

Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering and
Natural Sciences
of the
Ruprecht - Karls - University
Heidelberg

Presented by

M.Sc. Maxime Killer

born in Vienne

Oral examination: 3rd of March 2023

Molecular insights into peptide transporters

Referees:

Prof. Dr. Britta Brügger

Dr. Thomas R. Schneider

The work presented in this dissertation was performed between September 2018 and September 2022 in the group of Dr. Christian Löw at the European Molecular Biology Laboratory (EMBL) in Hamburg, and at the Centre for System Structural Biology (CSSB), in Germany.

Table of content

I	List of publications.....	I
II	List of abbreviations	II
III	List of tables.....	VI
IV	List of figures.....	VII
1	Summary.....	1
2	Zusammenfassung.....	2
3	Preamble	4
4	Introduction.....	5
4.1	Biological membranes.....	5
4.2	Proteins in biological membranes.....	6
4.3	Transport of small molecules across biological membranes.....	8
4.4	Passive and active transport	10
4.5	Structural biology on integral membrane protein.....	13
4.6	Producing integral membrane protein samples for MX and SPA cryo-EM....	15
4.7	Macromolecular X-ray crystallography, on integral membrane proteins	16
4.8	Single particle analysis cryogenic electron microscopy, on integral membrane proteins.....	18
4.9	Proton coupled oligopeptide transporters (POTs).....	23
4.10	Peptide absorption and reabsorption by PepT1 and PepT2.....	24
4.11	Role of PepT1 and PepT2 in drug disposition.....	25
4.12	Role of SLC15 transporters in disease and inflammation	26
4.13	Previous structural biology data, and open questions on POTs.....	29
5	Aim of the work.....	32
6	Results	33
6.1	Promiscuity in prototypical POTs.....	33

6.1.1	Structure determination of DtpB in complex with 14 peptides.....	33
6.1.2	Plasticity of the binding site of DtpB.....	38
6.1.3	Discussion.....	44
6.2	Transport in the SLC15 family.....	45
6.2.1	Crystallisation trials on HsPepT2.....	45
6.2.2	Structure determination of HsPepT2.....	50
6.2.3	Structure determination of HsPepT1.....	59
6.2.4	Structural basis for peptide transport by PepT1 and PepT2	69
6.2.5	Discussion.....	74
6.3	Structure determination of PHT1	77
6.3.1	SPA cryo-EM structure of PHT1 in complex with a synthetic nanobody	77
6.3.2	Structural basis for Sb27 binding to PHT1	80
6.3.3	Putative binding of TASL to PHT1.....	81
6.3.4	Discussion.....	84
6.4	Selectivity in atypical POTs	85
6.4.1	Crystallisation trials on DtpC	85
6.4.2	Structure determination of DtpC by SPA cryo-EM	87
6.4.3	Discussion.....	96
7	Conclusion and outlook.....	97
8	Materials and methods	98
8.1	Materials	98
8.2	Methods.....	108
8.2.1	Transformation of bacteria with DNA:	108
8.2.2	Polymerase chain reaction (PCR):	108
8.2.3	Gene sequencing.....	108
8.2.4	Insertion of the genes of interest into DNA vectors	108
8.2.5	Expression and purification of saposin A protein.....	109
8.2.6	Expression and purification of nanobodies and macrobodies	109
8.2.7	Expression and purification of prokaryotic membrane proteins.....	110
8.2.8	Expression and purification of mammalian membrane proteins.....	111
8.2.9	Reconstitution of HsPepT2 in saposin lipid nanoparticles (SapNP)	111
8.2.10	Protein concentration determination	112

8.2.11	Electrophoresis.....	112
8.2.12	Transport assays with HsPepT2.....	112
8.2.13	Thermal unfolding of proteins.....	113
8.2.14	3D structure predictions.....	113
8.2.15	General crystallization trial workflow performed on bacterial and mammalian membrane proteins.....	113
8.2.16	Specifics of the co-crystallizations of DtpB-Nb132 with di- and tripeptides 114	
8.2.17	SPA cryo-EM on HsPepT1, HsPepT2, GgPHT1, and DtpC.....	115
9	References.....	120
10	Contributions.....	134
11	Acknowledgements	135

I) List of publications

1. Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes. **Maxime Killer**, Jiri Wald, Joanna Pieprzyk, Thomas C. Marlovits, and Christian Löw. *Science Advances*. 2021. DOI: 10.1126/sciadv.abk3259.
2. Cryo-EM structure of an atypical proton-coupled peptide transporter: Di- and tripeptide permease C. **Maxime Killer**, Giada Finocchio, Haydyn D.T. Mertens, Dmitri I. Svergun, Els Pardon, Jan Steyaert, Christian Löw. *Frontiers in Molecular Biosciences*. 2022. DOI: 10.3389/fmolb.2022.917725

II) List of abbreviations

°	Degree
°C	Degree Celsius
µg	Microgram
µL	Microliter
µM	Micromolar
µs	Microseconds
2D	Two-dimensional
3D	Three-dimensional
Å	Angstrom
ABC transporter	ATP binding cassette transporter
AK-AMCA	β-Ala-Lys-N-7-amino-4-methylcoumarin-3-acetic acid
BBH	Bundle bridge helix
BL	Brain lipid extract
BRIL	Apocytochrome b562
CAC	Colitis associated cancer
CD	Crohn's disease
CMC	Critical micellar concentration
CSSB	Centre for system structural biology
CTF	Contrast transfer function
DAMP	Dmage associated molecular pattern
DDM	n-Dodecyl-β-D-maltopyranoside
DM	n-Decyl-β-D-maltopyranoside
DNA	Deoxyribonucleic acid
ECD	Extracellular domain
EMBL	European molecular biology laboratory
dNTP	Deoxyribonucleoside triphosphate
DtpA, B, C, D	Dipeptide and tripeptide permease A, B, C, D
<i>E. coli</i>	<i>Escherichia coli</i>
g	Gram or gravitational force
GPCR	G-protein-coupled receptor
h	Hours
H ⁺	Proton

HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICH	Intracellular helices domain
IBD	Inflammatory bowel disease
IMAC	Immobilized metal affinity chromatography
IMP	Integral membrane protein
IPTG	Isopropyl β -d-1-thiogalactopyranoside
IRF	Interferon regulatory factor
K ⁺	Potassium ion
K _D	Dissociation constant
kDa	Kilodalton
L	Liter
LB	Lysogeny broth medium
LCP	Lipidic cubic phase
LMNG	Lauryl maltose neopentyl glycol
mAU	Milli absorption units
MDP	Muramyl dipeptide
MFS	Major facilitator superfamily
mg	Milligram
min	Minutes
mL	Millilitre
mM	Millimolar
ms	Milliseconds
mV	Millivolt
MX	Macromolecular X-ray crystallography
Nb	Nanobody
nanoDSF	Nano differential scanning fluorimetry
NF κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NG	n-Nonyl β -D-glucopyranoside
ng	Nanogram
Ni-NTA	Nickel nitrilotriacetic acid
nL	Nanolitre
NLR	NOD like receptor
nm	Nanometre

NM	n-Nonyl- β -D-maltopyranoside
NOD	Nucleotide Binding Oligomerization Domain Containing
OD600 nm	Optical density at 600 nm
PAMP	Pathogen associated molecular pattern
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein data bank
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PepT1/2	Peptide transporter 1/2
PHT1/2	Peptide and histidine transporter 1/2
PI	Phosphatidylinositol
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POT	Proton-dependent oligopeptide transporte
PS	Phosphatidylserine
psi	Pound-force per square inch
PRR	Pattern recognition receptor
RIPK	Receptor interacting protein kinase
RNA	Ribonucleic acid
RT	Room temperature
s	Seconds
ssRNA	Single stranded RNA
Salipro	Saposin lipid protein
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLC15	Solute carrier 15 family
SM	Sphingomyelin
SPA cryo-EM	Single particle analysis cryogenic electron microscopy
TAK1	Transforming growth factor beta-activated kinase 1
TASL	TLR adaptor interacting with endolysosomal SLC15A4
TB	Terrific broth medium
TBST	Tris-buffered saline with Tween20

TCEP	Tris(2-carboxyethyl)phosphine
TLR	Toll like receptor
TM	Transmembrane helix
T _m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
Tri-DAP	diaminopimelic acid containing tripeptides
UC	Ulcerative colitis
keV	Kiloelectron volt
v/v	Volume per volume
WT	Wildtype
λ	Wavelength

III) List of tables

Table 1. MX data collection and refinement statistics of DtpB-Nb132-peptide datasets	37
Table 2. List of chemicals used in this work	98
Table 3. List of enzymes used in this work.....	101
Table 4. List of consumables used in this work	102
Table 5. List of commercial crystallisation screens used in this work	104
Table 6. List of devices and instruments used in this work	104
Table 7. List of cellular strains used in this work.....	107
Table 8. Data collection and refinement statistics of the deposited HsPepT1 and HsPepT2 structures	117
Table 9. Data collection and refinement statistics of the deposited DtpC structure	118
Table 10. Data collection and refinement statistics of the PHT1-Sb27 structure.....	119

