Multispecific substrate recognition in a Proton-Dependent Oligopeptide Transporter

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Chemistry
Institute of Biochemistry and Molecular Biology

Maria Martinez Molledo

Hamburg, September 2018

Thesis evaluators

Dr. Christian Löw
Group leader at the European Molecular Biology Laboratory (EMBL) and the Center of Structural Systems Biology (CSSB)

Prof. Dr. Henning Tidow
Group leader at the Institute of Biochemistry and Molecular Biology (University of Hamburg)

Defense committee (21.09.2018)

Dr. Christian Löw
Group leader at the European Molecular Biology Laboratory (EMBL) and the Center of Structural Systems Biology (CSSB)

Prof. Dr. Christian Betzel
Group leader at the Institute of Biochemistry and Molecular Biology (University of Hamburg)

Prof. Dr. Andrew Torda
Group leader at the Center for Bioinformatics (University of Hamburg)

The work described in this dissertation was performed between August 2014 and July 2018 in the group of Dr. Christian Löw at the European Molecular Biology Laboratory (EMBL).

Acknowledgments

Table of contents

List of publications

1. Tripeptide binding in Proton-Dependent Oligopeptide Transporters. **Martinez Molledo M**, Quistgaard EM, Löw C. (Manuscript submitted).

2. Multispecific substrate recognition in a proton-dependent oligopeptide transporter. **Martinez Molledo M**, Quistgaard EM, Flayhan A, Pieprzyk J, Löw C. Structure 2018 Mar 6; 26 (3): 467-476.

3. Saposin lipid nanoparticles: a highly versatile and modular tool for membrane protein research. Flayhan A, Mertens HDT, Ural-Blimke Y, **Martinez Molledo M**, Svergun DI and Löw C. Structure 2018 Feb 6; 26 (2): 345-355.

4. Lipid-like Peptides can Stabilize Integral Membrane Proteins for Biophysical and Structural Studies. Veith K, **Martinez Molledo M**, Almeida Hernandez Y, Josts I, Nitsche J, Löw C, Tidow H. Chembiochem. 2017 Sep 5; 18 (17): 1735-1742.

5. Structure determination of a major facilitator peptide transporter: Inward facing PepT$_{St}$ from *Streptococcus thermophilus* crystallized in space group P3121. Quistgaard EM, **Martinez Molledo M**, Löw C. PLoS One 2017 Mar 6; 12 (3): e0173126.

List of abbreviations

7.8 MAG: 1-(7Z-pentadecenoyl)-rac-glycerol

9.9 MAG: monoolein or 1-(9Z-octadecenoyl)-rac-glycerol

Å: Ångström

ABC transport superfamily: ATP-binding cassette transport superfamily

APC: acid-polyamine-organocation transporter

CMC: critical micelle concentration

DDM: n-dodecyl-β-D-maltoside

LMNG: 2,2-didecylpropane-1,3-bis-β-D-maltopyranoside

DLS: dynamic light scattering

DM: n-decyl- β-D-maltoside

DMPC/ CHAPSO: 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (lipid)/ 3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate (detergent) bicelle

DSF: differential scanning fluorimetry

E.coli: *Escherichia coli*

EM: electron microscopy

ER: endoplasmic reticulum

F(*hkl*): structure factor

FT: Fourier transformation

GPCR: G-protein coupled receptor

HGNC: HUGO gene nomenclature committee

HPLC: high-performance liquid chromatography

I/ σ(I): Signal to noise ratio

IC_{50}: half maximal inhibitory concentration

IMPs: integral membrane protein

ITC: isothermal titration calorimetry

K_D: dissociation constant

K_i: inhibitory constant

LacY: lactose permease or lactose/ H^+ symporter

LC-MS: liquid chromatography-tandem mass spectrometry

LCP: lipidic/ lipid cubic phase method

LeuT: leucine transporter

LLP: lipid-like peptide or peptergents

MFS: major facilitator superfamily

mM: millimolar

mM: millimolar

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer

Tris-HCl: 2-Amino-2-(hydroxymethyl)propane-1,3-diol buffer

mPEG-5K: maleimide polyethyleneglycol 5000

MR: molecular replacement

MS: mass spectrometry

MSP: membrane scaffold protein

MST: microscale thermophoresis

MST: microscale thermophoresis

MX: macromolecular crystallography

NaP$_i$: sodium phosphate buffer

nl: nanoliter

NM: n-nonyl- β-D-maltoside

NMR: nuclear magnetic resonance

P1: pocket one

P2: pocket two

PA: phosphatidic acid

PCR: polymerase chain reaction

PEG: polyethylene glycol

PI: phosphatidylinositol

POTs/ PTRs/ PepTs: proton-dependent oligopeptide transporters

rmsd: root-mean-square deviations

SANS: small angle neutron scattering

SapNP: saposin-derived lipid nanoparticles

SAXS: small X-ray scattering

SEC: size exclusion chromatography

SEC: size exclusion chromatography

SLC: solute carrier

SPR: signal recognition particle

SPR: surface plasmon resonance

TCDB: transport classification database

Tm: transition midpoint, apparent melting temperature

TM: transmembrane

TRAM: translocating chain-associated membrane protein

TSA: thermal shift assay

β-Ala-Lys-AMCA/ AK-AMCA: β-alanyl-lysyl-N-7-amino-4-methylcoumarin-3-acetic acid

μg: microgram

μM: micromolar

Amino acid nomenclature code:

Amino acid	Three letter code name	One letter code name
Alanine	Ala	A
Arginine	Arg	R
Aspartic acid/ Aspartate	Asp	D
Asparagine	Asn	N
Cysteine	Cys	C
Glutamic acid/ Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of tables

List of figures

Chapter 1. Abstract

Proton-dependent oligopeptide transporters (POTs) are members of the major facilitator superfamily (MFS) of transporters, one of the largest transporter families in nature. They are widely distributed in all evolutionary lineages, also with representatives in humans. POTs have been typically described as highly promiscuous transporters, as they move di- and tripeptides across the membrane independently of their amino acid composition. Furthermore, humans POTs are also involved in drug absorption and thus have great pharmacological significance. Previous to our studies, there were very few reported structures of bacterial POTs in complex with natural ligands. Furthermore, these ligands presented poor variability in terms of physicochemical features and therefore, we aimed to explore how substrates with different characteristics can bind to these transporters.

Here we present a screening method to identify potential ligands for the bacterial POT from *Streptococcus thermophilus* (PepT$_{St}$) as well as its characterization by microscale thermophoresis (MST). We established a workflow that firstly involved the screening of a library of different di- and tripeptides by differential scanning fluorimetry (DSF). The potential substrates for PepT$_{St}$ were identified as they increased the stability of the transporter. We observed that PepT$_{St}$ favorably binds dipeptides over tripeptides, especially those with non-polar residues and polar but not charged side chains. We further characterized the identified ligands by MST, determining the K$_D$ values for individual di- and tripeptides. The resulting affinities for the measured ligands were in the millimolar range. Finally, this information was used to obtain crystallographic structures of PepT$_{St}$ in complex with chemically diverse ligands using the lipidic cubic phase (LCP) crystallization method. In total we determined nine X-ray crystallographic structures of PepT$_{St}$ (resolution range of 2.0-2.7 Å) in complex with natural peptide ligands (Ala-Leu, Ala-Gln, Asp-Glu, Phe-Ala and the tripeptide Phe-Ala-Gln), non-peptidic substrates (phosphate ions and HEPES molecule) and an apo structure.

From the structures of PepT$_{St}$ in complex with the peptidic ligands, we conclude that the coordination of the peptides from the N- and C-termini is conserved but we also realized that these are not the only relevant interactions to bind a substrate in PepT$_{St}$. As a result, the main principles of substrate binding in PepT$_{St}$ were established as follows: (1) Dipeptides can place their peptide backbone at least in two different positions in the binding cavity. (2) The size of the PepT$_{St}$ binding cavity can be fine-tuned by the highly conserved residue Tyr-68. (3) The solvation state changes in

the binding cavity upon ligand binding and water molecules play an important role in ligand coordination. (4) Tripeptide binding was only observed for peptides with Phe-Ala-Xxx sequence, which is enabled by the bending of the peptide backbone in the middle position to fit into the reduced size of the binding cavity.

In conclusion, our studies shed light on the molecular mechanisms of peptide recognition and binding in bacterial POTs, which can be extrapolated to other members of the family due to the high conservation of the binding site residues among POT transporters. Furthermore, we have now established a workflow for the study of peptide binding in solution, which is applicable to other members of this transporter family.

Kapitel 1. Zusammenfassung

Protonen-abhängige Oligopeptidtransporter (POTs) gehören zur Gruppe der Major Facilitator Superfamily (MFS) Transporter, die eine der größten Transporter Proteinfamilien in der Natur darstellt. POTs sind weit verbreitet in allen evolutionären Abstammungslinien und kommen auch im Menschen vor. Sie sind typischer Weise als sehr substrat-unspezifisch beschrieben worden, weil sie Di- und Tripeptide unabhängig von ihrer Aminosäurezusammensetzung durch die Zellmembran transportieren können. Darüber hinaus sind POTs im Menschen auch bei der Aufnahme von Medikamenten beteiligt und sind daher von hoher pharmakologischer Bedeutung. Vor Beginn meiner Studien waren wenige Röntgenkristallstrukturen von POTs in Verbindung mit natürlichen Substraten beschrieben worden. Die verwendeten Ligandsubstrate repräsentieren zudem eine geringe Variabilität in Beziehung auf ihre physikochemischen Eigenschaften und ein Ziel meiner Arbeit war es daher herauszufinden, wie Substrate mit verschieden Eigenschaften an diese Transporter binden können.

In meiner Arbeit beschreibe ich eine Screening-Methode um potenzielle Substrate eines bakteriellen POT aus *Streptococcus theromophilus* (PepT$_{St}$) zu identifiezieren, sowie die Charakterisierung dieses Transporters mittels Microscale Thermophorese (MST). Es wurde ein Workflow etabliert, der im ersten Schritt ein Screening einer Bibliothek von verschiedenen Di- und Tripeptiden mittels Differential-Scanning-Fluorimetrie (DSF) beinhaltet. Die potentiellen Substrate für PepT$_{St}$ konnten durch die Erhöhung der Stabilität des Transporters identifiziert werden. Beim Vergleich zwischen Di- und Tripeptiden konnte beobachtet werden, dass PepT$_{St}$ Dipeptide bevorzugt bindet, und vor allem solche Peptidketten, die unpolare Aminosäuren und polare, nicht-geladene Seitenketten besitzen. Die identifizierten Ligandsubstrate konnten mittels MST Methodik weiter charakterisiert werden, um K$_D$ Enzymkonstanten für individuelle Di- und Tripeptide zu bestimmen. Die ermittelten Affinitäten für die gemessenen Liganden waren im millimolaren Bereich. Diese Information wurde im letzten Schritt dazu benutzt um kristallographische Strukturen von PepT$_{St}$ in Verbindung mit chemisch-diversen Liganden mittels Lipidic Cubic Phase (LCP) Kristallographie zu bestimmen. Insgesamt konnten dadurch acht individuelle Röntgenkristallstrukturen mit einer Auflösung von 2.0 – 2.7 Å bestimmt werden und beinhalten Kristallstrukturen in Verbindung mit natürlichen Peptidliganden (Ala-Leu, Ala-Gln, Asp-Glu, Phe-Ala, Phe-Ala-Gln), Strukturen in Verbindung mit nicht-Peptid Puffersubstraten (Phosphationen, HEPES Molekül) und eine Apoproteinstruktur.

Anhand der Struktur von PepT$_{St}$ in Verbindung mit Peptidliganden können wir rückschließen, dass die N- und C-terminale Koordination der Peptide in der Bindefalte konserviert ist, aber wir können auch feststellen, dass diese Koordination nicht die einzige relevante Wechselwirkung für die Bindung eines Substrates durch PepT$_{St}$ ist. Es wurden infolgedessen die folgenden Hauptprinzipien der Substratbindung durch PepT$_{St}$ wie folgt etabliert: (1) Das Rückgrat von Dipeptiden kann in mindestens zwei verschieden Positionen in der Bindefalte untergebracht werden. (2) Die Größe der PepT$_{St}$ Bindefalte kann durch den hochkonservierten Aminosäurerest Tyr-68 feinreguliert werden. (3) Der Grad der Solvatisierung der Bindefalte verändert sich im Zuge der Bindung eines Liganden und Wassermoleküle spielen eine Rolle bei der Koordination eines Liganden. (4) Die Bindung von Tripeptiden konnte nur für Peptide mit Phe-Ala-Xxx Sequenzen beobachtet werden und wird durch die Krümmung der Peptidkette in der mittleren Position gewährleistet, wodurch das Peptidmolekül an die kleine Bindefalte angepasst werden kann.

Insgesamt konnten meine Untersuchungen neue Einsichten in die molekularen Mechanismen der Erkennung von Peptiden durch bakterielle protonen-abhängige Oligopeptidtransporter (POTs) liefern, die auf Grundlage der hohen Konservierung der Bindestellen von POTs extrapoliert, und auf andere Mitglieder dieser Proteinfamilie angewendet werden können. Darüber hinaus konnte ein Workflow für die Untersuchung der Peptidbindung in Lösung etabliert werden, der sich auf andere Mitglieder dieser Transporter-Proteinfamilie anwenden lässt.

Chapter 2. Introduction

2.1 Biological membranes

Biological membranes are essential for life. They are necessary for the existence of an independent cell, as they separate the cellular components from the environmental ones. Prokaryotic cells have one cell membrane as for Gram-positive microorganisms, or an inner and an outer membrane in the case of Gram-negative organisms. Eukaryotes have also a single cell membrane but they have further evolved to compartmentalize the cell into function-specific organelles (nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, transport vesicles, etc.). Thus, the membrane in the organelles varies in terms of composition to accomplish their function specifically.

Membranes are a physical barrier between the intracellular media and the exterior but on the other hand, a cell needs to be in close contact with the environment too. Therefore, biological membranes are permeable, allowing the cell to obtain external nutrients to fulfill their energetic requirements. Thus, the maintenance of the cell homeostasis is possible due to the active transport across the plasma membrane and the responses the cell gives to the environmental stimuli.

In terms of composition, biological membranes consist of lipids, proteins and carbohydrates (Figure 1). Lipids are arranged in a bilayer. This arrangement is possible due to amphipathic nature of lipids: the polar head groups are oriented towards the aqueous medium, whereas the hydrophobic tails are interacting among each other.
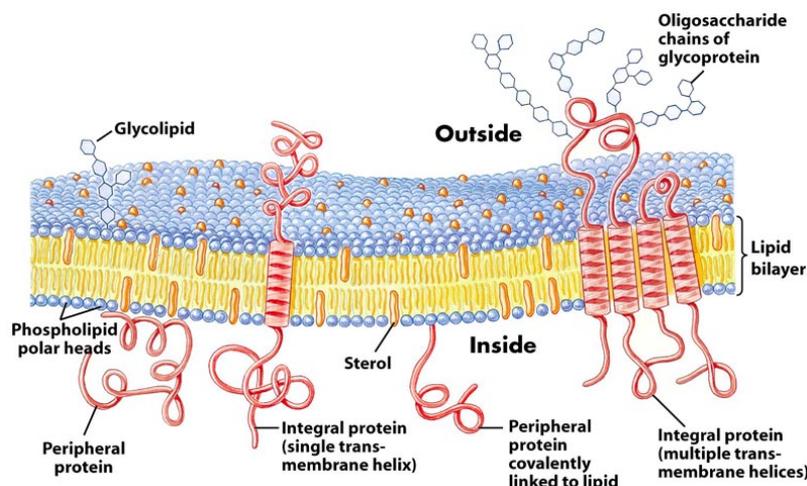
Figure 1: Biological membrane components. Membranes are composed of lipids, proteins and carbohydrates. Proteins can span through the membrane by one or several transmembrane regions, or can interact with the periphery of the membrane. Figure from Nelson and Cox, 2000.

Proteins can be bound in the periphery of the membrane (peripheral proteins) or integrated in the membrane itself (integral membrane proteins or IMPs). IMPs are integrated in the membrane due to their amphipathic structure; that is, their ionic and highly polar groups are protruding from the membrane into the aqueous phase, whereas their large nonpolar regions or transmembrane regions (TMs) are buried in the hydrophobic interior of the membrane. Furthermore, in eukaryotic cells proteins can be post-translationally modified and incorporate carbohydrates bound to their residues.
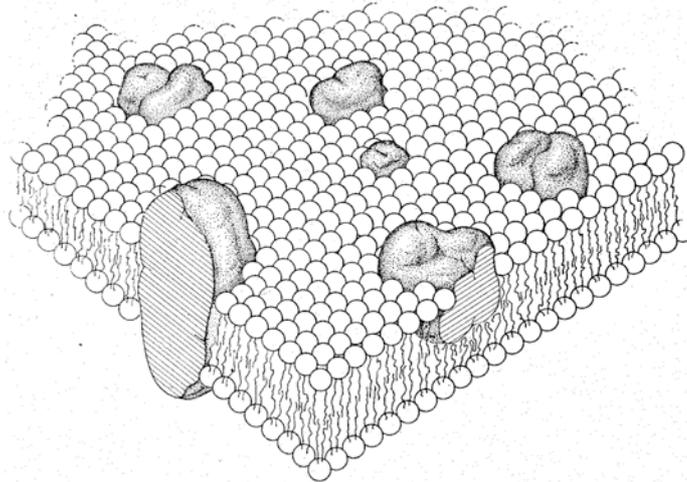
Figure 2: The fluid mosaic model as described by S. J. Sanger and G. L Nicolson. Here IMPs are represented as solid bodies, randomly distributed in a phospholipid matrix. Phospholipids are represented with their polar heads as white circles, and wavy lines represent the fatty acid chains. Figure from Sanger and Nicolson, 1972.

2.1.1 Biological membrane models

Membranes are highly dynamic structures. Their components can diffuse laterally and from one leaflet of the membrane to the other, resulting in a highly asymmetric structure. In 1972, Sanger and Nicolson proposed the *'Fluid Mosaic Model'* to explain the main properties of biological membranes (Sanger and Nicolson, 1972). Here they described the membrane components and their arrangement in the membrane (see section 2.1.2). In brief, biological membranes were defined as thermodynamically favored arrangements of lipids, proteins and oligosaccharides, which constitute a fluid, heterogeneous and asymmetric assembly (Figure 2). Thus, phospholipids are arrange in a discontinuous bilayer with apparent no long-range order and in a fluid rather than crystalline state.

On the other hand, it was already observed at this point that a membrane might not always be a mosaic with randomly distributed lipids and proteins, as certain experimental evidence suggested: For example, in the case of myelin, it was described that the membrane is highly enriched in cholesterol and this could be correlated to the cellular function. Furthermore, IMPs can also present short-range order in the membrane. There are proteins that congregate in a particular cell type

membrane and in high abundance, even with a particular stoichiometry. Some examples are the rhodopsins in the retinal cells, the components of the electron transport chain in mitochondria or the synapse membrane components.
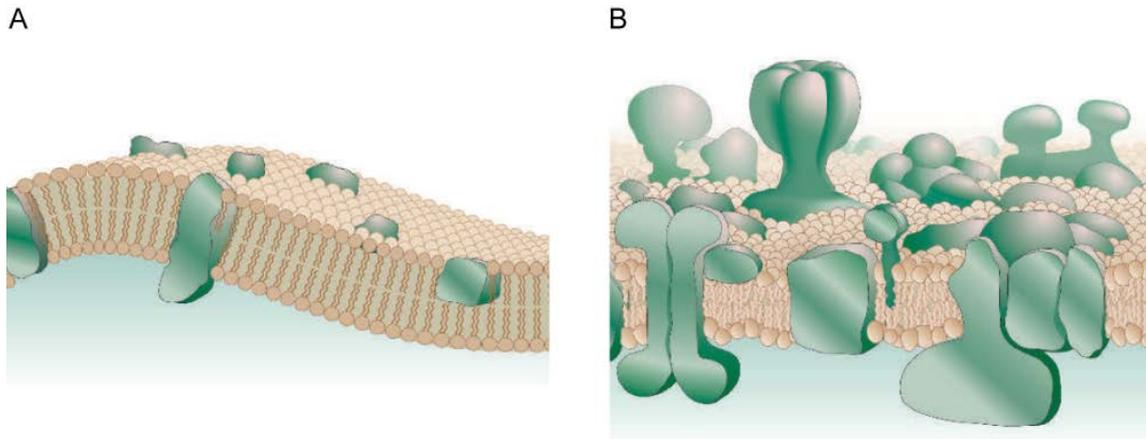
Figure 3: General models for membrane structure. (A) Biological membrane as described by Sanger and Nicolson. (B) Biological membrane as described by Engelman, with a variable patchiness, thickness and high protein occupancy. Figure from Engelman, 2005.

This model did not anticipate that there are areas in the membrane that are crowded with proteins and lipids interacting in a specific manner. Typically, these areas are enriched in cholesterol and sphingolipids (known as lipid rafts) and lipid-anchored proteins. They fulfill important roles in the cell, such as membrane trafficking or cell signaling. As a result, the membrane is thicker and less fluidic (Simons and Ikonen, 1997).

Finally, as a later contribution to biological membranes description, Engelman proposed a reviewed model (Engelman, 2005). Here, three main concepts were revised: (1) membrane protein distribution is dispersed in a lipid matrix, (2) membrane proteins appear in low concentration in membranes, and (3) the hydrophobic dimensions of membrane proteins match the imperturbable thickness of the lipid membrane. Instead, a membrane was described as a patchy structure, with oligomeric membrane proteins or complexes, as well as regions with specific lipid-lipid and lipid-protein interactions (Figure 3). Furthermore, the membrane thickness is not constant and there are areas in which the protein or the lipids distort the membrane dimensions. Thus, this would further affect membrane fluidity. Finally, the occupancy of proteins and lipids might be different according to the cell type and cell function, but in general great protein occupancy is observed. Additionally, membrane proteins can present large hydrophilic ectodomains, expanding out of the membrane (Figure 3B).

2.1.2 Elements in biological membranes

As already mentioned, biological membranes consist of lipids, proteins (peripheral or integral) and carbohydrates. In this section, the main features of these components are summarized.
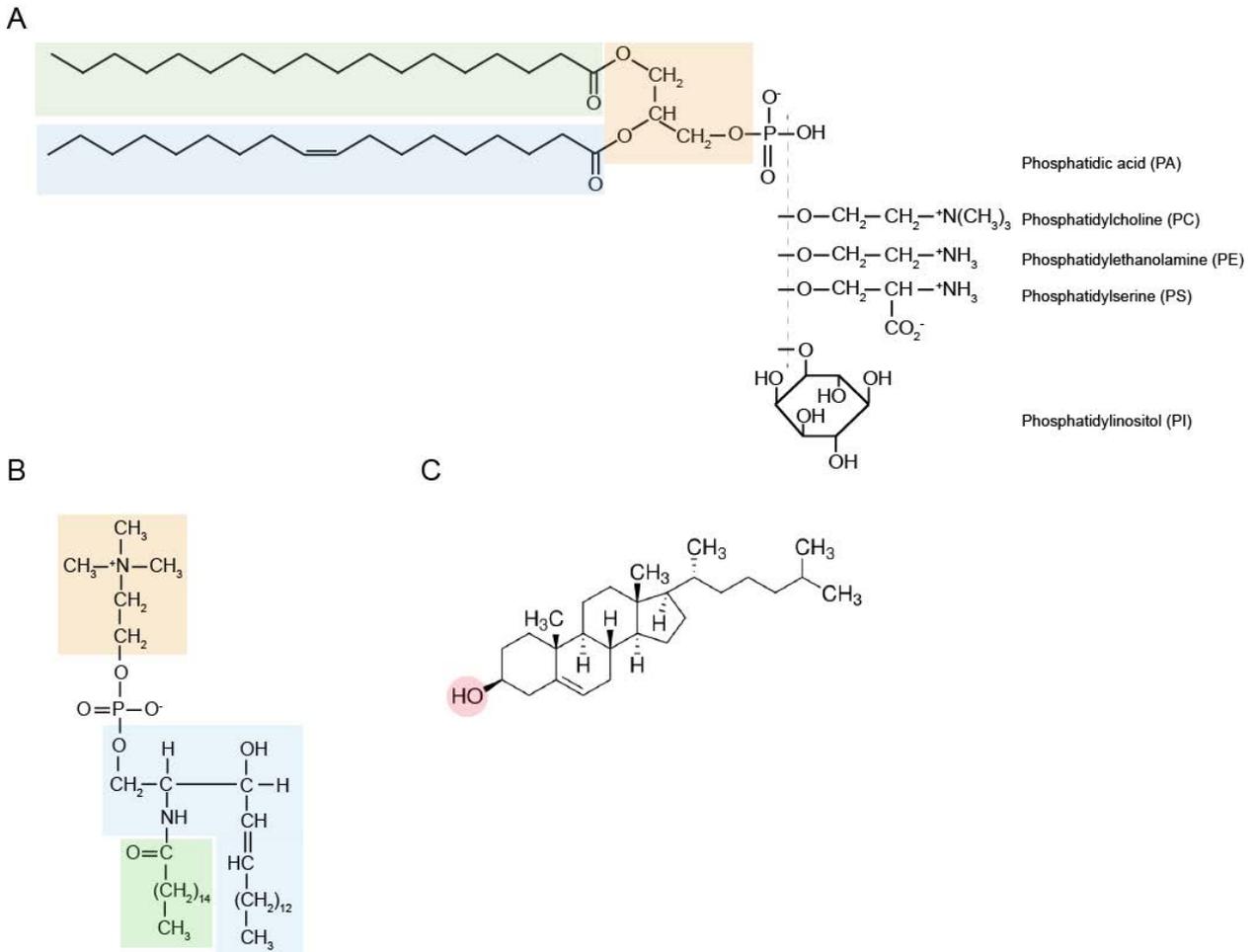
Figure 4: Elements in biological membranes. (A) Glycerophospholipids. The glycerol molecule (in the orange box) is bound to a phosphate group, forming glycerol-3-phosphate. Positions one and two are esterified by fatty acids: the fatty acid in position one (in the green box) has 16 or 18 carbons while the fatty acid in position two (in the blue box) is longer and unsaturated. Different alcohols generate the polar head group of the phospholipid. (B) Sphingolipids. The sphingosine molecule (in the blue box) is bound to a fatty acid (green box) by an amide bond. Typical head groups in sphingolipids are either a phosphocholine (as in the example, in orange box) or a phosphoethanolamine. (C) Cholesterol. Cholesterol is a cyclic polyisoprenoid, common in animal membranes. It has a small polar head group, indicated by the red cycle.

2.1.2.1 Lipid diversity in biological membranes

Lipids are a heterogeneous group of biomolecules. They are soluble in nonpolar solvents and apparently insoluble in water. Furthermore, they are often amphipathic with a large nonpolar

fraction and a polar component, which can vary in terms of size, hydration and charge. Both polar and nonpolar lipids are found in biological membranes. Polar lipids comprise phospholipids, glycolipids and sphingolipids, whereas among the nonpolar lipids, isoprenoids and acylglycerols are found. The latter, even though still amphipathic, have a very small polar head.

- Phospholipids (or glycerophospholipids). They are built on a glycerol molecule derivatized to glycerol-3-phosphate by incorporating a phosphate group in the third carbon of the glycerol molecule (becoming phosphatidic acid, PA). Carbons one and two are esterified by fatty acids. Typically in biological membranes, the fatty acid in position one has 16 or 18 carbons while the fatty acid in position two is longer and unsaturated. The esterification of the phosphoric group by different alcohols generates the polar head group of the phospholipid (Figure 4A). Furthermore, membranes present different phospholipid proportions based on the function they accomplish. To cite few examples, membranes enriched with phosphatidylinositol (PI) are typically involved in signaling processes; phosphatidylethanolamine (PE) participates in membrane fusion events and regulates membrane curvature (Dowhan, 1997); phosphatidylcholines are abundant in the lungs, reducing the surface tension (Bernhard, 2016).

- Sphingolipids. Sphingolipids are not built on a glycerol molecule but a sphingosine one instead. Additionally, a fatty acid is attached via an amide bond. Typical head groups in sphingolipids are either a phosphocholine or a phosphoethanolamine, also known as sphingomyelins (Figure 4B).

- Glycolipids. Even though classified in a separate group, they only differ with respect to phospholipids and shingolipids in the head group, an oligosaccharide. Therefore, there are glyceroglycophospholipids and sphingoglycophospholipids, such as cerebrosides (with monosaccharide head groups) and gangliosides (with oligosaccharides instead).

- Isoprenoids. Lipids in this group derive from isoprene. Linear isoprenyl groups, such as farnesyl (with 15 carbons) or geranylgeranyl (with 20), are involved in anchoring proteins to the membrane. Dolicols are long polyisoprenoid lipids (with 90 carbons) involved in the attachment of sugar moieties to newly synthetized membrane proteins in the endoplasmic reticulum (ER). Finally, sterols are cycled polyisoprenoids; in animal membranes, cholesterol (Figure 4C) is the dominant one and its content can vary from 0-25% in a membrane influencing the biophysical properties of the membrane (de Meyer and Smit, 2009). Typically, membranes enriched in

cholesterol are thicker and less fluid. Other eukaryotic sterols are ergosterol (in yeast and fungi), and sitosterol and stigmasterol (in plants).

Another consideration of lipids in the membrane is the possibility of diffusing laterally and transversely. Whereas the lateral diffusion is relatively fast, the transverse movement of lipids is slow and energetically unfavorable. Thus, it requires enzyme catalysis or vesicle transport (Sharom, 2011). As a result, a membrane is asymmetric and the thickness and fluidity can vary depending in the lipid composition.

2.1.2.2 Carbohydrates in biological membranes

Glycosylation is a universal feature in biological cells (Esko et al., 2009). For example, in bacteria, archaea, and fungi, glycans form the cell wall. They are involved in resisting osmotic differences between the cytoplasm and the environment and in bacteria, they may also have a role in defense against bacteriophages or antibiotics generated by other microorganisms in the environment. In eukaryotic cells, glycans are exposed to the extracellular side of the cell membrane. They can be bound to proteins (glycoproteins or peptidoglycans) or lipids (glycolipids). Glycans are synthetized in the ER and the Golgi apparatus; here, enzymes such as glycosyltransferase and sulfotransferases allow the binding and extension of glycan chains to proteins (Sprovieri et al., 2018). Glycosylated proteins have relevant functions in cell-cell recognition and interaction.
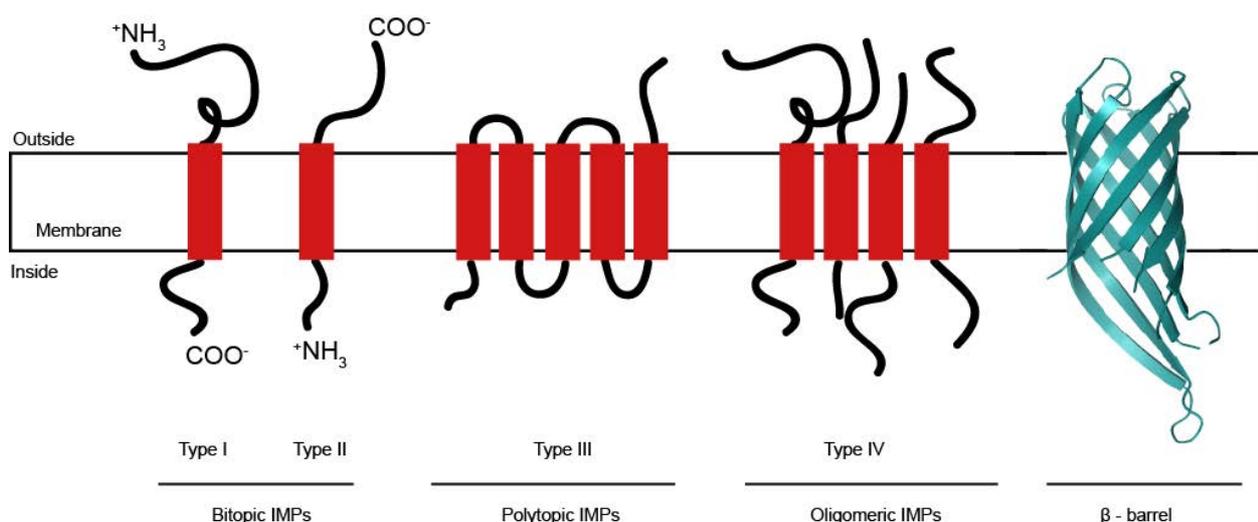
Figure 5: IMPs topology in biological membranes. Bitopic IMPs have a single TM spanning through the membrane. They are classified as type I or II depending on the orientation of the N- and C-termini. Type I IMPs have the N-terminus in the extracellular side of the membrane or towards the lumen or the organelle. On the contrary, type II have the N-terminus oriented to the cytosol. Polytopic IMPs have several TMs spanning through the membrane, with

connecting loops between them. Finally, type IV IMPs are oligomers formed by several polypeptide chains, each of them with a single TM. A different fold in IMPs are β-barrels, here illustrated by the porin OmpA (Pautsch et al., 2000; PDB ID 1QJP). Figure adapted from Nelson and Cox, 2000.

2.1.2.3 Proteins in biological membranes

Proteins that are somehow linked to the membrane can be classified in two groups: (1) peripheral or extrinsic proteins bind to the membrane surface by electrostatic interactions established to lipids or other membrane proteins. This group also comprises proteins anchored to the membrane via direct covalent bonds to lipids and secondly, amphitropic proteins which are weakly and reversibly bound to the membrane. (2) On the contrary, integral membrane proteins (IMPs) span completely through the membrane one or several times. As the research described further in this dissertation focuses on IMPs, only this type of membrane proteins will be described in more detail.

IMPs are embedded (totally or partially) in the membrane. Within this group, there are IMPs oriented in parallel to the membrane plane and they are integrated only in one leaflet of the membrane. These proteins are called monotopic and they have loops or amphipathic helices. Here the hydrophobic residues face the membrane and interact with the lipids. More common are the IMPs spanning once (bitopic) or several times (polytopic) through the complete membrane (Ott and Lingappa, 2002). The latter can adopt a β-barrel or α-helical structural fold (Figure 5).

An interesting consideration about IMPs is their biogenesis and insertion in the lipid bilayer. *In vivo*, insertion of α-helical IMPs is performed by the translocation apparatus (White and von Heijne, 2004). The translocon is conserved in all kingdoms of life, although there are differences in the complex components. Following-up, the biogenesis of plasma membrane α-helical IMPs is summarized.

2.1.2.3.1 α-helical IMPs biogenesis

Since IMPs are not synthetized by a special ribosomal population, the nascent polypeptide is targeted towards the membrane as soon as it exits the ribosome. In prokaryotes, IMPs destination is the plasma membrane. Furthermore, in Gram-negative bacteria the outer membrane is an additional end for IMPs. In eukaryotes, IMPs destined to the plasma membrane are synthetized in the ER, whereas proteins with mitochondrial, chloroplastal and nuclear locations might follow different synthesis and trafficking pathways.

There are different types of signals that target a nascent peptide towards the membrane: (1) signal sequences, (2) signal-anchors and (3) reverse signal-anchors. Signal sequences located in N-terminus have ~20 residues sometimes followed by a cleavable site. For those IMPs lacking a signal sequence, the first TM might function as a signal-anchor, initiating the topogensis of the IMP. In other cases, the signal-anchor might be located later in the protein sequence or even at C-terminus. Finally, reverse signal-anchors impose a particular orientation of the IMPs relative to the membrane. They are often found in type I IMPs (N-terminus oriented towards the extracellular side of the membrane and C-terminus towards the cytoplasm).
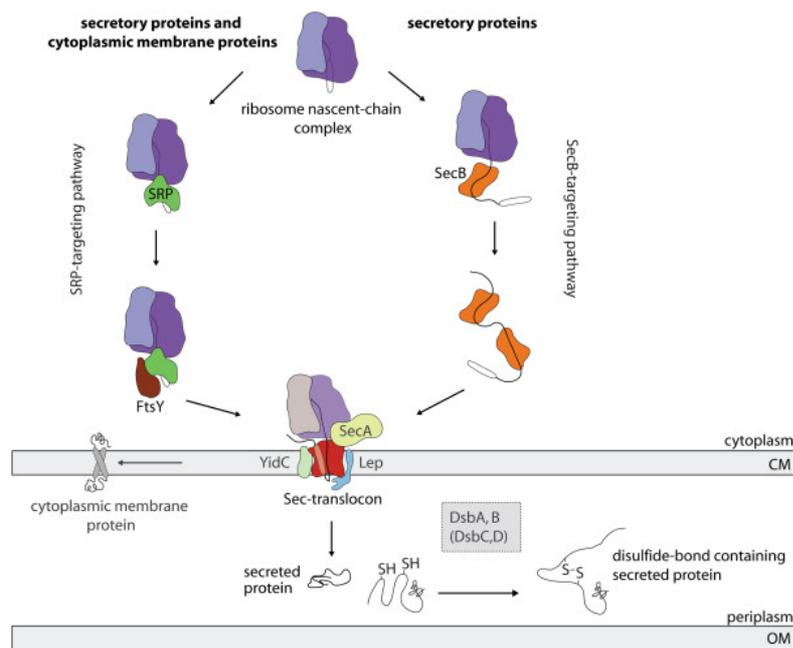
Figure 6: IMPs biogenesis in bacteria. In *E. coli*, most secretory and cytoplasmic membrane proteins require the SecYEG translocon for their biogenesis. Secretory proteins or proteins destined to the outer membrane are equipped with a signal sequence. The signal sequence determines whether a protein is targeted to the translocon via the post-translational SecB-targeting pathway or the co-translational signal recognition particle (SRP)-targeting pathway, which is comprised of the SRP and its receptor FtsY. The latter pathway is common to IMPs that are inserted in the plasma membrane. Figure from Schlegel et al., 2013.

Signal sequences are highly variable but despite the diversity, the function is preserved. Once recognized, the nascent polypeptide binds to chaperones, protecting the nascent peptide from aggregation. Then, it is targeted towards the membrane: a ribonucleoprotein complex called the signal recognition particle (SRP) moves the nascent peptide to the ER in eukaryotes or to the inner membrane in bacteria. Once in the vicinity of the membrane, the SPR is recognized by its receptor and finally, the ribosome associates to the translocon complex. The translocon is a protein-conducting channel with a hydrophilic interior. It is form by a heterotrimeric membrane protein

complex called Sec61 in eukaryotes and SecYEG complex in prokaryotes (Osborne et al., 2005). The Sec61/ SecYEG complex is a passive pore that requires different coupled proteins to drive the translocation of the nascent polypeptide. This mechanism is common between IMPs and soluble proteins that are either exported to the periplasm in bacteria or translocated to the lumen of the ER in eukaryotes (Schlegel et al., 2013). For IMPs, the polypeptide is inserted into the membrane. Here, the TM segments have to move from the aqueous and hydrophilic pore of the translocon to the hydrophobic environment of the lipid bilayer. The N-terminus of a protein can be on either side of the membrane, but the insertion of the first TM of a polytopic protein will affect the orientation of the rest of the TMs. In the case of eukaryotes, the Sec61 system works together with the translocating chain-associated membrane protein (TRAM) to facilitate the insertion of the TM segments; they are recognized and integrated laterally into the membrane. Finally, it is also subject of discussion whether the insertion of the first TM is followed by the subsequent integration of the rest of TMs or not. It is plausible as well that certain TMs have higher tendency to insert into the membrane and would, therefore, help other TMs.

The previous mechanism of IMPs insertion into the membrane is the one that has been most extensively studied, but it is the only (Osborne et al., 2005). For example, in yeast but it is likely to happen in higher eukaryotes, the signal sequence of the nascent polypeptide is not recognized by the SPR but instead, the newly synthetize polypeptide chain is stabilized in the cytoplasm by a set of cytosolic chaperones. Later, the complex is moved to the Sec61 complex. There, the cytosolic chaperones are released and the polypeptide chain is translocated in a process driven by ATP hydrolysis. In this case, a set of additional proteins is required in the translocation together with Sec61. Also in bacteria, mostly for synthesis of the outer membrane proteins, there is posttranslational translocation mode. In this case, the cytosolic ATPase SecA binds to the newly synthetized polypeptide chain; the ATP hydrolysis leads to a conformational change in SecA, helping the translocation of the protein through the SecYEG channel. This mechanism is commonly used for secretory proteins in bacteria and it has common features to the β-barrel IMPs biogenesis.

2.1.2.3.2 Functions accomplished by IMPs

IMPs fulfill multiple functions in the membrane, primarily responding to the stimuli from the environment. IMPs can function as enzymes, transporters, channels and receptors:

(1) Membrane enzymes catalyze important reactions for transport and signaling. In this group, it is worth mentioning the IMPs constituting the electron-transport chain and others involved in metabolism. (2) Transporters are involved in the movement of molecules across the membrane. As later discussed, there are different mechanisms of transport (see section 2.2) adapted to the chemical features of the molecule to be transported. (3) Channels are selective ion pores that drive ions from one side of membrane to the other. They play a crucial role in the polarization and depolarization of membranes, and have important functions such as the transmission of the nerve impulse through the nervous system. (4) Finally, receptors are IMPs involved in signal transmission. Upon a ligand-binding event, an intracellular response starts triggering different cellular effects, such as change in Ca^{2+} concentrations, phosphorylation events, activation or deactivation of downstream signaling proteins, etc.

For the rest of the dissertation, membrane proteins involved in nutrient transport across the membrane will be described in more detail.
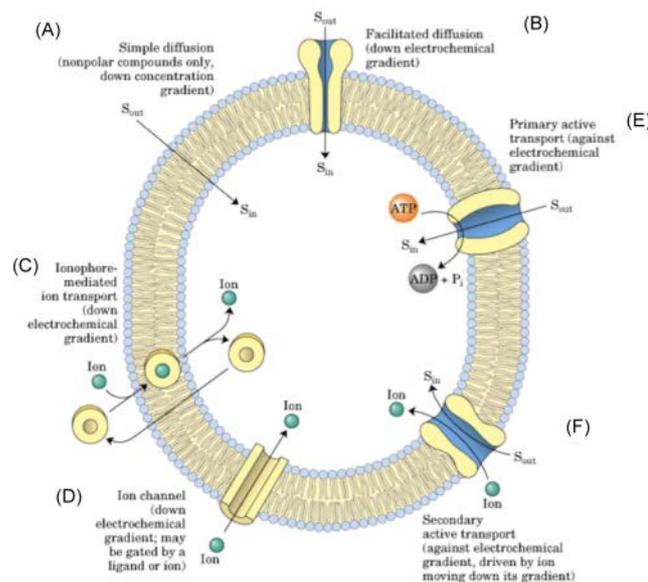
Figure 7: Transport through biological membranes. (A) Lipidic or non-polar molecules can cross the membrane down a concentration gradient, without the participation of transporters or channels; this mechanism of transport is called diffusion. Passive transport or facilitated diffusion involves the participation of transporters (B), ionophores (C) or channels (D). All of them allow the movement of high molecular weight and/or charged molecules down the electrochemical gradient. In contrast, active transport involves the movement of molecules against an electrochemical gradient, either by using energy (primary transport, E) or driven by an ion moving down its electrochemical gradient (secondary transport, F). Figure from Nelson and Cox, 2000.

2.2 Transport through biological membranes

As described above, IMPs fulfill a broad number of functions, including molecular transport. Ions, nutrients and other molecules are constantly moving from one side to other of the cellular and organellar membranes. This molecular trafficking needs to be tightly regulated to meet the metabolic cell requirements and maintain cell homeostasis.

Considering the selective permeability of membranes, different mechanisms are required to allow nutrients to cross the lipid bilayer. Therefore, transport through membranes can be classified into different categories (Figure 7):

- <u>Simple diffusion.</u> Molecules cross the membrane down a concentration gradient, without the participation of transporters or channels. In this case, the movement of molecules is spontaneous and it will continue until equilibrium is reached. In general, the movement is faster when the concentration difference between one side of the membrane and the other is higher, but it is also influenced by the molecular weight of the molecule and, in the case of organic compounds, their liposolubility.

- <u>Facilitated diffusion.</u> As in simple diffusion, it is also a passive transport modality, as it does not require energy. Thus molecules move down a concentration gradient. It involves the participation of transporters, channels or ionophores to allow the movement of high molecular weight and/or charged molecules down the electrochemical gradient. Additionally, these IMPs could be further regulated chemically or by the membrane potential.

- <u>Active transport.</u> It involves the movement of molecules against an electrochemical gradient, either by using energy (primary transport) or driven by a second molecule or co-substrate (typically an ion) moving down its electrochemical gradient (secondary transport). Primary transport requires ATP hydrolysis to move the substrate, whereas secondary transport involves the co-transport of substrate and co-substrate either in the same direction (symport) or in opposite directions (antiport).

2.3 Major Facilitator Superfamily (MFS)

The Major Facilitator Superfamily (MFS) of transporters is one of the largest transport families in nature, together with the ATP-binding cassette superfamily of transporters (ABC transporters) (Beis, 2015). MFS are present in all living organisms (Reddy et al., 2012): for example, in prokaryotes they can comprise up to 25% of the entire membrane proteome. In humans, it is estimated that there are more than 100 genes encoding MFS transporters (Quistgaard et al., 2016). MFS are involved in the uptake of nutrients (Augustin, 2010; Cura et al., 2012) and the export of harmful compounds (Gbelska et al., 2006; Fluman et al., 2012). Moreover, they have also evolved additional transport functions, such as movement of metabolites and signaling molecules. Thus, there is a huge pharmacological interest in this family of transporters as many of them have been described to transport toxins and drugs (Smith et al., 2013; Roth et al., 2012; Hagenbuch et al., 2013; Koepsell, 2013; Halestrap et al., 2012).

2.3.1 General features of MFS transporters

Most of the molecular features and functional principles of MFS transporters have been elucidated by the intense study of the *E. coli* lactose permease transporter (lactose/ H^+ symporter or LacY) (Guan and Kaback, 2006). As the field gained more attention, the common features of MFS transporters could be elucidated. They are listed as follows:
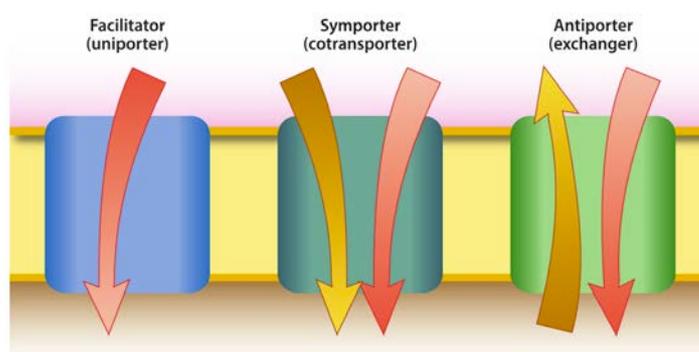
Figure 8: Substrate transport mechanism in MFS transporters. Facilitators (represented as a blue box) allow substrate diffusion across the membrane down its concentration gradient. Symporters (in dark green) and antiporters (in light green) are secondary active transporters that use the energy released from the translocation of a substrate down the concentration gradient to uptake another substrate against its concentration gradient. The transport of both substrates is either in the same direction (symporters) or opposite ones (antiporters). Note that in the figure, the color gradient of the arrows represents the substrate gradient. Figure from Yan, 2015.

▪ Substrate transport mechanism. MFS transporters are divided in three categories according to their transport mechanism (Marger and Saier, 1993): Uniporters, also called facilitators, permit the transport of a single molecule down its concentration gradient. Symporters and antiporters are secondary active transporters and they move two molecules in the same or opposite directions respectively, using the electrochemical potential of a co-substrate, typically an ion (Figure 8).
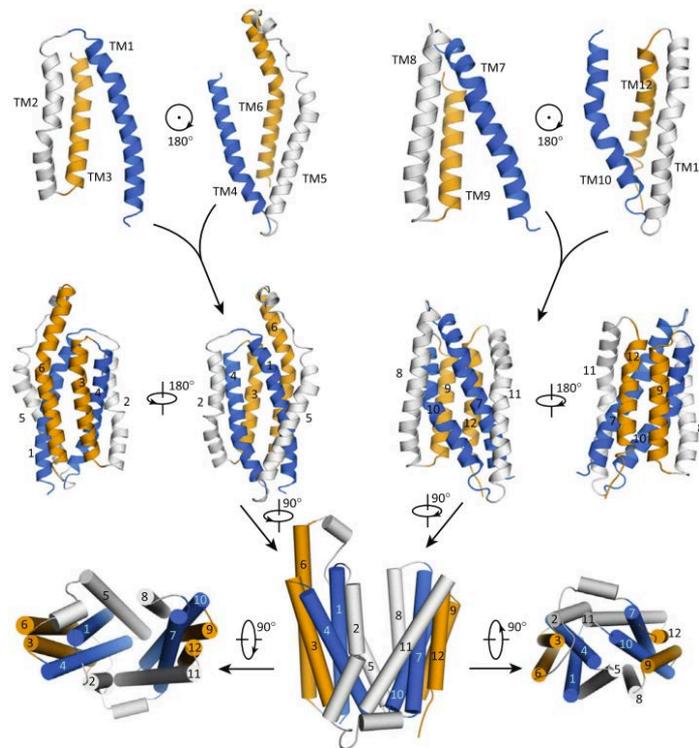
Figure 9: Canonical fold in MFS transporters. The so-called MFS fold compromises 12 TM α-helices organized in two bundles of six helixes each (middle row in the figure). Additionally, each bundle can be subdivided in two inverted repeats, with three consecutive helices each (top row, consecutive helices colored in blue, white and orange). In the bottom row of the figure, the color of the helices reflects the position in the structure: TMs 1, 4, 7 and 10 are in the center of the transporter; TMs 2, 5, 8 and 11 connect the domains; these two groups are involved in ligand coordination and transport. TMs 3, 6, 9 and 12 are located in the periphery and maintain the integrity of the structure. Figure from Yan, 2013.

▪ Conformational changes during transport. Differently from channels, transporters alternately expose the ligand-binding cavity to each side of the membrane during a transport cycle (Jardetzky, 1966). Therefore, transporters adopt different structural conformations to facilitate the movement of the ligand across the membrane: in the inward-open state, the binding cavity is accessible from the cytoplasmic side of the membrane. Once the ligand is bound, the transporter adopts an occluded conformation; consequently, the binding side is not accessible from either side of the membrane. Finally, in the outward-open state, the binding cavity is open to the

periplasmic side of the membrane in bacteria or the extracellular environment in eukaryotes. The transport cycle will be discussed with more detail in section 2.3.3 of the introduction.

- <u>Canonical structural fold.</u> The so-called MFS fold is highly conserved among the members of this group (Law et al., 2008). It comprises typically 12 TMs, with the N- and C-termini commonly oriented towards the cytoplasmic side of the membrane. The TMs are organized in two bundles, the N-terminal and C-terminal bundle. Each bundle is formed by six consecutive TM α-helices, with the binding cavity located in between (Figure 9). Additionally, each bundle could be subdivided in two groups of three α-helices, which have an inverted orientation (Radestock and Forrest, 2011). Furthermore, the TMs have similar positions or functions throughout the MFS family: TMs 1, 4, 7 and 10 are positioned in the center of the transporter and they are involved in ligand coordination and co-transport coupling. TMs 2, 5, 8 and 11 are located in the periphery of the transporter and they connect the different domains. Additionally, they might participate in ligand binding. Finally, TMs 3, 6, 9 and 12 are surrounding TMs 1, 4, 7 and 10 and they maintain the structural integrity of the transporter.

- <u>A-motif.</u> MFS transporters present a highly conserved amino acid sequence in the loops connecting TM2 and TM3 in the N-terminal bundle and/or between TM8 and TM9 in the C-terminal domain. This sequence consists of the following residues: GX_3-(D/E)-(R/K)-X-G-[X]-(R/K)-(R/K), where 'X' indicates any amino acid and '[X]' indicates that the residue might be missing. Typically, these loops are short, restricting the relative movement of the TMs on the cytoplasmic side (Yamaguchi et al., 1992; Jessen-Marshall et al, 1995; Jiang et al., 2013). The A-motif is involved in transport; the residues in this sequence form different gating interactions in the inward-open conformation, occluded and outward-open to the TMs that bend to allow the conformational rearrangements (Quistgaard et al., 2016).

2.3.2 Classification and nomenclature systems for MFS

Considering the large number of MFS transporters, a classification system is required. There are three major nomenclature and grouping systems applicable to MFS transporters (Table 1): (1) The Pfam database (protein families database) is a classification system of sequenced protein domains from all organisms (Finn et al., 2014). Here, related protein families are grouped into clans. To date, according to this classification system, the MFS clan (CL0015) consists of 24 families and 320665 sequenced domains. (2) The HUGO Gene Nomenclature Committee (HGNC) uses the solute

carrier (SLC) system to classify human genes that encode membrane transport proteins, excluding channel proteins, ABC transporters, and ion pumps (Gray et al., 2015). According to this system, 14 SLC families comprising more than 100 genes belong to the MFS. (3) A third classification system is presented by the Transport Classification Database (TCDB), which classifies representative transporters from all organisms on the basis of transport mechanism, phylogenetic relations and substrates (Ren et al., 2007). In the TCDB, 8 of over 600 families belong to the major facilitator.

Members of the MFS clan (CL0015) in the Pfam database		
Acatn	MFS 1-like	Nucleoside tran
ATG22	MFS 2	OATP
BT1	MFS 3	PTR2
Folate carrier	MFS 4	PUCC
FPN1	MFS5	Sugar tr
FTR1	MFS *Mycoplasma*	TLC
LacY symp	Nodulin-like	TRI12
MFS1	Nuc H symport	UNC-93
MFS families according to the SLC nomenclature system		
Family	*Family description*	
SLC2	Facilitative glucose transporters	
SLC15	Proton oligopeptide transporters	
SLC16	Monocarboxylate transporter	
SLC17	Vesicular glutamate transporter	
SLC18	Vesicular amine transporter	
SLC19	Folate/ thiamine transporter	
SLC21/ SLCO	Organic anion transporter	
SLC22	Organic cation/ anion/ zwitterion transporter	
SLC29	Facilitative nucleoside transporter	
SLC33	Acetyl-CoA transporter	
SLC37	Sugar-phosphate/ phosphate exchanger	
SLC43	Sodium-independent, system-L like amino acid transporter	
SLC45	Putative sugar transporter	
SLC46	Folate transporter	
MFS families according to the TCDB		
Family	*Family description*	
2.A.1	Major facilitator superfamily (MFS)	
2.A.2	Glycoside-pentoside-hexuronide:cation symporter (GPH) family	
2.A.12	ATP:ADP antiporter (AAA) family	
2.A.17	Proton-dependent oligopeptide transporter (POT/ PTR) family	
2.A.48	Reduced folate carrier (RFC) family	
2.A.60	Organo anion transporter (OAT) family	
2.A.71	Folate-biopterin transporter (FBT) family	
9.B.111	6TMs Lysyl tRNA synthetase (LysS) family	

Table 1: MFS classification according to the Pfam, SLC and TCDB databases. Table adapted from Yan, 2015.

