. This shows that the transition time for molecules from the minor to the major peak is ~ 1 ms (Figure 33c). However, even when increasing the time delay to 6 ms, the original histogram with a major peak fraction (Fraction Open) of 0.88 cannot be obtained (Figure 33c). This indicates that processes slower than 6 ms contribute to the interconversion between molecules in the minor and major peak as well. Consequently, both populations are in dynamic exchange. The high FRET population of the minor peak is thus not a misfolded state.

When RASP is performed for N00 bound W203C/Q487C, here as well both FRET populations are in dynamic exchange (Figure 33d). Taken together, these results suggest that even with N00 bound, meaning the periplasmic side of the protein closed, DtpA is still able to adopt different conformations on the cytoplasmic side.

So far, in all tested samples the majority of the proteins adopt the inward-open state and neither a different mutation or different detergent was able to change the ratio between the conformational states. Using smFRET, it should be possible to identify conditions with an increased proportion of molecules in the outward-open state in detergent. First, it was tested whether the addition of different ligands would shift the ratio of high and low FRET populations. Nine different di- and tripeptides as well as the prodrug valaciclovir, were added to the W203C/Q487C variant. Although all ligands were shown to bind W203C/Q487C in a thermal shift assay, none altered the ratio of high and low FRET peaks (Figure 34a,c). This agrees with the available crystal structures of DtpA in the apo- and ligand-bound form [50]. In both structures, DtpA is in the inward-open state and the rsmd is 0.33 Å over 3809 atoms [50]. This shows that ligand binding is possible in the inward-open state and that ligand binding does not necessarily induce a shift towards an outward-open conformation.

As DtpA is a proton-coupled transporter [53], it was assumed that protonation or both, protonation and substrate binding, might be required for a conformational shift towards the outward-open state. Therefore, three different pH values were tested for their effect on the conformational states of the FRET mutant. W203C/Q487C was shown to be stable and able to bind the ligand LL at the tested pH values of 5.3, 7.5 and 9.3 (Figure 34b). But similarly to the addition of ligand, neither a change of pH nor the combination of a change in pH and ligand addition altered the populations observed in the FRET histograms (Figure 34d). This suggests that in detergent the outward-open state cannot be adopted by DtpA.

To summarize, the investigation of conformational states adopted by detergent-solubilized DtpA showed that (i) the majority of molecules adapt the inward-open state, (ii) this inward-open state is in fact an ensemble of states with different degrees opening of the cytoplasmic side and not a single defined state, (iii) a minority of molecules exists in an outward-open and/or occluded state, (iv) the exchange rate between the two states is on a millisecond timescale, and (v) the ratio between the states cannot be considerably altered by variation of detergent, pH and/or ligand.
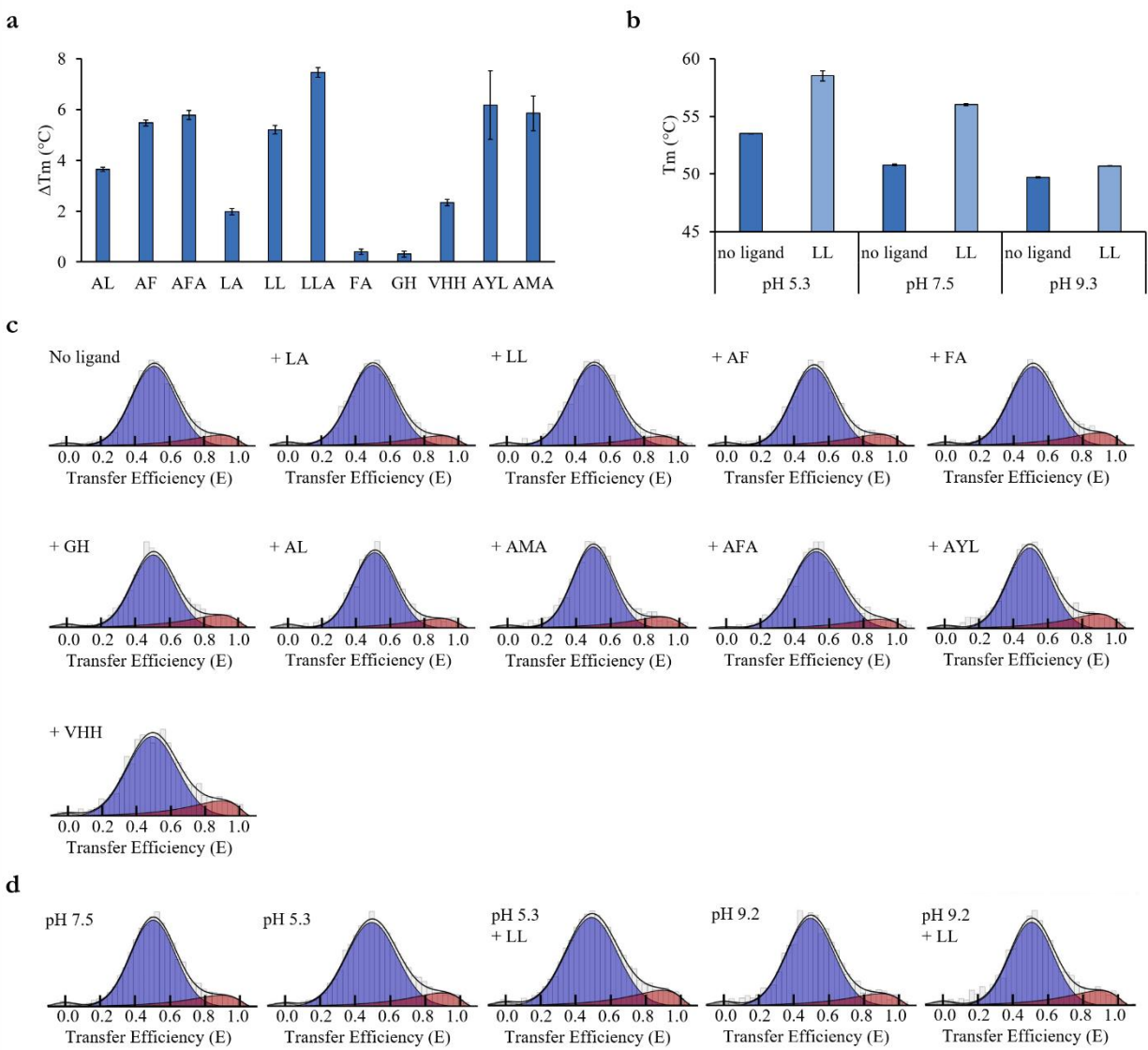
Figure 34: Ligand binding and stability at different pH for FRET mutant W203C/Q487C in LMNG. (a) Ligand binding of FRET mutants in LMNG. The tested ligands are Leu-Ala (LA), Leu-Leu (LL), Ala-Phe (AF), Phe-Ala (FA), Gly-His (GH), Ala-Leu (AL), Ala-Met-Ala (AMA), Ala-Tyr-Leu (AYL) and valaciclovir (VHH). The difference of the melting temperature ΔTm between W203C/Q487C without and with ligand addition was plotted. (b) Thermal stability at different pH of FRET mutant W203C/Q487C in LMNG with and without addition of the ligand LL. W203C/Q487C was measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. All measurements were done in triplicates. (c) FRET histograms of FRET mutant W203C/Q487C in LMNG without and with addition of different ligands, Leu-Ala (LA), Leu-Leu (LL), Ala-Phe (AF), Phe-Ala (FA), Gly-His (GH), Ala-Leu (AL), Ala-Met-Ala (AMA), Ala-Tyr-Leu (AYL) and valaciclovir (VHH) and (d) at different pH with and without addition of ligand LL. Final ligands concentration was 2 mM. Two FRET populations are observed. A low FRET population at a FRET efficiency around 0.5 (blue) and a high FRET population at a FRET efficiency of around 0.9 (red). Contribution of molecules without an active acceptor are shown in gray. Figure from [188].

5.1.2.2. SmFRET of DtpA in SapNPs

In detergent solution, no conditions were found where the outward-open state was accessible to DtpA. But based on the transport model and on available crystal structures from other MFS subfamilies [9], this state does exist. Assuming that the artificial detergent environment might hinder the protein to adopt all functionally relevant states, smFRET measurements were repeated with the protein in a more native-like environment. Therefore, the FRET mutants were reconstituted into SapNPs. Initial experiments were performed with the W203C/Q487C variant reconstituted into POPE SapNPs. POPE is the most abundant lipid in the *E. coli* inner membrane [189], the native environment that hosts DtpA (Figure 35a). Compared to the FRET histograms of W203C/Q487C in LMNG, in the histograms of W203C/Q487C in POPE SapNPs the same two populations, a major peak at E = 0.55 nm and a minor peak at E = 0.90 nm, were observed (Figure 35b). The analysis of the fluorescence lifetime shows that, as it was the case in detergent, the major peak is an ensemble of different inward-open states and not a single static state (Figure 35c). However, in contrast to the detergent data, the histogram of the W203C/Q487C variant in POPE SapNPs differs in two points, (i) the peaks are broader and (ii) the ratio between the peaks is smaller. Broadening of peaks can be due to quenching of the fluorophores, a strong interaction of the FRET dyes with the surface of the protein limiting their ability to rotate freely, or the different conformers of the inward-open ensemble are sampled at a slower timescale than in detergent.

To exclude quenching effects, nanosecond Fluorescence correlation spectroscopy (nsFCS) was performed [190]. The nsFCS of the acceptor signal in W203C/Q487C shows that no quenching is taking place (Figure 55 Appendix). Limited rotational freedom of the dyes in SapNPs compared to detergent can be ruled out because the anisotropy of the donor fluorescence does not change between those conditions. The values range between 0.173 and 0.234, with 0.2 being considered to represent a good dimensional freedom [146] (Figure 55 Appendix). Thus, the observed broadening of the peak is most likely caused by a decreasing speed with which the different inward-open conformations are sampled in a lipidic environment compared to detergent.

The most striking difference between the histogram in detergent and SapNPs is the ratio of the two populations. Whereas in detergent, only 12 % of molecules belonged to the minor peak, in SapNPs it is 29%, meaning the amount more than doubled (Figure 35a). Similar to the situation in detergent, it was investigated which conformation could be assigned to the minor FRET population. For this, N00 was added to the sample and smFRET was measured on the complex. The addition of N00 to W203C/Q487C reconstituted into POPE SapNPs shifts the high FRET population to the low FRET population (Figure 35a), indicating conformational change from an occluded or outward-open state to the inward-open state. In comparison to smFRET data measured in detergent, where the occluded state does not change upon N00 interaction, the data in SapNPs suggests that the high FRET population is representing the outward-open state. These results show that the native-like lipidic environment enables DtpA to adopt all functionally relevant conformations.

RASP analysis of the two populations showed that they are in dynamic exchange, as is the case in detergent (Figure 35c).
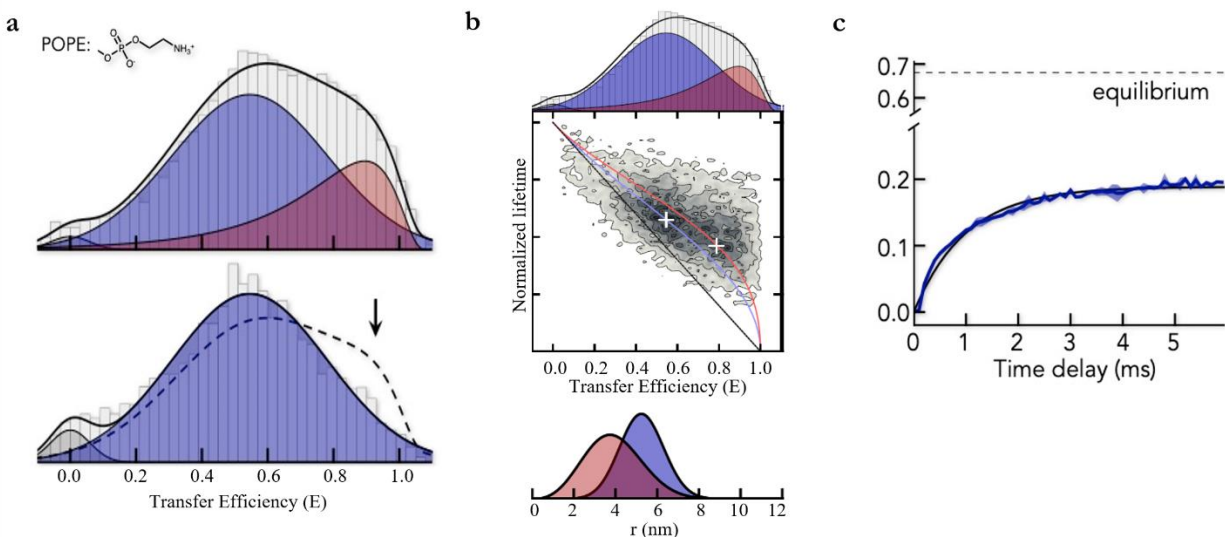
Figure 35: SmFRET of FRET mutant W203C/Q487C in POPE SapNPs. (a) FRET histograms of FRET mutant W203C/Q487C in POPE SapNPs without and with addition of N00. The final concentration of N00 was 8 μM. Without the addition of N00, two FRET populations are observed. A medium FRET population at a FRET efficiency of 0.5 (blue) and a high FRET population at a FRET efficiency of around 0.9 (red). After addition of N00 only the medium FRET population around a FRET efficiency of 0.5 is present. Contribution of molecules without an active acceptor are shown in gray. (b) 2D correlation map of the fluorescence lifetime of FRET mutant W203C/Q487C in POPE SapNPs. The dependence for a single donor-acceptor distance (black line) and the observed distance distribution for each population (blue and red line) is shown. At the bottom the distance distribution of the donor-acceptor distance for each population. (c) Formation of the low-FRET population (blue) with a single-exponential fit (black line). The dashed line indicates the equilibrium fraction of the low-FRET population. Figure from [188].

Next, it was investigated if simply the presence of a lipidic environment had this effect on DtpA or if the protein would behave differently depending on which lipid was used to form the nanoparticle. For this, W203C/Q487C was reconstituted into SapNPs of different lipid compositions: POPS, POPA and BL. These lipids were chosen because they have already been shown to be able to reconstitute DtpA in to SapNPs [50].

A strong effect of the lipid choice on the FRET histograms was observed (Figure 36a-c). In POPS and POPA, the same low and high FRET peaks as for POPE are present (Figure 36a-b). But in addition, there is a third population at even lower FRET efficiency present, E = 0.25 (Figure 36a-b). This third FRET population corresponds to a conformational state where the helices on the cytoplasmic side are even further apart than in the inward-open structure of DtpA. Thus, this conformation exhibits an even larger opening of the transporter on this side and will be referred to from here on as the extreme inward-open state. POPS and POPA are not found in the inner membrane of *E. coli*, thus they do not represent the native environment of DtpA as well as POPE. As the extreme inward-open state is only observed in POPS and POPA but not POPE, it could originate from this unnatural lipid surrounding. It is worth mentioning that such an extreme inward-open state was not predicted by the current transport model and its relevance for the transport cycle is unclear.
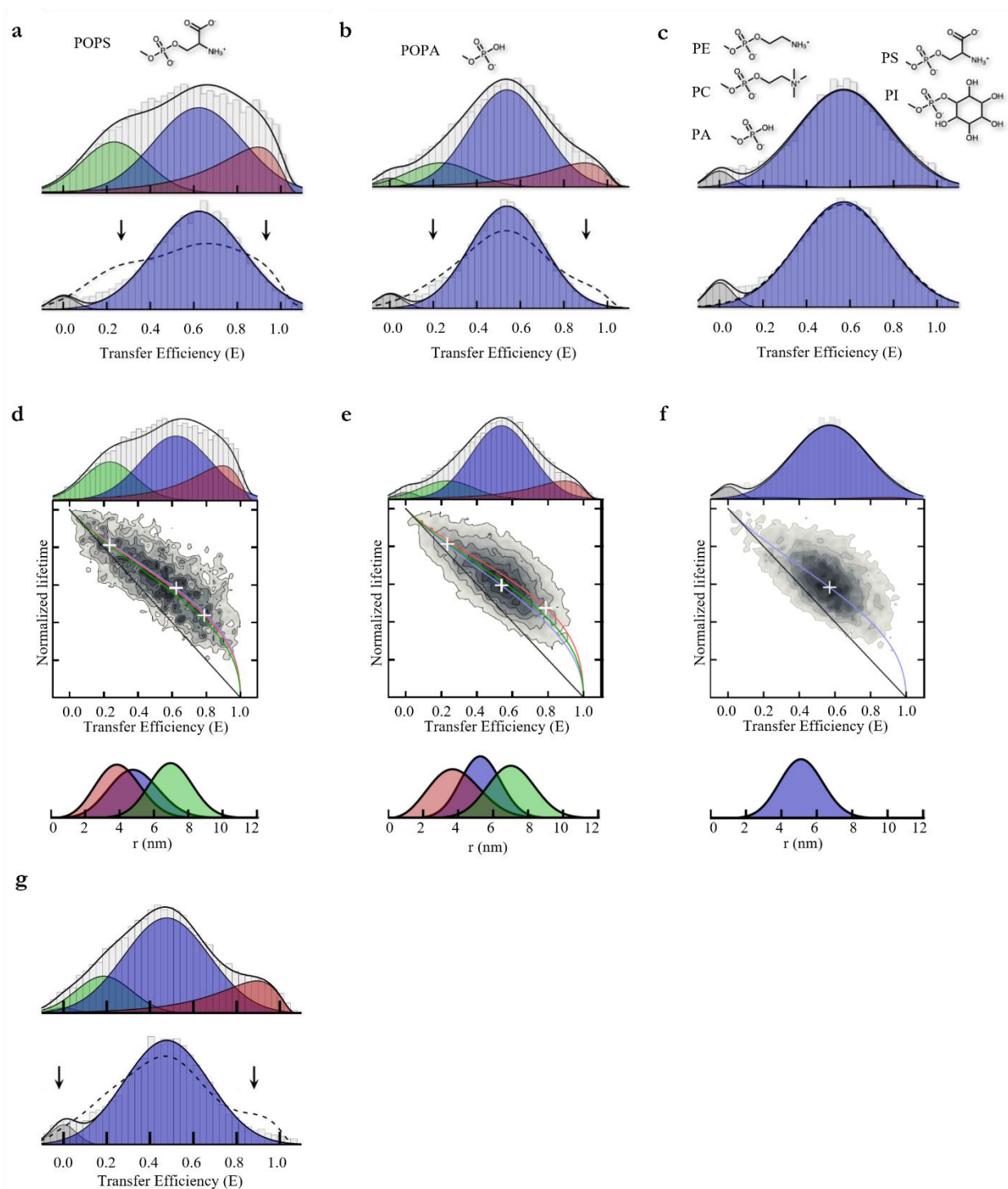
Figure 36: SmFRET of FRET mutant W203C/Q487C in SapNPs of different lipid composition. (a-c) SmFRET histograms of W203C/Q487C in (a) POPS, (b) POPA and (c) BL SapNPs without (top) and with (bottom) the addition of 8 μM N00. The position and width of the low-FRET peak in the presence of N00 was used to fit the histogram in the absence of N00. FRET populations observed, are a low FRET population at a FRET efficiency of 0.2 (green), a medium FRET population at a FRET efficiency of 0.5 (blue) and a high FRET population at a FRET efficiency of around 0.9 (red). (d-f) Fluorescence lifetime of FRET mutant W203C/Q487C in (d) POPS, (e) POPA and (f) BL SapNPs. The dependence for a single donor-acceptor distance (black line) and the

observed distance distribution for each population (blue, green and red line) is shown. At the bottom the distance distribution of the donor-acceptor distance for each population. (g) SmFRET histogram of W203/T351C in POPS SapNPs in the without (top) and with (bottom) 8 µM N00. Populations were fitted as described for (a-c). Figure from [188].

Unlike POPS and POPA, where a third additional FRET population is found, in BL SapNPs only one FRET population corresponding the inward-open state is observed (Figure 36c). In contrast to SapNPs from POPE, POPS, and POPA which contain one defined lipid type, BL is a mixture of lipids containing PC, PE, PI, PS, PA and 58.7 % unknown lipids [191]. Furthermore, BL is isolated from eukaryotic membranes. Similar to POPS and POPA, this could be the reason why the conformational states observed in BL differ from those in POPE. It seems that in BL, DtpA is not able to adopt any other state than the inward-open state.

These results indicate that the conformational states, DtpA is able to adopt, are strongly influenced by the composition of the lipidic environment. Depending on the lipid, up to three FRET population can be observed and can be assigned to the following conformational states: extreme inward-open, inward-open and outward-open. The ratio of the observed FRET populations depends on the chemistry of the lipid headgroup. In all tested SapNPs, addition of N00 shifts all FRET populations to the inward-open state, showing that in all lipid compositions, regardless of the headgroup chemistry, the cooperativity of the cytoplasmic and periplasmic side is restored (Figure 36a-c). Analogously to the W203C/Q487C variant in POPE SapNPs, the analysis of the fluorescence lifetime of W203C/Q487C in POPS, POPA and BL SapNPs shows that the major peak is an ensemble of different inward-open states (Figure 36d-f).

For the FRET mutant W203/T351C in POPS SapNPs, the same populations as for W203C/Q487C in POPS SapNPs were present, showing again that this is a feature of DtpA and not that of a specific mutant (Figure 36g).

Similar to the observations for W203C/Q487C in detergent, in SapNPs the addition of ligands, a change in pH, or the combination of both did not considerably alter the FRET histograms (Figure 37). This was not expected as in SapNPs the outward-open state is present in the sample. Ligand binding and protonation of the transporter is expected to change its conformation from the outward-open to inward-open state according to the transport model. However, this is not the case here. In the experimental set-up, the electrochemical gradient and substrate gradient the transporter would encounter in a native condition, when embedded in the membrane is not recreated. Thus, although the presence of a lipidic environment enables the transporter to access all functionally relevant conformational states for a full transport cycle, the presence of an electrochemical and substrate gradient might still be necessary for the transporter to actually perform a transport cycle.

Overall, the investigation of DtpA in a lipidic environment revealed that (i) besides the inward-open and occluded states observed for detergent-solubilized protein, an extreme inward-open state and most importantly, the outward-open state is sampled by DtpA. (ii) The number of conformational states DtpA is able to adopt and the ratio between them depends on the chemistry of the lipid headgroup, (iii) as in detergent the ratio between the states cannot be considerably altered by pH and/or ligand.
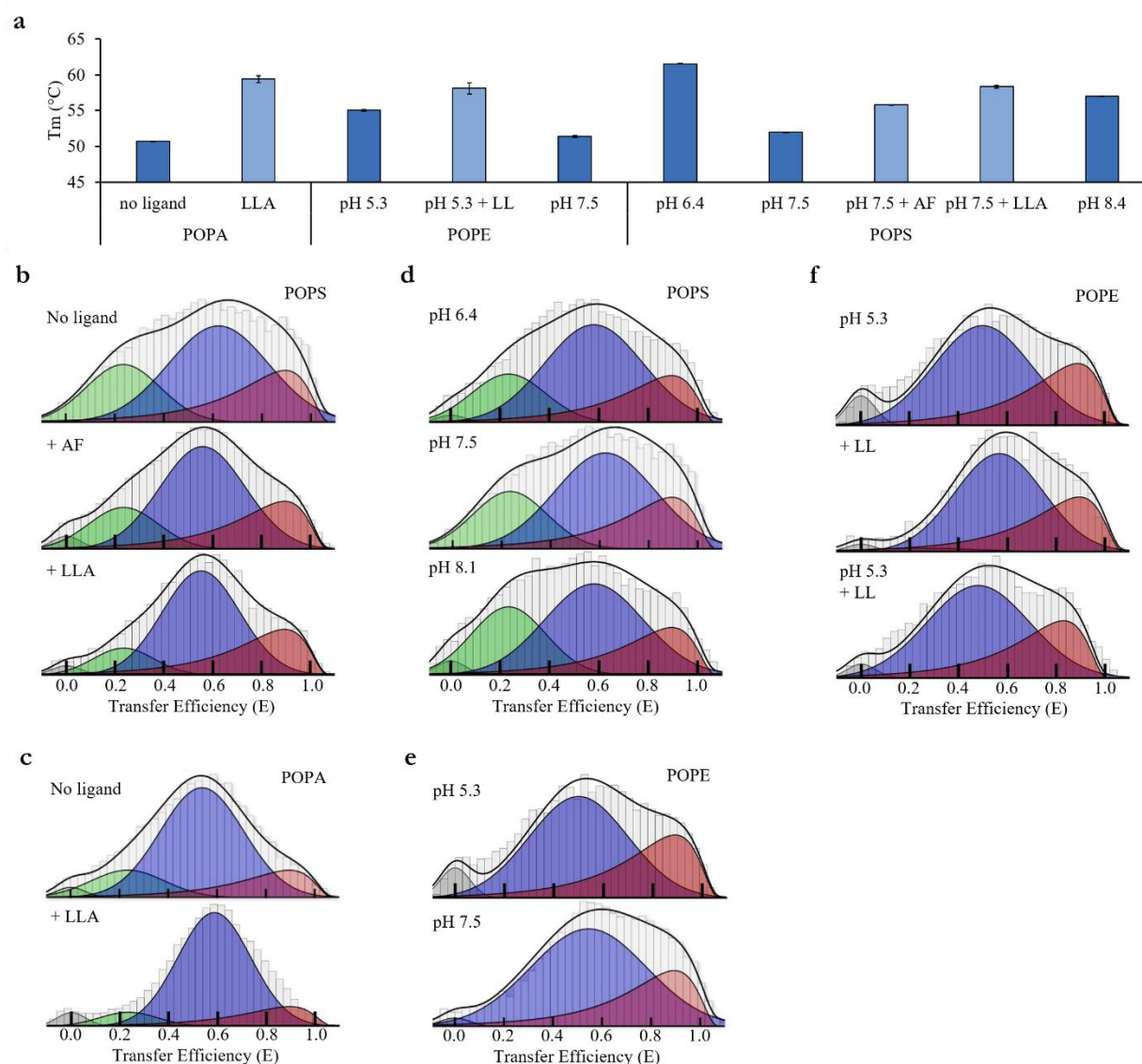
Figure 37: Ligand binding and stability at different pH for FRET mutant W203C/Q487C in SapNPs. (a) Ligand binding and thermal stability at different pH of FRET mutant W203C/Q487C in SapNPs made with POPA, POPS or BL. The tested ligands are Leu-Leu-Ala (LLA), Leu-Leu (LL) and Ala-Phe (AF). All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands (no ligand). For comparison, of the stability at different pH, the stability of W203C/Q487C at pH 7.5 was measured as this is the pH of the buffer used for purification. All measurements were done in triplicates. (b-c) FRET histograms of FRET mutant W203C/Q487C in (b) POPS SapNPs and (c) POPA SapNPs with and without the addition of ligands Ala-Phe (AF) and Leu-Leu-Ala (LLA). (d-e) FRET histograms of FRET mutant W203C/Q487C at different pH in (d) POPS SapNPs and (e) in POPE SapNPs. (f) FRET histograms of FRET mutant W203C/Q487C in POPE SapNPs at acidic pH with and without ligand Leu-Leu (LL) addition. Up to three FRET populations are observed. A low FRET population at a FRET efficiency around 0.2 (green), a medium FRET population at a FRET efficiency of 0.5 (blue) and a high FRET population at a FRET efficiency of around 0.9 (red). Ligand addition did not influence the smFRET histograms. Contribution of molecules without an active acceptor are shown in gray. Figure from [188].

5.1.2.2.1. Modeling of DtpA conformations observed in SapNPs

To date, structures of POTs have only been solved in the inward-open or occluded state (Table 2). As described above, DtpA adopts two additional states, the outward-open and the extreme inward-open state (Figure 38). To investigate how the domains are arranged in those two states, rigid body rotations of the N- and C-terminal domain with respect towards another were performed. Then, the expected FRET efficiency of the resulting rotamers was calculated. The current transport model suggests that the protein rotates around a central axis through the substrate binding site, which is located between the N- and C-terminal domain. Starting with the crystal structure in the inward-open state, the domains were rotated around this axis. To obtain a rotamer with a FRET efficiency E = 0.82, which would correspond to the outward-open state, the domains had to be rotated by -40° (Figure 38a-b).
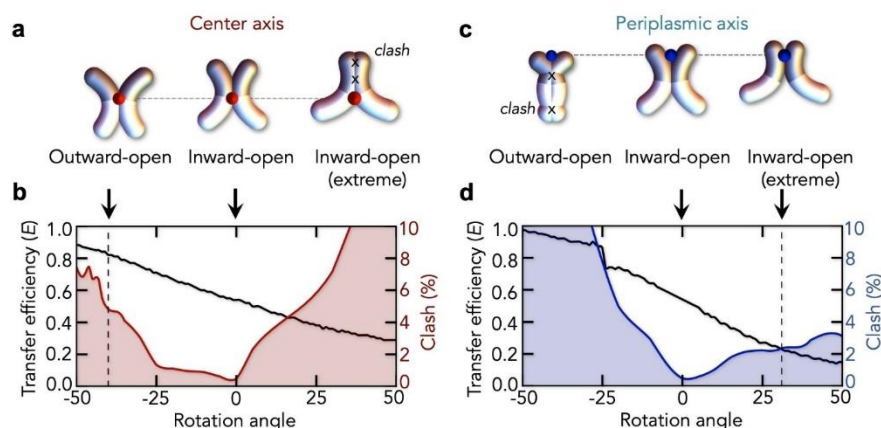
Figure 38: Modeling the different DtpA conformers observed at smFRET experiments. (a) Schematics of the rigid-body rotations of DtpA with an axis through the substrate binding site (center axis). (b) Calculated FRET efficiency (black) and clash score (red) for rotamers obtained through rotations around the center axis. (c-d) Same as (a-b) for rotations around a periplasmic axis. Figure from [188].