IV) List of figures

Figure 1. Principal lipid components of biological membranes.....	5
Figure 2. Types of proteogenic amino acids	7
Figure 3. Membrane protein classification and typology.....	8
Figure 4. Selective permeability of lipid bilayers.....	9
Figure 5. Small molecule transport mechanisms across biological membranes.....	10
Figure 6. Examples of transport proteins.....	11
Figure 7. Alternate access mechanism in MFS transporters	12
Figure 8. Methods of choice for high resolution structure determination of integral membrane proteins.....	14
Figure 9. Role of the SLC15 transport gene family in metabolism and inflammation.....	28
Figure 10. Characterization and complex formation of selected nanobodies with DtpB	33
Figure 11. Structure of DtpB bound to Nb132, and the dipeptide Ala-Ile.....	35
Figure 12. OMIT maps excluding the 14 modelled peptides.....	36
Figure 13. Definition of the binding pocket of DtpB.....	38
Figure 14. Fitting of side chains in the P1 pocket	40
Figure 15. Structural determinants of multi-substrate recognition in DtpB.....	41
Figure 16. Versatility of the P2 pocket of DtpB	43
Figure 17. Substrate's C-termini, and side chains fitting in the P3 pocket	44
Figure 18. Transport assays in human embryonic kidney (HEK) cells	46
Figure 19. Purification strategies of HsPepT	48
Figure 20. Stabilisation of purified HsPepT2 in presence of the dipeptide Ala-Phe.....	49
Figure 21. Strategies to tackle glycan heterogeneity in HsPepT2.....	50
Figure 22. Initial SPA cryo-EM trial on HsPepT2 purified in DDM-CHS.....	51
Figure 23. Reconstitution of HsPepT2 in Salipro	51
Figure 24. SPA cryo-EM on HsPepT2 in Salipro.....	53
Figure 25. SPA cryo-EM on HsPepT2 with the detergent LMNG	54
Figure 26. Final SPA cryo-EM trial on HsPepT2 with the detergent DDM-CHS	55
Figure 27. Cryo-EM data processing for inward facing partially occluded HsPepT2 bound to Ala-Phe...56	
Figure 28. Modelling of HsPepT2 inside the density map.....	57
Figure 29. Atomic model IF partially occluded HsPepT2 bound to Ala-Phe, shown as ribbon representation	58
Figure 30. Amphipathic nature of the bundle bridge.....	59
Figure 31. Initial SPA cryo-EM trials on HsPepT1	60
Figure 32. Cryo-EM structures of HsPepT1	61
Figure 33. Cryo-EM data processing for outward facing, open, apo-HsPepT1 (OF, open, apo-HsPepT1).....	62
Figure 34. Modelling of OF, open apo-HsPepT1 inside the density map.....	63
Figure 35. Cryo-EM data processing for outward facing, open and occluded, Ala-Phe bound-HsPepT1 (OF, open, and occluded Ala-Phe HsPepT1).....	64
Figure 36. Modelling of OF, open Ala-Phe bound-HsPepT1 inside the density map	66
Figure 37. Architectural differences between bacterial and human POTs.....	67

Figure 38. N-Glycans present on the extracellular domain of HsPepT1.....	68
Figure 39. Superposition of HsPepT2 and the prokaryotic POTs DtpA and PepT_{Sb} in grey bound to Valganciclovir and Valaciclovir	68
Figure 40. Side by side view of the experimentally determined structures of HsPepT1 and HsPepT2.....	70
Figure 41. Structural comparison between the outward and inward facing states observed in apo HsPepT1 and substrate bound HsPepT2	71
Figure 42. Mechanism for substrate recognition and transport in human POTs based on the presented structures.....	72
Figure 43. Development of the central cavity upon substrate binding and its transition to the inward facing state coloured by the electrostatic potential	74
Figure 44. Side by side comparison of PHT1 samples imaged by cryo-EM	77
Figure 45. High resolution structure determination of PHT1-Sb27.....	78
Figure 46. Cryo-EM structure of PHT1-Sb27.....	79
Figure 47. Stabilisation of the OF, open state in PHT1	79
Figure 48. Overview of the binding site of PHT1.....	80
Figure 49. Binding interface between PHT1 and Sb27.....	81
Figure 50. Putative interaction between TASL and PHT1.	82
Figure 51. Crystallisation trials on DtpC.....	86
Figure 52. Utilization of different fiducial markers to improve particle alignment and 2D averaging from cryo-EM images	88
Figure 53. Enlarged representative micrographs and 2D class averages	88
Figure 54. Characterization of the different fiducial markers	89
Figure 55. High resolution structure determination of DtpC-Nb26.....	90
Figure 56. Cryo-EM structure of DtpC-Nb26.....	91
Figure 57. Quaternary structure of complex assembly	92
Figure 58. Structural basis for the stabilization of the inward facing state in DtpC and other POT homologues.....	93
Figure 59. Structural basis for ligand selectivity in DtpC and atypical POTs	95

1 Summary

The uptake of peptides plays a crucial role in metabolism and inflammation. In mammals, peptide absorption and reabsorption are mediated by the proton coupled oligopeptide transporters (POTs) PepT1(SLC15A1) and PepT2 (SLC15A2), of the solute carrier family 15. POTs are one the most promiscuous transporters among solute carriers and constitute the main route of entry for orally administrated peptidomimetic drugs. SLC15 transporters are involved in various inflammatory diseases, and the paralogue PHT1 (SLC15A4) was recently identified as therapeutic target in Systemic Lupus Erythematosus (SLE). The three dimensional structures of several bacterial homologues have been determined in the past 10 years, but how these shuttle systems adapt to such an array of substrate remains poorly understood on the molecular level. In addition, these past snapshots represented exclusively ‘inward facing’ conformations, therefore limiting our molecular understanding of the transitions required to complete an entire transport cycle. In a first step, we determined high resolution three dimensional structures of the prototypical POT DtpB from *E. coli*, bound to 14 different di- and tripeptides, using macromolecular crystallography (MX). This work provides a profound basis for understanding promiscuity and ligand recognition in POTs at the molecular level. Second, I used single particle analysis cryogenic electron microscopy (SPA cryo-EM), to determine the first structures of the human peptide transporters 1 (HsPepT1) and 2 (HsPepT2). Human PepT1 and PepT2 were captured in four different states throughout the transport cycle, providing a dynamic molecular understanding of substrate uptake within the SLC15 family. Third, we continued using SPA cryo-EM to determine the first structure of PHT1, in an outward facing conformation. This work provides a framework to determine the structure of this newly identified target of SLE, which could be used to obtain high resolution data with various therapeutics. Last, I determined the first structure of the atypical POT DtpC, from *E. coli*. In this work, we explored various fiducial strategies, to improve the resolution of the reconstruction of MFS transporters in general, and provided a molecular explanation for the selectivity of DtpC towards positively charged dipeptides. In summary, this work delivers new insights into the working principles of proton coupled oligopeptide transporters, and will serve as a reference for future structure based drug design (SBDD) studies targeting members of this family.

2 Zusammenfassung

Die Aufnahme von Peptiden spielt eine entscheidende Rolle im Stoffwechsel und bei Entzündungen. Bei Säugetieren werden Peptidabsorption und -rückresorption durch die protonengekoppelten Oligopeptid-Transporter (POTs) PepT1 (SLC15A1) und PepT2 (SLC15A2) der Solute-Carrier-Familie 15 vermittelt. POTs gehören zu den promiskuitisten Transportern unter den Solute Carriern und stellen den Haupteintrittsweg für oral verabreichte peptidomimetische Arzneimittel dar. SLC15-Transporter sind an verschiedenen Entzündungskrankheiten beteiligt, und das Paralog PHT1 (SLC15A4) wurde kürzlich als therapeutisches Ziel bei Systemischem Lupus Erythematoses (SLE) identifiziert. Die dreidimensionalen Strukturen mehrerer bakterieller Homologe wurden in den letzten 10 Jahren ermittelt, aber wie sich diese Shuttle-Systeme an eine solche Vielzahl von Substraten anpassen, ist auf molekularer Ebene noch kaum verstanden. Darüber hinaus stellten diese bisherigen Momentaufnahmen ausschließlich nach innen gerichtete Konformationen dar, was unser molekulares Verständnis der Übergänge einschränkt, die für den Abschluss eines gesamten Transportzyklus erforderlich sind. In einem ersten Schritt haben wir mit Hilfe der makromolekularen Kristallographie (MX) hochauflösende dreidimensionale Strukturen des prototypischen POT DtpB aus *E. coli*, gebunden an 14 verschiedene Di- und Tripeptide, bestimmt. Diese Arbeit bietet eine solide Grundlage für das Verständnis von Promiskuität und Ligandenerkennung in POTs auf molekularer Ebene. Zweitens habe ich mit Hilfe der kryogenen Einzelpartikelanalyse-Elektronenmikroskopie (SPA cryo-EM) die ersten Strukturen der menschlichen Peptidtransporter 1 (HsPepT1) und 2 (HsPepT2) bestimmt. Menschliche PepT1 und PepT2 wurden in vier verschiedenen Zuständen während des Transportzyklus erfasst, was ein dynamisches molekulares Verständnis der Substrataufnahme innerhalb der SLC15-Familie ermöglicht. Drittens setzten wir weiterhin die SPA-Kryo-EM ein, um die erste Struktur von PHT1 in einer nach außen gerichteten Konformation zu bestimmen. Diese Arbeit liefert einen Rahmen zur Bestimmung der Struktur dieses neu identifizierten SLE-Ziels, der genutzt werden könnte, um hochauflösende Daten mit verschiedenen Therapien zu erhalten. Schließlich habe ich die erste Struktur des atypischen POT DtpC aus *E. coli* bestimmt. In dieser Arbeit haben wir verschiedene Referenzstrategien untersucht, um die Auflösung der Rekonstruktion von MFS-Transportern im Allgemeinen zu verbessern, und eine molekulare Erklärung für die Selektivität von DtpC gegenüber positiv geladenen Dipeptiden geliefert. Zusammenfassend liefert diese Arbeit neue Einblicke in die

Funktionsweise von protonengekoppelten Oligopeptid-Transportern und wird als Referenz für zukünftige strukturbasierte Wirkstoffdesign-Studien (SBDD) dienen, die auf Mitglieder dieser Familie abzielen.

3 Preamble

Life has the quality of sustaining biological processes in an individual entity. It maintains order in an organism, and increases disorder in the universe, inevitably obeying the second law of thermodynamics. I doubt it is possible, to grasp the totality of Life by *thinking*. This process is, partial, conditioned, and therefore limited in itself. Nonetheless, the scientific approach provides insight fragments, into Nature, including biological processes. These fragments, documented, and stored in the collective memory, are the source of technological progress in medicine. In fact, knowledge expands, as theories and technologies refine each other, through experience.

As of today, the precise sequence of events leading to abiogenesis is still investigated and debated. However, it is clear that the early forms of Life consisted in a self-replicating organic molecule, enclosed in a self-assembling amphipathic bilayer. Cellular membranes - together with physical and chemical sources of energy present in the environment (geothermal heat, sunlight, chemical gradients) - allowed to concentrate prebiotic material within open microsystems, endowed with the remarkable ability to maintain a highly ordered state, to replicate, to grow, and therefore, to live.

During this thesis, I have tried to obtain clear three dimensional representations of a particular class of proteins, present in cellular membranes of currently living organisms. These proteins are called proton coupled oligopeptide transporters (POTs), and they mediate the uptake of molecules (i.e., peptides, drugs, protons) from the environment, into the cell.

In the introduction, I will first describe fundamental aspects of biological membranes, proteins, transporters, and POTs. Then, I will explain some methodological and technological aspects relevant to the experiments, performed in this work. In the results and discussion section, the major findings and implications of the investigation will be presented. This work was performed between September 2018 and September 2022 in the group of Dr. Christian Löw at the European Molecular Biology Laboratory (EMBL), and at the Centre for System Structural Biology (CSSB), in Hamburg, Germany. The contribution of each person, generating results presented in this dissertation, is detailed in the contribution section.