2.3.3 The alternate access mechanism in MFS transporters

Transporters change their conformation during a transport cycle, exposing the binding cavity to the cytoplasmic side of the membrane and the extracellular side alternatively, as postulated by Jardetzky in the alternate access mechanism model (Jardetzky, 1966). According to this model, transporters only adopted the inward-open and the outward-open conformations and it was confirmed by many MFS reported structures (Figure 10A). The motion explaining the change between the inward-open and the outward-facing conformation was explained by the rocker-switch model (Huang et al., 2003; Law et al., 2008). It postulates a rotation of the N-domain and the C-domain over an axis that crosses the binding cavity, perpendicular to the membrane surface.
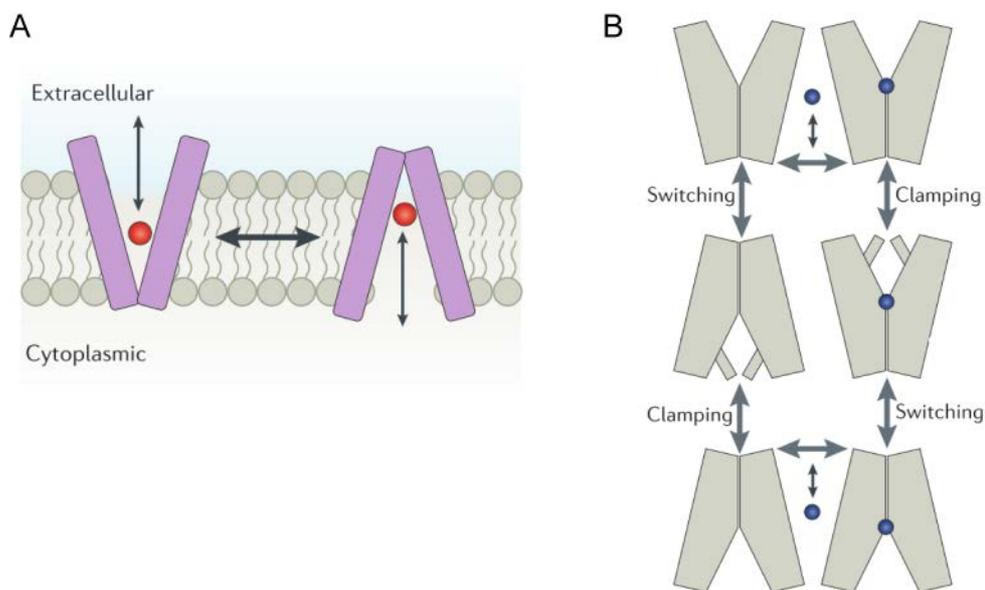
Figure 10: Alternate access in MFS transporters. (A) Alternate access mechanism proposed by Jardetzky (Jardetzky, 1966). According to this model, the ligand binding cavity is alternatively accessible to one or the other side of the membrane, but not both at the same time. (B) Clamp-and-switch model as proposed by Quistgaard *et al.* (Quistgaard et al., 2016). According to this model, the ligand binding induces the bending of some helices in the transporter, occluding the access of the binding cavity to large molecules. Still water molecules or small ions could bind. After the clamping step, the transporter opens to the other side of the membrane by a rocker-switch type motion. Figures from Quistgaard et al., 2016.

As intermediate states have also been reported, more details were introduced to this initially simple model. The "clamp-and-switch" model (Quistgaard et al., 2016) is one of the latest suggested mechanisms. According to this model, first the bending of some helices mediates the occlusion of the binding site from the cytoplasmic side or the extracellular side. Secondly, a rocker-switch-type rotation of the N- and the C-domain results in the exposure of the binding cavity to the other side of the membrane (Figure 10B).

To date, there are only few examples in which the same transporter has been reported in different conformations: the xylose/ H^+ symporter XylE (Quistgaard et al., 2013) and the mammalian fructose transporter GLUT5 (Nomura et al., 2015) are good examples. Apart from crystallographic structures, there are also results coming from molecular dynamics studies that predict the outward-open structure of transporters, such as for LacY (bacterial lactose permease transporter) (Kumar et al., 2014; Stelzl et al., 2014), PepT$_{So}$ (POT from bacterium *Shewanella oneidensis*) (Fowler et al., 2015), GkPOT (POT from *Geobacillus kaustophilus*) (Newstead et al., 2011), GLUT1 (human glucose transporter 1) (Fu et al., 2016) and GlpT (glycerol-3-phosphate transporter) (Moradi et al., 2015).

In Table 2, the reported X-ray structures from MFS transporters are summarized (note that the available structures from family 2.A.17, corresponding to proton dependent oligopeptide transporters, are not included here but in table 3 instead):

Transporter	Organism	TCDB	Substate	Conformation
GlcP$_{Se}$	*Staphylococcus epidermidis*	2.A.1.1	Glucose/ H^+ symporter	Inward open
GLUT1	*Homo sapiens*	2.A.1.1	Glucose uniporter	Inward open
GLUT3	*Homo sapiens*	2.A.1.1	Glucose uniporter	Outward open and outward open partially occluded
GLUT5	*Bos taurus*	2.A.1.1	Fructose uniporter	Inward open
	Rattus norvegicus	2.A.1.1		Outward open
XylE	*Escherichia coli*	2.A.1.1	Xylose/ H^+ symporter	Outward open partially occluded, inward open and inward open partially occluded
EmrD	*Escherichia coli*	2.A.1.2	Multidrug/ H^+ antiporter	Inward open partially occluded
MdfA	*Escherichia coli*	2.A.1.2	Multidrug/ H^+ antiporter	Inward open
YajR	*Escherichia coli*	2.A.1.2	-	Outward open
GlpT	*Escherichia coli*	2.A.1.4	Glycerol-3-phosphate antiporter	Inward open
LacY	*Escherichia coli*	2.A.1.5	Lactose/ H^+ symporter	Inward open and outward open partially occluded
FucP	*Escherichia coli*	2.A.1.7	Fucose/ H^+ symporter	Outward open
NarK	*Escherichia coli*	2.A.1.8	Nitrate/ nitrite antiporter	Inward open and inward open partially occluded
NarU	*Escherichia coli*	2.A.1.8	Nitrate or nitrite symporter or antiporter	Partially inward open and inward open partially occluded

PiPT	*Piriformospora indica*	2.A.1.9	Phosphate/ H$^+$ symporter	Inward open partially occluded
MelB	*Salmonella typhimurium*	2.A.2	Melbiose/ Na$^+$ or Li$^+$ symporter	Outward open and outward open partially occluded
NRT1.1	*Arabidopsis thaliana*	2.A.17	Nitrate/ H$^+$ symporter	Inward open

Table 2: X-ray structures of MFS transporters. Table adapted from Quistgaard et al., 2016.

2.3.4 MFS transporters physiology, disease and pharmacological perspectives

MFS transporters have physiological relevance, as they are involved in nutrient uptake, metabolism, growth, neurotransmission, and signaling or drug resistance, among other physiological roles. Their malfunction has been associated to diseases (Hediger et al., 2013), such as cancer, gout, schizophrenia, epileptic seizure, amyotrophic lateral sclerosis and Alzheimer's disease (El-Gebali et al., 2013; Rask-Andersen et al., 2013; Reimer et al., 2013). For example, the glutamate transporters (SLC1 family) have a critical role in the nervous system as they maintain the extracellular glutamate concentrations below cytotoxic level. From this family, GLT1 transporter is involved in the pathogenesis of amyotrophic lateral sclerosis and Alzheimer's disease (Kanai et al., 2013). Another example is the urate transporter (SLC2 family). Here, genetic defects of the transported are related to nephropathies and metabolic syndromes (Mueckler and Thorens, 2013). The functionally diverse SLC17 family of organic anion transporters has been associated with risk for gout, and possibly schizophrenia, amyotrophic lateral sclerosis, Alzheimer disease and Huntington disease (Reimer, 2013). A different case is the one described for the human SLC26 genes. They are associated with congenital or early onset Mendelian diseases, such as chondrodysplasias for *SLC26A2*, chloride diarrhea for *SLC26A3* and deafness for *SLC26A4* (Alper and Sharma, 2013). These are just few examples but there are many more reported (see review from Hediger et al., 2013 for more examples).

In addition to the roles of transporters in the diseases noted above, several transporters are also of great importance from a pharmaceutical perspective. MFS transporters can serve as drug targets or as drug delivery systems to cells and tissues. Here, neurotransmitter transporters (SLC6 family), intestinal bile acid transporters (SLC10 family) and cation-Cl cotransporters (SLC12 family) have been reported as drug targets. Furthermore, the intestinal oligopeptide transporter PepT1 (SLC15A1) or transporters at the blood-brain barrier are proving to be important drug delivery systems.

2.4 Proton-dependent oligopeptide transporters

Proton coupled oligopeptide transporters (POTs, PTRs or PepTs) are part of the MFS; namely, they are classified into the SLC15 family (according to the SLC system nomenclature) or the 2.A.17 (according to the TCDB). They are present in all kingdoms of life, from bacteria to humans (Daniel et al., 2006).
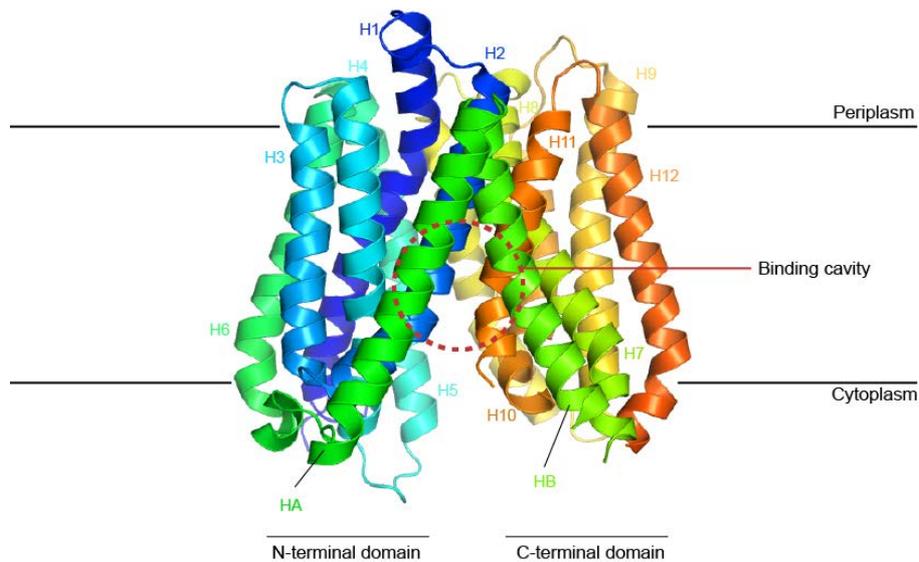
Figure 11: Crystallographic structure of bacterial POT transporter from species *Streptococcus thermophilus* (PepT$_{st}$). Each α-helix is represented in a different color and correspondingly labeled. The TM helices 1-6 comprise the N-terminal domain, whereas TM helices 7-12 are part of the C-terminal domain. The ligand binding cavity is located between the N- and C-terminal domains. Exceptionally, bacterial POTs have two additional helices, namely HA and HB, located between the N- and C-terminal bundles. Structure reported by Martinez Molledo *et al.* (Martinez Molledo et al., 2018; PDB ID 5OXO).

POTs fulfill the previously described features of MFS transporters: (1) they are symporters involved in the uptake of di- and tripeptides by using the proton gradient as driving force (Paulsen and Skurray, 1994). These short peptides are used as nitrogen sources or building blocks for other molecules in the cell. Human POTs, hPepT1 and hPepT2, are mainly expressed in the intestine and the kidney, respectively. In the intestine, they are involved in the uptake of di- and tripeptides, products of dietary protein degradation (Covitz et al., 1998; Fei et al., 1994; Liang et al., 1995; Leibach et al., 1996; Knütter et al., 2004). In the kidney, hPepT2 reabsorbs peptides from the glomerular filtrate (Daniel and Rubio-Aliaga, 2003; Biegel et al., 2006). (2) POTs share with MFS transporters the so-called MFS fold (Figure 11). They comprise 12 TM α-helices, organized in the N- and C-terminal bundles. Exceptionally, prokaryotic POTs have two additional helices (HA and

HB) between the N- and C-domains of unknown function. On the contrary, human POTs present a large extracellular soluble domain (Beale at al, 2015), which was hypothesized to recruit proteases in close proximity to the transporter and cleave longer peptides into di- and tripeptides.

Transporter	Organism	PDB ID	Ligand	Resolution (Å)
PepT$_{St}$	*Streptococcus thermophilus*	4APS		3.3
		4D2C	Ala-Phe	2.47
		4D2D	Ala-Ala-Ala	2.52
		4D2B		2.35
		4XNJ		2.3 (cryo)
		4XNI		2.8 (RT)
		5MMT		3.4
		5D58	Ala-Phe	2.4
		5D59	Ala-Phe	2.4
		5D6K		2.4
		5OXL	Ala-Leu	2.66
		5OXK	Ala-Gln	2.38
		5OXM	Asp-Glu	2.30
		5OXN	Phe-Ala	2.20
		6GHJ	Phe-Ala-Gln	2.26
		6EIA	HEPES	2.00
		5OXQ	HEPES	2.19
		5OXP	Phosphate	2.37
		5OXO		1.95
PepT$_{So2}$	*Shewanella oneidensis*	4LEP	Alafosfalin	3.2
		4TPG	Ala-Tyr(Br)-Ala	3.91
		4TPH	Ala-Tyr(Br)	3.15
		4TPJ	Tri-Ala	3.20
PepT$_{So}$	*Shewanella oneidensis*	4UVM		3
		2XUT		3.62
YbgH	*Escherichia coli*	4Q65		3.4
YePEPT	*Yersinia enterocolitica*	4W6V		3.01
GkPOT	*Geobacillus kaustophilus*	4IKV		1.9
		4IKW	Sulfate	2.0
		4IKX		2.1
		4IKY	Sulfate	2.1
		4IKZ	Alafosfalin	2.4
PepT$_{Xc}$	*Xanthomonas campestris*	6EI3		2.1

Table 3: Reported bacterial POT structures. For each transporter, the organism of origin, PDB ID, the ligand (if any) and the resolution of the model are stated.

(3) To date, all the reported POTs structures are from bacterial sources (Solcan et al., 2012; Lyons et al., 2014; Huang et al., 2015, 2016; Ma et al., 2017; Martinez Molledo et al., 2018; Guettou et al., 2013, 2014; Fowler et al., 2015; Newstead et al., 2011; Zhao et al., 2014; Boggavarapu et al., 2015; Doki et al., 2013; Parker et al., 2017). Furthermore, they have been crystallized in the same structural conformation: the inward-open conformation, in a ligand-bound or ligand-free state, or inward-open partially occluded (Table 3). These models are in agreement with the previously explained clamp-and-switch model (Quistgaard et al., 2016). (4) Finally, the A-motif is located in the loop between TM2 and TM3, facing the cytoplasmic side of the membrane. Additionally, POTs have a conserved signature on TM1, with the sequence E-X-X-E-R/K. Moreover, there are conserved acidic residues in TMs 1, 4, 7 and 10 which were identified as important sites of proton and/or peptide binding (Newstead, 2017).

Apart from their physiological role as peptide uptake systems, it was observed early on that these transporters are able to uptake drugs and antibiotics that have peptidic features. For example, hPepT1 can transport pro-drugs with different clinical applications: oral β-lactam antibiotics for the treatment of bacterial infections, antiviral pro-drugs such as valacyclovir used for the treatment of herpes virus disease or varicella zoster, or the angiotensin-converting enzyme inhibitors used for the treatment of hypertension. Therefore, they can function as drug delivery systems (Ganapathy et al., 1998; Luckner and Brandsch, 2005; Brandsch, 2013).

In this dissertation, there are two aspects of POTs that will be addressed with more detail: the ligand preferences and the conformational changes occurring in a transport cycle.

2.4.1 POTs ligand specificities

The substrate range of POTs is generally very wide, for example human PepT1 and PepT2 are predicted to transport almost any di- and tripeptide (Brandsch et al., 2008). However, the preferences for different peptide side chains vary significantly among different POTs (Boggavarapu et al., 2015; Brandsch et al., 2008; Chiang et al., 2004; Ernst et al., 2009; Fang et al., 2000; Guettou et al., 2014; Ito et al., 2013; Sharma et al., 2016; Solcan et al., 2012) and some strongly prefer dipeptides to tripeptides (Boggavarapu et al., 2015; Ernst et al., 2009; Solcan et al., 2012). As listed in Table 3, several structures have been determined for different bacterial POTs in both apo and substrate-bound forms (Boggavarapu et al., 2015; Doki et al., 2013; Fowler et al., 2015; Guettou et al., 2013, 2014; Lyons et al., 2014; Newstead et al., 2011; Quistgaard et al., 2017; Solcan et al., 2012; Zhao et al., 2014). Substrate-bound structures include GkPOT from *Geobacillus kaustophilus* in complex with

the peptidomimetic alafosfalin (Doki et al., 2013), $PepT_{So2}$ from *Shewanella oneidensis* in complex with alafosfalin, Ala-Ala-Ala, Ala-Tyr(Br) and Ala-Tyr(Br)-Ala (where 'Br' indicates bromination) (Guettou et al., 2013, 2014) and finally $PepT_{St}$ from *Streptococcus thermophilus* in complex with Ala-Phe and Ala-Ala-Ala (Lyons et al., 2014), plus the dipeptides Phe-Ala, Ala-Leu, Asp-Glu and Ala-Gln (Martinez Molledo et al., 2018) and the tripeptides Phe-Ala-Gln, Phe-Ala-Thr and Phe-Ala-Ala. The later di- and tripeptides are part of the results in this dissertation and they will be discussed in detail in the results section.

These structures have revealed that peptides generally bind across the binding cavity between the N- and the C-domain, although an alternative vertical binding mode was proposed for $PepT_{St}$ in complex with Ala-Ala-Ala (Lyons et al., 2014). In the $PepT_{So2}$ structures, peptides Ala-Tyr(Br) and Ala-Tyr(Br)-Ala were found to bind horizontally across the binding cavity, as also observed for all dipeptides in the rest of the described complexes (Guettou et al., 2014). However, as the resolution was rather low (3.2 Å and 3.9 Å, for di- and tripeptide complex respectively), the peptide backbone geometry and binding mode could not be described in detail. The structure of $PepT_{St}$ in complex with Ala-Ala-Ala was determined at a moderately high resolution of 2.5 Å, and suggested an alternative vertical orientation for the peptide. This model has a considerable impact in the field and many simulation studies used this binding mode as a reference (Parker at al., 2014; Fowler et al., 2015; Samsudin et al., 2016; Immadisetty et al., 2017). However, in our recent publication (Martinez Molledo et al., 2018), we characterized binding of different dipeptides to $PepT_{St}$, and we obtained data suggesting that the Ala-Ala-Ala observed vertically bound was actually a misinterpretation of the electron density map. Instead a HEPES buffer molecule could be reliably modeled. Therefore, there was a substantial need for more structural insights into how di- and tripeptides are recognized by POTs.

Additionally, structural studies are complemented with functional assays. In the transporters research field, protein activity is studied in liposomes (Solcan et al., 2012; Parker et al., 2014) or with *in vivo* assays (Weitz et al., 2007; Prabhala et al., 2014). In both cases, radioactively or fluorescently labeled ligands are required to detect transport. It should be taken into consideration that the setup in either experiment is considerably different from the scenario in a crystallization drop: in the functional assays, a membrane with tension, curvature and a specific lipid composition surrounds the transporter. In crystallization, proteins are solubilized in mixed micelles of protein and detergent or reconstituted in a membrane-like assembly, as in the case of lipidic cubic phase (LCP) crystallization (Aherne et al., 2012; Caffrey, 2015). As a result, the environment might influence the

function of the transporter and its ligand preferences. These aspects will be discussed in more detailed in the section 2.5.

For PepT$_{St}$, there are liposome-based functional studies available (Solcan et al., 2012; Lyons et al., 2014). Under these conditions, PepT$_{St}$ was reported to transport di- and tripeptides with different chemical properties, although a certain preference of dipeptides above tripeptides was observed. It should be pointed out that in these studies, the number of tested ligands was rather low and limited to the commercially available peptides.

2.4.2 POTs conformational changes upon a transport cycle

As described before for MFS transporters, POTs also change their conformation during a transport cycle, exposing the binding cavity to the extracellular side of the membrane or the cytoplasmic side alternatively. The clamp-and-switch model (Quistgaard et al., 2016) also applies in this case, as there are inward-open (GkPOT, PepT$_{St}$, PepT$_{So2}$, PepT$_{So}$, YbgH, YePEPT and PepT$_{Xc}$) and inward-open partially occluded (PepT$_{St}$, PepT$_{So}$) conformations described. To date, no crystallographic structure of a POT has been reported in outward-facing conformation. Determining a structure in this conformation would be valuable to have a structural overview of a complete transport cycle. However, molecular dynamics studies predicted an outward-open structure for PepT$_{So}$ (Fowler et al., 2015) and so, we have a rough idea of how this conformation might appear.
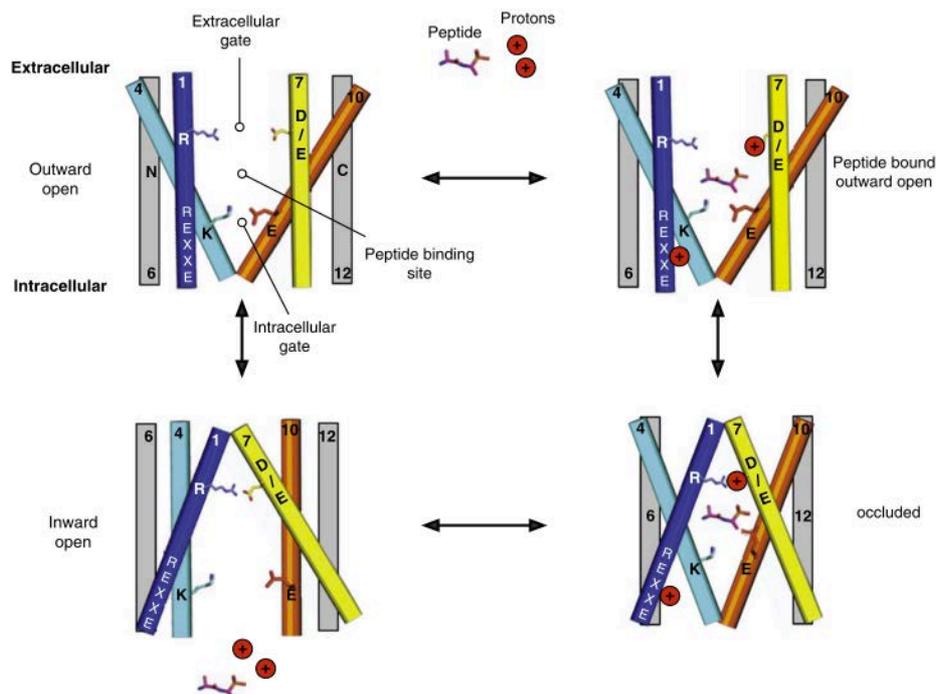
Figure 12: Model of proton coupled transport by the POT family. Proton binding/release is possible due to conserved residues that work as proton binding sites. As a result, ligands are driven towards the inside of the cell. Here,

the main residues involved in ligand coordination and proton binding are mapped within the TM helices of a POT. Each helix is colored differently (except for TMs 6 and 12, in grey) and correspondingly numbered. In TM1, the motif E-X-X-E-R/K is highlighted. Figure from Newstead, 2017.

Another question to be addressed is whether protonation and ligand binding are sequential events, if they have a particular order or even if they depend one another. The latest model proposed by Newstead (Newstead, 2017; Figure 12) suggests that transport is initiated by proton binding to both the E-X-X-E-R/K motif in TM1 and the carboxylate groups located in TM7 in the outward-facing conformation of the transporter. Then, the ligand would bind, leading to a large conformational change. As a result, the transporter would adopt an occluded conformation, closing the extracellular gates. Then, protons would move from the extracellular salt bridge to the intracellular gate, releasing the peptide and the protons in the interior of the cell. Moreover, according to this model, the rate-limiting step in peptide movement would be the reestablishment of the intracellular gate and the opening of the transporter towards the periplasm to accept a new ligand. Even though there are evidences supporting this model, still the details of proton coupled peptide translocation remain to be determined.

2.5 Membrane protein structural and functional characterization

The results of this dissertation focus on the characterization of bacterial POTs in terms of their structure and function as peptide transporters. In this section, the main challenges and progresses in membrane protein structural determination by X-ray crystallography will be discussed, as well as the main methods to study ligand transport (*in vivo*, in membrane-based systems) and ligand binding in solution.

2.5.1 Tools for studying membrane proteins

The study of membrane proteins in terms of structural and/ or functional characterization involves the expression of the protein of interest in a host organism or particular cell line and, except for the *in vivo* experiments, it requires the purification of the protein. Despite the improvements in many of the experimental steps, there are still major bottlenecks in the process (Figure 13): (1) overexpressing a membrane protein can be challenging and might require the design of multiple constructs, test

different expression systems and culture conditions (Wagner et al., 2008; Schlegel et al., 2012; Hays et al., 2010; Clark et al., 2010; Drew et al., 2005; Tate, 2001; Lyons et al., 2016). Furthermore, the expression level can still be low even after the previous optimization. (2) The purification IMPs requires the extraction of the protein from the membrane; as a result, the stability, activity and integrity of the IMP might be compromised. Detergents are typically used to perform this step; choosing a suitable detergent for the IMP increases the chances to maintain its stability and functionality in solution (Seddon et al., 2004; Privé, 2007; Serebryany et al., 2012). (3) When the aim of the experiment is using the sample for structural determination purposes, consideration include the incorporation of (thermo)stabilizing mutations, complex formation with monoclonal antibodies, addition of ligands or inhibitors, and testing different detergents, additives or crystallization methods (Serrano-Vega et al., 2008; Serrano-Vega and Tate, 2009; Tate et al., 2009; Chung et al., 2012). Next, some of the mentioned aspects will be considered with more detail.
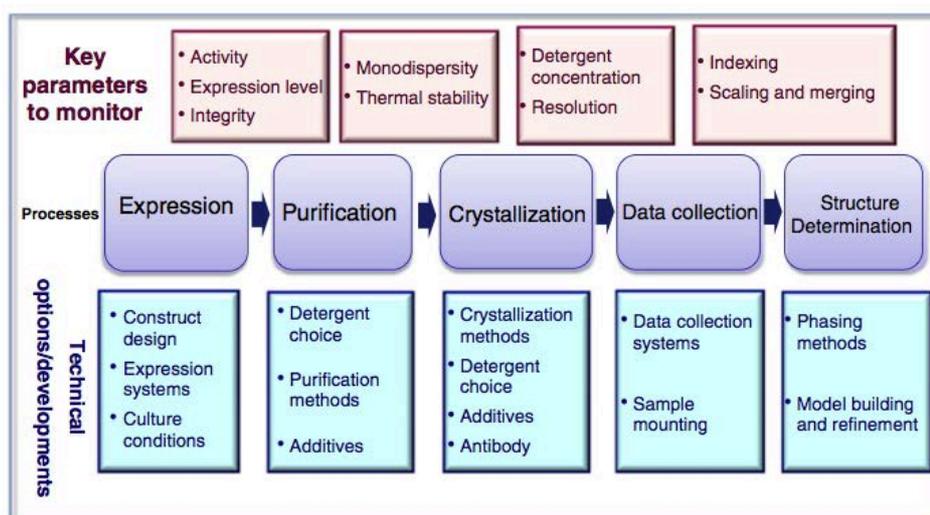
Figure 13: Bottlenecks in membrane protein structure determination. Figure from Moraes et al., 2014.

2.5.1.1 Use of detergents in IMPs solubilization

As mentioned before, detergents are used to solubilize the membrane and extract the IMPs in the form of mixed-micelles of detergent, IMP and remaining membrane lipids. In this step, the protein will be removed from its natural environment, which can compromise its structure, function or activity level (Linke, 2009). Later, the detergents are maintained for the remaining steps in the purification and also in crystallization or other structural characterization methods. Therefore, it is required to invest some time in selecting an appropriate detergent to ensure the sample quality (Arachea et al., 2012; Parker and Newstead, 2012).

Detergents have amphipathic properties, with a hydrophobic tail and a hydrophilic head, and they mimic the membrane environment surrounding the IMPs. The ability of a certain detergent to

solubilize and form micelles is related to their critical micelle concentration (CMC), which is the minimal detergent concentration required to form detergent micelles. Below the CMC, only monomeric molecules of detergent exist in solution whereas above the CMC, there is equilibrium between monomers and micelles. This concept should be carefully considered during the purification of the IMP: the membrane solubilization and IMP extraction requires excess of detergent, whereas in the following purification steps, it should be maintained above the CMC but an excess of detergent could inactivate the protein, masking it in a huge detergent belt and preventing the crystallization of the sample. Additionally, detergents are classified according to the charge of their polar head. Therefore, there are ionic detergents (cationic or anionic, depending on if the net charge is positive or negative, respectively), non-ionic (with uncharged head groups, typically based on sugars moieties) and zwitterionic (with combined properties of the previous groups). Non-ionic detergents are the most commonly used ones as they are mild detergents and do not drastically alter protein-protein interactions (Figure 14A).
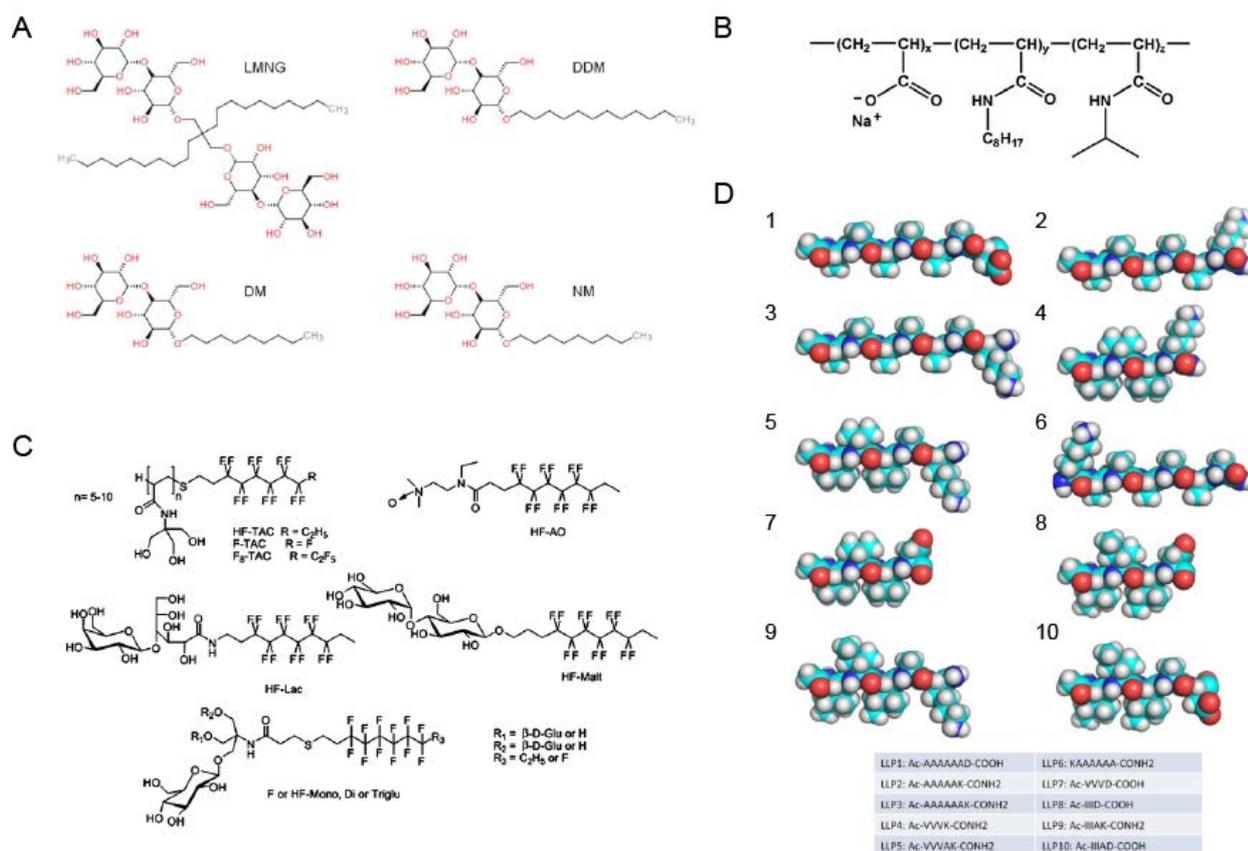
Figure 14: Tools for studying IMPs. (A) Examples of maltopyranoside detergents. These are non-ionic detergents used in membrane solubilization and IMPs purification. Here the following detergents are represented: LMNG (lauryl maltose neopentyl glycol), DDM (n-dodecyl-β-D-maltopyranoside), DM (n-decyl-β-D-maltopyranoside), and NM (n-nonyl-β- D-maltopyranoside). These detergents were used in our studies. Figure from Quistgaard et al., 2017. (B) Example of amphipoles (A8-35). (C) Chemical structures of fluorinated and hemifluorinated surfactants. Here several fluorinated surfactants are represented, showing the diversity in the polar head groups and the hydrophobic acyl chain

carrying multiple fluorine moieties. Figures of panels B and C from Breyton et al., 2010. (D) Peptergents used in the studies of Veith *et al.* (Veith et al., 2017). The chemical composition of the molecules is indicated in the table. Molecules are color coded: turquoise, carbon atoms; red, oxygen; blue, nitrogen and gray, hydrogen. Figure from Veith et al., 2017.

Besides detergents, the incorporation of lipids in the purification process can also improve IMP stability. In these regards, cholesterol is the most commonly used lipid and it has been successfully used in the purification and crystallization of GPCRs, such as the human β2-adrenergic receptor (β2-AR) (Cherezov et al., 2007). There are also examples in the literature in which other lipids are added in the purification and crystallization, such as the mammalian voltage-dependent K$^+$ channel Kv1.2 (Long et al., 2005).

2.5.1.2 Alternatives to detergents in the work with IMPs

Although extensively used, there are also alternatives to detergents that keep the IMP in a close to natural lipid environment. Other options include the use of amphipoles and fluorinated surfactants (Breyton et al., 2010), lipic-like peptides (LLPs) (Veith et al., 2017), or the reconstitution of the IMPs in membrane-like particles, such as bicelles, nanodiscs (Denisov et al., 2016) or the recently described saposin-derived lipid nanoparticles (SapNPs) (Frauenfeld et al., 2016; Lyons et al., 2017; Flayhan et al., 2018). The latter will be described in the next section (2.5.1.3).

Amphipoles are small (5-20 kDa) flexible amphipathic polymers (Figure 14B), with high water solubility and a dense hydrophobic region to tightly bind to the TM domains of the IMPs. In general, amphipoles themselves do not solubilize and extract the IMPs from the membrane. This step is still performed with detergents and later, the mixed micelles are incubated with the amphipoles. Fluorinated surfactants have the same structure as detergents but instead of having hydrogenated aliphatic chains, the acyl chain carries multiple fluorine substituents (Figure 14C). As a drawback, fluorinated chains are more rigid and they present a reduced affinity towards the hydrophobic TMs of the IMPs compared to detergents. On the contrary, this feature is advantageous in the sense that these surfactants are less aggressive than detergents and will have little effect on the protein-protein interactions. Finally, LLPs or peptergents (Figure 14D) consist of a short hydrophobic tail of consecutive nonpolar amino acids and a polar head group, which can be neutral, positively- (Lys, Arg, His) or negatively-charged (Glu, Asp). In all the previously mentioned methods, the membrane solubilization step is performed in the presence of detergents and later, either the amphipoles, LLPs or fluorinated surfactants are incorporated in the IMP purification.

Finally, the excess of detergent is removed by dialysis, using polystyrene beads or by cycles of dilution/ concentration of the sample.
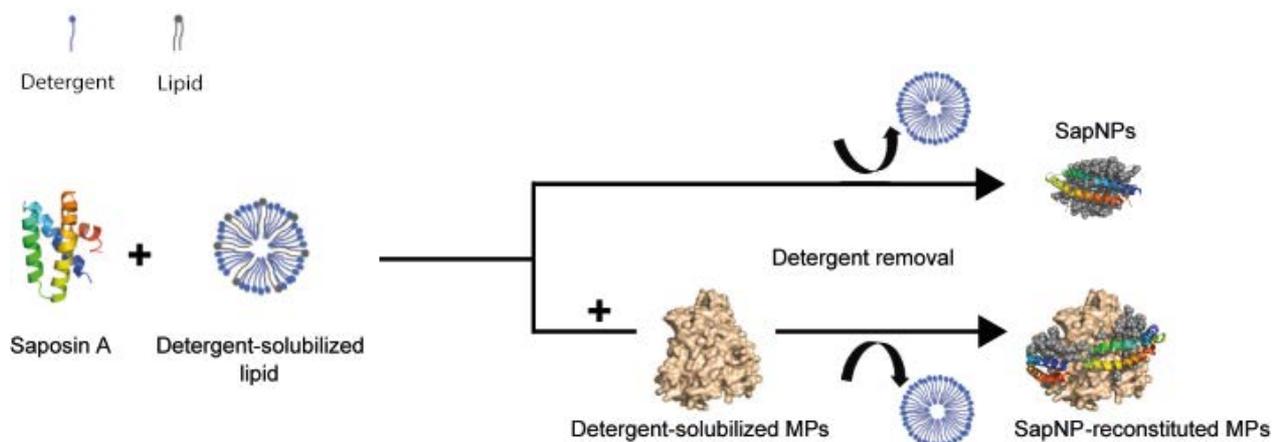
Figure 15: Reconstitution of IMPs in saposin nanoparticles. Workflow for the reconstitution of an IMP using the saposin proteins. Here, as an example saposin A is mixed with the detergent-solubilized lipids. If no IMPs are present, empty discs of lipids surrounded by saposins are formed (upper part of the diagram). When an IMP is added, it can be incorporated in to saposin disc. This method requires the removal of detergent excess. Figure from Flayhan et al., 2018.

2.5.1.3 Reconstitution of IMPs in membrane-like particles

Among larger IMPs-lipid complexes, bicelles, nanodiscs and SapNPs are found. Bicelles are flat discoidal structures composed by lipid and detergent mixed in particular ratios (Faham and Bowie, 2014; Poulos et al., 2015). The most commonly used bicelle is formed by DMPC/ CHAPSO, lipid and detergent, respectively. They have successfully been used for the reconstitution of IMPs and there are examples in the literature of structures solved by X-ray crystallography of IMPs in bicelles, such as GPCRs (β2-adrenergic receptor by Rasmussen et al., 2007; bacteriorhodopsin by Faham et al., 2005), β-barrels (VDAC structure by Ujwal et al., 2008) or transporters (as the LeuT by Wang et al., 2012).

Nanodiscs are discoidal lipid bilayers belted by two amphipathic helical proteins, called membrane scaffold proteins (MSPs) (Denisov et al., 2016). The length of the MSPs and the lipid ratio required to form the disc itself determines the size of the nanodisc. Despite of the many improvements to generate more homogeneous disc preparations or engineer MSPs of different sizes to accommodate larger IMPs, nanodisc formation requires optimization and there are limitations in the IMP size it can accommodate. As a follow-up of this approach, the SapNPs technology overcomes some of the nanodiscs limitations: saposin proteins are able to adopt a different conformation in the presence of liposomes or detergent-solubilized lipids at a pH range of 4.5-7.5. Under these conditions, the saposin opens up and recruits the lipids from the environment. If there is an IMP, it can be reconstituted within the particle (Figure 15). There seems to be certain specificity between the

32

saposin protein and the lipids it can embrace. An additional step to remove the excess of detergent is required after the reconstitution (Flayhan et al., 2018).

2.5.2 IMPs structure determination by X-ray crystallography, challenges and improvements

Protein structural studies are performed by different approaches: nuclear magnetic resonance (NMR), electron microscopy (EM), small X-ray scattering (SAXS), small angle neutron scattering (SANS) and X-ray crystallography. Despite the experimental and technical improvements in all the previously mentioned approaches, X-ray crystallography is still the method of choice to obtain molecular details at high-resolution. Before discussing in more detail IMPs crystallization, some basic concepts about X-ray crystallography will be address in the following section. Latter, the different approaches for crystallizing IMPs will be discussed, emphasizing the difficulties in IMPs crystallization and the recent improvements.

2.5.2.1 Basic principles of X-ray crystallography

Since the first protein structure was determined in 1958 (myoglobin at 6 Å resolution; Kendrew et al., 1958), macromolecular crystallography (MX) has rapidly evolved. Many improvements have been established in protein production and crystallization, reducing the sample volume requirements and automatizing the process. Moreover, new powerful X-ray sources have been built and equipped with fast detectors, reducing the time of data collection to seconds. In terms of data processing, new algorithms and computer software for diffraction data collection, structure solution, refinement and model representation are user friendly and accessible even to non-expert crystallographers. As a result, the available crystallographic data has increased to more than 120,000 PDB entries to date.

In this section, a brief description of the workflow in X-ray crystallography is provided. Even though with not much detail, the aim here is to get an overview on how from a protein crystal, we can obtain molecular information and visualize models representing the structure of a protein.

Starting from the crystallization of macromolecules, the first critical requirement is to have a good protein sample. Here the purity of the sample has to be evaluated to ensure that the preparation presents a high purity. As an alternative, if the sample is not as pure, further proceedings should be implemented in the sample purification to remove contaminants. Furthermore, the protein preparation should be homogeneous, avoiding aggregates or different conformations or oligomerization states of the protein of interest. The sample quality can be determined by different

methods, including dynamic light scattering (DLS), mass spectrometry (MS), thermal stability, size exclusion chromatography (SEC), etc. (Boivin et al., 2016). Thus, crystallography is a demanding technique in terms of sample requirements; despite the miniaturization and automatization of the crystallization setup, still ~1 mg of a protein is required to screen a wide range of conditions. This amount might be difficult to produce for challenging systems, including membrane proteins.

When a crystallization experiment starts, the protein of interest is initially in solution. By changing the conditions (salt concentration, pH, buffer, precipitants, temperature, etc.), the protein slowly comes out of the solution and forms crystals rather than a precipitate. The early-formed crystals can continue growing until equilibrium between the crystals, the remaining protein in solution and the crystallant components is reached (McPherson and Gavira, 2014). Once the first crystallization experiments are set up, the drops are carefully inspected to identify the first crystal hits. There is certainly no correlation between "good-looking" crystals and good diffraction qualities. Typically the first crystal hits have to be improved by a systematic screening of different conditions that might affect crystal growth. Furthermore, crystal harvesting and freezing is not a trivial activity; crystals need to be frozen very fast to avoid ice formation and cryo-protection might be required in some cases.

At the X-ray source, the MX experiment lies on the study of the X-rays scattering by the electrons in the crystal molecules. Protein crystals are highly ordered, where the molecules constitute a periodically repeated pattern in the three-dimensions or lattice. The minimal group of particles in the crystal in which a lattice can be divided is called unit cell. Furthermore, the smallest entity from which the entire unit cell can be generated by symmetry operations is called asymmetric unit (Wlodawer et al., 2013). Due to the periodic repetition of the unit cell in the crystal volume, the scattering of the X-rays is enhanced in selected directions. In a crystal, the X-ray scattering effect is known as diffraction and it is represented by well-defined and sharp reflections in the diffraction pattern collected by a detector. Moreover, the position of each atom in the crystal affects the intensity of the individual reflections. Furthermore, once the data collection is completed, a first quality assessment of the data can be performed (Wlodawer et al., 2008), evaluating certain statistical parameters:

In a diffraction experiment, a set of images with the reflections is collected. Most reflections are measured many times and they are represented by a set of pixels in an image, which are integrated, averaged and corrected to generate a list of unique reflection intensities, each accompanied by a standard error. The accuracy of the averaged intensities can be evaluated by the R factors (Diederichs and Karplus, 1997). Other desirable characteristics of a data set are (1) high redundancy

or multiplicity, (2) good completeness and (3) high signal to noise ratio ($I/\sigma(I)$). A highly redundant data set will improve the calculation of the intensities and the associated errors, although one should be careful with the associated radiation damage. Secondly, if some reflections are missing meaning that the data completeness is low, the electron density map will be difficult to interpret. Finally, $I/\sigma(I)$ evaluates the intensity of the reflection above the background. This parameter is typically used as a reference to define the resolution of the dataset. Although classically an $I/\sigma(I)$ of 2.0 was accepted to limit the resolution of a dataset, other criteria can also be considered, such as the correlation coefficients (Karplus and Diederichs, 2012, 2015).

In each reflection there is information concerning the position of each atom in the molecule but this information needs to be extracted, possible by using the Fourier transformation (FT):
Each reflection is defined by a structure factor (F(hkl)), which comprises the amplitude (obtained as the square root of the measured intensity) and the phase. The FT requires both parameters, amplitudes and phases to calculate the electron density map. In the diffraction experiment, the projection of a three-dimensional object (the crystal with the protein molecules) into a dimensional plane (the detector) causes the loss of the phases; this is often referred as the crystallography phase problem. Once the phases are recovered, they are combined with the structure factors and converted to a different representation of the molecule by the FT, moving from reciprocal to real space.

Considering that the phases cannot be measured, they need to be estimated indirectly. There are three major procedures to calculate the phases: (1) direct methods, (2) molecular replacement (MR) and (3) experimental phasing methods. Whereas direct methods are applicable to small-molecule crystallography and they require high-resolution data (close to atomic resolution), MR and experimental phasing are typically used in MX crystallography. In MR, a Patterson map is computed from the structure factors collected in the diffraction experiment. Furthermore, no phase information is required to calculate the Patterson map. Then, the same approach is followed for an already available structure of a closely related protein. Both maps will be correlated when the molecules are correctly oriented and located in the asymmetric unit. On the contrary, the experimental phasing methods involve the calculation of phases from the differences measured between native and heavy-atom derivatized crystals, where the presence of scatterers provides a strong and distinct signal (Hendrickson, 1991). This additional signal, absent in the native data is used to locate the heavy-atoms. Then, the phases of such atoms can be calculated and they are used

to estimate the phases for all the other atoms. The phasing approach to be used will determine in great extend the diffraction data collection.
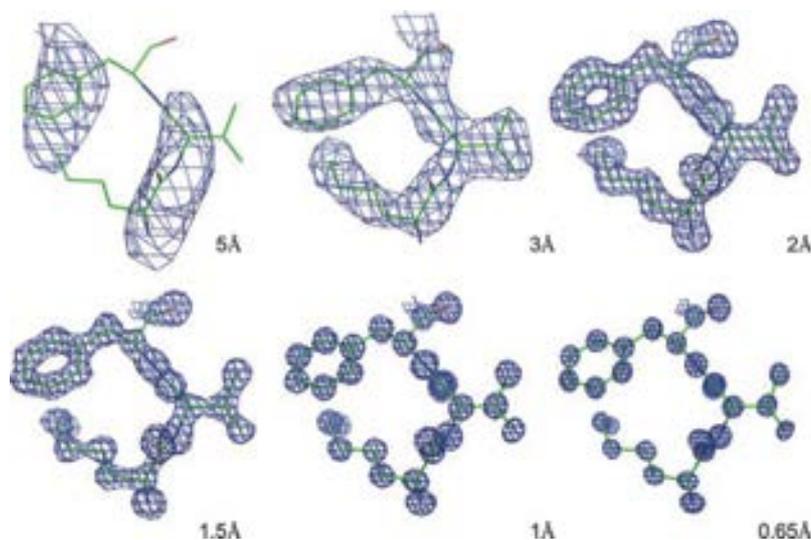
Figure 16: Electron density maps at different resolution limits. Electron density maps calculated at different resolution cut-offs for the N-terminal fragment (Lys1–Val2–Phe3) of lysozyme (PDB ID 2VB1). Figure from Wlodawer et al., 2008.

As a result of the FT of the structure factors, an electron density map (ρ) is obtained. Here the distribution of the electrons within the unit cell of the crystal is represented as a map. The basic electron density map can be calculated by FT of the experimentally determined amplitudes (Fobs) and their phases. However, because the phases are not available experimentally and they are calculated from the used model, a difference map is calculated using the differences between the observed and calculated amplitudes and calculated phases (2Fobs - Fcalc). Additionally, the parts existing in the structure but not included in the model are represented in the positive map (green density), whereas the parts wrongly introduced by the model and absent in the true structure are visible in the negative map (red density). Despite the error associated to the calculated phases, the accuracy and clarity of an electron density map depends on a greater extend to the resolution limit of the data, i.e. to the number of reflections used in the calculations. Therefore, maps obtained from higher resolution data are easier to interpret (Figure 16).

Once the initial model has been obtained, it is further refined using automatic refinement approaches or even manually correcting the model. There are many different parameters that can be optimized in the model refinement, including the B-factors, TLS parameters and stereochemical restrains. Finally, the quality of a model should also be evaluated before deposition (Kleywegt,

2000). Here different statistical parameters are good indicators: (1) the R-factors, (2) root-mean-square deviations (rmsd) and (3) Ramachandran outliers. The R-factors, namely R-work and R-free, evaluate the discrepancy between the crystallographic model and the experimental data. The rmsd value indicates the divergence of the model from geometrical parameters that are considered as normal. Finally, the Ramachandran plot evaluates the combination of the ϕ and φ angles of the folded peptide backbone.
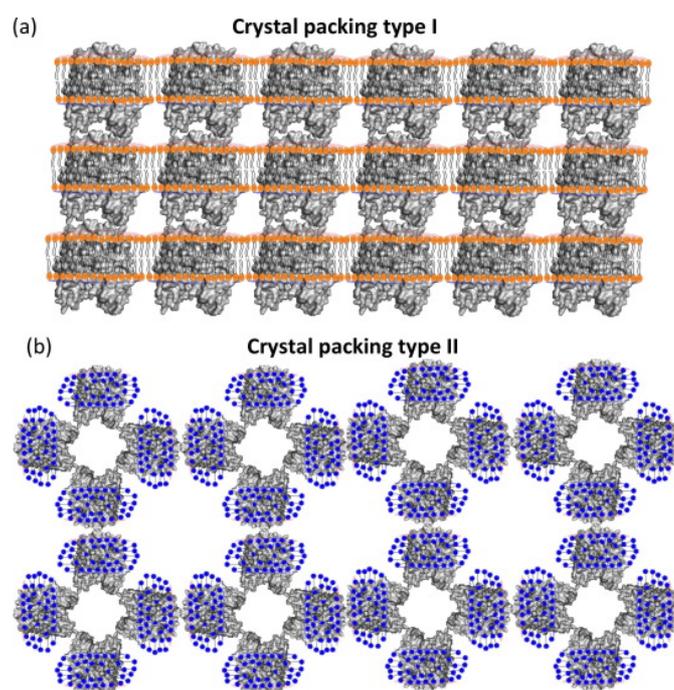
Figure 17: Types of crystal packing in IMPs crystallization. (A) In crystal packing type I, IMPs and lipid molecules are assembled as in a biological membrane. Protein-protein contacts can be established not only within the soluble domains of the IMPs but also the TM regions. (B) In crystal packing type II, crystals are formed by polar protein-protein interactions while the hydrophobic surface of the IMPs is covered by the detergent micelle (represented here by blue dots). Figure from Birch et al., 2018.

2.5.2.2 IMPs crystallization

Apart from the previously discussed difficulties related to membrane protein expression and purification, crystallization of IMPs can be a long-term and tedious project with several hundreds of crystals to be screened and many rounds of optimization steps. There are different methods to crystallize IMPs (Birch et al., 2018), which can be divided in two categories: *in surfo* methods and the *in meso* methods.

The *in surfo* methods make use of surfactants to generate mixed-micelles of IMPs, lipids carried from the membrane of origin and, in most of the cases, the detergents used in the solubilization of the membrane and the purification process. These mixed micelles of IMPs and detergents are directly

used in vapor diffusion crystallization, batch methods, microdialysis or counter diffusion. When crystallizing, the obtained crystals typically show weak X-ray diffraction and the reflections are often observed at low-resolution with anisotropic diffraction patterns. This is due to the packing of protein molecules within the crystal, called type II crystal packing (Figure 17). Here, as the detergent or other surfactant belt surrounds the IMPs, the exposed regions of the protein are rather limited and therefore, the intermolecular contacts are reduced. Typically these crystals have a high solvent content and they are very fragile.
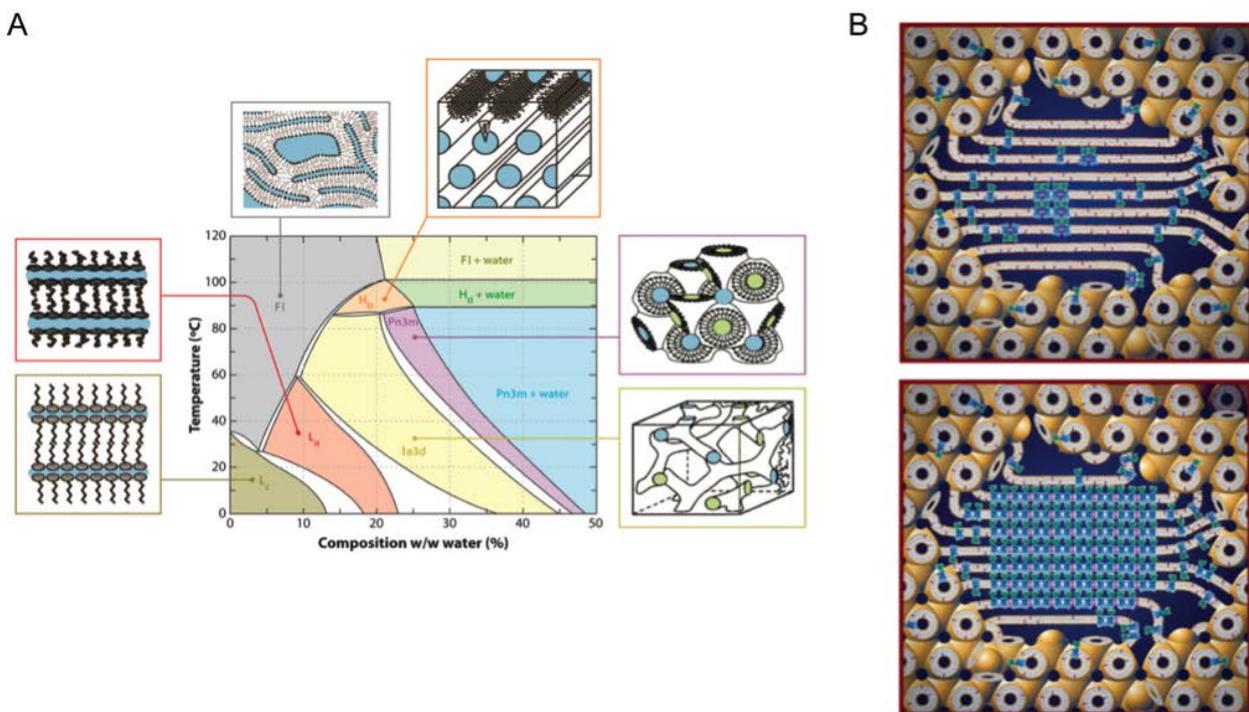
Figure 18: The lipidic cubic phase method. (A) Phase diagram of monoolein. The cubic phase (Pn3m) forms in the temperature range observed in the graph and mixing the lipid and the detergent-solubilized protein at a particular ratio. (B) Model representing the cubic phase bicontinuous system, in which the lipid bilayer and the water channel network are constant. In this case, a GPCR is reconstituted in the cubic phase. It can diffuse freely and form areas with high protein concentration.