The degree of atomic clashes for the outward-open model of 4.8% is still acceptable considering, that no energy optimization of the structures was performed (Figure 38a-b). Rotation around the same axis however, is not sufficient to explain the extreme inward-open state. If the opening on the cytoplasmic side is increased by rotation around the central axis, too many atomic clashes are introduced (Figure 38a-b). Hence, a different axis between the N- and C-terminal domain at the periplasmic side was chosen to rotate the domains. In this case, the extreme inward-open conformation can be reached by rigid-body rotation and introducing only 2.8% atomic clashes (Figure 38c-d). It has to be noted that at the same time, this particular rotation axis cannot be used to model the outward-open state (Figure 38c-d).

For the rigid-body rotations performed here, the linker between the Ha and Hb helices from the HaHb domain was omitted when modeling the outward-open state, since the HaHb domain hindered the conformational switch. When analyzing the interface of the HaHb domain with the core helices of DtpA, it was observed that this region is highly hydrophobic (Figure 39). The hydrophilicity was judged according to the hydrophobicity scale proposed by Kyte and Dolittle et al. [192].
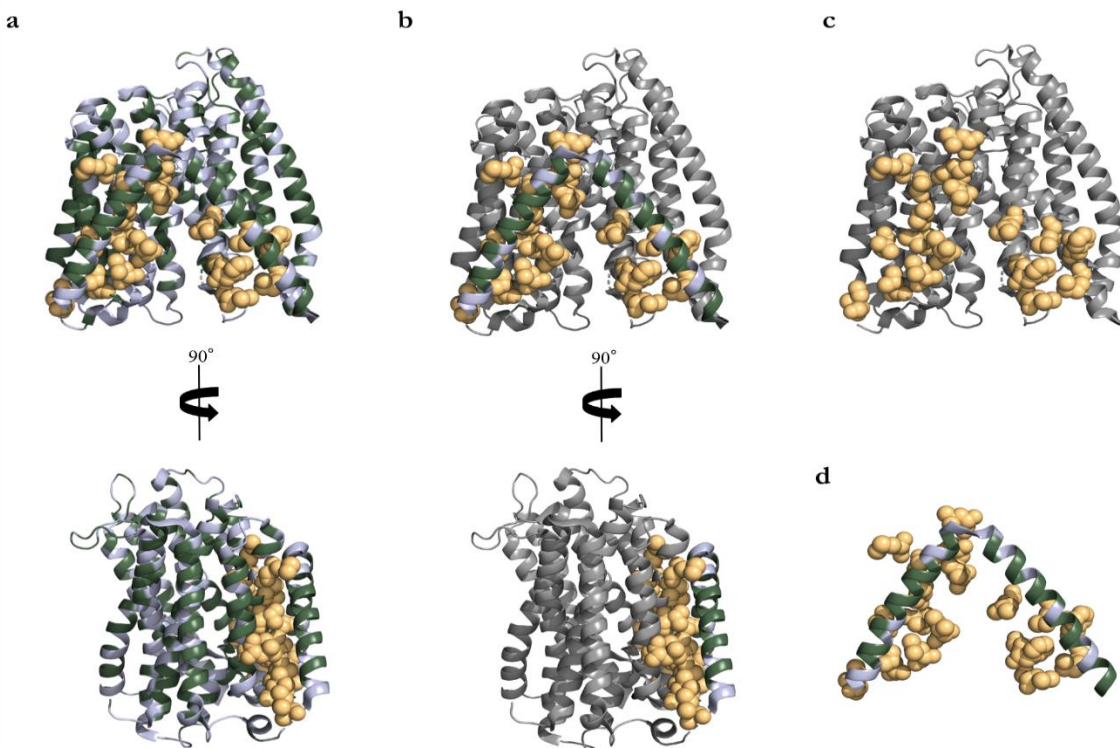
Figure 39: Hydrophobicity of DtpA. (a) Hydrophobicity of DtpA according to Kyte and Dolittle [192]. Negative values for amino acid residues on the Kyte-Dolittle scale are defined as hydrophobic and are colored green. Neutral and hydrophilic residues have positive values on the Kyte-Dolittle scale and are colored blue. Hydrophobic residues within a 4 Å distance of the HaHb domain are shown as yellow spheres. (b) Same coloring as before is used but the N-terminal and C-terminal domain are colored gray for better visibility of the HaHb domain. (c) Same coloring as before is used but the HaHb domain is omitted for better visibility of the hydrophobic residues within a 4 Å distance of the HaHb domain. (d) Same coloring as is used before but the N-terminal and C-terminal domains are omitted for better visibility of the hydrophobic residues within a 4 Å distance of the HaHb domain together with the HaHb domain.

Based on these results, it is assumed that in detergent a more water-rich environment exists compared to the lipid environment in SapNPs. Thus, the HaHb domain is strongly attached to the DtpA core, hindering the switch of the protein to an outward-open conformation in detergent. In the lipid environment however, this attachment could loosen, allowing the protein to adopt the outward-open state as observed for the FRET measurements.

5.1.3. Discussion

The human POTs, PepT1 and PepT2, are of high pharmacological interest because they can be utilized as drug delivery systems [12,56]. Therefore, POTs have been extensively studied in recent years. Often bacterial homologues are studied as a proxy for the human transporters, as those are challenging to work with due to their low stability and purification yield. DtpA from *E. coli* is a well-suited substitute for PepT1 as both transporters exhibit a similar substrate specificity [48,53]. Here, the dynamics of DtpA were studied using smFRET to gain insight in the mechanism of the transport cycle that goes beyond what can be learned from high-resolution structures.

Interestingly, in detergent solution, DtpA adopts mainly the inward-open state. Only a small fraction of the protein adopts an occluded state. This can explain the observed crystallization bias for POTs. Although over the past years numerous POT structures were solved, both in the apo and ligand bound form, only inward-open and occluded conformation were observed but no outward-open state (Table 2).

Furthermore, our experiments showed that the observed inward-open and occluded state in detergent convert into each other on a millisecond time scale. In addition, the inward-open conformation itself is not one static state but an ensemble of conformations with different cytoplasmic opening. The function of the fast sampling of different inward-open conformers is unknown, but it could be to aid the orientation of substrate in the binding site, as was already observed in the case of the enzyme adenylate kinase [193]. Thus, in detergent solution, the protein is highly dynamic, sampling different degrees of opening on the cytoplasmic side, but it cannot adopt the functionally relevant outward-open state.

In contrast, the outward-open state is observed when DtpA is in a lipid environment. This conformational difference of the protein between detergent and lipid environment could be caused by the HaHb domain. The HaHb domain was found to be tightly packed to the core helices of the transporter in the available crystal structure. In this conformation, it restricts a rigid body movement of the N- and C-terminal domains to adopt the outward-open state. In a lipid environment, the long aliphatic tails of the lipids could be able to insert between the 12 core helices of the transporter and the HaHb domain. This would weaken the attachment of this domain to the transporter core which in turn could enable the conformational change from the inward-open to the outward-open state. In POT structures solved from crystals that were grown in LCP, lipids could be observed between the transporter core and HaHb domain [62].

In addition, it was found that DtpA is not only sensitive toward a lipid or non-lipid environment but also shows a strong dependence on the lipid composition. POPA, POPE and POPS, which were tested here, only differ in their headgroup, with POPA and POPS lipids carrying a net negative charge, while POPE is uncharged. In all three lipid environments, the physiological relevant inward-open and outward-open state are adopted by DtpA but only with the negatively charged POPA and POPS, an extreme inward-open state is present. As those two lipids are not abundant in the *E. coli* membrane, the natural environment of DtpA, it might be that this extreme inward-open state has no functional relevance in the transport cycle. But its presence shows that the conformational space DtpA is able to sample strongly depends on its surroundings. In general, similar conformational differences between a protein sample in detergent and lipid environment were also observed for a scramblase using cryo-EM [194]. This suggests that caution is needed when interpreting structures of membrane proteins obtained from detergent-solubilized samples, as in detergent not all functionally relevant conformations might be adopted by the proteins.

5.2. Characterization of *E. coli* ACS transporters

In *E. coli*, six ACS transporters have been annotated, GarP, GudP, DgoT, LgoT, RhmT and ExuT. These transporters are poorly characterized with the exception of DgoT, the structure of which was published recently [101]. All six of the listed proteins transport acidic sugars, and for each one a specific substrate has been proposed [101,103–106,114]. But so far for none of them it has been shown experimentally that they are able to transport the proposed substrates, the exception being DgoT [101]. For this study, to further characterize the different ACS transporters in *E. coli*, the workflow was as following: (i) A suitable expression and purification protocol for the transporters had to be developed, (ii) the functional characterization was established by an *in vivo* growth assay, an *in vitro* ligand binding assay and a liposome-based transport assay, and (iii) the structural characterization of the transporters by X-ray crystallography.

5.2.1. Expression and purification of proteins

The coding genes for each of the six transporters were cloned into pNIC-CTHF vectors and, except for DgoT, into pNIC28-Bsa4 vectors for expression. The pNIC-CTHF vector adds a His-tag and FLAG-tag C-terminally to the protein and the pNIC28-Bsa4 vector adds a His-tag to the N-terminus of the protein. In both vectors the tags are cleavable by TEV protease [178] (Figure 40a-b). Expression was tested with immunoblotting against the His-tag of the constructs. With the exception of GarP, all constructs cloned into the pNIC28-Bsa4 vector could be expressed. All six transporters could be expressed in the pNIC-CTHF vector. Therefor the pNIC-CTHF constructs were chosen for subsequent experiments (Figure 40c).

The purification protocol for the investigated ACS transporters was the same as already described for DtpA in chapter 5.1.1.2. With this purification protocol, four out of the six transporters could be purified to reasonable amounts of 1.5 mg to 2.2 mg of protein per liter of cell culture. In the detergent DDM, DgoT, LgoT, GarP and GudP showed homogenous SEC elution profiles (Figure 40d and Figure 56 Appendix). The obtained material was then used to test ligand binding by nanoDSF as part of the functional characterization and for crystallization experiments. For the liposome-based uptake assays, DgoT, LgoT, GarP and GudP were purified in DM (Figure 40e and Figure 56 Appendix). To reconstitute the detergent-solubilized transporters into liposomes, the detergent has to be removed from the solution. Any residual detergent would make the liposome membrane more permeable for ions and thus dissipate the membrane potential. This has to be avoided since the conditions under which the transport of sugars is monitored, require a membrane potential. The high critical micelle concentration (CMC) of DM allows the removal via dialysis [176,195].

In addition, LgoT and DgoT were also purified in other detergents for crystallization experiments. In the case of DgoT, the detergents used were LMNG, DM, NM and NG (Figure 56 Appendix). In the case of LgoT, NM and a mixture of DDM and LDAO was used (Figure 40f-g). The aim of decreasing micelle size when moving from DDM to shorter chain detergents like DM or NM, is to facilitate better crystal packing and thus better diffraction quality of the resulting protein crystals [131,196].
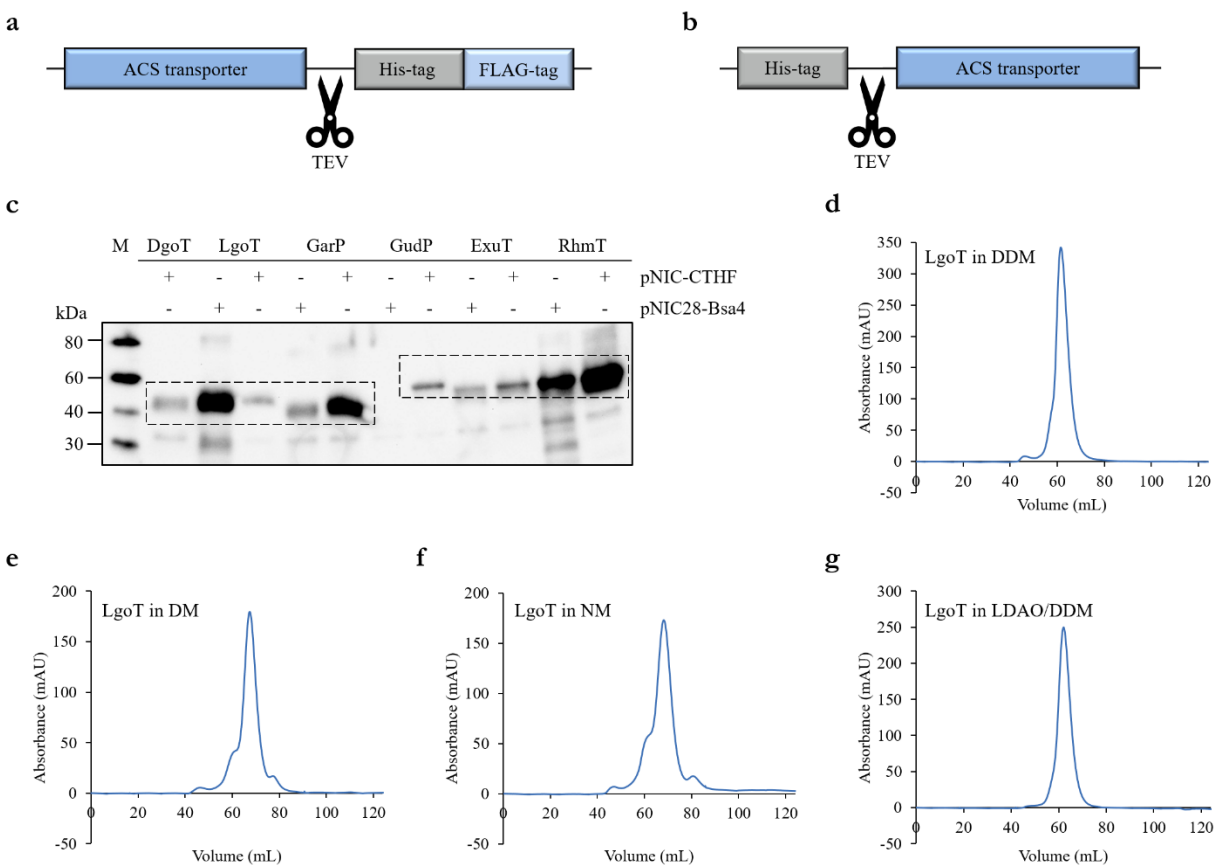
Figure 40: Expression and Purification of ACS transporters from _E. coli._ (a) Schematic of the pNIC-CTHF (pNIC) construct for the expression of ACS transporters. Here, the protein is expressed with a C-terminal His-tag and FLAG-tag. Both tags are cleavable by TEV protease, resulting in a protein sample without His-tag after purification. (b) Schematic of the pNIC28-Bsa4 construct for the expression of ACS transporters. Here, the protein is expressed with an N-terminal His-tag. The tag is cleavable by TEV protease, resulting in a protein sample without His-tag after purification. (c) Western blot analysis to detect the expression for the ACS transporters from _E. coli_ in both the pNIC-CTHF and pNIC28-Bsa4 construct. The membrane was stained with anti-His-tag antibodies. Besides the GudP pNIC28-Bsa4 construct, all other constructs are expressed. (d) SEC profile of LgoT in DDM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for ligand binding studies and crystallization trials. (e) SEC profile of LgoT in DM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for reconstitution into liposomes and crystallization trials. (f) SEC profile of LgoT in NM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for crystallization trials. (g) SEC profile of LgoT in a LDAO/DDM mixture. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for crystallization trials.

5.2.2. Functional characterization

The functional characterization of the _E. coli_ ACS transporters was divided into three parts. First, an _in vivo_ growth assay was performed. Here, knockout strains for all six _E. coli_ ACS transporters were tested for their ability to grow on different sugars and sugar derivatives as sole carbon source. Second, with the four transporters that could be purified, namely LgoT, DgoT, GarP and GudP, an _in vitro_ ligand binding assay using nanoDSF was carried out. Here, the binding of a library of different sugars and sugar derivatives to the transporters was tested. The library was composed in such a way that it

included a range of different sugars like monosaccharides, disaccharides, a trisaccharide and a polysaccharide, pentoses and hexoses, ketoses and aldoses, as well as different sugar derivatives like sugar alcohols, sugar acids and amino sugars (Table 7).

Table 7: Sugar and sugar derivatives used for the functional characterization of *E. coli* ACS transporters.

Compound name	Compound group	Remarks
Agarose	Sugar	Polysaccharide from galactose
Arabinose	Sugar	Monosaccharide, pentose, aldose
Dulcitol	Sugar alcohol	Derivative of galactose
Fructose	Sugar	Monosaccharide, hexose, ketose
Fucose	Sugar	Monosaccharide, hexose, aldose
Galactarate	Sugar acid	Aldaric acid of galactose
Galactonate	Sugar acid	Aldonic acid of galactose
Galactosamine	Amino sugar	Derivative of galactose with the C2 hydroxyl group replaced by an amino group
Galactose	Sugar	Monosaccharide, hexose, aldose
Galacturonate	Sugar acid	Uronic acid of galactose
Glucarate	Sugar acid	Aldaric acid of glucose
Gluconate	Sugar acid	Aldonic acid of glucose
Glucosamine	Amino sugar	Derivative of glucose with the C2 hydroxyl group replaced by an amino group
Glucose	Sugar	Monosaccharide, hexose, aldose
Glucuronate	Sugar acid	Uronic acid of glucose
Inositol	Sugar alcohol	Carbocyclic sugar
Lactose	Sugar	Disaccharide from galactose and glucose
Lyxose	Sugar	Monosaccharide, pentose, aldose
Maltose	Sugar	Disaccharide from glucose
Muramic acid	Sugar acid	Ether of lactic acid and glucosamine
N-Acetyl-galactosamine	Amino sugar	Derivative of galactose with the C2 hydroxyl group replaced by an acetylated amino group
N-Acetyl-glucosamine	Amino sugar	Derivative of glucose with the C2 hydroxyl group replaced by an acetylated amino group
Raffinose	Sugar	Trisaccharide from galactose, glucose and fructose
Rhamnonate	Sugar acid	Aldonic acid of rhamnose
Rhamnose	Sugar	Monosaccharide, hexose, aldose
Ribose	Sugar	Monosaccharide, pentose, aldose
Sialic acid	Sugar acid	N-Acetylneuramic acid

Sorbitol	Sugar alcohol	Derivative of glucose
Sucrose	Sugar	Disaccharide from glucose and fructose
Trehalose	Sugar	Disaccharide from glucose
Xylitol	Sugar alcohol	Derivative of xylose
Xylose	Sugar	Monosaccharide, pentose, aldose

With the proposed substrates and ligands that were shown to stabilize the transporters in nanoDSF measurements, liposome-based uptake assays were performed to distinguish between molecules that were binding the transporters and such that are being transported.

5.2.2.1. *In vivo* growth assay

Besides glucose, *E. coli* has been shown to grow on several different sugar derivatives as sole carbon source [106,111–113]. Among those are the sugar acids D-galactonate, L-galactonate, glucarate, glucuronate, galacturonate, and gluconate [106,111–113]. For the *in vivo* growth assay, *E. coli* strains from the KEIO collection were grown on a range of 23 different carbon sources. The strains of the KEIO collection are based on the *E. coli* K-12 strain [197]. As such, they are not able to utilize sucrose, raffinose, xylitol and myo-inositol as carbon source [198–201]. The strains used were BW25113, the parental strain, and the knockout strains for the individual transporters, JW5859-1 (ΔdgoT), JW4319-1 (ΔlgoT), JW3096-4 (ΔgarP), JW2760-1 (ΔgudP), JW3064-1 (ΔexuT), JW2204-2 (ΔrhmT) [197]. The cells were initially grown in a minimal medium with glucose as carbon source. After washing the cells, to remove glucose from the buffer, all samples were diluted to the same optical density and a dilution series was spotted onto agarose plates with only a specific sugar derivative as carbon source (Figure 41). Glucose serves as positive control, as all strains should be able to grow on this carbon source (Figure 41 and Table 8). The parental strain carries mutations which prevent it from metabolizing arabinose, lactose or rhamnose, thus inhibiting growth when these sugars are the sole carbon source. Subsequently, all knockout strains derived from this parental strain are also not able to grow on these sugars. Additionally, agarose was tested that it does not on its own support growth of the cells (Figure 41 and Table 8).

In the assay, as expected, neither the parental nor the knockout strains show growth on lactose, rhamnose and arabinose due to the mutations and additionally no growth on myo-inositol, raffinose, sucrose and xylitol due to the origin from *E. coli* K-12 (Table 8). In contrast, on glucose and several other sugars and sugar derivatives, such as the monosaccharide galactose, disaccharide maltose, the sugar acid gluconate, the amino sugar N-acetyl-glucosamine and the sugar alcohol sorbitol, all tested strains grew well (Table 8).

While the tested strains show the same growth pattern (growth or no-growth) for almost all tested carbon sources, differences are mainly observed for the sugar acids D-galactonate, L-galactonate, galactarate, glucarate, galacturonate and glucuronate (Table 8). In detail, ΔdgoT does not grow on D-galactonate, ΔlgoT does not grow on L-galactonate, ΔexuT does not grow on galacturonate and glucuronate and ΔgarP does not grow on glucarate (Figure 41 and Table 8). Additionally, the growth of ΔgarP on galactarate and ΔgudP on glucarate is impaired compared to the other strains, although not completely abolished (Figure 41 and Table 8).

If the knockout strain of a transporter is not able to grow on a specific compound but the other stains do grow, this suggests that the compound is a potential substrate for this transporter. Furthermore, it also hints that this transporter is

the only transporter of this compound, as there is no other system to rescue the knockout of this gene. Therefore, DgoT is potentially the transporter of D-galactonate, LgoT potentially the transporter of L-galactonate, ExuT potentially of galacturonate and glucuronate, and GarP potentially of glucarate. For DgoT, LgoT and ExuT, these findings confirm what has been predicted or shown before [101,105,106].

Although the role of GarP was suggested to be the uptake of glucarate, the protein was predicted to be a galactarate transporter as well [103,114]. As sole galactarate transporter in *E. coli*, the expected outcome would have been that Δ*garP* grows on glucarate but not on galactarate. In contrast, I observed, that Δ*garP* does not grow on glucarate and shows impaired growth on galactarate. As for GudP, which is predicted to transport glucarate, the knockout was expected to grow on galactarate and not on glucarate [103,114]. Although, normal growth is observed on galactarate as expected, only little growth impairment is observed with glucarate. To interpret these findings it is important to know that both, glucarate and gluconate, share the same metabolic pathway albeit with the small exception of the dehydrogenase [103,114]. This mutual metabolic pathway explains why, although the genes for GarP and GudP are in two different operons, they are found to be induced by both galactarate and glucarate. The induction by galactarate is stronger than by glucarate [119]. Recognizing this relationship, the findings can be interpreted as follows. GarP is the main transporter for galactarate in *E. coli*. As long as the *garP* gene is present, *E. coli* can grow normally when galactarate is a carbon source. In the case of Δ*garP*, the induction and subsequent expression of GudP can compensate for the loss of GarP albeit with limitations. This leads to the observation of impaired growth. Therefore, while GarP is possibly the main transporter with GudP being able to transport galactarate as well.

A similar explanation can be given for the growth pattern observed with glucarate (Figure 41 and Table 8). When no GudP is present, the growth of the cells on glucarate is impaired (Figure 41 and Table 8). This change in growth shows that GudP plays a part in the uptake of glucarate. Since growth of the knockout strain is not completely abolished, there seems to be another transporter which can, at least in part, rescue the Δ*gudP* phenotype. This transporter appears to be GarP. Δ*garP* does not grow on glucarate, although GudP should be expressed as explained above. The transport of glucarate by GudP alone does not seem to be sufficient for cell growth. GarP might be the main transporter for glucarate as well, but GudP is also able to transport glucarate. Taken together, both GarP and GudP, are potentially transporters for galactarate as well as glucarate.

No conclusion could be drawn from this assay with regard to the substrate of Δ*rhmT* as none of the strains were able to grow on rhamnonate. The reason might be that since the cells lack the metabolic enzymes to break down rhamnose, they might have also lost the ability to metabolize the corresponding aldonic acid, rhamnonate. Perhaps some metabolic enzymes are overlapping and are thus no longer expressed in any of the tested strains, including the parental strain. Therefore, to be able to perform this assay for Δ*rhmT*, a new parental strain is needed. This parental strain has to be able to express the metabolic enzymes for the breakdown of rhamnonate. Starting from this parental strain, a new Δ*rhmT* then would be generated.

Interestingly, while Δ*dgoT* and Δ*lgoT* do not grow on D- and L-galactonate, the aldonic acid of galactose, respectively, there is no growth impairment for any of the tested ACS transporters observed for gluconate, the aldonic acid of glucose. In *E. coli* several transporters for gluconate exist, but although they are MFS transporters they do not belong to the ACS family and were thus not tested here [202].

Table 8: *In vivo* growth assay of *E. coli* ACS transporter knockout strains on different carbon sources. Green denotes growth, yellow denotes impaired growth and red denotes no growth.

Carbon source	parental	ΔdgoT	ΔexuT	ΔgarP	ΔgudP	ΔlgoT	ΔrhmT
Agarose	red	red	red	red	red	red	red
Arabinose	red	red	red	red	red	red	red
Fructose	green	green	green	green	green	green	green
Galactarate	green	green	green	yellow	green	green	green
D-Galactonate	green	red	green	green	green	green	green
L-Galactonate	green	green	green	green	green	red	green
Galactose	green	green	green	green	green	green	green
Galacturonate	green	green	red	green	green	green	green
Glucarate	green	green	green	red	yellow	green	green
Gluconate	green	green	green	green	green	green	green
Glucose	green	green	green	green	green	green	green
Glucuronate	green	green	red	green	green	green	green
Inositol	red	red	red	red	red	red	red
Lactose	red	red	red	red	red	red	red
Maltose	green	green	green	green	green	green	green
N-Acetyl-glucosamine	green	green	green	green	green	green	green
Raffinose	red	red	red	red	red	red	red
Rhamnonate	red	red	red	red	red	red	red
Rhamnose	red	red	red	red	red	red	red
Ribose	green	green	green	green	green	green	green
Sorbitol	green	green	green	green	green	green	green
Sucrose	red	red	red	red	red	red	red
Trehalose	green	green	green	green	green	green	green
Xylitol	red	red	red	red	red	red	red

Figure 41: *In vivo* growth assay of *E. coli* ACS transporter knockout strains on different carbon sources. Dilution series of a cell suspension for the parental and knockout strains for the ACS transporters from *E. coli* were spotted onto agarose plates containing different carbon sources. The used knockout strains are the knockout strains for *gudP (ΔgudP), garP (ΔgarP), exuT (ΔexuT), rhmT (ΔrhmT), dgoT (ΔdgoT)* and *lgoT (⊥lgoT)*. Here, only selected agarose plates with glucose, agarose, D-galactonate, L-galactonate, glucarate and galactarate are shown. An overview of the growth of the different strains on all tested carbon sources is shown in Table 8.

5.2.2.2. *In vitro* ligand binding

After it was investigated which knockout strain, each targeting a particular ACS transporter, was no longer able to grow on specific sugar acids as carbon source, the ability of the corresponding purified transporter to bind that sugar acid was tested. The same technique was used that has already been described for ligand binding of DtpA mutants for FRET experiments (chapter 5.1.1.4.1.1.2). Again, nanoDSF was used to test a set of compounds for their ability to stabilize the transporters during thermal unfolding. The stabilization of a transporter is interpreted as binding of the compound tested. The tested ligand library consisted of 37 different compounds. In addition to sugars, sugar acids, sugar alcohols and amino sugars, the amino acids alanine, asparagine, aspartate, glutamine and glutamate were tested. The amino acids were included because some of the mammalian homologues like the VGLUTs but also Sialin, were shown to transport aspartate and/or glutamate [82,120].