4 Introduction

4.1 Biological membranes

The dry mass of biological membranes is constituted of about 50% of lipids (Alberts, 2022). Lipids are fundamental to living organisms because they constitute the matrix of membranes, which are necessary to maintain order, and regulate what enters and exits the cell. In addition, they act as signalling molecules, in concert with proteins. Lipids are composed of variable building blocks, which increases the toolbox accessible to cells and organelles, to make up their membranes. In humans and other mammals, membranes are mainly constituted of glycerophospholipids, cholesterol, and sphingolipids (Ingólfsson *et al.*, 2014; Casares, Escribá and Rosselló, 2019) (**Figure 1A**). Lipid molecules possess polar headgroups, and hydrophobic tails of variable length and composition. Most hydrophobic molecules, when present in an aqueous environment, collapse, to decrease the surface exposed to the solvent (i.e., hydrophobic effect). In the case of amphiphilic membrane lipids, the hydrophobic tails dock to each other to obey the same entropic driving force, but the polar headgroup can form favourable electrostatic interactions or hydrogen bonds with water, conferring a specific orientation and organization to the system. The most stable three-dimensional arrangement for this system is to assemble in a double layer sheet (**Figure 1B**), which closes on itself (Alberts, 2022).

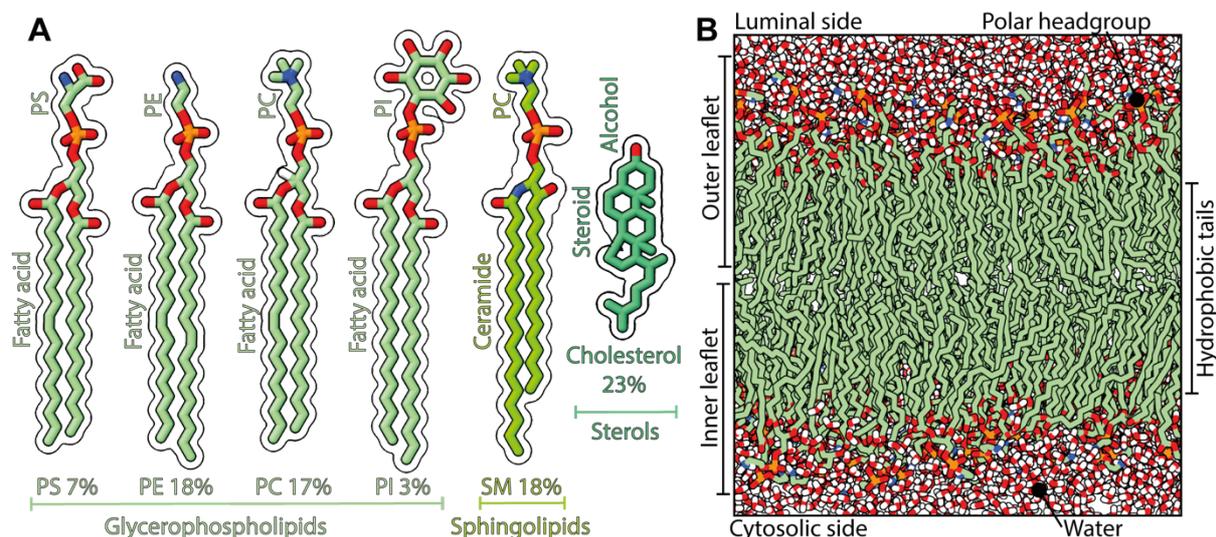


Figure 1. Principal lipid components of biological membranes. (A) Principal lipid components of eukaryotic cell plasma membranes. PS=phosphatidylserine, PE=phosphatidylethanolamine, PC=phosphatidylcholine, PI=phosphatidylinositol, SM=sphingomyelin. The estimate proportions of the various lipid types are indicated as percentages. (B) Organisation of a POPE membrane bilayer in water, simulated by molecular

dynamics. Lipids are coloured as in (A). Water oxygen molecules are coloured in red, and hydrogens in white.

Adding another level of complexity, different lipid types are spread asymmetrically along the inner and outer leaflets of the bilayer, and their stoichiometry can vary across the spherical plane. The lipid composition in a particular region of the membrane affects viscosity, fluidity, and therefore planar diffusion of membrane components. Notably, patchworks of domains or so called ‘lipid rafts’ (named as such, because their increased thickness makes them appear as floating islands on the surface of the bilayer), are enriched in cholesterol, sphingolipids, and specific proteins (Karnovsky *et al.*, 1982; Sezgin *et al.*, 2017). This spatial organization can help to cluster proteins working toward a common goal, and therefore favours the overall kinetics of certain signalling processes.

4.2 Proteins in biological membranes

Proteins constitute the other half of membranes’ dry mass. They also crowd the cytoplasm, and are extremely versatile macromolecules performing the majority of all biological processes. The identity of a protein is encoded genetically in the DNA, transcribed into RNA and translated into a polypeptide chain of L-amino acid residues (**Figure 2**) linked together by peptide bonds (Alberts, 2022). Importantly, the sequence of amino acid residues confers structural and surface properties to a protein. Within a given environment (physical, chemical, biological), a protein can interconvert between an ensemble of more or less stable akin conformations. This dynamic aspect of proteins is key to functionality, but requires first proper folding of the polypeptide chain. The folding of a protein into a biologically relevant state, is driven by the hydrophobic collapse effect, similarly as previously described with lipids in aqueous solvents. Folding is also influenced by many weak non covalent bonds (i.e., hydrogen bonds, van der Waals attractions, and electrostatic interactions) (Anfinsen, 1973; Dobson, 2003). Domains, composed of different secondary structure elements such as α helices, β sheets, or loops, can fold more or less individually within a protein, and work either autonomously, or collectively, to perform a given function. This holds true for entire proteins which oftentimes interact with other proteins to form complexes. Because sequence, structure, dynamics, and function are genuinely associated, the analogy of molecular machines is often made when referring to proteins or protein complexes. The wide chemical diversity of amino acids (**Figure 2**), combined with the immense possibilities of three and four dimensional configurations that a polypeptide chain can adopt, make proteins the most complex and

sophisticated biological molecules to exist. In fact, the plethora of apparatus, motors, and mechanisms we witness today has evolved through natural genetic modifications and selective pressure over the course of 3.5 billion years.

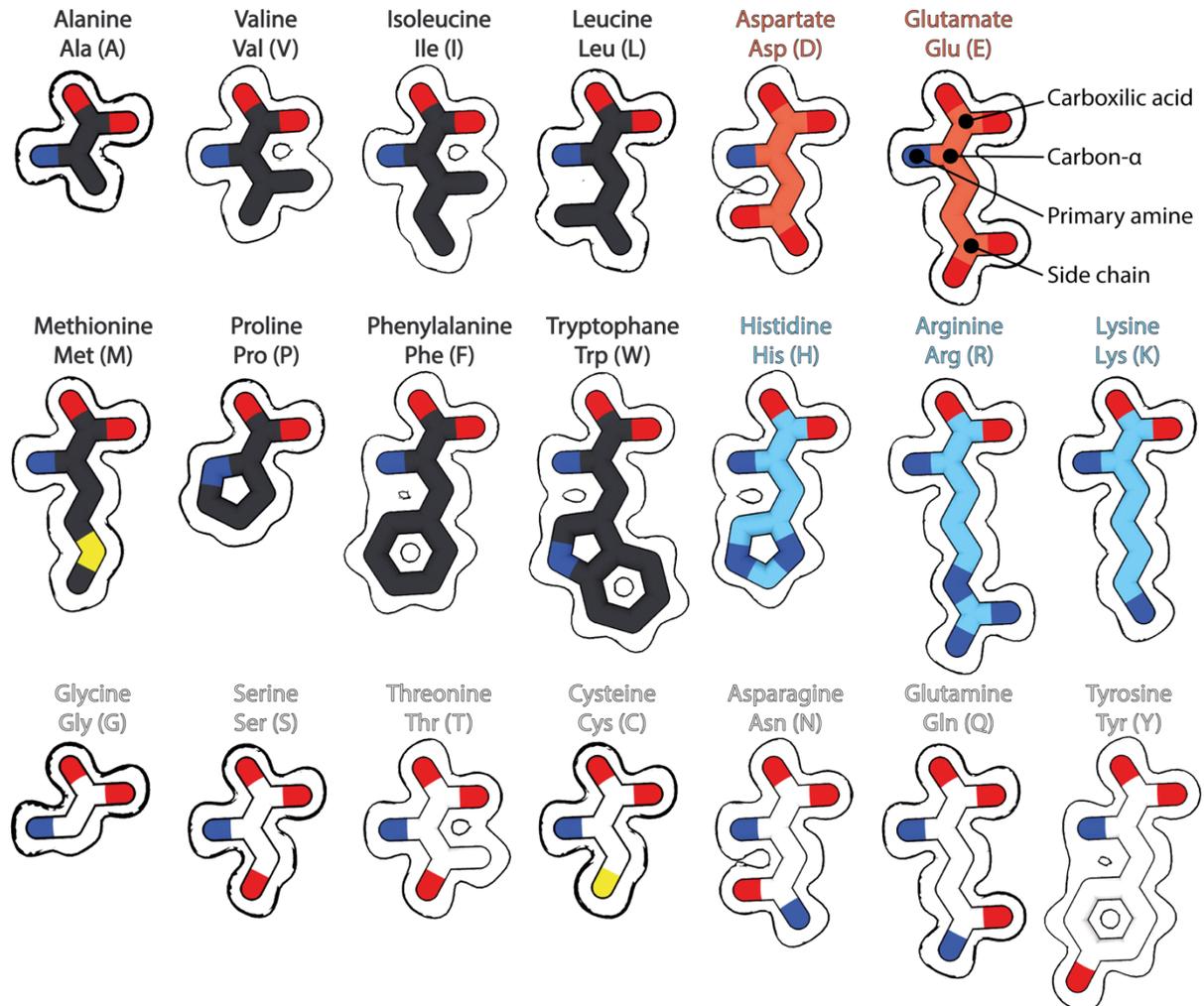


Figure 2. Types of proteogenic amino acids. Carbon atoms of apolar, polar, negatively charged, and positively charged amino-acids are coloured in black, white, orange, and light blue, respectively.