On the other hand, the *in meso* methods require the reconstitution of the IMPs in a lipid environment. The lipidic cubic phase (LCP) method is, to date, a well-established crystallization method for IMPs (Caffrey and Cherezov, 2009). It involves the mixture of the detergent-solubilized protein sample with the lipid-like molecules in a particular ratio, prior the setup of the crystallization trials. Monoacylglycerol 1-(9Z-octadecenoyl)-rac-glycerol (also known as monoolein or 9.9 MAG) is the most commonly used molecule for this purpose; it is mixed in a 3:2 lipid to protein ratio (w/w) (Figure 18). Here, the IMP is reconstituted in a so-called cubic phase or mesophase, which is a bicontinuous system, in which the lipid bilayer and the water channel network are constant in three

dimensions (Figure 18B). Under these conditions, the IMPs can diffuse freely through the membrane-like environment and concentrate, form nucleation points and crystallize. Under these conditions, the protein-protein interactions are not only established through the loops or soluble protein domains but also through the TMs segments.

Furthermore, this method has been extensively developed and automatized, reducing the sample volume considerably (down to 50 nl mesophase or *bollus* drops) and the required time to set up a crystallization plate. On the contrary, the LCP method still presents a major bottleneck in the crystal harvesting step: the mesophase is viscous, and thus difficult to manipulate and extract individual crystals from it. Opening the wells in the crystallization plate is not trivial, and many times the crystals are lost in this step. Furthermore, when frozen, the mesophase can turn opaque, difficulting the location of the crystal in the harvesting loop and the subsequent screening at the X-ray source. Despite the drawbacks, the LCP method is imposing itself as a regular technique in IMPs crystallization, mostly due to the fact that the diffraction of these crystal goes to higher resolution if compared to the *in surfo* methods; these crystals present a type I crystal packing (Figure 17). Additionally, new crystal screening methods are being developed to avoid crystal harvesting, such as the studies reported by Huang and colleagues (Huang et al., 2015, 2016). Here the authors developed a special type of crystallization setup compatible with data collection at room temperature and in cryo-conditions.

As mentioned earlier, there are several reported structures of bacterial POTs (see Table 3). Most of them were determined at moderate resolutions (3.9-3.0 Å) from crystals grown in vapor diffusion. Additionally, higher resolution structures were obtained for $PepT_{St}$, GkPOT and $PepT_{Xc}$ from crystals grown by the LCP method, reaching even the impressive resolution of 1.9Å in the case of GkPOT (PDB ID: 4IKV).

2.5.3 Study of transport and ligand binding

There are multiple experimental approaches to study the interaction of a protein and its ligand, such as isothermal titration calorimetry (ITC), thermal shift assay (TSA), microscale thermophoresis (MST), surface plasmon resonance (SPR), mass spectrometry (MS), as well as many of the previously mentioned structural determination approaches (NMR or SAXS, for instance). In the upcoming section, different experimental approaches to study transport will be discussed, taking as a reference experiments performed for POTs.
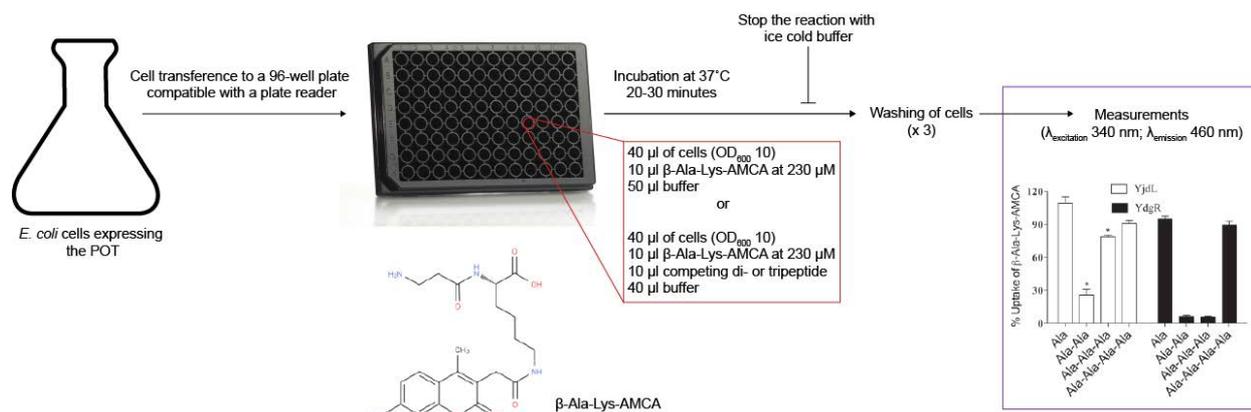
Figure 19: Workflow for the *in vivo* transport assays using the β-Ala-Lys-AMCA reporter molecule. *E. coli* cells are grown and the expression of the POT to study is induced. Then, the culture is centrifugated and the cells are resuspended to a final OD_{600} of 10. The cells are transferred to a 96-well plate compatible to the plate reader to use. They are incubated with the β-Ala-Lys-AMCA or the β-Ala-Lys-AMCA plus the competing di- or tripeptide. After an incubation time at 37°C, the cells are washed to remove the excess of AMCA in the media and the measurements are performed in the plate reader. As an example, the plot in the purple box shows the results of an *in vivo* transport assay for YjdL (DtpC) and YdgR (DtpA). Here single alanine peptides of increasing length were used as competitors against β-Ala-Lys-AMCA. Whereas Ala-Ala and Ala-Ala-Ala inhibit the uptake of β-Ala-Lys-AMCA, it is still transported in the presence of Ala and Ala-Ala-Ala-Ala, showing the preference of POTs for di- and tripeptides. Plot from Prabhala et al., 2014.

2.5.3.1 *In vivo* transport studies

In the context of POTs, there is a well-established *in vivo* uptake assay in which a fluorescently labeled dipeptide-mimetic molecule is used as reporter: β-Ala-Lys-AMCA (β-alanyl-lysyl-N-7-amino-4-methylcoumarin-3-acetic acid) (Malle et al., 2011; Jensen et al., 2012). This experimental approach was successfully used for the characterization of the *E.coli* POTs (DtpA, DtpC and DtpD) in terms of ligand preferences. Here, the transporter is overexpressed in bacterial cells and incubated with β-Ala-Lys-AMCA. After incubation, the cells are washed to remove the excess of the fluorophore in the media and finally, the fluorescence is measured using a plate reader (the fluorophore is excited at a wavelength ($\lambda_{excitation}$) of 340 nm and it emits ($\lambda_{emission}$) at 460 nm). If fluorescence is detected after the washing step, it corresponds to β-Ala-Lys-AMCA transported by the POTs. This assay can be performed in competition with natural di- and tripeptides: if the uptake of β-Ala-Lys-AMCA is reduced in the presence of a peptide, it means that the peptide is competing with the fluorophore and the transporter has a higher affinity towards this ligand compared to the β-Ala-Lys-AMCA (Figure 19). The IC_{50} (half maximal inhibitory concentration) and K_i (inhibitory constant) values can be calculated for different peptides, establishing the ligand preferences for a particular POT.

It should be mentioned that with this experimental setup, not real transport is measured by rather competition: as only the fluorescence of the reporter molecule can be detected, when a second ligand is added, the decrease in the recorded fluorescence might result from the uptake of the non-labeled ligand or just that the later inhibited the uptake of the fluorophore, blocking the transporter. An example that illustrates the second situation was reported in the studies of the xylose transporter (Sun et al., 2012). Using a liposome-based assay, it was shown that the xylose transporter was strongly inhibited by glucose. Additionally, to complement the previous setup, the study reported by Prabhala and colleagues (Prabhala et al., 2017) on the *E.coli* POT DtpA (also known as YdgR) combined an *in vivo* cell-based assay with liquid chromatography-tandem mass spectrometry (LC-MS): bacterial cells overexpressing YdgR were incubated in presence of the probe only or also with different ligands. In addition to the fluorescence measurements, cells were lysed and their content was studied by LC-MS, to detect the presence in the cytoplasm of the transported ligands.

2.5.3.2 Liposome-based assays

A second approach to study transport is by reconstituting the transporter of interest in liposomes. Then, the ligand uptake can be determined by measuring the amount of compound that is contained inside the liposome. In this case, a reported ligand is also required; typical probes carry fluorescence dyes or radioactively labeled atoms (Solcan et al., 2012; Lyons et al., 2014). Additionally, as in this case the transporter is not in a cellular context, peptide uptake is driven by a proton gradient established between the liposomes membranes: the liposomes are preloaded with a buffer at higher pH (low proton concentration) and diluted in a buffer with lower pH value (higher proton concentration).

Ligand preferences in $PepT_{So2}$ (Guettou et al., 2014), GkPOT (Doki et al., 2013), DtpA and DtpB (Weitz et al., 2007; Harder et al., 2008), DtpC (Ernst et al., 2009), DtpD (Casagrande et al., 2009), and $PepT_{St}$ have been studied by liposome-based assays. For $PepT_{So2}$, the radioactively-labeled dipeptide L-[^{3}H]Ala-Ala was used as reporter ligand. The kinetics of $PepT_{So2}$-mediated L-[^{3}H]Ala-Ala uptake reported that, at pH 6.0, the rate of substrate uptake fitted a standard Michaelis-Menten equation, with a K_m value of ~20 µM. Furthermore, L-[^{3}H]Ala-Ala uptake could be inhibited in the presence of tripeptides; a library of different tripeptides was tested in a systematic way, proving that some of them could inhibit L-[^{3}H]Ala-Ala uptake in almost a complete manner.

Using a similar setup and the same dipeptide L-[^{3}H]Ala-Ala, $PepT_{St}$ was shown to transport specifically di- and tripeptides (Solcan et al., 2012). In this study, polyAla peptides of increasing length (from two to five residues) were tested in competition with L-[^{3}H]Ala-Ala. Only di- and

tripeptides inhibited the uptake of the labeled dipeptide, and overall dipeptides appeared to compete better than tripeptides. Thus, a small set of dipeptides with different chemical features was tested; here, nonpolar dipeptides, aromatic and polar but non-charged residues were preferred over charged residues. Furthermore, Parker and colleagues provided some hints regarding the stoichiometry the transport (Parker et al., 2014a): it was proposed that PepT$_{St}$ requires three protons to transport a tripeptide, whereas dipeptide transport requires four to five protons.

2.5.3.3 Study of binding in detergent-solubilized membrane protein samples

A different approach to quantify the affinities of a transporter to different ligands is by studies in solution. In this case, transport itself is not measured but binding instead.

For POTs, there have been several publications in which this experimental approach was used:

Arabidopsis thaliana nitrate transporter NTR1.1 (Parker et al., 2014b) is a member of the proton-couple peptide transporter family (PTRs or POTs); this transporter has evolved as a nitrate uptake system in plants. In this study, nitrate uptake was measured in detergent-solubilised GFP-fusion NTR1.1 transporter by microscale thermophoresis (MST) (see methods section for more details on this technique). As a result, the dissociation constant (K$_D$) of nitrate was estimated to the low millimolar range (~ 1 mM).

Clémençon and coworkers (Clémençon et al., 2018) recently published a workflow to detect binding in solution for the yeast peptide transporter PepT$_{Pp}$ (from yeast *Physcomitrella patens*). In this case, PepT$_{Pp}$ was not purified but instead the MST measurements were performed in purified and solubilized membranes. Western blot analysis was used to confirm the expression of the transporter.

Finally, Flayhan *et al.* reported an interesting study (Flayhan et al., 2018) where ligand binding of PepT$_{St}$ and PepT$_{So2}$ was studied in SapNPs-reconstituted transporters. In both cases, a K$_D$ value could be estimated and it was similar or below the estimated K$_D$ to the detergent-solubilized PepT$_{St}$ and PepT$_{So2}$.

Chapter 3. Aim of the work

Proton dependent oligopeptide transporters (POTs) are important for the uptake of dietary di- and tripeptides in many organisms and are also involved in drug absorption in humans. These proteins transport a wide range of substrates and thus, they have typically been described as highly promiscuous transporters. However, the structural basis on how different peptides are accommodated in the same binding pocket has so far remained obscure. Previous to our studies, only few structures of bacterial POTs in complex with natural ligands were reported. Furthermore, the physicochemical diversity of the ligands was limited to small non-polar and aromatic residues only. These structural data were complemented with liposome-based assays, but even there the number of screened ligands was rather narrow. Moreover, the binding mode for di- and tripeptides has been heavily debated in literature and needs further clarification.

To obtain molecular insights into the mechanism of substrate recognition in peptide transporters, the POT from *Streptococcus thermophilus* (PepT$_{St}$) was used as a model system to address the following questions:

    1. Does PepT$_{St}$ have particular ligand preferences among all the possible di- and tripeptides?

    2. How are substrates coordinated in the promiscuous binding site? Are there any general mechanisms applicable to other peptide transporters?

    3. How can PepT$_{St}$ be stabilized in different conformations to obtain insights into the transport cycle?

Chapter 4. Results and discussion

In this section, we will discuss the aspects related to the characterization of PepT$_{St}$ transporter in terms of ligand preferences and ligand binding. Moreover, we will describe the structural insights on ligand binding and the additional structures we reported in the apo- form or in complex with non-peptidic molecules. First, the ligand binding studies in solution are examined. Here the experimental approaches used were differential scanning fluorimetry (DSF) and microscale thermophoresis (MST). Both of these techniques have been recently introduced in our lab and they are still not common approaches in the study of transporters. Secondly, the structures in complex with different ligands will be described, concluding with the mechanisms PepT$_{St}$ has to accommodate different ligands in the same binding cavity. The structures presented in this dissertation have been determined by X-ray crystallography. Despite the challenges of membrane protein crystallization, we present here high-resolution structures (2.0-2.7 Å) possible to determine as the crystals were grown by the LCP method.

4.1 Ligand binding studies in PepT$_{St}$

4.1.1 PepT$_{St}$ ligand binding studies in solution

Ligand preferences for PepT$_{St}$ were previously characterized using proteoliposome competition assays, making use of radioactively labeled peptides (Solcan et al., 2012). Here, the authors tested a set of peptides based only on alanine residues of increasing lengths, from one residue up to five. PepT$_{St}$ was reported to transport exclusively di- and tripeptides, being Ala-Ala preferred over Ala-Ala-Ala. Additionally, from these experiments it could be concluded that peptides carrying polar residues, such as Glu-Glu or Lys-Lys, are not transported preferentially.

Despite the efforts characterizing substrate preferences in peptides transporters, there were no screening methods to identify potential substrates that required little protein material and that were relatively fast and not so laborious as liposome assays. Therefore, in our ligand binding studies in solution, we aimed to develop a systematic and high-throughput method to screen a large ligand library and identify those ligands that could potentially be transported by PepT$_{St}$. We used differential scanning fluorimetry (DSF) for this purpose. The advantage of the setup designed by NanoTemper technologies is that up to 48 samples can be simultaneously measured, in a capillary-based format. The sample volume per capillary is reduced to ~10 μl, with low protein concentration (10 μM for PepT$_{St}$). Additionally, no protein labeling is required. The principle here is to measure

PepT$_{St}$ stabilization induced by the peptide upon heat denaturation (Niessen et al., 2007). As a second approach, we complemented these results with the determination of dissociation constants for those ligands that were identified as better binders by the DSF screening.

4.1.1.1 Peptide library screening by differential scanning fluorimetry (DSF)

A peptide library (with 28 different di- and tripeptides) was built aiming to screen a broad range of ligands in terms of physicochemical features. We tested peptide binding *in vitro* to detergent solubilized PepT$_{St}$ (Figure 20A). At a first glance, dipeptides induce a higher stabilization effect compared to tripeptides, which is consistent with the results reported in the liposome-based experiments (Solcan et al., 2012).
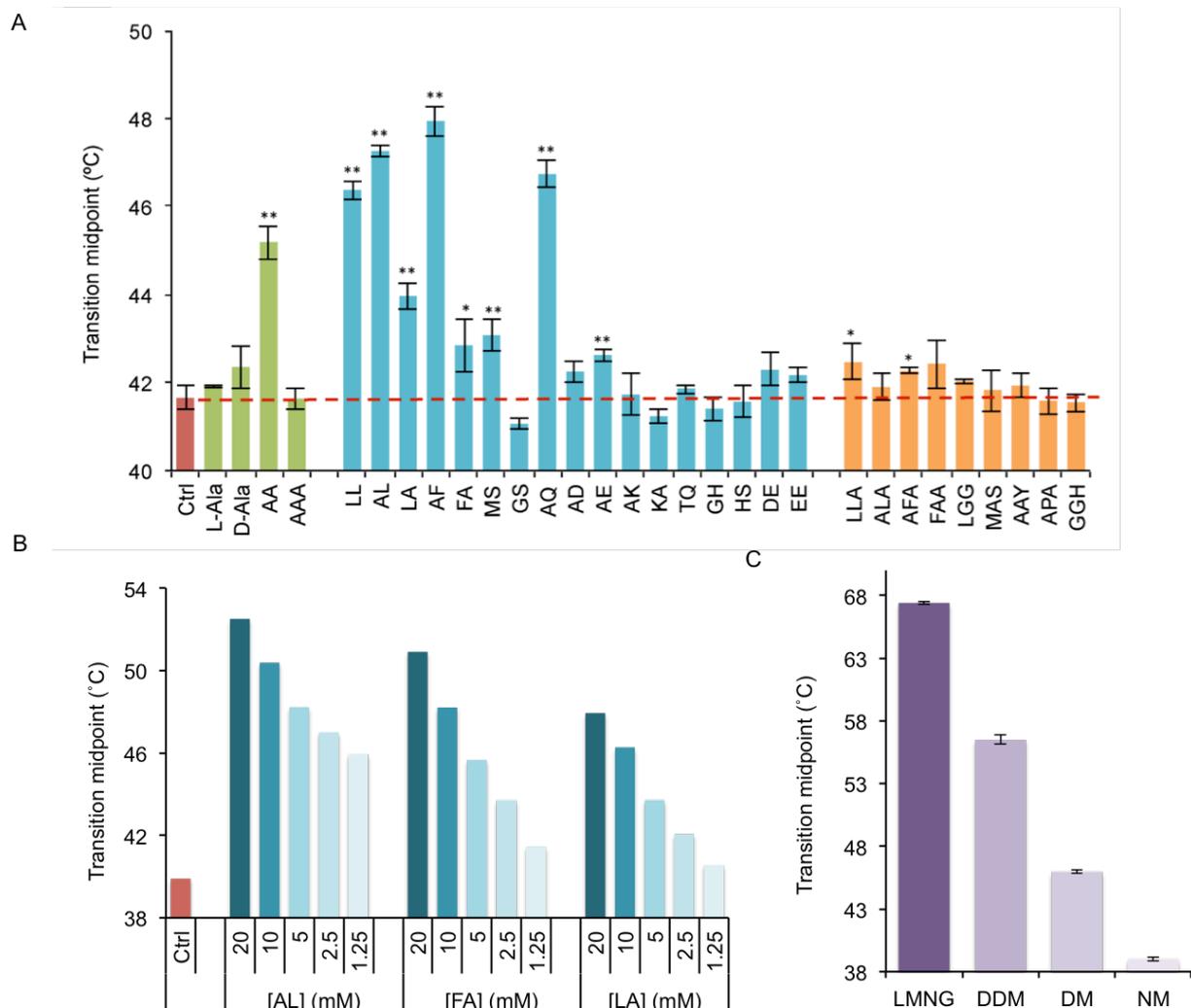
Figure 20: Peptide binding studies by DSF. (A) Screening of the peptide library. The peptide concentration was 5 mM in all the measurements. Green bars indicate peptides containing only alanine, dipeptides are represented by blue bars, and orange bars represent tripeptides. The red horizontal dashed line shows the Tm of the control sample, with no substrate present. Error bars indicate the standard deviation (SD) calculated from three independent measurements. The

level of significant difference between the control and the rest of the samples is indicated by asterisks (*p < 0.05 for significantly different samples and **p < 0.01 for highly significant differences). (B) Example of concentration dependence for peptides Ala-Leu, Ala-Phe and Leu-Ala. Dipeptides are labeled according to the one-letter code amino acid nomenclature. (C) Comparison of the stability of PepT$_{St}$ depending on the selected detergent for the purification: LMNG (2,2-didecylpropane-1,3-bis-β-D-maltopyranoside), DDM (n-dodecyl-β-D-maltoside), DM (n-decyl-β-D-maltoside) and NM (n-nonyl- β-D-maltoside). Figures in panels A and B were adapted form Martinez Molledo et al., 2018; figure in panel C was adapted from Quistgaard et al., 2017.

Among dipeptides, the ones inducing higher stabilization were those with nonpolar side chains of different sizes (such as Leu-Leu, Ala-Phe) or polar but not charged (Ala-Gln). On the contrary, dipeptides with charged side chains have either no effect or very little. This tendency is also conserved among tripeptides, with tripeptides like Leu-Leu-Ala having a significative stabilization effect compared to the rest of the tested tripeptides. Additionally, for a selected group of peptides, measurements were also performed at various concentrations to show that the stabilization effect is dependent on the presence of the peptide and not due to other factors. An example of a set of dipeptides measured by DSF at different concentrations is in Figure 20B.

It should be noted that the DSF measurements were performed in PepT$_{St}$ purified in n-nonyl- β-D-maltoside (NM) detergent. The reason behind this is that PepT$_{St}$ purified in n-dodecyl-β-D-maltoside (DDM) detergent is more stable in comparison to the protein purified in NM (Quistgaard et al., 2017), and under those experimental conditions the effect of peptide binding was reduced. Therefore, PepT$_{St}$ was destabilized by using a shorter acyl chain detergent (NM); when NM is used to purify PepT$_{St}$, the Tm value decays from ~56°C in DDM to ~40°C in NM (Figure 20C) and therefore, the addition of the peptides had a stronger stabilization effect.

4.1.1.2 Determining peptide affinities by MST

Next, we used microscale thermophoresis (MST) to quantify binding affinities of a subset of peptides (Table 4), based on the results of the DSF experiments. All the tested dipeptides presented a dissociation constant value (K_D) in the millimolar range, which varied from the low millimolar as in the case of Ala-Leu (0.56 mM) to high millimolar, as for Ala-Asp (here, the K_D could not be confidently estimated but it was above 100 mM). Furthermore, for other charged dipeptides, no K_D could be calculated. On the other hand, all tested tripeptides exhibited very low affinity or no binding at all under these experimental conditions, and therefore no K_D value could be estimated for any tripeptide. It can be then concluded that residues with either small or large hydrophobic side

chains can be accommodated at the N- or C-termini of a dipeptide, whereas binding of charged residues is generally disfavored. Moreover, Ala-Gln was also identified as a good PepT$_{St}$ binder, so that a glutamine could be a residue to fit in the binding pocket of PepT$_{St}$ when present at the C-terminus of a dipeptide. Note that these conclusions are in agreement with both the previous DSF data (Figure 20A), and the published uptake results based on liposomes studies (Solcan et al., 2012).

Peptide	Protein	pH	Buffer	K$_D$ (mM)	±
Dipeptides					
Ala-Leu	WT	7.5	100 mM HEPES	0.56	0.08
Ala-Phe	WT	7.5	100 mM Tris-HCl	0.95	0.24
Leu-Leu	WT	7.5	100 mM Tris-HCl	3.56	0.83
Ala-Gln	WT	7.5	100 mM Tris-HCl	6.82	1.33
Ala-Ala	WT	7.5	100 mM Tris-HCl	8.59	0.92
Phe-Ala	WT	7.5	100 mM Tris-HCl	10.95	2.22
Leu-Ala	WT	7.5	100 mM HEPES	13.34	0.87
Asp-Glu	WT	7.5	100 mM Tris-HCl	> 50	–
Ala-Asp	WT	7.5	100 mM Tris-HCl	> 100	–
Glu-Glu	WT	7.5	100 mM Tris-HCl	–	–
Ala-Lys	WT	7.5	100 mM Tris-HCl	–	–
Lys-Ala	WT	7.5	100 mM Tris-HCl	–	–
Tripeptides					
Ala-Ala-Ala	WT	7.5	100 mM Tris-HCl	> 100	–
Ala-Pro-Ala	WT	7.5	100 mM Tris-HCl	–	–
Leu-Leu-Ala	WT	7.5	100 mM Tris-HCl	–	–
Ala-Phe-Ala	WT	7.5	100 mM Tris-HCl	–	–
Ala-Leu-Ala	WT	7.5	100 mM Tris-HCl	–	–
pH screening					
Ala-Leu	WT	4.5	100 mM Citrate	> 50	–
Leu-Ala	WT	4.5	100 mM Citrate	> 100	–
Leu-Ala	WT	5.0	100 mM Citrate	47	5.1
Leu-Ala	WT	5.5	100 mM Citrate	23	1.87
Leu-Ala	WT	6.0	100 mM Citrate	20.45	1.37
Leu-Ala	WT	6.5	100 mM Citrate	20.0	4.24
Leu-Ala	WT	7.0	100 mM HEPES	21.33	2.50
Leu-Ala	WT	7.5	100 mM HEPES	21.8	6.83
Mutant screening					
Leu-Ala	R26A	7.5	100 mM Tris-HCl	> 50	–
Leu-Ala	E299A	7.5	100 mM NaP$_i$	> 50	–
Leu-Ala	E300A	7.5	100 mM Tris-HCl	–	–
Leu-Ala	E400A	7.5	100 mM NaP$_i$	–	–
Buffer screening					
Leu-Ala	WT	7.5	100 mM Tris-HCl	13.80	1.31
Leu-Ala	WT	7.5	100 mM NaP$_i$	9.56	1.24
Leu-Ala	WT	7.5	100 mM HEPES	13.34	0.87

Table 4: MST results for peptide binding to PepT$_{St}$. For each measurement, the following information is provided: Peptide used, PepT$_{St}$ wild type (WT) or protein carrying a point mutation, pH in which the measurement was performed, used buffer and concentration, estimated K$_D$ value (in mM) and associated error. The buffer for each measurement was additionally supplemented with 150 mM NaCl and 0.03% DDM. The table is divided in different blocks: measured dipeptides, tripeptides, measurements performed at different pH values, measurements performed using PepT$_{St}$ carrying single mutations in the binding pocket (later discussed in the results section) and the buffer screening at pH 7.5.

4.1.1.3 Effect of pH in ligand binding

POTs are secondary active transporters and they uptake peptides using a proton gradient as a driving force instead of ATP. Thus, for different bacterial POTs the residues that might play an important role in protonation have been described (Lyons et al., 2014; Parker et al., 2014). Under these considerations, we studied the effect of the pH on the stability of $PepT_{St}$ and also in ligand binding.

Again, using DSF $PepT_{St}$ thermostability was determined in a pH range between 4 and 9. Surprisingly, $PepT_{St}$ presents the highest Tm value at pH of 4.5 (46.5°C) after which it progressively decreases as the pH raises (Figure 21A). Secondly, we studied the influence of pH on substrate binding by measuring binding of Leu-Ala at pH 4.5, 5.0, 5.5, 6.0, 6.5, 7.0 and 7.5. Here, we found that binding affinities remain constant above 5.5, but they significantly drop as the pH is decreased from 5.5 (Figure 21 panels B and C, and Table 4). Thus, acidic pH stabilizes the protein but it does not favor substrate binding. As it will be explained later with more detail, this information was crucial to obtain the structures of $PepT_{St}$ in complex with peptides. For the MST measurements, the buffer concentration was kept at 100 mM to avoid any fluctuation of the pH due to the presence of dipeptides at high concentration.

Finally, we also studied the effect of different buffers (HEPES, Tris-HCl and NaP_i) at pH 7.5, to make sure that the buffer did not influence the observed binding. As indicated in Table 4, the K_D values for Leu-Ala binding to $PepT_{St}$ at pH 7.5 were in the same range regardless the used buffer.
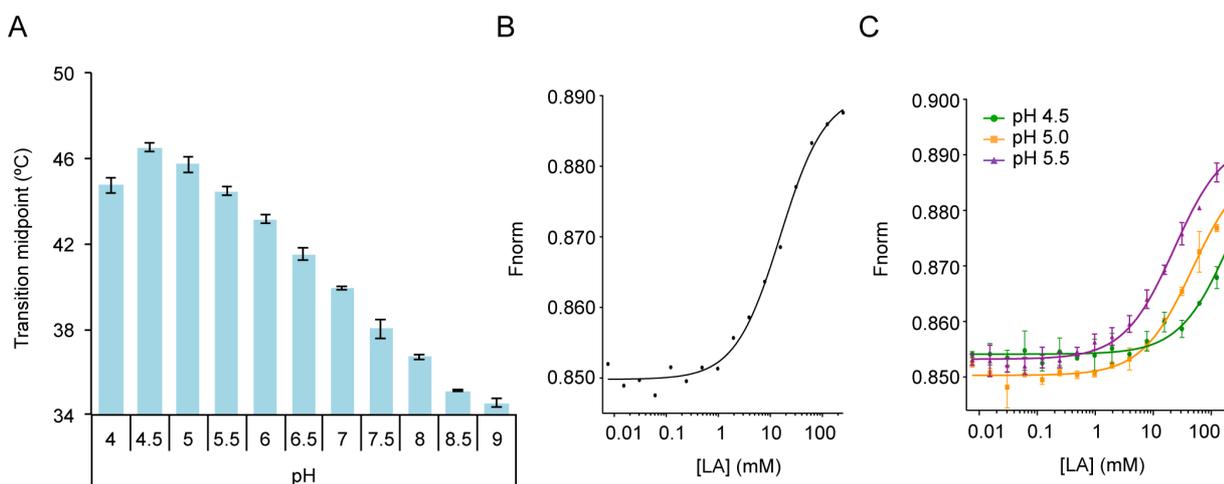
Figure 21: pH effect on PepT$_{St}$ stability and binding. (A) PepT$_{St}$ thermostability in a pH range from 4.0 - 9.0. (B) Representative binding curve for Leu-Ala at pH 7.5 (C) Binding of Leu-Ala at pH 5.5 (purple), 5.0 (orange) and 4.5 (green). Error bars indicate the standard deviation (SD) from three independent measurements. Figures form Martinez Molledo et al., 2018.
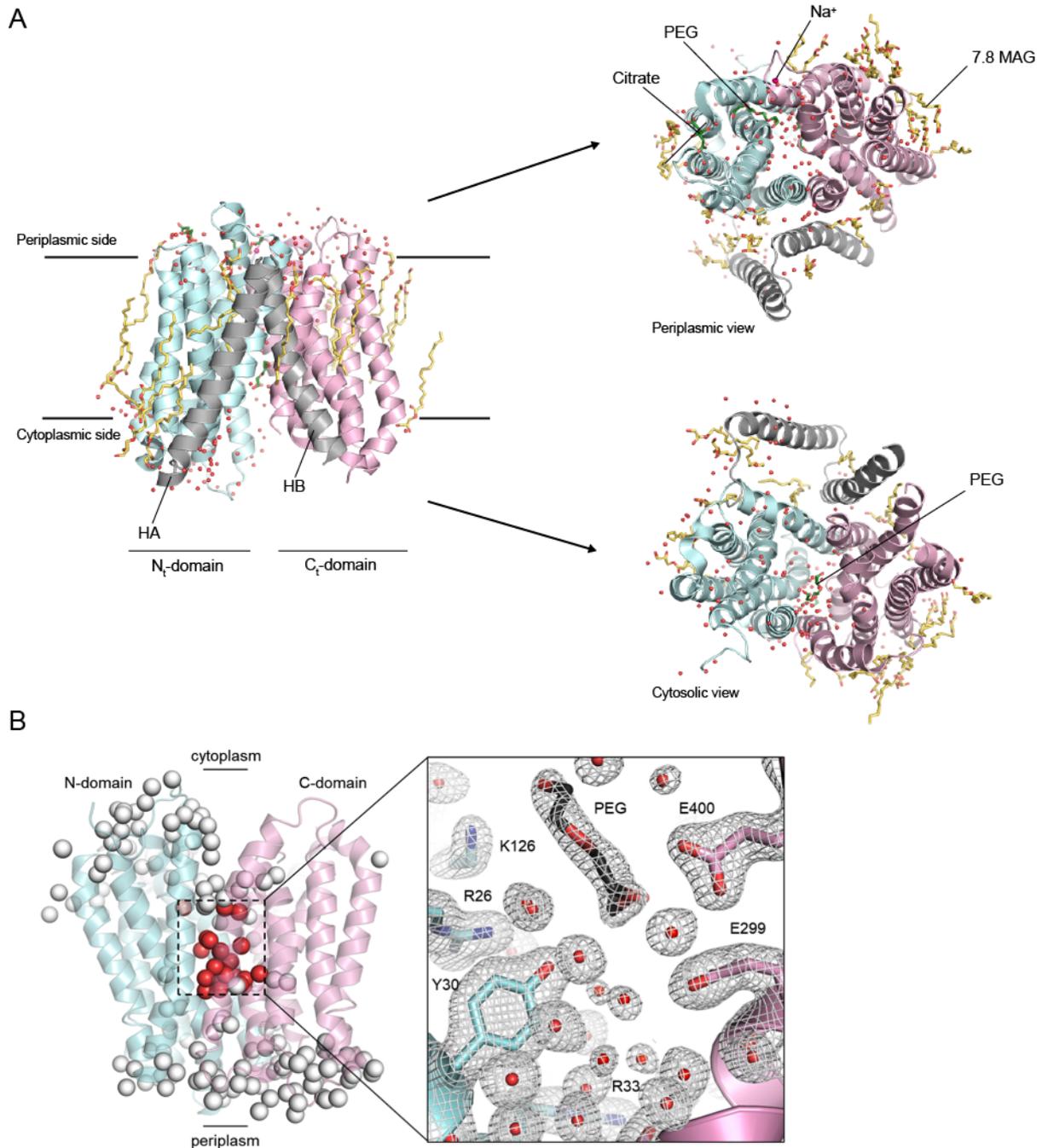
Figure 22: PepT$_{St}$ apostructure. (A) PepT$_{St}$ apostructure in the inward-open conformation. The N-terminal domain is colored in light cyan, the C-terminal domain is in light pink and the HA-HB domains are in grey. The black horizontal lines indicate the approximate position of the membrane relative to the transporter. On the right hand side of panel A, PepT$_{St}$ is shown from the periplasm (top) and the cytoplasm (bottom). Additionally, the rest of the molecules that could be modeled in the electron density maps are here indicated: a citrate molecule, the PEG molecules (one located on the periplasmic side of the transporter and the second one, in the binding cavity), a sodium ion (Na$^+$), the crystallization lipids (7.8 MAG, indicated as yellow sticks) and the water molecules (red spheres). (B) Distribution of the ordered water molecule network. Water molecules are represented as grey spheres or red spheres, for the water molecules located in the binding cavity. A detail image of the binding pocket shows the water molecules present in the binding cavity and the density identified as a PEG molecule.

4.1.2 PepT$_{St}$ apostructure

Previous to our studies, there were several PepT$_{St}$ apostructures reported: the first one by Solcan and colleagues (Solcan et al., 2012) was a rather low resolution structure (3.3 Å, PDB ID 4APS) but still valuable in the field as it was the first available model for PepT$_{St}$. Later, Lyons and colleagues (Lyons et al., 2014) (PDB ID 4D2B) reported a high-resolution crystallographic structure of PepT$_{St}$ obtained by the LCP method. Here, even though the model is considerably better, a careful inspection of the density map denotes that it might not be an apostructure in a strict sense as in the binding cavity, there is some additional density that was not assigned. Since then, several additional models of PepT$_{St}$ with no ligands bound have been reported (see Table 3 in the introduction for more information).

Our studies of ligand binding under different pH conditions reveled that ligand binding was indeed impaired under acidic conditions. Furthermore, PepT$_{St}$ is more stable under acidic conditions, which might indicate the stabilization of a particular state of PepT$_{St}$. We set up crystallization trials to screen for non-reported crystallization conditions, particularly focusing on crystallant components with an overall pH below 5.0. First hits were observed and after some optimization, a rather simple condition with sodium citrate at pH 4.5 and PEG300 (15-25%), yielded high-diffracting crystals. The data set was processed and a structure of PepT$_{St}$ could be determined at a resolution of 2.0Å (PDB ID 5OXO) (see Table 32 for the crystallographic data processing and refinement statistics in the appendix I).

In this model, PepT$_{St}$ is also in fully inward-open conformation. To date, this is the most complete available model of PepT$_{St}$, where only six residues could not be modeled in the loop between the HB domain and the TM7 and 10 more in the loop between TM10 and TM11. Furthermore, in the electron density map 20 molecules of the monoacylglycerol used in the LCP crystallization (7.8 MAG) could be modeled as well as two PEG molecules, a citrate molecule and a sodium ion.

Impressively, a high number of well-ordered water molecules were observed not only in this model but also in all the structures determined in the presence of ligands (Figure 22). The water molecules are located around the transporter and they are also occupying the binding cavity. Additionally, in the binding pocket, a PEG molecule was modeled in a small twisted electron density blob. It may represent either a PEG molecule or the aliphatic tail of a lipid or a detergent molecule. As the binding cavity is quite hydrophilic, it is more likely that it is a PEG molecule.

4.1.3 PepT$_{St}$ in complex with ligands

Crystallization trials were set up for PepT$_{St}$ in the presence of those ligands that according to the DSF screening and the MST characterization could be potential binders. The objective here was to generate PepT$_{St}$-dipeptides structures where the dipeptides presented certain variety in terms of size and chemical features. Here we attempted to understand how the same binding pocket could accommodate ligands with considerably different features. As a result, we determined four structures of PepT$_{St}$ in complex with dipeptides and we could conclude which are the adaptation mechanisms that the binding cavity has to accommodate different ligands.

4.1.3.1 Complexes with dipeptides Ala-Leu, Ala-Gln, Asp-Glu and Phe-Ala

For crystallization of the complexes, we used the LCP method and the crystallization condition reported previously that yielded the structures PepT$_{St}$[Ala-Phe] and PepT$_{St}$[Ala-Ala-Ala] (the latter will be discussed in the section corresponding to tripeptide complexes) (Lyons et al., 2014).
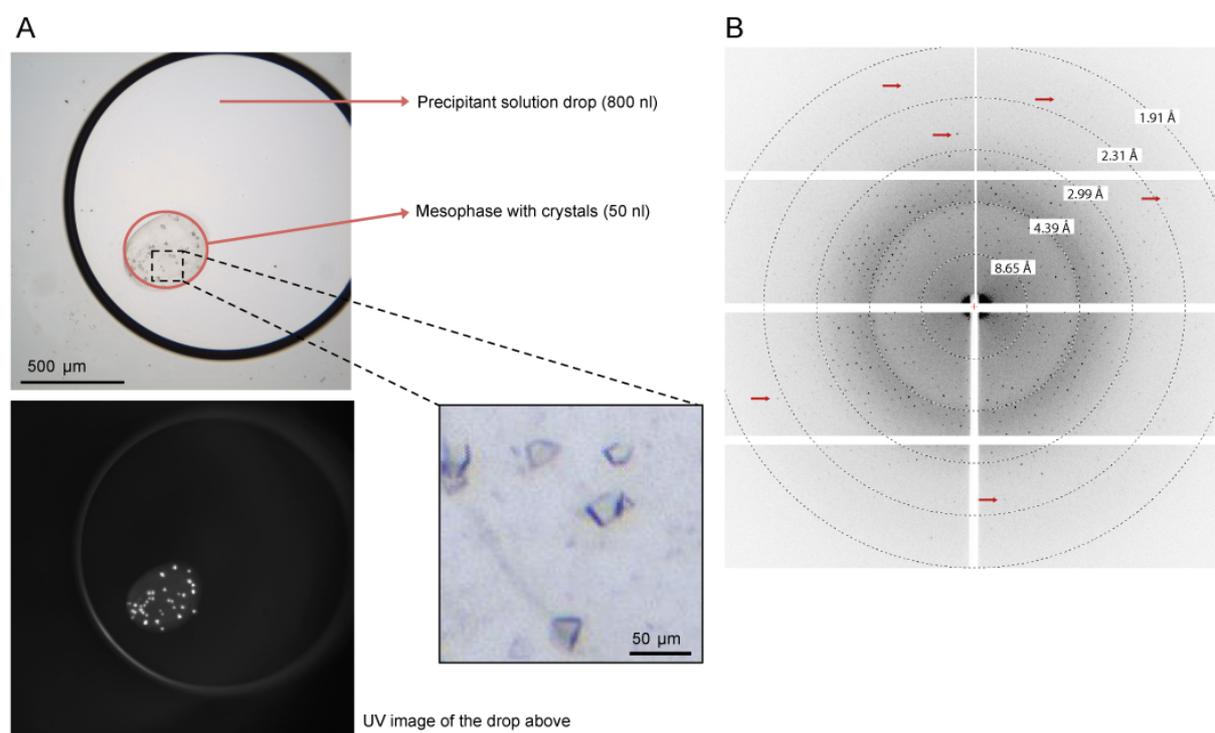
Figure 23: LCP crystallization of PepT$_{St}$ with dipeptides. (A) Example of a crystallization drop of a LCP plate for PepT$_{St}$. The upper image was taken under visible light, whereas the bottom corresponds to the same well but imaged under UV light. Here, the crystals are shining. (B) Example of a diffraction pattern of PepT$_{St}$. Crystals diffracted to an average resolution of 2.5 Å. The red arrows indicate diffraction beyond 2.5 Å.

This condition contained HEPES as a buffering system, a phosphate salt (ammonium phosphate monobasic, $NH_4H_2PO_4$), and the precipitant PEG400 at low concentrations (15-25%). Despite that

the HEPES buffer was at pH 7.0, the incorporation of the rest of the components decreased the overall pH of the system to between 5.0 and 5.8, depending on their concentration. As it was mentioned before, acidic pH stabilizes PepT$_{St}$ but does not favor binding of peptides. Therefore, the pH was the key parameter to be tightly controlled in the crystallization. The crystallant was measured and kept in a range between 5.5-6.0 to ensure ligand binding. As an alternative approach, we also tried to find new crystallization conditions at higher pH values, which would not restrict ligand binding. Unfortunately, these attempts failed, yielding only poorly diffracting crystals or no crystals at all. As a result, the pH range to be used in crystallization was considerably restricted (less than a unit) and it had to be carefully monitored in all our experiments.
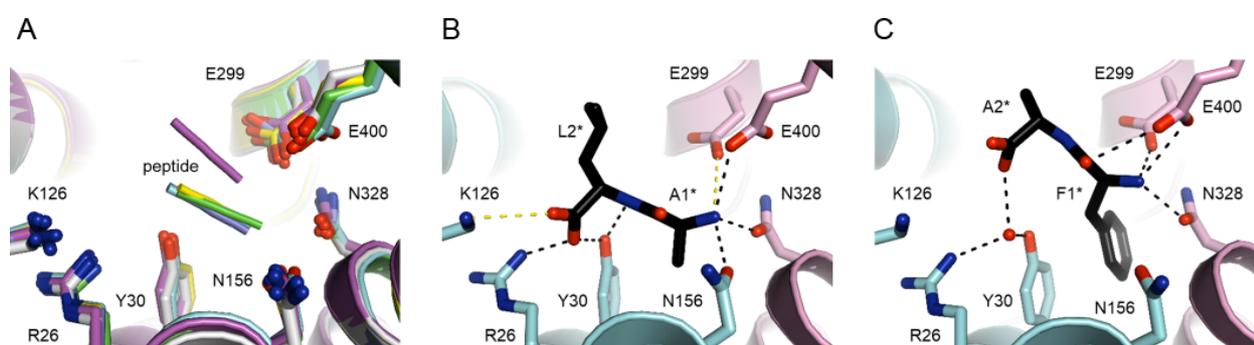
Figure 24: Interactions with the backbone, N- and C-termini of the dipeptides. (A) Differences in position of the peptide backbone and interacting residues. An overlay is shown of PepT$_{St}$[apostructure] (white), PepT$_{St}$[Ala-Leu] (light blue), PepT$_{St}$[Phe-Ala] (purple), PepT$_{St}$[Ala-Gln] (green) and PepT$_{St}$[Asp-Glu] (yellow). The main residues involved in the coordination of the peptide in the binding pocket are shown in sticks representation. (B) Backbone interactions of Ala-Leu, in the same orientation as in panel A. The N-domain of the transporter is colored in light blue and the C-domain is in pink. Black dashes indicate potential hydrogen bonds or salt bridges with lengths ≤ 3.2 Å, and yellow dashes indicate potential salt bridges with lengths of 3.2 - 4.0 Å. (C) Backbone interactions of Phe-Ala, shown as in panel B. Figures form Martinez Molledo et al., 2018.

Crystals grew in a reproducible manner under this crystallization condition and they diffracted X-rays to an average resolution of 2.5 Å (Figure 23). We determined structures of PepT$_{St}$ in complex with four dipeptides: Ala-Leu, Phe-Ala, Ala-Gln and Asp-Glu (Figure 24). From an experimental point of view, the structures in complex with dipeptides were obtained from the previously mentioned crystallization condition supplemented with the corresponding dipeptide: Ala-Leu and Phe-Ala were added to the screen in concentrations of 5 mM and 30 mM, respectively. Ala-Gln was premixed with PepT$_{St}$, reaching a final concentration of 100 mM in the mesophase, which was incubated at room temperature for one hour. The idea behind this approach was to reduce the pipetting that involved adding the peptide in the crystallization screen; unfortunately, this approach

has the limitation that once the crystallization drop is set up, the ligand might be further diluted. In this case, it might have worked due to the high affinity of the ligand. For Asp-Glu, a dry coating approach (Gelin et al., 2015) was used in order to reach the highest possible peptide concentration, considering the high K_D for this peptide.

One of the main challenges we faced when crystallizing PepT$_{St}$ by LCP was crystal harvesting. Even though PepT$_{St}$ crystallized in a reproducible manner under this condition, harvesting the crystals was not trivial, not only for the small crystal size but also because the environment where they grew was highly viscous. In many cases, just opening the drop caused the lost of the crystals; in other cases, the crystals were displaced in the well and it was hard to localize them. To increase the chances of success when freezing crystals, we designed the crystallization screens with a narrow variation around one condition. As a result, many of the drops had crystals and we could try many times to harvest them. Despite the improvement in harvesting as practicing more, it is evident that crystal freezing is still to date the major bottleneck in this crystallization approach.

The structures of PepT$_{St}$ in complex with dipeptides were refined to maximum resolutions of 2.7 - 2.2 Å, all in space group C222$_1$ and similar unit cell dimensions. The information related to the crystallographic data processing and refinement statistics is shown in Table 33 in the appendix I.
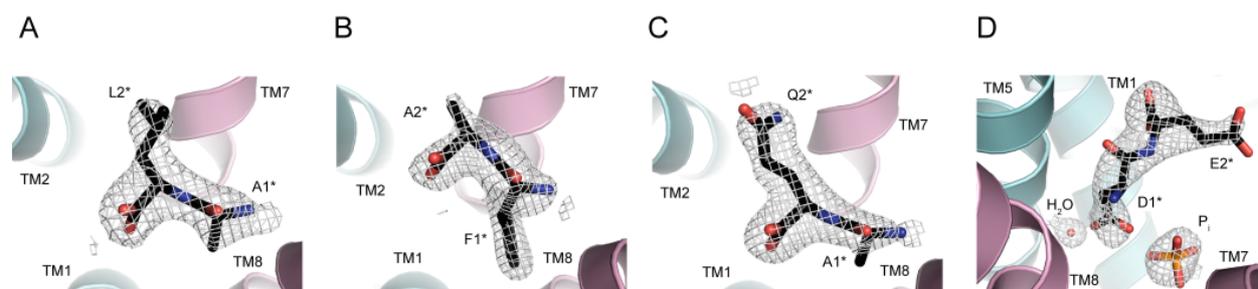
Figure 25: Electron density maps for bound dipeptides. (A) Binding of Ala-Leu in PepT$_{St}$[Ala-Leu]. N-domain is colored in light blue, C-domain is pink, the peptide is black and the 2Fo-Fc electron density map contoured at 1 σ around the peptide is grey. Transmembrane helices (TM) and peptide residues are correspondingly labeled. (B) Binding of Phe-Ala in PepT$_{St}$[Phe-Ala]. The color coding in the figure and orientation are the same as in panel A. (C) Binding of Ala-Gln in PepT$_{St}$[Gln-Ala]. (D) Binding of Asp-Glu in PepT$_{St}$[Asp-Glu]. Orientation of the complex is a bit different from the previous ones to get a better view of the peptide and the phosphate ion in the binding cavity.

It should be noted that these peptides already represent certain ligand diversity as they have side chains of different sizes and with different chemical features, either at N- or C-terminus of the

peptide. Moreover, they also vary in the binding affinities determined by the MST experiments: Ala-Leu presents a K_D value in the low millimollar range (0.56 mM), for Ala-Gln (6.82 mM) and Phe-Ala (10.95 mM) the K_D is still moderate, but Asp-Glu binds rather weakly to the transporter (>50 mM). This is reflected in the quality of the electron density maps: the density for the peptides is very well defined (Figure 25) with the exception of the glutamate side chain of Asp-Glu (Figure 25D); the poor density here might be caused by the low occupancy of the ligand in the binding cavity or the flexibility of the side chain.

Even though in these ligand bound complexes $PepT_{St}$ is in inward open conformation, there are small differences in the transmembrane helices 10 and 11 (TM10 and TM11): whereas complexes $PepT_{St}$[Phe-Ala] and $PepT_{St}$[Asp-Glu] are in the inward open conformation, $PepT_{St}$[Ala-Leu] and $PepT_{St}$[Ala-Gln] complexes present an inward facing partially occluded form. The partial occlusion is due to the pronounced bending of TM10 and TM11, which has also been reported for other members of the MFS (Lyons et al., 2014; Quistgaard et al., 2017). These details will be later addressed in the results chapter (see sections 4.2.1 and 4.2.2).

4.1.3.1.1 Coordination of N- and C-termini and the peptide backbone

The position in the binding cavity of the peptides was analyzed in terms of the N- and C-termini location, the coordination of the peptide backbone and the coordination of the side chains.

The first striking observation was the fact that the peptide backbones overlap nicely in the binding pocket except for dipeptide Phe-Ala, which has a markedly different position (Figure 24A): the backbone of Ala-Leu, Asp-Glu and Ala-Gln is located similarly to the one described for Ala-Phe before (Lyons et al., 2014), where Arg-26, Lys-126 and Tyr30 coordinate the C-terminus of the ligand; the later also interacts with the peptide backbone. The N-terminus is coordinated by Asn-156, Glu-299, Asn-328 and Glu-400. This ligand binding is represented in Figure 24B by the complex of Ala-Leu. On the other hand, for complex with Phe-Ala, the coordination differs mostly at the C-terminus of the ligand: Phe-Ala is tighter bound from its N-terminus by residues Asn-156, Glu-299 and Glu-400, but its C-terminus is only weakly coordinated and indirectly via a water molecule to residues Tyr-30 and Arg-26 (Figure 24C). It should be mentioned that an additional interaction present in this complex is the one that the side chain of the Tyr-30 forms with the phenylalanine ring of the ligand. This π stacking facilitates the approximation of the N-terminus of the ligand to the TM helices of $PepT_{St}$ C-domain.
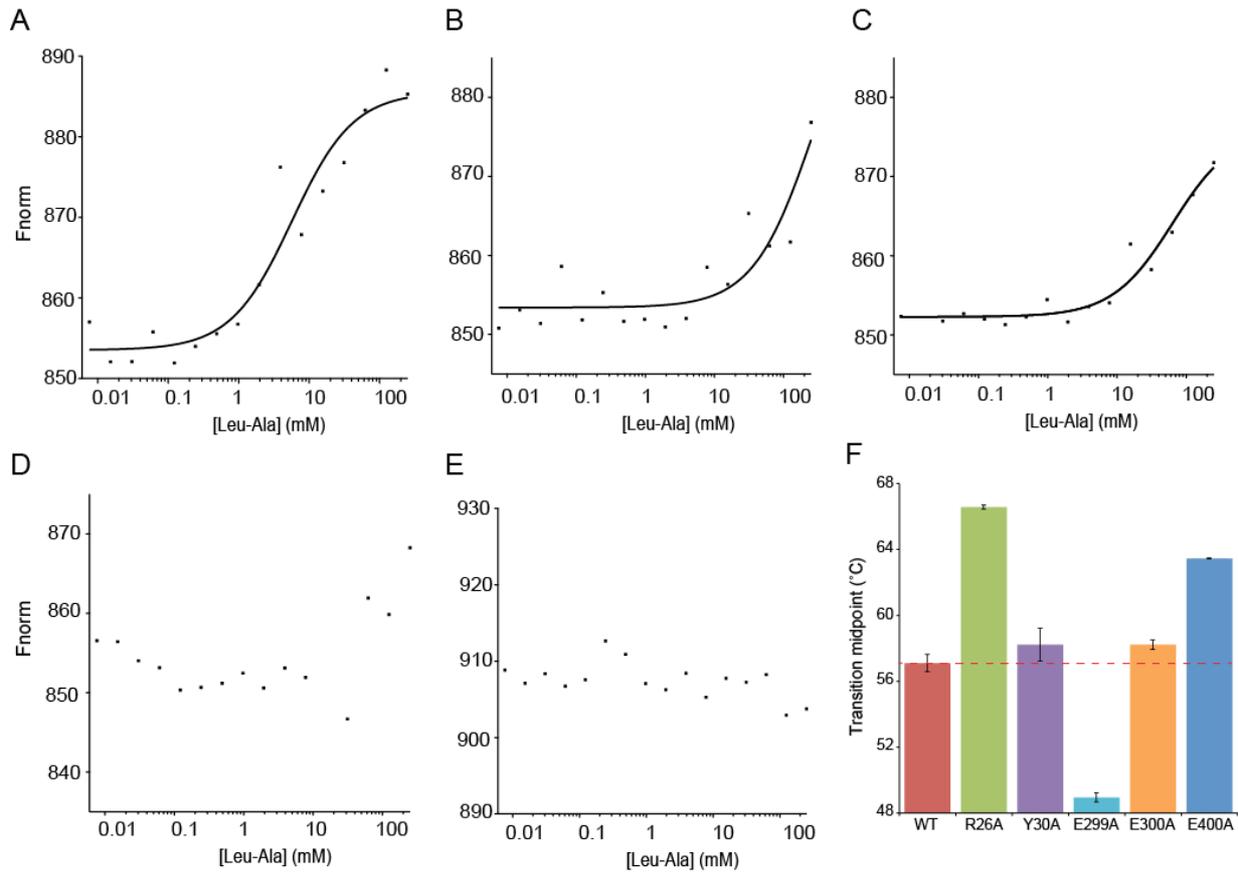
Figure 26: Leu-Ala binding to PepT$_{St}$ single mutants. (A) MST curve for Leu-Ala binding to detergent-solubilized PepT$_{St}$ WT, (B) PepT$_{St}$ R26A, (C) PepT$_{St}$ E299A, (D) PepT$_{St}$ E300A and (E) PepT$_{St}$ E400A. Mutants PepT$_{St}$ R26A and E299A present K$_D$ values > 50 mM, whereas it was not possible to estimate a K$_D$ value for mutants PepT$_{St}$ E300A and E400A. (F) Thermostability for several PepT$_{St}$ single mutants. Mutants PepT$_{St}$ Y30A (to be discussed in the tripeptide binding section) and E300A present a Tm value comparable to de WT, whereas mutant PepT$_{St}$ E299A is considerably unstable. Interestingly, mutations R26A and E400A are stabilizing.