This assay could only be performed for GarP, GudP, DgoT and LgoT, since ExuT and RhmT could not be purified. Strikingly, the transporters which were tested, showed a pronounced ligand specificity, only showing stabilization with a few of the screened ligands. GarP was stabilized by galactarate (ΔTm = 8.8 °C) and D-galactonate (ΔTm = 3.2 °C), respectively. GudP was stabilized by galactarate (ΔTm = 9.4 °C) and glucarate (ΔTm = 3.8 °C) and DgoT by D-galactonate (ΔTm = 5.5 °C) (Figure 42a-c). Interestingly, LgoT did not show a stabilization effect for any of the tested compounds (Figure 42d). Based on the results from the *in vivo* growth assay, stabilization by L-galactonate had been expected. However, it is still possible that L-galactonate binds to LgoT but does not further stabilize the protein.
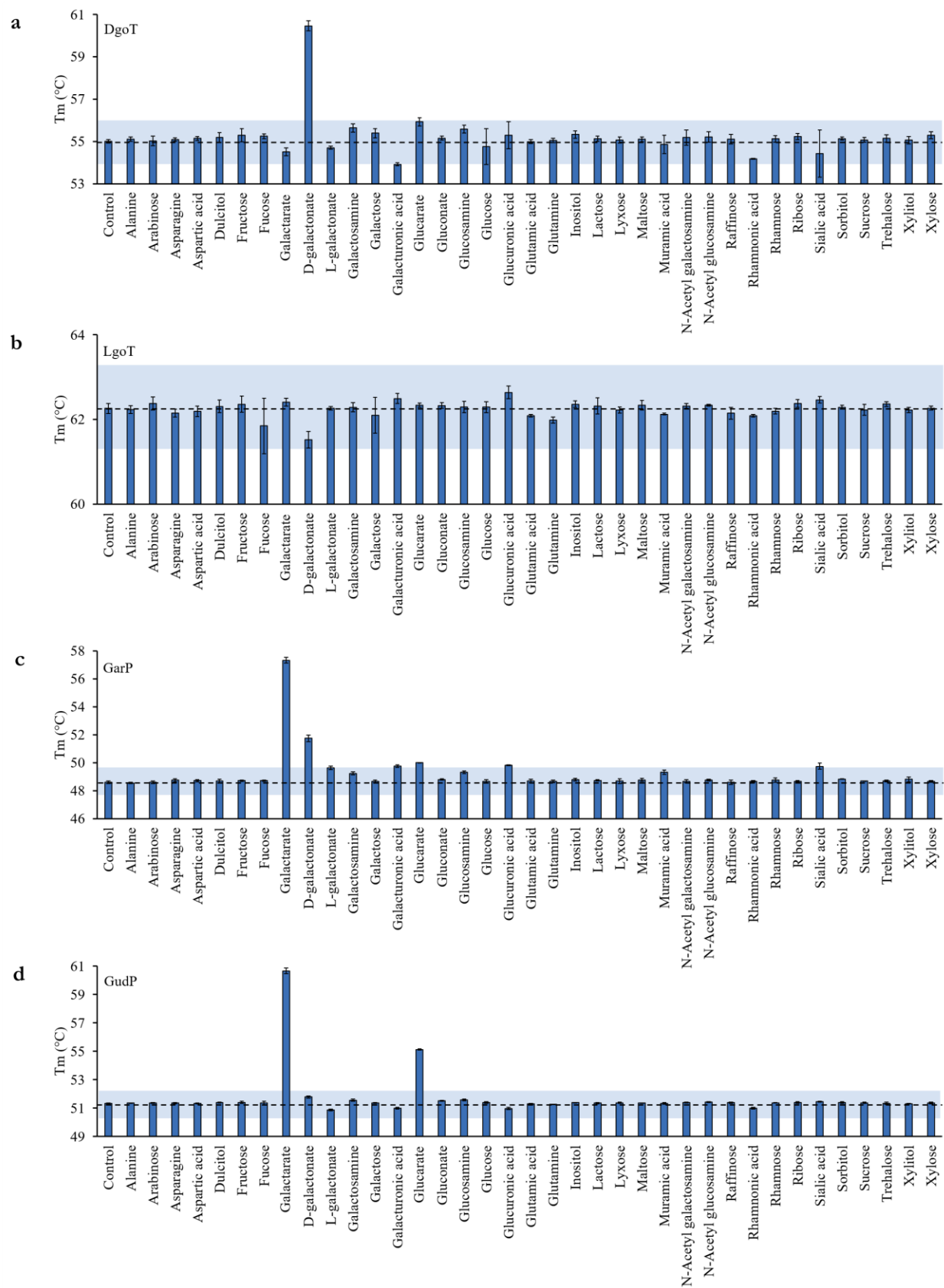
Figure 42: Ligand library screening of ACS transporters from *E. coli*. (a) Thermal stability and ligand binding of DgoT in DDM. (b) Thermal stability and ligand binding of LgoT in DDM. (c) Thermal stability and ligand binding of GarP in DDM. (d) Thermal stability and ligand binding of GudP in DDM. All proteins were measured at 0.5 mg/mL. Ligands were added to a final concentration of 2.5 mM. As a control, water was added instead of ligands. All measurements were done in triplicates. A range of ± 1 °C deviation from the melting temperature of the apo protein is colored in light blue. Changes of the melting temperature within this range were not considered as hits.

The stabilization of GarP by galactarate, GudP by galactarate and glucarate and DgoT by D-galactonate agrees with the results from the *in vivo* growth assay which suggests that these compounds are substrates for the respective transporters. The fact that D-galactonate has a stabilizing effect on GarP could be due to the structural similarity of the two molecules D-galactonate and galactarate. They only differ in the C6 position, where galactarate has a carboxyl and D-galactonate a hydroxyl group. If GarP is able to accommodate galactarate with the larger carboxyl group in its substrate binding site, the smaller D-galactonate might also be able to bind. Glucarate, which was suggested to be a substrate for GarP based on the *in vivo* growth assay, does not show any stabilization effect which could be due to the same reason LgoT does not show stabilization with L-galactonate. Here, it could also be the case that although glucarate binds GarP, it is not stabilizing the protein.

A concentration-dependent stabilization effect could also be observed for selected ligands (Figure 43a-c). For each transporter, those ligands that had shown a stabilization effect larger than 1 °C in the library screen previously, were tested. The dilution series for each ligand ranged from 0 mM to 12.5 mM final concentration of the ligand. DgoT was tested with D-galactonate, GarP with galactarate and D-galactonate and GudP with galactarate and glucarate. The concentration-dependent stabilization observed in all cases further indicates that the initially observed stabilization is due to ligand binding. As for LgoT, no ligand had been found in the ligand library screen, but the *in vivo* growth assay indicated L-galactonate as ligand, a pH screen was performed (Figure 43d). The reasoning here being that L-galactonate might only bind to the transporter at a specific pH because it requires a certain protonation state of key amino acids of LgoT. But no stabilization effect could be observed in the pH range from 6.0 to 8.0 (Figure 43d). As LgoT without any ligand is already relatively stable in DDM, (Tm = 62.3 °C), it could be that the further stabilization effect of a ligand binding is too small to be observed. To test this hypothesis, L-galactonate binding to LgoT that had been purified in different detergent was measured (Figure 43d). In the detergents used, DM, NM and a mix from LDAO and DDM, the melting temperature of the apo protein was lower than in DDM: 45.7 °C, 44.7 °C and 51.1 °C respectively (Figure 43d). Still the addition of L-galactonate to LgoT in any of the detergents did not change the melting temperature (Figure 43d). Next the binding of L-galactonate to LgoT reconstituted into SapNPs was tested (Figure 43d). It was assumed that in detergent the protein adopts a conformation that does not allow the ligand to access the binding site, hence no stabilization effect upon ligand addition could be observed. It has already been shown for other membrane proteins that they are able to adopt different conformation when in a lipidic environment compared to detergent [194]. However, for LgoT in SapNPs made from POPS no stabilization effect by L-galactonate could be observed (Figure 43d). It seems, that if L-galactonate binds to LgoT, this does not induce further stabilization of the protein. Although the results from the *in vivo* uptake assay suggest LgoT to transport L-galactonate, the *in vitro* binding assays did not confirm nor rule out this observation. Further experiments, such as liposome-based uptake assays, are necessary to further investigate this.

Additionally, for ligands that did lead to stabilization of the transporters, it remained to be tested, if those ligands not only bind the proteins, but are also transported by them.
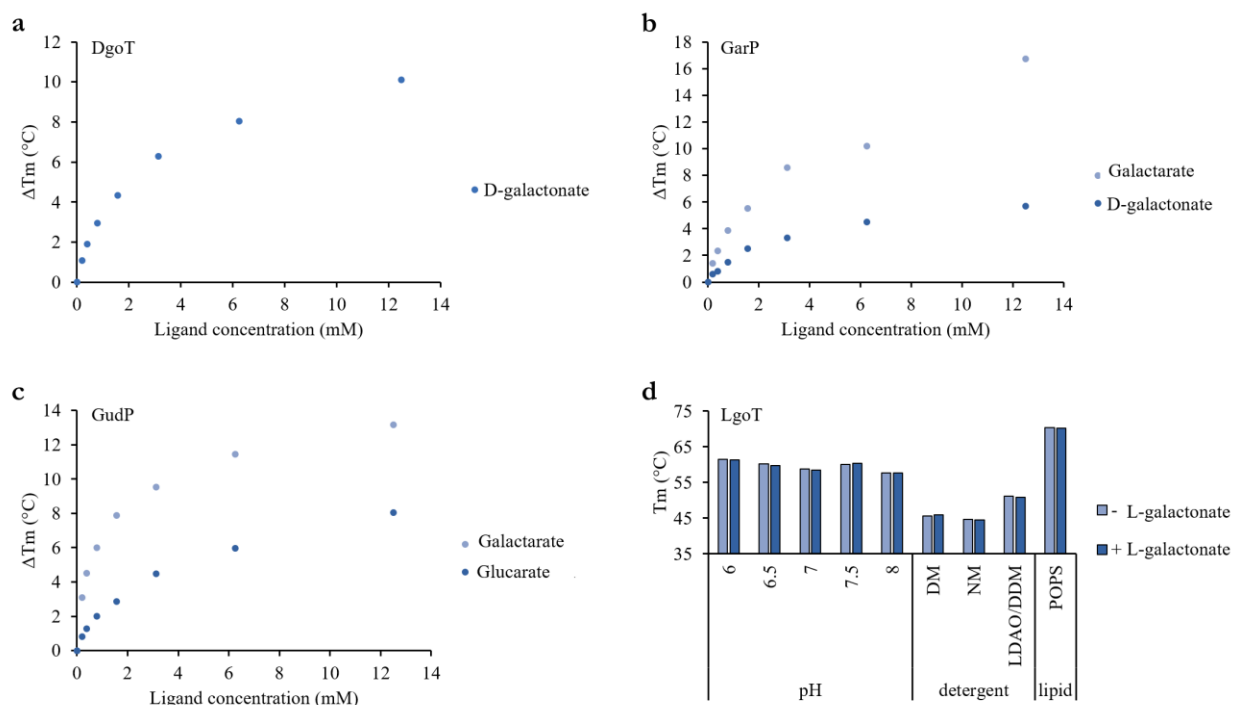
Figure 43: Thermal stability of DgoT, GarP and GudP depending on the ligand concentration and ligand binding of LgoT under different conditions. (a) Thermal stability of DgoT with different concentrations of D-galactonate. (b) Thermal stability of GarP with different concentrations of D-galactonate or galactarate. (c) Thermal stability of GudP with different concentrations of galactarate or glucarate. All proteins were measured at 0.5 mg/mL. A 1:1 dilution series of the ligands were added to a maximal final concentration of 12.5 mM. As a control, water was added instead of ligands. (d) Thermal stability of LgoT with L-galactonate in sodium phosphate buffer of different pH, purified in different detergents or reconstituted into SapNPs with POPS. LgoT was measured at 0.5 mg/mL. Ligand was added to a final concentration of 2.5 mM. As a control, water was added instead of ligand.

5.2.2.3. Liposome-based transport assays

With the sugar acids found to bind to the transporters in the *in vitro* ligand binding assay and the proposed substrates from literature [101,103,105], liposome-based uptake assays were performed. This technique helps to distinguish between ligands that bind to the transporter but are not transported, and substrates which are transported, since both of these types of compounds might show a stabilization effect in nanoDSF measurements.

Liposomes are spherical vesicles composed of a lipid bilayer. This allows different conditions to be generated, e.g. pH or substrate concentrations inside and outside the liposomes, thus creating electrochemical and substrate gradients. Reconstitution of a transporter of interest into these vesicles would allow to study the substrate uptake of a transporter.

Here, liposomes made from a POPE and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG) mixture in a 3:1 (w/w) ratio were used to best resemble the *E. coli* inner membrane composition [189]. Additionally, these liposomes have been shown to be able to withstand a pH gradient of two units [7].

To monitor transport, liposomes are loaded with a pH-sensitive fluorescent dye called pyranine[203]. Pyranine is hydrophilic and thus not able to cross the lipid bilayer when loaded into liposomes [203]. It can be excited at two different wavelength, 415 nm and 460 nm, while the fluorescence is measured in both cases at 510 nm. The signal at the excitation wavelength of 415 nm but not at 460 nm is pH-dependent [203]. Therefore, the ratio of the observed fluorescence of pyranine, when excited at both wavelengths, can be used to measure the internal pH of the liposomes and changes thereof in response to proton influx through transport proteins [203]. Because the *E. coli* ACS transporters are proton-coupled symporters, the transport of their substrates can be monitored by following the movement of protons.

The buffer inside the liposomes has a high potassium concentration (120 mM), whereas the buffer on the outside has the same amount of sodium (120 mM) but no potassium. Thus, a potassium gradient is established but no membrane potential yet, as the same number of cations are found in both buffers (Figure 44). Then, valinomycin is added. Valinomycin is an ionophore selectively allowing potassium, but not sodium, to cross the membrane of the liposomes [204]. Thereby, the potassium gradient is abolished but the sodium gradient is upheld. This generates a membrane potential with a negative charge inside the liposome (Figure 44). This membrane potential can then drive the uptake of protons from the outside buffer into the liposomes through the reconstituted transporters in the presence of substrate (Figure 44). The influx of protons alters the fluorescence of pyranine and can thus be measured [7,203]. Finally, the addition of CCCP abolishes the membrane potential, because CCCP is an ionophore which allows protons to cross the membrane [205] (Figure 44).
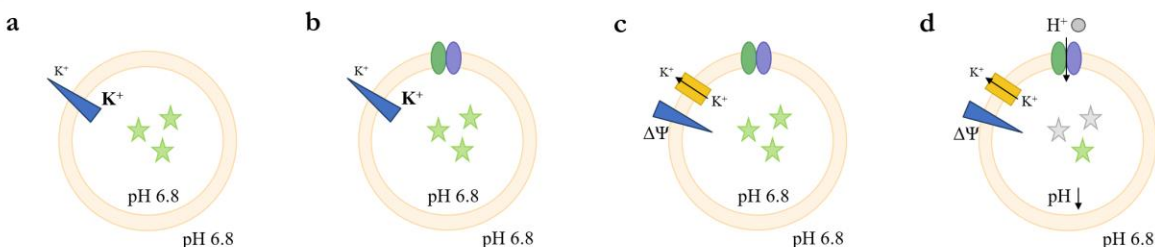
Figure 44: Schematic of a liposome-based uptake assay using pyranine to monitor proton-coupled transport. (a) Liposomes are loaded with fluorescent pyranine dye (green stars). The pH of the buffer inside and outside the liposomes is constant, but the buffer inside contains 120 mM potassium whereas the buffer on the outside contains 120 mM sodium. Thus, a potassium gradient across the membrane is established but no charge gradient. (b) the same set-up as in (a) but here a transporter is reconstituted into the liposomes (green and blue ovals representing the N-terminal and C-terminal domain of an MFS transporter). (c) Addition of the ionophore valinomycin (yellow box) allows for potassium ions to cross the liposome membrane. This leads to abolishing of the potassium gradient which in turn generates a membrane potential ($\Delta\Psi$) with the inside of the liposomes charged negatively. (d) The transporter takes up protons and substrate (gray sphere) in response to the established membrane potential. The influx of protons lowers the pH inside the liposomes which in turn leads to a decrease in pyranine fluorescence (grey stars).

Several control experiments are performed to ensure that the measured change of pH is due to proton-coupled substrate uptake of the transporters and not e.g. through proton leakage or caused by the addition of the substrates. To further

ensure that the presence of the transporter in the liposomes is responsible for the change in pH, the same set of measurements were performed for liposomes with and without reconstituted protein (empty liposomes). Empty liposomes should not exhibit proton-influx in the presence of substrate without the assistance of the respective transporter protein. For liposomes with reconstituted transporters, the change of pH inside the liposomes in the presence of substrate only, valinomycin only or both substrate and valinomycin, is measured. In the first two cases, no pH change should be observed for proton-coupled transport. If only substrate is added, there is no membrane potential generated to drive transport, and if only valinomycin is added, the substrate, which is necessary for symport, is missing. Only if both, valinomycin and substrate, are added, the transporters should be able to move proton and substrate across the membrane, thus changing the pH and causing the pyranine fluorescence to change.

Proton-coupled uptake is observed for D-galactonate by DgoT, L-galactonate by LgoT, galactarate by GarP and galactarate and glucarate by GudP (Figure 45). In contrast, D-galactonate and glucarate, which had been shown to bind to GarP in the *in vitro* binding assay, were not taken up in this assay (Figure 45). As for LgoT, although L-galactonate is transported, the stereoisomer D-galactonate is not a substrate of the transporter, hinting at a similar specificity for this transporter as has already been observed for DgoT (Figure 45).

In summary, while D-galactonate is a substrate for DgoT, L-galactonate for LgoT, galactarate for GarP and galactarate and glucarate for GudP, D-galactonate and glucarate seem only to bind GarP but are not transported under the experimental conditions used here.
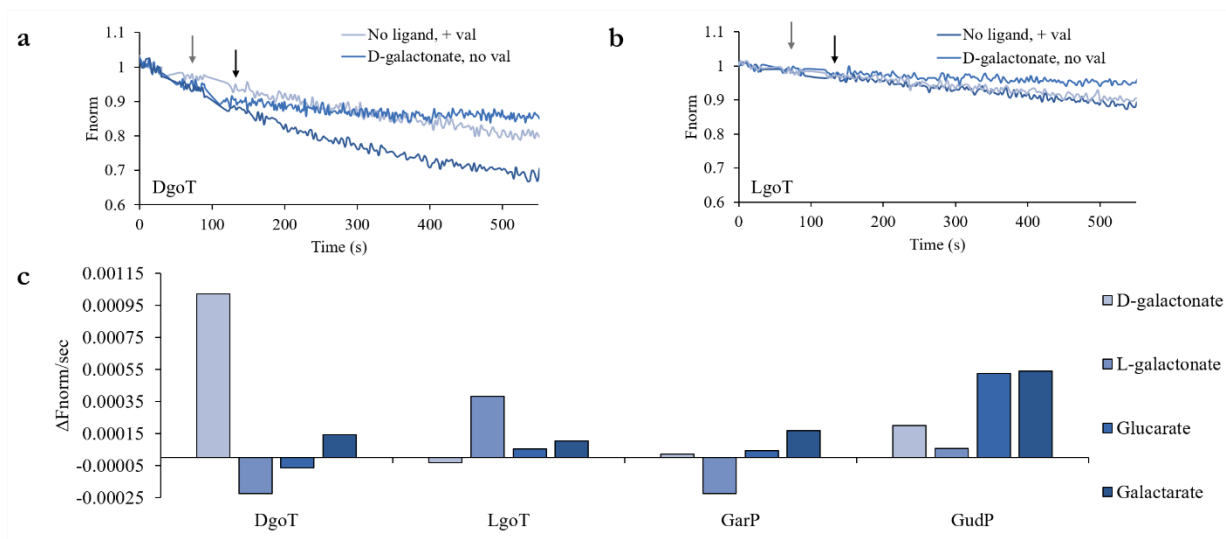
Figure 45: Liposomes-based uptake assay of *E. coli* ACS transporters to monitor proton-coupled transport. Fluorescence trace of pyranine in liposomes with reconstituted (a) DgoT and (b) LgoT with (+ val) and without (- val) the addition of valinomycin in the absence (no ligand) and presence of D-galactonate. Ligands were added to a final concentration of 250 μM (gray arrow). The uptake was initiated by addition of valinomycin to a final concentration of 0.05 μM (black arrow). (c) Changes of the normalized fluorescence of pyranine in liposomes without reconstituted transporters for the first 100 seconds after addition of valinomycin in the presence of ligands, compared to addition of valinomycin to liposomes without reconstituted transporters in the presence of ligands.

5.2.3. Discussion

Six different ACS transporters are annotated in the genome of *E. coli*, DgoT, LgoT, GarP, GudP, ExuT and RhmT. These transporters are proposed to be proton-coupled symporters of sugar acids. With the exception of DgoT, none of these transporters are well characterized and only for DgoT there are high-resolution structures available [101]. To further functionally characterize the ACS transporters from *E. coli* several experiments were performed, the results of which are summarized in Table 9.

Table 9: Summary of the functional characterization of *E. coli* ACS transporters. + means this experiment was successful, - means this experiment was not successful, n.d. means that this experiment was not conducted. For each transporter the sugar acids found to be ligands and therefore potential substrates (nanoDSF) or confirmed substrates (liposome-based assay) are listed, as well as those where the transporter is involved in the metabolic pathway (growth assay).

Protein	Expression	Purification	Metabolic pathway (growth assay)	Ligand binding (nanoDSF)	Substrate transport (liposome-based assay)
DgoT	+	+	D-galactonate	D-galactonate	D-galactonate
ExuT	+	-	Galacturonic acid Glucuronic acid	n.d.	n.d.
GarP	+	+	Galactarate Glucarate	Galactarate D-galactonate	Galactarate
GudP	+	+	Galactarate Glucarate	Galactarate Glucarate	Galactarate Glucarate
LgoT	+	+	L-galactonate	-	L-galactonate
RhmT	+	-	n.d.	n.d.	n.d.

An *in vivo* growth assay was performed for all six transporters and with the four transporters, DgoT, LgoT, GarP and GudP, which could be purified to homogeneity, additionally *in vitro* ligand binding and liposome-based uptake assays were carried out (Table 9).

No further functional or structural characterization of RhmT was possible as this transporter could not be successfully purified. The *in vivo* growth assay was not interpretable because the parental strain did not grow on the test condition. Therefore, a *ΔrhmT* strain in a different parental strain is required.

For ExuT, limited information could be gained. Although this transporter could not be purified, the *in vivo* growth assay strengthens the prediction of galacturonic acid and glucuronic acid as this transporter's substrates. Furthermore, this seems

to be the main transport system in *E. coli* for those sugar acids, as the knockout of *exuT* abolished growth on galacturonic and glucuronic acid (Table 9).

In the case of DgoT, the results from all assays agreed well. This transporter is specific for D-galactonate. Only D-galactonate is a substrate for DgoT. It was found to bind and be transported by the transporter (Table 9). In contrast, the stereoisomer L-galactonate or the aldonic acid from glucose, gluconate, are not found to be transported or to bind to DgoT. This high selectivity can be explained by the coordination of the substrate in the binding site observed in the high-resolution structure [101]. Here, each hydroxyl group of D-galactonate is coordinated which makes its stereoisomer or closely related compounds such as galactonate weaker binders (Table 10).

For LgoT, L-galactonate could be assigned as substrate. Although L-galactonate could not be shown to bind to LgoT in the ligand binding assay, it was transported in the liposome-based uptake assay and also the knockout strain lost the ability to grow on L-galactonate (Table 9). Additionally, LgoT is highly substrate-specific, transporting only L-galactonate but not D-galactonate. L-galactonate binding might not lead to a stabilization of the protein, explaining the lack of stabilization observed by nanoDSF.

For GudP, the results from all experiments agreed well in that its substrates are galactarate and glucarate (Table 9).

Finally, GarP was implicated in the uptake of galactarate and glucarate in the *in vivo* growth assay and shown to bind to galactarate as well as D-galactonate in the *in vitro* ligand binding assay. Nevertheless, only galactarate was found to be a substrate in the liposome-based uptake assay (Table 9). Here, galactarate can be assigned as the substrate of GarP. Although glucarate was proposed to be a substrate for GarP based on the *in vivo* growth assay, this could not be confirmed by the uptake assay. To further investigate if glucarate is a low affinity substrate, the liposome-based uptake should be repeated with higher ligand concentration. D-galactonate was shown to be a ligand, binding to GarP, but not a substrate as it is not transported.

Table 10: Comparison of binding site residues in DgoT, LgoT, GarP and GudP. Numbering according to DgoT as used by Leano et al. [101].

Function in DgoT	Residue no. in DgoT	Residue no. in LgoT	Residue no. in GudP	Residue no. in GarP
Proton-coupling	D46	D57	D35	D33
Proton-coupling	E133	E144	E128	E126
Coordination of the C1 carboxyl group	Y44	Y55	Y33	Y31
Coordination of the C1 carboxyl group	R47	R58	R36	R34
Coordination of the C1 carboxyl group	Y79	Y90	Y68	Y66
Coordination of the C2 and C4 hydroxyl groups	N393	N397	N390	N388
Coordination of the C3 hydroxyl group	Q164	S175	Q159	Q157
Coordination of the C4 and C6 hydroxyl groups	Q264	F266	Q261	Q259
Coordination of the C5 hydroxyl group	S370	T374	A367	A365

For comparison of the substrate binding site and potential residues involved in proton-coupling among all four transporters, sequence alignment with Clustal Omega [206] was performed and homology models based on DgoT [101] were

built using SWISS-MODEL [207–212]. Among the *E. coli* ACS transporters, the proton-coupling residues D46 and E133 in DgoT are highly conserved (Table 10 and Figure 46b). This hints that the transport mechanism is conserved among those four transporters as well. Second, the binding-site residues are organized in two groups. The first one, coordinating the functional groups of the substrates C1 and C2 atoms, is highly conserved as well (Table 10 and Figure 46c). In DgoT these residues are Y44, R47, Y79 an N393. This conservation is not surprising, considering that all substrates of these transporters have a carboxyl group at the C1 and a hydroxyl group at the C2 position. Thus, that part of the molecule is identical for all four substrates, D-galactonate, L-galactonate, galactarate and glucarate. But they differ in the position and type of the functional groups in C3, C4, C5 and C6 position. This explains why the residues coordinating this part of the substrates are not conserved among all transporters (Table 10).

Between DgoT, GarP and GudP, further residues are conserved, namely the ones equivalent to Q164 and Q264 in DgoT. Only the residue equivalent to S370 in DgoT is replaced by A367 and A365 in GudP and GarP, respectively (Table 10 and Figure 46d-e). All three transporters were shown to transport sugar acids of galactose. But while DgoT transports the aldonic acid, GarP and GudP transport the aldaric acid. This might explain, why they only differ in the residue coordinating the functional group of the C6 atom of the substrate. For the aldonic acid of galactose that is a hydroxyl group, but for the aldaric acid this is a carboxyl group. The binding of D-galactonate by GarP, but not its transport, could thus be due to the smaller size of D-galactonate compared to galactarate. If the binding site of GarP can accommodate galactarate, with its larger carboxyl group, then it might be possible for D-galactonate, with the smaller hydroxyl group, to bind. But due to the smaller size, D-galactonate cannot be coordinated in the same way as galactarate, preventing its transport.

Among those four transporters, LgoT has the smallest number of conserved residues in the substrate binding site (Table 10 and Figure 46f). Of the four transporters, LgoT is the only one transporting an L-enantiomer, L-galactonate, in contrast to the other three transporters whose substrates are the D-enantiomers. Here, not only one hydroxyl group changes its position but L-galactonate as a whole is the mirror image of D-galactonate. Therefore, the binding site of LgoT is quite different from DgoT and LgoT is highly specific for L-galactonate, as it cannot coordinate D-enantiomers in its binding site.
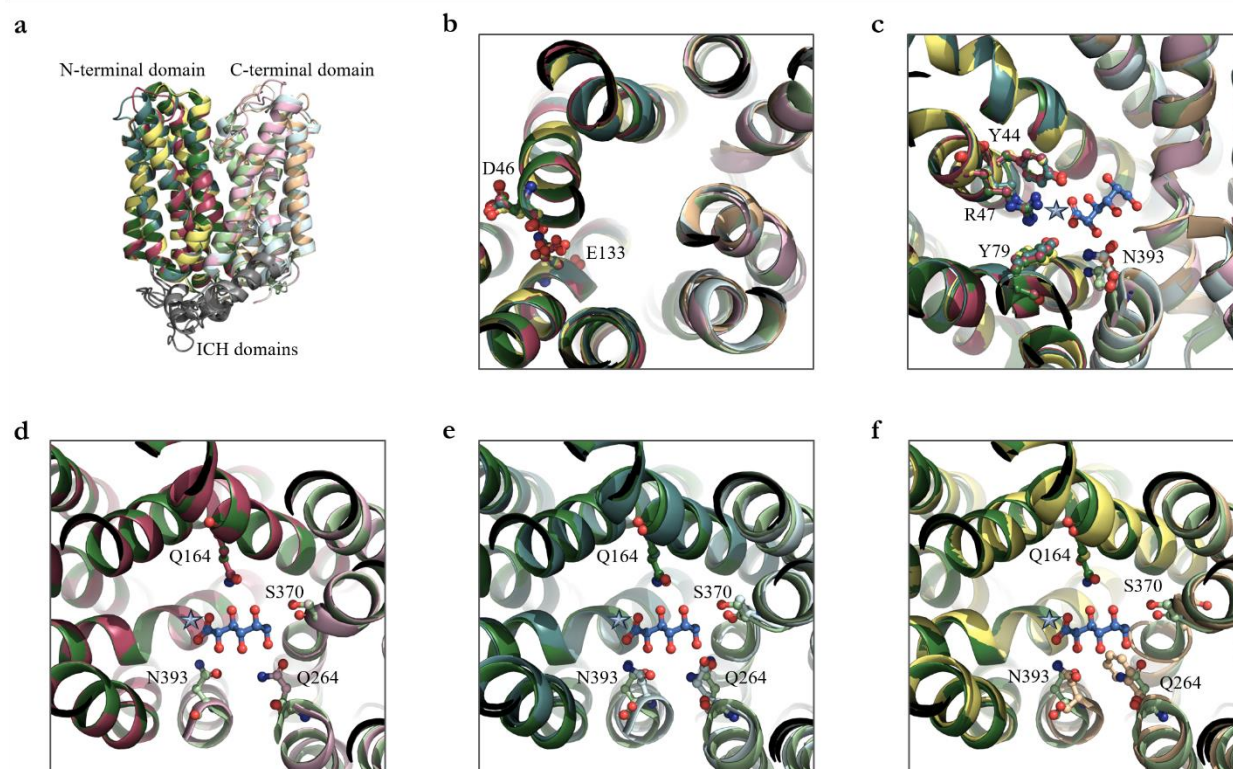
Figure 46: Substrate coordination by _E. coli_ ACS transporters. Numbering according to DgoT as used by Leano et al. [101]. (a) Overlay of the structure of DgoT (green) and the models for LgoT (yellow), GarP (red) and GudP (blue). N-terminal domains shown in darker colors, C-terminal domain shown in lighter colors, the ICH domains are colored gray. (b) Close up on the proton-coupling residues D46 and E133. (c) Conserved substrate coordinating residues between DgoT, LgoT, GarP and GudP. D-galactonate shown in light blue. The star marks the carboxyl group at the C1 atom of D-galactonate. (d-f) Substrate coordinating residues which are not conserved between DgoT, LgoT, GarP and GudP. (d) DgoT and GarP, (e) DgoT and GudP (f) DgoT and LgoT. D-galactonate shown in light blue, the star marks the carboxyl group at the C1 atom of D-galactonate.