For instance, proteins can catalyse chemical reactions (enzymes), allow communication between cells and their environment (hormones and receptors), transfer molecules inside and outside the cell (transporters), protect against pathogens (antibodies and complement system), gather ions and molecules (storage proteins), or support cellular and extracellular architecture (cytoskeleton, fibrils). Cell membranes harbour proteins carrying out such tasks. Proteins can interact with the lipid bilayer in various ways (Alberts, 2022) (**Figure 3**). Membrane associated proteins or peripheral proteins are docked to the membrane by weak interactions with other membrane proteins or lipids. Lipid-anchored membrane proteins are covalently bound to lipids

with myristoyl, palmitoyl, and farnesyl anchors. Most membrane proteins however, are inserted inside the bilayer. Monotopic proteins are embedded in the membrane but do not span it. Integral membrane proteins are constituted of single (i.e., bitopic type I and II IMPs) or multiple alpha α helical transmembrane helices (i.e., polytopic type III IMPs), and β -barrels (rolled-up β sheets) extending across the orthogonal axis of the membrane. IMPs can also oligomerize (type IV IMPs). In type III and IV IMPs, the side chains of apolar amino acids are often pointing towards the hydrophobic membrane, while polar and charged residues are shielded towards the interface between transmembrane helices, where they can play important functional roles.

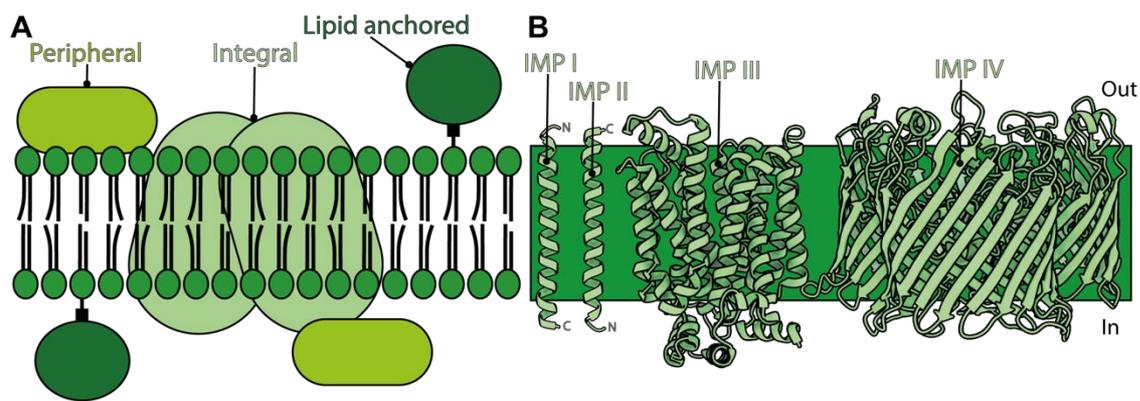


Figure 3. Membrane protein classification and typology. (A) The various types of associations between membranes and proteins. (B) Types of integral membrane proteins (IMP).

4.3 Transport of small molecules across biological membranes

As previously stated, membranes allow the cell or the organelle to establish and maintain a different molecular environment within the delimited space. Given enough time, any molecule would eventually diffuse through a membrane to reach thermodynamic equilibrium. However, cells require to uptake specific metabolites, and discard waste products in defined quantities and narrow time frames. For instance, while gases and hydrophobic molecules such as O_2 , CO_2 , and steroid hormones, flow spontaneously through lipid bilayers, hydrated, polar, and charged molecules, have extremely slow rates of diffusion (**Figure 4**).

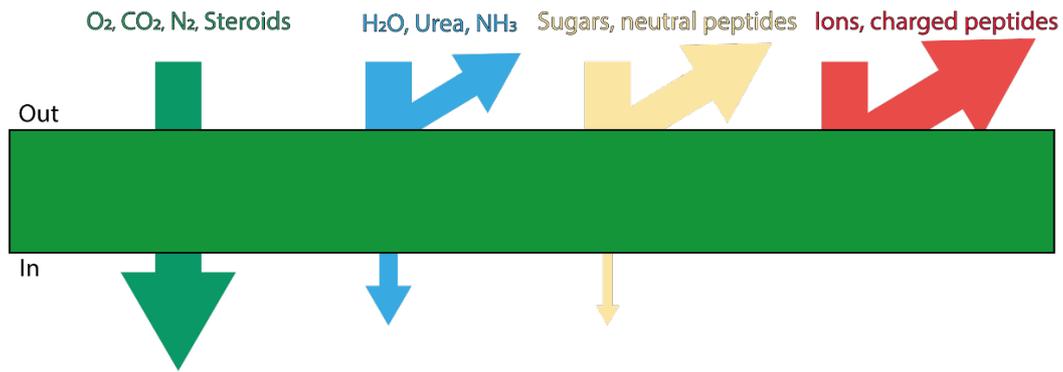


Figure 4. Selective permeability of lipid bilayers. Lipid bilayers are more or less permeable to small molecules, as a function of their size, and polarity.

To tackle this challenge, cells have evolved with transmembrane transport proteins which constitute up to a third of all membrane protein coding genes in certain organisms, and which consume up to two third of the total metabolic energy, in epithelial cells and in neurons (Engl and Attwell, 2015). Transporters and channels are the two types of membrane transport proteins (**Figure 5**). Transporters bind to small molecules and translocate them across the membrane by undergoing large conformational changes (Dyla *et al.*, 2020; Drew *et al.*, 2021). The energetic cost of such movements can be paid by downhill concentration gradients, or ATP hydrolysis . Channels mediate extremely fast transport (up to 100 million ions per second) of inorganic ions, downhill their concentration gradients (Di Resta and Becchetti, 2010). Channels are highly selective gated pores, and can open, or close, upon binding of a ligand (e.g., neurotransmitter, ion, nucleotide), mechanical stress (e.g., physical touch, contact), or change of membrane potential (i.e., voltage across the bilayer, electrochemical gradients). Ion channels trigger all the processes requiring electrical activity in our body. They spark - in fact, they *gate* - our senses, our feelings, heartbeats, our movements, thoughts and memories. They fundamentally condition our experiences, pains, pleasures, addictions; and defects in their regulation, or activity, lead to severe diseases. Since this thesis focuses on peptide transporters, we will now leave ion channels behind, and elaborate on transporters.

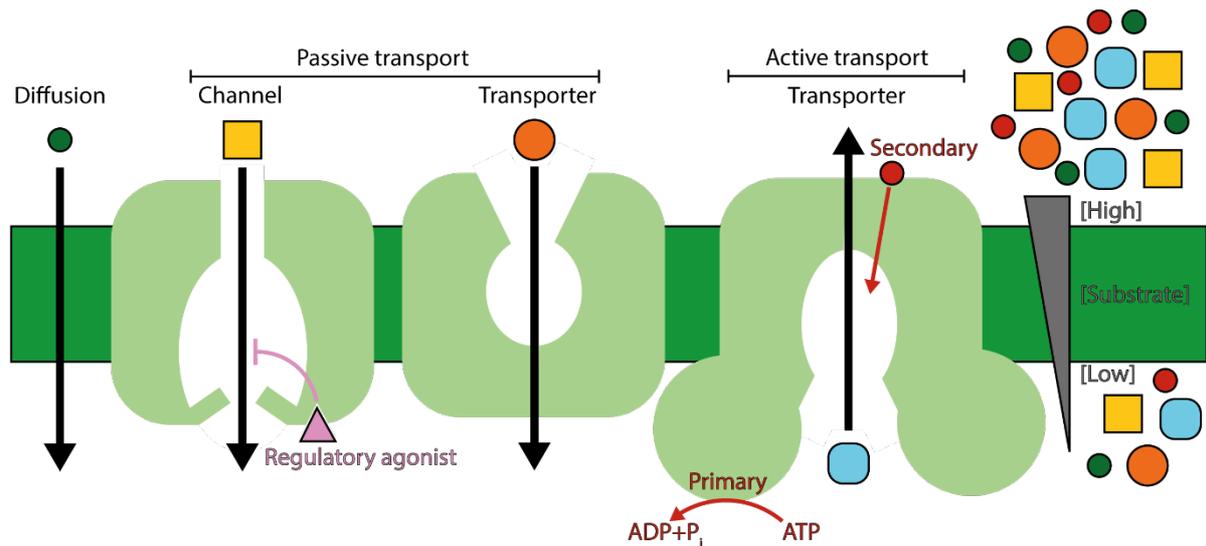


Figure 5. Small molecule transport mechanisms across biological membranes. From left to right: small hydrophobic molecules, such as steroids can cross the bilayer. This process is called simple diffusion. Polar molecules can be transported by channels and transporters, downhill their concentration gradients. This process is called facilitated, or passive transport. The gating of channels can be regulated by agonist molecules, membrane potential, or mechanical stress. When the substrate is to be transported uphill its concentration gradient, an additional source of energy is required. This process is called active transport. The potential energy powering the conformational changes of the transporter can be stored as ATP (primary active transport), or as electrochemical gradient of a coupled ion (secondary active transport).

4.4 Passive and active transport

Transporters can use light (Grotjohann and Fromme, 2005), ATP hydrolysis (Dyla *et al.*, 2020), electrochemical gradients (Bosshart and Fotiadis, 2019), or substrate concentration gradients (Mueckler and Thorens, 2013) as driving force to power large conformational changes (**Figure 5**). These motions alternatively expose and shield a substrate specific central binding cavity to the extracellular and intracellular milieu. When the transport is mediated from a compartment with lower concentration of substrate, to a compartment with higher concentration (uphill gradient), the transporter requires either light, ATP hydrolysis (i.e., primary active transport), or a coupled downhill electrochemical gradient (i.e., secondary active transport, or coupled transport). Otherwise, downhill substrate concentration gradient can constitute a sufficient onset for facilitated transport (i.e., passive transport).