4.1.3.1.2 Structure validation by mutational studies

The structures of the complexes were validated studying the binding in solution of peptide Leu-Ala to PepT$_{St}$ protein carrying point mutations in the binding pocket. Mutants R26A, E299A, E300A and E400A were successfully cloned, expressed and purified, and characterized by analytical size exclusion chromatography (aSEC), DSF and MST. We selected these mutants as these residues directly coordinate the ligand in the structures we determined except for Glu-300. SEC profiles were comparable to the WT protein (data not shown). In terms of stability, mutant E300A presented a Tm value comparable to the WT protein, E299A was more unstable and mutations R26A and E400A stabilized the protein. As indicated in Table 4, these mutants showed a considerable weaker binding to Leu-Ala, as in the case of PepT$_{St}$ R26A and E299A, or no binding at all, as for mutants PepT$_{St}$ E400A and E300A (Figure 26). Even though no direct ligand coordination was observed for

Glu-300, this residue has been described in literature to be involved in proton coupling. Furthermore, the equivalent glutamate residue (Glu-310) in GkPOT was described as a protonation site and when mutated into alanine, ligand binding and transport were impaired (Doki et al., 2013).

Figure 27: Pocket architecture in PepT$_{St}$-dipeptide structures. (A) Differences in the position of the peptide backbone in the structures in complex with dipeptides relative to the residues constituting P1 and P2. An overlay is shown of PepT$_{St}$[apostructure] (white), PepT$_{St}$[Ala-Leu] (light blue), PepT$_{St}$[Phe-Ala] (purple), PepT$_{St}$[Ala-Gln] (green) and PepT$_{St}$[Asp-Glu] (yellow). Tyr-68 is the only residue that presents greater variations in the different structures. (B) Location of residues forming P1 and P2 in PepT$_{St}$[Ala-Leu] complex structure. Residues forming P1 are represented in sticks and colored in green, P2 residues are colored in yellow and P2-lid is colored in orange. Ala-Leu ligand is colored in black. (C) Corresponding representation of PepT$_{St}$[Phe-Ala], (D) PepT$_{St}$[Ala-Gln] and (E) PepT$_{St}$[Asp-Glu]. Figures form Martinez Molledo et al., 2018.

4.1.3.1.3 Coordination of the side chains

When the peptide is bound to PepT$_{St}$, the N-terminal residue of the peptide occupies the so-called pocket one (P1), whereas the second residue is located in pocket two (P2). P1 is mostly hydrophobic and residues Tyr-30 and Ala-159 from the N-domain, as well as Asn-328 and Pro-329 from the C-domain compose it; P2 is more hydrophilic and it is formed by Tyr-68, Trp-296, Glu-300 from the N-domain and Ser-431 from the C-domain (Figure 27). In addition, two residues from the TM11, Trp-427 and Phe-428, form a flexible region of this helix and they can restrict the access to the binding cavity from the cytoplasmic side of the membrane. This part of the TM11 we have named the lid of pocket two (P2-lid) and it plays an important role in the partial occlusion of the transporter, which will be discussed in more detail in section 4.2.

Figure 27 describes how the ligands are located relative to the residues forming P1, P2 and P2-lid. There are two notable details here: First, the pockets are rich in aromatic residues, in particular P2 and P2-lid. As described elsewhere (Nishio et al., 2014; Salonen et al., 2011; Harder et al., 2013), aromatic residues have the ability of forming multiple types of interactions and they are ideal to promote the recognition of different ligands, as in the case of POTs (Guettou et al., 2013; Ito et al., 2013; Pieri et al., 2009). Secondly, the comparison of all structures in complex with peptides reveals no significant variation in the positions of the P1 binding site residues, but substantial variation in P2 (Figure 27A), particularly Tyr-68 adopts a range of different positions, which effectively tunes the size of P2. Tyr-68 is a very well conserved residue among POTs.
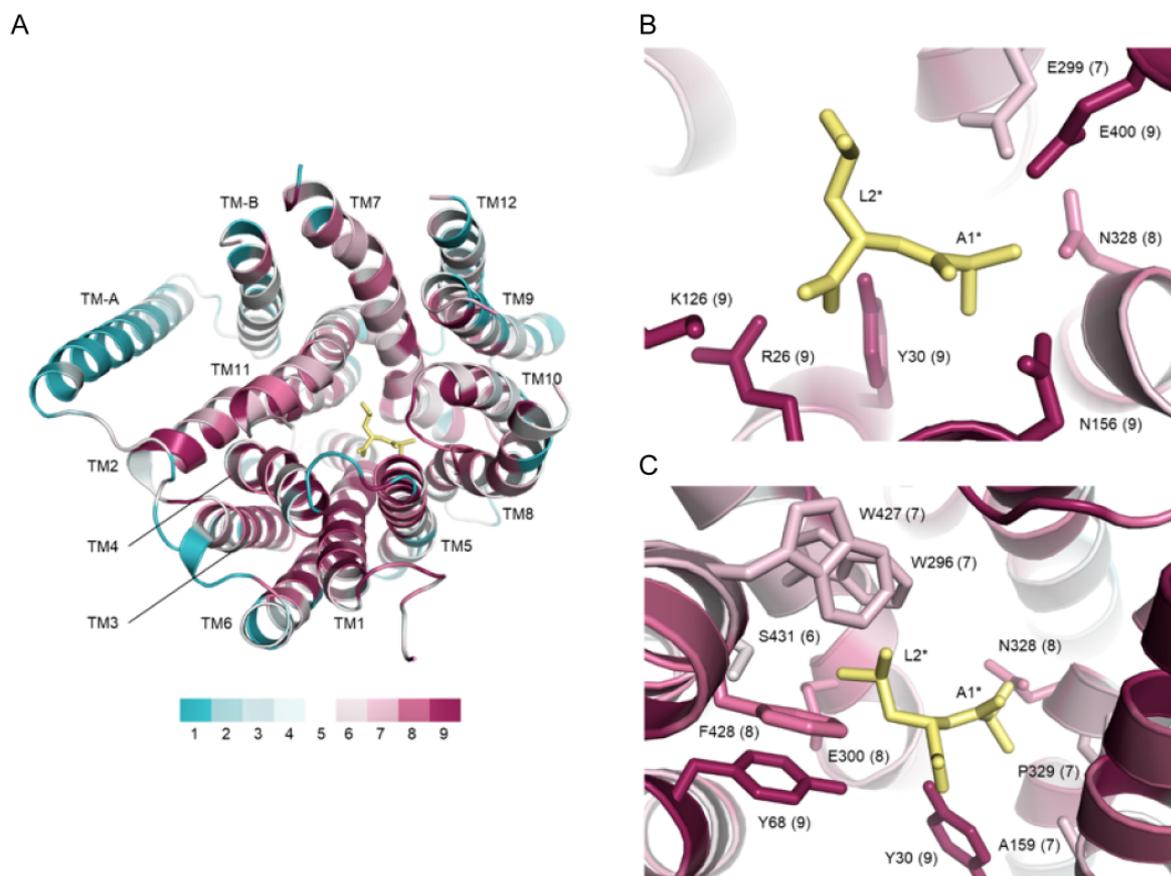
Figure 28: Residue conservation in PepT$_{St}$. (A) PepT$_{St}$ in complex with Ala-Leu viewed from the cytoplasmic side of the membrane. The ligand is represented in stick model and it is colored in yellow. PepT$_{St}$ is color-coded by conservation using ConSurf; the color is ramped from teal (low conservation, score = 1) over white to burgundy (high conservation, score = 9). (B) Conservation of binding site residues interacting with the peptide backbone. (C) Conservation of binding site residues forming P1, P2 and P2-lid. For panels B and C, each residue is numbered and the TM helix where they are located is also indicated. Figures form Martinez Molledo et al., 2018.

The rest of the residues in P1 and P2 are overall less well conserved than the residues interacting with the peptide backbone (Figure 28), but this is not surprising considering that POTs have broad substrate preferences (Boggavarapu et al., 2015; Brandsch et al., 2008; Chiang et al., 2004; Ernst et al., 2009; Fang et al., 2000; Guettou et al., 2014; Ito et al., 2013; Sharma et al., 2016; Solcan et al., 2012).
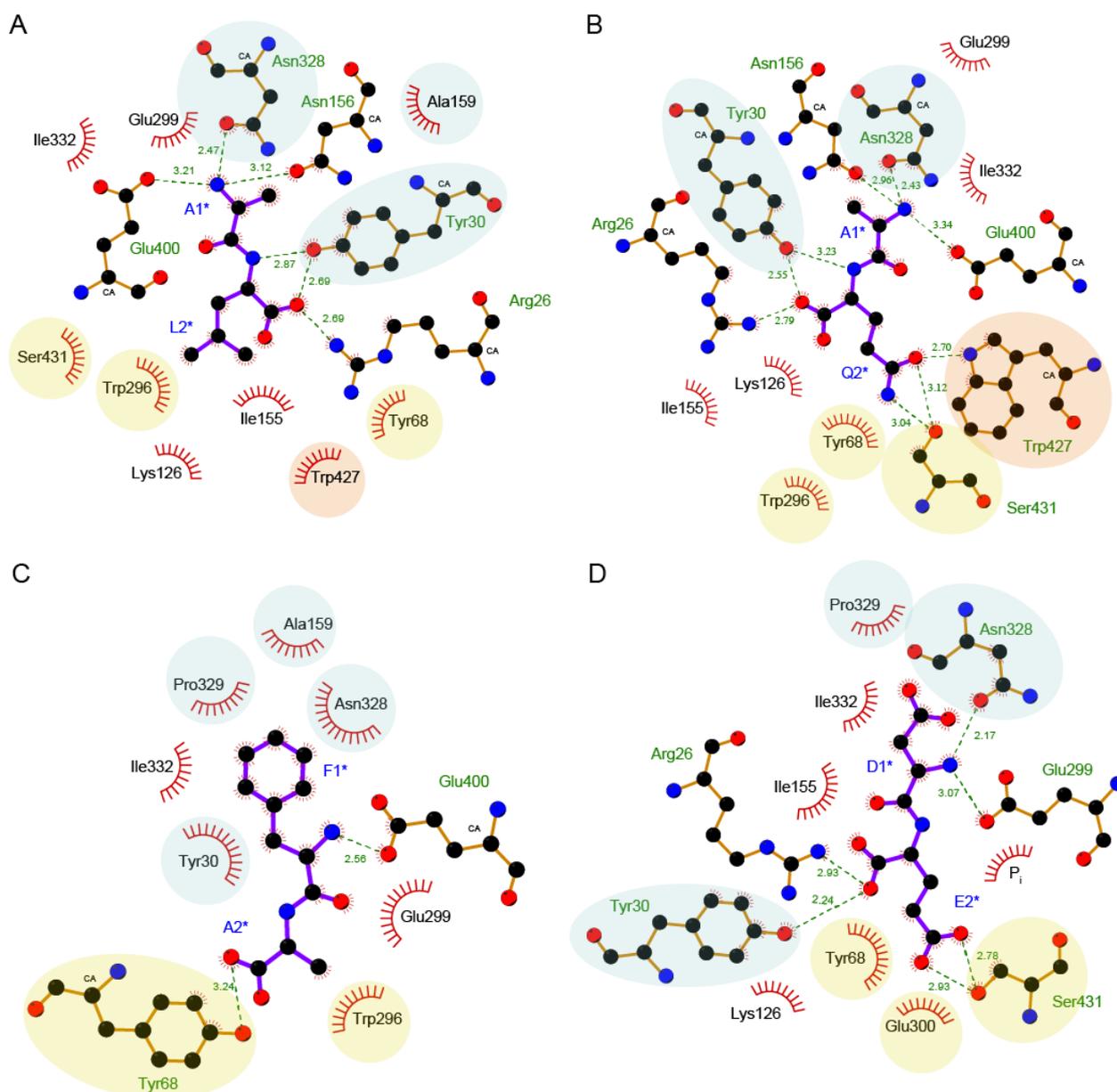
Figure 29: Coordination of side chains in PepT_St in complex with dipeptides. (A) LigPlot+ analysis and representation of the ligand Ala-Leu bound in complex PepT_St[Ala-Leu]. Here, the ligand is represented in purple sticks and labeled in the one letter code for amino acids. The main residues forming salt bridges or hydrogen bonds are represented in golden sticks and labeled in green; here, distances between these residues and the ligand are indicated by green dashed lines and measured in Å. Finally, the residues forming Van der Waals interactions (including arene interactions) are labeled in black and surrounded by a red semi-circle. Also, some of the binding pocket residues are circled and colored according to the pocket they form: green for P1, yellow for P2 and orange for P2-lid. (B)

Corresponding LigPlot⁺ analysis of the binding of Phe-Ala in PepT$_{St}$[Phe-Ala] complex, (C) PepT$_{St}$[Ala-Gln] complex and (D) PepT$_{St}$[Asp-Glu] complex.

4.1.3.1.3.1 Coordination of hydrophobic side chains

Among the dipeptides used for the complex generation, Ala-Leu and Phe-Ala are both hydrophobic. Here the difference is mainly the size of the residue side chains located at the N- or C-termini of the ligand: Ala-Leu has a small side chain in the residue at the N-terminus and a larger one at the C-terminal residue, whereas Phe-Ala has a large side chain at the N-terminus and a small one at the C-terminus. It was already described earlier in section 4.1.3.1 that the peptide backbone of Ala-Leu and Phe-Ala presents a slightly different position in the binding cavity. This is due to the role of Tyr-30 in the ligand coordination: there is a π stacking between the Tyr-30 ring and the side chain of the phenylalanine residue of ligand Phe-Ala, whereas Tyr-30 coordinates only the C-terminus of ligand Ala-Leu (Figure 29, panels A and B).

Apart from the coordination of the N- and C-terminus through salt bridges and hydrogen bonds, the side chains of these ligands are mostly coordinated by hydrophobic and Van der Waals interactions (Figure 29, panels A and B), as well as arene interactions: For Ala-Leu, the side chain of the leucine residue interacts with Trp-296 and Trp-427; for Phe-Ala, the side chain of the phenyalanine interacts with Tyr-30 via a π stacking and with Asn-328 via an amide-π stacking. The side chain of the alanine residue in Phe-Ala forms also a C-H··π interaction to Trp-296.

Furthermore, the size of the side chain of the residue at the C-terminus of the ligand determines the conformation of PepT$_{St}$: In PepT$_{St}$[Phe-Ala], the alanine side chain might be too small to interact with the residues forming the P2-lid, whereas in PepT$_{St}$[Ala-Leu] the side chain of the leucine interacts with the P2-lid and thus, TM11 is bended in this area. As a result the PepT$_{St}$[Ala-Leu] has an inward open partially occluded conformation.

4.1.3.1.3.2 Coordination of polar side chains

For complexes PepT$_{St}$[Ala-Gln] and PepT$_{St}$[Asp-Glu], the position of the peptide backbone is equivalent to Ala-Leu in PepT$_{St}$[Ala-Leu] complex, but the former complexes have side chains with polar groups and therefore, the coordination is different when compared to Ala-Leu (Figure 29, panels C and D): the amide group of the glutamine side chain in the Ala-Gln peptide forms hydrogen bonds to Ser-431 and Trp-427. Furthermore, the glutamine side chain forms Van der Waals interactions to Tyr-68 and Trp-296 as well as a N-H··π bond to Trp-296. In PepT$_{St}$[Asp-Glu], the aspartate Cβ atom forms similar Van der Waals interactions with P1 as the alanine residues

of Ala-Leu and Ala-Gln, while the carboxylic acid moiety is directed out of the groove. Here it forms a hydrogen bond with a well ordered water molecule (not shown in the LigPlot$^+$ in Figure 29D) and potentially, it forms an additional hydrogen bond with a larger molecule, which we have interpreted as phosphate. Unfortunately, for the glutamate residue the electron density for the side chain is rather poor, so no conclusions can be made on how this residue is coordinated in the binding pocket in detail.
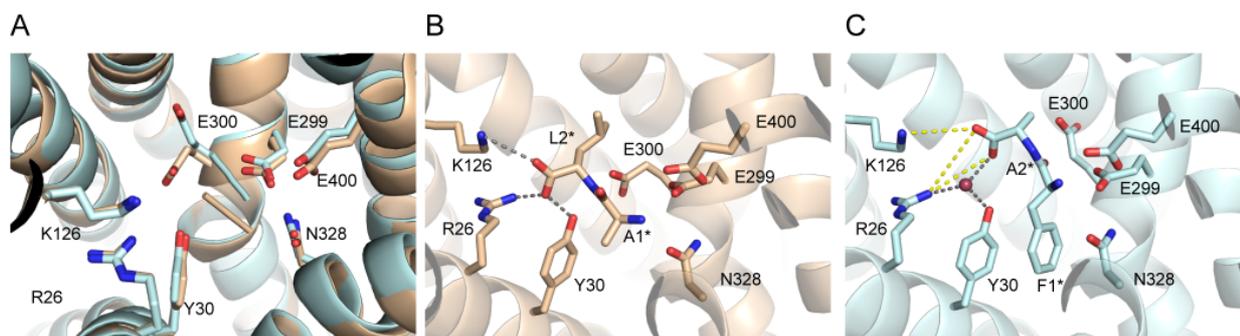
Figure 30: Alternative binding modes for dipeptides in PepT$_{St}$. (A) Comparison of the position of the dipeptide backbone for complex PepT$_{St}$[Ala-Leu] (PDB ID 5OXL) in wheat color and PepT$_{St}$[Phe-Ala] (PDB ID 5OXN) in light blue. Highlighted as sticks are the main residues coordinating the dipeptide backbone at N- and C-termini. (B) Coordination of the C-terminus of the ligand in the PepT$_{St}$[Ala-Leu] complex. Marked distances in dashed grey lines indicate the hydrogen bonds and salt bridges between the C-terminus and residues Lys-126, Arg-26 and Tyr-30 of the transporter, all of them in distances ranging from 2.7-3.4 Å. (C) Coordination of the C-terminus of the ligand in PepT$_{St}$[Phe-Ala] complex. The C-terminus is interacting through a water molecule (red sphere) with Tyr-30 and Arg-26 (distances 3.1-3.2 Å). There is no interaction with Lys-126.

4.1.3.2 Complexes with tripeptides Phe-Ala-Ala, Phe-Ala-Thr and Phe-Ala-Gln

From the dipeptide binding description before, we observed that the dipeptide backbone binds at least in two different positions: the first option is exemplified by complexes PepT$_{St}$[Ala-Leu], PepT$_{St}$[Ala-Gln] and PepT$_{St}$[Asp-Glu] (binding mode 1), and the second possibility was the one observed for complex PepT$_{St}$[Phe-Ala] (binding mode 2). It should be noted that, in the second binding mode, the dipeptide is tighter bound from the N-terminus, mostly due to the π stacking of the phenylalanine ring of the ligand and residue Tyr-30 in the binding cavity. As a result, there is additional space created at the C-terminus of the ligand (Figure 30). As a follow-up from the previous studies, we investigated whether peptides of sequence Phe-Ala-Xxx could bind to PepT$_{St}$, placing the third ligand residue in the additional space created in the binding cavity. It should be reminded at this point that among all the tested tripeptides in our original library (see Figure 20A),

none of them contained the sequence Phe-Ala-Xxx. Therefore, it was not tested experimentally up to that moment the hypothesis that an aromatic residue is necessary in position one of the ligand to see tripeptide binding. Furthermore, to date only three structures have been reported in which a tripeptide is bound to a bacterial POT (Figure 31): two structures of PepT$_{So2}$ from *Shewanella oneidensis* in complex with Ala-Ala-Ala (PDB ID 4TPJ) and Ala-Tyr(Br)-Ala (PDB ID 4TPG) ('Br' denotes that the residue is brominated) (Guettou et al., 2014), and a single structure of PepT$_{St}$ from *Streptococcus thermophilus* in complex with Ala-Ala-Ala (PDB ID 4D2D) (Lyons et al., 2014).
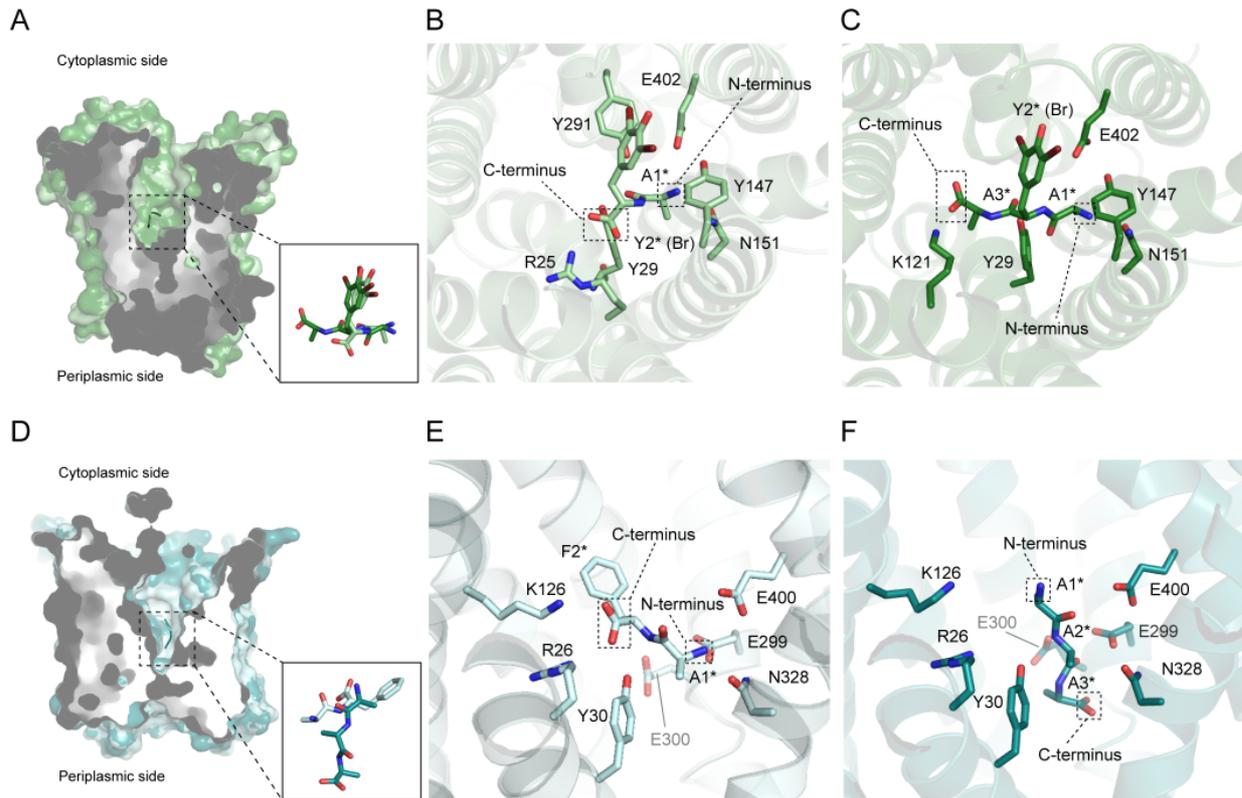
Figure 31: Binding of di- and tripeptides in PepT$_{So2}$ and PepT$_{St}$. (A) Overlay of PepT$_{So2}$ in complex with Ala-Try(Br) (PDB ID 4TPH) (in light green) and Ala-Tyr(Br)-Ala (PDB ID 4TPG) (in dark green) in surface representation. The zoom shows the different overlap of the dipeptide Ala-Tyr (Br) and the tripeptide Ala-Tyr (Br)-Ala. (B) PepT$_{So2}$ in complex with Ala-Try(Br). (C) PepT$_{So2}$ in complex with Ala-Try(Br)-Ala. The N- and C-termini of the ligand are indicated with boxes. The same orientation for both complexes is shown in panels B and C. (D) Overlay of PepT$_{St}$ in complex with Ala-Phe (PDB ID 4D2C) (in light cyan) and Ala-Ala-Ala (PDB ID 4D2D) (in dark cyan) in surface representation. The zoom shows the different position of the di- and the tripeptide in the binding cavity, where Ala-Phe is oriented horizontally and Ala-Ala-Ala is positioned vertically. (E) PepT$_{St}$ in complex with Ala-Phe. The N- and C-termini of the ligand are indicated with boxes. (F) Same orientation for complex PepT$_{St}$ [Ala-Ala-Ala].

In the PepT$_{So2}$ structures, both peptides were found to extend horizontally across the binding cavity, as also observed for all dipeptides (Figure 31). However, as the resolution was rather low (3.2 Å and 3.9 Å, for the di- and tripeptide complex respectively), the peptide backbone geometry and binding

mode could not be described in detail (Figure 31, panels A-C). The structure of PepT$_{St}$ in complex with Ala-Ala-Ala was determined at a moderately high resolution of 2.5 Å, and suggested an alternative vertical orientation for the tripeptide (Figure 31, panels D-F). However, as it will be later explained in the results section, we have reliable data supporting that this model should be revised; we have observed that the previously defined Ala-Ala-Ala vertically bound tripeptide might instead have been a misidentified HEPES buffer molecule (Martinez Molledo et al., 2018) (see section 4.1.5.2 for more details). Therefore, as these examples illustrate, there is a disagreement on how tripeptides are bound in POTs: there are two low-resolution structures for PepT$_{So2}$ and a questionable vertically bound tripeptide for PepT$_{St}$.
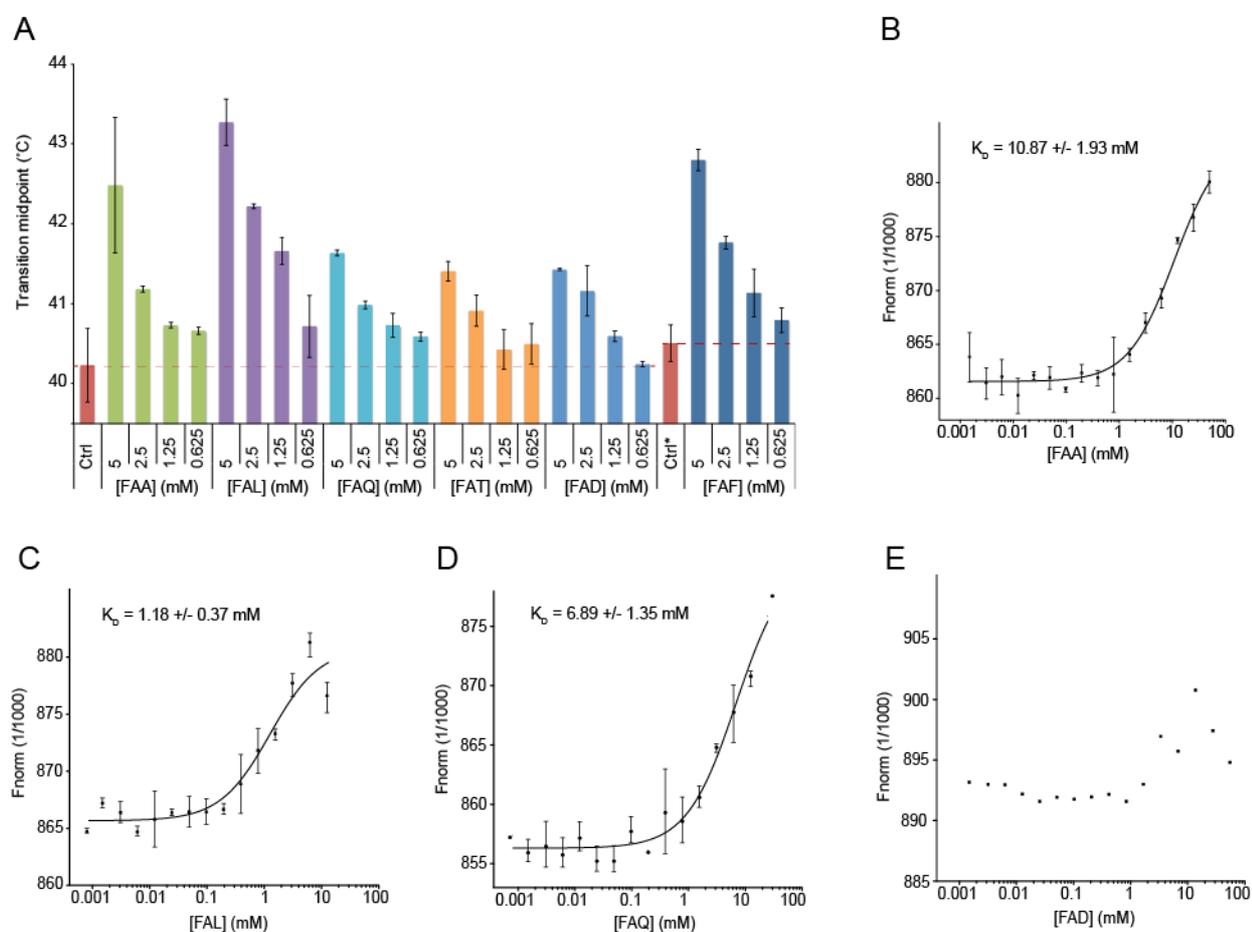
Figure 32: Tripeptide binding on detergent-solubilized PepT$_{St}$ measured by nanoDSF and MST. (A) Thermostability data of PepT$_{St}$ measured by nanoDSF. Each ligand was measured at four different concentrations (5, 2.5, 1.25 and 0.625 mM), as stated on the x-axis of the plot. Control samples (Ctrl) did not contain any tripeptide (Ctrl without DMSO, Ctrl* with 5% DMSO); they are shown in red bars, and the red horizontal dashed lines indicate the transition midpoint (Tm) for the control samples. Data corresponding to each ligand are color-coded as follows: Phe-Ala-Ala in light green, Phe-Ala-Leu in purple, Phe-Ala-Gln in light blue, Phe-Ala-Thr in orange, Phe-Ala-Asp in blue, and Phe-Ala-Phe in dark blue. Phe-Ala-Phe bars are referred to Ctrl* as this ligand was solubilized in DMSO. Average Tm value for each condition was calculated from three independent measurements. The error bars correspond to the standard deviation from the corresponding three independent measurements. (B) MST binding curve for tripeptide Phe-Ala-Ala, (C) Phe-Ala-Leu, (D) Phe-Ala-Gln and (E) Phe-Ala-Asp. The corresponding estimated dissociation constants

(K_D) are indicated for each tripeptide (when possible to calculate). All K_D values are in the millimolar range, except for Phe-Ala-Asp. Error bars indicate the standard deviation of two independent measurements.

Additionally, transport competition assays and binding studies have suggested that PepT$_{St}$ prefers dipeptides over tripeptides, although only a tiny subset tripeptides were tested, namely Ala-Ala-Ala, Ala-Pro-Ala, Leu-Leu-Ala, Ala-Phe-Ala and Ala-Leu-Ala (Solcan et al., 2012; Martinez Molledo et al., 2018). Note once again that none of the previously tested tripeptides conforms to the hypothesis that PepT$_{St}$ might bind tripeptides with the sequence pattern Phe-Ala-Xxx. We used differential scanning fluorimetry (DSF) (Niesen et al., 2007) for studying the binding of Phe-Ala-Ala, Phe-Ala-Leu, Phe-Ala-Gln, Phe-Ala-Thr, Phe-Ala-Asp and Phe-Ala-Phe to detergent solubilized PepT$_{St}$ (Figure 32A). These peptides were all found to stabilize the protein against thermal unfolding and aggregation, which is indicative of binding. The strongest effect was observed for peptides with a bulky apolar residue in the third position, Phe-Ala-Leu and Phe-Ala-Phe. These two peptides were also more stabilizing than any of the previously tested tripeptides, although markedly less than the best performing dipeptides (Martinez Molledo et al., 2018). Next, we used microscale thermophoresis (MST) (Seidel et al., 2013) to further characterize the binding of Phe-Ala-Xxx peptides to PepT$_{St}$. Here, we obtained the following affinity constants (Figure 32, panels B-D and Table 5): K_D (Phe-Ala-Ala) = 10.87 ± 1.9 mM, K_D (Phe-Ala-Leu) = 1.18 ± 0.3 mM, K_D (Phe-Ala-Gln) = 6.89 ± 1.3 mM, and K_D (Phe-Ala-Thr) = 28.09 ± 12.3 mM (data not shown). In the case of Phe-Ala-Asp, binding was insufficiently strong to allow for a K_D value to be determined (Figure 32E). The tripeptide Phe-Ala-Phe was only soluble in DSMO (100 mM stock), which did not allow us to determine a full binding isotherm using MST.

As a experimental detail, MST experiments were performed in a buffer with 400 mM HEPES pH 7.5 as the pH of the solubilized peptides in water was 2.0 - 2.5 (see section 6.3.6 in the materials and methods chapter). To compensate such acidic pH and maintain it constant despite the peptide concentration, HEPES buffer was required in high concentration. The corresponding controls were performed to ensure that high HEPES was not affecting peptide binding (see Table 5).

Peptide	Protein	pH	Buffer	K_D (mM)	±
Controls (different HEPES concentrations)					
Leu-Ala	WT	7.5	100 mM HEPES	15.02	1.35
Leu-Ala	WT	7.5	200 mM HEPES	17.97	1.72
Leu-Ala	WT	7.5	300 mM HEPES	15.63	2.54
Leu-Ala	WT	7.5	400 mM HEPES	17.65	3.92
Tripeptides					
Phe-Ala-Ala	WT	7.5	400 mM HEPES	10.87	1.9
Phe-Ala-Leu	WT	7.5	400 mM HEPES	1.18	0.3
Phe-Ala-Gln	WT	7.5	400 mM HEPES	6.89	1.3
Phe-Ala-Thr	WT	7.5	400 mM HEPES	28.09	12.3
Phe-Ala-Asp	WT	7.5	400 mM HEPES	-	-
Phe-Ala-Phe	WT	7.5	400 mM HEPES	-	-
Mutants					
Phe-Ala-Gln	R26A	7.5	400 mM HEPES	> 40 mM	
Phe-Ala-Gln	Y30A	7.5	400 mM HEPES	-	
Phe-Ala-Gln	E299A	7.5	400 mM HEPES	-	
Phe-Ala-Gln	E300A	7.5	400 mM HEPES	-	
Phe-Ala-Gln	E400A	7.5	400 mM HEPES	-	

Table 5: MST results for tripeptide binding to PepT$_{St}$. For each measurement, the following information is provided: Peptide used, PepT$_{St}$ wild type (WT) or protein carrying a point mutation, pH in which the measurement was performed, used buffer and concentration, estimated K_D value (in mM) and associated error. Furthermore, all solutions used for the measurements contained 150 mM NaCl and 0.03% DDM. The table is divided in different blocks: control experiments, measured tripeptides, and measurements performed using PepT$_{St}$ carrying single mutations in the binding pocket (later discussed in the results section).

As stated in Table 4, the dipeptide Phe-Ala dipeptide displays a K_D value of 10.95 ± 2.2 mM (Martinez Molledo et al., 2018). We conclude that: 1) Extending this peptide with an extra alanine residue has no effect on the binding affinity, suggesting that it can be accommodated in the binding site, but does not contribute significantly to the interaction. 2) Adding instead a threonine or aspartate residue moderately or strongly reduces affinity, respectively, indicating that especially the latter is clashing within the binding site. 3) Extending the peptide with a glutamine residue in the third position improves the affinity slightly, while adding a leucine residue improves it markedly, indicating that in particular the latter contributes significantly to the interaction of the tripeptide with the binding site. We conclude that both the DSF and MST results indicate that bulky apolar residues are preferred over smaller polar/charged ones in the third position of Phe-Ala-Xxx peptides. While it is possible that a systematic analysis of all di- and tripeptides would show in general that PepT$_{St}$ prefers dipeptides, some tripeptides evidently also bind with relatively high affinity.
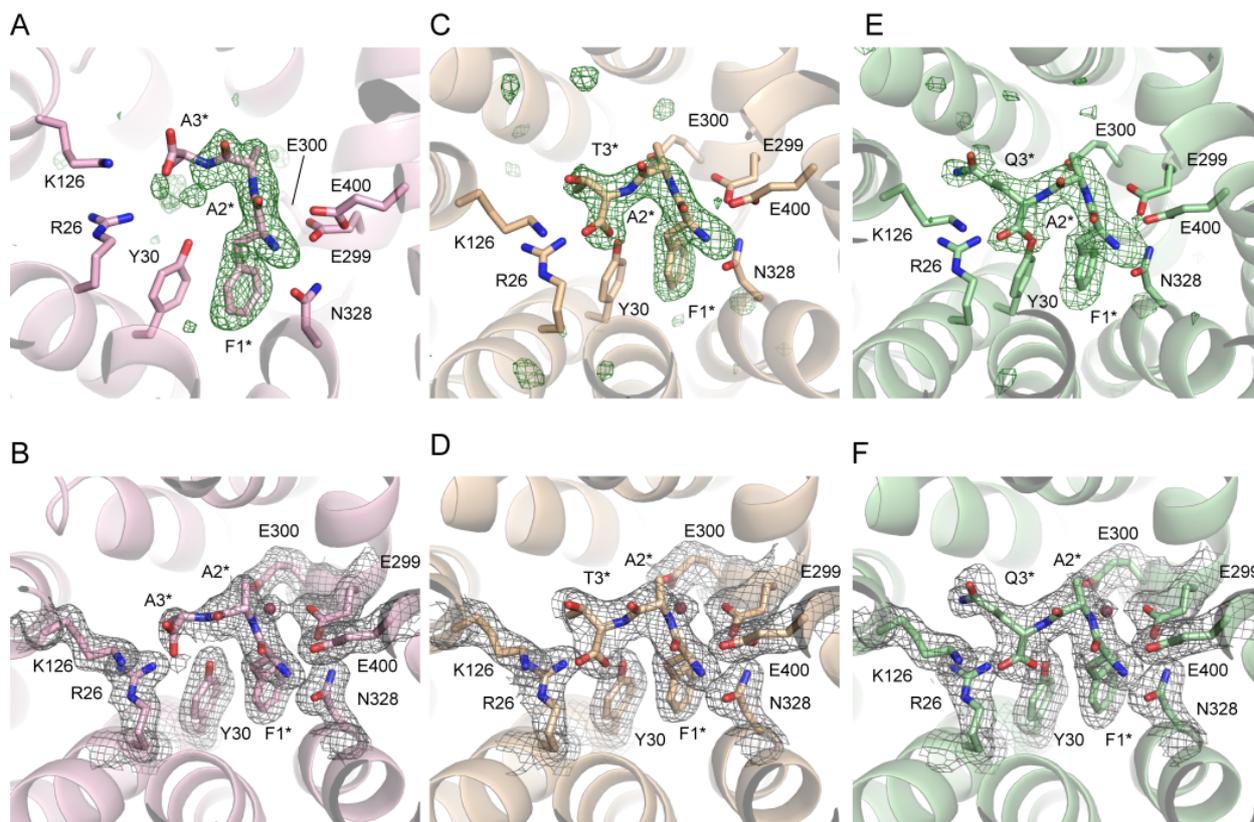
Figure 33: 1Fo-Fc composite omit maps and 2Fo-Fc electron density maps for tripeptide complex structures.
(A) Omit map for PepT$_{St}$[Phe-Ala-Ala] complex. The omit map is shown as a green mesh at a sigma level (σ) of 3. The alanine in the C-terminal position of the tripeptide is weekly overlapping with the omit map, indicating a certain degree of flexibility. Main residues coordinating the ligand are shown in sticks and labeled. Note that the orientation of this complex is slightly rotated in comparison to the rest of the panels, so that the 2Fo-Fc map is clearly visible at the C-terminus of the ligand. (B) Refined structure for the PepT$_{St}$[Phe-Ala-Ala] complex. The 1Fo-Fc electron density map is shown as a grey mesh, at a σ level of 1. (C) Omit map for PepT$_{St}$[Phe-Ala-Thr] complex at 3 σ level. Improved difference density is observed for the threonine residue. (D) Refined structure for PepT$_{St}$[Phe-Ala-Thr] complex. (E) Omit map for PepT$_{St}$[Phe-Ala-Gln] complex at 3 σ level. As in the case of PepT$_{St}$[Phe-Ala-Thr], the difference density covers the entire tripeptide. (F) Refined density map for PepT$_{St}$[Phe-Ala-Gln] complex at 1 σ level.

To obtain crystals with bound tripeptides, we used the previously published lipid cubic phase (LCP) crystallization conditions for PepT$_{St}$ peptide complexes (Lyons et al., 2014; Martinez Molledo et al., 2018). Also in this case the pH had to be carefully maintained in the range of 5.5-6.0 by using a concentration gradient of HEPES buffer in the crystallization screen. It should be reminded that this tripeptides were chemically synthetized and they presented a pH 2.0-2.5 when solubilized in water. In the crystallization trials, the tripeptide Phe-Ala-Leu had a high tendency to form crystals itself under the given conditions, and could therefore not be used further. Phe-Ala-Phe was poorly soluble in aqueous solutions and was therefore dissolved in DMSO. However, the presence of 5% DMSO was found to have a negative impact on LCP crystallization. Crystallization succeeded for

the following tripeptides: Phe-Ala-Ala, Phe-Ala-Thr and Phe-Ala-Gln and high-resolution diffraction data could be collected for all complexes (maximum resolution in the range of 2.3-2.0 Å). The crystal form was the same as observed for the previously determined structures grown under similar conditions (Lyons et al., 2014; Martinez Molledo et al., 2018).
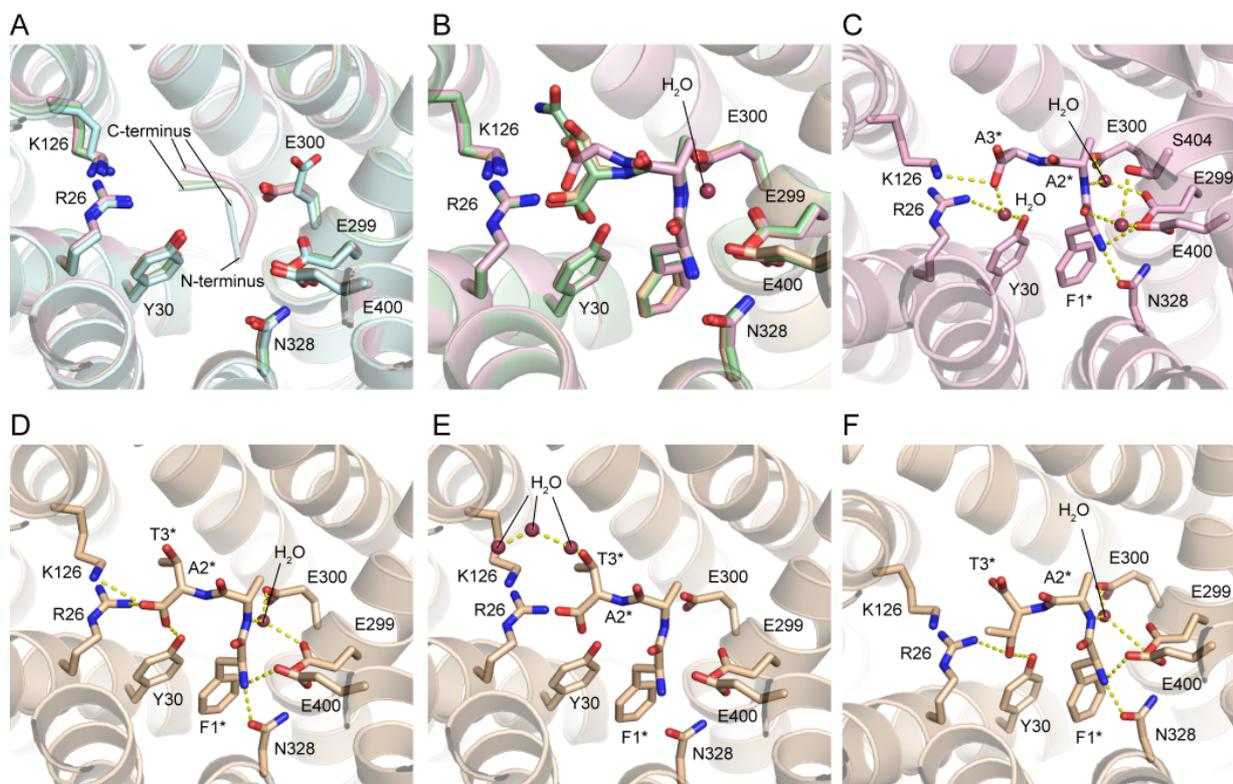
Figure 34: PepT$_{St}$ in complex with tripeptides. (A) Overlay of complex structures PepT$_{St}$[Phe-Ala] (in light blue), PepT$_{St}$[Phe-Ala-Ala] (in light pink), PepT$_{St}$[Phe-Ala-Thr] (in wheat) and PepT$_{St}$[Phe-Ala-Gln] (in light green). The main residues coordinating the ligands are represented as sticks. The ligands are represented as a cartoon, where the position of the N- and C-termini is labeled. Note the bending of the peptide backbone for the tripeptides. (B) Overlay of tripeptides Phe-Ala-Ala, Phe-Ala-Thr and Phe-Ala-Gln in the binding pocket. Complexes are colored as in panel A. (C) Coordination of the ligand Phe-Ala-Ala in PepT$_{St}$[Phe-Ala-Ala] structure. (C) Coordination of the peptide backbone and the N- and C-termini in PepT$_{St}$[Phe-Ala-Thr] structure. (D) Coordination of the side chain in the PepT$_{St}$[Phe-Ala-Thr] structure. (E) Alternative position of the C-terminus of ligand Phe-Ala-Thr for complex PepT$_{St}$[Phe-Ala-Thr].

In all the reported complexes with tripeptides, the conformation of PepT$_{St}$ was fully inward open; no occlusion due to bending of TM11 was observed. For all three complexes, clear density was observed for the bound tripeptides in both the electron density maps (1Fo-Fc) and omits maps (2Fo-Fc) (Figure 33). It is clear that the peptides all extend across the binding cavity formed in the space between the two MFS domains of the protein, and this supports an horizontal binding of ligands instead of the vertical binding mode proposed for tripeptdes by Lyons and colleagues (Lyons et al., 2014). It is furthermore clear that the first two residues of the ligand adopt the same position

in all three structures (Figure 34, panels A and B). However, the third residue could only be confidently modeled in the case of Phe-Ala-Gln. For Phe-Ala-Ala, the density was very weak for the C-terminal alanine residue (Figure 33A), implying that it interacts poorly with the protein, as was also indicated by the binding experiments. For Phe-Ala-Thr, the electron density map supports the same peptide backbone configuration as the one observed for Phe-Ala-Gln (Figure 33, panel C and D). However, it could also support an alternative position in which the threonine side chain and the C-terminal carboxylate moiety are interchanged (Figure 34, panels D-F). Despite the good statistics for the structures in complex with Phe-Ala-Ala and Phe-Ala-Thr (data not shown), we should be cautious when describing the coordination for the third ligand residue as the map is not good for this position: For complex PepT$_{St}$[Phe-Ala-Ala], the C-terminus might be directly coordinated by Lys-126 and via a water molecule to Arg-26 and Tyr-30 (Figure 34, panel C). On the other hand, the C-terminus of the complex PepT$_{St}$[Phe-Ala-Thr] is coordinated by Lys-126, Arg-26 and Tyr-30, and the side chain of threonine residue could interact with water molecules located in the binding cavity (Figure 34, panels D-F).
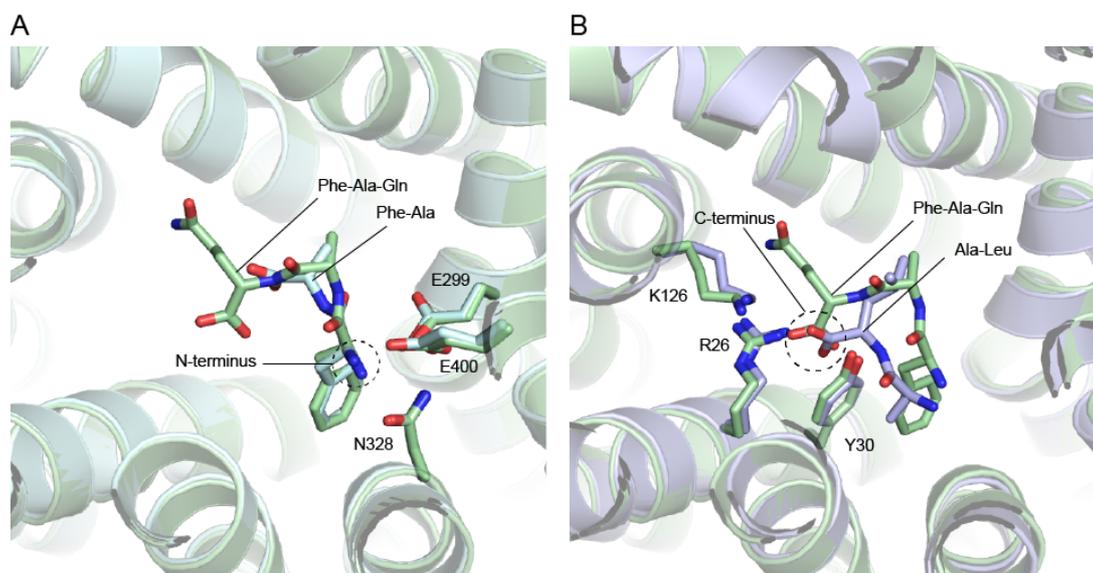
Figure 35: Coordination of tripeptide Phe-Ala-Gln in comparison to dipeptides Ala-Leu and Phe-Ala. (A) The coordination of the N-terminus of Phe-Ala-Gln (colored in light green) is similar to the one observed for dipeptide Phe-Ala (colored in cyan). In the figure, only the residues of the binding cavity involved in the coordination of the N-terminus are represented in sticks and correspondingly labeled. A circle highlights the N-terminus of both peptides. (B) On the other hand, the coordination of the C-terminus of Phe-Ala-Gln is comparable to the one observed for dipeptide Ala-Leu (in light purple). The C-terminus of the peptides is indicated by a circle.
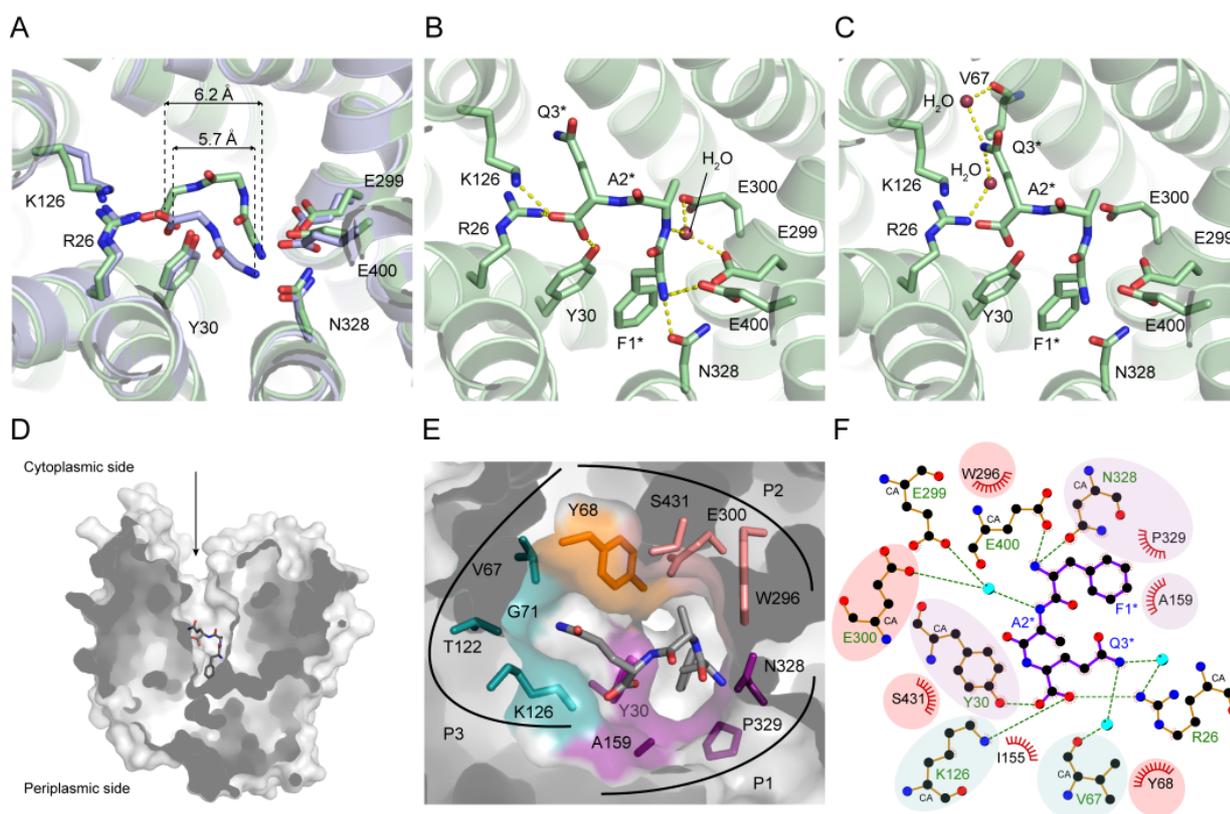
Figure 36: Structural basis for binding of Phe-Ala-Gln to PepT$_{St}$. (A) Comparison of the peptide backbone configurations in PepT$_{St}$[Ala-Leu] (light blue) and PepT$_{St}$[Phe-Ala-Gln] (light green). Main residues coordinating the N- and C-termini and the peptide backbone are highlighted in sticks and labeled, except for residue Glu-300 that was removed for clarity. The main chain of the peptides is represented, while the side chains have been removed. Here, the distance between the N-terminus and the C-terminal carboxylate carbon of each peptide is correspondingly indicated and measured. (B) Interactions with the peptide backbone and termini in PepT$_{St}$[Phe-Ala-Gln]. (C) Interactions with the peptide glutamine side chain in PepT$_{St}$[Phe-Ala-Gln]. (D) PepT$_{St}$[Phe-Ala-Gln] in surface representation. The arrow indicates the accessibility to the binding cavity from the cytoplasmic side of the membrane (the protein is in the inward open conformation). (E) Binding pockets in PepT$_{St}$. Pocket 1 (P1) is colored purple, pocket 2 (P2) is salmon, pocket 3 (P3) is cyan, and the peptide Phe-Ala-Gln is grey. Note that Tyr-68 (colored orange) is not only part of P2 but also P3. The peptide and the residues forming the pockets are represented as sticks. (F) LigPlot+ diagram for the binding of Phe-Ala-Gln to PepT$_{St}$ with the pockets indicated by different background colors (using the same color scheme as in panel E). Hydrogen bonds and ionic interactions are indicated by green dashes, and the residues involved are labeled in green. Residues forming hydrophobic contacts with the peptide are labeled in black.