Altogether, for four out of six ACS transporters, their substrates could be successfully assigned. D-galactonate for DgoT, L-galactonate for LgoT, galactarate and possibly glucarate for GarP as well as galactarate and glucarate for GudP. It seems that all _E. coli_ ACS transporters have a narrow substrate-specificity range, transporting only up to two structurally related compounds, e.g. differing in the position of a hydroxyl group, as is the case for GudP and possibly GarP and ExuT or transporting only one substrate, as is the case for DgoT and LgoT.

6. Outlook

6.1. Dynamics of POTs

POTs are members of the MFS, one of the largest family of transporters in nature [8,40]. They are proton-coupled transporter of di- and tripeptides [19,40]. Especially the human POTs PepT1 and PepT2 are of pharmacological interest, as they have been shown to transport prodrugs in addition to their natural substrates [12]. In the absence of high-resolution structures for mammalian POTs, bacterial POTs are used as a proxy to study the transport mechanism. Among the bacterial POTs, DtpA exhibits a remarkably similar substrate specificity compared to human PepT1 [48].

Studying the conformational dynamics of DtpA with smFRET, shows that the conformational states that the transporter is able to adopt strongly depend on its surrounding. In detergent solution, DtpA mainly adopts the inward-open state, with a fraction of the molecules in the occluded state. Strikingly, only when reconstituted in a lipid environment, is the transporter able to adopt the outward-open conformation as well, and thus all functionally relevant states. The ratio of transporter in different conformational states is dependent on the chemistry of the headgroup of the lipid used.

To complement the data that was recorded for the cytoplasmic side, labels are placed on the periplasmic side. Here, the placement of the fluorophores is somewhat more difficult as there are more potentially quenching tryptophan residues on the periplasmic side than on the cytoplasmic of DtpA. In the end, two mutants are suitable to work with: L52C/D449C and D111C/D449C (Figure 25 - Figure 27 and Table 6). Inverse to the situation with labels on the cytoplasmic side, a high FRET efficiency around 0.6 for D111C/D449C or 0.9 for L52C/D449C would represent a closed periplasmic side and as such correspond to the inward-open or occluded state. A lower FRET efficiency would represent an open periplasmic side and thus be the inward-open state.

In addition, to attach FRET dyes on one side of DtpA, the labeling positions that proved to be viable for labeling on the cytoplasmic and periplasmic side are combined. This generated FRET mutants where dyes reside on opposite sides of the transmembrane helices of DtpA. These mutants will be used to investigate the mutual flexibility of the N- and C-terminal domains with respect to each other. Based on the initial quality control experiments, the FRET mutants D111C/W203C, T351C/W384C and D449C/Q487C will be used for the anticipated experiments (Figure 25 - Figure 27 and Table 6). In detail, D111C/W203C carries both labeling positions in the N-terminal domain, whereas T351C/W384C and D449C/Q487C carry them in the C-terminal domain. When the N- and C-terminal domains move with respect towards each other, but remain fairly rigid within themselves, one FRET population is expected for the investigated FRET mutants. Furthermore, one FRET mutant that measures conformational changes across the membrane was designed, D111C/T351C (Figure 25 - Figure 27 and Table 6). For this mutant, the labels are on either side of the membrane and in different domains, D111C in the N-terminal domain and T351C in the C-terminal domain. All above mentioned variants can be purified, labeled and reconstituted into SapNPs (Figure 51 - Figure 53). SmFRET measurements for these additional variants will allow a more detailed picture of the conformations DtpA is able to adopt in different environments, especially to differentiate between occluded and outward-open states.

All FRET measurement described in this thesis are done on freely diffusing protein, which allowed us to detect the conformational states present in the sample [156]. Nevertheless, compared to diffusion measurements, where dynamics on a millisecond timescale can be determined, longer timescales even up to minutes are accessible for immobilized samples [156]. Instead of observing different molecules while they diffuse through the confocal spot here, the microscope is focused

on a single immobilized molecule, recording its FRET efficiency over a long period of time, its so-called FRET trajectory [156]. This technique allows to follow a single transporter through its transport cycle which in turn enables us to study how frequently a transport cycle occurs and how long each different state in the transport cycle takes. To this end, the already established FRET mutant W203C/Q487C will be immobilized via Avidin-Biotin interaction on a surface. In short, the FRET mutants are site-specifically biotinylated by adding an AviTag to the protein sequence (Figure 47a). This AviTag is 15 amino acids long and is recognized by the *E. coli* Biotin Ligase BirA [213]. BirA then specifically biotinylates the lysine residue of the AviTag [213]. Thus, each protein will carry a single biotinylation at a specific site in contrast to e.g. chemically biotinylated sample where each accessible lysine residue of the protein can be biotinylated (Figure 47c). Successfully biotinylation is verified by immunoblotting using an anti-biotin antibody coupled to HRP (Figure 47d). Quartz cuvettes are biotinylated as well and treated with avidin followed by an incubation with biotinylated FRET mutants to immobilize the samples (Figure 47b). The resulting Avi-tagged W203C/Q487C can be purified, labelled and reconstituted into SapNPs (Figure 47). The smFRET measurements for immobilized FRET will allow to follow a single transporter through a transport cycle and thus give information on the length and frequency of the transport cycle.
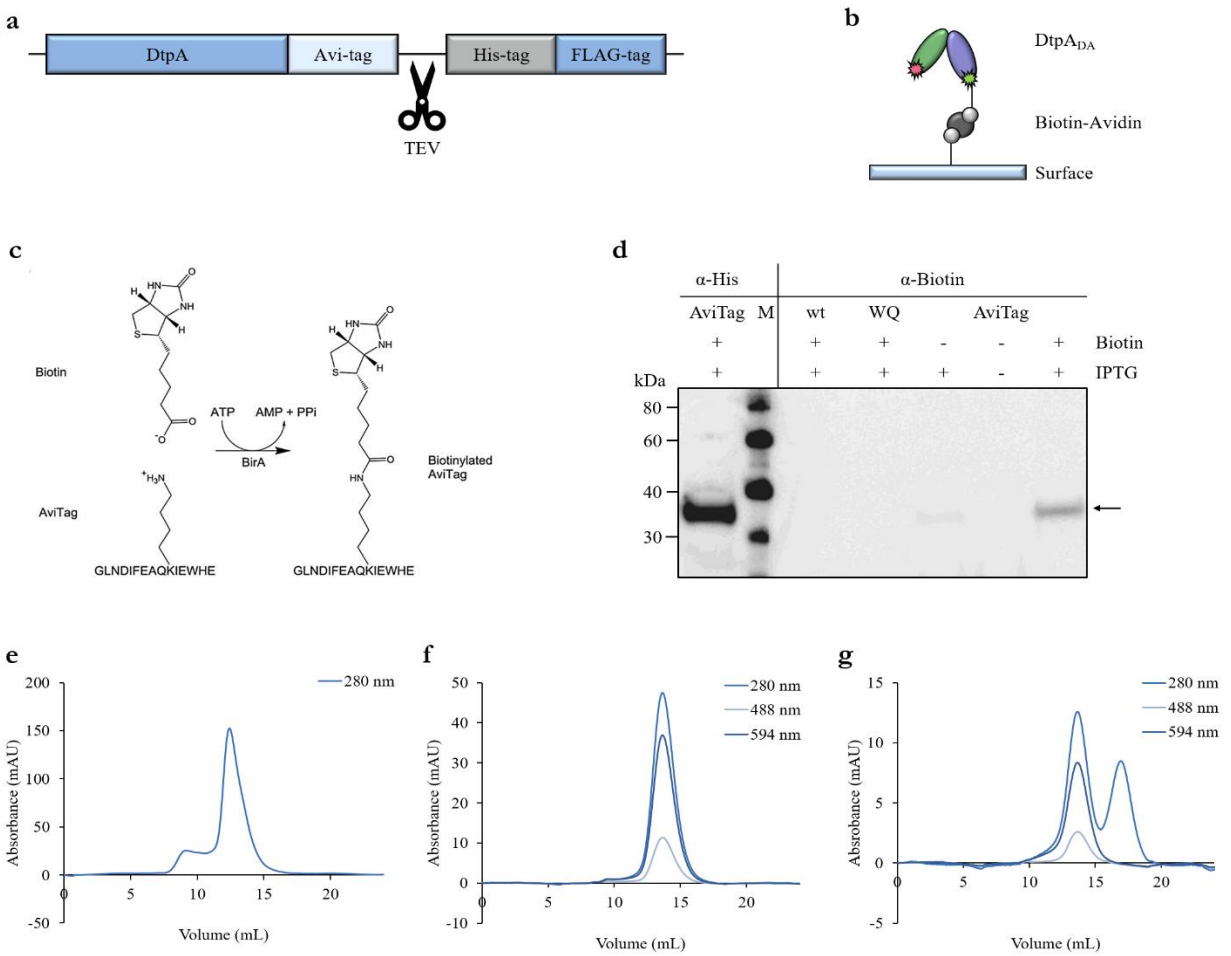
Figure 47: Avi-tagged FRET mutants for *in vivo* biotinylation and immobilization for smFRET measurements. (a) Schematic of the AviTag construct for FRET mutants. DtpA is expressed with C-terminal AviTag, His-tag and FLAG-tag. Between the AviTag

and the His-tag is a TEV protease cleavage site. (b) Schematic of surface immobilization of an Avi-tagged FRET mutant via Biotin-Avidin interaction. The N-terminal domain depicted in green, C-terminal domain in blue. The HaHb domain is omitted. The FRET dyes are shown as stars. Biotin is shown as light gray spheres and Avidin as dark gray spheres. (c) Biotinylation of the AviTag sequence by the BirA Ligase from *E. coli*. The AviTag sequence consists of 15 amino acids. Biotin is attached to the Lys residue of this sequence by the ligase BirA. (d) Western blot analysis to detected *in vivo* biotinylation of the FRET mutant W203C/Q487C with an AviTag (AviTag). Lysed cells were loaded and the membrane was stained with anti-His-tag antibodies (left) and Anti-Biotin antibodies (right). The transporter band is observed at 35 kDa, denoted by an arrow. Only if the expression of both the AviTag and BirA ligase are induced (+IPTG) and Biotin is added (+Biotin) a strong expression of biotinylated AviTag is observed. This is not the case for wildtype (wt) DtpA and the FRET mutant W203C/Q487C as both proteins do not have the AviTag sequence and thus cannot be biotinylated by BirA. (e-g) SEC profiles of (e) FRET mutant W203C/Q487C with AviTag (f) W203C/Q487C Q with AviTag after labeling with Alexa fluorophores and (g) of labeled W203C/Q487C with AviTag reconstituted into POPE SapNPs. Absorbance was measured at 280 nm for the protein and at 488 nm and 594 nm for the donor and acceptor FRET dye, respectively.

Based on our finding that the dynamics of DtpA strongly depend on its environment, the aim is then to gradually move from simple systems (detergent-solubilized protein) to more native environments (lipid bilayers in SapNPs). Since POTs utilize an electrochemical gradient to transport molecules across the cellular membrane, a similar system is required, under which it is possible to study the transporter using FRET. Liposomes provide such a two-compartment system. They allow to generate different conditions e.g. different pH or substrate concentrations inside and outside of the liposomes, thus creating electrochemical and substrate gradients. Reconstitution of a transporter of interest into these vesicles would allow for the study of the protein under numerous conditions and thus be a suitable system required for the FRET measurements that were anticipated.

The protein is already successfully reconstituted into POPE/POPG liposomes as described in chapter 5.2.2.3 and detergent was removed using bio-beads [214]. The efficiency of the reconstitution was judged by SDS-PAGE comparing the amount of protein in liposomes to a known concentration of detergent-solubilized sample. Typically for this protein, the reconstitution efficiency was around 56% (Figure 48). After this first reconstitution with unlabeled mutant, the labeled mutant will be reconstituted and used for smFRET measurements.
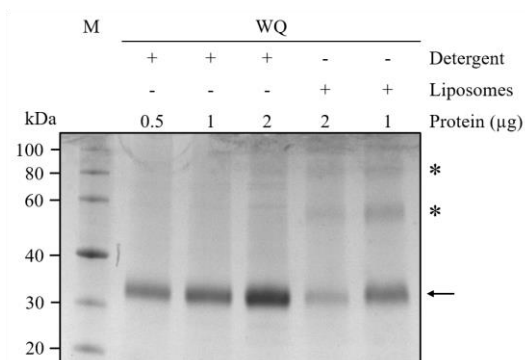
Figure 48: Reconstitution of FRET mutants into POPE/POPG liposomes. SDS-PAGE of W203C/Q487C before and reconstitution into POPE/POPG liposomes. W203C/Q487C ran at 35 kDa (black arrow) and lipids can be observed at 10 kDa (black star). After reconstitution W203C/Q487C shows higher oligomers (black asterisks). Known concentrations of W203C/Q487C in

detergent were run on the gel to compare to the concentration in liposomes. W203C/Q487C proteoliposomes were loaded on the gel such that a 100% efficiency of reconstitution would result into 2 or 1 μg respectively. Judging from the detergent standard the reconstitution efficiency was around 56%.

In addition to studying the dynamics of the transport cycle, our smFRET set-up could also be used to facilitate the structure determination of the outward-open state of DtpA. So far, no structure of a POT has been determined in the outward-open conformation (Table 2). But as drugs bind at this site during their uptake, this is the pharmacologically most relevant conformation. With the aid of a selection protocol for sybodies, which was recently established in the group, in combination with smFRET, this could be approached. Sybodies are synthetic single domain antibodies which can be used as crystallization chaperones [215]. Crystallization chaperones are proteins that bind to their target protein and improve its crystallization behavior [216]. In the case of LgoT sybodies could improve crystal quality by locking the protein in one conformation, making for a more homogeneous sample or by providing additional crystal contacts. This was already observed for other membrane transporters crystallized with the aid of crystallization chaperones, e.g. DtpA and the nanobody N00 [50].

One could select for those sybodies using DtpA in SapNPs, as here the transporter is present in its outward-open conformation. Such generated sybodies could then be used in smFRET measurements. Those sybodies that upon addition to DtpA variants result in stabilization of the outward-open conformation could in turn be used in crystallization experiments.

6.2. Characterization of *E. coli* ACS transporters

Until recently, the members of the ACS family from *E. coli* were poorly characterized. In 2019, the high-resolution structure of DgoT from *E. coli* and in 2020 the high-resolution structure of VGLUT2 from *rattus norvegicus* pointed to key residues involved in substrate coordination and proton-coupling [101,102].

Here, by studying the *E. coli* members of this family, DgoT, LgoT, GarP and GudP could be assigned their substrates. Mutational studies are on-going to test the proposed transport mechanism [101]. The residues proposed to be involved in proton-coupling are conserved among the four transporters. Mutation of those residues is expected to abolish transport but not substrate binding. In addition, mutational studies of the binding-site residues are planned to understand the substrate-specificity.

Furthermore, crystallization trails are carried out on GarP, GudP and LgoT. LgoT is the most promising candidate with crystals obtained under several different conditions using vapor diffusion technique (Figure 49a-b). Nevertheless, the diffraction quality of those crystals did not allow for structure determination (Figure 49c).

To overcome this obstacle two approaches are followed up. First, LgoT is being crystallized using the LCP technique. LCP provides a continuous lipid bilayer for the protein to crystallize in [217]. This typically leads to crystals with better packing and thus better diffraction [217]. First crystals were already successfully grown using LCP (Figure 49d).

Second, crystallization chaperones could facilitate better crystal packing and diffraction quality of LgoT crystals. To this end the possibility to generate sybodies against LgoT could be explored.
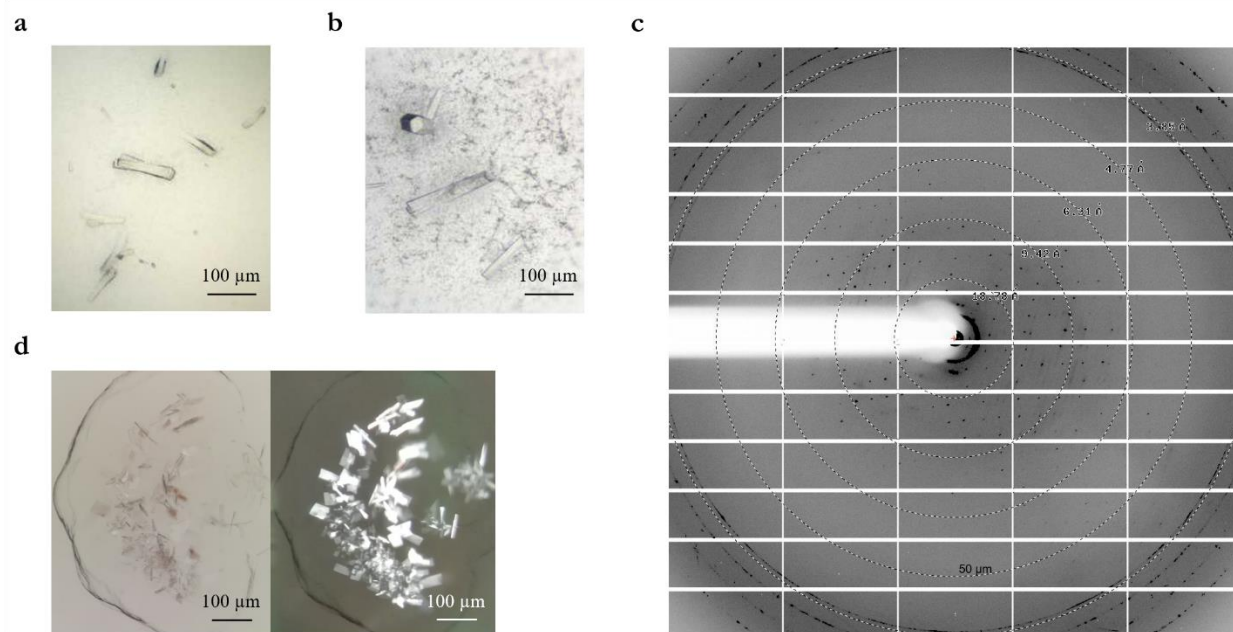
Figure 49: Crystallization of LgoT. (a) Crystals of LgoT from vapor diffusion grown at 19 °C with (a) 39% PEG 400, 0.05 M Glycine pH 9.5, 0.05 M Sodium sulphate and 0.05 M Lithium chloride and (b) 31% PEG 400, 0.05 M Glycine pH 9.5 and 0.15 M Calcium chloride. (c) Diffraction pattern of LgoT grown at 19 °C with 31% PEG 400, 0.05 M Glycine pH 9.5 and 0.15 M Calcium chloride. (d) Crystals of LgoT from LCP grown at 19 °C with 30% PEG 400, 0.1 M Tris pH 7.0, 0.2 M Lithium sulphate.

Additional high-resolution structures of highly substrate-specific ACS transporters could be compared to highly promiscuous transporter families, such as the POT family, to understand which features make a transporter highly specific or highly promiscuous.

7. Materials and methods

7.1. Materials

7.1.1. Chemicals

Table 11: Chemicals used in the described experiments.

Name	Supplier	Cat number
1,4-dithiothreitol (DTT), >99%	Roth	3483-12-3
10 x TBE	Roth	3061.2
2-Propanol	Roth	6752.3
Acetic acid	Roth	3738.4
Agarose for DNA electrophoresis	Serva	11404
Albumin, from Bovine Serum (BSA)	Sigma-Aldrich	A7906
Alexa Fluor 488 C5 maleimide	ThermoFisher	A10254
Alexa Fluor 594 C5 maleimide	ThermoFisher	A10256
Ammonium chloride >99,5%,p.a.,ACS	Roth	K298.1
Anti-biotin HRP-linked Antibody	Cell Signaling Technology	7075S
Avidin from egg white BioUltra, lyophilized powder	Sigma-Aldrich	A9275
Boric acid	Sigma-Aldrich	B0252
Calcium chloride dihydrate >99%,p.a.,ACS	Roth	5239.2
Calcium-D-galactonate hydrate	Carbosynth	FC57272
Carbenicillin disodium salt	Roth	6344.3
Carbonyl cyanide 3-chlorophenylhydrazone (CCCP)	Sigma-Aldrich	C2759
Chloramphenicol	Roth	3886,3
Chloroform,>= 99.8% for analysis	Sigma-Aldrich	288306
Cobalt(II)chloride hexahydrate	Sigma-Aldrich	C8661
cOmplete™ Protease Inhibitor Cocktail	Roche	5056489001
Copper(II)chloride dihydrate	Sigma-Aldrich	307483
D(-)-Fructose	Sigma-Aldrich	F-0127
D-(−)-Ribose ≥99%	Sigma-Aldrich	R7500
D(+) Sucrose	Roth	4621.1
D(+)-Biotin	Roth	3822.1
D(+)-Galactosamine hydrochloride	Sigma-Aldrich	G-1639
D(+)-Galactose >98%	Roth	4987.2
D(+)-Galacturonic acid	Fluka	48280
D(+)-Glucosamine hydrochloride	Sigma-Aldrich	G1514
D(+)-Glucose, p.a., ACS	Roth	X997.2
D(+)-Raffinose pentahydrate	Sigma-Aldrich	R7630
D(+)-Trehalose dihydrate	Fluka	90210
D-(+)-Xylose ≥99%	Sigma-Aldrich	X1500
D-Galactono-1,4-lactone	Carbosynth	MG01323
D-Glucuronic acid sodium salt monohydrate	Sigma-Aldrich	G8645
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D8418
Di-Potassium hydrogen phosphate, >99%, p.a.	Roth	P749.2
Di-Sodium hydrogen phosphate heptahydrate >98%, p.a., ACS	Roth	X987.2
D-Lyxopyranose	Sigma-Aldrich	220477
D-Saccharic acid potassium salt ≥98%	Sigma-Aldrich	S4140
D-Sorbitol	Sigma-Aldrich	6213,2
Dulcitol ≥99%	Sigma-Aldrich	D0256
EDTA Tetrasodiumsalt 86-88%	Roth	3619.1
Ethanol >=99,8%	Roth	9065,3

Ethidium Bromide	Roth	HP471
Ferric citrate	Sigma-Aldrich	F-6129
GelGreen Nucleic Acid Gel Stain, 10000x ,in water	Biotium	41005
Glycerol, Rotipuran >99,5%, p.a.	Roth	3783.1
Guanidine hydrochloride, >99,5%	Roth	0037.1
HEPES, Pufferan >99,5%, p.a.	Roth	7365-45-9
HisProbe-HRP conjugated antibody	ThermoFisher	15165
Hydrochloric acid 32%	Roth	X896.1
Imidazole, >99%, p.a.	Roth	X998.4
Isopropyl-β-d-thiogalactopyranoside (IPTG)	Roth	2316.4
Kanamycin sulfate	Roth	T832.4
L-(-)-Fucose	Sigma-Aldrich	F2252
L-(+)-Arabinose	Sigma-Aldrich	A-3256
Lactose monohydrate	Roth	8921.1
L-Ascorbic acid Sodium salt	Fluka	11140
LB Agar (Lennox)	Roth	X965.1
LB Broth Low Salt Granulated	Melford	GL1703
LDS Sample Buffer (4x), NuPAGE	Novex	NP0008
L-Galactono-1,4-lactone	Carbosynth	MG04084
L-Glutathione, reduced	Roth	6382,2
L-Rhamnonic acid lithium salt ≥90.0% (HPLC)	Sigma-Aldrich	78312
L-Rhamnose monohydrate ≥99%	Sigma-Aldrich	R3875
Magnesium chloride hexahydrate, >99%, ACS	Roth	2189.1
Magnesium Sulphate	Roth	0261.1
Maltose	Sigma-Aldrich	M-5885
Manganese(II)chloride tetrahydrate, >99%, p.a., ACS	Roth	T881.1
MES, Pufferan >99%	Roth	4256.4
Methanol ,for analysis ACS	Roth	4627.2
Methoxypolyethylene glycol maleimide 5000 (PEG-maleimide)	Sigma-Aldrich	63187
Mucic acid 97%	Sigma-Aldrich	M89617
Muramic acid ≥95%	Sigma-Aldrich	M2503
Myo-Inositol	Sigma-Aldrich	I5125
N-Acetyl D-Glucosamine	Sigma-Aldrich	A8625
N-Acetyl-galactosamine	Sigma-Aldrich	A2795
N-Acetylneuraminic acid synthetic, ≥95%	Sigma-Aldrich	A0812
Nickel(II)chloride hexahydrate	Roth	4489.2
Pentane	Merck	60089
Potassium chloride, >99,5%, p.a., ACS, ISO	Roth	6781.1
Potassium dihydrogen phosphate, >99%, p.a., ACS	Roth	3904.1
SDS Pellets	Roth	CN 30.3
Sekusept Plus	Ecolab	104372E
Sodium chloride, >99,5%, p.a., ACS, ISO	Roth	3957.2
Sodium D-galactonate ≥98.0% (TLC)	Sigma-Aldrich	44511
Sodium dihydrogen phosphate monohydrate, >98%, p.a., ACS	Roth	K300.2
Sodium gluconate	Sigma-Aldrich	S2054
Sodium hydroxide, pellets, >99%, p.a., ISO	Roth	6771
Sodium molybdate dihydrate	Sigma-Aldrich	M1003
Sodium selenite	Sigma-Aldrich	S5261
SuperSignal™ West Femto Maximum Sensitivity Substrate	ThermoFisher	34094
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	ThermoFisher	34580
TB powder	Melford	T1510-1000.0
Tetracycline hydrochloride	Roth	HP63.2
Tris buffered saline with Tween® 20 (TBST)	Sigma-Aldrich	91414

TRIS hydrochloride, Pufferan, >99%, p.a.	Roth	9090.3
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Soltec Bio Science	M115
TRIS, Pufferan, >99,9%, Ultra Qualitaet	Roth	5429.3
Valinomycin ≥98% (TLC), ≥90% (HPLC)	Sigma-Aldrich	V0627
Xylitol	Sigma-Aldrich	85,158-2
Zinc sulfate heptahydrate	Sigma-Aldrich	Z4750
β-Ala-(L)-Lys-N-7-amino-4-methylcoumarin-3-acetic acid (AK-AMCA)	Biotrend	BP0352

7.1.1.1. Peptides

Table 12: Peptides used in the described experiments.

Name	Supplier	Cat number
Ala-Ala-Ala (AAA)	Sigma-Aldrich	A9627
Ala-Leu (AL)	Sigma-Aldrich	A1878
Ala-Met-Ala (AMA)	GL Biochem	N/A
Ala-Phe (AF)	Sigma-Aldrich	A3128
Ala-Tyr-Leu (AYL)	GL Biochem	N/A
Gly-Gly (GG)	Sigma-Aldrich	G1002
Gly-His (GH)	Sigma-Aldrich	G1627
H-Ala-Phe-Ala-OH (AFA)	Bachem	4011362
H-Leu-Ala-OH (LA)	Bachem	G-2460
H-Leu-Leu-Ala-OH (LLA)	Bachem	4009785
H-Leu-Leu-OH (LL)	Bachem	4001608
H-Phe-Ala-OH (FA)	Bachem	G-2850

7.1.1.2. Lipids

Table 13: Lipids used in the described experiments.

Name	Supplier	Cat number
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (POPA)	Avanti Polar Lipids	840857
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG)	Avanti Polar Lipids	840457
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE)	Avanti Polar Lipids	850757
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS)	Avanti Polar Lipids	840034
Brain total lipid extract (BL)	Avanti Polar Lipids	131101
Monoolein 9.9 MAG	Molecular Dimensions	MD2-67

7.1.1.3. Detergents

Table 14: Detergents used in the described experiments.

Name	Supplier	Cat number
Lauryl maltose neopentyl glycol (LMNG)	Anatrace	NG310
n-Decyl-β-D-maltoside (DM)	Anatrace	D322
n-Dodecyl-N,N-dimethylamine-N-oxide (LDAO)	Anatrace	D360
n-Dodecyl-β-D-maltoside (DDM)	Anatrace	D310

Name	Supplier	Cat number
n-Nonyl-β-D-glucoside (NG)	Anatrace	N324S
n-Nonyl-β-D-maltoside (NM)	Anatrace	N330

7.1.1.4. Molecular cloning

Table 15: Chemicals used in the described molecular cloning experiments.

Name	Supplier	Cat number
6x DNA loading dye	ThermoFisher	R0611
dCTP	New England Biolab	N0446S
dGTP	New England Biolab	N0446S
dNTPs	New England Biolab	N0447S
Nuclease-free water	Qiagen	129115

7.1.1.4.1. Enzymes

Table 16: Enzymes used in the described experiments.

Name	Supplier	Cat number
Antarctic Phosphatase	New England Biolabs	M0289S
BirA	EMBL Hamburg	N/A
BfuAI	New England Biolabs	R0701S
BsaI	New England Biolabs	R0535S
DNase I	Appli-Chem	P10080A
DpnI	New England Biolabs	R0176S
Lysozyme	Roth	8259.2
Phusion HF DNA Polymerase	New England Biolabs	M0530S
T4 DNA Ligase	New England Biolabs	M0202S
T4 DNA Polymerase	New England Biolabs	M0203S
T4 Polynucleotide kinase	New England Biolabs	M0201S
Taq DNA polymerase	New England Biolabs	M0267S
TEV protease	EMBL Hamburg	N/A

7.1.2. Consumables

Table 17: Consumables used in the described experiments.