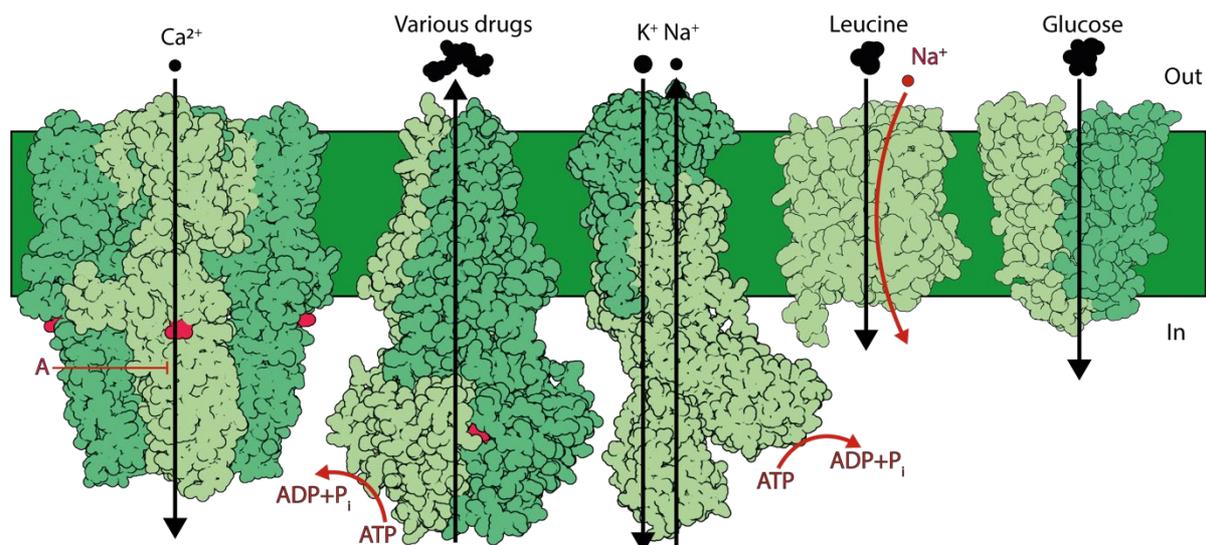


Figure 6. Examples of transport proteins. From left to right: the TRPA1 channel is a sensor for pain, cold, and itch. It mediates the fast influx of calcium ions, and is regulated by various agonist and antagonist, such as the one shown in red, labelled ‘A’. PDB accession number 6V9V. The multidrug resistance (MDR) P-glycoprotein uses the energy of ATP hydrolysis to pump molecules, such as chemotherapeutics, outside of the cell. PDB accession number 6C0V. The sodium-potassium pump establishes and maintains chemical gradients of K^+ and Na^+ , using ATP hydrolysis as energy source. PDB accession number 2ZXE. The Leucine transporter LeuT, transports leucine and other apolar amino-acids inside the cell, uphill their concentration gradients. LeuT uses the potential energy stored in sodium gradients to power alternate access. PDB accession number 3TU0. GLUT3 facilitates the downhill transport of glucose across the plasma membranes of neurons. PDB accession number 4zw9.

Primary active transport is mediated by pumps (**Figure 6**). Light or redox driven pumps (e.g., proton pumps and F-type ATP synthases, from photosynthetic, bacterial, and mitochondrial electron transport chains), are the fundamental generators of potential energy, stored in the form of proton gradients and ATP (Kühlbrandt, 2019). ATP driven pumps (e.g., P-type ion pumps and V-type proton pumps) use the free energy released upon hydrolysis of previously synthesised ATP into ADP and phosphate, to establish or maintain further proton, sodium, potassium, and calcium gradients (Dyla *et al.*, 2020). The latter, greatly increases the diversity of potential energy forms at hand, and is largely used by secondary active transporters (and ion channels!) (Di Resta and Becchetti, 2010; Bosshart and Fotiadis, 2019). Finally, ATP binding cassette (ABC) transporters pump organic molecules inside or outside compartments using - here as well - energy released upon the hydrolysis of ATP. An example is the P-glycoprotein, or multidrug resistance (MDR) protein, which pumps out a variety of compounds outside the cell, and is highly expressed in certain cancer cells. In such cases, chemotherapy can be rather ineffective (Waghray and Zhang, 2018).

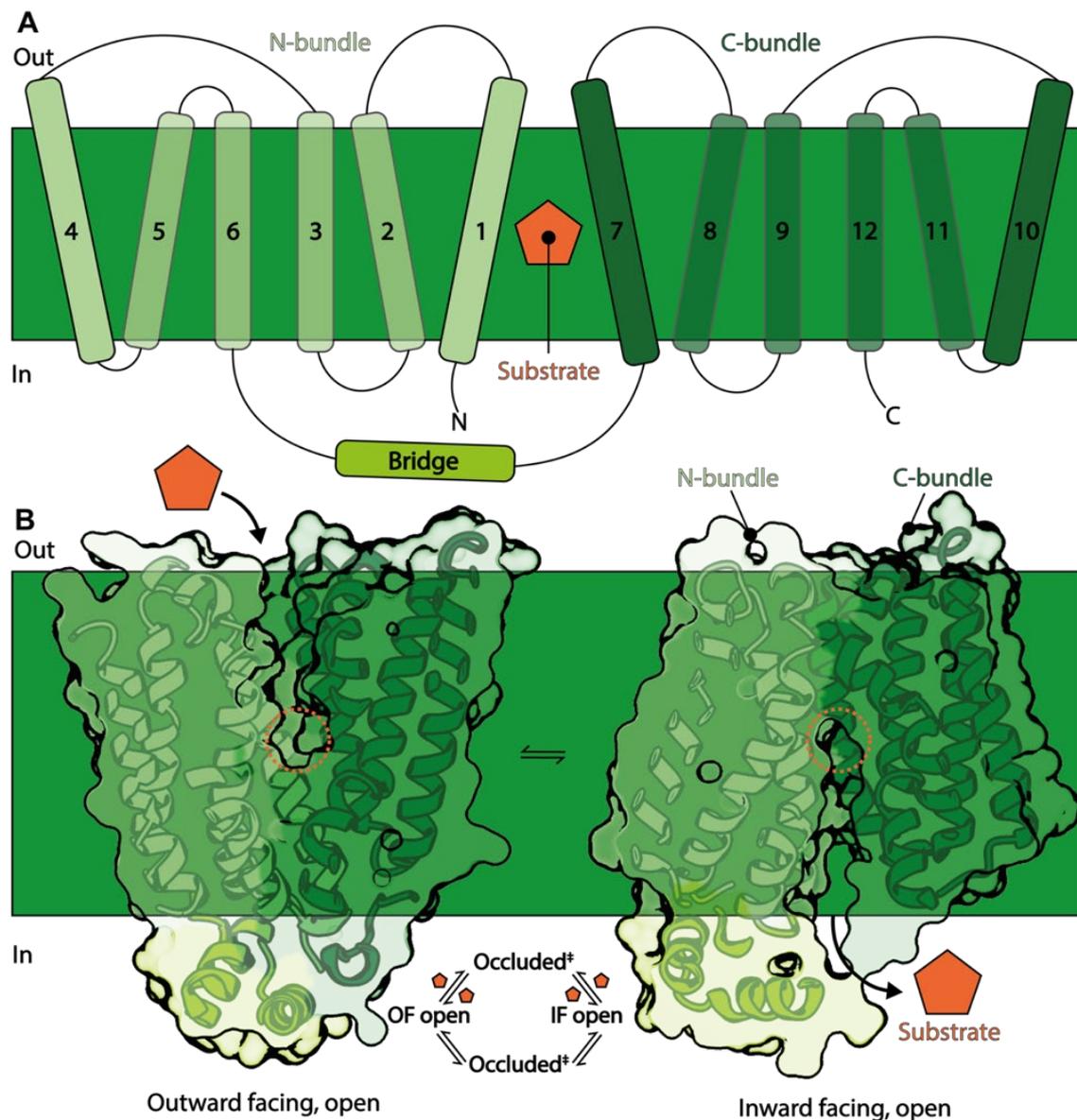


Figure 7. Alternate access mechanism in MFS transporters. (A) Typology of the MFS fold. The MFS fold is characterised by two six-helix bundles connected by a linker. The N-terminus, C-terminus, and the linker, are present on the cytosolic side of the membrane. (B) The substrate fits in the central cavity, which is alternatively exposed to the extracellular, or luminal environment (outward facing, open state), to the intracellular space (inward facing, open state). Alternate access is thought to be achieved through conformational changes in the transporter from the OF to the IF state, with occluded transition states, in the apo- or substrate-bound scenario. The displayed MFS transporter is the fructose transporter GLUT5, in the apo-OF, open and apo-IF, open states. PDB accession numbers are respectively 4YBQ and 4YB9 respectively.

Secondary active transport is mediated by ion gradient dependant, or proton gradient dependant, solute carriers (**Figure 6**). These transporters mediate the disposition of small molecules such as sugars, nucleotides, lipids, amino acids, vitamins, neurotransmitters, and

various drugs, to name a few (Bosshart and Fotiadis, 2019). Based on sequence similarity and functionality, mammalian solute carriers have been assigned into gene families of various sizes (SLC nomenclature) (Liu, 2019). For instance, the SLC15 gene family regroups solely proton coupled oligopeptide transporters, while the SLC6 family covers sodium coupled serotonin, dopamine, norepinephrine, GABA, taurine, creatine, as well as amino acid, transporters. The greater number of SLCs belong to the major facilitator superfamily (MFS.), which regroups transporters sharing a common three dimensional fold of twelve transmembrane helices organized in two six-helix bundles, and a central binding site (Quistgaard *et al.*, 2016a; Drew *et al.*, 2021) (**Figure 7A**). The first high resolution three dimensional structure of an MFS transporter was determined in 2003 for the lactose permease from the bacterium *Escherichia coli* (*E. coli*) (Abramson *et al.*, 2003). Since then, other related structures, have revealed some of the different conformations that MFS transporters can adopt in order to translocate their substrate across membranes (Drew and Boudker, 2016; Quistgaard *et al.*, 2016b; Drew *et al.*, 2021). Although the ensembles of states observed so far, can vary significantly between sub-families, or individual transporters, an overall alternate access mechanism still holds valid for all MFS members. Indeed, transport can be described as a cycle between at least three postulated states: (i) inward-facing open, (ii) occluded, and (iii) outward-facing open (**Figure 7B**). This allows one side of the transporter to be opened at one time, to bind the solute, use an energy input to power the transition to an unstable intermediate state, expose the binding site to the other side of the bilayer, and finally to release the substrate(s). The repetition of this sequence is called transport cycle.

4.5 Structural biology on integral membrane protein

Black holes, planets, facial expressions, cells, molecules, atoms; for every object present in the universe: observation, when attentive, provides information. Structural biology is a discipline, in which biological macromolecules are looked at, using advanced reconstruction and imaging techniques. The picture of a protein obtained in a controlled condition, provides a snapshot, of an event. There is intelligence in that. Various parameters can then be modified (i.e., time, energy, ligand), to facilitate the understanding of the protein's working principles, and the biochemical process it is involved in. Consecutively, fundamental knowledge in biology expands, which promotes the development of applied sciences, like drug discovery and medicine. When fine details of a protein are to be visualized (e.g., functional groups of a side chain interacting with a ligand), then the two method of choices are macromolecular X-ray

crystallography (MX) and single particle analysis cryogenic electron microscopy (SPA cryo-EM) (**Figure 8**). Both techniques have their own limitations and are therefore complementary (Wang and Wang, 2017). One common requirement is the production of a sample containing many copies of the protein, and preferably isolated from the cellular complexity. The latter tends to facilitate the crystallisation process in MX, and the image analysis in SPA cryo-EM.