Certainly, we are confident about the model with Phe-Ala-Gln as a ligand and here, we can assure the coordination of the ligand in the binding cavity (see Table 34 in the appendix I for the data collections and statistics for PepT$_{St}$[Phe-Ala-Gln] complex): The N- and C-termini of Phe-Ala-Gln interact with the same residues in PepT$_{St}$ as the termini of the dipeptides. Specifically, the N-terminus interacts with a subsite consisting of Glu-299, Asn-328 and Glu-400 in the same manner as observed for Phe-Ala (dipeptide binding mode 2) (Figure 35A), while the C-terminus interacts with

a subsite consisting of Arg-26, Tyr-30, and Lys-126 in a similar way as observed for Ala-Leu (dipeptide binding mode 1) (Figure 35B). Phe-Ala-Gln is able to fit in the space between these subsites because the backbone presents a torsion in the central alanine residue (Figure 36A). Indeed, the distance between the N-terminal nitrogen atom and the C-terminal carboxylate carbon atom is only slightly longer for Phe-Ala-Gln than for Ala-Leu (6.2 Å versus 5.7 Å). Apart from the interactions formed with the termini, the only other backbone interaction is a putative water-mediated hydrogen bond between the nitrogen atom of the peptide alanine residue and the side chains of Glu-299 and Glu-300 (Figure 36B). However the electron density for the implicated water molecule is rather weak, suggesting that it is only partially occupied and therefore probably not of great importance for the interaction with the peptide.

Concerning the Phe-Ala-Gln side chains, the first two fit into two mostly hydrophobic and aromatic pockets, denoted pocket 1 (P1) and pocket 2 (P2), respectively, in the same way as previously described for the Phe-Ala dipeptide (Martinez Molledo et al., 2018), while the third residue fits into a previously undescribed pocket 3 (P3). As stated above, the movement of the Tyr-68 side chain controls to some extent the space available for a third peptide residue in the binding site, as it fine-tunes the size of P2 (Martinez Molledo et al., 2018). Consistent with this, Tyr-68 was found at the border of P2 and P3, partially separating them from each other (Figure 36E).

The side chain of the peptide glutamine residue appears to form van der Waals interactions with Tyr-68 and Lys-126, as well as hydrogen bonds with two water molecules, one of which also interacts with the backbone carbonyl of Val-67, and the other of which also interacts with Arg-26 (Figure 36C). However, it should be pointed out that the electron density for the side chain of the peptide glutamine residue is rather weak beyond the Cβ atom (as shown in the difference map in Figure 33E). Although we can determine its general direction we cannot be fully confident about the rotamer, and thereby determine the exact interactions it forms within P3. Other residues delineating P3 include Val-67, Gly-71, Thr-122, and Lys-126 (Figure 36, panels E and F). As the polar and charged moieties of Thr-122 and Lys-126 point away from the pocket, P3 is mostly hydrophobic. However, in the inward open state of the protein, it is sufficiently open to the cytoplasmic side that water molecules can enter.

Finally, to validate our ligand-bound model, we mutated the residues interacting with the ligand and study the effect of these mutations in ligand binding. We used MST to test binding of Phe-Ala-Gln to the following five PepT$_{St}$ mutants: R26A, Y30A, E299A, E300A, and E400A. The affinity was markedly reduced in all cases, several fold for PepT$_{St}$ R26A while for the other mutants binding

could no longer be determined (Table 5 and Figure 37). In this case, binding of Phe-Ala-Gln to PepT$_{St}$ Y30A was measured as residue Tyr-30 was observed to be directly interacting with the C-terminus of the ligand. This mutant was generated, expressed, purified and characterized by DSF (Figure 26F). PepT$_{St}$ Y30A presented a Tm value comparable to the WT protein. The mutagenesis data thus supports the observations from the structural analysis.
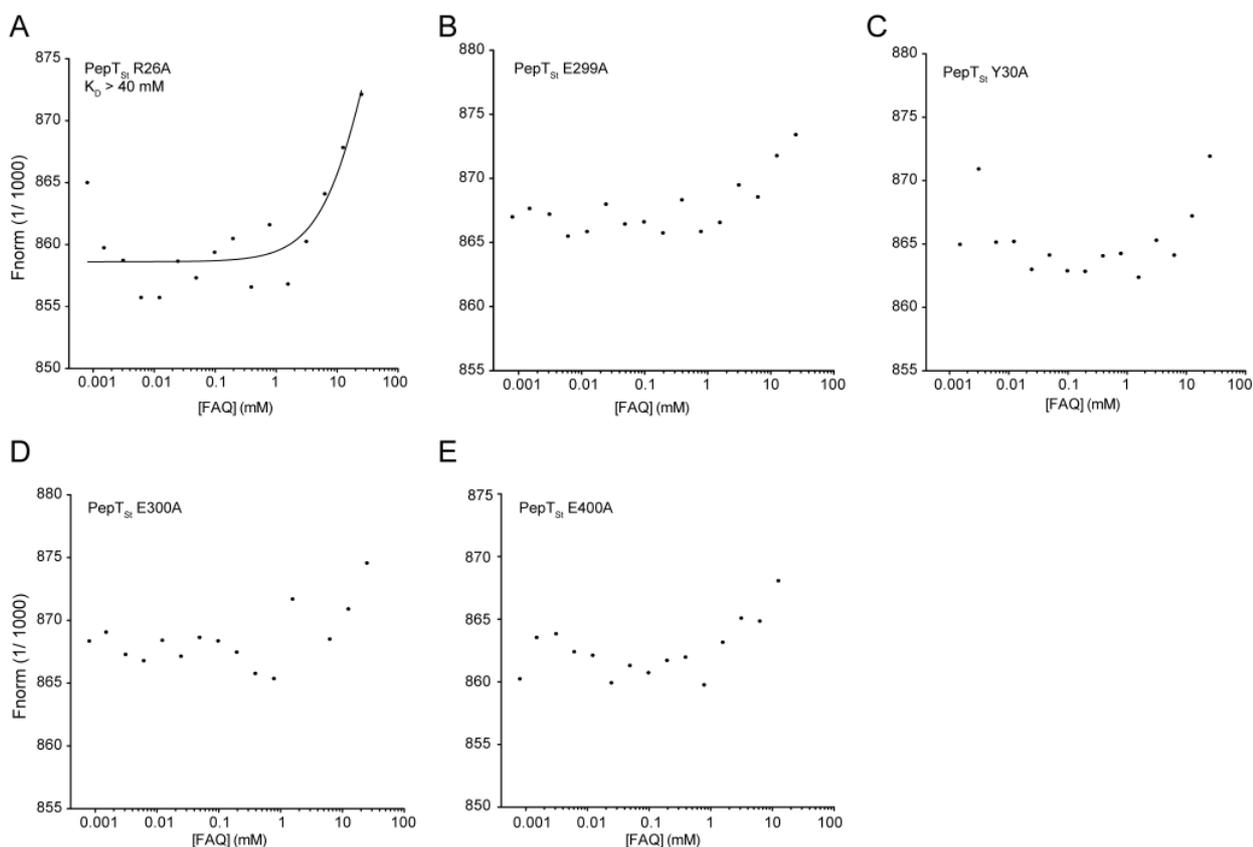
Figure 37: Phe-Ala-Gln binding to PepT$_{St}$ single mutants. (A) MST curve for Phe-Ala-Gln binding to detergent-solubilized PepT$_{St}$ R26A, (B) PepT$_{St}$ E299A, (C) PepT$_{St}$ Y30A, (D) PepT$_{St}$ E300A and (E) PepT$_{St}$ E400A. Mutant PepT$_{St}$ R26A presents a K_D value > 40 mM, whereas it was not possible to estimate a K_D value for mutants PepT$_{St}$ E299A, Y30A, E300A and E400A.

4.1.4 Role of Tyr-68 in ligand binding

As mentioned in the ligand binding section, Tyr-68 is a highly conserved residue in the binding cavity among POTs (see Figure 28 for detail information of the conservation of the residues in the binding site). Furthermore, the mutation of the equivalent residue in rabbit PepT1 (Tyr-64) into alanine impairs peptide transport (Chen et al., 2000).
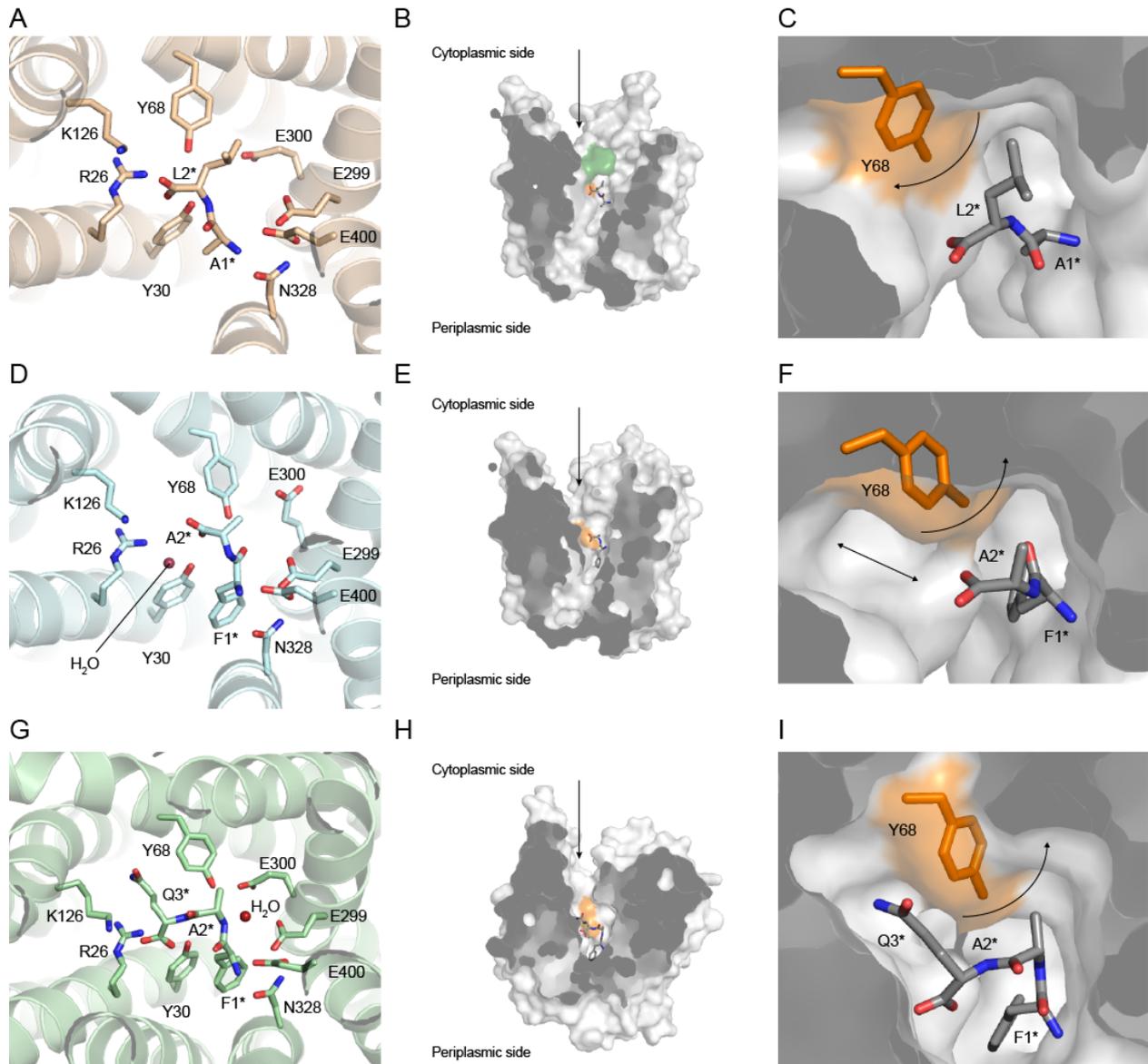
Figure 38: Role of Tyr-68 in di- and tripeptide binding in PepT$_{St}$. (A) Binding site of PepT$_{St}$ in complex with Ala-Leu (PepT$_{St}$[Ala-Leu]) (PDB ID 5OXL), which represents dipeptide binding mode 1. (B) Surface representation of PepT$_{St}$[Ala-Leu]. The protein is colored white except for Tyr-68, which is orange, and for Phe-428 and Trp-427, which are green. The latter two residues partially restrict access to the binding site from the cytoplasmic side in this structure. The grey sticks represent the Ala-Leu dipeptide, and the arrow indicates the position from which the figure in panel C was generated. (C) Detailed position of Tyr-68 in the binding cavity when Ala-Leu is bound. The side chain of Tyr-68 is represented by sticks and labeled. The arrow indicates that the residue is in a position that widens one pocket (pocket 2), while restricting access to a neighboring one (pocket 3). (D) Binding site of PepT$_{St}$[Phe-Ala] (PDB ID 5OXN), representing dipeptide binding mode 2. (E) Surface representation of PepT$_{St}$[Phe-Ala]. (F) Detailed position of Tyr-68 in the binding cavity when Phe-Ala is bound. The arrows indicate that Tyr-68 is in a position that narrows pocket 2, while allowing access to pocket 3. (G) Binding site of PepT$_{St}$[Phe-Ala-Gln]. (H) Surface representation of PepT$_{St}$[Phe-Ala-Gln]. (I) Detailed position of Tyr-68 in the binding cavity when Phe-Ala-Gln is bound. The position is similar to the one it adopts in PepT$_{St}$[Phe-Ala]. There is therefore access to pocket 3. Indeed, the C-terminal glutamine residue of the tripeptide is here occupying this pocket.

Tyr-68 has an important role tuning the size of P2 in dipeptide binding, and P2-P3 in tripeptide binding. As shown in the LigPlot$^+$ panels over the results section, Tyr-68 is mostly forming hydrophobic interactions with the bound peptides, particularly to the residue at the C-terminus of dipeptides or in the middle position of tripeptides, and it adopts a range of different positions in the described dipeptide complexes (see Figure 27A). Furthermore, its lateral displacement mostly depends on the binding mode the peptide adopts, and the size of the residue at the C-terminus for dipeptides or in the middle position for tripeptides. For dipeptides such as Ala-Leu, Ala-Gln and Asp-Glu (binding mode 1), as they have a bulkier residue in the C-terminus, Tyr-68 moves in a way that P2 is expanded to leave space enough to accommodate the side chain (Figure 38, panels A-C). On the contrary, as in the case of the complex with Phe-Ala (binding mode 2), the residue at the C-terminus of the peptide is rather small. Therefore, Tyr-68 reduces the size of P2, which at the same time generates additional space at the C-terminus of the peptide (Figure 38, panels D-F). Finally, for tripeptides with sequence Phe-Ala-Xxx, there are two factors determining the position of Tyr-68: the first is the small size of the alanine residue in the middle position of the peptide. Therefore, P2 has to be modulated to adopt a rather small overall size, just as in the Phe-Ala complex. Secondly, the residue at the C-terminus of the tripeptide expands towards the additional space generated, pushing Try-68 in a way that the newly form pocket, P3, surrounds the C-terminus of the tripeptide (Figure 38, panels G-I).

4.1.5 Ligand promiscuity of PepT$_{St}$

Up to this point of the results section, only the complexes with either di- or tripeptides have been discussed. However, in the described structures of PepT$_{St}$ in complex with peptides and in the apostructure, we observed additional molecules in the binding cavity. As an example, in the described apostructure, several water molecules could be confidently modeled in the binding cavity. In this section, the presence of water molecules, phosphate ions and other peptide-mimetic molecules will be discussed as well as their possible biological roles. Finally, we will conclude the section explaining the general mechanisms POTs have to work as promiscuous transporters.
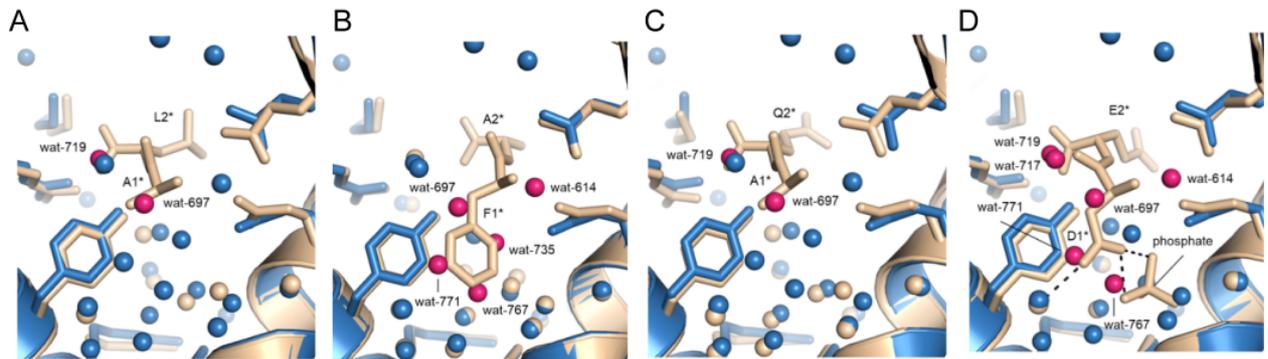
Figure 39: Displacement of water molecules in PepT_st apostructure upon ligand binding. (A) Comparison of water structure in PepT_st[apo] (blue) and PepT_st[Ala-Leu] (wheat). Water molecules in PepT_st[apo], which would clash with the peptide (less than 2 Å away from the peptide), as it is bound in PepT_st[Ala-Leu], are highlighted in red and correspondingly labeled. (B) Comparison of water structure in PepT_st[apo] and PepT_st[Phe-Ala]. (C) Comparison of water structure in PepT_st[apo] and PepT_st[Ala-Gln]. (D) Comparison of water structure in PepT_st[apo] and PepT_st[Asp-Glu]. Figures form Martinez Molledo et al., 2018.
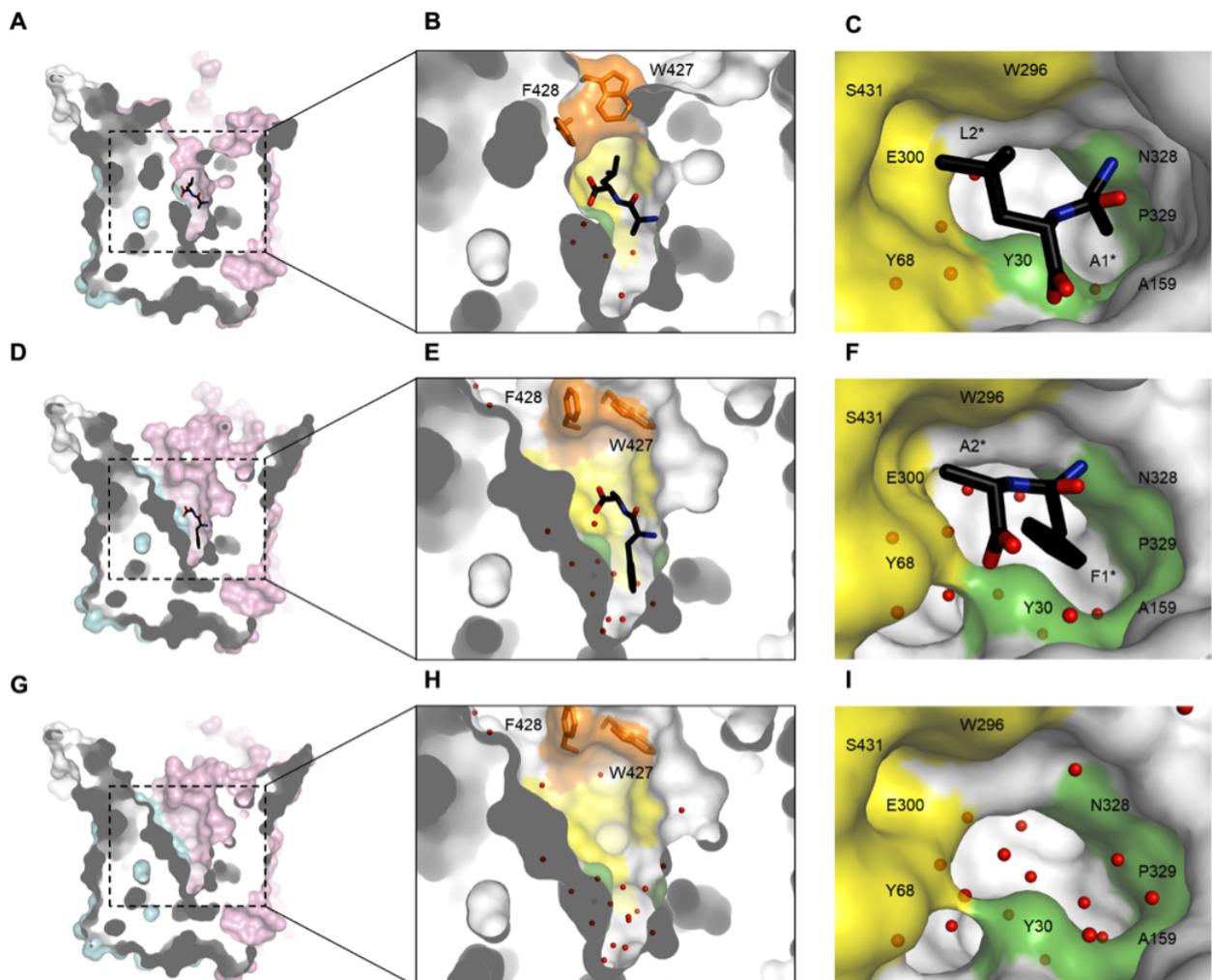
Figure 40: Surface views of the peptide binding site. (A) Side view of the binding site of PepT_st[Ala-Leu]. The protein is shown from the side in semitransparent surface representation and a part has been cut away to reveal the binding site in the middle of the protein. The N-domain is light blue, the C-domain is pink and the peptide is black. (B) Zoomed side view of the binding site of PepT_st[Ala-Leu]. Same orientation as in panel A, but zoomed in on the peptide

and colored differently – P1 is green, P2 is yellow and P2-lid is orange. In addition, the water molecules of the binding cavity are shown as small red spheres, and the two aromatic residues of P2-lid are shown, not only in semitransparent surface representation, but also as sticks. (C) Cytoplasmic view of the binding site of PepT$_{St}$[Ala-Leu]. Shown as in panel B, but in different orientation and zoomed in further on the peptide. In addition, P2-lid was omitted to allow an unobscured view of the peptide. (D-F) same as panels A-C, but for PepT$_{St}$[Phe-Ala]. (G-I) same as panels A-C, but for of PepT$_{St}$[apo]. The PEG molecule was omitted for clarity. Figures form Martinez Molledo et al., 2018.

4.1.5.1 Water molecule network

Due to the high data quality of the PepT$_{St}$ structures obtained from crystals grown by the LCP method, a numerous ordered water molecules could be modeled in the binding cavity not only in the apostructure of PepT$_{St}$ (see Figure 22) but also in the PepT$_{St}$ structures in complex with di- and tripeptides (Figures 39 and 40). In the apostructure, the water molecules are mostly located at the apex of the binding cavity, some distance away from where the peptide binds, in the center of the P1 (Figure 22 and Figure 40, panels G-I) but in contrast, there are not as many water molecules in P2. In the substrate-bound structures, it was not possible to model as many water molecules mostly due to the lower resolution of these models, but also because the position of some of the water molecules found in the binding cavity of PepT$_{St}$[apo] is incompatible with the presence of a peptide (Figure 39 and Figure 40, panels A-F). This suggests that solvation of the binding site might facilitate substrate release, as has recently been suggested for the amino acid-polyamine-organocation (APC) transporter AdiC (Ilgu et al., 2016). Furthermore, the displacement of certain water molecules of the apostructure compare to the position in the structures in complex with peptides strongly depends on the sequence of the peptide, mostly on the nature of the N-terminal residue occupying P1. Thus, while some water molecules are displaced by the peptide, others may instead facilitate its binding in a likewise sequence dependent manner. For example, we have seen that the aspartate residue of the Asp-Glu peptide forms a hydrogen bond with one of the water molecules in P1, which is not formed in any of the other structures (Figure 40D). A general way of achieving multispecific binding in P1 may thus be to displace or retain specific water molecules, in order to optimally match the size of P1 to the physicochemical properties of the incoming peptide side chain.
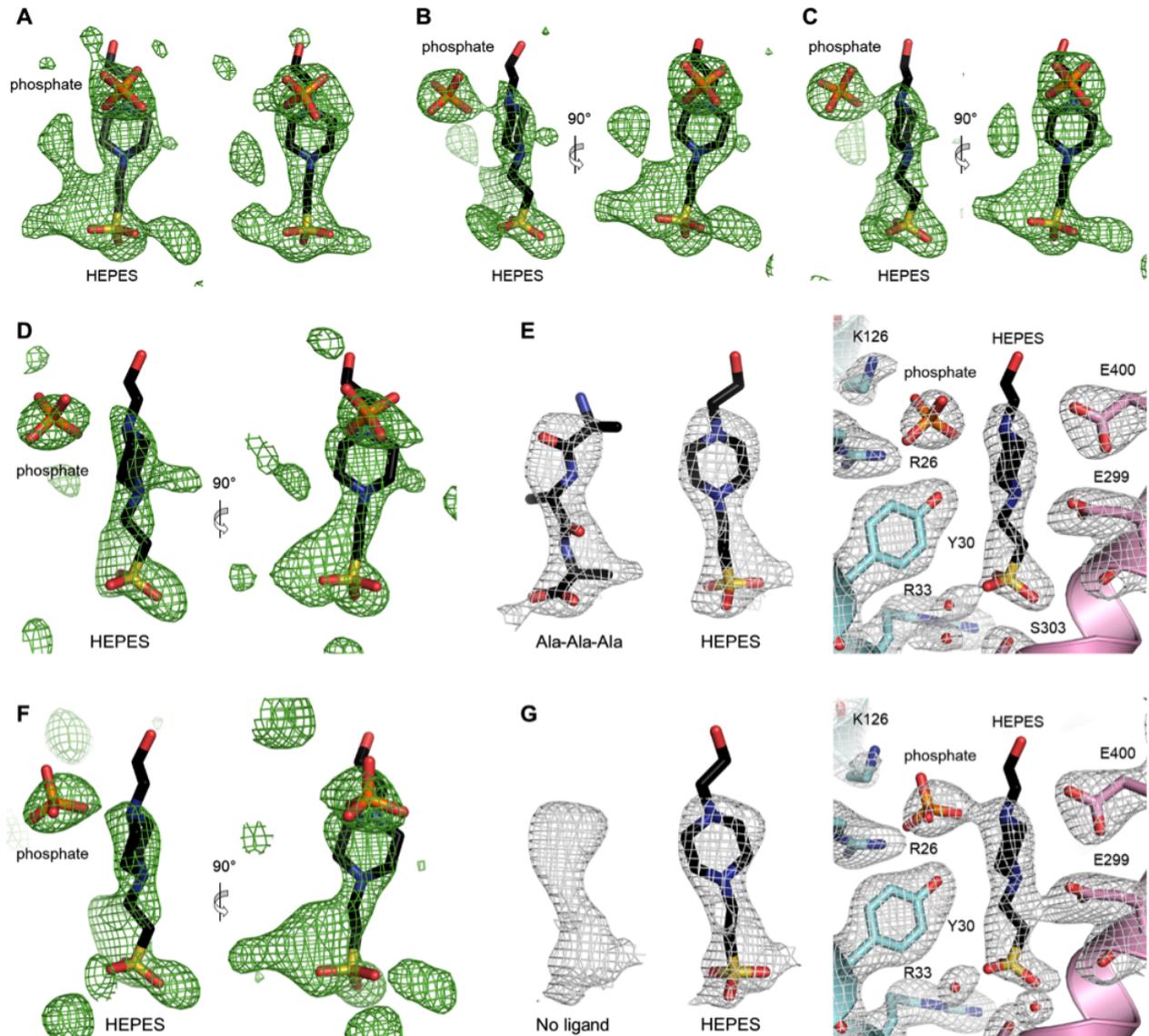
Figure 41: Additional electron density maps for HEPES and phosphate. (A) The 3-σ 1Fo-Fc omit maps for HEPES and phosphate are shown for PepT$_{St}$[100 mM HEPES] (left) and PepT$_{St}$[300 mM HEPES] (right). In the case of PepT$_{St}$[100 mM HEPES], HEPES fits fairly well in the electron density. The additional observed density could be due to alternative conformations or to competition with other molecules that may be present with low occupancy in overlapping positions. This structure comes from crystals grown in the presence of peptide Ala-Tyr. In the case of PepT$_{St}$[300 mM HEPES], much of this extra difference density has disappeared, which aligns well with the latter hypothesis. (B) 3-σ 1Fo-Fc omit map of HEPES and phosphate for PepT$_{St}$ co-crystallized with Val-Tyr-Val and (C) Ala-Ala-Ala. (D) Putative binding of HEPES and phosphate in re-refined PDB ID: 4D2D. This structure was originally refined with a vertically bound Ala-Ala-Ala peptide, but was here re-refined with HEPES and phosphate. The 3-σ 1Fo-Fc omit map is shown. (E) 1-σ 2Fo-Fc maps for re-refined PDB ID: 4D2D with Ala-Ala-Ala left in place (left), and with Ala-Ala-Ala replaced with HEPES and phosphate (middle and right). (F) Putative binding of HEPES and phosphate in re-refined PDB ID: 4D2B. This structure was originally refined as ligand-free, but was here re-refined with HEPES and phosphate. The 3-σ 1Fo-Fc omit map is shown. (G) 1-σ 2Fo-Fc maps for re-refined PDB ID: 4D2B with no ligands present (left), and with HEPES and phosphate added (middle and right). Figures form Martinez Molledo et al., 2018.
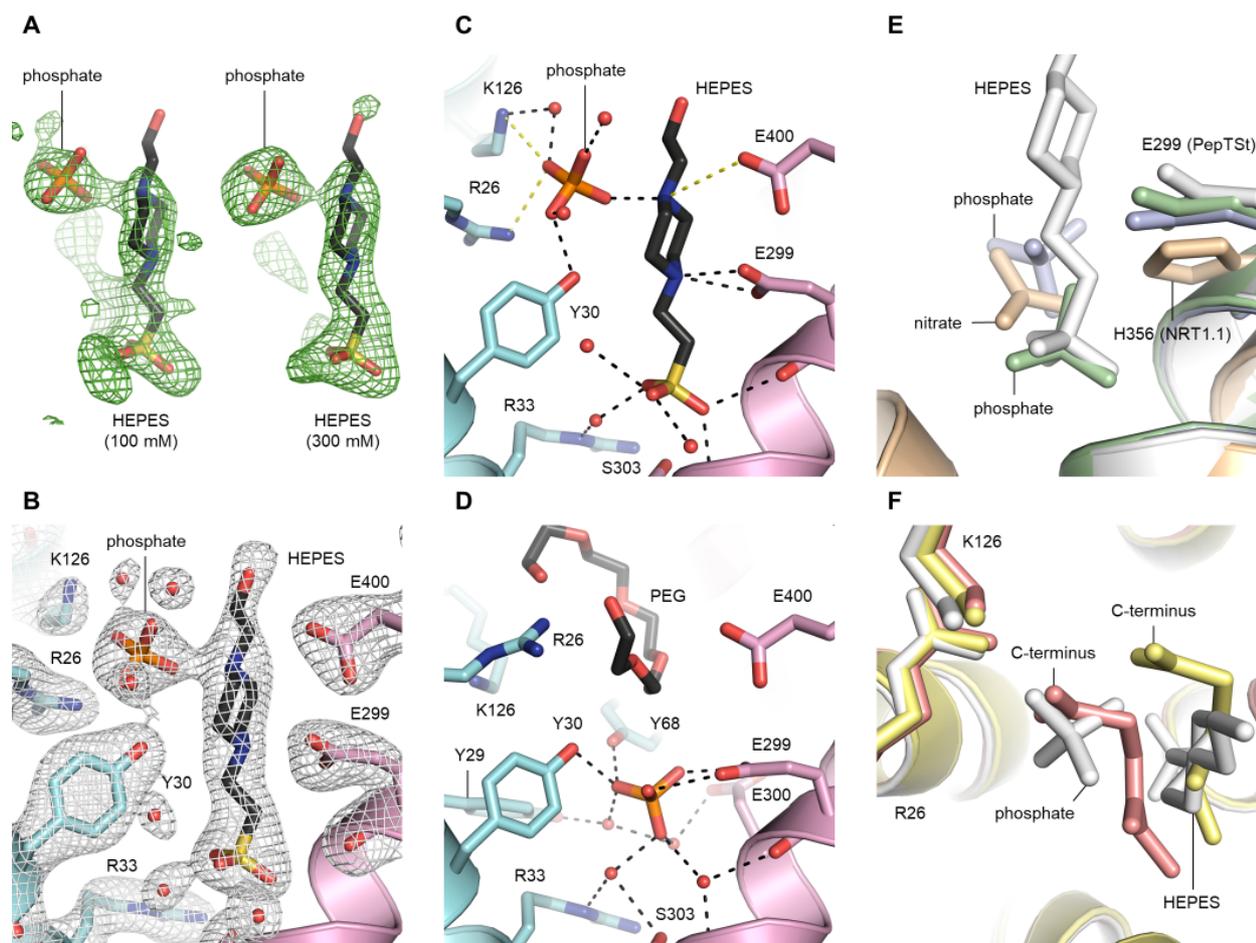
Figure 42: Binding of non-peptidic molecules. (A) Evidence for binding of HEPES and phosphate. The modeled HEPES and phosphate molecules in PepT$_{St}$[100 mM HEPES] (left) and PepT$_{St}$[300 mM HEPES] (right) are shown together with their Fo-Fc difference density omit maps contoured at 3 σ. (B) Refined electron density map for HEPES and the co-bound phosphate molecule. The 2Fo-Fc map for PepT$_{St}$[HEPES 300 mM] is contoured at 1 σ. (C) Interaction mode of HEPES and the co-bound phosphate molecule. HEPES, phosphate and interacting residues are shown in sticks, and water molecules involved in binding are shown as small red spheres. Black dashes indicate potential hydrogen bonds or salt bridges with lengths ≤ 3.2 Å, and yellow dashes indicate potential salt bridges with lengths of 3.2–4.0 Å. (D) Interaction mode of phosphate in PepT$_{St}$[phosphate]. (E) Upper acid/anion binding region. An overlay is shown of PepT$_{St}$[HEPES 300 mM] (white), PepT$_{St}$[phosphate] (pale violet), PepT$_{St}$[Asp-Glu] (pale green), and AtNRT1.1 (PDB ID: 4OH3 (Sun et al., 2014)) (wheat). (F) Lower acid/anion binding region. An overlay is shown of PepT$_{St}$[HEPES 300 mM] (white), PepT$_{St}$[Ala-Leu] (salmon) and PepT$_{St}$[Phe-Ala] (pale yellow). Side chains of Ala-Leu and Phe-Ala have been omitted for clarity purposes. Figures form Martinez Molledo et al., 2018.

4.1.5.2 Model of the vertically bound HEPES molecule

When co-crystallizing PepT$_{St}$ with substrates, we could still observe in some cases an elongated blob of positive difference density flanked by a smaller roundish one in the electron density maps (Figure 41), which is not consistent with a bound peptide in a horizontal manner between the two MFS domains similarly to all the previously described complexes with peptides. We initially thought that

the elongated vertical blob would represent the previously proposed vertical tripeptide binding mode (Lyons et al., 2014). However, the density looked essentially the same even when very different substrates in terms of compositions and dimensions were used for co-crystallization. As an example, very similar omit maps (1Fo-Fc) at a sigma level of three (3-σ) were obtained for data sets collected from crystals grown in the presence of Ala-Tyr, Val-Tyr-Val and Ala-Ala-Ala (Figure 41, panels A to C, respectively). In all cases, a HEPES molecule together with a phosphate ion could be modeled and refined. Furthermore, to ensure the certainty of this assumption, the concentration of the HEPES was increased to 300 mM in the crystallization condition, obtaining a considerably improved density in the binding pocket (Figure 41A and Figure 42, panels A and B). An additional experimental prove that a HEPES molecule occupies the binding site is that, whenever the buffer is replaced by citrate (such in the case of the apostructure) or phosphate buffer instead (as for the occluded conformation that will be later discussed), the vertical difference density was absent.

A structure was therefore refined with the HEPES molecule modeled in the vertical difference density blob and phosphate in the flanking position (see table Table 35 for the data collection and statistics in the appendix I), at two different concentrations of HEPES, 100 mM and 300 mM. According to our model, the HEPES molecule is bound as follows: the sulfonate group is located at the apex of the binding cavity where it interacts with several water molecules and the backbone of the kinked TM7 helix (Figure 42C), while the two nitrogen atoms of the HEPES ring interact with Glu-299 and potentially Glu-400, which are also used for binding the N-terminus of dipeptide substrates. The flanking phosphate molecule is paired with HEPES through an interaction with the nitrogen atom farthest away from the sulfonate group. Apart from that, it also interacts with Arg-26 and Lys-126 similarly to the C-terminus of Ala-Leu, as well as with several water molecules. It is thus clear that the HEPES-phosphate pair mimics some aspects of how peptides are bound in spite of not having a peptide-like composition.

Finally, we found that HEPES could also be modeled in two previously published structures that were obtained using the same crystallization conditions as employed in this study: PDB ID: 4D2D where Ala-Ala-Ala had originally been modeled, and PDB ID: 4D2B where no ligand had been modeled (Figure 41, panels D-G). This finding suggests that the previously proposed vertical binding mode for Ala-Ala-Ala should be reevaluated.

4.1.5.3 Ion binding sites in PepT$_{St}$

When crystallizing PepT$_{St}$ using a phosphate buffer, the binding site was occupied by a phosphate and a PEG molecule (this structure is discussed in more detail in section 4.2.2). Interestingly, the

phosphate molecule does not bind in the same place as in the structure with HEPES and phosphate, but is instead localized in the water-filled apex of the binding cavity (Figure 42D). Here it forms a double hydrogen bond with Glu-299, which must thus be protonated, and additional hydrogen bonds with Tyr-30 and Tyr-68 as well as several water molecules (Figure 42D).

It seems that phosphate ions can bind in two distinct positions: We have termed the position that is close to the apex of the cavity and the extracellular side, the 'upper region' (Figure 42E), and the one that accommodates the C-terminus of dipeptides, the 'lower region' (Figure 42F). In the upper region, the HEPES sulfonate group overlaps strongly with the phosphate molecule modeled in PepT$_{St}$[Asp-Glu], but only partially with the one found in PepT$_{St}$[phosphate] (Figure 42E). In the lower region, we find a strong overlap of the phosphate paired with HEPES and the C-terminus of Ala-Leu. The HEPES ring overlaps however not with Ala-Leu but with the backbone of the phenylalanine of Phe-Ala (Figure 42F). Thus, POTs are closely related to NRT1 nitrate transporters (Tsay et al., 2007). It is therefore intriguing that the nitrate binding site of AtNRT1.1 (Parker and Newstead, 2014; Sun et al., 2014) coincides with the upper acid/anion binding region identified here for PepT$_{St}$ (Figure 42E). In fact, the key nitrate binding residue in AtNRT1.1, His-356, is directly equivalent to the key phosphate binding residue in PepT$_{St}$, Glu-299 (Figure 42E). This prompts the question of whether binding of acids/anions in the upper region might also be biologically relevant in PepT$_{St}$ and/or other POTs.

4.1.6 Discussion

It was previously postulated that POTs are able to recognize a broad range of ligands due to the multiple hydrogen bonds and salt bridges they form mainly with the peptide backbone rather than the different side chains of the peptide substrate (Guettou et al., 2013). From our studies on ligand binding in solution as well as in the reported crystal structures, there is strong evidence that the sequence variation of the ligand directly influences its binding to the transporter. Thus, for PepT$_{St}$, the side chains of the ligand can be accommodated in the binding pockets P1 and P2 through different mechanisms: (1) the interaction between the side chains of the ligand and the residues constituting the binding pockets influence the position of the ligand backbone. Therefore, two different binding modes were observed for the dipeptides, the one represented by complexes with Ala-Leu, Ala-Gln and Asp-Glu, and the second one represented by the complex with Phe-Ala. (2) Furthermore, the solvation of the binding cavity changes upon ligand binding, particularly in P2. Thus, even though some water molecules are not compatible with the presence of the ligand in the binding site, other might even facilitate its coordination. (3) Try-68 is a highly conserved residue throughout the POT family. In our structures in complex with peptides, this residue fine-tunes the

size of P2, adapting its dimension to the residue at the C-terminus of dipeptides or in the middle position of tripeptides. Furthermore, the mutation of the corresponding residue in rabbit PepT1 to alanine (Tyr-64) impairs transport (Chen et al., 2000). Therefore, Tyr-68 is clearly functionally important. (4) Finally, it is remarkable the number of aromatic residues present in the binding cavity of POTs, particularly in P2 and P2-lid in the case of PepT$_{St}$. Aromatic residues can establish multiple types of interactions, such as hydrogen bonds, π stacking, arene-arene interactions, arene-amide …, which is perfect for the recognition of different ligands.

Additionally, it is important to highlight the role of residue Glu-300; even though it does not interact directly with the ligand, our mutational studies support that its change to an alanine residue affects ligand binding. It might have a role in charge balance and/or proton coupling during a transport event. A role in proton coupling has thus been found for the equivalent Glu-310 residue in GkPOT (Doki et al., 2013; Immadisetty et al., 2017).

Our data also show that the previously published vertical binding mode proposed for tripeptides should be revised (Lyons et al., 2014), as it might have been the result of the misinterpretation of the electron density map. Alternatively, we have reported the structure of PepT$_{St}$ in complex with Phe-Ala-Gln. Here, the tripeptide is located horizontally across the binding cavity between the two MSF domains, which is thus in agreement with the previously reported PepT$_{So2}$ structures in complex with tripeptides Ala-Ala-Ala and Ala-Tyr(Br)-Ala (Guettou et al., 2014). We have shown that tripeptides of sequence Phe-Ala-Xxx bind to PepT$_{St}$ with dissociation constants comparable to the ones reported for most of the studied dipeptides. Whether tripeptides of sequence Phe-Ala-Xxx are the only tripeptides that can be transported by PepT$_{St}$ should be carefully considered. The current data point out that the presence of an aromatic residue at the N-terminus of the ligand favors tripeptide binding by positioning the peptide backbone in a comparable manner to the dipeptide Phe-Ala (binding mode 2). Furthermore, a small residue in the middle position of the tripeptide might facilitate the torsion of the peptide backbone and the displacement of Tyr-68, enlarging P3 and making it accessible to a third residue at the C-terminus of the ligand.

4.2 Transport cycle in PepT$_{St}$

POTs change their conformation in order to move substrates across the membrane: they adopt an outward-open conformation when their binding cavity is accessible from the extracellular environment (or the periplasm, in Gram negative bacteria). Once the ligand is bound, they cycle from the outward-open state, through an occluded conformation, to the inward-open state. There, the ligand is released to the cytoplasm of the cell. Later, the transporter cycles back and exposes its ligand binding cavity to the exterior of the cell again, so that a new substrate can be transported.

As summarized in Table 2, among MFS there are structures reported form transporters in the three hitherto mentioned conformations; additionally, there are also intermediate states, with partial occlusions due to the bending of certain TM helices. Quite the opposite is the case for POTs (see section 1.4.2 in the introduction): only inward-open conformations or inward-open partially occluded states have been reported. For PepT$_{St}$, our studies complemented the already available conformations, contributing to get new insights into partially occluded inward-open states and also a somehow different occluded conformation. Even though one of our initial aims was to determine the structure of PepT$_{St}$ in an outward open conformation, here we can only show some preliminary results that would require further optimization to fulfill the objective.

In this section, the different conformations that were identified for PepT$_{St}$ will be discussed as well as the recent advances to determine the outward-open state.

PDB ID	Conformation	Ligand	Publication
4APS	Inward-open		Solcan et al., 2012
4D2C	Inward-open partially occluded	Ala-Phe	Lyons et al., 2014
4D2D	Inward-open	Ala-Ala-Ala	Lyons et al., 2014
4D2B	Inward-open		Lyons et al., 2014
4XNJ	Inward-open		Huang et al., 2015
4XNI	Inward-open		Huang et al., 2015
5MMT	Inward-open		Quistgaard et al., 2017
5D58	Inward-open partially occluded	Ala-Phe	Huang et al., 2016
5D59	Inward-open partially occluded	Ala-Phe	Huang et al., 2016
5D6K	Inward-open partially occluded		Ma et al., 2017
5OXL	Inward-open partially occluded	Ala-Leu	Martinez Molledo et al., 2018
5OXK	Inward-open partially occluded	Ala-Gln	Martinez Molledo et al., 2018
5OXM	Inward-open	Asp-Glu	Martinez Molledo et al., 2018
5OXN	Inward-open	Phe-Ala	Martinez Molledo et al., 2018
-	Inward-open	Phe-Ala-Ala	Martinez Molledo et al., submitted

-	Inward-open	Phe-Ala-Thr	Martinez Molledo et al., submitted
6GHJ	Inward-open	Phe-Ala-Gln	Martinez Molledo et al., submitted
6EIA	Inward-open	HEPES	Martinez Molledo et al., 2018
5OXQ	Inward-open	HEPES	Martinez Molledo et al., 2018
5OXP	Occluded	Phosphate	Martinez Molledo et al., 2018
5OXO	Inward-open		Martinez Molledo et al., 2018

Table 6: Reported structures for PepT$_{St}$. For each reported conformation, the following information is provided: PDB ID, crystallized conformation for PepT$_{St}$, bound ligand (if any) and corresponding publication.

4.2.1 Inward-open and inward-open partially occluded complex structures

Previous to the publication of our results (Martinez Molledo et al., 2018), there were several structures reported of PepT$_{St}$ either in inward-open conformation or inward-open partially occluded (table 10), but the insights explaining the restricted access to the binding cavity in the partially occluded conformations were not yet established for PepT$_{St}$.

Among the structures of PepT$_{St}$ in complex with either di- or tripeptides, we observed that TM helix 11 is bended towards the binding cavity for dipeptides Ala-Phe (Lyons et al., 2014), Ala-Leu and Ala-Gln (Martinez Molledo et al., 2018), whereas in complexes with Phe-Ala, Asp-Glu and tripeptide Phe-Ala-Ala, Phe-Ala-Thr and Phe-Ala-Gln, PepT$_{St}$ is in fully inward-open state (see Figure 38 and Figure 40 for the different conformations, and Figure 43).
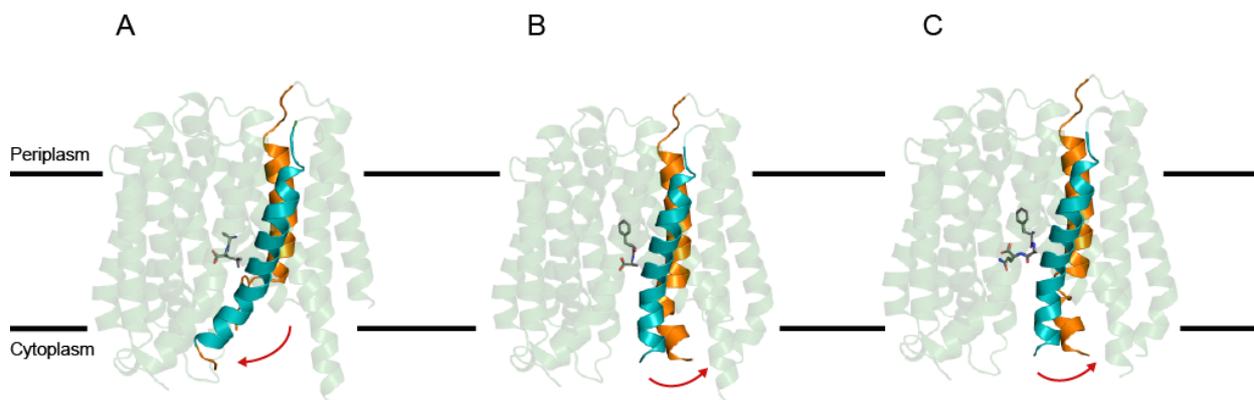
Figure 43: Inward-open partially occluded and inward-open conformations for PepT$_{St}$. (A) PepT$_{St}$[Ala-Leu], PepT$_{St}$[Ala-Gln] and PepT$_{St}$[Ala-Phe] structures are in inward-open partially occluded state. Here PepT$_{St}$[Ala-Leu] illustrates such conformation. PepT$_{St}$ is represented in green, where TM10 is highlighted in orange and TM11 is in cyan. The red arrow underneath indicates the direction of the bending of these helices. In the binding pocket, ligand Ala-Leu is represented in grey sticks. HA and HB helices have been removed for clarity of the figure. The horizontal lines indicate the approximate position of the membrane relative to the transporter. (B) Same representation for structure PepT$_{St}$[Phe-Ala]. Here the transporter is in inward-open conformation. Helices TM10 and TM11 are not bended as in structure PepT$_{St}$[Ala-Leu]. (C) Inward-open structure PepT$_{St}$[Phe-Ala-Gln].

For the occluded conformations, two residues in TM helix 11, Trp-427 and Phe-428, form what we have defined as P2-lid. We proposed that the residues forming P2-lid interact with ligands with bulky residues at the C-terminus position: in the Ala-Leu complex, the leucine at C-terminus forms hydrophobic Van der Waals interactions with Trp-427 and Phe-428 from the P2-lid as well as other aromatic residues in P2 (see figure 27B and LigPlot$^+$ representation in figure 29A). In the complex with Ala-Gln, the glutamine at C-terminus interacts by a N-H … π bond to Trp-296, and by hydrophobic interactions also to Tyr-68 and Trp-296 (see figure 27D and LigPlot$^+$ representation in figure 29C). As a result, the P2 presents a size adapted to the side chains of the ligand and thus, these residues are large enough to interact with the P2-lid.

For the structures with ligands with a smaller residue at C-terminus, such as Phe-Ala, the alanine is too small to interact with the residues of the P2-lid but still it interacts with Tyr-68 to reduce the size of P2. For complex with Asp-Glu, no certain conclusions can be accomplished: even though the glutamante size chain has a considerable size, the electron density is rather weak for it in the 2Fo-Fc map. This might be due to the low ligand occupancy or the flexibility of the side chain.

In the structures in complex with tripeptides Phe-Ala-Xxx, the C-terminus position is equivalent to the one adopted by dipeptides binding by the mode 1 (Ala-Leu, Asp-Glu and Ala-Gln), due to the bending of the backbone of the ligand. The side chain of the residue at the C-terminus of the ligand expands to the P3. Therefore, the alanine in the middle position of the tripeptide is the one in close vicinity of the P2-lid but it is also too small to interact with the residues of P2-lid.