Name	Supplier	Cat number
Anti-Penta-HIS (HIS1K) biosensors	fortéBIO	18-5120
Bio-Beads™ SM-2 Resin	Biorad	152-8920
CapureSelect beads for EPEA-tag	ThermoFisher	194288010
Disposable PD 10 Desalting Columns	GE Healthcare	GE17-0851-01
EZ-Link™ NHS-PEG4-Biotin kit	ThermoFisher	21455
Gene Ruler 1 kb DNA ladder	ThermoFisher	SM0311
Greiner CELLSTAR® 96 well plates flat bottom black polystyrene wells	Sigma	M0312
Greiner CELLSTAR® 96 well plates flat bottom clear wells	Sigma	M0812
illustra MicroSpin G-25 Columns	GE Healthcare	27532501
Instant Blue (coomassie based staining solution)	Expedeon	ISB
LiposoFast Liposome factory 400 nm filter	Avestin, Inc.	Z373435-50EA
Loading Dye Purple	New England Biolab	B7024S

Mix & Go! *E. coli* Transformation Kit and Buffer Set	Zymo Research	T3001
Ni-NTA agarose	Invitrogen	R901-15
Novex NuPAGE LDS Sample Buffer	ThermoFisher	NP0008
Prometheus NT.48 Standard grade nanoDSF capillaries	Nanotemper	PK002
QIAquick Miniprep Kit	Qiagen	27104
QIAquick PCR Purification Kit	Qiagen	28106
Roti®-Mark 10-150	Roth	T850.1
RunBlue™ Bis-Tris Protein Gels 4-12% 12 well	Expedeon	NBT41212
RunBlue™ Bis-Tris Protein Gels 4-12% 17 well	Expedeon	NBT41227
SnakeSkin Dialysis Tubing MWCO 10 kDa	ThermoFisher	68100
Spectra/Por 6 Standard Regenerated Cellulose (RC) Dialysis Tubing, Pre-treated, 38 mm Flat Width, MWCO: 2 kD	Spectrum Laboratories	132625
Spin-X® UF 20 mL Centrifugal Concentrator, 10,000 MWCO Membrane	Corning	431488
Spin-X® UF 20 mL/6mL Centrifugal Concentrator, 100,000 MWCO Membrane	Corning	431491/ 431486/ 431481
Spin-X® UF 20 mL/6mL/0.5mL Centrifugal Concentrator, 5,000 MWCO Membrane	Corning	431487/ 431482/ 431477
Spin-X® UF 20 mL/6mL/0.5mL Centrifugal Concentrator, 50,000 MWCO Membrane	Corning	431490/ 431485/ 431480
Streptavidin (SA) biosensors	fortéBIO	18-5019
Strep-Tactin® Sepharose® 50% suspension	iba Lifesciences	2-1201-010
Trans-Blot® Turbo™ Mini PVDF Transfer Packs	Biorad	1704156

7.1.3. Crystallization screens

Table 18: Crystallization screens used in the described experiments.

Name	Supplier	Cat number
JCSG+ Suite	Qiagen	130720
MemAdvantage	Molecular Dimensions	MD1-70
MemGold2	Molecular Dimensions	MD1-64
MemMeso	Molecular Dimensions	MD1-87
MemTrans	Molecular Dimensions	MD1-112
NeXtal Stock Kit Salt	Qiagen	132985

7.1.4. Equipment

Table 19: Equipment used in the described experiments.

Name	Supplier
Agilent 1260 with autosampler and fraction collector	Agilent technologies
ÄKTA Pure with F9C fraction collector	GE Healthcare
Analytical scale	Sartorius
Avanti JXB-26 Centrifuge	Beckmann-Coulter
CaptureSelect™ C-tagXL Pre-packed Column 1 mL	ThermoFisher
Centrifuge 5424 R	Eppendorf
Centrifuge 5810 R	Eppendorf
Electrophoresis chamber for agarose gels	NeoLab
Electrophoresis chamber for SDS gels	Invitrogen

EmulsiFlex-C3 homogenizer	Aventin
Glass homogenizer 55 mL	Wheaton
Freezer -20 °C	Liebherr
Freezer -80 °C	Eppendorf
Fridge 4 °C	Liebherr
Fume hood	Waldner
AnalytikJena UVP Chemstudio	AnalytikJena AG
Heatblock	Eppendorf
JLA 25.50 rotor for Avanti JXB-26 centrifuge	Beckmann-Coulter
JLA 8.1000 rotor for Avanti JXB-26 centrifuge	Beckmann-Coulter
Magnetic stirr plate	Roth
Photometer	Implen
Microscope	Nikon
Microwave	Severin
MilliQ machine	Millipore
MLA-130 rotor for Optima MAX-XP ultracentrifuge	Beckmann-Coulter
MonoQ HiTrap Q HP, 5 ml	GE Healthcare
Mosquito-LCP	ttplabtech
nanodrop 2000c	ThermoScientific
Nanotemper Prometheus NT.48	Nanotemper
New Brunswick™ Innova® 42 Incubator Shaker	Eppendorf
New Brunswick™ Innova® 44 Incubator Shaker	Eppendorf
Octet RED96 System	Molecular Devices
Optima MAX-XP Benchtop Ultracentrifuge	Beckmann-Coulter
Optima XE-90 Ultracentrifuge	Beckmann-Coulter
PCR cycler	Eppendorf
Peristaltic pump	Medorex
pH meter	Mettler Toledo
Power supply	Consort
Rotating wheel	Stuart
Rock imager	Formulatrix
Rotary Evaporator Rotavapor R-114	Büchi
Scales	Sartorius
Superdex® 200 HiLoad 16/600 pg	GE Healthcare
Superdex® 200 Increase 10/300 GL	GE Healthcare
Superdex® 75 HiLoad 16/600 pg	GE Healthcare
Shaking platform	Edward Bühler GmbH
Scorpion Screen Builder	ARI-arts Robbins Instruments
Sonicator waterbath	Elmasonic
LiposoFast Liposome factory	Avestin
TECAN Spark 20M multimode plate reader	TECAN
Ti 45 rotor for Optima XE-90 ultracentrifuge	Beckmann-Coulter
Vortex	Scientific Industries
Heating waterbath	VWR
MiniSpin® centrifuge	Eppendorf
MiniStar Microcentrifuge	VWR

7.1.5. Computational resources

Table 20: Computational resources used in the described experiments.

Name	Source
GIMP 2.8.18	https://www.gimp.org/
Microsoft Office 2019	https://www.microsoft.com/
Multialign	http://multalin.toulouse.inra.fr/multalin/

PDB	https://www.rcsb.org/
Protparam	https://web.expasy.org/protparam/
PyMOL	https://pymol.org/2/
ReverseComplement	http://www.bioinformatics.org/sms/rev_comp.html
Uniprot	https://www.uniprot.org/
SWISS-MODEL	https://swissmodel.expasy.org/
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/

7.1.6. Plasmids

Table 21: Plasmids used in the described experiments.

Name	Supplier	Remarks
pBirAcm	Avidity, LLC	Plasmid for *in vivo* biotinylation of proteins
pMES4y	Vrije Universiteit Brussel	Expressionplasmid nanobodies
pNIC28-Bsa4	EMBL Hamburg	Expressionplasmid membrane proteins and SapsosinA
pNIC-CTHF	EMBL Hamburg	Expressionplasmid membrane proteins
pTH	Karolinska Institutet	Expressionplasmid TEV protease
pTH24	Karolinska Institutet	Expressionplasmid membrane proteins

7.1.7. Bacterial strains

Table 22: Bacterial strains used in the described experiments.

Name	Organism	Supplier	Remarks
AVB99	*E. coli*	Avidity, LLC	Isolating pBirAcm plasmid
BL21 (DE3)	*E. coli*	EMBL Hamburg	Expression strain for TEV protease
BW25113	*E. coli*	KEIO collection	Parental strain from knockout KEIO collection
C41 (DE3)	*E. coli*	EMBL Hamburg	Expression strain for membrane proteins
DH5a	*E. coli*	EMBL Hamburg	Cloning strain
JW2240-2	*E. coli*	EMBL Heidelberg	RhmT knockout strain
JW2760-1	*E. coli*	EMBL Heidelberg	GudP knockout strain
JW3064-1	*E. coli*	EMBL Heidelberg	ExuT knockout strain
JW3096-4	*E. coli*	EMBL Heidelberg	GarP knockout strain
JW4319-1	*E. coli*	KEIO collection	LgoT knockout strain
JW5859-1	*E. coli*	KEIO collection	DgoT knockout strain
Rosetta-gami 2 (DE3)	*E. coli*	EMBL Hamburg	Expression strain for saposin A
WK6	*E. coli*	Vrije Universiteit Brussel	Expression strain for nanobodies

7.1.8. Buffers and media

Table 23: Buffers and media used in the described experiments.

Stock solutions	Composition
10 x M9	475 mM $Na_2HPO_4 \cdot 12\ H_2O$
	220 mM KH_2PO_4
	86 mM g/L NaCl
	187 mM NH_4Cl
TS2+Fe-citrate solution	696 μM $ZnSO_4 \cdot 7H_2O$
	238 μM $MnCl_2 \cdot 7H_2O$
	9.7 mM H_3BO_3

	1.7 mM $CoCl_2 \cdot 6H_2O$
	1.7 mM $NiCl_2 \cdot 6H_2O$
	117 µM $CuCl_2 \cdot 2H_2O$
	7 µM $Na_2MoO_4 \cdot 2H_2O$
	165 µM Na_2SeO_3
	10 mM Fe(III)citrate
20 x MES	1 M MES
	1 M Tris Base
	69 mM SDS
	21 mM EDTA free acid
50 x TAE	2 M Tris Base
	50 mM EDTA
	1 M Acetic acid
10 x TBE	890 mM Tris Base
	890 mM Boric acid
	20 mM EDTA
1 x PBS	137 mM NaCl
	2.7 mM KCl
	10 mM Na_2HPO_4
	1.8 mM KH_2PO_4

Media	**Composition**
LB medium	10 g/L tryptone
	5 g/L yeast extract
	5 g/L NaCl
TB medium	12 g/L tryptone
	24 g/L yeast extract
	9.4 g/L K_2HPO_4
	2.2 g/L KH_2PO_4
	8 mL/L glycerol
LB agar	15 g/L agar
	10 g/L tryptone
	5 g/L yeast extract
	5 g/L NaCl
Minimal medium	100 mL/L 10X M9
	1 mL/L TS2+Fe-citrate solution
	2 mM $MgSO_4$
	100µM $CaCl_2$
	10 mL/L 100x BSE vitamin solution
	2 g/L sugar
Minimal agar	100 mL/L 10X M9
	1 mL/L TS2+Fe-citrate solution
	2 mM $MgSO_4$
	100µM $CaCl_2$
	10 mL/L 100x BSE vitamin solution
	2 g/L sugar
	10 g/L agar

Buffer	**Composition**
Lysis buffer	20 mM NaPi 7.5
	300 mM NaCl
	5 % (v/v) glycerol
	15 mM imidazole
	0.5 mM TCEP
	5 units/mL of Dnase I
	1 mg/mL lysozyme
	1 tablet/100 mL protease inhibitor mix
IMAC wash buffer 1	20 mM NaPi 7.5
	300 mM NaCl

	5 % (v/v) glycerol
	15 mM imidazole
IMAC wash buffer 2	20 mM NaPi 7.5
	300 mM NaCl
	5 % (v/v) glycerol
	25 mM imidazole
IMAC elution buffer	20 mM NaPi 7.5
	150 mM NaCl
	5 % (v/v) glycerol
	250 mM imidazole
SEC buffer POT	20 mM NaPi 7.5
	150 mM NaCl
	5 % (v/v) glycerol
SEC buffer ACS	20 mM HEPES 7.0
	200 mM NaCl
	5 % (v/v) glycerol
SEC buffer SapNP	20 mM NaPi 7.5
	150 mM NaCl
TES buffer	200 mM Tris 8.0
	0.5 mM EDTA
	500 mM sucrose
CaptureSelect elution buffer	20 mM NaPi 7.5
	150 mM NaCl
	5 % (v/v) glycerol
	2 M MgCl2
Octet buffer	20 mM NaPi 7.5
	150 mM NaCl
Liposome buffer	50 mM KPi 7.0
Pyranine inside buffer	5 mM HEPES 6.8
	120 mM KCl
	2 mM $MgSO_4$
Pyranine outside buffer	5 mM HEPES 6.8
	120 mM NaCl
	2 mM $MgSO_4$
AMCA assay buffer	20 mM NaPi 7.5
	150 mM NaCl
	5 mM glucose

7.2. Methods

7.2.1. Molecular cloning

7.2.1.1. Transformation

For transformation, 50 µL of chemically competent cells were mixed with typically 50 ng of expression plasmid and incubated on ice for 5 min before a heat-shock at 42 °C for 45 s and subsequent incubation on ice for 10 min. Cells were grown in 300 µL LB medium at 37 °C for 60 min before spreading on LB agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight.

7.2.1.2. Colony PCR

After mutagenesis, colony PCR is used to screen colonies growing on selective LB agar for those carrying the gene of interest. Here, the reaction was prepared as described in Table 24 with the addition of one colony and run according to the specification in Table 25. The same colony was used to inoculate a 5 mL LB culture supplemented with the appropriate antibiotic incubated at 37 °C overnight. As a control, the wildtype variant of the gene of interest is amplified by PCR as well. The PCR products sizes are compared on agarose gel.

From precultures of colonies that exhibited the right size of insert the plasmid was isolated using a Miniprep kit (Qiagen) and verified by sequencing as described in chapter 7.2.1.5.

Table 24: Reaction mix for colony PCR.

Reagent	Volume
5 x Thermo buffer	2.5 μL
dNTPs (10 mM)	0.5 μL
T7 forward primer	0.5 μL
T7 reverse primer	0.5 μL
Taq polymerase	0.125 μL
ddH$_2$O	Up to 25 μL

Table 25: PCR program for colony PCR.

Step	Temperature	Time	Cycles
1	95 °C	1 min	1
2	95 °C	30 s	
3	55 °C	30 s	30
4	68 °C	2 min	
5	68 °C	5 min	1
6	4 °C	∞	

Table 26: Primer sequences for colony PCR

Primer pair	Sequence 5' – 3'
T7_fw	TAATACGACTCACTATAGGG
T7_rev	CTAGTTATTGCTCAGCGGT

7.2.1.3. Site-directed mutagenesis

To generate point mutations, blunt-end PCR was used for site-directed mutagenesis. In short, the reaction was prepared as described in Table 27 and run according to the specification in Table 28. Primers are designed in such a way that the mutation is split between them, so one primer sequence includes one nucleotide of the mutation while the other includes the other two. Primers are shown in Table 31.

Table 27: Reaction mix for blunt-end PCR.

Reagent	Volume
5 x HF buffer	10 μL
dNTPs (10 mM)	1 μL
forward primer	2.5 μL
reverse primer	2.5 μL
plasmid DNA	5 ng final amount
Phusion polymerase	0.5 μL
ddH$_2$O	Up to 50 μL

Table 28: PCR program for blunt-end PCR. X is the annealing temperature that was modified according to the primer pair used in the PCR reaction as shown in Table 31. The annealing temperature was calculated using the NEB Tm calculator v. 1.12.0.

Step	Temperature	Time	Cycles
1	98 °C	1 min	1
2	98 °C	30 s	
3	X °C	30 s	35
4	72 °C	5 min	
5	72 °C	10 min	1
6	4 °C	∞	

The PCR reaction yields a linearized vector. Before circularization, the template vector, which does not carry the mutation, was digested by addition of 1 μL DpnI to the PCR reaction mix and incubation at 37 °C overnight. The PCR product was then purified using the PCR purification kit (Qiagen), according to the manufacturer's directions.

To circularize the vector, it is first phosphorylated for 2 h at 37C (Table 29), then the product is purified using the PCR purification kit (Qiagen). This is followed by a ligation reaction for 2 h at RT (Table 30). The ligation product is transformed into *E. coli* DH5a cells. Cells are then spread onto LB-agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight. Colonies are screened for insert of the proper size by colony PCR and verified by sequencing.

Table 29: Reaction mix for the phosphorylation of the linearized vector.

Reagent	Volume
PCR product	30 μL
10 x Ligation buffer	4 μL
T4 Polynucleotide kinase	1 μL
ddH$_2$O	5 μL

Table 30: Reaction mix for the ligation of the linearized vector.

Reagent	Volume
PCR product	30 μL
10 x Ligation buffer	4 μL
T4 DNA Ligase	1 μL
ddH$_2$O	5 μL

Table 31: Primers used for site-directed mutagenesis. The mutagen codon is marked in gray.

Protein	Mutation	Primer pair	Sequence 5' – 3'	Annealing temperature
DtpA	K21C	K21C_f	CACGGTTGTTTGAAAGCGTT	63 °C
		K21C_r	CGCGTTCTATCTCATCTTCTCG	
DtpA	L52C	L52C_f	CATTGTTTAACCAGGTAAACAGC	61 °C
		L52C_r	CGGTATGTCTGAAGCGGAT	
DtpA	W107C	W107C_f	CAGGCAACCAGAGCATAACC	63 °C
		W107C_r	CTCTGTTCACGACGCCG	
DtpA	D111C	D111C_f	CAGTGACCAGACCAGGC	59 °C
		D111C_r	CGCCGGTATCGTTTATATG	
DtpA	C140S	C140S_f	GATGTAGAAAGCAGAGAAGACG	59 °C
		C140S_r	GTATGAGAAAAACGACCCG	
DtpA	C200S	C200S_f	GAGAAGGCGAAGTTAACGATA	61 °C
		C200S_r	ACAACGCTGGGTTAAACAGTA	
DtpA	W203C	W203C_f	CAGCGTTGTGAGAAGG	58 °C
		W203C_r	CGTTAAACAGTACGGTTCAA	
DtpA	Y218C	Y218C_f	CAGTTGATAGGCTCGAAGTCT	61 °C
		Y218C_r	CCGTAACCTGCTGCTG	
DtpA	W237C	W237C_f	CAGGTGGCGATAGCGATCA	65 °C
		W237C_r	TCTGCTGCACAATCAGGAAGTT	
DtpA	T351C	T351C_f	ATGGCATCGGCAGGGTAT	61 °C
		T351C_r	GTAAGTTTGCAATCGGCATG	
DtpA	C360S	C360S_f	GACATCACCATGCCGATTGC	64 °C
		C360S_r	ATCTGGTGCGTTCCTGATTCT	
DtpA	W384C	W384C_f	CAGCTTACAGACACGATACCAGC	65 °C
		W384C_r	TCTGGTCGCAAGCTATGGC	
DtpA	D449C	D449C_f	CAGGTAACGTTATCCGGCACAG	66 °C
		D449C_r	CCCGCTGATGTCACTGGAAG	
DtpA	M452C	M452C_f	CACAGCGGATCGGTAACGTTAT	65 °C
		M452C_r	CTCACTGGAAGTCTATGGTCGC	
DtpA	K481C	K481C_f	CACGGCGCGGTCAG	62 °C
		K481C_r	CCTGCACCGCATGACG	
DtpA	Q487C	Q487C_f	CACGTCATGCGGTGCAG	63 °C
		Q487C_r	CGATGACGCTGCAGACAAA	

7.2.1.4. Ligand independent cloning

The wildtype sequence of DtpA was already available in the research group, cloned into the pTH24 vector. This vector possesses two His-tags, an N-terminal one that is cleavable by TEV protease and a C-terminal which is not cleavable. To generate a DtpA construct where the all His-tags can be cleaved off, in order to add a negative IMAC step to the purification protocol, the DtpA sequence was cloned into pNIC-CTHF vector using the ligand independent cloning (LIC) method [178]. Wildtype DgoT cloned into pNIC-CTHF vector was already available at the start of this project. For the remaining ACS transporters LgoT, GarP, GudP, ExuT and RhmT, the genes encoding for them were amplified from genomic DNA, and cloned into pNIC-CTHF and pNIC28-Bsa4 vectors using the LIC method.

For the LIC method first the genes (from here on called Insert) were amplified from genomic or plasmid DNA using the reaction mix described in Table 32 and the PCR protocol in Table 33. The PCR product was isolated using the PCR purification kit (Qiagen).

Table 32: Reaction mix for insert amplification PCR.

Reagent	Volume
5 x HF buffer	10 μL
dNTPs (10 mM)	1 μL
forward primer	2.5 μL
reverse primer	2.5 μL
plasmid DNA	5 ng final amount
genomic DNA	10 ng final amount
Phusion polymerase	0.5 μL
ddH$_2$O	Up to 50 μL

Table 33: PCR program for insert amplification PCR.

Step	Temperature	Time	Cycles
1	98 °C	1 min	1
2	98 °C	30 s	
3	65 °C	30 s	35
4	72 °C	5 min	
5	72 °C	10 min	1
6	4 °C	∞	

The pNIC-CTHF and pNIC28-Bsa4 vectors were prepared to incorporate the insert by removing the sacB gene by restriction digest with BfuAI restriction enzyme for pNIC-CTHF, and BsaI restriction enzyme for pNIC28-Bsa4 as described in Table 34 and Table 35. The sacB gene serves as a negative selection marker. When expressed, the gene product SacB converts sucrose to the polysaccharide levans which is toxic for *E. coli*. The digestion was performed at 50 °C for 2 h and followed by heat-inactivated of the restriction enzymes for 20 min at 65 °C. The cleaved vector was isolated using the PCR purification kit (Qiagen).

Table 34: Reaction mix for pNIC-CTHF digestion by BfuAI.

Reagent	Volume
pNIC-CTHF	5 ng final amount
10 x NEB 3.1 buffer	10 μL
BfuAI	3 μL
ddH$_2$O	Up to 100 μL

Table 35: Reaction mix for pNIC28-Bsa4 digestion by BsaI.

Reagent	Volume
pNIC28-Bsa4	5 ng final amount
10 x NEB 3.1 buffer	10 μL
BsaI	3 μL
ddH$_2$O	Up to 100 μL

The cleaved vector was incubated with T4 DNA Polymerase in the presence of dCTP for pNIC-CTHF and dGTP for pNIC28-Bsa4 respectively. The incubation with T4 DNA polymerase in the presence of excess of one dNTP generates sticky overhangs due to the polymerase 3'→5' -exonuclease activity (Table 36 and Table 37). The insert was treated in a similar way but in the presence of the complementary dNTP used to treat the vector in order to generate matching

overhangs to the vector (Table 38). The T4 DNA Polymerase treatment lasted 30 min at 22 °C before the enzyme is heat inactivated at 75 °C for 20 min.

Table 36: Reaction mix for T4 DNA polymerase treatment of BfuAI cleaved pNIC-CTHF vector.

Reagent	Volume
cleaved pNIC-CTHF	50 μL
10 x NEB 2.1 buffer	10 μL
dCTP	10 μL
T4 DNA Polymerase	5 μL
ddH$_2$O	25 μL

Table 37: Reaction mix for T4 DNA polymerase treatment of BsaI cleaved pNIC28-Bsa4 vector.

Reagent	Volume
cleaved pNIC28-Bsa4	50 μL
10 x NEB 2.1 buffer	10 μL
dGTP	10 μL
T4 DNA Polymerase	5 μL
ddH$_2$O	25 μL

Table 38: Reaction mix for T4 DNA polymerase treatment of insert.

Reagent	Volume
PCR product	50 μL
10 x NEB 2.1 buffer	10 μL
dGTP/dCTP	10 μL
T4 DNA Polymerase	5 μL
ddH$_2$O	25 μL

As a final step, the same volume of treated vector and insert are mixed and after incubation at RT for 20 min transformed into *E. coli* DH5a cells. Cells are then spread onto LB-agar plates supplemented with kanamycin and 5% sucrose and incubated at 37 °C overnight. Colonies are screened for insert of the proper size by colony PCR and send for sequencing to identify the right construct.

Table 39: Primers used for LIC.

Protein	Plasmid	Primer pair	Sequence 5' – 3'
DtpA	pNIC-CTHF	DtpA_C_f	TTAAGAAGGAGATATACTATGTCCACTGCAAACCAAAAA
		DtpA_C_r	GATTGGAAGTAGAGGTTCTCTGCCGCTACGGCTGCTTT
W203C/	pNIC-CTHF	DtpA_Avi_f	TTAAGAAGGAGATATACTATGTCCACTGCAAACCAAAAA
Q487C		DtpA_Avi_r	CTTCAAAGATGTCGTTCAGGCCCGCTACGGCTGCTTTCGCC
Avitag		DtpA_Avi_r2	CTCAGAAAATCGAAGTGCATGAAGCAGAGAACCTCTACTTCCAATCGCA
ExuT	pNIC-CTHF	ExuT_C_f	TTAAGAAGGAGATATACTATGCGTAAAATTAAAGGGTTACGT
		ExuT_C_r	GATTGGAAGTAGAGGTTCTCTGATGTTGCGGTGCGGGATCGTTA
ExuT	pNIC28-Bsa4	ExuT_N_f	TACTTCCAATCCATGCGTAAAATTAAAGGGTTACGT
		ExuT_N_r	TATCCACCTTTACTGTTAATGTTGCGGTGCGGGAT
GarP	pNIC-CTHF	GarP_C_f	TTAAGAAGGAGATATACTATGATTCTGGACACCGTTGACGAA
		GarP_C_r	GATTGGAAGTAGAGGTTCTCTGCTTTCTGCAATTCCATACGT
GarP	pNIC28-Bsa4	GarP_N_f	TACTTCCAATCCATGATTCTGGACACCGTTGACGAA
		GarP_N_r	TATCCACCTTTACTGTTTCTGCAATTCCATACGTTTA
GudP	pNIC-CTHF	GudP_C_f	TTAAGAAGGAGATATACTATGAGTTCTTTAAGTCAGGCTGCG
		GudP_C_r	GATTGGAAGTAGAGGTTCTCTGCTTGCCCCGCAACAGGTTTC
GudP	pNIC28-Bsa4	GudP_N_f	TACTTCCAATCCATGAGTTCTTTAAGTCAGGCTGCG
		GudP_N_r	TATCCACCTTTACTGTTGCCCCGCAACAGGTTTCAAC
LgoT	pNIC-CTHF	LgoT_C_f	TTAAGAAGGAGATATACTATGGAAAAAGAAAATATCACC
		LgoT_C_r	GATTGGAAGTAGAGGTTCTCTGCATCTTTACGTGGGTC
LgoT	pNIC28-Bsa4	LgoT_N_f	TACTTCCAATCCATGGAAAAAGAAAATATCACCATC
		LgoT_N_r	TATCCACCTTTACTGATCTTTACGTGGGTCGTTGATC
RhmT	pNIC-CTHF	RhmT_C_f	TTAAGAAGGAGATATACTATGAGCACCGCTTTGCTTGACG
		RhmT_C_r	GATTGGAAGTAGAGGTTCTCTGATGATGTGCCACGTCGGTCT
RhmT	pNIC28-Bsa4	RhmT_N_f	TACTTCCAATCCATGAGCACCGCTTTGCTTGAC
		RhmT_N_r	TATCCACCTTTACTGATGATGTGCCACGTCGGTCT

7.2.1.5. Sequencing

For sequencing samples were send either to Eurofins Genomics Germany or Microsynth AG and sequenced using the Sanger sequencing method. As sequencing primers standard primers T7 and T7 term provided by Eurofins and Microsynth respectively.

Table 40: Primer pairs used for sequencing.

Primer pair	Sequence 5' – 3'	Supplier
T7	TAATACGACTCACTATAGGG	Eurofins
T7 term	CTAGTTATTGCTCAGCGGT	
T7	TAATACGACTCACTATAGGG	Microsynth
T7 term	TGCTAGTTATTGCTCAGCGG	

7.2.2. Protein expression

7.2.2.1. Expression of membrane proteins

Membrane proteins were expressed and purified as previously described [50,218]. Briefly, the constructs were transformed into *E. coli* strain C41(DE3), and grown in TB medium, supplemented with 30 μg/mL kanamycin at 37 ⁰C. The cells were

induced by 0.2 mM IPTG at OD600 nm of 0.6, incubated further for 16 h at 18 °C, and harvest by centrifugation (9379 xg, 15 min, 4 °C). Cell pellets were either used for purification directly or stored at -20 °C until further use.

7.2.2.2. Expression of biotinylated membrane proteins

Membrane proteins carrying an AviTag were expressed and purified as previously described [50,218]. Briefly, the constructs were co-transformed into *E. coli* strain C41(DE3) with pBirAcm [213], the plasmid coding for the biotin ligase BirA. Cells were grown in TB medium, supplemented with 30 µg/mL kanamycin and 10 µg/mL chloramphenicol at 37 °C. The cells were induced with 0.2 mM IPTG at OD600 nm of 0.6 and the medium was supplemented with 50 µM D-Biotin. Cells were incubated further for 16 h at 18 °C and harvest by centrifugation (9379 xg, 15 min, 4 °C). Cell pellets were either used for purification directly or stored at -20 °C until further use.

7.2.2.3. Expression of nanobodies

The nanobody selection was previously described [218]. The nanobody expression plasmid was transformed into *E. coli* strain WK6. The cells were grown at 37 °C in TB medium supplemented with 100 µg/mL carbenicillin. At an OD600 nm of 0.7 the cells were induced with 1 mM IPTG and incubated for 16 h at 27 °C. Cells were harvested (9379 xg, 15 min, 4 °C). Cell pellets were either used directly for purification or stored at -20 °C until further use.

7.2.2.4. Expression of saposin A

SapA expression plasmid was transformed into *E. coli* strain Rosetta gami-2(DE3). Transformed bacteria were grown at 37 °C in TB medium supplemented with 30 µg/mL kanamycin, 34 µg/mL chloramphenicol and 10 µg/mL tetracycline. At an OD600 nm of 0.7, the cells were induced with 1 mM IPTG and incubated for 4 h at 37 °C. Cells were harvested by centrifugation (9379 xg, 15 min, 4 °C). Cell pellets were either used for purification directly or stored at -20 °C until further use.