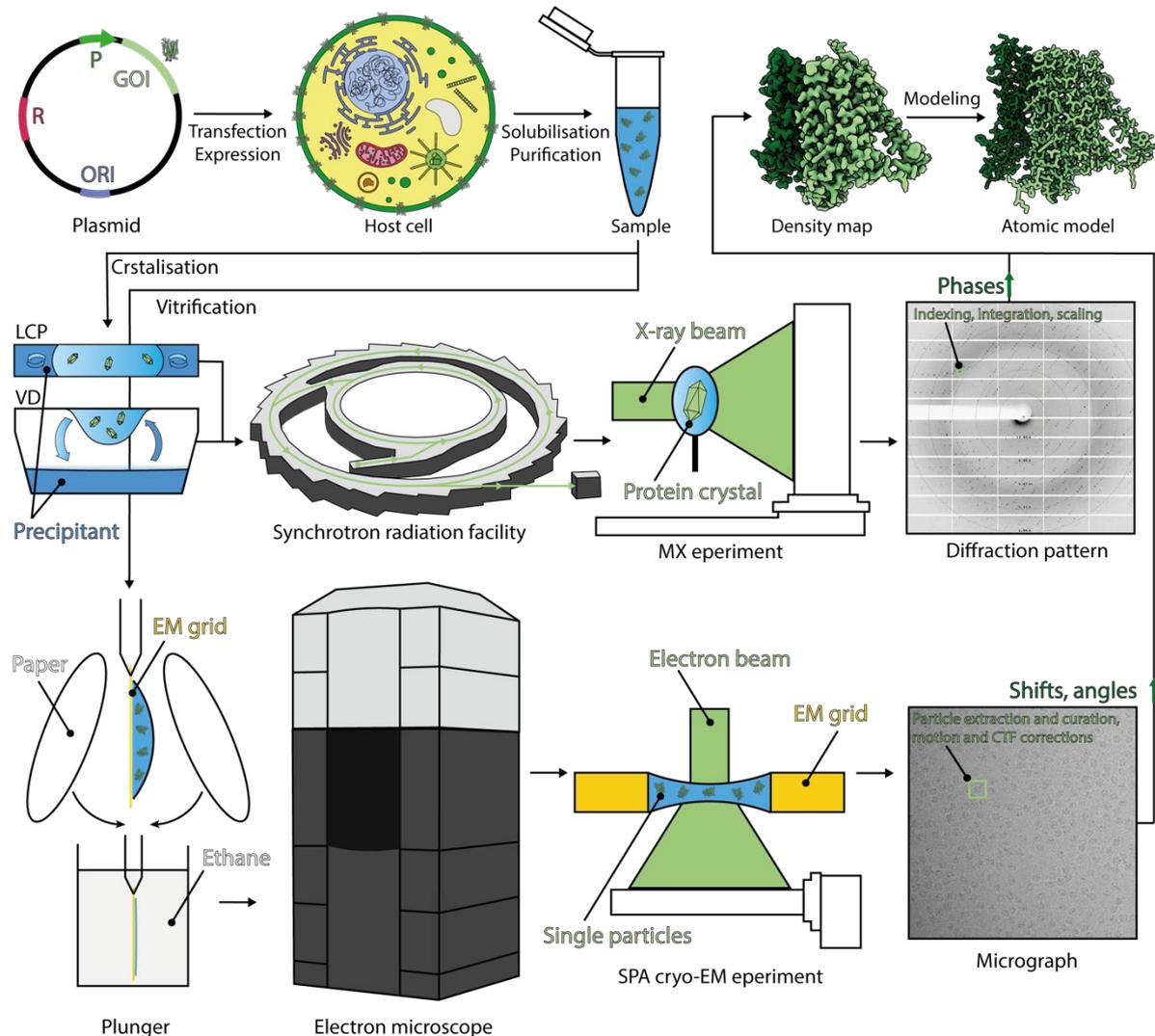


Figure 8. Methods of choice for high resolution structure determination of integral membrane proteins. The workflow typically starts with construct design and cloning into a DNA vector (plasmid), containing the gene of interest (GOI), a selection marker such as an antibiotic resistance gene (R), and an origin of replication (ORI). The plasmid is then transfected in a host cell (here, an animal cell is depicted, but bacterial hosts are common for prokaryotic targets, and cheaper to grow), to produce the protein through transcription-translation. The cell is then lysed and the membrane fraction solubilised with mild detergents. Direct extraction in nanodiscs or Salipro can be performed instead, for SPA cryo-EM only. The protein of interest is typically enriched by affinity purification techniques, using affinity ‘tags’ encoded in the designed construct. If well diffracting crystals can be obtained for the protein

of interest (either by vapour diffusion, VD, or by the lipidic cubic phase method, LCP), MX should be the method of choice. MX benefits from the great signal amplification by the crystal, and over fifty years of hardware and software development, making the technique extremely fast and effective. Typical MX experiments take place in advanced synchrotron radiation facilities. The data can be analysed on a local computer. If no crystals can be obtained, SPA cryo-EM is the technique to turn to. There, the purified sample is directly vitrified onto an EM grid, and imaged using an electron microscope. The computational cost of SPA cryo-EM is much heavier, and requires high performance computing machines.

4.6 Producing integral membrane protein samples for MX and SPA cryo-EM

The major difficulties in producing membrane protein samples for MX and SPA cryo-EM, are to (i) obtain high copy numbers of the target inserted in a host cell membrane and (ii) maintain the integrity of the protein after extraction from the host cell membrane (Walian, Cross and Jap, 2004; Choy *et al.*, 2021). To tackle the first challenge, different host organisms can be used, deepening on the natural origin of the protein. For instance, PepT1 and PepT2 of human origin, should preferably be produced in mammalian host cells. There, particular machineries involved in protein glycosylation and cellular trafficking are present, and match the needs of such targets. Naturally, bacterial host systems are compatible with prokaryotic targets, such as the *E. coli* POTs DtpB and DtpC. The gene coding for a membrane protein are usually cloned into an expression vector, containing engineered promoter DNA sequences, boosting the transcription process, and consequently, the number of copies of the protein of interest. The gene of interest is often supplemented with upstream or downstream polypeptide coding tags, allowing affinity chromatography purification techniques. To tackle the second challenge, a large portfolio of membrane mimetics has been developed in the field (Zhang and Cherezov, 2019). Detergents are widely used to perform this step. Detergents are amphipathic molecules, like lipids, but the structure of their hydrophobic tails forces them to form micelles in aqueous solvents. An important characteristic is the critical micellar concentration (CMC). This value indicates the concentration at which a particular detergent molecule will arrange in micelles. Detergents, (particularly: maltosides and glucosides, followed by amine oxides and polyoxyethylene glycols), are used to shield the transmembrane region of the protein from water, and thus allow the transfer of the protein in aqueous solvents. This extraction can however result in unfolding or loss of function of the protein. In such cases, several detergents can be tried out, and sample integrity can be quickly checked by steric exclusion chromatography (SEC) and biophysical methods (e.g., thermal unfolding, ligand binding). For

mammalian membrane proteins in particular, the addition of lipids (such as cholesterol) to the detergent of choice, can make a critical difference (Choy *et al.*, 2021). When detergent-lipid mixtures do not yield folded samples, good alternatives are membrane scaffold protein (MSP) nanodiscs (Bayburt and Sligar, 2010), and saposin-lipid-protein nanoparticles (Salipro) (Frauenfeld *et al.*, 2016). Both methods rely on amphiphilic proteins, wrapping around lipids, or profitably, around the transmembrane region of a protein, imbedded in lipids. The experimental conditions (e.g., molar ratio of lipid to protein, temperature, incubation time, etc.), for successful reconstitution, can be tedious to determine, and often requires a trial and error. It should also be added, that nanodiscs and Salipro are not suitable for macromolecular x-ray crystallography.

4.7 Macromolecular X-ray crystallography, on integral membrane proteins

Quantities of purified protein, in the milligram range, are sufficient to get started, and screen for a few hundreds of crystallisation conditions, using liquid dispensing robots. Yet, obtaining such amounts can be a major challenge for many integral membrane proteins (Kermani, 2021). A crystal nucleates when the protein slowly precipitates out of solution, under specific conditions (Dessau and Modis, 2011). The system ‘protein-solution’ is set by (i) the nature of the purified sample (i.e., unique properties of the protein, type of detergent, and the concentrations of these two components), and (ii) the crystallisation conditions it is incubated in (mother liquor, temperature). The chemical composition of the system is then allowed to evolve slowly by osmosis or vapor diffusion (VD), hopefully meeting a condition favourable to nucleation (McPherson and Cudney, 2014). If so, the crystal grows until thermodynamic equilibrium between the crystal, the remaining protein in solution, and the mother liquor components is reached. Crystals constitute a network, or lattice, of periodically repeated volumes, called unit cells. Unit cells are constituted of one or several molecules adopting similar conformations and orientations. Typical useful protein crystals in MX have a size ranging between 1 μm to 1 mm, and can therefore contain several billions of copies of the protein, all related by symmetry operations, and adopting a homogenous conformational ensemble. This constitute an immense signal amplifier, and is one of the greatest asset of MX. Ironically, this also constitutes the main bottleneck of the technique. Without a crystal, the experiment does not take place. In fact, most eukaryotic IMPs have long flexible loops linking transmembrane helices, and are heavily glycosylated. This brings a great deal of flexibility and

heterogeneity to the protein, and usually prevents crystallisation (Choy *et al.*, 2021). In such cases, construct optimisation is necessary and requires more work. It should be noted as well, that no protein crystal is ever ordered perfectly, and typically requires some degree of optimization of the growing conditions, anyhow (McPherson and Cudney, 2014). However, integral membrane protein crystals commonly never reach any satisfactory compromise. As a reminder, the purified IMPs are surrounded by large and unstructured detergent micelles. This reduces in turn: (i) surface accessibility between proteins to form stable contacts, (ii) packing of the crystal, (iii) consistency of the periodicity, (iv) order, (v) diffracting power of the crystal. An alternative method to overcome (i) and (ii), and the ensuing issues, is the lipidic cubic phase crystallisation (LCP) method (Caffrey, 2015). In the LCP method, the protein is also purified in detergent, but then homogenized with monoacylglycerol (MAG) to form a continuous three dimensional cubic phase bilayer. The protein can diffuse freely within the cubic phase, and a precipitant is added to create local disturbances feeding growth of tightly packed protein lattices. Although LCP crystals are very small and difficult to harvest, their characteristic tight packing often results in better diffraction. Another approach to obtain well diffracting crystals of IMPs, is to use conformation specific nanobodies as crystallisation chaperones (Pardon *et al.*, 2014). The latter can increase the stability of the protein, and mediate protein-protein contacts, therefore acting as crystallisation chaperone.