From these observations we conclude that (1) the different ligands might stabilize a different conformation of the transporter based on the interactions they establish with the residues in the P2-lid. (2) The transporter cycles through a series of conformations when moving from the outward-open conformation to the inward-open, so these partially occluded conformations might represent some intermediate states.
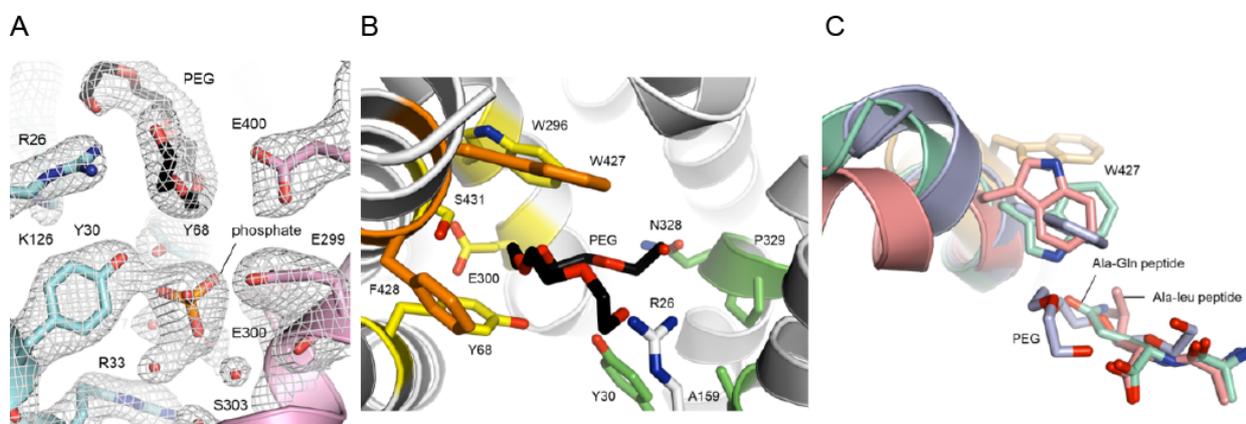
Figure 44: PepT$_{St}$ occluded conformation. (A) 1-σ 2Fo-Fc electron density map for the binding cavity. The PEG molecule and the phosphate ion are clearly well defined in the map. (B) Position of the PEG molecule relative to the P1 (green), P2 (yellow) and P2-lid (orange) residues. The PEG molecule inserts deeply into P2/P2-lid where it packs against

the faces of the aromatic rings of Tyr-68, Trp-296, Trp-427 and Phe- 428, thus resulting in the formation of numerous van der Waals interactions and several C- H $\cdots \pi$ interactions. (C) Overlay of inward-open PepT$_{St}$[apo] (light orange) and all inward facing occluded structures PepT$_{St}$[phosphate] (pale violet), PepT$_{St}$[Ala-Leu] (salmon) and PepT$_{St}$[Ala-Gln] (mint). Note that the various substrates all interact differently with Trp-427, which correlates with differences in the bending of TM11. Figures form Martinez Molledo et al., 2018.
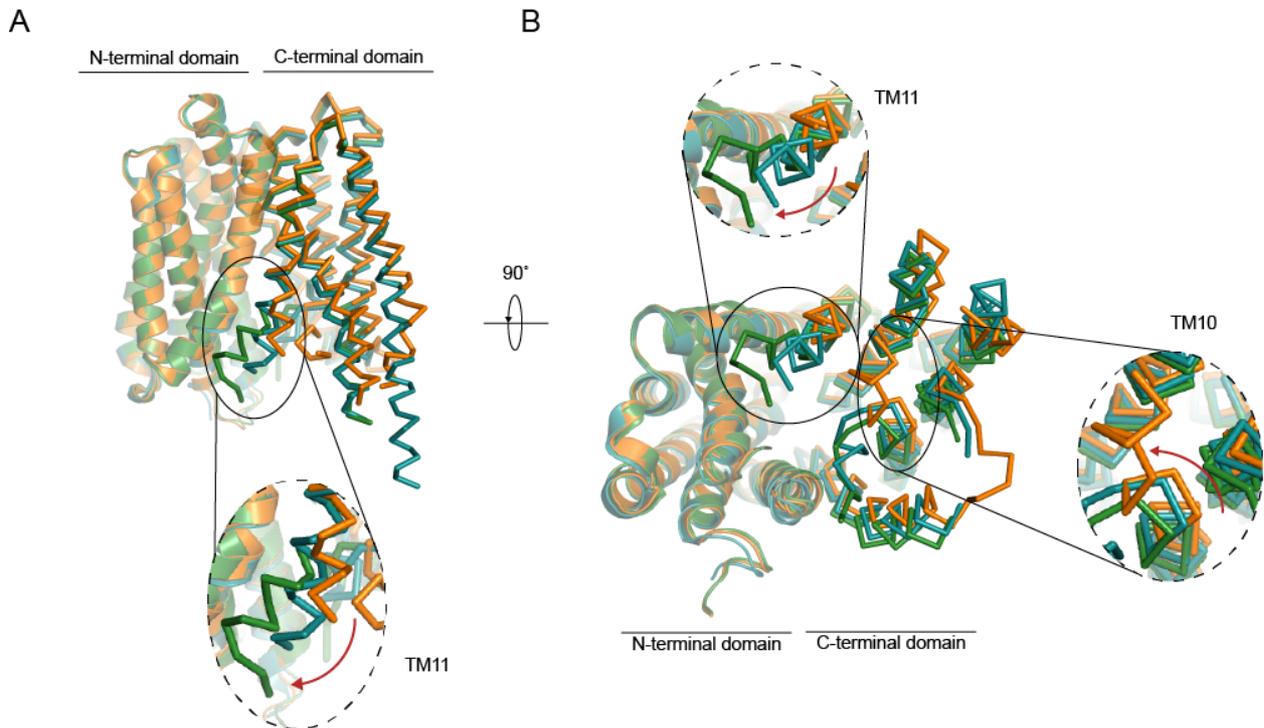
Figure 45: Bending of TM10 and 11. (A) Alignment of the N-terminal bundle in structures PepT$_{St}$[apo] (orange), PepT$_{St}$[Ala-Leu] (cyan) and PepT$_{St}$[phosphate] (green) to show the main rearrangements at the C-terminal bundle upon partial occlusion of the transporter. The N-terminal domain is represented as cartoon, whereas in the C-terminal domain only the trace of the Cα atoms is shown. The HA and HB helices have been removed for clarity. In a detail, the bending of TM11 in the different structures is shown. (B) The same overlay from panel A but viewed from the cytosolic side of the membrane. Here, in the details the differences in TM10 and 11 are highlighted.

4.2.2 Occluded conformation

The substitution of HEPES buffer by NaP$_i$ buffer in crystallization yielded a slightly different conformation of the transporter. In this case, a structure was refined to 2.4 Å maximum resolution (see Table 36 in appendix I for the data processing and refinement statistics), and in the binding cavity two molecules could be built in the electron density map: a PEG molecule, which interacts with Arg-26 and the P2/P2-lid, and a phosphate ion, located in the so-called lower ion binding site (Figure 44). As the PEG molecule interacts with the P2-lid, the transporter is in occluded state but different from the reported one when the ligands are bound. Even though TM11 is bended, it is not

as drastic as in the ligand-bound forms. The difference resides in the way Trp-427 interacts with the ligand or the PEG molecule (Figure 44C and Figure 45). This structure together with the inward-open partially occluded states might represent intermediate conformations between the inward-open and the not yet defined outward-open conformation.
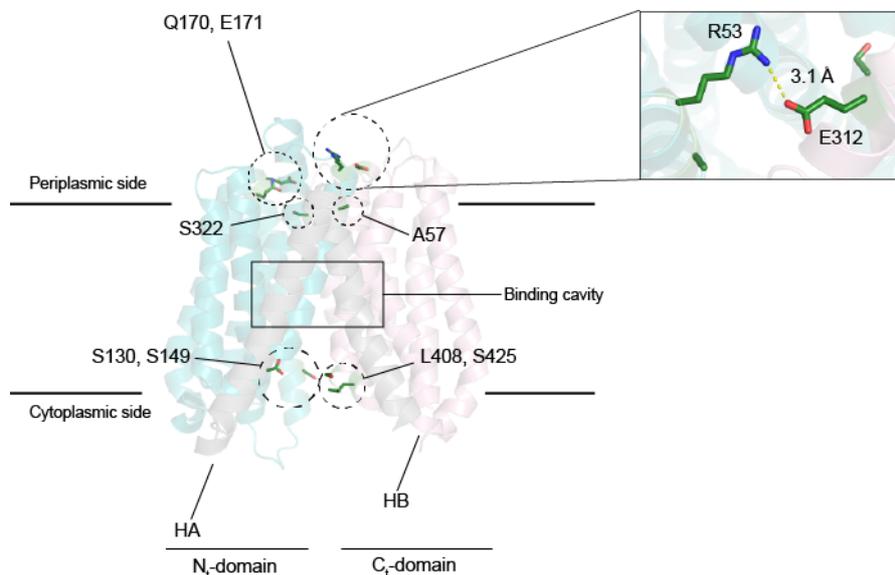
Figure 46: Location of the mutated residues in PepT$_{St}$. The residues represented in sticks were mutated in the attempt to stabilize the outward-open state. In the zoom view, residues R53 and E312 are highlighted; they constitute a periplasmic gate, closing the access to the binding cavity from the periplasmic side of the membrane.

4.2.3 Efforts to determine the structure of PepT$_{St}$ in an outward-open conformation

Among the structures we determined for PepT$_{St}$ as well as the already available ones, there are only inward-open, inward-open partially occluded and occluded conformations. Therefore, there is a lack of information about the structure of PepT$_{St}$ in the outward-open state. On these regards, there are studies based on molecular dynamics predicting the overall structure of other bacterial POTs, such as the peptide transporter of *Shewanella oneidensis* PepT$_{So}$ (Fowler et al., 2015). The three crystallization conditions we reported (HEPES buffer for ligand complexes, the NaP$_i$ of the occluded conformation and the citrate one for the apostructure) favored the inward-open conformation of PepT$_{St}$. To promote the stability of the outward-open conformation in PepT$_{St,}$ we followed three approaches: (1) Introduction of mutations in the periplasmic gates, (2) substitution of residues in the periplasmic side of the transporter by bulky residues, and (3) arrangement of a disulphide bond in the cytoplasmic side of the transporter. Furthermore, several of these mutations were combined. In Figure 46, the residues that were mutated with different purposes are mapped within PepT$_{St}$ structure.
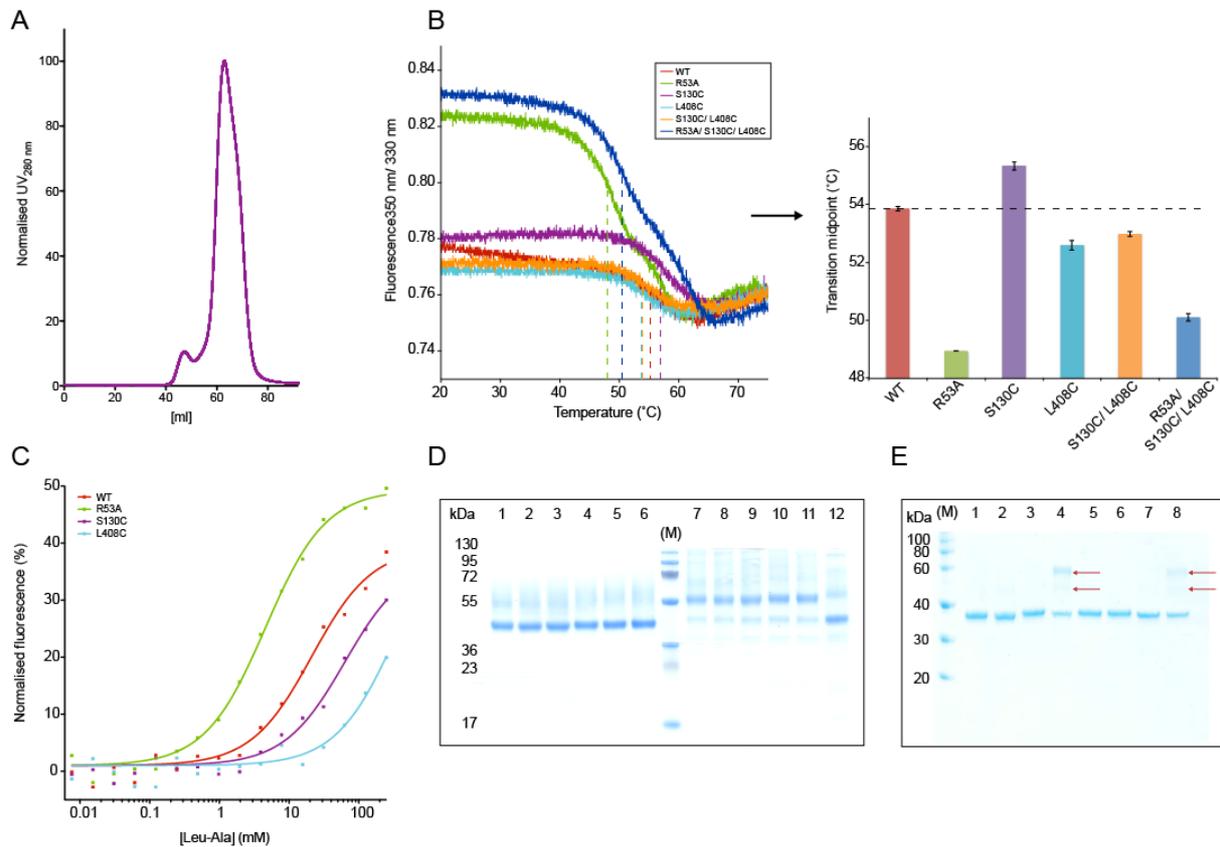
Figure 47: Outward-open structure determination. (A) Size-exclusion chromatogram for PepT$_{St}$ R53A/ S130C/ L408C. The peak is not completely homogeneous, which denotes the presence of different protein populations. (B) DSF characterization of the single mutants and the combination of the mutations. On the left, the raw data from the DSF measurements is shown. The dashed vertical bars indicate the Tm for each protein. On the right, the bar representation of the DSF measurements is shown. The average Tm value was calculated from three independent measurements and the error bars represented the corresponding standard deviation. The horizontal dashed line represents the Tm value of the WT protein used as a control. (C) MST data for the single mutants. PepT$_{St}$ WT presents a K$_D$ of 21.26 ± 3.39 mM. Interestingly PepT$_{St}$ R53A presents a K$_D$ of 4.59 ± 0.36 mM, considerably reduced compared to the WT protein. Finally, mutants PepT$_{St}$ S130C and L408C present a reduced binding affinity to ligand Leu-Ala: PepT$_{St}$ S130C has a K$_D$ > 50 mM and PepT$_{St}$ L408C has a K$_D$ > 100 mM. Furthermore, the combination of these mutations results in a protein that cannot bind ligands at all (data not shown). (D) Pegylation test results for the control samples. PepT$_{St}$ WT (samples 1-6) was used as negative control as the WT protein does not contain cysteine residues. Single mutant PepT$_{St}$ S130C (samples 7-12) was the positive control. Samples were incubated for 45' at RT with different concentrations of mPEG5K (10, 5, 2, 1, 0.5, 0 mM, from left to right). (E) Pegylation test results for PepT$_{St}$ R53A/ S130C/ L408C. The protein was analyzed under denaturing (addition of 1% SDS) and non-denaturing conditions, in oxidized or reduced conditions (addition of 0.5mM TCEP). SDS-PAGE gel legend is summarized as follows:

M: Protein marker. 1: Non-treated protein (control). Samples 1-4 were treated under non-denaturing conditions. Samples 5-8 were treated under denaturing conditions. Samples 2 and 6 were incubated 45' with mPEG5K. Samples 3 and 7 were reduced by the addition of 0.5mM TCEP. Samples 4 and 8 were reduced prior pegylation. Only in these samples higher molecular weight conjugates are detectable (indicated with a red arrow).

In the inward-open conformation, Arg-53 and Glu-312 form a salt bridge, closing the access of the binding cavity from the periplasmic side of the membrane (Solcan et al., 2012; Lyons et al., 2014). Additionally, a second salt bridge in the periplasmic side of the transporter was identified: Gln-170 is interacting with Glu-171, and also Ser-322 is in the close vicinity forming hydrogen bonds to the previous residues. To introduce more flexibility, one or the other residues involved in the salt bridge were mutated to alanine. The typical workflow once the mutation was introduced was to test the mutant expression in small-scale and to characterize the corresponding protein by SEC, DSF and MST. Here, we aimed to identify mutants that were stable enough to be purified in larger amounts and that could be used for crystallization purposes. Even though most of the single mutants were stable to be purified, the combination of several mutations resulted in highly unstable proteins, prone to precipitate. Intriguingly, PepT$_{St}$ R53A showed interesting binding properties as the Leu-Ala K_D was reduced to 4.59 ± 0.36 mM (binding of Leu-Ala to the WT protein presented a K_D of 21.26 ± 3.39 mM) (Figure 47); this result was reproducible and was also observed for other peptides. Therefore, this mutant was positively considered for further experiments.

As a second approach and following the example of LacY transporter (Kumar et al., 2014), two bulky residues were introduced in the close vicinity of the periplasmic gate (A57W and S322W) in order to separate the N- and C-terminal bundles of the transporter. Moreover, these mutants were further combined with R53A. Unfortunately, the insertion of tryptophan residues destabilized the protein in great extend and this approach was discarded.

As an alternative method, we tried to engineer a disulphide bond on the cytoplasmic side of the transporter. To select the residues to mutate into cysteines, the available literature on MFS transporters crystallized in outward-open structures was studied and a list of the residues involved in maintaining this conformation was created. These structures were overlaid with PepT$_{St}$ WT to identify the residues in the equivalent positions. We assumed that interactions between residues located in the N- and the C-terminal bundles are involved in stabilizing the outward-open conformation, precisely between TM4 or TM5 and TM10 or TM11. Several pairs of residues were identified and cysteine mutations were inserted at these points of the sequence. We reasoned that cysteines would form a disulphide bond under oxidizing conditions, locking the transporter in the outward-open conformation. It should be noted that PepT$_{St}$ WT protein does not have any cysteine in its sequence. Several double cysteine mutants were generated and the proteins were expressed and purified. Then, these mutants were used to perform a SDS-PAGE analysis under reducing and oxidizing conditions, to detect a difference in the protein migration behavior. From this first examination, one protein looked more promising than the rest: PepT$_{St}$ S130C/ L408C (Figure 47).

To favor disulphide bond formation, the R53A mutation was also introduced. Even though the protein carried three mutations, its stability was only reduced by ~6°C (50°C compared to the 56°C of the WT protein purified in DDM; Figure 47B). Mutant PepT$_{St}$ R53A/ S130C/ L408C was subsequently purified at pH 8.0, as basic pH promoted thiol group oxidation to thiolate and therefore, higher reactivity to form a disulphide bond (Murray et al., 2012). The mutated protein was subjected to the thiol-maleimide reaction to detect free cysteine residues: PepT$_{St}$ WT was used as negative control, as it does not contain any cysteine residues, and the single mutant S130C was used as positive control (Figure 47D). The analysis was performed under denaturing and non-denaturing conditions, in an oxidizing or reducing environment. Samples presented the same behavior independently if treated in denaturing or non-denaturing conditions, meaning that unfolding of the protein did not affect cysteine residue accessibility. Addition of mPEG-5K did not generate higher molecular weight conjugates. Only when the sample was reduced prior to pegylation, these conjugates were detectable (Figure 47E). Incubation under reducing conditions with 1 mM mPEG-5K did not prevent the presence of the band corresponding to the native protein. Increasing the concentration of mPEG-5K did not change the results. Therefore, there was a population in the protein batch with neither a disulphide bond nor accessible thiol group in their cysteine residues. We think these groups might be oxidized to sulfinic or sulfonic groups (Murray et al., 2012) and thus, not accessible for the maleimide derivative. Additional test to improve the formation were also performed: ion-exchange chromatography was tested to separate different oxidative states in the population but the separation of the peaks were not well resolved. Moreover, we tested cross-liking experiments with divalent metals such as Cu^{2+} or Cd^{2+}; still after incubation at different concentrations and time points, the cysteines were reactive to mPEG-5K meaning that they were accessible and no cross-linking occurred.

On the other hand, it should be mentioned that single mutants PepT$_{St}$ S130C or PepT$_{St}$ L408C presented a considerable reduced binding capability (Figure 47C): PepT$_{St}$ S130C had a K$_D$ > 50 mM and PepT$_{St}$ L408C had a K$_D$ > 100 mM. These results indicate that the presence of the cysteine residues is somehow affecting the accessibility of ligands to the binding cavity and we actually questioned at this point if these mutants were the best candidates to proceed with this approach.

4.2.4 Discussion

To move substrates across the membrane, POTs cycle through a series of different conformations. During this process, the binding cavity is accessible to one or the other side of the membrane, but not both at the same time (Jardetzky, 1966). Therefore, according to the alternate access mechanism, a transporter adopts at least two different conformations over a transport cycle: the inward-facing

state and the outward-facing, with the binding cavity accessible from the cytoplasm or the periplasm, respectively. The additional crystallographic structures reported for other members of MFS transporters (POTs included) ensured that intermediate conformations might occur in nature, such as occluded states or partially occluded.

Previous to our studies, several structures of POTs were reported in inward-open states and inward-open partially occluded (Table 3), but the mechanisms underlying the partial occlusion upon ligand binding were poorly understood. Our structures of PepT$_{St}$ in complex with different ligands, peptidic and non-peptidic molecules, shed light on the multispecifity of peptide transporters and the structural arrangements the transporter undergoes over a transport event.

Even though our PepT$_{St}$ structures in complex with di- or tripeptides are in inward-open conformation, some of them, namely PepT$_{St}$[Ala-Leu], PepT$_{St}$[Ala-Gln] and PepT$_{St}$[phosphate], showed a severe bending of TM11. As a result, the access to the binding cavity from the cytosolic side of the membrane is restricted to ligands but not to small molecules, such as water. The occlusion is caused by the interaction of the ligand with the so-called P2-lid, constituted by residues Trp-427 and Phe-428. These aromatic residues form arene interactions and interact with ligands of different nature. In the structures with Ala-Leu, Ala-Gln and the previously reported structure with Ala-Phe by Lyons and colleagues (Lyons et al., 2014), the residues located at C-terminus of the substrate are large enough to interact with Trp-427 and Phe-428, leading to the bending of TM11. On the contrary, ligands with small residues at C-terminus, such as Phe-Ala, are not able to form such interactions with the P2-lid and as a result, the transporter is in fully inward-facing conformation. It might be that different ligands stabilize a particular conformation of the transporter, depending on the interactions they built with the residues in the binding cavity and surrounding areas. Furthermore, the occlusion can be mimicked by non-peptidic molecules, such as the PEG observed in the binding pocket of PepT$_{St}$[phosphate]. Here, the bending of TM11 is somehow different to the one observed in the structures in complex with peptides. Finally, we reported an apostructure (with a low occupancy PEG in the binding cavity) in fully inward-open conformation.
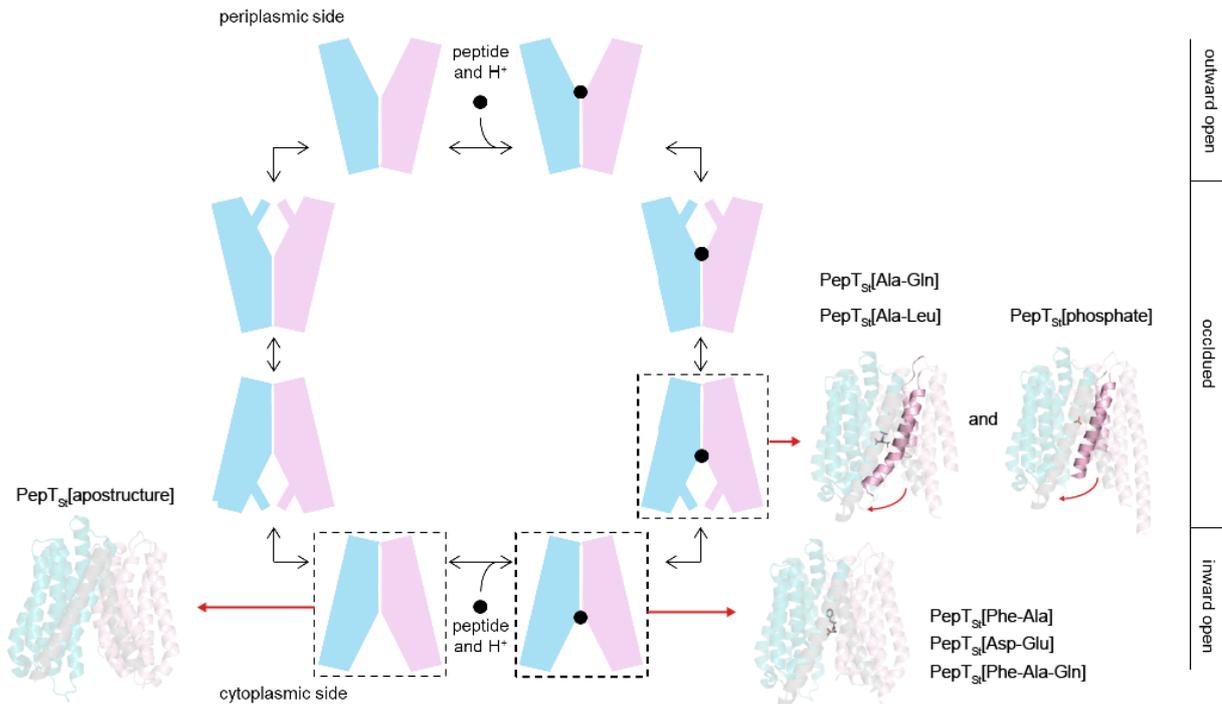
Figure 48: PepT_St transport cycle. The two MFS domains (N- and C-terminal domain) are shown in light blue and pink, respectively. Similarly to other MFSs, POTs are believed to mediate substrate transport alternating between an outward open state (binding site only accessible from the extracellular or periplasmic space) and an inward open state (binding site only accessible from the cytoplasm) via outward-facing and inward-facing occluded forms (access to the binding site restricted from both sides). Here, the structures we reported are represented next the theoretical states highlighted by stippled boxes. Our structures are color-coded as in the representation, and the HA and HB helices are colored in grey. Figure adapted from Martinez Molledo et al., 2018.

Our results show that peptide transporters can adopt intermediate conformations with partial occlusions, which is in good agreement with the recently proposed switch-and-clamp model (Quistgaard et al., 2016). An overall picture of the current knowledge of PepT_St transport cycle is summarized in Figure 48. Here, our reported structures are located next to the theoretical conformations over a transport cycle. Despite the different strategies followed to stabilize the transporter in an outward-open conformation, we did not succeed in out attempts. We have some preliminary data showing that the formation of an engineered disulphide bond in the cytoplasmic side of the transporter might be possible, but further investigation is required to improve the sample homogeneity. Moreover, other possibilities are also under consideration, such as crosslinking the introduced cysteine residues via small metal ions or inserting an unnatural amino acid instead.

4.3 Side projects and collaborative work

Apart from the work in the PepT$_{St}$ project described on the previous sections of the dissertation, I also participated in additional side projects. In this section, a short summary of my contribution to each project is provided.

4.3.1 Lipid-like peptides (LLPs) and saposin-derived lipid nanoparticles (SapNPs)

Typically to work with IMPs, membranes are solubilized with detergents and the IMPs are extracted from the membrane as mixed-micelles of protein, detergent and the remaining membrane lipids. As discussed in the introduction (see section 2.5 for more details), the choice of the detergent is crucial to maintain the stability of the IMPs once it has been extracted from the membrane, but they are not the only possible molecules to keep the IMPs stable in solution. During my studies, I was able to participate in two project in this research line: the first was the use of lipid-like peptides or peptergents as tools for IMPs stabilization and crystallization (Veith et al., 2017), and the second was the use of the salipro proteins to reconstitute IMPs in the presence of lipids (Flayhan et al., 2018). Both projects are now finished and the results were published.

In the LLPs project, ten LLPs were synthetized and screened against several IMPs studied in our lab and also in the lab of Prof. Dr. Henning Tidow. Here the stability of the protein in the presence of the LLPs was measured by DSF. Additionally, a systematic study of the effect of LLPs in IMP crystallization was performed using the LPPs inducing a stronger stabilization effect over the hypothetical sugar transporter from *E.coli* used in these studies. Here the number of hits in the commercial crystallization screen MemGod2 was counted in presence and absence of the LLPs. Whereas 20–30% of the conditions grew crystals when using LMNG as the detergent and LLP7 or LLP8 as the additive,without the presence of the LLPs, crystals only appeared in seven (out of 96) conditions. Therefore, a positive effect of the LLPs on the crystallizability of the *E. coli* transporter was observed. My role here was to screen the crystals grown for the *E. coli* transporter at the synchrotron beamline.

In the SapNPs project, also several proteins studied in our lab (including PepT$_{St}$ and PepT$_{So2}$) were reconstituted in the salipro discs using different lipids. Once reconstituted, we tested the stability of the IMPs in SapNPs by DSF and we studied the functionality of the reconstituted IMPs by MST. Here we measured the ability of the reconstituted PepT$_{St}$ and PepT$_{So2}$ to bind peptide ligands. Both PepT$_{St}$ and PepT$_{So2}$ were more stable when reconstituted in salipro discs (for example, PepT$_{St}$ was almost 10°C more stable when reconstituted in salipro discs with brain lipids). Furthermore, they were still able to bind peptides in the same K$_D$ range as the detergent-solubilized protein. My role in

this project was to purify PepT$_{St}$ and PepT$_{So2}$ and provide these proteins for the reconstitution and to perform the MST measurements for the detergent-solubilized proteins and the reconstituted proteins.

4.3.2 LCP crystal growth in CrystalDirect plates. Crystal harvesting and in situ data collection

The main bottleneck of LCP crystallization is still harvesting crystals, as opening the plate often results is crystal breaking or just finding the crystals is very difficult as the drop components move around. Furthermore, the crystals grow in a highly viscous environment, making very challenging the harvesting of individual crystals. Moreover, the mesophase can turn opaque when it is frozen and therefore, the crystal is not further visible once mounted at the beamline. Then, the crystals cannot be center properly, resulting in a data set of low completeness.

To avoid crystal manipulation, we explore the possibilities of LCP crystallization in CrystalDirect (CD) plates (Cipriani et al., 2012). We used PepT$_{St}$ and the described crystallization condition in HEPES to grow crystals in CD plates. The top and the bottom of these plates consist of a thin optically transparent COC film. A plastic frame divides the plate in 96 wells with their corresponding reservoir space. We set up the plates using the same experimental approach as the one described for the standard LCP plates, but in this case the reservoir was pre-filled with 30-40 μl of crystallization solution. Drops contained 50 nl of *bollus* and 800 nl of the crystallant. Crystals grew within a day but the size was slightly smaller compared to the crystals grown in the regular LCP plates.

The next step involved testing two different approaches to bring the crystals to the beamline: either harvesting the crystals automatically or testing them *in situ* (at room temperature) at the beamline. For the crystal harvesting, we used the automatized crystal harvester (Zander et al., 2016) system develop by EMBL Grenoble and actually commercialized by Arinax. This project was carried out with Dr. Sophie Zimmermann (former postdoctoral fellow in the group of Dr. Thomas Schneider at the EMBL Hamburg). Here, the COC film of the plate is cut through laser photoablation and the COC film with the crystal or crystals remains glued to the crystal-mounting pin. Finally, the robot is also able to freeze the harvested crystals using a cryo-stream. The first tests were performed for crystal harvesting and even though the crystallant was aspired, the mesophase was too dense and viscous to be removed by the same procedure. Still, the whole *bollus* could be automatically harvested and frozen. Later, at the beamline, we faced the difficulty of centering at the beam position individual crystals for screening purposes. Therefore, this approach still requires some optimization. Alternatively, together with Doris Jahn from the instrumentation group at the EMBL

Hamburg, we designed a prototype of a three-D printed frame to adapt a normal LCP plate to the harvesting robot. Here the LCP plate was sealed with a COC film to permit the crystal harvesting. The first tests were carried out with success.

As a second approach, CD plates can be mounted directly at P14 beamline using a special goniometer head. The data collection can be then performed at room temperature, with no crystal harvesting and freezing. As a limitation, the plate cannot be rotated with much freedom and therefore, a full data set cannot be collected from a single crystal. Instead, several partial data sets from different crystals need to be merged to reach completeness. Together with Dr. Gleb Bourenkov, we used this approach for PepT$_{St}$ and a structure at 2.5 Å resolution was determined from crystals screened at room temperature (manuscript in preparation).

4.3.3 LCP-fluorescence recovery after photobleaching (FRAP)

Another aspect related to LCP crystallization is the choice of the lipid to form the mesophase. Considering that to date there are several available lipids, this is one more variable to consider in crystallization. Together with the Sample Preparation and Crystallization (SPC) facility at the EMBL Grenoble, we started a collaborative project in which the effect of the lipids used for LCP crystallization was systematically study for PepT$_{St}$. This work is still in progress but we aimed to understand how the different monoacylglycerols (MAGs) affect PepT$_{St}$ crystallization. For this purpose, we make use of the LCP-FRAP approach (Cherezov et al., 2008; Fenalti et al., 2015). Here, PepT$_{St}$ was labeled with Cy3 dye prior mesophase formation with the different MAGs available. Then, the LCP-FRAP experiment is set up using a particular type of glass plates. The plates were incubated overnight before the FRAP experiment and for performing the measurements, they were placed in the Formulatrix LCP-FRAP robot (Xu et al., 2011). The procedure to follow involved the bleaching of the fluorophore attached to the protein in a particular point of the mesophase. Then, the fluorescence recovery in this particular position is tracked overtime until it is recovered, if recovered at all. This information can be correlated with the mobility of the protein in the cubic phase. LCP-FRAP has been proposed as a pre-screening method for LCP crystallization (Cherezov et al., 2008), where the most appropriate crystallization MAG can be selected and also the crystallization conditions can be pre-screened. This method does not required as high protein concentrations as crystallization, it can provide valuable pre-screening information which can be later translated to the crystallization experiment. Furthermore, it can be performed in high-throughput and it is highly automatized.

With this collaboration, we aim to (1) get experience testing different MAGs in crystallization. (2) Study how MAGs affect mesophase formation for PepT$_{St}$ and its crystallizability. (3) Ultimately, we

would like to screen $PepT_{St}$ crystals grown with different MAGs and correlate this to its diffraction properties.

4.3.4 DtpC-Nb26 complex

In the lab, we work with the peptide transporters of other bacterial species: $PepT_{So2}$ and $PepT_{So}$ from *Shewanella oneidensis*, DtpA, -B, -C and –D from *E. coli* and $PepT_{St}$, in terms of structural determination and characterization of ligand preferences by DSF and MST. Together with Dr. Sophie Zimmerman, the purification of DtpC in complex with the nanobody 26 (Nb26) was performed and used for crystallization purposes. The complex could be purified in LMNG and DDM detergents and the first crystallization trials were set up. Crystals of the complex obtained in LMNG were screened at the synchrotron beamlines P14 and P13, and they diffracted anisotropically to 6 Å in the best direction. For the crystals obtained from the complex purified in DDM diffraction properties were considerably worst, but still optimization of the crystallization condition is ongoing. Ultimately, we aimed to determine the structure of DtpC and compare it to the other POTs, and understand how the binding pocket features correlate to the ligand preferences.

Yonca Ural-Blimke established the protocol described in section 6.3.13 in our lab. It is a robust approach to generate complexes between nanobodies and bacterial POTs, resulting in pure and homogenous complexes that can be used for crystallization purposes.

4.3.5 Systematic ligand screening for bacterial POTs

Considering the good correlation we observed for $PepT_{St}$ in terms of ligand screening in solution and ligand binding in crystallization, we decided to extend the use of the library we created originally for $PepT_{St}$ characterization and use it to characterize the rest of the bacterial POTs we have in our lab. The student Rolf Munk Nielsen under the supervision of Dr. Christian Löw and myself performed this experiments. We used DSF to do the first screening of the library and the effect of the pH on POT stability. Secondly, we tested a subset of peptides (the same in all the cases) and test their binding affinities to the different POTs by MST. Interestingly, we observed considerably different ligand preferences for the studied POTs, even among the ones of *E.coli*. Additionally, some transporters such as $PepT_{So}$ or DtpD did not show any responsiveness to peptides in solution. We only observed a strong stabilization effect due to pH in the case of $PepT_{St}$. This systematic study contributes to the overall knowledge of ligand preferences in POTs and correlates well with already published transport studies.

Chapter 5. Summary and future prospect

Peptide transporters (POTs, PRTs or PepTs) are members of the MFS family of transporters, one of the largest families of transporters in nature. They are involved in the uptake of di- and tripeptides, using a proton gradient to direct transport (Daniel et al., 2006). Once transported into the cell, peptides are used as a nitrogen source to build other biomolecules and as a source of amino acids. POTs are present in organisms from all evolutionary lineages. The human representatives, hPepT1 and hPepT2, are primarily expressed in the smooth intestines and the kidneys, respectively. In the former, hPepT1 is involved in the uptake of dietary peptides; in the latter, hPepT2 controls the reabsorption of peptides (Knütter et al., 2004; Biegel et al., 2006). Furthermore, hPepT1 can transport peptide-mimetic drugs and therefore, these transporters are of pharmacological interest (Luckner and Brandsch, 2005; Brandsch, 2013).

Despite the efforts to functionally characterize hPepTs, to date there is no crystallographic structure available but instead bacterial homologues have been subject of intense research to shed light on molecular mechanisms of peptide transport. As summarized in Table 3, there are several structures of bacterial POTs, either reported in ligand-free (apo-form) or with ligands (peptides, peptidomimetic molecules and non-peptidic molecules) (Solcan et al., 2012; Lyons et al., 2014; Huang et al., 2015, 2016; Ma et al., 2017; Martinez Molledo et al., 2018; Guettou et al., 2013, 2014; Fowler et al., 2015; Newstead et al., 2011; Zhao et al., 2014; Boggavarapu et al., 2015; Doki et al., 2013; Parker et al., 2017). Together these results provided information on the structural fold of POTs and basic concepts on ligand coordination: POTs share the canonical fold of MFS transporters, with 12 TM α-helices divided in two bundles of six consecutive TM α-helices each (N- and C-terminal domains), with the binding cavity in between these domains (Newstead, 2017). Bacterial POTs have two additional α-helices, denoted HA and HB, in between the conserved TM domains. These helices have still an unknown function. Peptide-bound structures showed that the substrate is bound mainly by its N- and C-terminus, mostly by salt bridges and hydrogen bonds. It was believed that POTs transport peptides with different composition given the main coordination contacts can be established via the N- and C-termini of the substrate. This conclusion was based on moderate resolution crystallographic models and the study of few ligands, with poor variability in the physicochemical features of the substrate. Even though these interactions are essential, the side chains of the residues in the ligand also play an important role in ligand binding.

Prior to our research, only two crystallographic models of PepT$_{St}$ in complex with Ala-Phe and Ala-Ala-Ala were available (Lyons et al., 2014). Additionally, Solcan and coworkers (Solcan et al., 2012) and later Lyons and colleagues (Lyons et al., 2014) screened few more ligands in liposome-based assays, investigating the transport of Ala-Phe, Phe-Phe, Ala-Ala, Ala-Tyr, Gly-Gln, Glu-Glu and Lys-Lys. Already in these studies, PepT$_{St}$ was reported to transport preferentially dipeptides over tripeptides, whereas non-polar residues or polar but not charged (at least at C-terminus) were preferred over polar residues. Still a more systematic approach was necessary to draw reliable conclusions.

The aim of our ligand binding studies in solution was to develop a rapid screening method to test a large library of compounds with as little sample requirement as possible. Here, the capillary-based DSF setup of NanoTemper technologies turned to be very suitable: measurements could be performed on detergent-solubilized PepT$_{St}$ at low concentration (~10 μM protein concentration) and reduced volume (~10 μl of sample volume per capillary). Furthermore, several samples could be tested in parallel without the need for protein labeling, but instead the intrinsic fluorescence of the protein was recorded upon heat unfolding. Ligands were identified as they stabilized PepT$_{St}$, increasing the transition midpoint of the protein by few degrees. We assembled a peptide library of commercially available di- and tripeptides, trying to include the widest range of physicochemical features possible (e.g., chemical properties of amino acids, molecular sizes, N- or C-terminal positioning of amino acids residues in peptide ligands). It became apparent that PepT$_{St}$ preferentially binds to dipeptides compared to tripeptides, as it was earlier reported (Solcan et al., 2012; Lyons et al., 2014). Within dipeptides, those carrying small and bulky non-polar side chains were good binders, regardless of the position in the substrate (either N- or C-terminus). Furthermore, Ala-Gln was also found to bind well, implying that a glutamine residue can be accommodated in the second position. On the contrary, binding of charged residues was highly disfavored.

Then, microscale thermophoresis (MST) was used to further characterize those substrates that were identified as binders by the peptide library screening. Here, we aimed to estimate binding affinities for different peptides. These results revealed binding with low millimolar affinity for several dipeptides, while all tested tripeptides exhibited very low affinity or no binding at all under the employed experimental conditions (see Table 4). Thus, the results obtained by MST were in agreement with our DSF results and the liposome-based studies (Solcan et al., 2012; Lyons et al., 2014). Based on this information and the observation that acidic pH impaired binding, a pH-controlled crystallization of PepT$_{St}$ in presence of different potential peptide ligands was carried out.

Despite the difficulties and challenges of membrane protein crystallization, our studies reported structures with high resolutions ranging from 2.0 to 2.7 Å. We determined a total number of nine structures: four structures of PepT$_{St}$ in complex with dipeptides, one structure in complex with tripeptide Phe-Ala-Gln, an apostructure in fully inward-open conformation, an occluded conformation (with a phosphate ion and a PEG molecule in the binding cavity) and finally, two models of PepT$_{St}$ with a HEPES molecule vertically positioned in the binding cavity.

From the studies related to dipeptide binding (Martinez Molledo et al., 2018), the general mechanisms underlying the adaptability of the binding cavity to substrates with different features could be established: (1) there are two possible positions (modes) how the dipeptide backbone can be accomodated in the binding pocket: the binding mode 1 is represented by the structures in complex with Ala-Leu, Ala-Gln and Asp-Glu, whereas the binding mode 2 is represented by the structure in complex with Phe-Ala. Due to the tighter interaction of Phe-Ala to the N-terminal bundle of PepT$_{St}$ in the latter, there is an additional space in the binding cavity at the C-terminus of the ligand, which can potentially fit a third residue of a tripeptide. (2) The first residue of the peptide occupies the pocket 1 (P1), whereas the second is located in pocket 2 (P2). P1 provides a more hydrophobic environment, whereas P2 is more hydrophilic. The residues forming P1 are at very similar position in all the structures, but in P2 the Tyr-68 residue shows a range of different positions, which correlate to the size of the substrate residue located in P2 and is involved in fine-tuning the size of P2 according to the substrate. Interestingly, Tyr-68 is also highly conserved in the POT family highlighting its importance in coordinating peptide binding in the binding cavity (Chen et al., 2000). (3) Based on the high quality of our data sets, we identified a well-ordered water molecule network in the binding cavity, not only in the apostructure (PDB ID 5OXO) but also in the ligand-bound structures. Certainly, not all the observed positions of water molecules in the apostructure are compatible with binding of ligands. Indeed, in some ligand-bound structures, the water molecules contributed to the overall coordination of the substrate. Therefore, the solvation of the binding cavity might facilitate the binding capability of the pockets (particularly P1) to the different substrates. (4) Finally, POTs have a binding cavity that is rich in aromatic residues, particularly P2 and the lid of P2. Aromatic residues have the capacity to form hydrophobic interactions and specific arene interactions, such as arene-arene, arene-amide, π-stacking, cation-π interactions, and X-H…π-bonds (X = C, N, O) (Harder et al., 2013; Nishio et al., 2014; Salonen et al., 2011), which makes them ideally suited for multispecific recognition.

Additionally, we also observed non-peptidic molecules occupying the binding cavity, such as phosphate ions, and PEG or HEPES molecules provided in the crystallant. This observation further highlights the multispecificity of peptide transporters, which are able to bind peptide-mimetic molecules with low affinity. We observed that a phosphate ion paired with a HEPES molecule could mimic a peptide as it strongly overlaps with the C-terminus of the peptides bound in mode 1. The HEPES molecule itself partially overlaps with the peptide backbone of the Phe-Ala (substrate bound in mode 2). On the other hand, the phosphate ion located in the binding cavity in the PepT$_{St}$[phosphate] model is bound in a slightly different position compared to the phosphate ion of the PepT$_{St}$[HEPES] models. In this case, the phosphate does not coincide with any of the peptides, but instead, we observed a strong overlap of this phosphate ion to the nitrate ion of the *A. thaliana* nitrate transporter NRT1.1 (Parker and Newstead, 2014; Sun et al., 2014). Here, the residue coordinating the nitrate ion (His-356) is equivalent to the residue interacting with the phosphate ion in PepT$_{St}$ (Glu-299). Regulatory functions of ion binding in MFS transporters have been described for members of the SLC17 family (Juge at al., 2006). However, for POTs, we have currently no evidence for the biological relevance of ion binding but it would be of interest to investigate this further.

Thus, the obtained structures of PepT$_{St}$ in complex with HEPES suggested that the previously published model for tripeptide binding should be revised. Apart from our structure of PepT$_{St}$ in complex with tripeptide Phe-Ala-Gln, there are three more models of bacterial POTs available with a tripeptide bound: PepT$_{So2}$ was reported in complex with Ala-Ala-Ala and Ala-Try(Br)-Ala (where 'Br' denotes a bromination site) (Guettou et al., 2014) and the vertically bound Ala-Ala-Ala model proposed for PepT$_{St}$ (Lyons et al., 2014). The PepT$_{So2}$-tripeptide bound models, even though at relatively low resolution (3.2-3.9 Å), proposed that tripeptide binding was conserved, as both di- and tripeptides were modeled horizontally in the binding cavity. The later proposed model (Lyons et al., 2014) of vertically bound tripeptides was surprising in the field and had a considerable impact, as several publications followed this observation (Parker at al., 2014; Fowler et al., 2015; Samsudin et al., 2016). Our structure of PepT$_{St}$[Phe-Ala-Gln] reliably shows that PepT$_{St}$ also binds tripeptides similar to dipeptides. Although PepT$_{St}$ seems to preferentially transport dipeptides, at least for tripeptides of the Phe-Ala-Xxx sequence it shows affinities comparable to dipeptides. However, it should be noted that Phe-Ala-Xxx peptides were never tested prior to this study, even in the previously reported liposome-based assays (Solcan et al., 2012; Lyons et al., 2014).

Finally, we report PepT$_{St}$ models in three different conformations: the inward facing state and occluded conformation, and an intermediate state that we called inward-open partially occluded. The partially occluded conformations show a bending of TM11 that restricts the access of di- or tripeptides to the binding cavity but not to small molecules, such as water.

Among the ligand-bound structures, PepT$_{St}$[Ala-Leu] and PepT$_{St}$[Ala-Gln] are fully inward-open partially occluded state, whereas PepT$_{St}$[Asp-Glu], PepT$_{St}$[Phe-Ala] and PepT$_{St}$[Phe-Ala-Gln] are in inward-open. We conclude that the occlusion of the former structures is due to the interaction between the residues at the C-terminus of the ligand with the P2-lid (Trp-427 and Phe-428).

The occluded structure with phosphate in the binding cavity is somehow different to the partial occluded models with a peptide bound. Therefore, we propose that the occluded conformations might be intermediate states between the inward and the outward facing conformations. Finally, despite the work accomplished to establish the outward-open conformation, we did not succeed in the crystallization of the transporter in this conformation but we gathered preliminary data of interesting mutants that should be considered for further experimentation.

Even though our studies have provided considerable insights into ligand recognition and binding in bacterial peptide transporters, there are still open questions and additional aspects that should be investigated further.

As already mentioned above, it appears promising to further explore the role of ion binding in either transport or regulation of transport. Particularly, under consideration of the structures in complex with phosphate, the biological relevance of the phosphate ion binding should be explored.

Even though the ligand screening and characterization methods that we have developed are reproducible and correlate well with the conditions in a crystallization drop, certain limitations remain. First of all, as we used label-free approaches, ligands carrying tyrosine or tryptophan residues were excluded from our library and MST measurements. Therefore, we have no information about the effect these residues might have in ligand coordination. Secondly, in our studies we observed binding but not transport and therefore in our experimental setup there is neither pH gradient nor ligand gradient. Moreover, we have no experimental data to discern whether the transporter has a preferred conformation in its detergent-solubilized form or not. Consequently, it is unclear if we are currently measuring binding to a particular conformation or once the ligand is bound, the transporter might adopt a particular state.

In terms of tripeptide binding, our studies were restricted to substrates with a Phe-Ala-Xxx sequence, but we could still explore the influence of residues with different chemical features and sizes in the middle position of the tripeptide. Furthermore, the role of PepT$_{St}$ in drug uptake

remains unexplored. Although several studies have reported transport of drugs for several of the *E. coli* POTs (Prabhala et al., 2017) and hPepT1 and hPepT2, a more systematic screening should be performed for PepT$_{St}$.

Thus, it would be of high interest to compare the binding data to actual transport, either in liposome-based assays or *in vivo*. For the liposome approach, there are reported studies in which radioactively-labeled peptides were used. It should be mentioned that the main limitation of reconstituting proteins in liposomes is that the protein of interest can be inserted in two different orientations into the membrane of the liposome, which could affect the transport readout of the experiment. Furthermore, most of the conclusions made from this experiment are coming from competition assays, meaning that the same radioactively-labeled ligand is used as a reporter molecule and then, a second peptide is provided to study how it affects the uptake of the labeled one. Even though this experimental approach is widely used in the field, it should be noted that the fact that a ligand is competing with other does not imply that the non-labeled ligand is transported. It could happen that the competitor is somehow inhibiting the uptake of the probe.

For the *in vivo* approach, we tested whether PepT$_{St}$ could take up the reporter molecule β-Ala-Lys-AMCA but the results were not promising. Alternatively, the peptide transporter of interest could be expressed in bacterial cells and different peptides could be provided in the growth media. Then, the cellular content could be analyzed by mass-spectrometry to make conclusions about the ligand preferences of the peptide transporter (Prabhala et al., 2017). A similar approach could be used to study the content of liposomes. Furthermore, to gain a better understanding of the function of POTs in a cellular context, more research could be carried out to study the effect of the lipid environment in the function of the transporter. Here, native mass-spectrometry has provided interesting observation for other IMPs and could be an approach to consider (Laganowsky et al., 2013, 2014). Additionally, it would be interesting to identify interaction partners of peptide transporters, which could be involved in the regulation of these transporters or couple their function to proton pumps, for example.

Finally, an open question that still remains is to solve the structure of the outward-facing state of POTs. Here, an alternative to the applied mutational approach, the synthesis of nanobodies could be considered, as they could stabilize a particular conformation as reported for GPCRs (Manglik et al., 2017). Furthermore, our preliminary results on the double cysteine mutant showed that these residues can interact and therefore, this strategy could be further optimized or used as a starting

point for crosslinking experiments with divalent metals or chemical cross-linkers. Additionally, a more systematic residue screening should be performed as only few mutations were tested.

Overall, our studies have established a workflow to study peptide transporters and have shed new light on the molecular mechanisms of peptide recognition and binding in bacterial POTs, which can be extrapolated to other members of this biologically and medically important protein family.

Chapter 6. Materials and methods

6.1 Materials

6.1.1 Chemicals and consumables

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Peptides (nomenclature with the three letter code for amino acids)		
L-Ala	Fluka	Cat# 5129
D-Ala	Sigma-Aldrich	Cat# 162655
Ala-Ala	Sigma-Aldrich	Cat# A502
Ala-Ala-Ala	Sigma-Aldrich	Cat# A9627
Leu-Leu	Bachem	Cat# M-1535
Ala-Leu	Sigma-Aldrich	Cat# A1878
Leu-Ala	Bachem	Cat# G-2460
Ala-Phe	Sigma-Aldrich	Cat# A3128
Phe-Ala	Bachem	Cat# G-2850
Met-Ser	Sigma-Aldrich	Cat# M9380
Gly-Ser	Sigma-Aldrich	Cat# G3127
Ala-Gln	Sigma-Aldrich	Cat# G8541
Ala-Asp	Bachem	Cat# G-1195
Ala-Glu	Bachem	Cat# G-1200
Ala-Lys	Bachem	Cat# G-1290
Lys-Ala	Bachem	Cat# G-2630
Thr-Gln	Sigma-Aldrich	Cat# T3275
Gly-His	Sigma-Aldrich	Cat# G1627
His-Ser	Sigma-Aldrich	Cat# H3129
Asp-Glu	Sigma-Aldrich	Cat# A1916
Glu-Glu	Sigma-Aldrich	Cat# G3640
Leu-Leu-Ala	Bachem	Cat# H-3905
Ala-Leu-Ala	Bachem	Cat# H-5975
Ala-Phe-Ala	Bachem	Cat# H-5420
Phe-Ala-Ala	GL Biochem	N/A
Phe-Ala-Leu	GL Biochem	N/A
Phe-Ala-Gln	GL Biochem	N/A
Phe-Ala-Thr	GL Biochem	N/A
Phe-Ala-Phe	GL Biochem	N/A
Phe-Ala-Asp	GL Biochem	N/A
Leu-Gly-Gly	Sigma-Aldrich	Cat# L9750
Met-Ala-Ser	Sigma-Aldrich	Cat# M1004
Ala-Ala-Tyr	Bachem	Cat# H-1445
Ala-Pro-Ala	Bachem	Cat# H-1595
Gly-Gly-His	Sigma-Aldrich	Cat# G4541
Detergents and lipids		

n-nonyl- β-D-maltoside (NM) detergent	Anatrace	Cat# N330
n-decyl- β-D-maltoside (DM) detergent	Anatrace	Cat# NG322
n-dodecyl-β-D-maltoside (DDM) detergent	Anatrace	Cat# D310
2,2-didecylpropane-1,3-bis-β-D-maltopyranoside (LMNG) detergent	Anatrace	Cat# NG310
1-(7Z-pentadecenoyl)-rac-glycerol (7.8 MAG)	Avanti Polar Lipids, Inc.	Cat# 850531O
Chemicals		
Acetic acid 100 %, Rotipuran® 100 %, p.a.	Roth	Cat# 3738
Acrylamide/Bis stock sol. 29:1 (40% w/v)	Roth	Cat# A515
Agarose for DNA electrophoresis	SERVA	Cat# 11404
Ammonium peroxodisulphate	Roth	Cat# 9592
Ammonium phosphate monobasic	Sigma-Aldrich	Cat# 216003
Ampicillin disodium salt	Roth	Cat# K029
Boric acid ≥ 99,8 %, p.a., ACS, ISO	Roth	Cat# 6943
Citric acid monohydrate	Sigma-Aldrich	Cat# C1909
cOmplete, EDTA-free protease inhibitor cocktail tablets	Roche	Cat# 5056489001
D(+)-Saccharose ≥ 99,5 %, p.a.	Roth	Cat# 4621
di-Sodium hydrogen phosphate heptahydrate	Roth	Cat# X987
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat# D2650
EDTA Tetrasodiumsalt	Roth	Cat# 3619
Ethanol ≥ 99,8%	Roth	Cat# 9065
Ethidium bromide Solution 0,025 %	Roth	Cat# HP47
Glucose	Roth	Cat# X997
Glycerol, Rotipuran® ≥ 99,5%, p.a.	Roth	Cat# 3783
Guanidine hydrochloride ≥ 99,5%	Roth	Cat# 0037
HEPES, Pufferan® ≥ 99,5%, p.a.	Roth	Cat# 9105
Hydrochloric Acid 32%	Roth	Cat# X896
Imidazol, Pufferan® ≥ 99%, p.a.	Roth	Cat# X998
Iodoacetamide	Sigma	Cat# I1149
Isopropyl β-D-1-thiogalactopyranoside	Roth	Cat# 2316
Kanamycin sulphate	Roth	Cat# T832
LB Agar (Lennox)	Roth	Cat# X965
LB Broth Low Salt Granulated	Melford	Cat# GL1703
Magnessium chloride hexahydrate ≥ 99 %, p.a., ACS	Roth	Cat# 2189
MES, Pufferan® ≥ 99%	Roth	Cat# 4256
Methanol Rotipuran® ≥ 99,9 %, p.a., ACS, ISO	Roth	Cat# 4627
Methoxypolyethylene glycol maleimide 5000	Sigma-Aldrich	Cat# 63187
Nickel(II) chloride hexahydrate ≥ 98 %, p.a.	Roth	Cat# 4489
Polyethylen glycol 300 (PEG 300)	Fluka	Cat# 90878
Polyethylen glycol 400 (PEG 400)	Fluka	Cat# 91893

Sekusept Plus	Ecolab	Cat# 104372E
Sodium chloride, >99,5%, p.a., ACS, ISO	Roth	Cat# 3957
Sodium hydroxide ≥ 99	Roth	Cat# 9356
Sodium phosphate dibasic dodehydrate	Sigma-Aldrich	Cat# 71649
TEMED ≥ 99%, p.a.	Roth	Cat# 2367
Terrific Broth (TB, modified), granular	Melford	Cat# GT1702
tri-Sodium citrate dehydrate	Roth	Cat# 3580
Tris Acetate-EDTA (TAE) buffer 10X	Sigma-Aldrich	Cat# T8280
TRIS hydrochloride, Pufferan® ≥ 99%, p.a.	Roth	Cat# 9090
tris(2-carboxyethyl)phosphine	Solvec Ventures	Cat# M115
Urea ≥ 99,5%, p.a.	Roth	Cat# 3941
Molecular cloning		
DNase, RNase-free PCR tubes	Sarstedt	
dNTPs	New England BioLabs® Inc.	Cat# N0447S
dGTP, dCTP, dATP, dTTP	New England BioLabs® Inc.	
ATP	Sigma-Aldrich	Cat# A9187
Nuclease-free water	Qiagen	Cat# 129115
Customised lyophilized primers	Eurofins Genomics	
6x DNA Loading Dye	Thermo Fisher Scientific	Cat# R0611
Gene Ruler 1 kb DNA Ladder	Thermo Fisher Scientific	Cat# SM0311
Enzymes		
DpnI restriction enzyme	New England BioLabs® Inc.	Cat# R0176S
Taq DNA Polymerase	New England BioLabs® Inc.	Cat# M0267S
Phusion ® High-Fidelity DNA Polymerase	New England BioLabs® Inc.	Cat# M0530S
T4 DNA Polymerase	New England BioLabs® Inc.	Cat# M0203S
T4 DNA Ligase	New England BioLabs® Inc.	Cat# M0202S
T4 Polynucleotide Kinase	New England BioLabs® Inc.	Cat# M0201S
DNase I recombinant	Roche	Cat# 4536282001
Lysozyme	Roth	Cat# 8259
Commercial kits		
QIAquick PCR Purification Kit	Qiagen	Cat# 28106
QIAquick Miniprep Kit	Qiagen	Cat# 27104
QIAquick Gel Extraction Kit	Qiagen	Cat# 28706
Mix & Go! *E. coli* Transformation Kit and Buffer Set	Zymo Research	Cat# T3001
Protein purification and characterization consumables		
Ni-NTA Agarose	Invitrogen	Cat# R901-15
CaptureSelect beads for EPEA-tag	Thermo Fisher Scientific	Cat# 194288010
Econo-Pac® Chromatography columns	BioRad	Cat# 7321010
Roti®-Mark 10-150 Protein-Marker	Roth	Cat# T850
NuPAGE® LDS Sample Buffer (4X)	Thermo Fisher Scientific	Cat# NP008
NuPAGE® Novex® 4-12% gels	Thermo Fisher Scientific	

InstantBlue™	Expedeon	Cat# ISB1L
Spin-X® UF 100 kDa concentrator (20 and 6 ml, and 500 µl)	Corning	Cat# 431491, 431486 and 431481
Spin-X® UF 50 kDa concentrator (20 and 6 ml, and 500 µl)	Corning	Cat# 431490, 431485 and 431480
Prometheus NT.48 nanoDSF Grade Standard Capillaries	Nanotemper	Cat# PR-C002
Monolith NT.LabelFree Zero Background MST Premium Coated Capillaries	Nanotemper	Cat# MO-Z005
Crystallization consumables		
Laminex Plastic Base 100 micron (100 µm well depth)	Molecular Dimensions	Cat# MD11-51-100
Laminex Plastic Cover 200 MicronUV (200 µm)	Molecular Dimensions	Cat# MD11-53-200
100 µl Hamilton syringe	Hamilton	
50 µl Hamilton syringe	Hamilton	
10 µl Hamilton syringe	Hamilton	
mosquito® LCP syringe coupling	TTP Labtech	Cat# 3072-01050
mosquito® LCP needles	TTP Labtech	Cat# 4150-05902
Fishing loops 20, 30, 35, 50, 75 µm	MiTeGen	
Additional consumables for crystal harvesting and manipulation	MiTeGen and Molecular Dimensions	
MemGold and MemGold2 crystallization screen	Molecular Dimensions	Cat# MD1-41 and MD1-64
Cubic crystallization screen	Jena Bioscience	Cat# CS-EB-LCP-B
MemTrans crystallization screen	Molecular Dimensions	Cat# MD1-112
JCSG+ Suite crystallization screen	Qiagen	Cat# 130720
NeXtal Stock Kit Salt	Qiagen	Cat# 132985

Table 7: Chemicals and consumables used in the experiments described in the dissertation.