7.2.3. Protein purification

7.2.3.1. Purification of membrane proteins

Cell pellets were either used directly after harvesting or when stored frozen, those pellets were thawed for purification. The cells were resuspended in Lysis Buffer supplemented with 0.5 mM TCEP. Typically, 5 mL of Lysis Buffer were used for 1 g of cell pellet. Cells were lysed by using an EmulsiFlex-C3 (Aventin) at 10,000 psi for three cycles. After a low-speed centrifugation (10,000 xg, 15 min, 4 °C) that separated the undisrupted cells and debris, an ultracentrifugation of the supernatant was performed to extract the membrane fraction (142,400 xg, 50 min, 4 °C). The membrane-pellet was resuspended in Lysis buffer, supplemented with 1% (w/v) DDM or LMNG and 0.5 mM TCEP, and gently stirred for 60 min at 4 °C. After additional ultra-centrifugation (104,600 g, 50 min, 4 °C) the supernatant was purified by IMAC using Ni-NTA agarose (ThermoFisher). Typically, 2 mL of a 50% resin in 20% ethanol slurry was added per 1 L of cell culture. His-tagged proteins were bound to the resin for 60 min at 4 °C on a rotating wheel, extensively washed with IMAC wash

buffer 1 and IMAC wash buffer 2 supplemented with detergent (0.01% (w/v) LMNG or 0.03 % (w/v) DDM) and 0.5 mM TCEP and eluted in IMAC elution buffer supplemented with detergent (0.01% (w/v) LMNG, 0.03 % (w/v) DDM) and 0.5 mM TCEP. TEV protease was added to the eluate at 1 mg per 3 L of cell culture. The sample was dialyzed overnight at 4 °C against 750 mL SEC buffer POT or SEC buffer ACS supplemented with detergent (0.01% (w/v) LMNG, 0.03 % (w/v) DDM) and 0.5 mM TCEP for POT and ACS proteins respectively. The next day, negative IMAC was performed to recover the cleaved protein. The sample was incubated for 45 min with 1 mL of a 50% Ni-NTA agarose in 20% ethanol slurry per 1 L of cell culture. SEC was performed on an ÄKTA Pure system (GE Healthcare Life Sciences). Here, protein-containing fractions were loaded on a Superdex 200 Increase 10/300 GL column or Superdex 200 HiLoad 16/600 pg (GE Healthcare Life Sciences) equilibrated with SEC buffer POT or SEC buffer ACS. Both SEC buffers were supplemented 0.5 mM TCEP and detergent. The detergents used were either (0.01% (w/v) LMNG, 0.03 % (w/v) DDM, 0.3 % (w/v) DM, 0.4 % (w/v) NM, 0.25 % (w/v) NG or a mixture from 0.03 % (w/v) DDM and 0.02 % (w/v) LDAO). Fractions containing the protein were pooled and concentrated using 100 MWCO concentrator (Corning Spin-X UF concentrators). Typically, proteins were concentrated to 5 mg/mL.

7.2.3.1.1. Labeling of DtpA variants for FRET measurements

For labeling, the eluate of the IMAC step was recovered, concentrated to 1 mL using a 50 MWCO concentrator (Corning Spin-X UF concentrators) and incubated with 10 mM DTT for 30 min at 4 °C. Protein was loaded on a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences) equilibrated with SEC buffer POT supplemented with 0.01% (w/v) LMNG. Fractions containing the protein were pooled and concentrated to 500 μL. For labeling the variants were incubated for 2 h at RT under gentle agitation with 1:1 molar ratio mixture of Alexa Fluor 488 C5 maleimide and Alexa Fluor 594 C5 maleimide dyes (ThermoFisher) in a 1:2.4 molar ratio of protein to dyes. To stop the labeling reaction the sample was incubated with 1 mM L-Glutathione for 30 min, at RT. Free dyes were removed using a PD-10 desalting column (GE Healthcare Life Sciences). An additional IMAC step was performed, where the eluate from the desalting column was incubated with Ni-NTA agarose for 60 min at 4 °C on a rotating wheel, extensively washed with increasing imidazole concentration from 15 mM to 25 mM, and eluted in IMAC elution buffer. From here the normal purification protocol for membrane proteins proceeded, starting with addition of TEV protease and overnight dialysis.

7.2.3.2. Purification of nanobodies

Cells were resuspended in 5 mL TES buffer per 1 g of pellet. Four-times diluted TES buffer was added to induce an osmotic shock at 4 °C. Cell debris were removed by centrifugation (10,000 xg, 15 min, 4 °C) and the supernatant was recovered and applied on a Capture Select column (ThermoFisher), equilibrated previously with SEC buffer POT and the protein was eluted in CaptureSelect elution buffer. The eluted fractions were loaded on SEC HiLoad 16/600 Superdex 75 pg column (GE Healthcare Life Sciences), equilibrated with SEC buffer POT. Fractions containing protein were pooled and concentrated to typically 10 mg/mL, using a 5 MWCO concentrator (Corning Spin-X UF concentrators).

7.2.3.3. Purification of saposin A

Typically, 1 g of pellet was resuspended in 5 mL Lysis buffer. Cells were lysed by using an EmulsiFlex-C3 (Aventin) at 10,000 psi for three cycles. The lysate was incubated at 80 °C for 10 min. Non-lysed cells, debris and aggregated proteins were removed by centrifugation (10,000 xg, 15 min, 4 °C). The supernatant was purified by IMAC using Ni-NTA agarose. His-tagged samples were bound to the resin for 60 min at 4 ºC on a rotating wheel. 2 mL of a 50% Ni-resin in 20% ethanol slurry per 3 L of cell culture was used. The sample was extensively washed with IMAC wash buffer 1 and IMAC wash buffer 2. SapA was eluted in IMAC elution buffer. TEV protease was added to the eluate, typically 2 mg were used per 3 L of cell culture. The sample was dialyzed overnight at 4 ºC against 1L of SEC buffer SapNP. The next day, negative IMAC was performed to recover the cleaved protein. The sample was incubated for 45 min with 2 mL of a 50% Ni-NTA agarose in 20% ethanol slurry per 3 L of cell culture. The last purification step included a SEC run on a HiLoad 16/600 Superdex 75 pg column equilibrated with SEC buffer SapNP. Fractions containing protein were pooled, and concentrated to typically 2.5 mg/mL, using a 5 MWCO concentrator (Corning Spin-X UF concentrators).

7.2.3.4. Proteins concentration determination

Protein concentrations were determined from the absorbance at 280 nm using the extinction coefficient at 280 nm calculated from the sequence via Expasy Protparam [219]. The proteins were flash-frozen in liquid nitrogen and stored at -80 ºC until further use.

7.2.4. Reconstitution of membrane proteins into lipid environment

7.2.4.1. Reconstitution into saposin-derived lipid nanoparticles

Labeled DtpA variants were reconstituted into SapNPs as previously described, using a 1:20:35 molar ratio of membrane protein-to-SapA-to-lipids [185]. The following lipids were used: POPS, POPE, POPA and BL (Avanti Polar Lipids). Lipids were incubated at 37 ºC for 10 min, membrane protein was added and incubated at RT for 15 min. After addition of SapA, the mixture was incubated at RT for another 20 min. 50 mg of Biobeads (Biorad) were added per 100 μL of sample and the sample was incubated overnight at 4ºC on a rotating wheel. The reconstituted membrane protein was recovered by SEC on a Superdex 200 Increase 10/300 GL column equilibrated with SEC SapNP.

7.2.4.2. Preparation of POPE/POPG liposomes

Liposomes were prepared from a 3:1 (w/w) mixture of POPE and POPG powder (Avanti Polar Lipids). The lipid powder was dissolved into 10 mL chloroform and dried using rotary evaporator while heating the lipid mixture to 35 °C in a water bath. The lipid mixture was washed twice with 10 mL pentane. After each wash the lipids were dried using rotary evaporator while heating the lipid mixture to 35 °C in a water bath. Following the washing steps the lipids were resuspended in degassed liposome buffer to a final lipid concentration of 20 mg/mL. After three freeze-thaw cycles liposomes were stored at -80 °C until further use.

7.2.4.3. Reconstitution of membrane proteins into liposomes

POPE/POPG liposomes were diluted with liposome buffer to 5 mg/mL and extruded 11 times through a LiposoFast Liposome factory 400 nm filter (Avestin, Inc.). Extruded liposomes were mixed with membrane protein at a protein concentration of 0.5 mg/mL in a 1:60 or 1:30 ratio of protein-to-lipids and incubated for 1 h at RT. For empty liposomes the same amount of SEC buffer with detergent, as had been used for the protein sample, was incubated with the liposomes. For membrane proteins solubilized in LMNG, 50 mg/mL Biobeads (Biorad) were added and the sample was incubated overnight at RT. Biobeads were removed and washed with 3 cv of liposome buffer. For membrane proteins solubilized in DM the sample was dialyzed against 3 L liposome buffer at 4 °C using SnakeSkin Dialysis Tubing (ThermoFisher) with a MWCO of 10 kDa. The sample was dialyzed first for 2 h and then three times for 12 h. Proteoliposomes were harvested by centrifugation (100,000 g, 30 min, 4 °C). After two freeze-thaw cycles, liposomes were stored at -80 °C until further use. To determine the efficiency of membrane protein reconstitution into liposomes, the same protein reconstituted into liposomes was loaded on an SDS gel at three different known concentrations as well as the corresponding proteoliposomes at a known volume. Using densitometry, a calibration curve for each protein was determined and the amount in the proteoliposomes calculated.

7.2.5. Gel electrophoresis and blotting

7.2.5.1. DNA electrophoresis

For DNA electrophoresis, a 1 % (w/v) agarose gel was prepared in 1 x TAE or 1 x TBE buffer. Samples were mixed with Loading Dye Purple (NEB) and the gels were run at 110 mV for 55 min in 1 x TAE or 1 x TBE running buffer. Gels were imaged using an AnalytikJena UVP Chemstudio (AnalytikJena AG) system.

7.2.5.2. SDS-PAGE

For SDS-PAGE, 4-12% Bis-Tris gel (Expedeon) were used. Samples were mixed with Novex NuPAGE LDS Sample Buffer (ThermoFisher) supplemented with 71 mM TCEP and the gels were run at 150 V for 45 min in 1 x MES running buffer. Gels were imaged using an AnalytikJena UVP Chemstudio (AnalytikJena AG) system.

7.2.5.3. Western blotting

For western blotting, SDS-PAGE using a 4-12% Bis-Tris gel (Expedeon) was performed and transferred on a PVDF membrane (Biorad). 2% BSA in TBST (Sigma) was used for blocking for 1 h at RT. TBST was used as washing buffer. The membrane was incubated with HisProbe-HRP conjugated antibody (ThermoFisher) or Anti-biotin HRP-linked Antibody (CST) for 1 h at RT. The blot was developed using Super Signal West Pico Substrate (ThermoFisher) and Super Signal West Femto Substrate (ThermoFisher) in a 1:10 ratio. Blots were imaged using an AnalytikJena UVP Chemstudio (AnalytikJena AG) system.

7.2.6. Functional characterization of membrane proteins

7.2.6.1. *In vivo* uptake assay

To test the activity of newly generated DtpA mutant the uptake of a fluorescent dye coupled to a dipeptide, which is a substrate for DtpA wildtype, was tested. The uptake assay of the fluorescent N-7-amino-4-methylcoumarin-3-acetic acid coupled to AK dipeptide was performed as described previously with minor changes [53]. For *in vivo* uptake the DtpA pNIC-CTHF expression plasmid was transformed into C41(DE3) *E. coli* strain. Transformed bacteria were grown at 37°C in TB medium supplemented with 30 µg/mL kanamycin. At an OD600 nm of 0.6 cells were induced with 0.2 mM IPTG and incubated for further 3 h at 37°C. Cells were harvested by centrifugation at 3214 xg for 15 min at 4°C and resuspended in AMCA assay buffer to reach a final OD600 nm of 10. In a final reaction volume of 100 µL 50 µL AMCA assay buffer and 40 µL cells at OD600 nm of 10 were incubated with 10 µL of 1 mM AK-AMCA for 20 min at 37°C. As negative control, double distilled water instead of AK-AMCA was added. The reaction was stopped by addition of 200 µL ice-cold AMCA assay buffer. Cells were washed twice with 200 µL ice-cold AMCA assay buffer and finally resuspended in 200 µL ice-cold AMCA assay buffer. Remaining fluorescence was measured using a Tecan Spark multimode microplate reader (Tecan Life Sciences) with an $\lambda_{excitation} = 350$ nm and $\lambda_{emission} = 450$ nm. To correct for the number of the cells, contributing to the fluorescence signal, the OD600 nm for each sample was measured. Western blot was performed as described in chapter 7.2.5.3. to account for differences in expression level of the variants and the wild type.

7.2.6.2. *In vitro* growth assay

For *in vitro* growth assays, the parental (BW25113) and knockout strains for six ACS transporters (DgoT, GarP, GudP, ExuT, LgoT, RhmT) were grown at 37 °C in LB medium overnight. For the knockout strains the medium was supplemented with 30 µg/mL kanamycin. Next day M9 minimal medium supplemented with 2 g/L glucose was inoculated with LB culture to an OD 600 nm of 0.05 and grown at 37 °C overnight. For the knockout strains, the medium was supplemented with 30 µg/mL kanamycin. The next day, cells were harvested by centrifugation (3163 xg, 5 min, 4 °C) washed twice with 5 mL 1 x PBS. Cells were resuspended in 1 x PBS to reach a final OD 600 nm of 10. The cell suspension was spotted on Minimal agar plates supplemented with 2 g/L per carbon source and incubated at 37 °C overnight.

7.2.6.3. Thermal shift assay

The stability of proteins and protein complexes were monitored by nanoDSF. Here, the intrinsic fluorescence of tryptophan residues from the protein is recorded at 330 nm and 350 nm while heating the sample heated from 20 °C to 95 °C with increments of 1 °C/min. For ligand binding assays 0.5 mg/mL of membrane proteins was incubated with transporter substrates at 2.5 mM final ligand concentration for 10 min at RT. To measure nanobody binding, membrane proteins at 0.5 mg/mL were incubated with nanobody at a molar ratio of 1:1.2 for 10 min at RT. Standard grade nanoDSF capillaries (Nanotemper) were loaded into a Prometheus NT.48 device (Nanotemper) controlled by PR. Therm-Control (version 2.1.2). Excitation power was adjusted to 20% for POT proteins and 10 % for ACS proteins. All samples were analyzed using Excel (Microsoft).

7.2.6.4. Biolayer interferometry

K_D values of nanobodies binding to membrane proteins were measured using an Octet RED96 System (fortéBIO). Briefly, nanobodies were biotinylated using an EZ-Link™ NHS-PEG4-Biotin kit (ThermoFisher). Nanobodies were diluted to 5 µg/mL with Octet Buffer supplemented with 0.01% (w/v) LMNG or 0.03 % (w/v) DDM and loaded onto streptavidin (SA) biosensors (fortéBIO), previously hydrated with the same buffer. Unbound nanobody was washed off with Octet Buffer. Membrane protein at 200 nM was then bound to the nanobody. Experiments were performed at 22 °C and shaking (1000 rpm). Data was analyzed using Data Analysis v9.0 software (fortéBIO) assuming a 1:1 stoichiometry of the membrane protein nanobody complex. A Savitzky-Golay filter was applied to smooth the data.

7.2.6.5. Cysteine accessibility assay

The accessibility of cysteine residues was tested by incubating 0.15 mg/mL of the protein with Methoxypolyethylene glycol maleimide 5000 (PEG-maleimide) (Sigma-Aldrich) at 2 mM final PEG-maleimide concentration for 30 min at RT. SDS-PAGE was performed using a 4-12% Bis-Tris gel (Expedeon) and stained using InstantBlue Coomassie Protein Stain (Expedeon).

7.2.6.6. Liposome-based uptake assay

To test the activity of membrane proteins *in vivo* pyranine assays were performed. Pyranine is a pH-sensitive dye, thus allowing to measure the influx of protons into liposomes as a result of proton-dependent transport [203].

Typically, 5 ug of reconstituted protein per data point was used. Proteoliposomes were Thawed at RT and the amount necessary for several assays was taken and pelleted by centrifugation (100,000 xg, 30 min, 4 °C). The pellet was then resuspended into 450 µL pyranine inside buffer supplemented with 10 mM pyranine and transferred into a freh tube. After 7 freeze-thaw cycles the liposomes were extruded 11 times through a LiposoFast Liposome factory 400 nm filter (Avestin, Inc.) to form unilamellar vesicles. The liposomes were harvested by centrifugation (60,000 xg, 30 min, 15 °C). The pellet was washed by gently pipetting 100 µL pyranine inside buffer over the pellet surface. The pellet was then resuspended into 100 µL inside buffer. A G-25 spin column was used to remove dye from the outside of the liposomes and the liposomes were again harvested by centrifugation (60,000 xg, 30 min, 15 °C). Finally, liposomes were resuspended to a final protein concentration of 1.25 mg/mL in pyranine inside buffer.

To follow the proton-movement, the fluorescence of the pyranine dye was monitored for 12 min while exciting at a wavelength of 415 nm and 460 nm and measuring emission at a wavelength of 510 nm in a TECAN spark multiplate reader. While the fluorescence signal at 460 nm is dependent on the pH, the signal at 415 nm is not [203]. Therefore, it serves as internal standard to normalize for the amount of pyranine loaded into liposomes.

For the uptake assay, 4 µL of the liposome solution was diluted in 200 µL pyranine outside buffer. Both pyranine inside and pyranine outside buffer contain the same concentration of monovalent cations as to prevent establishing of a membrane potential. The outside buffer contains 120 mM potassium whereas the inside buffer contains 120 mM sodium, thus establishing a potassium gradient across the liposome membrane. After 30 s substrate was added to a final substrate concentration of 250 µM, water was used for control. Valinomycin was added after 90 s to a final concentration of 0.1 µM. Valinomycin allows potassium ions to cross the liposome membrane, thus abolishing the potassium gradient but

creating a negatively charged inside of the liposomes. This allows proton influx into the liposomes. As the incorporated membrane proteins are proton-coupled transporters, the measurable influx of protons through these transporters serves as readout for the transport of substrate. After 10 min the ionophore CCCP is added to abolish the proton-gradient. To control for proton leakage through the liposome membrane, the uptake assay is performed for liposomes without reconstituted transporters (empty liposomes) as well.

7.2.7. Crystallization

7.2.7.1.   Vapor diffusion method

Typically, ACS transporters were concentrated to 5 mg/mL. In the case of co-crystallization with substrates, ligands were added to 2 mM final concentration, followed by incubation on ice for 30 min. Before setting up crystallization plates, protein samples were centrifuged (13,000 xg, 30 min, 4 °C) to remove any aggregates. Crystallization plates were set-up using a Mosquito-LCP robot (TTP Labtech) dispensing three ratios of protein-to-reservoir of 200:100 nL, 150:150 nL and 100:200 nL into an Intelliplate-96plate (Hampton Research). Plates were incubated at 19 °C or 4 °C, while being imaged by a RockImager system (Formulatrix).

7.2.7.2.   LCP

In brief, LgoT was concentrated to 40 mg/mL and mixed with the lipid Monoolein in a ratio, using two coupled Hamilton syringes. To ensure the lipid remained liquid it was previously heated to 37 °C and filled in a prewarmed syringe. The protein was pipetted into a syring at RT. Both syringes were connoted using an LCP syringe coupler (TTP Labtech) without introducing air bubbles. The content of both syringes was mixed until the resulting mesophase was clear and stable over time. Crystal plates were set-up using a Mosquito-LCP robot (TTP Labtech) dispensing 100 nL mesophase and 1000 nL reservoir buffer into 96-well glass sandwich plates. Plates were incubated at 19 °C. Crystals grew after 7 days and were harvested after 14 days.

8. References

[1]   D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry*, W.H. Freeman & Co. Ltd, **2017**.

[2]   J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, W.H. Freeman & Co. Ltd, **2012**.

[3]   J. K. Volkman, *Org. Geochem.* **2005**, *36*, 139–159.

[4]   H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, *Molecular Cell Biology*, W.H. Freeman & Co. Ltd, **2001**.

[5]   M. Frick, C. Schmidt, *Chem. Phys. Lipids* **2019**, *221*, 145–157.

[6]   B. A. Wallace, *J. Struct. Biol.* **1998**, *121*, 123–141.

[7]   J. L. Parker, J. A. Mindell, S. Newstead, *Elife* **2014**, *3*, e04273.

[8]   S. S. Pao, I. T. Paulsen, M. H. Saier, *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1–34.

[9]   N. Yan, *Annu. Rev. Biophys.* **2015**, *44*, 257–283.

[10]   V. Naroditskaya, M. J. Schlosser, N. Y. Fang, K. Lewis, *Biochem. Biophys. Res. Commun.* **1993**, *196*, 803–809.

[11]   K. Nishino, A. Yamaguchi, *J. Bacteriol.* **2001**, *183*, 5803–5812.

[12]   M. Brandsch, I. Knütter, E. Bosse-Doenecke, *J. Pharm. Pharmacol.* **2008**, *60*, 543–585.

[13]   N. Yan, *Curr. Opin. Struct. Biol.* **2013**, *23*, 483–491.

[14]   X. Bai, T. F. Moraes, R. A. F. Reithmeier, *Mol. Membr. Biol.* **2017**, *34*, 1–32.

[15]   Y. Zhang, Y. Zhang, K. Sun, Z. Meng, L. Chen, *J. Mol. Cell Biol.* **2018**, *11*, 1–13.

[16]   M. Gregor Madej, S. Dang, N. Yan, H. Ronald Kaback, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5870–5874.

[17]   Y. Shi, *Annu. Rev. Biophys.* **2013**, *42*, 51–72.

[18]   M. D. Marger, M. H. Saier, *Trends Biochem. Sci.* **1993**, *18*, 13–20.

[19]   H. Daniel, B. Spanier, G. Kottra, D. Weitz, *Physiology* **2006**, *21*, 93–102.

[20]   R. Höhner, A. Aboukila, H. H. Kunz, K. Venema, *Front. Plant Sci.* **2016**, *7*, 218.

[21]   F. W. Verheijen, E. Verbeek, N. Aula, C. E. M. T. Beerens, A. C. Havelaar, M. Joosse, L. Peltonen, P. Aula, H. Galjaard, P. J. Van Der Spek, G. M. S. Mancini, *Nat. Genet.* **1999**, *23*, 462–465.

[22]   J. R. Casey, S. Grinstein, J. Orlowski, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 50–61.

[23]   J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, S. Iwata, *Science (80-.).* **2003**, *301*, 610–615.

[24]   Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, *Science (80-.).* **2003**, *301*, 616–620.

[25] B. P. Pedersen, H. Kumar, A. B. Waight, A. J. Risenmay, Z. Roe-Zurz, B. H. Chau, A. Schlessinger, M. Bonomi, W. Harries, A. Sali, A. K. Johri, R. M. Stroud, *Nature* **2013**, *496*, 533–536.

[26] D. Deng, C. Xu, P. Sun, J. Wu, C. Yan, M. Hu, N. Yan, *Nature* **2014**, *510*, 121–125.

[27] V. S. Reddy, M. A. Shlykov, R. Castillo, E. I. Sun, M. H. Saier, *FEBS J.* **2012**, *279*, 2022–2035.

[28] S. Newstead, *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 488–499.

[29] C. J. Law, P. C. Maloney, D. N. Wang, *Annu. Rev. Microbiol.* **2008**, *62*, 289–305.

[30] O. Jardetzky, *Nature* **1966**, *211*, 969–970.

[31] D. Drew, O. Boudker, *Annu. Rev. Biochem.* **2016**, *85*, 543–572.

[32] S. Dang, L. Sun, Y. Huang, F. Lu, Y. Liu, H. Gong, J. Wang, N. Yan, *Nature* **2010**, *467*, 734–738.

[33] L. Sun, X. Zeng, C. Yan, X. Sun, X. Gong, Y. Rao, N. Yan, *Nature* **2012**, *490*, 361–366.

[34] E. M. Quistgaard, C. Löw, F. Guettou, P. Nordlund, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 123–132.

[35] N. Solcan, J. Kwok, P. W. Fowler, A. D. Cameron, D. Drew, S. Iwata, S. Newstead, *EMBO J.* **2012**, *31*, 3411–3421.

[36] P. W. Fowler, M. Orwick-Rydmark, S. Radestock, N. Solcan, P. M. Dijkman, J. A. Lyons, J. Kwok, M. Caffrey, A. Watts, L. R. Forrest, S. Newstead, *Structure* **2015**, *23*, 290–301.

[37] H. Kumar, V. Kasho, I. Smirnova, J. S. Finer-Moore, H. R. Kaback, R. M. Stroud, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1784–1788.

[38] H. R. Kaback, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 1259–1264.

[39] N. Yan, *Trends Biochem. Sci.* **2013**, *38*, 151–159.

[40] H. -Y Steiner, F. Naider, J. M. Becker, *Mol. Microbiol.* **1995**, *16*, 825–834.

[41] R. Boggavarapu, J. M. Jeckelmann, D. Harder, Z. Ucurum, D. Fotiadis, *BMC Biol.* **2015**, *13*, 58.

[42] S. Léran, K. Varala, J. C. Boyer, M. Chiurazzi, N. Crawford, F. Daniel-Vedele, L. David, R. Dickstein, E. Fernandez, B. Forde, W. Gassmann, D. Geiger, A. Gojon, J. M. Gong, B. A. Halkier, J. M. Harris, R. Hedrich, A. M. Limami, D. Rentsch, M. Seo, Y. F. Tsay, M. Zhang, G. Coruzzi, B. Lacombe, *Trends Plant Sci.* **2014**, *19*, 5–9.

[43] Y. Boursiac, S. Léran, C. Corratgé-Faillie, A. Gojon, G. Krouk, B. Lacombe, *Trends Plant Sci.* **2013**, *18*, 325–333.

[44] H. H. Nour-Eldin, T. G. Andersen, M. Burow, S. R. Madsen, M. E. Jørgensen, C. E. Olsen, I. Dreyer, R. Hedrich, D. Geiger, B. A. Halkier, *Nature* **2012**, *488*, 531–534.

[45] G. Krouk, B. Lacombe, A. Bielach, F. Perrine-Walker, K. Malinska, E. Mounier, K. Hoyerova, P. Tillard, S. Leon, K. Ljung, E. Zazimalova, E. Benkova, P. Nacry, A. Gojon, *Dev. Cell* **2010**, *18*, 927–937.

[46]   Y. F. Tsay, C. C. Chiu, C. B. Tsai, C. H. Ho, P. K. Hsu, *FEBS Lett.* **2007**, *581*, 2290–2300.

[47]   T. Terada, K. I. Inui, in *Curr. Top. Membr.*, **2012**, pp. 257–274.

[48]   B. K. Prabhala, N. G. Aduri, H. Hald, O. Mirza, *Int. J. Pept. Res. Ther.* **2015**, *21*, 1–6.

[49]   N. Sharma, N. G. Aduri, A. Iqbal, B. K. Prabhala, O. Mirza, *J. Mol. Microbiol. Biotechnol.* **2016**, *26*, 312–319.

[50]   Y. Ural-Blimke, A. Flayhan, J. Strauss, V. Rantos, K. Bartels, R. Nielsen, E. Pardon, J. Steyaert, J. Kosinski, E. M. Quistgaard, C. Löw, *J. Am. Chem. Soc.* **2019**, *141*, 2404–2412.

[51]   H. A. Ernst, A. Pham, H. Hald, J. S. Kastrup, M. Rahman, O. Mirza, *Biochem. Biophys. Res. Commun.* **2009**, *389*, 112–116.

[52]   F. Guettou, E. M. Quistgaard, M. Raba, P. Moberg, C. Löw, P. Nordlund, *Nat. Struct. Mol. Biol.* **2014**, *21*, 728–731.

[53]   D. Weitz, D. Harder, F. Casagrande, D. Fotiadis, P. Obrdlik, B. Kelety, H. Daniel, *J. Biol. Chem.* **2007**, *282*, 2832–2839.

[54]   B. K. Prabhala, N. G. Aduri, N. Sharma, A. Shaheen, A. Sharma, M. Iqbal, P. R. Hansen, C. Brasen, M. Gajhede, M. Rahman, O. Mirza, *J. Biol. Chem.* **2018**, *293*, 1007–1017.

[55]   D. E. Smith, B. Clémençon, M. A. Hediger, *Mol. Aspects Med.* **2013**, *34*, 323–336.

[56]   V. H. Thomas, S. Bhattachar, L. Hitchingham, P. Zocharski, M. Naath, N. Surendran, C. L. Stoner, A. El-Kattan, *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 591–608.

[57]   D. Jung, A. Dorr, *J. Clin. Pharmacol.* **1999**, *39*, 800–804.

[58]   S. Newstead, *Curr. Opin. Struct. Biol.* **2017**, *45*, 17–24.

[59]   S. Newstead, D. Drew, A. D. Cameron, V. L. G. Postis, X. Xia, P. W. Fowler, J. C. Ingram, E. P. Carpenter, M. S. P. Sansom, M. J. McPherson, S. A. Baldwin, S. Iwata, *EMBO J.* **2011**, *30*, 417–426.

[60]   Y. Zhao, G. Mao, M. Liu, L. Zhang, X. Wang, X. C. Zhang, *Structure* **2014**, *22*, 1152–1160.

[61]   F. Guettou, E. M. Quistgaard, L. Trésaugues, P. Moberg, C. Jegerschöld, L. Zhu, A. J. O. Jong, P. Nordlund, C. Löw, *EMBO Rep.* **2013**, *14*, 804–810.

[62]   M. Martinez Molledo, E. M. Quistgaard, A. Flayhan, J. Pieprzyk, C. Löw, *Structure* **2018**, *26*, 467-476.e4.

[63]   J. H. Beale, J. L. Parker, F. Samsudin, A. L. Barrett, A. Senan, L. E. Bird, D. Scott, R. J. Owens, M. S. P. Sansom, S. J. Tucker, D. Meredith, P. W. Fowler, S. Newstead, *Structure* **2015**, *23*, 1889–1899.

[64]   J. Sun, J. R. Bankston, J. Payandeh, T. R. Hinds, W. N. Zagotta, N. Zheng, *Nature* **2014**, *507*, 73–77.

[65]   C. Y. Huang, V. Olieric, N. Howe, R. Warshamanage, T. Weinert, E. Panepucci, L. Vogeley, S. Basu, K. Diederichs, M. Caffrey, M. Wang, *Commun. Biol.* **2018**, *1*, 124.