Practically, X-rays can be produced in advanced facilities such as synchrotrons and X-ray free electron lasers (XFELs). Since X-rays have much shorter wavelength than visible light (the practical range for MX is 0.5 to 1.5 Å), individual atoms can be resolved. In fact, when X-rays interact with atoms they make electronic clouds vibrate. These oscillations of charges, emit electromagnetic waves of comparable frequency (i.e., elastic scattering). With periodic structures such as crystals, interference effect occurs between the elastically scattered waves. Under certain experimental settings (i.e., fulfilling Bragg's law), X-rays, elastically scattered by a family of parallel lattice planes (defined with a set of coordinates according to their relation to the primitive unit cell), can interfere constructively, and their intensity can be recorded on a detector, as a diffraction spot, also called a reflection (Bragg, 1956). On the detector area, the recorded spots constitute a diffraction pattern, or diffraction image. However, all families of Bragg planes present in the crystal, cannot be represented in one single two dimensional image. Instead each image corresponds to the section of a sphere, called Ewald sphere, which is an image of the crystal lattice in reciprocal space. Rotating the crystal allows to probe the entirety of the Ewald sphere, and to measure replicates. The recorded diffraction

spots are then assigned to the set of planes they originate from (indexing), and their intensities are integrated (Powell, 2017, 2021). Because the periodicity in the crystal can be estimated after indexing, redundant measurements allow to correct for some systematic errors related to the experiment, and detect outliers. In reciprocal space, the electrons present in the unit cell can be described by a sum of sinusoidal functions, consisting of phases and amplitudes. The amplitudes can be derived easily from the integrated intensities of the diffraction spots but, the phases are not recorded during the MX experiment. A common way to obtain the phase information is to initially approximate it, from a similar protein structure (i.e., molecular replacement) (Ilari and Savino, 2008). This approach benefits greatly from the recent breakthroughs in protein structure prediction tools (Baek *et al.*, 2021; Jumper *et al.*, 2021). The probabilistic distribution of electrons in every point of the unit cell can then be calculated in real space (i.e., electron density map), with Fourier transform. By fitting an atomic model inside the electron density, while respecting fundamental stereochemistry knowledge, the phase information is estimated more accurately, allowing better interpretation of the experimental data (i.e., model refinement).

The whole MX pipeline benefits from more than sixty years of hardware and software developments (Kendrew *et al.*, 1958; Wlodawer, 2021). Nowadays, routine data collection and processing can be done without any user intervention, within minutes for each protein crystal. And despite the difficulties in producing highly diffracting IMP crystals, MX was still the only technique, providing high resolution structural data, on MFS transporters (Choy *et al.*, 2021) - until 2020.

4.8 Single particle analysis cryogenic electron microscopy, on integral membrane proteins

The main advantage of SPA cryo-EM over MX is that it does not rely on crystals, but on a liquid sample. Other attractive aspects, are the lower requirements in protein quantities, and the possibility of studying heterogeneity (i.e., large conformational changes, various quaternary states), within the same sample. Yet, it contains its own pitfalls and limitations.

In SPA cryo-EM, a drop of the purified protein sample is first applied onto a cryo-EM grid. The grid is constituted of a number of squares, each containing multiple holes. When the sample is applied, the protein and the solvent diffuse more or less homogeneously onto the

whole surface of the grid, and the excess is blotted away with filter papers. The purpose of blotting is to obtain a thin layer of sample to facilitate vitrification (i.e., transformation of a liquid into an amorphous solid, or glass). Vitrification is achieved by rapidly dropping the grid in a high heat transfer capacity liquid, typically liquid ethane, cooled down at its melting point (-188 °C), with liquid nitrogen (Dubochet *et al.*, 1988). Vitrification is an essential aspect of SPA cryo-EM, for two main reasons. Firstly, it allows to immobilize the sample. Since, the electron microscope is under vacuum, liquid samples would otherwise evaporate, and proteins need to be solvated to acquire their three dimensional structures. Chemical fixation limits the achievable resolutions to roughly 15 Å, so it is not a desirable alternative. And secondly, vitrification prevents the formation of crystalline water ice (e.g., cubic and hexagonal ice), which would inevitably form, upon slow freezing of the sample. Crystalline features strongly scatter electrons, and therefore interfere with imaging of the specimen. Crystallisation of solvation water molecules can also destroy the protein structure. Typical challenges during grid vitrification are (i) the production of optimum ice thickness across the grid holes, (ii) the particle distribution inside the meniscus (i.e., number of particles, and repartition alongside the ice gradient), (iii) the immobilisation of randomly oriented and intact particles (i.e., protein adopting various angular poses, and not accumulated at the air water interface). With the introduction of semi-automatic blotting and plunging devices, grid vitrification became rather straightforward and reproducible, however, modifications have often to be brought to the sample in the case of preferred orientation, poor particle distribution, or accumulation at the air water interface (Sgro and Costa, 2018; Weissenberger, Henderikx and Peters, 2021). In the case of IMP, the choice of detergent or membrane mimetic, can trigger or address these issues. For instance, certain proteins, when reconstituted in nanodiscs, or particular detergents, will tend to fall into one specific orientation. High concentration of detergents might change the viscosity of the sample or the surface tension of the meniscus, thereby impacting particle distribution or accumulation at the air water interface. Such aspects should be optimised, and can vary from sample to sample.

The main limit of SPA cryo-EM actually resides in the nature of the sample itself, radiation damage, and beam induced motions. There is no alternative to it, but compromises. Biological macromolecules, are highly sensitive to ionizing radiation sources such as the electron beam of an electron microscope (or the X-ray beam of a synchrotron source) (Henderson, 1995). In MX, radiation damage is definitely a big issue and also requires trade-off, but there, again, the

crystal amplifies the signal immensely; so the exposure to X-rays can be reduced to a minimum, while still measuring relatively high signal over noise ratio (SNR). In SPA cryo-EM however, the protein molecules are spread in the vitreous ice, as *single particles* being destroyed and displaced, as few electrons hit the sample. The dose of electrons illuminating the sample, has to be minimal to limit damages, and there is no elegant trick, other than brute force data collection and computational image averaging, to overcome the low SNR (Sigworth, 2016). In the next paragraphs, I will try to explain some of the basic principles in SPA cryo-EM which, help to understand how low dose imaging, impacts high resolution structure determination of small IMPs, such as MFS transporters.

In current electron microscopes used for SPA cryo-EM (i.e., Titan Krios G3 or G4 from ThermoFisher), electrons are generated by a field emission gun and pulled towards an accelerator stack, thereby gaining a particular energy of 300 keV. High vacuum is maintained inside the column to avoid interaction of electrons with gas. The column is constituted of three lens systems, each composed of a set of deflectors, lenses, stigmators, and an aperture. After being accelerated, the electron beam is first focused onto the sample by the condenser lens system. The sample resides within the next lens system: the objective lens system. The objective lens produces a magnified image of the sample, which can be further magnified by the projector lens system, and sent onto a camera (i.e., a direct electron detector). SPA cryo-EM therefore operates in brightfield, transmission electron microscopy (TEM), acquisition mode. As electrons travel through the sample, some will pass between the atoms and therefore, will not interact with the sample (i.e., unscattered = transmitted electrons). Some electrons will interact with atoms and leave some of their energy in the sample, causing radiation damage (i.e., inelastic scattering), and some will interact with atoms, and leave the sample with the same energy (i.e., elastic scattering). Inelastically scattered electrons come out of the sample with different velocities, and cannot be focused into a single point, leading to blurring of the 2D images (i.e., chromatic aberration). Therefore, microscopes are outfitted with energy filters to let only, unscattered, and elastically scattered electrons, reach the camera. For the sake of understanding image formation, unscattered, and elastically scattered electrons can be considered as a plane waves (or two dimensional arrays), moving down the column, passing through the sample, and hitting the detector. In fact, each electron, or wave plane, travels down the column, one at a time, probes the sample in its entirety, and contribute to the image all by itself. The electron contributing to the image keeps the same energy (300 keV) and wavelength

(2 pm), but as wave plane, it *carries* – through the scattering event – all the information (i.e., spatial frequencies, or Fourier components) required to produce a magnified image of the sample on the detector. Whether the different spatial frequency components, contribute to the final image or not, depends mostly, on the phase shift they undergo through the scattering event, and the path length they travel through the microscope, as Fourier components of the two dimensional wave function. Unfortunately, in an electron microscope, some of the frequency components of the object, are not transferred in the image because, the particular phase shift they undergo, does not provide any contrast with the recorded intensity (amplitude) of the unscattered beam. The latter phenomenon can be described by a sinusoidal function, called contrast transfer function (CTF). The CTF is the Fourier transform of the point spread function (PSF), of the objective lens. It relates the contrast present in the image (i.e., power spectrum, or thon ring pattern), with the different aberrations affecting it. Thus, contrast is, as expected, reduced by background noise (e.g., thick ice); but by the CTF as well. And importantly, the CTF is modulated by two main factors: envelope functions and defocus. Envelope functions result from experimental defects, such as spatial and temporal incoherence, which cause image blurring. Increments in defocus value contributes to (i) the total envelope function, and therefore dampening of higher spatial frequencies signal, (ii) decreasing the periodicity of the CTF at higher spatial frequencies, (iii) increasing the signal contribution of lower spatial frequencies, (iv) delocalizing signal in real space, due to the PSF. All together, these CTF effects have several implications on data collection and processing. The first point, is that acquiring multiple images of the object, at various defocus values, allows to recover some of the missing Fourier components in the CTF. The second point is that, a dataset recorded at high ranges of defocus values, will contain particles with strong low resolution features, such as the contour of the protein, or of a micelle. This can be beneficial to certain image processing steps. However, the high resolution signal will be significantly flawed. The third point, is that CTF effects have to be considered to calculate a 3D reconstruction from the images. This is done by first estimating the CTF and then restoring the signal in the image. The estimation can be done by computing a power spectrum of the image, and fitting a calculated CTF. Since an object (i.e., the protein), is the inverse Fourier transform, of the Fourier transform of its two dimensional projections, divided by the CTF, some corrupt signal can also later be recovered, after obtaining an initial 3D reconstruction. However, the decreasing periodicity of the CTF at high spatial frequencies, brought about by the higher defocus range,

makes it harder to generate a good CTF fit, and consequently, to perform any useful image correction.