6.1.2 Equipment

REAGENT or RESOURCE	SOURCE
Molecular biology	
PCR cycler	Eppendorf
Electrophoresis chamber for agarose gels	NeoLab
Power supply	Consort
Gel imaging and documentation system	Bio-Rad
Protein expression	
Water bath at 42°C	VWR
Centrifuge Avanti JXN-26	Beckman-Coulter
JLA 8.1000 rotor	Beckman-Coulter
New Brunswick™ Innova® 42 small shaking incubator	Eppendorf

New Brunswick™ Innova® 44 shaking incubator for large scale cultures	Eppendorf
Protein purification and characterization	
EmulsiFlex-C3 cell homogenizer	Avestin
Ultracentrifuge Optima XE-90	Beckman-Coulter
Ti 45 rotor	Beckman-Coulter
Ultracentrifuge tubes (polycarbonate, 70 ml tubes)	Beckman-Coulter
ÄKTA Pure with fraction collector F9C	GE Healthcare
Analytical gel filtration system	Agilent technologies
Superdex200 Increase 10/300 GL (gel filtration column)	GE Healthcare
HiLoad 16/600 Superdex200 (gel filtration column)	GE Healthcare
HiLoad 16/600 Superdex75 (gel filtration column)	GE Healthcare
Electrophoresis chamber for SDS-PAGE gels	Invitrogen
Monolith NT.LabelFree microscale thermophoresis (MST)	Nanotemper
Prometheus NT.48 nano differential scanning fluorimetry (DSF)	Nanotemper
Crystallization	
Mosquito-LCP	TTP Labtech
Thermoblock (at 45°C)	Eppendorf
Rock Imager	Formulatrix
Scorpion dispenser	ARI-Art Robbins Instruments
Microscope	Nikon
General laboratory equipment	
pH meter	Mettler Toledo
Peristaltic pump	Medorex
Balances	Sartorius
Microwave	
Fine balance	Sartorius
MilliQ water machine	Millipore
Magnetic stirrer	Roth
Thermoblock	Eppendorf
Vortex	Janke und Kunkel
Refrigerated bench-top microcentrifuge (5424R)	Eppendorf
Refrigerated bench-top centrifuge (5810R)	Eppendorf
Freezer -20°C	Liebherr
Freezer -80°C	Eppendorf
Fridge +4°C	Liebherr
Nanodrop UV-Vis spectrophotometer	Thermo Fisher Scientific
Fume hood	Waldner

Table 8: Equipment used in the experiments described in the dissertation.

6.1.3 Plasmids and bacterial cell strains

REAGENT or RESOURCE	SOURCE	COMMENTS
Plasmids		
pNIC-CTHF-PepT$_{St}$	*Streptococcus thermophilus*	POT
Tobacco Etch Virus (TEV) protease	Karolinska Institutet	Protease
pMES4y-Nb26	Immunized lamas	Nanobody
pTH24-DtpC	*E.coli*	POT
Bacterial cell strains		
E.coli DH5α	EMBL Hamburg	Cloning strain
E.coli C41 (DE3)	EMBL Hamburg	Expression strain
E.coli BL21 (DE3)	EMBL Hamburg	Expression strain
E.coli WK6	Universiteit Brussels	Nanobody expression strain

Table 9: Plasmids and bacterial cell strains used in the experiments described in the dissertation.

6.1.4 Computational resources

REAGENT or RESOURCE	SOURCE
Software and Algorithms	
XDS / XSCALE / XDSCONV (version January 26, 2018)	(Kabsch, 2010)
PHENIX (version 1.11.1-2575-000)	(Adams et al., 2010)
PHASER (version 2.1)	(McCoy et al., 2007)
COOT (version 0.8.2)	(Emsley et al., 2010)
MolProbity (in PHENIX package version 1.11.1-2575-000)	(Chen et al., 2010)
Consurf webserver	http://consurf.tau.ac.il/2016/
Open-Source PyMol™ (version 1.7.x)	Schrödinger LLC; www.pymol.org
Protein Data Bank (PDB)	www.pdb.org
GraphPad Prism (version 5.0f)	www.graphpad.com
LigPlot+ (version v.1.4.5)	(Laskowski et al., 2011)
SnapGene® Viewer (version 4.1.4)	www.snapgene.com/

Table 10: Computational resources used in the experiments described in the dissertation.

6.2 Buffers and media

6.2.1 Bacterial growth culture media

Luria-Bertani (LB) broth media	
LB broth low salt granulated	200 g
MilliQ water	Up to 1 L
Each liter of LB agar contains: 10 g Tryptone 5 Sodium chloride 5 g Yeast extract LB media was autoclaved prior usage.	

Table 11: LB media preparation.

Terrific broth (TB)	
TB modified granulated	47.6 g
100% Glycerol (v/ v)	4 ml
MilliQ water	Up to 1 L
Each liter of TB contains: 12 g Tryptone 24 g Yeast extract 9.4 g Potassium phosphate dibasic 2.2 g Potassium phosphate monobasic TB media was autoclaved prior usage.	

Table 12: TB media preparation.

To grow the bacterial cultures, LB or TB media were supplemented with the required antibiotic depending on the construct to express. For example, 30 μg/ml kanamycin were required to grow bacteria carrying the pNIC-CTHF-PepT$_{St}$ vector. The kanamycin antibiotic stock had a concentration of 30 mg/ml. To prepare the antibiotic stock, the lyophilized powder was weighted in an analytical balance and dissolved in the corresponding volume of ultrapure water to reach the desired concentration. The solution was sterile filtered using a 0.22 μm pore size filter. The filtering was performed either under a laminar flow hood or using sterile environment created close to a Bunsen burner. Finally, 500 μl aliquots were prepared and they were stored at -20°C until further use. Other antibiotics used in these studies to express different constructs were ampicillin (stock at 100 mg/ml) and carbenicillin (stock at 100 mg/ml). The protocol for the antibiotic stock preparation was the same to the one described for kanamycin. Antibiotics were diluted 1000-times to supplement the media.

LB-agar (Lennox)	
LB-agar	35 g
MilliQ water	Up to 1 L
Each liter of LB agar contains: 10 g Tryptone 5 g Yeast extract 5 g Sodium chloride 15 g Agar-agar pH adjusted to 7.0 ± 0.2 LB-agar was autoclaved prior usage.	

Table 13: LB-agar preparation.

To prepare the agar plates, the LB-agar was melted by heating in the microwave. It was left stirring with a magnetic stirrer to fully dissolve the agar and mix the components. Subsequently it was cool down prior addition of the corresponding antibiotics (final concentration of the antibiotics as previously stated). To avoid the antibiotic degradation, it was necessary to wait until the LB-agar had cooled to approximately 40°C. Typically, the glass bottle should be easily manipulated with ware hands. Once ready, the antibiotic-supplemented LB-agar was poured in Petri dishes in a sterile atmosphere (either using a Bunsen burner or under a laminar flow hood). LB-agar solidified as it cooled down. For long-term storage, plates were kept at +4°C until further use.

The procedure is the same when preparing plates supplemented with 5% sucrose (negative selection for the LIC procedure. See section 6.3.1.1 for more details). In this case, the sucrose was added at the same time as the antibiotic.

6.2.2 Buffers for DNA electrophoresis

50x Tris-Acetic-EDTA (TAE) buffer	
Tris base (MW: 121.1 g/ mol)	242 g
EDTA	18.6 g
Acetic acid (MW: 60.05 g/ mol)	57.1 ml
MilliQ water	Up to 1 L
Filter using an 0.45 μm pore size filter	

Table 14: 50x TAE buffer preparation.

5x Tris-Borate-EDTA (TBE) buffer	
Tris base (MW: 121.1 g/ mol)	54 g
Boric acid (MW: 61.83 g/ mol)	27.5 ml
0.5 M EDTA pH 8.0	20 ml
MilliQ water	Up to 1 L
Filter using an 0.45 μm pore size filter	

Table 15: 5x TBE buffer preparation.

Agarose gels were prepared weighting the agarose and dissolving it in 1x buffer, either TAE or TBE. To dissolve the agarose, the mixture was boiled in the microwave. The gel was cast in the molecular cloning lab. Few drops of ethidium bromide were added before the gel solidified. Agarose gels were run at 100 V for 40-50 minutes at room temperature. To prepare the samples, they were mixed with the loading dye (6x DNA Loading Dye, Thermo Fisher Scientific) to have a final dye concentration of 1x. The Gene Ruler 1 kb DNA Ladder was loaded on the gel as a reference.

1% agarose gels (w/ v)	
Agarose	1 g
1x TAE or TBE buffer	100 ml

Table 16: 1% agarose gels preparation.

6.2.3 Buffers for protein acrylamide gel electrophoresis (PAGE)

20x MES running buffer	
MES	195.2 g
Tris base	121.2 g
SDS	20 g
EDTA-free acid	6 g
MilliQ water	Up to 1 L
1x running buffer should have a pH of 7.3	

Table 17: 20x MES running buffer preparation.

For 140 µl of loading dye	
NuPAGE® LDS Sample Buffer (4X)	100 µl
0.5 M TCEP	20 µl
MilliQ water	20 µl

Table 18: SDS-PAGE sample loading dye preparation.

Membrane protein samples were mixed in a 1:3 dilution with the loading dye and incubated for 5 minutes at room temperature prior loading the gel. In the case of soluble proteins, samples were heated at 75°C for 5 minutes prior loading the gel.

Preparation of four polyacrylamide gels				
Required chemicals	Resolving gel			Stacking gel
	10%	12%	15%	
5x Bis-Tris buffer (pH 6.5- 6.8), ml	4	4	4	1.6
MilliQ water, ml	10.9	9.9	8.4	5.4
40% acrylamide, ml	5	6	7.5	1
20% APS, µl	100	100	100	40
TEMED, µl	20	20	20	20

Table 19: Polyacrylamide gel preparation.

Gels were cast under the fume hood due to the toxicity of some chemicals. Gels were run at constant voltage of 180 V for 40-45 minutes at room temperature. For staining the gels, 5-10 ml of InstantBlue™ were poured on the gels and they were incubated with the staining solution for 10 minutes on a shaking platform. Excess of stain was removed by washing the gel with water.

6.3 Methods

6.3.1 Molecular cloning

Full-length cDNA of $PepT_{St}$ wild type (WT) was already available in the lab and it was previously amplified from the *Streptococcus thermophilus* genome (Löw et al., 2013), and cloned into pNIC-CTHF vector (Figure 49) by a ligation-independent cloning (LIC) procedure (Aslanidis and de Jong, 1990). Protein expression in this vector is under the control of the T7 promoter and the *lac* operon. Therefore, protein expression can be induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture media. This vector contains a 6xHis-tag in C-terminus position and a Tobacco Etch virus (TEV) cleavage site. The vector carries a kanamycin resistance gene as selectable marker.

Furthermore, a blunt-end cloning procedure or Quikchange polymerase chain reaction (PCR) was used to generate the required mutants for this study.
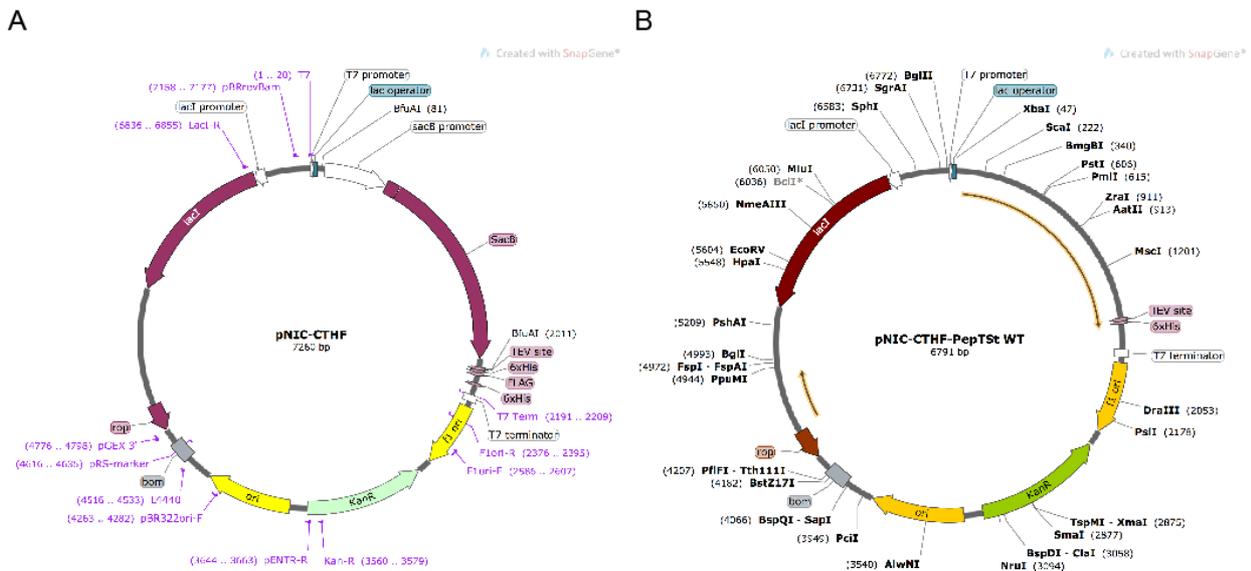
Figure 49: Vector map of pNIC-CTHF and pNIC-CTHF-PepT$_{St}$. (A) Empty pNIC-CTHF vector and (B) pNIC-CTHF-PepT$_{St}$ vector. The main features are indicated in the map, such as promoter and terminator, kanamycin resistance gene, *lac* operon and PepT$_{St}$ WT gene.

6.3.1.1 Ligation-independent cloning (LIC) procedure

In brief, the LIC protocol involves the amplification of the gene of interest and the plasmid using primers that overlap partially with them and have terminal complementary overhangs. This procedure involves the preparation of the plasmid and the amplification of the gene and the plasmid with the corresponding set of primers:

The pNIC-CTHF vector (Figure 49A) is treated and prepared to incorporate the gene of interest in a specific region, replacing the *sacB* gene. The *sacB* gene works as a negative selective marker. This gene is under the control of the *sacB* promoter. The *sacB* gene product is involved in sugar metabolism; it processes sucrose and generates levan, a toxic metabolite for *E.coli* that accumulates in the periplasm. To remove this gene, the plasmid is digested with the BfuAI restriction enzyme (see the plasmid digestion reaction in Table 20), obtaining a blunt-ended double strand linearized plasmid (Figure 50).

Figure 50: pNIC-CTHF vector and preparation for LIC procedure. Linearized pNIC-CTHF vector with the two BfuAI cleavage sites indicated by stripped lines and the scissor cartoons. After the digestion of the *sacB* gene by the BfuAI restriction enzyme, the *sacB* gene is removed from the vector and it is replaced by the gene of interest, in this case PepT$_{St}$.

pNIC-CTHF plasmid preparation	
Plasmid (~ 50 μg)	60 μl
10x NEB 3.1 buffer	10 μl
BfuAI restriction enzyme	3 μl
Nuclease-free water	Up to 100 μl
Incubation of the reaction at 50°C for 2-3 hours	
Heat inactivation of the enzyme at 65°C for 20 minutes	
Purification of the cleaved plasmid with the Qiagen PCR purification kit (plasmid elution in 55 μl of nuclease-free water)	

Table 20: pNIC-CTHF plasmid digestion.

Next, the plasmid is treated with the T4 DNA polymerase. This enzyme has the property of working as a polymerase or as a depolymerase depending on the deoxynucleotide (dNTP) availability in the reaction mixture; i.e., if there are dNTPs in the reaction mix, the polymerase will add them to the 3'

end of the DNA strand. If there are none in the medium, the DNA strand will be depolymerized from the 3' end until there are enough dNTPs free in solution and the equilibrium between the polymerization and depolymerization reaction is reached. In case of the plasmid, dCTP is added to the mixture to perform this reaction in a controlled manner. Therefore, the enzyme depolymerases the vector from the 3' ends until cytosine nucleotides are found. As a result, the vector has two sticky ends. The reaction with the T4 DNA polymerase is summarized in Table 21.

T4 DNA polymerase vector treatment	
BfuAI cleaved plasmid	55 µl
10x NEB2.1 buffer	10 µl
25 mM dCTP	10 µl
T4 DNA polymerase	5 µl
Nuclease-free water	Up to 100 µl
Incubation of the reaction at room temperature for 30 minutes Heat inactivation of the enzyme at 75°C for 20 minutes Addition of 100 µl of nuclease-free water to have a final volume of 200 µl. The vector can be stored at -20°C until further use	

Table 21: T4 DNA polymerase vector treatment.

The gene to be inserted in the vector requires some preparation too. First, it is amplified with primers carrying the corresponding overhangs. As a result of the amplification, a blunt end gene is obtained, which requires the T4 DNA polymerase treatment to generate sticky ends. The overhangs to be added to the gene primers are listed in Table 22.

Primer	Primer sequence (5' → 3')
Forward	TTA AGA AGG AGA TAT ACT
Reverse	GAT TGG AAG TAG AGG TTC TCT GC

Table 22: Overhangs for the primers required for the gene amplification.

In this case, the depolymerisation reaction is performed in the presence of dGTP, so that the reaction stops once a guanosine nucleotide is found. It is summarized in Table 23.

T4 DNA polymerase gene treatment	
Purified PCR product	5 µl
10x NEB2.1 buffer	1 µl
25 mM dGTP	1 µl
T4 DNA polymerase	0.5 µl
Nuclease-free water	Up to 25 µl
Incubation of the reaction at room temperature for 30 minutes Heat inactivation of the enzyme at 75°C for 20 minutes It is recommended to spin down the sample once it has cooled down	

Table 23: T4 DNA polymerase gene treatment.

Finally, both gene and plasmid are mixed in a 1 to 1 volume ratio and they are incubated for 15 minutes at room temperature. The plasmid is now ready for transformation into *E. coli* DH5α cells. Transformed cells are plated on LB-agar kanamycin + 5% sucrose plates, to confirm that the colonies that grow contain the pNIC vector (with a kanamycin-resistance gene as selectable marker) and do not contain the *sacB* gene. Grown colonies are screened and sequenced to identify the ones carrying the desired construct. Figure 51 summarizes the whole LIC procedure.
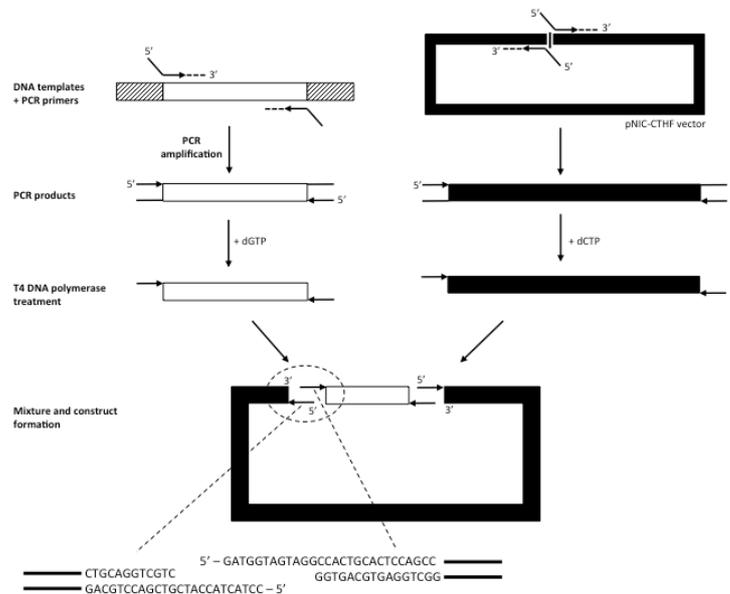
Figure 51: LIC protocol. The gene of interest and the digested vector are amplified using the designed primers carrying complementary overhangs. Next, the PCR products are treated with the T4 DNA polymerase to generate sticky ends. Finally, plasmid and gene are inclubated together. As the primer overhangs are complementary, the gene of interest is integrated in the plasmid, replacing the *sacB* gene position. Figure adapted from (Aslanidis and de Jong, 1990).

6.3.1.2 Polymerase chain reaction (PCR) for full vector amplification

In the strategies for introducing point mutations that are later explained (see sections 6.3.1.3 and 6.3.1.4), it is required to design a primer pair with the mutated codon and run a PCR reaction. The PCR reaction requires the following components (Table 24) and protocol (Table 25):

Master mix for PCR of 25 µl volume	
5x Phusion HF buffer	5 µl
10 mM dNTPs	0.5 µl
10 µM forward primer	1.25 µl
10 µM reverse primer	1.25 µl
100% DMSO (optional)	0.75 µl
DNA Phusion polymerase	0.25 µl
DNA template (< 250 ng)	1 µl (plasmid concentration ~ 100 ng/µl)
Nuclease-free water	Up to 25 µl

Table 24: Master mix for a PCR reaction of 25 µl volume.

PCR program for amplification of a pNIC-PepT$_{St}$ vector	
Temperature	*Duration of step*
98°C	1'
98°C	30"
ta (*)	30"
72°C	5-6'
72°C	10'
4°C	Hold

Table 25: PCR program for the amplification of the pNIC-CTHF-PepT$_{St}$ vector.

(*) ta: annealing temperature. This temperature was modified according to the primer pair.

After the PCR reaction and running an agarose gel to determine the success of the PCR, the amplified vectors of the correct size are digested with DpnI restriction enzyme. This incubation is performed at 37°C for at least one hour but preferably overnight, to make sure that the original methylated plasmid is digested. The newly synthetized plasmid does not contained methylated nucleotides, therefore is not digested by DpnI. Next, the two used procedures to introduced point mutations will be described.
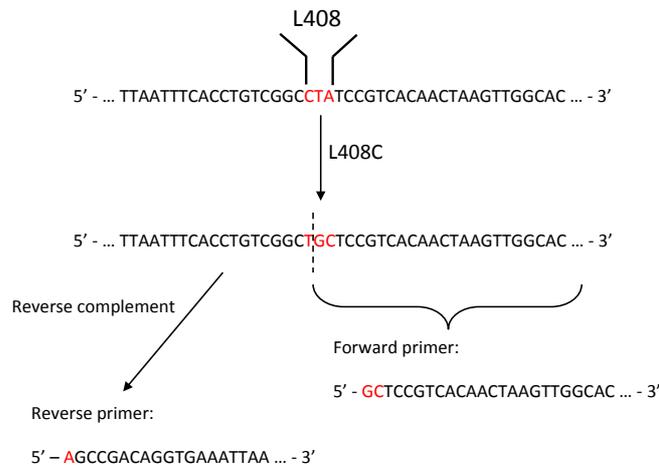
Figure 52: Example of primer design for a blunt end site-directed mutagenesis. In this case, nucleotides encoding for Leu-408 (CTA, highlighted in red) are mutated into the corresponding ones encoding for cysteine (TGC). Codon is divided between the forward primer and the reverse one. For the reverse primer, the reverse complement sequence is necessary.

6.3.1.3 Blunt-end protocol for site-directed mutagenesis

In a blunt-end PCR, the codon containing the point mutation is split between the two primers, meaning that two nucleotides are part of the sequence of a primer and the third nucleotide is part of

the other primer. One of the primers is amplifying the vector from the point where the mutation is inserted towards 3' (forward primer) and the second one, from that point towards 5' (reverse primer), but using the complementary strand for the amplification (Figure 52). As a result, a double-stranded blunt end linear vector is generated. To perform the ligation of the ends and obtain a circular plasmid, a phosphorylation step followed by a ligation step are required. After the PCR, an agarose gel is run with the samples and those with the correct size will be cleaned using a PCR product purification kit (elution of the product in 30 µl of nuclease-free water). To circularize the plasmid, the T4 polynucleotide kinase catalyzes the phosphorylation step by using ATP as a phosphorous source (Table 26), and the T4 DNA ligase is ligating both phosphorylated ends together (Table 27).

Phosphorylation step in a sample volume of 30 µl	
Purified PCR product	30 µl
10x polynucleotide kinase buffer	4 µl
10 mM ATP	4 µl
T4 polynucleotide kinase	1 µl
Nuclease-free water	Up to 40 µl
Incubation of the reaction at 37˚C for 1-2 hours	

Table 26: Phosphorylation reaction in a blunt-end PCR procedure.

Between the phosphorylation and the ligation step, the product is purified once again using a PCR product purification kit (elution of the product in 30 µl of nuclease-free water).

Ligation step in a sample volume of 30 µl	
Purified PCR product	30 µl
10X ligation buffer	4 µl
T4 polynucleotide kinase	2 µl
Nuclease-free water	Up to 40 µl
Incubation of the reaction at room temperature for 1-2 hours (or even overnight)	

Table 27: Ligation reaction in a blunt-end PCR procedure.

Finally, to test whether the mutation has been successfully introduced, *E. coli* DH5α competent cells are transformed. Several colonies growing in the agar plate will be subjected to a colony PCR (see section 6.3.1.5).

6.3.1.4 Quikchange PCR protocol for site-directed mutagenesis

Unlike in blunt-end cloning, the quikchange PCR involves the design of a set of primers that are overlapping completely or almost completely and where the codon containing the mutation is located in the center of the primers sequence (Figure 53). After the PCR and the digestion of the original plasmid with DpnI restriction enzyme, the PCR product can directly be used for the transformation of *E. coli* DH5α competent cells and proceed with the colony screening. A difference between the blunt-end and the quikchange PCR is that in the latter only the original plasmid is used as a template, as the generated copies have a nick in each DNA strand; it is repaired once the plasmid is used for transforming *E. coli* cells. In the blunt-end PCR, all the generated plasmid copies are used as template so the amount of DNA increases exponentially with each PCR cycle.
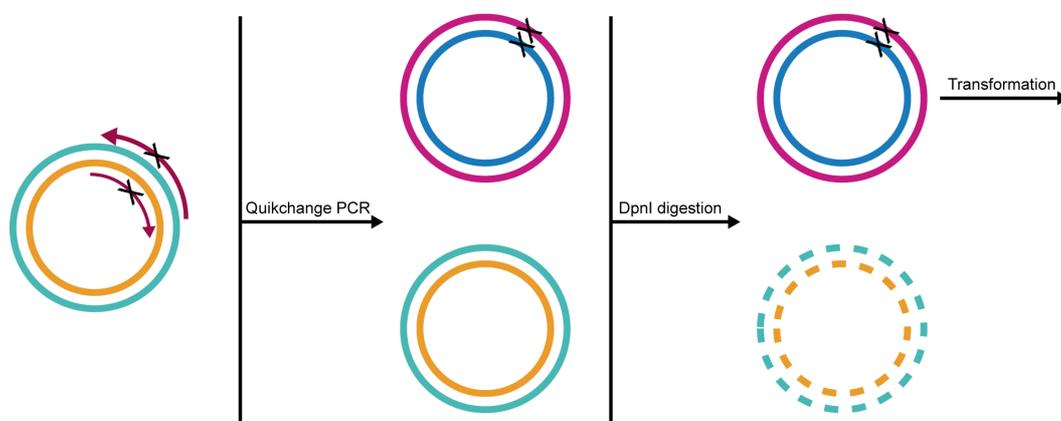
Figure 53: Quikchange PCR reaction. The overlapping primers (in red with a cross indicating the point where the mutation is located) anneal and amplify only the parental DNA, represented as the green and orange circles. The generated plasmid (in magenta and blue) has a nick. After digestion of the parental methylated plasmid, the amplified vector will be used for *E. coli* DH5α cells transformation, where the nick is repaired. Figure adapted from Agilent Technologies.

6.3.1.5 PCR for colony screening

The colony PCR is used to identify among the colonies that grew in the selective agar plate those carrying the gene of interest. Apart from the selected colonies, a control sample is also amplified (PepT$_{St}$ wild-type) and used as a reference for the products of the colonies PCRs on the agarose gel. Those colonies with the correct gene size are sent for sequencing to confirm that the mutation has been correctly inserted. The PCR reaction requires the following components (Table 28) and protocol (Table 29):

Master mix for a colony PCR of 25 μl volume	
10x Thermo buffer	2.5 μl
10 mM dNTPs	0.5 μl
10 μM forward primer (T7 promoter)	0.5 μl
10 μM reverse primer (T7 terminator)	0.5 μl
DNA Taq polymerase	0.125 μl
DNA template	1 colony (*)
Nuclease-free water	Up to 25 μl

Table 28: Colony PCR reaction master mix.

(*) To add the colony to the PCR reaction tube, the colony is picked from the plate with a tip and it is inserted into the PCR reaction tube. Afterwards, the tip is transferred to a well of a 24 well plate; on each well, 2 ml of LB supplemented with kanamycin (dilution 1:1000 from the stock) were previously added. The plate is incubated overnight in the 37°C shaker, so that a small culture is grown and the plasmid can be extracted from here in amounts in the order of 50-100 ng/ μl.

PCR program for amplification of PepT$_{St}$ gene	
Temperature	*Duration of step*
95°C	1'
95°C	30"
55°C (*)	30"
68°C (*)	1'
68°C	10'
4°C	Hold

Table 29: Colony PCR protocol.

(*) Both the annealing temperature for the primer pair (55°C) and the polymerization temperature (68°C) are optimized for the T7 primers and the Taq polymerase, respectively.

6.3.1.6 Sequencing of the clones

To confirm the correct introduction of the mutation, plasmids are sequenced from the T7 promoter and T7 terminator. We used the Eurofins Genomics sequencing service.

Primer	Primer sequence (5' → 3')
T7 promoter	TAA TAC GAC TCA CTA TAG GG
T7 terminator	CTA GTT ATT GCT CAG CGG T

Table 30: T7 promoter and T7 terminator primer sequences.

6.3.2 Protein expression

Protein expression requires the transformation of the construct in a suitable bacterial expression strain and the growth of cultures either in small- or large-scale, depending on the required amount of protein for the later experiments.

6.3.2.1 Transformation of the bacterial expression strain

The bacterial competent cells were transformed with the plasmid by a heat shock protocol, which involved a pre-incubation on ice of the cells (50 µl) with 1 µl of the plasmid at high concentration (~ 50 ng/ µl). Afterwards, the cells were placed at 45°C in a water bath for 42 seconds and placed on ice immediately after for a minute. 500 µl of Luria-Bertani broth (LB) media were provided to the cells and they were incubated for at least one hour at 37°C before spreading them on an agar plate enriched with the correct antibiotic or selective compounds. Only the cells successfully transformed with the vector would grow overnight.

6.3.2.2 Culture growth

To express the protein, a Luria-Bertani broth (LB) pre-culture was grown overnight at 37°C. The inoculation was done picking a colony from the agar plate. The LB was supplemented with the adequate antibiotic and additional chemical requirements if any. The LB pre-culture was used for the inoculation of the terrific broth (TB) cultures supplemented with the antibiotic, with a starting OD_{600} of 0.05. Cultures were grown at 37°C until induction with 200 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} of 0.8-1.0. After the induction, culture growth continued at 18°C for 16-18 hours. Cells were harvested by centrifugation (7000 rpm, 15 minutes, 4°C in JLA 8.1 rotor of Avanti JXN-26 centrifuge, Beckman Coulter) and the pellet was stored at -20°C until purification.

6.3.3 PepT$_{St}$ expression and purification protocol

6.3.3.1 PepT$_{St}$ expression

PepT$_{St}$ was expressed in *E. coli* C41(DE3) cells (Wagner et al., 2008). The pNIC-CTHF-PepT$_{St}$ has a kanamycin resistance gene as a selectable marker, so kanamycin was added at 30 µg/ ml in the LB pre-culture and the TB cultures. The transformation and culture growth protocols used were the same as the ones described in sections 6.3.2.1 and 6.3.2.2. Typically, PepT$_{St}$ purification started from a pellet coming from three liters of culture (~40-50 g of pellet).

6.3.3.2 Required buffers and stocks

- Lysis buffer: 20 mM NaP$_i$ pH 7.5, 300 mM NaCl, 5% glycerol, 15 mM Imidazol

- Wash buffer 1 or binding buffer: 20 mM NaP$_i$ pH 7.5, 300 mM NaCl, 5% glycerol, 15 mM Imidazol, 0.03% DDM, 0.5 mM TCEP

- Wash buffer 2: 20 mM NaP$_i$ pH 7.5, 300 mM NaCl, 5% glycerol, 30 mM Imidazol, 0.03% DDM, 0.5 mM TCEP

- Elution buffer: 20 mM NaP$_i$ pH 7.5, 150 mM NaCl, 5% glycerol, 250 mM Imidazol, 0.03% DDM, 0.5 mM TCEP

- Dialysis buffer: 20 mM Tris-HCl at pH 7.5 (pH adjusted at room temperature), 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, 0.03% DDM

- Gel filtration buffer: 20mM Tris-HCl pH 7.5 (RT), 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, 0.03% DDM.

- Lysozyme stock 100 mg/ ml

- TCEP stock 0.5 M

- DNases stock 2000 U/ ml

- DDM recommended stock 3-10%

- LMNG stock at 5%

- NM stock 4-10%

- TEV protease stock 1 mg/ ml, aliquots of 500 μl stored at -80°C

6.3.3.3 Purification protocol

Protein purification was performed as previously described (Löw et al., 2013; Quistgaard et al., 2017; Martinez Molledo et al., 2018). In brief, the pellet was resuspended in lysis buffer (5 ml of lysis buffer per gram of wet weight pellet), supplemented with 1 mg/ml of lysozyme, protease inhibitors (one tablet of cOmplete™ EDTA-free Protease Inhibitor Cocktail per 100 ml lysis buffer). The pellet was stirred in the cold room for 30 minutes, until it was resuspended. Then, DNases (5 U/ml final concentration) and 0.5 mM TCEP were added. The suspension was stirred in the cold room for 30 minutes to 1 hour more. Cell lysis was performed by three cycles in the EmulsiFlex-C3 (Avestin) with 10000-15000 psi of pressure. Recovered material was centrifuged to remove non-lysed cells (7000 rpm, 15 minutes, 4°C in JLA 8.1 rotor of Avanti J-20 XP centrifuge) and the recovered supernatant was subjected to an ultracentrifugation to separate the membrane fraction (35000 rpm, 1 hour, 4°C in 45 Ti rotor of Optima XE-90 centrifuge, Beckman Coulter). At this point, the membranes were recovered as a pellet and they were resuspended in lysis buffer supplemented with protease inhibitors and they were solubilized by adding 1% n-dodecyl-β-D-

maltoside (DDM) detergent. Membranes were stirred in the cold room for 40 minutes to 1 hour until the solution turned clear, a sign that membranes have been solubilized. A second ultracentrifugation was performed (30000 rpm, 1 hour, 4°C in 45 Ti rotor of Optima XE-90 centrifuge, Beckman Coulter) to remove non-solubilized material.

Detergent-solubilized PepT$_{St}$ was firstly purified by immobilized-metal affinity chromatography (IMAC) on gravity column. As a rule of thumb, two ml of seated beads (HisPur™ Ni-NTA resin, Thermo Scientific) were used per liter of purified culture. The beads were pre-equilibrated in lysis buffer supplemented with 0.03% DDM and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP) and they were incubated with the solubilized PepT$_{St}$ for one hour at 4°C in a rotating wheel. Loaded beads were poured onto the gravity columns and extensively washed with wash buffer 1 (typically, two washes of 15 ml per column) and wash buffer 2. The protein was eluted from the column with a buffer with high imidazole concentration (elution buffer) and combined with 1 ml of TEV protease at 1 mg/ml to perform the His-tag cleavage during dialysis overnight at 4°C. Typically, 1 mg of TEV protease was sufficient to cleave the purified protein from 3 liters of culture.

Cleavage was successful up to 90% and the cleaved protein was recovered by negative IMAC. A second purification step was done by size-exclusion chromatography (SEC). Cleaved protein was concentrated to 5 ml in 100 kDa concentrator (Corning® Spin-X® UF concentrators) and run in an ÄKTA Pure system (GE Healthcare Life Sciences), using a HiLoad 16/ 600 Superdex 200 column (GE Healthcare Life Sciences). Fractions containing the protein were pooled together and concentrated to 10 mg/ml, flash frozen and stored at -80°C until further use.

PepT$_{St}$ WT was also purified in different detergents to study the oligomerization state of the protein in solution and its thermostability. In this case, the purification procedure was identical to the one previously described with the exception of the detergent added to each buffer. In brief, the purification with 2,2-didecylpropane-1,3-bis-β-D-maltopyranoside (LMNG) involved the use of 1% detergent for membrane solubilisation and buffers were supplemented with 0.01%. In the purification with n-decyl-β-D-maltopyranoside (DM) and n-nonyl-β-D-maltoside (NM), the membrane solutibilisation was performed by adding 1% DM and the buffers used for the IMAC purification were supplemented with 0.1% DM. In the SEC step, the gel filtration buffer contained 0.1% DM in the case of the protein purified in DM, but was replaced with 0.4% NM for the protein purified in NM. As the micelle size in this last case is smaller, the concentration was performed using 50 kDa concentrators (Corning® Spin-X® UF concentrators). As before, the protein was flash frozen and stored at -80°C until further use.

6.3.4 Peptide stock preparation

Peptides were mainly purchased from Sigma-Aldrich and Bachem, except for tripeptide with sequence Phe-Ala-Xxx (where X stands for any amino acid), which were chemically synthetized by GL Biochem (Shanghai). Peptide stocks were prepared by weighing the lyophilized powder in an analytical balance and dissolving them in ultrapure water at the highest possible concentration. The approach published by Huang and colleagues was used for estimating the solubility of the peptides (Huang et al., 2012). For those peptides not soluble in water, 100% dimethyl sulfoxide (DMSO) was used instead.

6.3.5 PepT$_{St}$ stability characterization by Differential Scanning Fluorimetry (DSF)

PepT$_{St}$ transition midpoint (Tm) for thermal unfolding was determined with the nanoDSF Prometheus NT.48 devise (NanoTemper technologies). An advanced DSF technique with a capillary-based setup was used to monitor the protein intrinsic fluorescence upon temperature-induced denaturation. The fluorescence intensity of tryptophan and tyrosine residues in a protein strongly depends on the environment. As the protein unfolds, the change in the fluorescence is recorded at 330 nm and 350 nm wavelengths over the temperature ramp (from 20 to 90 °C, with one degree of temperature increasing per minute). The nanoDSF analysis software automatically determines the Tm calculating the first derivative of the unfolding curve.

To load a capillary, the required sample volume was 10 μl with a protein concentration of 0.5 mg/ml. For this experiment, PepT$_{St}$ was solubilized in n-nonyl- β-D-maltoside (NM), as the effect of peptide addition was more significant when using this detergent than when using the markedly more stabilizing n-dodecyl-β-D-maltoside (DDM) (Quistgaard et al., 2017). Additionally, PepT$_{St}$ Tm was measured in the presence of a library of peptides. The principle here is to measure the stabilization effect of the peptides on the transporter upon heat unfolding (Niesen et al., 2007). In the initial screen, the final peptide concentration was 5 mM but for a selected group of peptides, measurements were also carried out at additional concentrations in the range of 1.25-20 mM. Experiments were performed in triplicates and the standard deviation for each peptide was calculated. The unfolding curves were plotted using GraphPad Prism (GraphPad Software, San Diego, CA). A Student's t-test was performed to determine significant differences between the control sample and those in presence of a peptide at 5 mM. The level of significant difference between the control and the rest of the samples is indicated by * ($P=<0.05$) for significantly different samples, and ** for highly significant differences ($P=<0.01$).

6.3.6 PepT$_{St}$ binding studies *in vitro* by MicroScale Thermophoresis (MST)

PepT$_{St}$ peptide binding in solution was measured with the Monolith NT.LabelFree devise (NanoTemper technologies) (Seidel et al., 2013). Microscale thermophoresis (MST) is a technique in which the intrinsic fluorescence of a protein or a labeled protein (the protein can also be labeled with a fluorophore) is recorded as it moves in a temperature gradient generated in a small sample volume. The experimental procedure involves the preparation of 16 different samples, where the protein concentration was kept constant (125 nM) in all of them, whereas the peptide was added in 16 different concentrations in the mM range in a dilution series.

The measurement starts with the readout of the fluorescence in each individual capillary. Then, an infrared laser induces the temperature gradient; the greater the distance from the point where the laser hits the capillary, the lower the temperature is. Molecules typically diffuse from the highest temperature point to the lower ones. The fluorescence of the sample is recorded in the same point the laser induces the hot spot in the capillary. Therefore, in this point, the fluorescence is progressively decreasing as the molecules diffuse away. In a typical MST experiment, the laser is on for 30 seconds. Once it turns off, molecules diffuse back as the temperature gradient vanishes. The movement of the protein or the complex (protein-ligand) in this temperature gradient is different depending on the molecular weigh, the charge and the hydration shell. This procedure is repeated for the 16 capillaries. To represent the data and obtain the dissociation constant (K$_D$), the normalized fluorescence of the thermophoresis event in each capillary is plotted as a single point against the ligand concentration in logarithmic scale.

Ligand K$_D$ values were calculated with the following equation:

$$Kd = Unbound + \frac{(Bound - Unbound) \times (Concentration + TargetConc + Kd - \sqrt{(Concentration + TargetConc + Kd)^2 - 4 \times Concentration \times TargetConc})}{2 \times TargetConc}$$

In this equation, the term 'Concentration' corresponds to the ligand concentration, 'TargetConc' to PepT$_{St}$ protein concentration and 'Unbound' and 'Bound' to the capillary fluorescence before the temperature gradient is induced and 30 seconds after, respectively.
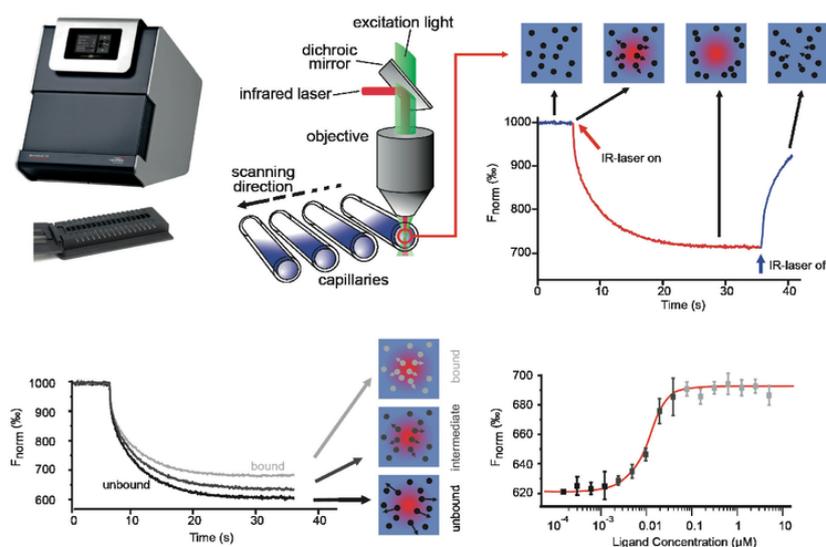
Figure 54: Microscale thermophoresis (MST) protocol. Samples are loaded in the glass capillaries and the infrared laser induces the temperature gradient. The sample fluorescence is recorded in the same point where the laser heats the capillary. Afterwards, the normalized fluorescence is plotted against the ligand concentration in logarithmic scale to calculate the dissociation constant (K_D). Figure from NanoTemper technologies.

For the ligand binding experiments in $PepT_{St}$, 16 different samples were prepared for each measurement. Here $PepT_{St}$ protein stock was diluted in different buffers depending on the condition to explore. It should be noted that the ligands were used in high millimolar concentrations. In order to keep the pH of the solution constant regardless the ligand concentration, the buffer used in the measurement had either a concentration of 100 mM (for the measurements with dipeptides) or 400 mM (for the measurements with tripeptides Phe-Ala-Xxx, which presented a pH of 2.0-2.5 in water-solubilised form).

Then, the samples were individually loaded in NT.LabelFree standard glass capillaries and placed horizontally on the device tray. Measurements were performed at 22°C. LED power was set to 20%, and MST power to 20% and 40%. MST data was exported and the binding curves were plotted using GraphPad Prism.

6.3.7 PepT$_{St}$ crystallization

All structures reported in this thesis were obtained from crystals grown using the lipidic cubic phase (LCP) method (Aherne et al., 2012). In brief, $PepT_{St}$ protein (10 mg/ml) was mixed with lipid 1-(7Z-pentadecenoyl)-rac-glycerol (MAG 7.8, Avanti Lipids) in a one to one volume ratio, using two

coupled Hamilton syringes. To facilitate the manipulation of the lipid, it was placed in a thermoblock at 45°C for 3 minutes. Once the lipid was fluid, it could be pipetted in a pre-warmed syringe. The protein was pipetted into a second syringe but at room temperature. The syringes were carefully connected avoiding air bubbles using a mosquito LCP syringe coupler (TTP Labtech) and the content was manually mixed until the mesophase reached homogeneity, it was optically clear and it remained stable over time (Figure 55).
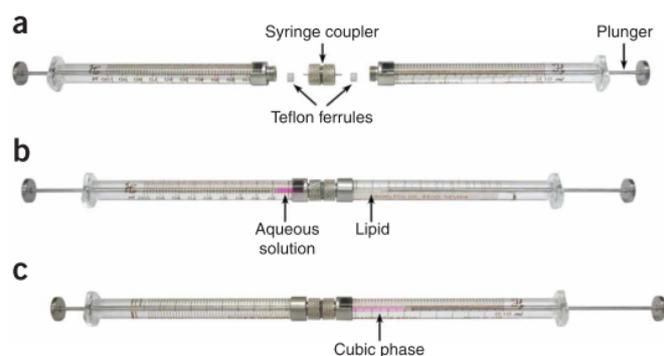
Figure 55: Lipidic cubic phase (LCP) syringes. (A) Two syringes connected by a coupler. The teflon ferules avoid leakage when the content of the syringes is mixed. (B) The lipid is placed on a pre-warmed syringe (right) and the protein in the second one (left), at room temperature. The syringes are connected using a coupler and (C) pushing the plunger of one syringe towards the other one mixes the content. This operation is repeated until the content is homogeneous and optically clear. Figure from (Caffrey and Cherezov, 2009).

Crystallization plates were set up using the Mosquito-LCP robot (TTP Labtech). The used protocol dispensed 50 nl of mesophase in the wells of one row of the plate and straight after, the mesophase was covered with 800 nl of precipitant solution. The plates used for crystallization were the Laminex™ UV Plastic Bases with wells of 100 μm depth (Molecular Dimensions). The plates were manually sealed with plastic covers Laminex™ UV Plastic 200 micron Film Covers (Molecular Dimensions), which were 200 μm thickness and UV compatible. Plates were stored at 19°C in a Rock Imager (Formulatrix) for automatic inspections.

Crystals grew within 24 hours and the size increased during the following 3 days. Crystals were harvested 7-10 days after appearing, flash frozen and stored until screened at a synchrotron beamline.

$PepT_{St}$ apostructure ($PepT_{St}$ [apo]) crystals grew in 0.1 M citrate buffer pH 4.5 and 15-30% PEG 300. Crystals of $PepT_{St}$ in complex with didpetides ($PepT_{St}$ [Ala-Leu], $PepT_{St}$ [Ala-Gln], $PepT_{St}$ [Asp-Glu], $PepT_{St}$ [Phe-Ala]) as well as the complex with HEPES at different concentrations ($PepT_{St}$ [100 mM HEPES] and $PepT_{St}$ [300 mM HEPES]) grew in 0.1 M HEPES pH 7.0, 15-25% PEG 400 and

0.15-0.55 M ammonium phosphate monobasic ($NH_4H_2PO_4$), the same crystallization condition reported for the complex PepT$_{St}$[Ala-Phe] (Lyons et al., 2014). Different strategies were tested to obtain complexes of PepT$_{St}$ and the dipeptides of interest: Ala-Leu and Phe-Ala were added to the screen in concentrations of 5 mM and 30 mM, respectively. Ala-Gln was premixed with PepT$_{St}$, reaching a final concentration of 100 mM, and once the mesophase was produced, it was incubated at room temperature for one hour. For Asp-Glu, a dry-coating approach was used (Gelin et al., 2015): Each well of the crystallization plate was coated with 1 µl of 10 mM Asp-Glu, and the drops were then left drying for two days at room temperature before setting up the screen. The crystallization with tripeptides Phe-Ala-Ala, Phe-Ala-Thr and Phe-Ala-Gln involved a tight control of the pH in the crystallization condition. These peptides were chemically synthetized and purified by HPLC; the peptide was eluted in a trifluoroacetic acid gradient. As a result, when the peptides were dissolved in ultrapure water, the pH of the solution was 2.0-2.5. As the pH was the key parameter to obtain PepT$_{St}$ -peptide complexes, the crystallization screen contained a concentration range of HEPES buffer at pH 7.5 from 0.1-0.3 M, 15-25% PEG 400 and 0.15-0.55 M $NH_4H_2PO_4$.

6.3.8 Data collection

Crystal screening and data collection were done at the EMBL P13 and P14 beamlines at the PETRA III storage ring (c/o DESY, Hamburg, Germany) and ID30A and B at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). A single crystal was enough to collect a complete data set. Typical data collection strategy was 0.1-0.2° oscillation range, 0.05-0.1 seconds exposure time and 20-40% beam transmission. Either 1800 or 3600 images were collected, depending on crystal sensitivity to radiation damage.

6.3.9 Data processing and structure determination

Typically, membrane protein crystals grown in vapor diffusion tend to have highly anisotropic diffraction patterns as they present a crystal packing type II. Our crystals were grown by the LCP method and therefore, crystals had a type I packing resulting in more isotropic diffraction. As a result, no anisotropy correction was required to process the data.

The data were indexed, reduced and scaled using the XDS program suite (Kabsch, 2010). The XDS program runs several jobs sequentially to generate a list of unique reflections. From the collected intensities, it identifies equivalent intensities due to the crystal symmetry and it averages them. Each intensity value has an error associated to it (R factor). At this point, the data set was be evaluated in

terms of completeness, $I/\sigma(I)$ and CC1/2 values to cut the resolution. The maximum resolution at which to cut the data was decided based on completeness and the CC1/2 correlation coefficient according to the definition of Karplus and Diederich (Karplus and Diederichs, 2012, 2015). Therefore, we typically used a CC1/2 value of 0.5 to cut the data set. For all the structures reported here, a single crystal was sufficient to reach completeness at the highest resolution shell. Only in the case of PepT$_{St}$[Phe-Ala], three complete data sets were merged, which improved the density for the ligand considerably. Regardless of the buffer system used, the crystals consistently presented the same space group as previously described by Lyons *et al.* (Lyons et al., 2014), and similar unit cell dimensions.