[66] C. Y. Huang, V. Olieric, P. Ma, E. Panepucci, K. Diederichs, M. Wang, M. Caffrey, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71*, 1238–1256.

[67] J. A. Lyons, J. L. Parker, N. Solcan, A. Brinth, D. Li, S. T. Shah, M. Caffrey, S. Newstead, *EMBO Rep.* **2014**, *15*, 886–893.

[68] C. Y. Huang, V. Olieric, P. Ma, N. Howe, L. Vogeley, X. Liu, R. Warshamanage, T. Weinert, E. Panepucci, B. Kobilka, K. Diederichs, M. Wang, M. Caffrey, *Acta Crystallogr. Sect. D Struct. Biol.* **2016**, *72*, 93–112.

[69] P. Ma, D. Weichert, L. A. Aleksandrov, T. J. Jensen, J. R. Riordan, X. Liu, B. K. Kobilka, M. Caffrey, *Nat. Protoc.* **2017**, *12*, 1745–1762.

[70] E. M. Quistgaard, M. M. Molledo, C. Löw, *PLoS One* **2017**, *12*, e0173126.

[71] M. Martinez Molledo, E. M. Quistgaard, C. Löw, *FEBS Lett.* **2018**, *592*, 3239–3247.

[72] R. Nagamura, M. Fukuda, A. Kawamoto, K. Matoba, N. Dohmae, R. Ishitani, J. Takagi, O. Nureki, *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2019**, *75*, 348–358.

[73] S. Doki, Hideaki E Kato, N. Solcan, M. Iwaki, M. Koyama, M. Hattori, N. Iwase, T. Tsukazaki, Y. Sugita, H. Kandori, S. Newstead, R. Ishitani, Osamu Nureki, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11343–11348.

[74] J. L. Parker, S. Newstead, *Nature* **2014**, *507*, 68–72.

[75] J. L. Parker, C. Li, A. Brinth, Z. Wang, L. Vogeley, N. Solcan, G. Ledderboge-Vucinic, J. M. J. Swanson, M. Caffrey, G. A. Voth, S. Newstead, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 13182–13187.

[76] G. S. Minhas, S. Newstead, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 804–809.

[77] G. S. Minhas, D. Bawdon, R. Herman, M. Rudden, A. P. Stone, A. G. James, G. H. Thomas, S. Newstead, *Elife* **2018**, *7*, e34995.

[78] S. Tynkkynen, G. Buist, E. Kunji, J. Kok, B. Poolman, G. Venema, A. Haandrikman, *J. Bacteriol.* **1993**, *175*, 7523–7532.

[79] V. Monnet, *Cell. Mol. Life Sci.* **2003**, *60*, 2100–2114.

[80] J. M. Jensen, F. Ismat, G. Szakonyi, M. Rahman, O. Mirza, *PLoS One* **2012**, DOI 10.1371/journal.pone.0047780.

[81] B. K. Prabhala, N. G. Aduri, J. M. Jensen, H. A. Ernst, N. Iram, M. Rahman, O. Mirza, *FEBS Lett.* **2014**, *588*, 560–565.

[82] R. J. Reimer, *Mol. Aspects Med.* **2013**, *34*, 350–359.

[83] J. Biber, M. Custer, A. Werner, B. Kaissling, H. Murer, *Pflügers Arch. Eur. J. Physiol.* **1993**, *424*, 210–215.

[84] H. Yabuuchi, I. Tamai, K. Morita, T. Kouda, K. I. Miyamoto, E. Takeda, A. Tsuji, *J. Pharmacol. Exp. Ther.* **1998**, *286*, 1391–1396.

124

[85]   S. Sreedharan, J. H. A. Shaik, P. K. Olszewski, A. S. Levine, H. B. Schiöth, R. Fredriksson, *BMC Genomics* **2010**, *11*, 17.

[86]   P. Jutabha, N. Anzai, K. Kitamura, A. Taniguchi, S. Kaneko, K. Yan, H. Yamada, H. Shimada, T. Kimura, T. Katada, T. Fukutomi, K. Tomita, W. Urano, H. Yamanaka, G. Seki, T. Fujita, Y. Moriyama, A. Yamada, S. Uchida, M. F. Wempe, H. Endou, H. Sakuraia, *J. Biol. Chem.* **2010**, *285*, 35123–35132.

[87]   H. Murer, N. Hernando, I. Forster, J. Biber, *Physiol. Rev.* **2000**, *80*, 1373–1409.

[88]   M. Iharada, T. Miyaji, T. Fujimoto, M. Hiasa, N. Anzai, H. Omote, Y. Moriyama, *J. Biol. Chem.* **2010**, *285*, 26107–26113.

[89]   Y. Shigeri, R. P. Seal, K. Shimamoto, *Brain Res. Rev.* **2004**, *45*, 250–265.

[90]   R. J. Reimer, R. H. Edwards, *Pflugers Arch. Eur. J. Physiol.* **2004**, *447*, 629–635.

[91]   R. T. Fremeau, K. Kam, Y. Qureshi, J. Johnson, D. R. Copenhagen, J. Storm-Mathisen, F. A. Chaudhry, R. A. Nicoll, R. H. Edwards, *Science (80-.).* **2004**, *304*, 1815–1819.

[92]   Å. Wallen-Mackenzie, H. Wootz, H. Englund, *Ups. J. Med. Sci.* **2010**, *115*, 11–20.

[93]   J. Ruel, S. Emery, R. Nouvian, T. Bersot, B. Amilhon, J. M. Van Rybroek, G. Rebillard, M. Lenoir, M. Eybalin, B. Delprat, T. A. Sivakumaran, B. Giros, S. El Mestikawy, T. Moser, R. J. H. Smith, M. M. Lesperance, J. L. Puel, *Am. J. Hum. Genet.* **2008**, *83*, 278–292.

[94]   R. P. Seal, O. Akil, E. Yi, C. M. Weber, L. Grant, J. Yoo, A. Clause, K. Kandler, J. L. Noebels, E. Glowatzki, L. R. Lustig, R. H. Edwards, *Neuron* **2008**, *57*, 263–275.

[95]   P. Morin, C. Sagné, B. Gasnier, *EMBO J.* **2004**, *23*, 4560–4570.

[96]   C. C. Wreden, M. Wlizla, R. J. Reimer, *J. Biol. Chem.* **2005**, *280*, 1408–1416.

[97]   M. Renlund, F. Tietze, W. A. Gahl, *Science (80-.).* **1986**, *232*, 759–762.

[98]   N. Pietrancosta, C. Anne, H. Prescher, R. Ruivo, C. Sagné, C. Debacker, H. O. Bertrand, R. Brossmer, F. Acher, B. Gasnier, *J. Biol. Chem.* **2012**, *287*, 11489–11497.

[99]   L. M. Prolo, H. Vogel, R. J. Reimer, *J. Neurosci.* **2009**, *29*, 15355–15365.

[100]   B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, Garland Science, **2002**.

[101]   J. B. Leano, S. Batarni, J. Eriksen, N. Juge, J. E. Pak, T. Kimura-Someya, Y. Robles-Colmenares, Y. Moriyama, R. M. Stroud, R. H. Edwards, *PLoS Biol.* **2019**, *17*, e3000260.

[102]   F. Li, J. Eriksen, J. Finer-Moore, R. Chang, P. Nguyen, A. Bowen, A. Myasnikov, Z. Yu, D. Bulkley, Y. Cheng, R. H. Edwards, R. M. Stroud, *Science (80-.).* **2020**, *368*, 893–897.

[103]   T. F. Moraes, R. A. F. Reithmeier, *Biochim. Biophys. Acta - Biomembr.* **2012**, *1818*, 2687–2706.

[104] J. F. Rakus, A. A. Fedorov, E. V. Fedorov, M. E. Glasner, B. K. Hubbard, J. D. Delli, P. C. Babbitt, S. C. Almo, J. A. Gerlt, *Biochemistry* **2008**, *47*, 9944–9954.

[105] J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, C. D. Herring, O. T. Bui, E. M. Knight, S. S. Fong, B. O. Palsson, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 17480–17484.

[106] G. Nemoz, J. Robert Baudouy, F. Stoeber, *J. Bacteriol.* **1976**, *127*, 706–718.

[107] R. A. Cooper, *Arch. Microbiol.* **1978**, *118*, 199–206.

[108] N. J. Rancour, E. D. Hawkins, W. W. Wells, *Arch. Biochem. Biophys.* **1979**, *193*, 232–241.

[109] C. Ficicioglu, C. Yager, S. Segal, *Mol. Genet. Metab.* **2005**, *84*, 152–159.

[110] K. Lai, M. I. Klapa, *Metab. Eng.* **2004**, *6*, 239–244.

[111] H. D. Kay, *Biochem. J.* **1926**, *20*, 321–329.

[112] J. Deacon, R. A. Cooper, *FEBS Lett.* **1977**, *77*, 201–205.

[113] R. A. Cooper, *FEBS Lett.* **1979**, *103*, 216–220.

[114] B. K. Hubbard, M. Koch, D. R. J. Palmer, P. C. Babbitt, J. A. Gerlt, *Biochemistry* **1998**, *37*, 14369–14375.

[115] N. J. Sweeney, D. C. Laux, P. S. Cohen, *Infect. Immun.* **1996**, *64*, 3504–3511.

[116] N. Peekhaus, T. Conway, *J. Bacteriol.* **1998**, *180*, 3495–3502.

[117] T. Conway, P. S. Cohen, in *Metab. Bact. Pathog.*, **2015**, pp. 343–362.

[118] B. Singh, G. Arya, N. Kundu, A. Sangwan, S. Nongthombam, R. Chaba, *J. Bacteriol.* **2019**, *201*, e00281-18.

[119] R. Monterrubio, L. Baldoma, N. Obradors, J. Aguilar, J. Badia, *J. Bacteriol.* **2000**, *182*, 2672–2674.

[120] K. Sawada, N. Echigo, N. Juge, T. Miyaji, M. Otsuka, H. Omote, A. Yamamoto, Y. Moriyama, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5683–5686.

[121] J. F. Robyt, *Essentials of Carbohydrate Chemistry*, Springer New York, New York, NY, **1998**.

[122] I. A. Suvorova, M. N. Tutukina, D. A. Ravcheev, D. A. Rodionov, O. N. Ozoline, M. S. Gelfand, *J. Bacteriol.* **2011**, *193*, 3956–3963.

[123] R. A. Cooper, *FEBS Lett.* **1979**, *103*, 216–220.

[124] E. C. McCusker, C. Bagnéris, C. E. Naylor, A. R. Cole, N. D'Avanzo, C. G. Nichols, B. A. Wallace, *Nat. Commun.* **2012**, *3*, 1102.

[125] D. M. Rosenbaum, S. G. F. Rasmussen, B. K. Kobilka, *Nature* **2009**, *459*, 356–363.

[126] Z. E. R. Newby, J. D. O'Connell, F. Gruswitz, F. A. Hays, W. E. C. Harries, I. M. Harwood, J. D. Ho, J. K. Lee,

D. F. Savage, L. J. W. Miercke, R. M. Stroud, *Nat. Protoc.* **2009**, *4*, 619–637.

[127]   E. Wallin, G. Von Heijne, *Protein Sci.* **1998**, *7*, 1029–1038.

[128]   I. Moraes, G. Evans, J. Sanchez-Weatherby, S. Newstead, P. D. S. Stewart, *Biochim. Biophys. Acta - Biomembr.* **2014**, *1838*, 78–87.

[129]   J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nat. Rev. Drug Discov.* **2006**, *5*, 993–996.

[130]   S. Wagner, M. L. Bader, D. Drew, J. W. de Gier, *Trends Biotechnol.* **2006**, DOI 10.1016/j.tibtech.2006.06.008.

[131]   A. Stetsenko, A. Guskov, *Crystals* **2017**, *7*, 197.

[132]   C. Chipot, F. Dehez, J. R. Schnell, N. Zitzmann, E. Pebay-Peyroula, L. J. Catoire, B. Miroux, E. R. S. Kunji, G. Veglia, T. A. Cross, P. Schanda, *Chem. Rev.* **2018**, *118*, 3559–3607.

[133]   H. Michel, *Trends Biochem. Sci.* **1983**, *8*, 56–59.

[134]   M. Caffrey, *Acta Crystallogr. Sect. FStructural Biol. Commun.* **2015**, *71*, 3–18.

[135]   C. Ostermeier, H. Michel, *Curr. Opin. Struct. Biol.* **1997**, *7*, 697–701.

[136]   I. R. Krauss, A. Merlino, A. Vergara, F. Sica, *Int. J. Mol. Sci.* **2013**, *14*, 11643–11691.

[137]   G. G. Privé, *Methods* **2007**, *41*, 388–397.

[138]   M. A. O'Malley, M. E. Helgeson, N. J. Wagner, A. S. Robinson, *Biophys. J.* **2011**, *101*, 1938–1948.

[139]   E. M. Landau, J. P. Rosenbusch, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 14532–14535.

[140]   M. Caffrey, *Annu. Rev. Biophys.* **2009**, *38*, 29–51.

[141]   J. Briggs, M. Caffrey, *Biophys. J.* **1994**, *66*, 573–587.

[142]   K. Sikic, S. Tomic, O. Carugo, *Open Biochem. J.* **2010**, *4*, 83–95.

[143]   K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, D. Kern, *Nature* **2007**, *450*, 838–844.

[144]   X. Wang, L. Vukovic, H. R. Koh, K. Schulten, S. Myong, *Nucleic Acids Res.* **2015**, *43*, 7566–7576.

[145]   C. D. Kinz-Thompson, A. K. Sharma, J. Frank, R. L. Gonzalez, D. Chowdhury, *J. Phys. Chem. B* **2015**, *119*, 10888–10901.

[146]   A. E. Johnson-Buck, M. R. Blanco, N. G. Walter, in *Encycl. Biophys.* (Ed.: G.C.K. Roberts), Springer Berlin Heidelberg, Berlin, Heidelberg, **2013**, pp. 2329–2335.

[147]   T. Förster, *Ann. Phys.* **1948**, *437*, 55–75.

[148]   X. Michalet, S. Weiss, M. Jäger, *Chem. Rev.* **2006**, *106*, 1785–1813.

[149]   E. Haas, *Methods Mol. Biol.* **2012**, *895*, 467–498.

[150]   W. R. Algar, N. Hildebrandt, S. S. Vogel, I. L. Medintz, *Nat. Methods* **2019**, *16*, 815–829.

[151]   M. Börsch, T. M. Duncan, *Biochem. Soc. Trans.* **2013**, *41*, 1219–1226.

[152]   H. P. Lu, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6734–6749.

[153]   M. Stracy, S. Uphoff, F. Garza De Leon, A. N. Kapanidis, *FEBS Lett.* **2014**, *588*, 3585–3594.

[154]   F. Husada, K. Bountra, K. Tassis, M. Boer, M. Romano, S. Rebuffat, K. Beis, T. Cordes, *EMBO J.* **2018**, *37*, e100056.

[155]   Y. Gambin, A. A. Deniz, *Mol. Biosyst.* **2010**, *6*, 1540–1547.

[156]   E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Y. Chung, X. Michalet, S. Weiss, *Science (80-.).* **2018**, *359*, eaan1133.

[157]   D. P. Nair, M. Podgórski, S. Chatani, T. Gong, W. Xi, C. R. Fenoli, C. N. Bowman, *Chem. Mater.* **2014**, *26*, 724–744.

[158]   R. B. Mujumdar, L. A. Ernst, S. R. Mujumdar, C. J. Lewis, *Bioconjug. Chem.* **1993**, *4*, 105–111.

[159]   N. Panchuk-Voloshina, R. P. Haugland, J. Bishop-Stewart, M. K. Bhalgat, P. J. Millard, F. Mao, W. Y. Leung, R. P. Haugland, *J. Histochem. Cytochem.* **1999**, *47*, 1179–1188.

[160]   B. Schuler, *J. Nanobiotechnology* **2013**, *11 Suppl 1*, S2.

[161]   L. Stryer, R. P. Haugland, *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 719–726.

[162]   T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, S. Weiss, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6264–6268.

[163]   A. Hoffmann, D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, B. Schuler, *Phys. Chem. Chem. Phys.* **2011**, *13*, 1857–1871.

[164]   B. K. Müller, E. Zaychikov, C. Bräuchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522.

[165]   E. M. Quistgaard, C. Löw, F. Guettou, P. Nordlund, *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 123–132.

[166]   R. Vafabakhsh, J. Levitz, E. Y. Isacoff, *Nature* **2015**, *524*, 497–501.

[167]   S. Wang, R. Vafabakhsh, W. F. Borschel, T. Ha, C. G. Nichols, *Nat. Struct. Mol. Biol.* **2016**, *23*, 31–36.

[168]   Y. Zhao, D. S. Terry, L. Shi, M. Quick, H. Weinstein, S. C. Blanchard, J. A. Javitch, *Nature* **2011**, *474*, 109–113.

[169]   N. Akyuz, R. B. Altman, S. C. Blanchard, O. Boudker, *Nature* **2013**, *502*, 114–118.

[170]   H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*,

16155–16160.

[171] "Alexa Fluor™ 594 C$_5$ Maleimid," can be found under https://www.thermofisher.com/order/catalog/product/A10256#/A10256, **n.d.**

[172] "Alexa Fluor™ 488 C$_5$ Maleimid," can be found under https://www.thermofisher.com/order/catalog/product/A10254#/A10254, **n.d.**

[173] E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, C. A. M. Seidel, in *Methods Enzymol.*, **2010**, pp. 455–514.

[174] H. Chen, S. S. Ahsan, M. B. Santiago-Berrios, H. D. Abruña, W. W. Webb, *J. Am. Chem. Soc.* **2010**, *132*, 7244–7245.

[175] L. E. Jansen, L. J. Negrón-Piñeiro, S. Galarza, S. R. Peyton, *Acta Biomater.* **2018**, *70*, 120–128.

[176] P. S. Chae, S. G. F. Rasmussen, R. R. Rana, K. Gotfryd, R. Chandra, M. A. Goren, A. C. Kruse, S. Nurva, C. J. Loland, Y. Pierre, D. Drew, J. L. Popot, D. Picot, B. G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka, S. H. Gellman, *Nat. Methods* **2010**, *7*, 1003–1008.

[177] J. Phan, A. Zdanov, A. G. Evdokimov, J. E. Tropea, H. K. Peters, R. B. Kapust, M. Li, A. Wlodawer, D. S. Waugh, *J. Biol. Chem.* **2002**, *277*, 50564–50572.

[178] P. Savitsky, J. Bray, C. D. O. Cooper, B. D. Marsden, P. Mahajan, N. A. Burgess-Brown, O. Gileadi, *J. Struct. Biol.* **2010**, *172*, 3–13.

[179] F. H. Niesen, H. Berglund, M. Vedadi, *Nat. Protoc.* **2007**, *2*, 2212–2221.

[180] J. Concepcion, K. Witte, C. Wartchow, S. Choo, D. Yao, H. Persson, J. Wei, P. Li, B. Heidecker, W. Ma, R. Varma, L.-S. Zhao, D. Perillat, G. Carricato, M. Recknor, K. Du, H. Ho, T. Ellis, J. Gamez, M. Howes, J. Phi-Wilson, S. Lockard, R. Zuk, H. Tan, *Comb. Chem. High Throughput Screen.* **2009**, *12*, 791–800.

[181] "BLI • Biolayer Interferometry • Kinetics Alternative to SPR and Biacore," can be found under https://2bind.com/bli/, **n.d.**

[182] M. R. Arana, M. C. Fiori, G. A. Altenberg, *Biochem. Biophys. Res. Commun.* **2019**, *512*, 448–452.

[183] D. M. Kehlenbeck, I. Josts, J. Nitsche, S. Busch, T. Forsyth, H. Tidow, *Biol. Chem.* **2019**, *400*, 1509–1518.

[184] M. Landreh, C. V. Robinson, *J. Physiol.* **2015**, *593*, 355–362.

[185] J. Frauenfeld, R. Löving, J. P. Armache, A. F. P. Sonnen, F. Guettou, P. Moberg, L. Zhu, C. Jegerschöld, A. Flayhan, J. A. G. Briggs, H. Garoff, C. Löw, Y. Cheng, P. Nordlund, *Nat. Methods* **2016**, *13*, 345–351.

[186] A. Flayhan, H. D. T. Mertens, Y. Ural-Blimke, M. Martinez Molledo, D. I. Svergun, C. Löw, *Structure* **2018**, *26*, 345-355.e5.

[187] K. Kanonenberg, S. H. J. Smits, L. Schmitt, *Sci. Rep.* **2019**, *9*, 8436.

[188] T. Lasitza-Male, K. Bartels, J. Jungwirth, F. Wiggers, G. Rosenblum, H. Hofmann, C. Löw, *Angew. Chemie Int. Ed.* **2020**, *n/a*, DOI 10.1002/anie.202008226.

[189] C. R. H. Raetz, W. Dowhan, *J. Biol. Chem.* **1990**, *265*, 1235–1238.

[190] R. Rigler, E. S. Elson, *Fluorescence Correlation Spectroscopy*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2001**.

[191] "Brain Extract Total | Avanti Polar Lipids," can be found under https://avantilipids.com/product/131101, **n.d.**

[192] J. Kyte, R. F. Doolittle, *J. Mol. Biol.* **1982**, *157*, 105–132.

[193] H. Y. Aviram, M. Pirchi, H. Mazal, Y. Barak, I. Riven, G. Haran, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 3243–3248.

[194] V. Kalienkova, V. C. Mosina, L. Bryner, G. T. Oostergetel, R. Dutzler, C. Paulino, *Elife* **2019**, *8*, e44364.

[195] J. L. Rigaud, D. Lévy, *Methods Enzymol.* **2003**, *372*, 65–86.

[196] Y. Sonoda, S. Newstead, N. J. Hu, Y. Alguel, E. Nji, K. Beis, S. Yashiro, C. Lee, J. Leung, A. D. Cameron, B. Byrne, S. Iwata, D. Drew, *Structure* **2011**, *19*, 17–25.

[197] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Mol. Syst. Biol.* **2006**, *2*, 2006.0008-2006.0008.

[198] J. W. Lee, S. Choi, J. H. Park, C. E. Vickers, L. K. Nielsen, S. Y. Lee, *Appl. Microbiol. Biotechnol.* **2010**, *88*, 905–913.

[199] K. Schmid, R. Schmitt, *Eur. J. Biochem.* **1976**, *67*, 95–104.

[200] T. Te Wu, *BBA - Gen. Subj.* **1976**, *428*, 656–663.

[201] B. R. Bochner, P. Gadzinski, E. Panomitros, *Genome Res.* **2001**, *11*, 1246–1255.

[202] N. Peekhaus, S. Tong, J. Reizer, M. H. Saier, E. Murray, T. Conway, *FEMS Microbiol. Lett.* **1997**, *147*, 233–238.

[203] K. Kano, J. H. Fendler, *BBA - Biomembr.* **1978**, *509*, 289–299.

[204] T. E. Andreoli, M. Tieffenberg, D. C. Tosteson, *J. Gen. Physiol.* **1967**, *50*, 2527–2545.

[205] P. G. Heytler, *Biochemistry* **1963**, *2*, 357–361.

[206] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.

[207] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296–W303.

[208] S. Bienert, A. Waterhouse, T. A. P. De Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, *Nucleic Acids Res.* **2017**, *45*, D313–D319.

[209]   G. Studer, C. Rempfer, A. M. Waterhouse, R. Gumienny, J. Haas, T. Schwede, *Bioinformatics* **2020**, *36*, 1765–1771.

[210]   P. Benkert, M. Biasini, T. Schwede, *Bioinformatics* **2011**, *27*, 343–350.

[211]   M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, T. Schwede, *Sci. Rep.* **2017**, *7*, 10480.

[212]   N. Guex, M. C. Peitsch, T. Schwede, *Electrophoresis* **2009**, *30*, S162-73.

[213]   S. S. Ashraf, R. E. Benson, E. S. Payne, C. M. Halbleib, H. Grøn, *Protein Expr. Purif.* **2004**, *33*, 238–245.

[214]   J. L. Rigaud, B. Pitard, D. Levy, *BBA - Bioenerg.* **1995**, *1231*, 223–246.

[215]   I. Zimmermann, P. Egloff, C. A. J. Hutter, F. M. Arnold, P. Stohler, N. Bocquet, M. N. Hug, S. Huber, M. Siegrist, L. Hetemann, J. Gera, S. Gmür, P. Spies, D. Gygax, E. R. Geertsma, R. J. P. Dawson, M. A. Seeger, *Elife* **2018**, *7*, e34317.

[216]   R. L. Lieberman, J. A. Culver, K. C. Entzminger, J. C. Pai, J. A. Maynard, *Methods* **2011**, *55*, 293–302.

[217]   M. Caffrey, *J. Struct. Biol.* **2003**, *142*, 108–132.

[218]   C. Löw, P. Moberg, E. M. Quistgaard, M. Hedrén, F. Guettou, J. Frauenfeld, L. Haneskog, P. Nordlund, *Biochim. Biophys. Acta - Gen. Subj.* **2013**, *1830*, 3497–3508.

[219]   E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, in *Proteomics Protoc. Handb.*, **2005**, pp. 571–607.

[220]   I. Avanti Lipids Polar, "Phase Transition Temperatures for Glycerophospholipids | Avanti Polar Lipids," can be found under https://avantilipids.com/tech-support/physical-properties/phase-transition-temps%0Ahttps://avantilipids.com/tech-support/physical-properties/phase-transition-temps/, **2018**.

[221]   ChemViews, "Hazard Pictograms," DOI 10.1002/chemv.202000035can be found under https://www.conceptdraw.com/How-To-Guide/hazard-pictograms, **2020**.

9. Appendix

9.1. Additional Figures
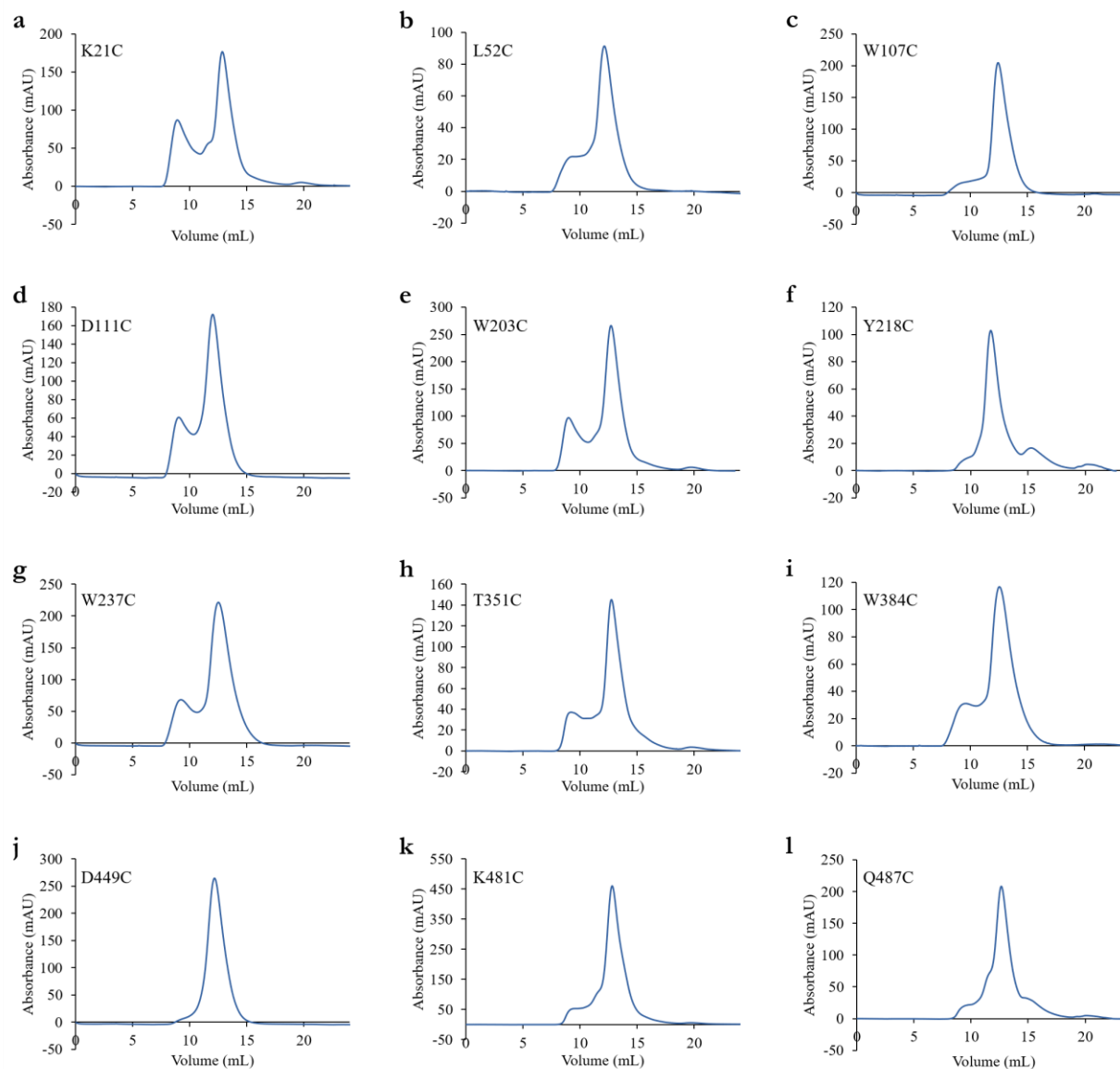
Figure 50: Purification of single mutants of DtpA. SEC profiles of single mutants. (a) K21C, (b) L52C, (c) W107C, (d) D111C, (e) W203C, (f) Y218C, (g) W237C, (h) T351C, (i) W384C, (j) D449C, (k) K481C and (l) Q487C. Absorbance was measured at 280 nm for the protein. All mutations were done in a cysteine free DtpA background (C140S/C200S/C360S).
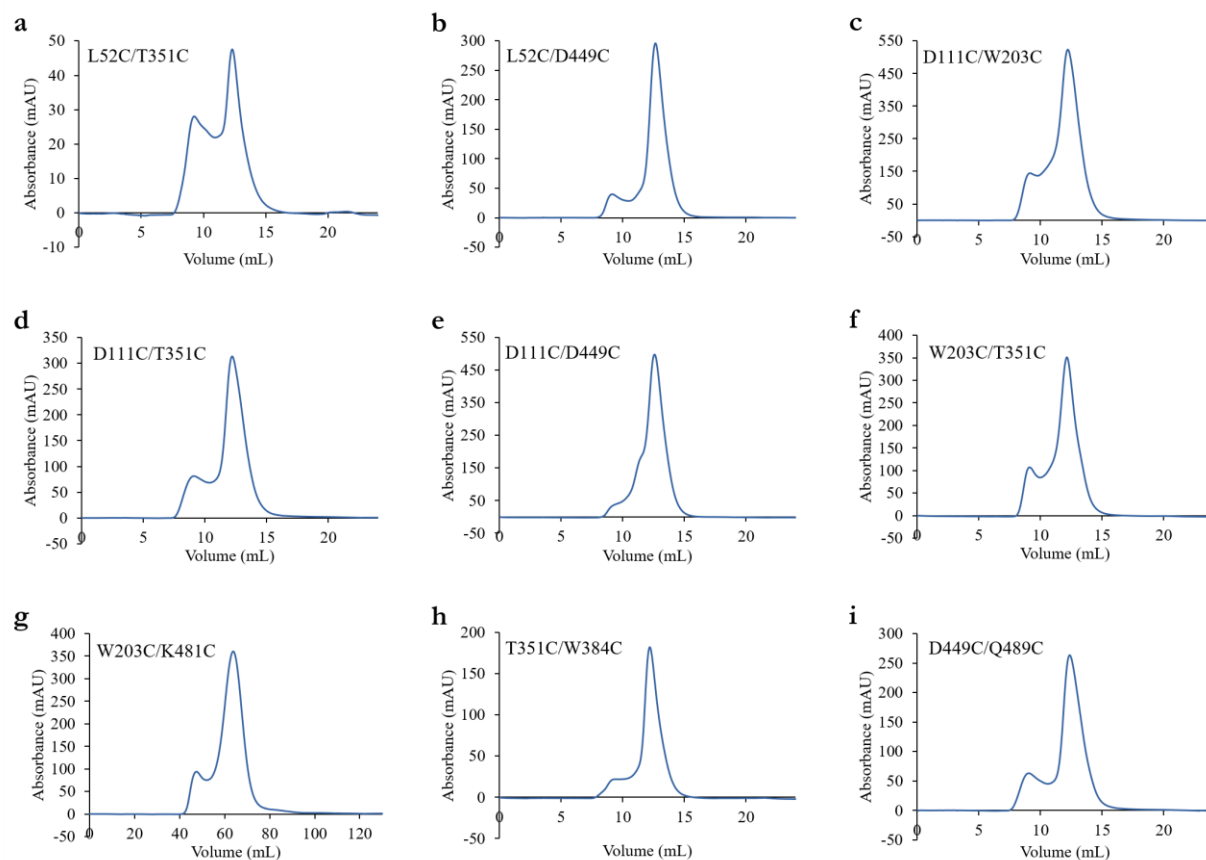
Figure 51: Purification of FRET mutants of DtpA. SEC profiles of FRET mutants. (a) L52C/T351C, (b) L52C/D449C, (c) D111C/W203C, (d) D111C/T351C, (e) D111C/D449C, (f) W203C/T351C, (g) W203C/K481C, (h) T351C/W384C and (i) D449C/Q487C. Absorbance was measured at 280 nm for the protein. All mutations were done in a cysteine free DtpA background (C140S/C200S/C360S).
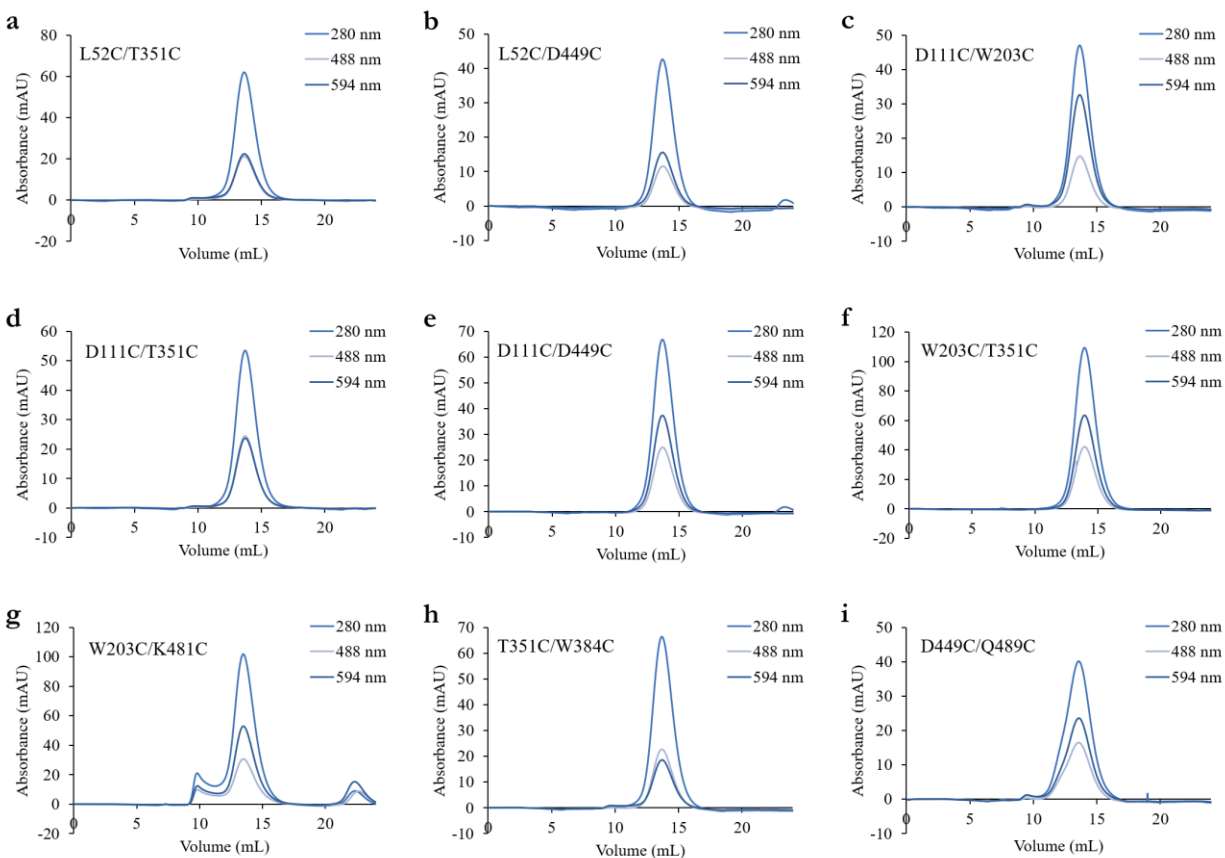
Figure 52: Labeling of FRET mutants of DtpA for smFRET experiments. SEC profiles of FRET mutants after labeling with Alexa fluorophores. (a) L52C/T351C, (b) L52C/D449C, (c) D111C/W203C, (d) D111C/T351C, (e) D111C/D449C, (f) W203C/T351C, (g) W203C/K481C, (h) T351C/W384C and (i) D449C/Q487C. Absorbance was measured at 280 nm for the protein and at 488 nm and 594 nm for the donor and acceptor FRET dye respectively. All mutations were done in a cysteine free DtpA background (C140S/C200S/C360S).
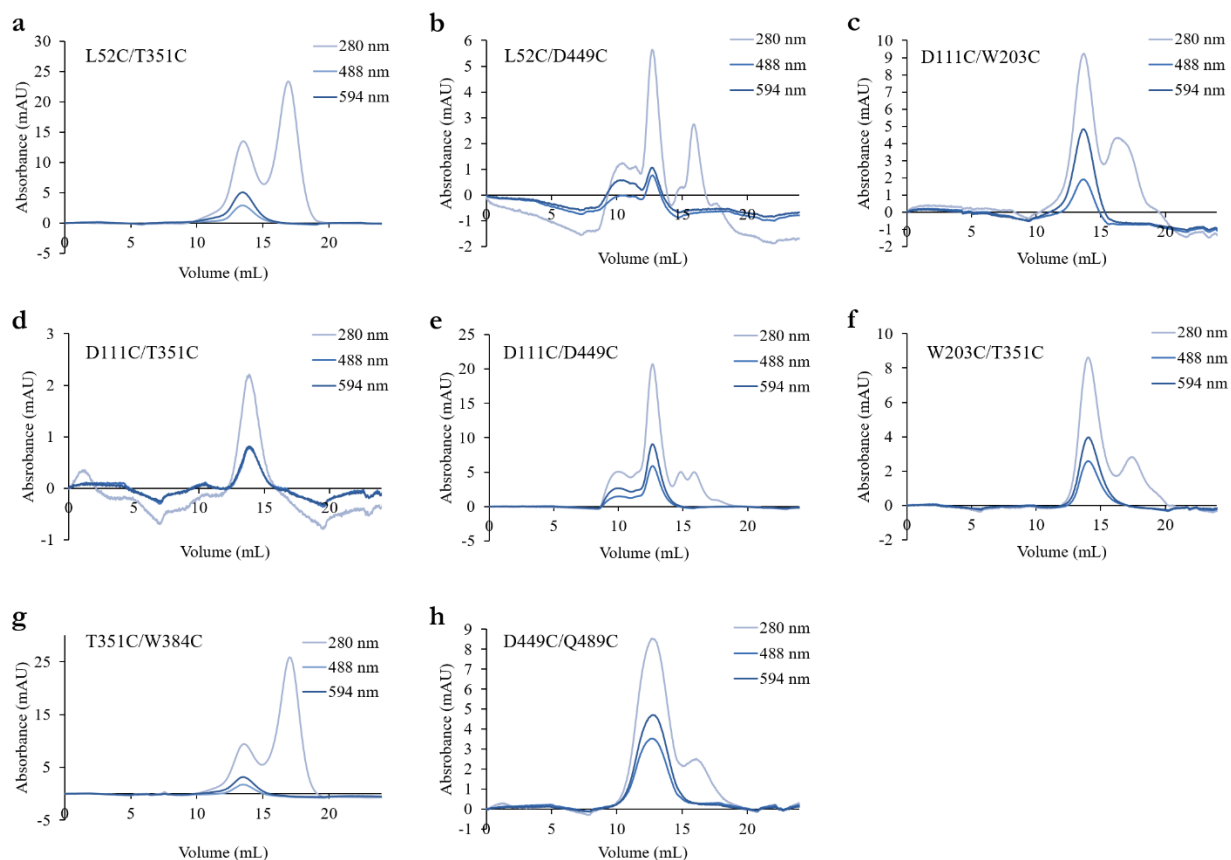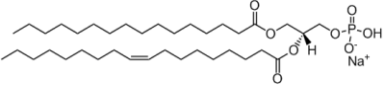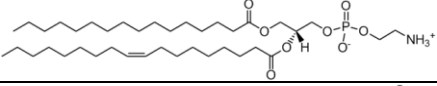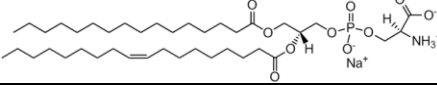
Figure 53: Reconstitution of labeled FRET mutants into SapNPs containing POPE. SEC profile of labeled (a) L52C/T351C mutant in POPE SapNPs, (b) L52C/D449C mutant in POPE SapNPs, (c) D111C/W203C mutant in POPE SapNPs, (d) D111C/T351C mutant in POPE SapNPs, (e) D111C/D449C mutant in POPE SapNPs, (f) W203C/T351C mutant in POPE SapNPs, (g) T351C/W384 mutant in POPE SapNPs and (h) D449C/Q487C mutant in POPE SapNPs. Absorbance was measured at 280 nm for the protein and at 488 nm and 594 nm for the donor and acceptor FRET dye respectively.
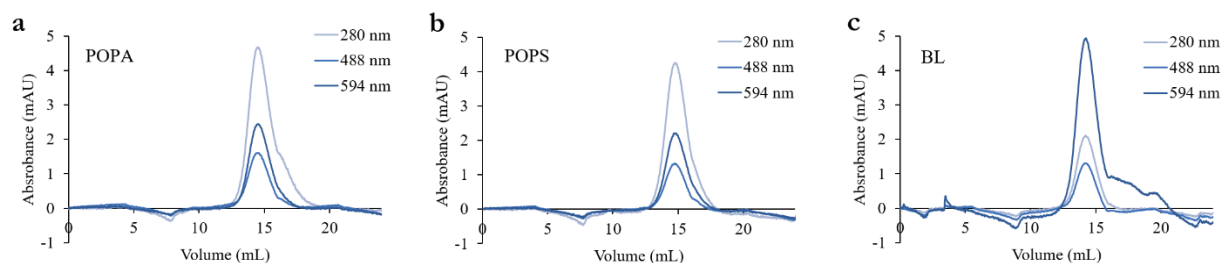
Figure 54: Reconstitution of the labeled FRET mutant W203C/Q487C into SapNPs of different lipid composition. SEC profile of labeled W203C/Q487C mutant reconstituted into (a) POPA SapNPs, (b) POPS SapNPs and (c) in BL SapNPs. Absorbance was measured at 280 nm for the protein and at 488 nm and 594 nm for the donor and acceptor FRET dye respectively.
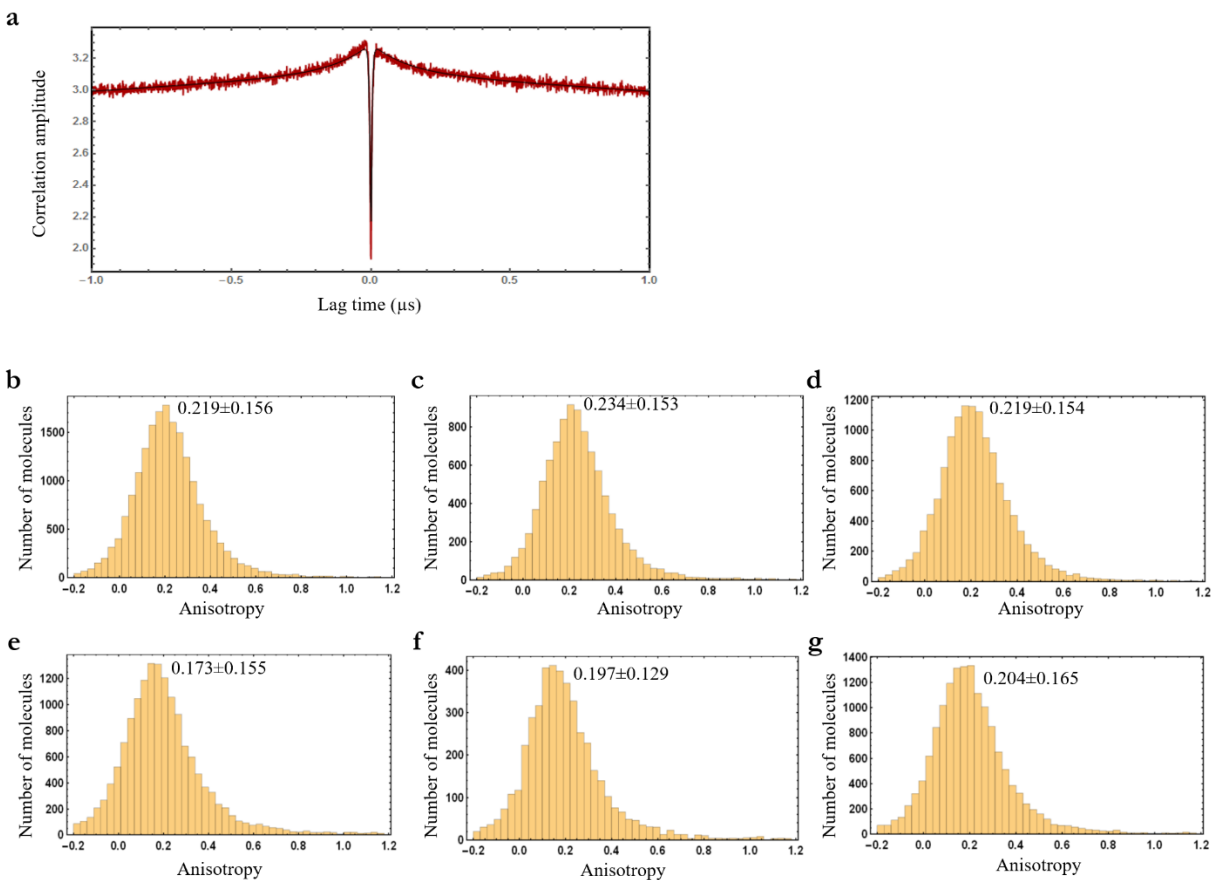
Figure 55: Nanosecond FCS (nsFCS) and anisotropy measurements for smFRET measurements. (a) NsFCS of the acceptor signal of W203C/Q487C. Measured of W203C/Q487C in POPS SapNPs after excitation at 594 nm (30 µW). Solid line is a fit including components for anti-bunching (3.8 ns), the correlated decay due to quenching (127 ns, amplitude 8.1%), and triplet blinking of the dye (3.5 s). (b-g) FRET anisotropy of the labeled species (excluding molecules with inactive acceptor). (b) W203C/Q487C in LMNG, (c) W203C/T351C in LMNG, (d) W203C/Q487C in LMNG supplied with N00, (e) W203C/Q487C in POPE SapNPs, (f) W203C/Q487C in POPS SapNPs, (g) W203C/Q487C in POPE SapNPs supplied with N00.
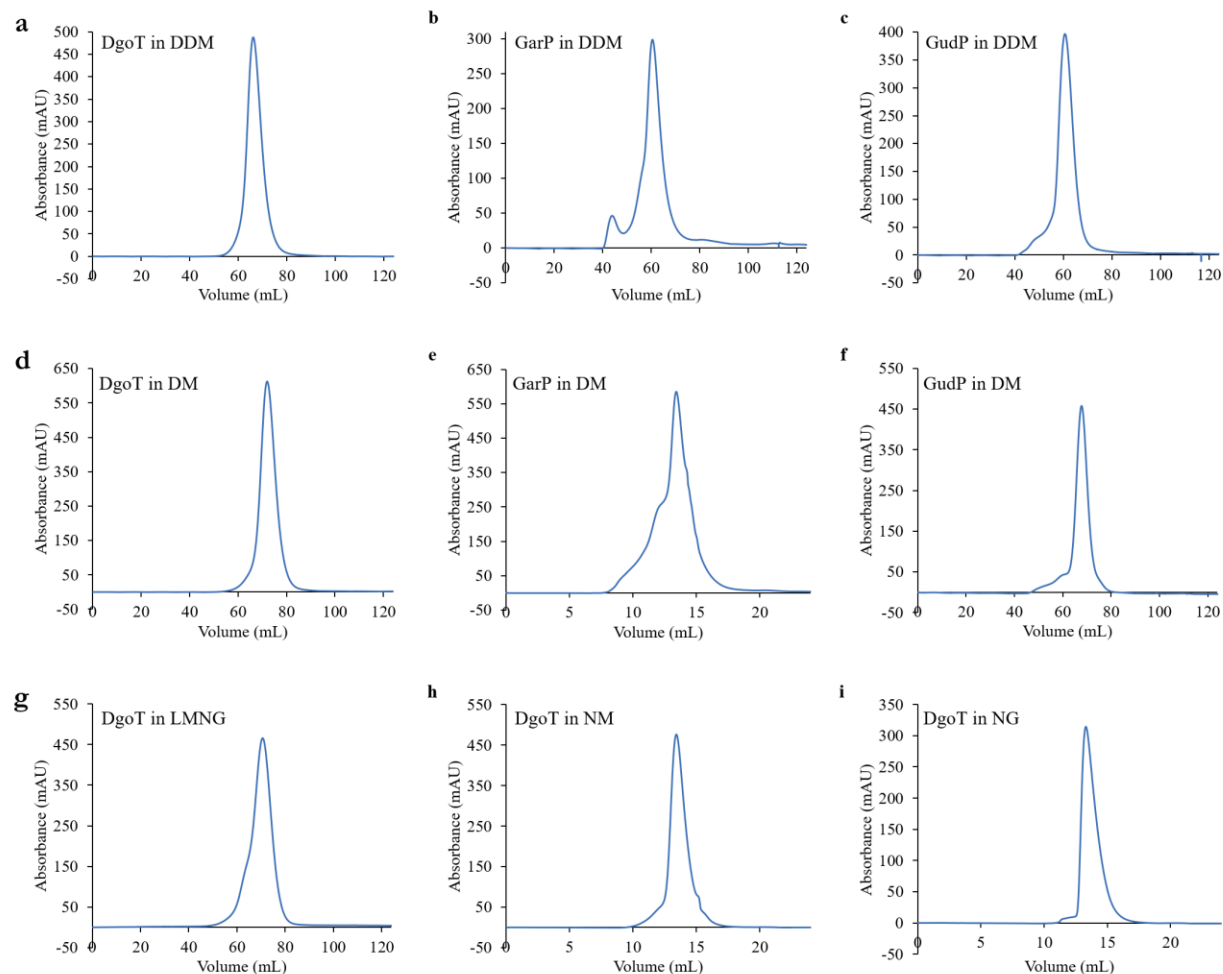
Figure 56: Purification of ACS transporters in different detergents. (a-c) SEC profile of (a) DgoT in DDM, (b) GarP in DDM and (c) GudP in DDM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for ligand binding studies and crystallization trials. (d-f) SEC profile of (d) DgoT in DM, (e) GarP in DM and (f) GudP in DM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used for reconstitution into liposomes and crystallization trials. (g) SEC profile of DgoT in LMNG. Absorbance was measured at 280 nm to detect the protein. The protein sample was used in crystallization trials. (h) SEC profile of DgoT in NM. Absorbance was measured at 280 nm to detect the protein. The protein sample was used in crystallization trials. (i) SEC profile of DgoT in NG. Absorbance was measured at 280 nm to detect the protein. The protein sample was used in crystallization trials.

9.2. Additional tables

Table 41: Lipids used for reconstitution of transporters into SapNPs.

Name	Lipid	Synthetic/Extract	Transition temperature	Structure
POPA	16:0-18:1 PA	Synthetic	28 °C [220]	
POPE	16:0-18:1 PE	Synthetic	25 °C [220]	
POPS	16:0-18:1 PS	Synthetic	14 °C [220]	
BL	Brain Total Lipid Extract	Extract	-	Mixture: 6.9 (w/w) % PC 16.7 (w/w) % PE 1.6 (w/w) % PI 10.6 (w/w) % PS 2.8 (w/w) % PA 58.7 (w/w) % unknown [191]

9.3. List of hazardous substances

Table 42: List of hazardous substances. For more information, check the 'Regulation (EC)No 1272/2008' of the European Agency for Safety and Health at Work.

Name	GHS pictogram	GHS hazardous statements (H)	GHS precautionary statements (P)
1,4-dithiothreitol	GHS07	302, 315, 319, 335	261, 305+351+338
2-Propanol	GHS02, GHS07	225-319-336	210-233-240-305+351+338-403+235
Acetic acid	GHS02, GHS05	226, 290, 314	210, 280, 303+361+353, 305+351+338, 310
Alexa Fluor 488 C5 maleimide	GHS05	302, 314, 317	260, 264, 270, 272, 280, 301+312, 301+330+331, 303+361+353
Alexa Fluor 594 C5 maleimide	GHS05, GHS07	302, 314, 317	261, 272, 301+312, 304+340, 305+351+338, 303+361+353
Ammonium chloride	GHS07	302, 319	305+351+338
Ampicillin disodium salt	GHS08	317, 334	261, 280, 302+352, 342+331
Boric acid	GHS08	360FD	201, 280, 308+313
Calcium chloride	GHS07	319	305+351+338
Calcium D-galactonate	GHS07	302+312+332, 315, 319, 335	270, 208, 301+312, 304+340, 305+351+338, 403+233, 501
Carbenicillin disodium salt	GHS08	317, 334	280, 302, 352
Carbonyl cyanide 3-chlorophenylhydrazone	GHS06	301, 311, 315, 319, 331, 335	261, 280, 301+310, 305+351+338, 311
Chloramphenicol	GHS08	351	280
Cobalt(II)chloride hexahydrate	GHS07, GHS08, GHS09	350i, 360F, 302, 317, 334, 341, 410	201, 273, 280, 302+352, 304+340, 342+311
cOmplete™ Protease Inhibitor Cocktail	GHS05	314	260, 280, 301+330+331, 303+361+353, 304+340+310, 305+351+338+310
Copper(II)chloride dihydrate	GHS05, GHS07, GHS09	302+312, 315, 318, 410	273, 280, 305+351+338
EDTA Tetrasodiumsalt	GHS07, GHS08	319, 332, 373	280, 304+340-312, 305+351+338-337+313
Ethanol	GHS02, GHS07	225, 319	210, 240, 305+351+338, 403+223
Ethidium Bromide	GHS06, GHS08	302, 330, 341	206, 281, 284, 310
GelGreen Nucleic Acid Gel Stain	GHS07	227	210, 280, 403+233, 501
Guanidine hydrochloride	GHS07	302+332, 315, 319	261, 280, 301+312, 330, 304+340+312, 305+351+338, 337+313
Hydrochloric acid	GHS05, GHS07	290, 314, 335	260, 280, 303+361+353, 304+340+310, 305+351+338
Imidazole	GHS05, GHS07, GHS08	302, 314, 360D	260, 280, 301+330+331, 303+361+353, 305+351+338, 308+313
InstantBlue™	GHS05	301, 317, 334	261, 208, 301+310, 342+311
Kanamycin sulfate	GHS07	360	201, 280, 308+313
Manganese(II)chloride tetrahydrate	GHS05, GHS07, GHS08, GHS09	302, 318, 373, 411	260, 264, 270, 280, 305+351+338+310

MES	GHS07	315, 319, 335	261, 305+351+338
Methanol	GHS02, GHS06, GHS08	225, 331, 311, 301, 370	210, 233, 208, 302+352, 304+340, 308+310, 403+235
Nickel(II)chloride hexahydrate	GHS06, GHS08, GHS09	350i, 360D, 341, 301, 331, 372, 315, 317, 334, 410	273, 281, 302+352, 304+340, 309+310
Pentane	GHS02, GHS07, GHS08, GHS09	225, 304, 336, 411	210, 240, 273, 301+330+331, 403+233
SDS	GHS02, GHS05, GHS07	228, 302+332, 315, 318, 335, 412	210, 261, 280, 301+312+330, 305+351+338+310, 370+378
Sekusept Plus	GHS05, GHS07, GHS09	302+332, 314, 400	273, 280, 303+361+353, 305+351+338, 310
Sodium hydroxide	GHS05	290, 314	280, 301+330+331, 305+351+338, 308+310
Sodium selenite	GHS06, GHS09	300+330, 315, 319, 317, 411	260, 280, 301+330+331+310, 304+340+310, 305+351+338, 403+233
Tetracycline hydrochloride	GHS07, GHS08	302, 315, 361, 362, 413	201, 202, 260, 263, 264, 273, 280, 301+312, 302+352, 308+313, 330, 332+313, 362+364
TRIS hydrochloride	GHS07	315, 319, 335	280, 302+352, 305+351+338
Tris(2-carboxyethyl)phosphine hydrochloride	GHS05	314	280, 305+351+338, 310
Valinomycin	GHS06	300, 310, 330	362, 364, 280, 301+310+330, 302+352+310, 361+364
Zinc sulfate heptahydrate	GHS05, GHS07, GHS09	302, 318, 410	273, 280, 305+351+338, 313
β-Ala-(L)-Lys-N-7-amino-4-methylcoumarin-3-acetic acid	GHS07	315, 319, 335	302+352, 305+351+338
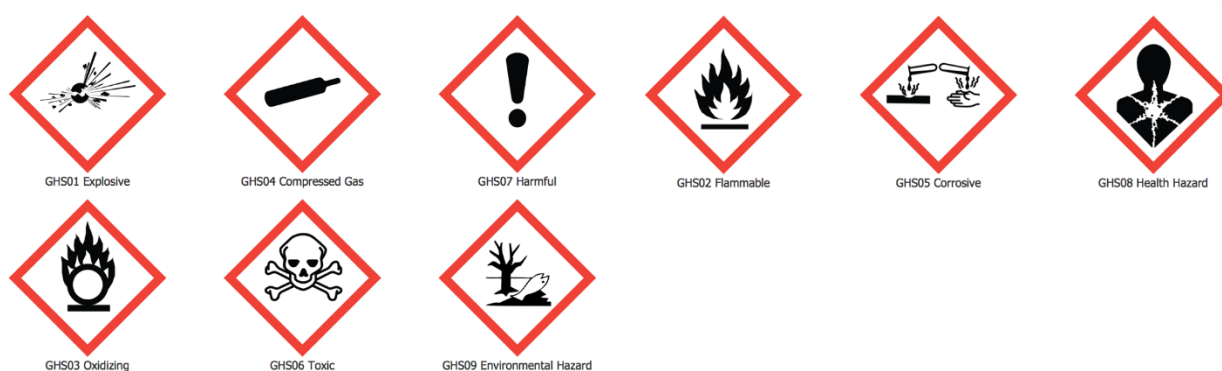
Figure 57: GHS pictograms. Figure adapted from [221].

10. Acknowledgments

11. Eidesstattliche Versicherung

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

Datum 25.09.2020

Unterschrift

Kim Bartels