Phase estimation was performed by molecular replacement (MR) with Phaser molecular replacement program from the PHENIX suite (Adams et al., 2010; McCoy et al., 2007). As a reference model, we used the previously reported apostructure (Lyons et al., 2014; PDB ID 4D2B). A good solution was found as indicated by the log-likelihood gain (LLG) and translation-function Z-score (TFZ) parameters (Oeffner et al., 2013). Refinement was then carried out through several cycles of manual rebuilding in Coot and maximum likelihood refinement in PHENIX version 1.9_1692, where translation libration screw (TLS) was used (Adams et al., 2010). Ligands in the model, such as di- and tripeptides and MAGs molecules, were manually placed. The final models were validated using MolProbity (Chen et al., 2010). If necessary, the indicated outliers were fixed. Finally, the models were submitted to the Protein Data bank (PDB).

Additional programs used for analyzing the structure or preparing the figures were Consurf, Pymol and LigPlot[+]. ConSurf was used for analyzing sequence conservation and mapping it on the structures (http://consurf.tau.ac.il/2016/) (Ashkenazy et al., 2016). In total, 150 sequences that are 35 – 85% identical to PepT$_{St}$ were used in this analysis. All omit maps presented in the results section were generated by repeating the last round of refinement after setting the occupancies of the ligand to zero in the input PDB file. PyMol was employed for generating structural overlays and making structure figures (Schrödinger LLC; http://www.pymol.org). LigPlot[+](Laskowski et al., 2011) was used for representing the ligand interactions to the residues in the binding cavity of PepT$_{St}$.

6.3.10 PDB accession codes

Coordinates and structure factors for PepT$_{St}$ apostructure and PepT$_{St}$-peptide complexes have been deposited in the PDB with the following accession numbers:

Deposited data	PDB ID
Crystal structure of PepT$_{St}$[Phe-Ala]	5OXN
Crystal structure of PepT$_{St}$[100 mM HEPES]	6EIA
Crystal structure of PepT$_{St}$[300 mM HEPES]	5OXQ
Crystal structure of PepT$_{St}$[phosphate]	5OXP
Crystal structure of PepT$_{St}$[apo]	5OXO
Crystal structure of PepT$_{St}$[Ala-Leu]	5OXL
Crystal structure of PepT$_{St}$[Ala-Gln]	5OXK
Crystal structure of PepT$_{St}$[Asp-Glu]	5OXM
Crystal structure of PepT$_{St}$[Phe-Ala-Ala]	Data not deposited
Crystal structure of PepT$_{St}$[Phe-Ala-Thr]	Data not deposited
Crystal structure of PepT$_{St}$[Phe-Ala-Gln]	6GHJ

Table 31: PDB accession codes.

6.3.11 Thiol-maleimide 'click' reaction for free cysteine residue detection

PepT$_{St}$ double cysteine mutants were subjected to a maleimide-labeling assay to determine the accessibility of the thiol group in the cysteine residues. The maleimide derivative methoxypolyethylene glycol maleimide 5000 (mPEG-5K) was used for this assay. Reaction of mPEG-5K to the thiol groups generated conjugates of higher molecular weight compared to the native protein. These conjugates were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The experiment was performed under denaturing and non-denaturing conditions, in an oxidizing or reducing environment. The samples were denatured by incubation of one hour in the presence of 1% SDS; for reducing the samples, 0.5 mM TCEP was added and the samples were incubated for 30 minutes at room temperature. Finally, the samples were treated with 1 mM mPEG-5K for 45 minutes at room temperature, and analyzed by SDS-PAGE.
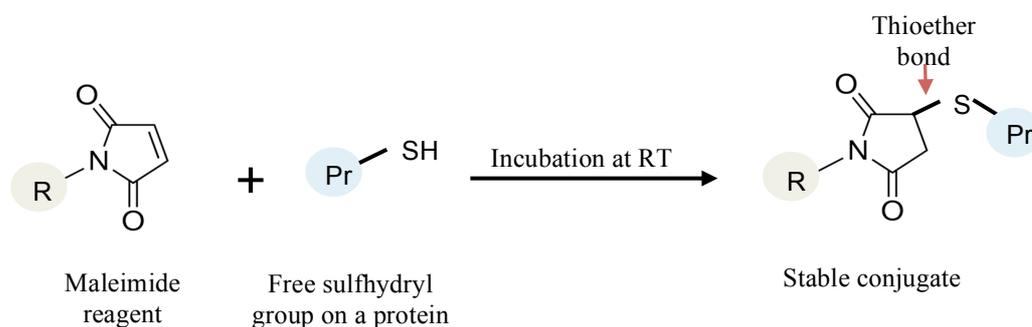
Figure 56: Thiol-maleimide 'click' reaction for free cysteine residue detection. Maleimide derivative used (left) for detecting the free thiol groups in the protein. The resulting stable conjugate carries a thioester bond between the sulfur atom of the thiol group of the protein and the maleimide ring. This conjugate has a molecular weight of 5 kDa more compared to the native protein. The separation of both species is possible by SDS-PAGE.

6.3.12 Analytical gel filtration

PepT$_{St}$ WT and mutants were run in an analytical gel filtration column as a quality control. Here the Agilent system for high-throughput sample screening was used. In brief, the samples were pipetted in a 96-well plate from where they were automatically injected. Typically, samples were run in duplicates; therefore 60 µl of protein at 0.5-1 mg/ml were pipetted into a well and two injections of 20-25 µl were performed in two consecutive runs. A 4 ml Superdex200 shelf-packed column was used in these experiments. As a running buffer, a 20mM Tris-HCl pH 7.5 (RT), 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, 0.03% DDM was used. Gel filtration runs were compared to the WT protein, used as a reference.

6.3.13 DtpC-Nanobody 26 complex expression and purification

The *E.coli* POT DtpC was purified in complex with the nanobody 26 (Nb26) for crystallization purposes. This experiment was carried out as a small side project together with Dr. Anna Sophie Zimmermann and using the protocol established by Yonca Ural-Blimke in our lab.

6.3.13.1 Protein expression

DtpC was expressed in *E. coli* C41(DE3) cells. It was cloned in a pTH24-gateway vector that has an ampicillin resistance gene as a selectable marker and two His-tags, at N- and C-terminus but only the one at C-terminus is cleavable by the TEV protease. For the culture growth, ampicillin was added at 100 µg/ ml in the LB pre-culture and the TB cultures. The transformation and culture growth protocols used were the same as the ones described in sections 6.3.2.1 and 6.3.2.2. For DtpC purification, we typically started from three liters of culture pellet (~40-50 g of pellet).

Nanobody 26 was cloned in a pMESy4 vector for periplasmic protein expression. It has a C-terminal EPEA-tag and a carbenicillin resistance gene. The protein was expressed in *E. coli* WK6 cells (Pardon et al., 2014). The protein expression is under the *lac* promoter, which is repressed by glucose and induced by IPTG. For the overnight LB pre-culture, the media was supplemented with carbenicillin at 100 µg/ ml, 2% glucose and 1 mM MgCl$_2$. The TB cultures were supplemented with carbenicillin at 100 µg/ ml, 0.1% glucose and 2 mM MgCl$_2$ and they inoculated with cells at a starting OD$_{600}$ of 0.05. Cultures were grown at 37°C until induced with 0.2 mM IPTG, at an OD$_{600}$ of 0.6-0.8. Then, the growth continued for 16-20 hours at 28°C.

6.3.13.2 Required buffers and stocks

For the Nb26 purification, the following buffers are required:

- TES buffer: 0.2 M Tris pH 8.0, 0.5 mM EDTA, 0.5 M sucrose

- Buffer A: 20 mM HEPES pH 7.5, 100 mM NaCl

- Elution buffer: 20 mM HEPES pH 7.5, 100 mM NaCl, 2 M MgCl$_2$

For the DtpC-Nb26 complex, the following buffers are required:

- Lysis buffer: 20 mM NaP$_i$ pH 7.5, 300 mM NaCl, 15 mM imidazol, 5% glycerol

- Buffer A: 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 0.03% DDM

- Elution buffer: 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 0.03% DDM, 2 M MgCl$_2$

- SEC buffer: 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 0.03% DDM, 5% glycerol

Additionally, the following stocks are also required in the purification of Nb26, the complex or both:

- Lysozyme stock 100 mg/ ml

- TCEP stock 0.5 M

- DNases stock 2000 U/ ml

- DDM recommended stock 3-10%

- TEV protease stock 1 mg/ ml, aliquots of 500 µl stored at -80°C

6.3.13.3 Purification protocol for Nb26

Nanobodies were expressed in the periplasm. To simplify the purification protocol, the cells were subjected to an osmotic shock in a way that the bacterial outer membrane was broken releasing the periplasmic content, but the plasma membrane remained intact. For this purpose, the cell plasma pellet was resuspended in TES buffer (5 ml of buffer per gram of cell pellet) and stirred in the cold room for one hour. This step allowed the disruption of the outer membrane. Later, the same volume of buffer was added but four times diluted in water, to avoid the breakage of the cell membrane. Stirring continued for ~ 1-2 more hours. The suspension was centrifuged to pellet the cells (7000 rpm, 45 minutes, 4°C in JLA 8.1 rotor of Avanti J-20 XP centrifuge), whereas the nanobodies were recovered in the supernatant. The periplasmic extract was loaded in a CaptureSelect affinity column, a shelf-packed column with beads that have a covalently bound antibody that specifically recognizes the EPEA-tag at the C-terminus of the nanobodies. The elution was possible in a MgCl$_2$ gradient. The eluted fractions were analyzed by SDS-PAGE and those fractions containing the nanobody were pooled and dialyzed overnight against buffer A. The day after, the dialyzed product was concentrated in a 5 kDa MWCO concentrator to 5 ml and run in SEC using the HiLoad Superdex75 column. Buffer A was used for the SEC. Protein purity was

analyzed by SDS-PAGE and the fractions containing pure Nb26 were pooled, concentrated to ~2-3 mg/ml, flash frozen and stored at -80°C until further use.

6.3.13.4 Purification protocol for DtpC in complex with Nb26

For the purification of DtpC-Nb26 complex, the previously purified Nb26 was required. DtpC was purified using the same protocol as the one explained for PepT$_{St}$ in section 6.3.3.3 until the membrane solubilization and second ultracentrifugation step. At that point, the detergent-solubilized DtpC was loaded in the CaptureSelect column, which was pre-loaded with Nb26. Typically, for purifying DtpC coming from three liters of culture, 3-4 mg of Nb26 should be bound to the column. As Nb26 binds specifically to DtpC, only DtpC remained bound in the column while the rest of the components in the solubilized membrane fraction were eluted directly. After a washing step with buffer A (in this case, supplemented with 0.03% DDM), the DtpC-Nb26 complex was eluted in a MgCl$_2$ gradient. The eluted fractions were analyzed by SDS-PAGE and those fractions containing the complex were pooled and dialyzed overnight against buffer A. Furthermore, in the dialysis, the His-tag of DtpC at C-terminus was cleaved by addition of TEV protease (1 mg of TEV protease was used for the purification of DtpC coming from three liter of culture). The day after, the dialyzed product was concentrated in a 100 kDa MWCO concentrator to 5 ml and run in SEC using the HiLoad Superdex200 column. As the complex stoichiometry is 1:1 and the Nb26 was loaded in excess for the CaptureSelect affinity column, this excess of Nb26 was coming at the end of the SEC column. The purity of the complex was analyzed by SDS-PAGE and the fractions containing the complex were pooled, concentrated to ~8-10 mg/ml, flash frozen and stored at -80°C until further use, or used directly for setting crystallization trials by vapor diffusion.

Chapter 7. References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 213-221.

Aherne, M., Lyons, J.A., Caffrey, M. A fast, simple and robust protocol for growing crystals in the lipidic cubic phase. *J. Appl. Crystallogr.* **2012**, 45, 1330-1333.

Alper, S.L., Sharma, A.K. The SLC26 gene family of anion transporters and channels. *Mol. Aspects Med.* **2013**, 34, 494-515.

Arachea, B.T., Sun, Z., Potente, N., Malik, R., Isailovic, D., Viola, R., E. Detergent selection for enhanced extraction of membrane proteins. *Protein Expr. Purif.* **2012**, 86, 12-20.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **2016**, 44, W344-W350.

Aslanidis, C., de Jong, P.J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **1990**, 18, 6069-6074.

Augustin, R. The protein family of glucose transport facilitators: it's not only about glucose after all. *IUBMB Life* **2010**, 62, 315-333.

Beale, J.H., Parker, J.L., Samsudin, F., Barrett, A.L., Senan, A., Bird, L.E., *et al.* Crystal structures of the extracellular domain from PepT1 and PepT2 provide novel insights into mammalian peptide transport. *Structure* **2015**, 23, 1889-1899.

Beis, K. Structural basis for the mechanism of ABC transporters. *Biochem. Soc. Trans.* **2015**, 43, 889-893.

Bernhard. W. Lung surfactant: Function and composition in the context of development and respiratory physiology. *Ann. Anat.* **2016**, 208, 146-150.

Biegel, A., Knütter, I., Hartrodt, B., Gebauer, S., Theis, S., Luckner, P., *et al.* The renal type H$^+$/peptide symporter PEPT2: structure-affinity relationships. *Amino Acids* **2006**, 31, 137-156.

Birch, J., Axford, D., Foadi, J., Meyer, A., Eckhardt, A., Thielmann, Y., Moraes, I. The fine art of integral membrane protein crystallisation. *Methods* **2018**.

Boggavarapu, R., Jeckelmann, J.-M., Harder, D., Ucurum, Z., Fotiadis, D. Role of electrostatic interactions for ligand recognition and specificity of peptide transporters. *BMC Biol.* **2015**, 13, 58.

Boivin, S., Kozak, S., Rasmussen, G., Nemtanu, I.M., Vieira, V., Meijers, R. An integrated pipeline for sample preparation and characterization at the EMBL@PETRA3 synchrotron facilities. *Methods* **2016**, 95, 70-77.

Brandsch, M. Drug transport via the intestinal peptide transporter PepT1. *Curr. Opin. Pharmacol.* **2013**, 13, 881-887.

Brandsch, M., Knutter, I., Bosse-Doenecke, E. Pharmaceutical and pharmacological importance of peptide transporters. *J. Pharm. Pharmacol.* **2008**, 60, 543-585.

Breyton, C., Pucci, B., Popot, J.-L. Amphipols and fluorinated surfactants: two alternatives to detergent for studying membrane proteins. *Methods Mol. Biol.* **2010**, 601, 219-245.

Caffrey, M. A comprehensive review of the lipid cubic phase or *in meso* method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr. F Struct. Biol. Commun.* **2015**, 71, 3-18.

Caffrey, M., Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **2009**, 4, 706-731.

Casagrande, F., Harder, D., Schenk, A., Meury, M., Ucurum, Z., Engel, A., *et al.* Projection structure of DtpD (YbgH), a prokaryotic member of the peptide transporter family. *J. Mol. Biol.* **2009**, 394, 708-717.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 12-21.

Chen, X.Z., Steel, A., Hediger, M.A. Functional roles of histidine and tyrosine residues in the H(+)-peptide transporter PepT1. *Biochem. Biophys. Res. Commun.* **2000**, 272, 726-730.

Cherezov, V., Liu, J., Griffith, M., Hanson, M.A., Stevens, R.C. LCP-FRAP assay for pre-screening membrane proteins for *in meso* crystallization. *Cryst. Growth Des.* **2008**, 8, 4307-4315.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T. S., *et al.* High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, 318, 1258-1265.

Chiang, C.S., Stacey, G., Tsay, Y.F. Mechanisms and functional properties of two peptide transporters, AtPTR2 and fPTR2. *J. Biol. Chem.* **2004**, 279, 30150-30157.

Chun, E., Thompson, A.A., Liu, W., Roth, C.B., Griffith, M.T., Katritch, V., Kunken, J., Xu, F., Cherezov, V., Hanson, M.A., Stevens, R.C. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* **2012**, 20, 967-976.

Cipriani, F., Röwer, M., Landret, C., Zander, U., Felisaz, F., Márquez, J.A. CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, 68, 1393-1399.

Clark, K.M., Fedoriw, N., Robinson, K., Connelly, S.M., Randles, J., Malkowski, M.G., DeTitta, G.T., Dumont, M.E. Purification of transmembrane proteins from *Saccharomyces cerevisiae* for X-ray crystallography. *Protein Expr. Purif.* **2010**, 71, 207-223.

Clémençon, B., Lüscherm, B.P., Hediger, M.A. Establishment of a novel microscale thermophoresis ligand-binding assay for characterization of SLC solute carriers using oligopeptide transporter PepT1 (SLC15 family) as a model system. *J. Pharmacol. Toxicol. Methods.* **2018**, 92, 67-76.

Covitz, K.M., Amidon, G.L., Sadée, W. Membrane topology of the human dipeptide transporter, hPEPT1, determined by epitope insertions. *Biochemistry* **1998**, 37, 15214-15221.

Cura, A.J., Carruthers, A. Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2012**, 2, 863-914.

Daniel, H., Rubio-Aliaga, I. An update on renal peptide transporters. *Am. J. Physiol. Renal Physiol.* **2003**, 284, 885-892.

Daniel, H., Spanier, B., Kottra, G., Weitz, D. From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology (Bethesda)* **2006**, 21, 93-102.

de Meyer, F., Smit, B. Effect of cholesterol on the structure of a phospholipid bilayer. *Proc. Natl. Acad. Sci. USA* **2009**, 106, 3654-3658.

Denisov, I.G., Sligar, S.G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **2016**, *23*, 481-486.

Diederichs, K., Karplus, P.A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **1997**, 4, 269-275.

Doki, S., Kato, H.E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N., Tsukazaki, T., Sugita, Y., Kandori, H., *et al.* Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl. Acad. Sci. USA* **2013**, 110, 11343-11348.

Dowhan, W. Molecular basis for membrane phospholipid diversity: Why Are There So Many Lipids? *Annu. Rev. Biochem.* **1997**, 66, 199-232.

Drew, D., Slotboom, D.J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl- Beinker, N.M., Lambert, W., Lerch, M., Daley, D.O., Van Wijk, J.-K., Hirst, J., Kunji, E., de Gier, J.-W. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* **2005**, 14, 2011-2017.

El-Gebali, S., Bentz, S., Hediger, M.A., Anderle, P. Solute carriers (SLCs) in cancer. *Mol. Asp. Med.* **2013**, 34, 719-734.

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

Engelman, D. M. Membranes are more mosaic than fluid. *Nature,* **2005**, 438, 578-580.

Ernst, H.A., Pham, A., Hald, H., Kastrup, J.S., Rahman, M., Mirza, O. Ligand binding analyses of the putative peptide transporter YjdL from E. coli display a significant selectivity towards dipeptides. *Biochem. Biophys. Res. Commun.* **2009**, 389, 112-116.

Esko, J.D., Varki, A., Cummings, R.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E. Essentials of Glycobiology. *Cold Spring Harbor Laboratory Press, New York* **2009**, 784 p.

Faham, S., Boulting, G.L., Massey, E.A., Yohannan, S., Yang, D., Bowie, J. U. Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci.* **2005**, 14, 836-840.

Faham, S., Bowie, J.U. Bicelle crystallization: A new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **2002**, 316, 1-6.

Fang, G., Konings, W.N., Poolman, B. Kinetics and substrate specificity of membrane-reconstituted peptide transporter DtpT of Lactococcus lactis. *J. Bacteriol.* **2000**, 182, 2530-2535.

Fei, Y.J., Kanai, Y., Nussberger, S., Ganapathy, V., Leibach, F.H., Romero, M.F., *et al.* Expression cloning of a mammalian proton-coupled oligopeptide transporter, *Nature* **1994**, 368, 563-566.

Fenalti, G., Abola, E.E., Wang, C., Wu, B., Cherezov, V. Fluorescence recovery after photobleaching in lipidic cubic phase (LCP-FRAP): A precrystallization assay for membrane proteins. *Methods Enzymol.* 2015, 557, 417-437.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., Punta, M. Pfam: the protein families database. *Nucleic Acids Res.* **2014**, 42, D222-D230.

Flayhan, A., Mertens, H.D.T., Ural-Blimke, Y., Martinez Molledo, M., Svergun, D.I., Löw, C. Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **2018**, *26*, 345-355.

Fluman, N., Ryan, C.M., Whitelegge, J.P., Bibi, E. *Mol. Cell* **2012**, 47, 777-787.

Fowler, P.W., Orwick-Rydmark, M., Radestock, S., Solcan, N., Dijkman, P.M., Lyons, J.A., *et al.* Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Structure* **2015**, 23, 290-301.

Frauenfeld, J., Löving, R., Armache, J.P., Sonnen, A.F.P., Guettou, F., Moberg, P., Zhu, L., Jegerschöld, C., Flayhan, A., Briggs, J. A., Garoff, H., Löw, C., Cheng, Y., Nordlung, P. A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* **2016** 13, 345-351.

Fu, X., Zhang, G., Liu, R., Wei, J., Zhang- Negrerie, D., Jian, X., Gao, Q. Mechanistic Study of Human Glucose Transport Mediated by GLUT1. *J. Chem. Inf. Model.* **2016**, 56, 517-526.

Ganapathy, M.E., Huang, W., Wang, H., Ganapathy, V. Leibach, F.H. Valacyclovir: a substrate for the intestinal and renal peptide transporters PEPT1 and PEPT2. *Biochem. Biophys. Res. Commun.* **1998**, 246, 470-475

Gbelska, Y., Krijger, J.-J., Breuning, K.D. Evolution of gene families: the multidrug resistance transporter genes in five related yeast species. *FEMS Yeast Res.* **2006**, 6, 345-355.

Gelin, M., Delfosse, V., Allemand, F., Hoh, F., Sallaz-Damaz, Y., Pirocchi, M., Bourguet, W., Ferrer, J.L., Labesse, G., Guichou, J.F. Combining 'dry' co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, 71, 1777-1787.

Gray, K.A., Yates, B., Seal, R.L., Wright, M.W., Bruford, E.A. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.* **2015**, 43, 1079-1085.

Guan, L., Kaback, H.R. Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, 35, 67-91.

Guettou, F., Quistgaard, E.M., Raba, M., Moberg, P., Löw, C., Nordlund, P. Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat. Struct. Mol. Biol.* **2014**, 21, 728-731.

Guettou, F., Quistgaard, E.M., Trésaugues, L., Moberg, P., Jegerschöld, C., Zhu, L., Jong, A.J., Nordlund, P., Löw, C. Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO Rep. 2013*, *14*, 804-810.

Hagenbuch, B., Stieger, B. The SLCO (former SLC21) superfamily of transporters. *Mol. Aspects Med.* **2013**, 34, 396-412.

Halestrap, A.P., Wilson, M.C. The monocarboxylate transporter family - role and regulation. *IUBMB Life* **2012**, 64, 109-119.

Harder, D., Stolz, J., Casagrande, F., Obrdlik, P., Weitz, D., Fotiadis, D., Daniel, H. DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of Escherichia coli. *FEBS J.* **2008**, 275, 3290-3298.

Harder, M., Kuhn, B., Diederich, F. Efficient stacking on protein amide fragments. *ChemMedChem.* **2013**, 8, 397-404

Hays, F.A., Zygy, R.-Z., Stroud, R.M. Overexpression and purification of integral membrane proteins in yeast. *Methods Enzymol.* **2010**, 470, 695-707.

Hediger, M.A., Clémençon, B., Burrier, R.E., Bruford, E.A. The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol. Asp. Med.* **2013**, 34, 95-107.

Hendrickson, W.A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **1991**, 254, 51-58.

Huang, C.-Y., Olieric, V., Ma, P., Panepucci, E., Diederichs, K., Wang, M., Caffrey, M. *In meso in situ* serial X-ray crystallography of soluble and membrane proteins. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, 71, 1238-1256.

Huang, C.Y., Olieric, V., Ma, P., Howe, N., Vogeley, L., Liu, X., *et al. In meso in situ* serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. *Acta Crystallogr. D Biol. Crystallogr.* **2016**, 72, 93-112.

Huang, H.L., Charoenkwan, P., Kao, T.F., Lee, H.C., Chang, F.L., Huang, W.L., Ho, S.J., Shu, L.S., Chen, W.L., Ho, S.Y. Prediction and analysis of protein solubility using a novel scoring card method with dipeptide composition. *BMC Bioinformatics.* **2012**, 13, Suppl 17, S3.

Huang, Y., Lemieux, M.J., Song, J., Auer, M., Wang, D.-N. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **2003**, 301, 616-620.

Ilgu, H., Jeckelmann, J.M., Gapsys, V., Ucurum, Z., de Groot, B.L., Fotiadis, D. Insights into the molecular basis for substrate binding and specificity of the wild-type L-arginine/agmatine antiporter AdiC. *Proc. Natl. Acad. Sci. USA* **2016**, 113, 10358-10363.

Immadisetty, K.C., Hettige, J.J., Moradi, M. What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *J. Phys. Chem. B.* **2017**, 121, 3644-3656.

Ito, K., Hikida, A., Kawai, S., Lan, V.T., Motoyama, T., Kitagawa, S., Yoshikawa, Y., Kato, R., Kawarasaki, Y. Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library. *Nat. Commun.* **2013**, 4, 2502.

Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **1966**, 211, 969–970.

Jensen, J.M., Ismat, F., Szakonyi, G., Rahman, M., Mirza, O. Probing the Putative Active Site of YjdL: An Unusual Proton-Coupled Oligopeptide Transporter from E. coli. *PLoS ONE* **2012**, *7*, e47780.

Jessen-Marshall, A.E., Paul, N.J., Brooker, R.J. The conserved motif, GXXX(D/E)(R/K)XG[X](R/K)(R/K), in hydrophilic loop 2/3 of the lactose permease. **1995**, 270, 16251-16257.

Jiang, D., Zhao, Y., Wang, X., Fan, J., Heng, J., Liu, X., Feng, W., Kang, X., Huang, B., Liu, J., Zhang, X. C. Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. *Proc. Natl Acad. Sci. USA* **2013**, 110, 14664-14669.

Juge, N., Yoshida, Y., Yatsushiro, S., Omote, H., Moriyama, Y. Vesicular glutamate transporter contains two independent transport machineries. *J. Biol. Chem.* **2006**, 281, 39499-39506.

Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132.

Kanai, Y., Clémençon, B., Simonin, A., Leuenberger, M., Lochner, M., Weisstanner, M., Hediger, M.A. The SLC1 high-affinity glutamate and neutral amino acid transporter family. *Mol. Aspects Med.* **2013**, 34, 108-120.

Karplus, P.A., Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **2015**, 34, 60-68.

Karplus, P.A., Diederichs, K. Linking crystallographic model and data quality. *Science* **2012**, 336, 1030-1033.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **1958**, 181, 662-666.

Kleywegt, G.J. Validation of protein crystal structures. *Acta Crystallogr. D* **2000**, 56, 249-265.

Knütter, I., Hartrodt, B., Theis, S., Foltz, M., Rastetter, M., Daniel, H., *et al.* Analysis of the transport properties of side chain modified dipeptides at the mammalian peptide transporter PEPT1. *Eur. J. Pharm. Sci.* **2004**, 21, 61-67.

Koepsell, H. The SLC22 family with transporters of organic cations, anions and zwitterions. *Mol. Aspects Med.* **2013**, 34, 413-435.

Kumar, H., Kasho, V., Smirnova, I., Finer- Moore, J.S., Kaback, H.R., Stroud, R.M. Structure of

sugar-bound LacY. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, 111, 1784-1788.

Laganowsky, A., Reading, E., Allison, T.M., Ulmschneider, M.B., Degiacomi, M.T., Baldwin, A.J., Robinson, C.V. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **2014**, 510, 172-175.

Laganowsky, A., Reading, E., Hopper, J.T., Robinson, C.V. Mass spectrometry of intact membrane protein complexes. *Nat. protoc.* **2013**, 8, 639-651.

Laskoswski, R.A., Swindells, M.B. LigPlot$^+$: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **2011**, 51, 2778-2786.

Law, C.J., Maloney, P.C., Wang, D.-N. Ins and outs of major facilitator superfamily antiporters. *Ann. Rev. Microbiol.* **2008**, *62*, 289-305.

Leibach, F.H., Ganapathy, V. Peptide transporters in the intestine and the kidney. *Annu. Rev. Nutr.* **1996**, 16, 99-119.

Liang, R., Fei, Y.J., Prasad, P.D., Ramamoorthy, S., Han, H., Yang-Feng, T.L., *et al.* Human intestinal H$^+$/peptide cotransporter. Cloning, functional expression, and chromosomal localization. *J. Biol. Chem.* **1995**, 270, 6456-6463.

Linke, D. Detergents: an overview. *Methods Enzymol.* **2009**, 463, 603-617.

Long, S.B., Campbell, E.B., Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K$^+$channel. *Science* **2005**, 309, 897-903.

Löw, C., Moberg, P., Quistgaard, E.M., Hedren, M., Guettou, F., Frauenfeld, J., Haneskog, L., Nordlund, P. High-throughput analytical gel filtration screening of integral membrane proteins for structural studies. *Biochim. Biophys. Acta* **2013**, 1830, 3497-3508.

Luckner, P., Brandsch, M. Interaction of 31 beta-lactamantibiotics with the H$^+$/peptide symporter PEPT2: analysis of affinity constants and comparison with PEPT1. *Eur. J. Pharm. Biopharm.* **2005**, 59, 17-24.

Lyons, J.A., Parker, J.L., Solcan, N., Brinth, A., Li, D., Shah, S.T., Caffrey, M., Newstead, S. Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO Rep.* **2014**, 15, 886-893.

Lyons, J.A., Shahasavar, A., Paulsen, P.A., Pedersen, B.P., Nissen, P. Expression strategies for structural studies of eukaryotic membrane proteins. *Curr. Opin. Struct. Biol.* **2016**, 38, 137-144.

Lyons, JA., Bøggild, A., Nissen, P., Frauenfeld, J. Saposin-Lipoprotein Scaffolds for Structure Determination of Membrane Transporters. *Methods Enzymol.* **2017**, 594, 85-99.

Malle, E., Zhou, H., Neuhold, J., Spitzenberger, B., Klepsch, F., Pollak, T., *et al.* Random mutagenesis of the prokaryotic peptide transporter YdgR identifies potential periplasmic gating residues. *J. Biol. Chem.* **2011**, 286, 23121-23131.

Manglik, A., Kobilka, B.K., Steyaert, J. Nanobodies to study G protein-coupled receptor structure and function. *Annu. Rev. Pharmacol. Toxin* **2017**, 57, 19-37.

Marger, M.D., Saier, MH.Jr. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem. Sci.* **1993**, 18, 13-20.

Martinez Molledo, M., Quistgaard, E.M., Flayhan, A., Pieprzyk, J., Löw, C. Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* **2018**, 26, 467-476.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, 40, 658-674.

McPherson, A., Gavira, J.A. Introduction to protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* **2014**, *70*, 2-20.

Moradi, M., Enkavi, G., Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate: phosphate antiporter. *Nat. Commun.* **2015**, 6, 8393.

Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., Shaw Stewart, P.D. Membrane protein structure determination-The next generation. *Biochim. Biophys. Acta* **2014**, 1838, 78-87.

Mueckler, M., Thorens, B. The SLC2 (GLUT) family of membrane transporters. *Mol. Aspects Med.* **2013**, 34, 121-138.

Murray, C.I., Van Eyk, J.E. Chasing cysteine oxidative modifications: proteomic tools for characterizing cysteine redox status. *Circ. Cardiovasc. Genet.* **2012**, 5, 591.

Nelson, D. L., Cox, M. M. Lehninger. Principles of Biochemistry *New York, Worth Publishers* **2000** (3rd edition).

Newstead, S. Recent advances in understanding proton coupled peptide transport via the POT family. *Curr. Opin. Struct. Biol.* **2017**, *45*, 17-24.

Newstead, S., Drew, D., Cameron, A.D., Postis, V.L.G., Xia, X., Fowler, P.W., Ingram, J.C., Carpenter, E.P., Sansom, M.SP., McPherson, M.J., Baldwin, S.A., Iwata, S. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **2011** 30, 417-426.

Niesen, FH, Berglund, H., Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2007**, 2, 2212-2221.

Nishio, M., Umezawa, Y., Fantini, J., Weiss, M., S., Chakrabarti, P. CH-pi hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **2014,** 16, 12648-12683.

Nomura, N., Verdon, G., Kang, H.J., Shimamura, T., Nomura, Y., *et al.* Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **2015**, 526, 397-401.

Oeffner, R.D., Bunkóczi, G., McCoy, A.J, Read, R.J. Improved estimated of coordinate error for molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* **2013**, 69, 2209-2215.

Osborne, A.R., Rapoport, T.A., van den Berg, B. Protein Translocation By the Sec61/SecY Channel. *Annu. Rev. Cell Dev. Biol.* **2005**, 21, 529-550.

Ott, C.M., Lingappa, V.R. Integral membrane protein biosynthesis: why topology is hard to predict. *J. Cell Sci.* **2002**, 115, 2003-2009.

Pardon, E., Laeremans, T., Triest, S., Rasmussen, S.G., Wohlkönig, A., Ruf, A., Muydermans, S., Hol, W.G., Kobilka, B.K., Steyaert, J. A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **2014**, 9, 674-693.

Parker, J., L., Newstead, S. Current trends in α-helical membrane protein crystallization: an update. *Protein Sci.* **2012**, *21*, 1358-1365.

Parker, J.L., Li, C., Brinth, A., Wang, Z., Vogeley, L., Solcan, N., *et al.* Proton movement and coupling in the POT family of peptide transporters. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 13182-13187.

Parker, J.L., Mindell, J.A., Newstead, S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLife* **2014a**, 3, 1-13.

Parker, J.L., Newstead, S. Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature* **2014b**, 507, 68-72.

Paulsen, I.T., Skurray, R.A. The POT family of transport proteins. *Trends Biochem. Sci.* **1994**, 19, 404.

Pautsch, A., Schulz, G.E. High-resolution structure of the OmpA membrane domain. *J. Mol. Biol.* **2000**, 298, 273-282.

Poulos, S., Morgan, J.L.W., Zimmer, J., Faham, S. Bicelles coming of age: An empirical approach to bicelle crystallization. *Methods Enzymol.* **2015**, 557, 393-416.

Prabhala, B.K., Aduri, N.G., Iqbal, M., Rahman, M., Gajhede, M., Hansen, P.R., Mirza, O. Several hPepT1-transported drugs are substrates of the Escherichia coli proton-coupled oligopeptide transporter YdgR. *Res. Microbiol.* **2017**, 168, 443-449.

Prabhala, B.K., Aduri, N.G., Jensen, J.M., Ernst, H.A., Iram, N., Rahman, M., Mirza, O. New insights into the substrate specificities of proton-coupled oligopeptide transporters from E. coli by a pH sensitive assay. *FEBS Lett.* **2014**, 14, 560-565.

Privé, G.G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **2007**, 41, 388-397.

Quistgaard, E.M., Löw, C., Guettou, F., Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell Biol.* **2016**, 17, 123-132.

Quistgaard, E.M., Löw, C., Moberg, P., Trésaugues, L., Nordlund, P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat. Struct. Mol. Biol.* **2013***, 20, 766-768.

Quistgaard, E.M., Martinez Molledo, M., Löw, C. Structure determination of a major facilitator peptide transporter: Inward facing PepT$_{St}$ from *Streptococcus thermophilus* crystallized in space group P3121. *PLoS ONE* **2017**, 12, 1-20.

Radestock, S., Forrest, L.R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* **2011**, 407, 698-715.

Rask-Andersen, M., Masuram, S., Fredriksson, R., Schiöth, H.B. Solute carriers as drug targets: current use, clinical trials and prospective. *Mol. Asp. Med.* **2013**, 34, 702-710.

Rasmussen, S.G.F., Choi, H.-J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., et al. Crystal structure of the human β2 adrenergic G-protein-coupled receptor. *Nature* **2007**, 450, 383-387.

Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., Saier, Jr.M.H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **2012**, 279, 2022-2035.

Reimer, R., J. SLC17: a functionally diverse family of organic anion transporters. *Mol. Aspects Med.* **2013**, 34, 350-359.

Ren, Q., Chen, K., Paulsen, I.T. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.* **2007**, 35, D274-D279.

Roth, M., Obaidat, A., Hagenbuch, B. OATPs, OATs and OCTs: The organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. *Br. J. Pharmacol.* **2012**, 165, 1260-1287.

Salonen, L.M., Ellermann, M., Diederich, F. Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed. Engl.* **2011**, 50, 4808-4842

Samsudin, F., Parker, J.L., Sansom, M.S.P., Newstead, S., Fowler, P.W. Accurate prediction of ligand affinities for a proton-dependent oligopeptide transporter. *Cell Chem. Biol.* **2016**, 23, 299-309.

Sanger, S.J., Nicolson, G.L. The fluid mosaic model of the structure of cell membranes. *Science* **1972**, 175, 720-731.

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M.M., Strous, M., Drew, D., Slotboom, D.J., de Gier, J.-W. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21 (DE3). *J. Mol. Biol.* **2012**, 423, 648-659.

Schlegel, S., Rujas, E., Ytterberg, A.J., Zubarev, R.A., Luirink, J., de Gier, J.-W. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb. Cell Fact.* **2013**, 12, 24.

Seddon, A.M., Curnow, P., Booth, P.J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* **2004**, 1666, 105-117.

Seidel, S.A., Dijkman, P.M., Lea, W.A., van den Bogaart, G., Jerabek- Willemsen, M., Lazic, A., Joseph, J.S., Srinivasan, P., Baaske, P., Simeonov, A., *et al.* Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* **2013**, 59, 301-315.

Serebryany, E., Zhu, G.A., Yan, E.C.Y. Artificial membrane-like environments for *in vitro* studies of purified G-protein coupled receptors. *Biochim. Biophys. Acta* **2012**, 1818, 225-233.

Serrano-Vega, M.J., Magnani, F., Shibata, Y., Tate, C.G. Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 877-882.

Serrano-Vega, M.J., Tate, C.G. Transferability of thermostabilizing mutations between β-adrenergic receptors. *Mol. Membr. Biol.* **2009**, 26, 385-396.

Sharma, N., Aduri, N.G., Iqbal, A., Prabhala, B.K., Mirza, O. Peptide selectivity of the proton-coupled oligopeptide transporter from Neisseria meningitidis. *J. Mol. Microbiol. Biotechnol.* **2016**, 26, 312-319.

Sharom, F.J. Flipping and flopping-lipids on the move. *IUBMB Life* **2011**, 63, 736-746.

Simons, K., Ikonen, E. Functional rafts in cell membranes. *Nature* **1997**, 387, 569-572.

Smith, D.E., Clémençon, B., Hediger, M.A. Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol. Aspects Med.* **2013**, 34, 323-336.

Solcan, N., Kwok, J., Fowler, P.W., Cameron, A.D., Drew, D., Iwata, S., Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J* **2012**, 31, 3411-3421.

Sprovieri, P., Martino, G. The role of the carbohydrates in plasmatic membrane. *Physiol. Res.* **2017**, 8408, 1-25.

Stelzl, L.S., Fowler, P.W., Sansom, M.S.P., Beckstein, O. Flexible gates generate occluded intermediates in the transport cycle of LacY. *J. Mol. Biol.* **2014**, 426, 735-751.

Sun, J., Bankston, J.R., Payandeh, J., Hinds, T.R., Zagotta, W.N., Zheng, N. Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* **2014**, 507, 73-77.

Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y., Yan, N. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature* **2012**, 18, 361-366.

Tate, C., G., Schertler, G., F., X. Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **2009**, 19, 386-395.

Tate, C.G. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett.* **2001**, 504, 94-98.

Tsay, Y.F., Chiu, C.C., Tsai, C.B., Ho, C.H., Hsu, P.K. Nitrate trans- porters and peptide transporters. *FEBS Lett.* **2007**, 581, 2290-2300.

Ujwal, R., Cascio, D., Colletier, J.-P., Faham, S., Zhang, J., Toro, L., *et al.* The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 17742-17747.

Veith, K., Martinez Molledo, M., Almeida Hernandez, Y., Josts, I., Nitsche, J., Löw, C., Tidow, H. Lipid-like Peptides can Stabilize Integral Membrane Proteins for Biophysical and Structural Studies. *ChemBioChem.* **2017**, 18, 1735-1742.

Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K.J., Slotboom, D.J., Persson, J.O., de Gier, J.-W. Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 14371-14376.

Wang, H., Elferich, J., Gouaux, E. Structures of LeuT in bicelles define conformation and substrate binding in a membrane-like context. *Nat. Struct. Mol. Biol.* **2012**, 19, 212-219.

Weitz, D., Harder, D., Casagrande, F., Fotiadis, D., Obrdlik, P., Kelety, B., Daniel, H. Functional and structural characterization of a prokaryotic peptide transporter with features similar to mammalian PEPT1. *J. Biol. Chem.* **2007**, 282, 2832-2839.

White, S. H., von Heijne, G. The machinery of membrane protein assembly. *Curr. Opin. Struct. Biol.* **2004**, 14, 397-404.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* **2013**, 280, 5705-5736.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **2008**, 275, 1-21.

Xu, F., Liu, W., Hanson, M.A., Stevens, R.C., Cherezov, V. Development of an automated high throughput LCP-FRAP assay to guide membrane protein crystallization in lipid mesophases. *Cryst. Growth Des.* **2011**, 11, 1193-1201.

Yamaguchi, A., Someya, Y., Sawai, T. Metal- tetracycline/H+ antiporter of Escherichia coli encoded by transposon Tn10: the role of a conserved sequence motif, GXXXXRXGRR, in a putative cytoplasmic loop between helices 2 and 3. *J. Biol. Chem.* **1992**, 267, 19155-19162.

Yan, N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.* **2013**, 38, 151-159.

Yan, N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu. Rev. Biophys.* **2015**, *44*, 257-283.

Zander, U., Hoffmann, G., Cornaciu, I., Marquette, J.P., Papp, G., Landret, C., Seroul, G., Sinoir, J., Röwer, M., Felisaz, F., Rodriguez-Puente, S., Mariaule, V., Murphy, P., Mathieu, M., Cipriani, F., Márquez, J.A. Automated harvesting and processing of protein crystals through laser photoablation. *Acta Crystallogr D Struct Biol* **2016**, 72, 454-466.

Zhao, Y., Mao, G., Liu, M., Zhang, L., Wang, X., Zhang, X. C. Crystal Structure of the *E. coli* Peptide Transporter YbgH. *Structure* **2014**, 22, 1152-1160.

Chapter 8. Appendix

I. Crystallographic data processing and refinement statistics

PepT$_{St}$ apostructure

Data collection	
Beamline	PETRA III, P13
Wavelength (Å)	1.0332
Space group	C222$_1$
Cell dimensions	
a, b, c (Å)	102.10, 110.30, 110.70
α, β, γ (°)	90, 90, 90
Resolution (Å)	49.36 – 1.95 (2.02 – 1.95)
R_{merge}	0.1223 (1.673)
$I/\sigma I$	16.20 (1.30)
CC1/2	0.999 (0.592)
Completeness (%)	99.63 (99.27)
Total no. reflections	603439 (60307)
Multiplicity	13.2 (13.5)
Wilson B-factor (Å^2)	37.29
Refinement	
R_{work} / R_{free}	0.181 / 0.197
No. atoms	
Protein	3546
Ligands/ions (binding site)	10
Ligands/ions (elsewhere)	27
Lipids	440
Water	177
B-factors	
Protein	41.6
Ligands/ions (binding site)	81.0
Ligands/ions (elsewhere)	81.4
Lipids	78.9
Water	45.8
R.m.s. deviations	
Bond lengths (Å)	0.005
Angles (°)	0.877
Ramachandran	
Favored (%)	98.9
Outliers (%)	0.0
Clash score	4.0
PDB accession	5OXO

Table 32: Crystallographic data processing and refinement statistics for PepT$_{St}$ apostructure.

PepT$_{St}$ complex with:	Ala-Leu	Ala-Gln	Asp-Glu	Phe-Ala
Data collection				
Beamline	ESRF ID30A-1	PETRA III P14	ESRF ID30B	PETRA III P14
Wavelength (Å)	0.9660	0.9762	1.0396	0.9763
Space group	C222$_1$	C222$_1$	C222$_1$	C222$_1$
Cell dimensions				
a, b, c (Å)	102.30, 110.60, 108.50	100.70, 110.20 104.20	100.60, 109.00, 107.00	100.80, 107.9 109.80
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	46.27 − 2.66 (2.75 − 2.66)	48.81 − 2.38 (2.465 − 2.38)	48.57 − 2.30 (2.377 − 2.295)	45.81 − 2.20 (2.275 − 2.196)
R_{merge}	0.1076 (0.5943)	0.1073 (1.121)	0.1188 (1.439)	0.147 (1.411)
$I/\sigma I$	9.58 (1.97)	18.75 (2.56)	10.33 (1.03)	14.61 (2.13)
CC1/2	0.994 (0.72)	0.999 (0.832)	0.998 (0.509)	0.997 (0.712)
Completeness (%)	98.20 (96.40)	99.74 (99.91)	98.99 (95.06)	99.53 (97.69)
Total no. reflections	53978 (5215)	314880 (31241)	172471 (17919)	568946 (41559)
Multiplicity	3.0 (2.9)	13.2 (13.4)	6.5 (6.8)	18.4 (13.9)
Wilson B-factor (Å)	49.82	47.07	53.66	47.21
Refinement				
R_{work} / R_{free}	0.224 / 0.233	0.208 / 0.222	0.215 / 0.235	0.194 / 0.214
No. atoms				
Protein	3463	3380	3394	3580
Ligands/ions (binding site)	14	15	23	17
Ligands/ions (elsewhere)	57	31	15	11
Lipids	220	286	352	264
Water	32	47	29	73
B-factors				
Protein	54.6	58.4	69.1	56.6
Ligands/ions (binding site)	47.0	88.3	110.0	85.8
Ligands/ions (elsewhere)	91.7	92.6	118.2	90.9
Lipids	74.8	89.5	92.8	84.5
Water	51.3	56.0	58.3	55.8
R.m.s. deviations				
Bond lengths (Å)	0.003	0.004	0.005	0.005
Angles (°)	0.755	0.833	0.916	0.804
Ramachandran				
Favored (%)	98.2	98.6	98.6	98.9
Outliers (%)	0.0	0.0	0.0	0.0
Clash score	3.4	3.0	6.7	4.7
PDB codes	5OXL	5OXK	5OXM	5OXN

Table 33: Crystallographic data processing and refinement statistics for PepT$_{St}$ in complex with dipeptides.

PepT$_{St}$ in complex with Phe-Ala-Gln

Data collection

Beamline	PETRA III
	P14
Wavelength (Å)	0.9143
Space group	C222$_1$
Cell dimensions	
a, b, c (Å)	101.55, 108.22, 111.61
α, β, γ (°)	90, 90, 90
Resolution (Å)	48.69 – 2.26 (2.341 – 2.26)
R_{merge}	0.083 (0.811)
$I/\sigma I$	19.44 (3.19)
CC1/2	0.999 (0.871)
Completeness (%)	99.73 (99.79)
Total no. reflections	299776 (29604)
Multiplicity	10.3 (10.4)
Wilson B-factor (Å^2)	53.66
Refinement	
R_{work} / R_{free}	0.193 / 0.219
No. atoms	
Protein	3529
Ligands/ions (binding site)	26
Ligands/ions (elsewhere)	41
Lipids	308
Water	101
B-factors	
Protein	41.19
Ligands/ions (binding site)	51.94
Ligands/ions (elsewhere)	96.49
Lipids	82.59
Water	53.04
R.m.s. deviations	
Bond lengths (Å)	0.006
Angles (°)	0.76
Ramachandran	
Favored (%)	98.89
Outliers (%)	0.0
Clash score	5.22
PDB accession	6GHJ

Table 34: Crystallographic data processing and refinement statistics for PepT$_{St}$ in complex with Phe-Ala-Gln.

PepT$_{St}$ complex with:	HEPES at 100 mM	HEPES at 300 mM
Data collection		
Beamline	PETRA III	PETRA III
	P13	P13
Wavelength (Å)	0.9796	0.9763
Space group	C222$_1$	C222$_1$
Cell dimensions		
$\quad$ a, b, c (Å)	102.49, 110.03, 110.56	102.21, 110.05, 109.50
$\quad$ α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	49.25 – 2.00 (2.072 – 2.00)	46.31 – 2.19 (2.27 – 2.19)
R_{merge}	0.0651 (1.139)	0.132 (2.175)
$I/\sigma I$	23.25 (2.33)	14.90 (0.95)
CC1/2	0.999 (0.766)	1.000 (0.408)
Completeness (%)	99.8 (99.52)	99.7 (97.4)
Total no. reflections	556226 (55100)	415514 (37484)
Multiplicity	13.1 (13.2)	13.1 (12.0)
Wilson B-factor (Å^2)	41.15	45.45
Refinement		
R_{work} / R_{free}	0.181 / 0.200	0.190 / 0.205
No. atoms		
$\quad$ Protein	3629	3519
$\quad$ Ligands/ions (binding site)	20	20
$\quad$ Ligands/ions (elsewhere)	43	41
$\quad$ Lipids	396	396
$\quad$ Water	122	98
B-factors		
$\quad$ Protein	50.1	57.6
$\quad$ Ligands/ions (binding site)	82.0	73.9
$\quad$ Ligands/ions (elsewhere)	101.4	100.4
$\quad$ Lipids	84.4	87.2
$\quad$ Water	51.5	53.8
R.m.s. deviations		
$\quad$ Bond lengths (Å)	0.003	0.010
$\quad$ Angles (°)	0.779	1.135
Ramachandran		
$\quad$ Favored (%)	99.3	98.7
$\quad$ Outliers (%)	0.0	0.0
Clash score	3.1	5.4
PDB accession	6EIA	5OXQ

Table 35: Crystallographic data processing and refinement statistics for PepT$_{St}$ in complex with HEPES at 100 mM and 300 mM concentration.

PepT$_{St}$ in occluded conformation (phosphate ion)

Data collection

Beamline	PETRA III P13
Wavelength (Å)	0.9763
Space group	C222$_1$
Cell dimensions	
a, b, c (Å)	101.60, 110.10, 107.90
α, β, γ (°)	90, 90, 90
Resolution (Å)	49.04 – 2.37 (2.46 – 2.37)
R_{merge}	0.094 (1.423)
$I/\sigma I$	14.09 (1.17)
CC1/2	0.999 (0.477)
Completeness (%)	99.6 (97.3)
Total no. reflections	162105 (15327)
Multiplicity	6.5 (6.4)
Wilson B-factor (Å^2)	58.04
Refinement	
R_{work} / R_{free}	0.194 / 0.208
No. atoms	
Protein	3572
Ligands/ions (binding site)	21
Ligands/ions (elsewhere)	33
Lipids	352
Water	37
B-factors	
Protein	67.1
Ligands/ions (binding site)	76.9
Ligands/ions (elsewhere)	123.1
Lipids	92.8
Water	57.2
R.m.s. deviations	
Bond lengths (Å)	0.009
Angles (°)	1.135
Ramachandran	
Favored (%)	98.7
Outliers (%)	0.0
Clash score	5.9
PDB accession	5OXP

Table 36: Crystallographic data processing and refinement statistics for PepT$_{St}$ in complex with a phosphate ion.

II. List of hazardous substances

Reagent	Pictogram	Hazardous (H)/ Precautions (P) statements
Acetic acid 100 %, Rotipuran® 100 %, p.a.	GHS02 GHS05	H226, H290, H314 P210, P280, P303 + P361 + P353, P305 + P351 + P338, P310
Acrilamide/ Bis stock sol. 29:1 (40% w/ v)	GHS07 GHS08	H301, H312 + H332, H315, H317, H319, H340, H350, H361f, H372 P201, P260, P280, P301+ P330 + P331 + P310, P305 + P351 + P338, P308 + P313
Ammonium peroxodisulphate	GHS03 GHS07 GHS08	H272, H302, H315, H317, H319, H334, H335 P261, P280, P302 + P352, P305 + P351 + P338, P332 + P313, P337 + P313
Ampicillin disodium salt	GHS08	H317, H334 P261, P280, P302 + P352, P342 + P311
Boric acid ≥ 99,8 %, p.a., ACS, ISO	GHS08	H360FD P280, P308 + P313
Citric acid monohydrate	GHS07	H319 P305 + P351 + P338
cOmplete, EDTA-free protease inhibitor cocktail tablets	GHS05	H314 P260, P280, P301 + P330 + P331, P303 + P361 + P353, P304 + P340 + P310, P305 + P351 + P338 + P310
EDTA tetrasodiumsalt	GHS05 GHS07 GHS08	H302 + H332, H318, H373 P261, P270, P280, P305 + P351 + P338, P310
Ethanol ≥ 99.8%	GHS02 GHS07	H225, H319 P210, P233, P305 + P351 + P338
Guanidine hydrochloride ≥ 99,5%	GHS07	H302, H315, H319 P280, P302 + P352, P305 + P351 + P338
Hydrochloric Acid 32%	GHS05 GHS07	H290, H314, H335 P280, P303 + P361 + P353, P304 + P340, P30 + P351+P338, P312
Imidazol, Pufferan ≥ 99%, p.a.	GHS05 GHS07 GHS08	H302, H314, H360D P260, P280, P301 + P330 + P331, P303 + P361 + P353, P305 + P351 + P338, P308 + P313
InstantBlue™	GHS05	H290, H315, H319 P305 + P351 + P338
Iodoacetamide	GHS06 GHS08	H301, H317, H334 P261, P280, P301 + P310, P342 + P311
Kanamycin sulfate	GHS07	H360 P201, P280, P308 + P313
Methanol Rotipuran® ≥99,9 %, p.a., ACS, ISO	GHS02 GHS06 GHS08	H225, H301 + H311 + H331, H370 P210, P270, P280, P303 + P361 + P353, P304 + P340, P308 + P311

Nickel(II) chloride hexahydrate ≥ 98 %, p.a.	GHS06 GHS08 GHS09	H301 + H331, H315, H317, H334, H341, H350i, H360D, H372, H410 P201, P260, P273, P304 + PP341, P308, P310
Sekusept Plus	GHS05 GHS07 GHS09	H302 + H332, H314, H400 P273, P280, P303 + P361 + P353, P305 + P351 + P338, P310
Sodium hydroxide ≥99	GHS05	H290, H314 P280, P301 + P330 + P331, P305 + P351 + P338, P310
TEMED ≥ 99%, p.a.	GHS02 GHS05 GHS07	H225, H302 + H332, H314 P210, P280, P301 + P330 + P331, P303 + P361 + P353, P305 + P351+ P338, P310
TRIS hydrochloride, Pufferan® ≥ 99%, p.a.	GHS07	H315, H319, H335 P280, P302 + P352, P305 + P351 + P338

Table 37: List of hazardous substances. For each chemical, the associated pictogram code is provided as well as the hazardous and precaution statements. For more information, check the "Regulation (EC) No 1272/2008" of the European Agency for Safety and Health at Work.

Figure 57: GHS pictograms.

IIII. Declaration of oath

I hereby declare on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids. The submitted written version corresponds to the version on the electronic storage medium. I hereby declare that I have not previously applied or pursued for a doctorate.

Date:

Signature:

Maria Martinez Molledo