Going further, data is recorded as movies. Each movie is a set of typically 30 to 60 frames. Typical frame exposures are within the range of 0.8 to 1.5 $e^-/\text{\AA}^2$. If enough signal is present within each frame, the beam induced motions occurring in the illuminated area, can be tracked and corrected, more or less accurately. Because most side chain chemical groups are deteriorated after 30 $e^-/\text{\AA}^2$, the exposures are “dose weighted”, and the high spatial frequency signal of the last frames are usually to be ignored, while low spatial frequency signal can be kept, to help to locate or to grossly align single particles.

Reconstructing a 3D volume from multiple 2D images is possible, at the condition of knowing the shifts and angles, relating each projection (Penczek, 2010). Another obvious condition is that angular sampling should reach a certain amount of completeness. To obtain a 3D reconstruction, the particle images are first extracted from the motion corrected micrographs. The experiment does not provide the relative in-plane transformations and projection angles and of these images (i.e., particle alignment and Euler angle assignment). Different algorithms have been developed to this end. A popular approach is to assign translational and angular probability weights to every 2D projection, in relation to a low resolution guess of the object (i.e., reference). As the probability distribution becomes sharper, the reference is refined, until convergence to a most likely 3D reconstruction (Scheres, 2012). In order to accelerate computation, these calculations are typically done in Fourier space, with global filtering operations applying spatial frequency weights as the refinement progresses. Still, the computing speed is inevitably lowered by large datasets, or large particle extraction box size. As MFS transporters are small size proteins, surrounded by noisy micelles, the number of averaged projections required to obtain sufficient SNR is extensive. If, in addition, high defocus range data acquisition is performed, the extraction box size has to be increased significantly to recover delocalized high resolution signal, and the calculations slow down considerably. One drawback of global filtering approaches, is that it doesn't account for non-uniformly resolved regions within the 3D map, and leads to local overfitting of flexible or disordered regions, and less accurate alignment of the projections. This is particularly relevant to small integral membrane proteins, for which, most of the signal corresponds to a large disordered detergent micelle, or nanodisc, while a smaller region corresponds to the structured

protein. Fortunately, procedures (Non Uniform Refinement (Punjani, Zhang and Fleet, 2020), and SIDESPLITTER (Ramlaul *et al.*, 2020)) have been developed to identify and down weigh high spatial frequencies in regions that are inconsistent between subsets of the data. These approaches reduce overfitting, and greatly improve the quality and interpretability of small IMP 3D reconstructions.

Finally, considering that most MFS transporters display no characteristic cytoplasmic or periplasmic features which are helpful to drive projection alignment, different strategies (i.e., fiducials) have been used to increase the overall size of these proteins and introduce structured features outside the micelle or nanodisc (Pardon *et al.*, 2014; Miyagi *et al.*, 2020; Mukherjee *et al.*, 2020; Nygaard, Kim and Mancina, 2020; Botte *et al.*, 2022). Indeed, almost every MFS transporter structure determined so far by SPA cryo-EM, has benefited from a conformation specific binder, such as a nanobody, or a fab fragment.

4.9 Proton coupled oligopeptide transporters (POTs).

Proton coupled oligopeptide transporters (POTs) are secondary active symporters, which belong to the major facilitator superfamily (MFS). They accumulate dipeptides and tripeptides across biological membranes, using the inward-directed electrochemical proton gradient as energy source. POTs are found in bacterial and eukaryotic organisms, but not in archaea. In humans and other mammalian species, the SLC nomenclature includes 4 paralogues: the peptide transporters 1 and 2 (PepT1, SLC15A1; PepT2, SLC15A2), and the lysosomal peptide and histidine transporters 1 and 2 (PHT1, SLC15A4; PHT2, SLC15A3). In bacteria, the number of paralogous POTs varies among species. The *E. coli* genome codes for four different POTs named di- and tripeptide permease A, B, C, and D (DtpA, DtpB, DtpC, DtpD). Most POTs, like PepT1, PepT2, PHT1, PHT2, DtpA, and DtpB, are extremely promiscuous, in the sense that they can transport a large variety of peptides, and other exogenous molecules. Such POTs have been classified as prototypical, or canonical POTs (Daniel and Adibi, 1993; Ganapathy *et al.*, 1995; Sugawara *et al.*, 2000; Brandsch, Knütter and Leibach, 2004; Luckner and Brandsch, 2005; Weitz *et al.*, 2007; Harder *et al.*, 2008; Brandsch, 2009; Prabhala *et al.*, 2017, 2018). DtpC and DtpD, favour dipeptides with a positively charged residue in the second position, and have been classified as atypical POTs (Ernst *et al.*, 2009; Jensen, Ismat, *et al.*, 2012; Jensen, Simonsen, *et al.*, 2012; Aduri *et al.*, 2015). In this chapter, I will first summarize some of the major steps and findings, in the discovery of peptide absorption and reabsorption

by PepT1 and PepT2 in humans and other mammals. I will then review some the main roles of SLC15 transporters in Human health and disease, focusing mainly on inflammatory disorders. Finally, I will summarize the contribution from structural biology in understanding POTs activity, and the questions that remain open to investigation.

4.10 Peptide absorption and reabsorption by PepT1 and PepT2

For many decades, it was believed that dietary proteins should undergo complete hydrolysis to free amino acids, before being absorbed by enterocytes (*i.e.*, cells lining the lumen of the intestine). In nutritional studies performed in the 1960s and 1970s however (Hellier *et al.*, 1972; Adibi *et al.*, 1975; Matthews, 1975; Adibi, 1976; Mathews and Adibi, 1976; Adibi and Morse, 1977), healthy patients would typically be orally administrated with equimolar quantities of amino acids, either in monomeric form, or as peptides. Blood samples would then be taken to measure the nutrient levels. Surprisingly, amino acid levels were higher in the blood of patients administrated with oligopeptides. These experiments suggested that not only monomers of ingested nutrients could be absorbed, and importantly, they indicated the existence of small peptide carriers (Adibi *et al.*, 1975). A decade later, studies on brush boarder membrane vesicles from the kidney demonstrated the role of proton gradients in this process, so these carriers were named proton coupled oligopeptide transporters (POTs) (Ganapathy and Leibach, 1983, 1985, 1986). In 1994, the first mammalian POT was cloned and therefore named PepT1 (Fei *et al.*, 1994). Its paralogue PepT2, was cloned in the following year (Liu *et al.*, 1995). While PepT2 shares almost seventy percent sequence similarity with PepT1, it differs mainly in expression levels throughout the body, and is was found in the kidney and in the brain (Liu *et al.*, 1995; Berger and Hediger, 1999; Shen *et al.*, 1999, 2004). PepT1 and PepT2 display similar substrates profiles, but the former is referred to as “low affinity – high capacity” while the latter, as “high affinity – low capacity” transporter (Rubio-Aliaga and Daniel, 2008). The primary function of PepT2 in the nephron, is to mediate reabsorption and clearance of di- and tripeptides, from the glomerular filtrate to renal proximal tubule epithelial cells . In the brain, the function of PepT2 was elucidated many years after, and I will come back to it later. In any case, it was clear at that time, that peptide transport, in humans and other mammals, was taking place in brush boarder epithelial cells of different organs, and was coupled to the proton motive force (PMF). Characteristically, epithelia are multicellular organizations that prevent the passage of molecules and ions through the space between cells. Adjacent plasma membranes are quasi imperviously sealed together by tight junctions. Thus,

the solutes must use the various regulated transport systems that we introduced earlier, to penetrate these barriers. The presence of tight junctions has also the function of establishing a polarity in every epithelial cell. The apical side is oriented toward the lumen of the epithelium, while the basal side is oriented towards capillaries of the circulatory system. PepT1 and PepT2 are exported on the apical side together with the sodium proton exchanger NHE3, which locally generate steeper PMF. As of today, no peptide transport protein was detected in the basolateral side of epithelia, but several single amino acid uniporters and antiporters were (Bröer and Fairweather, 2018). Therefore, it is believed that oligopeptides transported by PepT1 and PepT2, are hydrolysed inside the epithelial cells into single amino acids before entering in the circulatory system, and reaching various organs in the body for *de novo* protein synthesis (Daniel and Kottra, 2004). PepT1 is responsible for the great majority of protein digestion products uptake in the digestive system, and is highly upregulated after only 16 hours of fasting (Spanier and Rohm, 2018). Therefore, deletion of its gene was expected to have obvious phenotypic consequences. Surprisingly, PepT1 knocked out mice strains revealed no pathological phenotype, but normal body size, organ weight, and no significant difference in expression levels of any of the other SLC15 genes (Hu *et al.*, 2008). Similar observations were made with PepT2 knockout (Rubio-Aliaga *et al.*, 2003), and PHT1 knockout (Wang *et al.*, 2017) mice. This illustrates a certain biological redundancy of the SLC15 family, with amino acid transporters, in the various tissues and organs they are present. The advantage conferred by this redundancy, is illustrated in patients suffering from Hartnup disorder. The latter is a rare inherited disease, where mutations in the SLC6A19 gene cause the corresponding encoded neutral amino acid transporter, to be inactive, and therefore, abolish the uptake of essential amino acids such as tryptophan: the precursor of serotonin, melatonin, and niacin (vitamin B3) (Bröer, Cavanaugh and Rasko, 2005; Bröer, 2009). If the individual ingests a protein rich diet; PepT1 takes over the role of SLC9A19, and delivers the essential amino acids inside the body (Nässl *et al.*, 2011; Bröer and Fairweather, 2018). If not, however, the individual will suffer from severe symptoms of photosensitive skin rash, cerebellar ataxia, depression, and dementia (Hashmi and Gupta, 2022).

4.11 Role of PepT1 and PepT2 in drug disposition

The response to a given drug is conditioned with its effect on the organism (e.g., interaction and mode of action on a particular receptor), and by its absorption, distribution, metabolism and excretion (i.e. pharmacokinetics). *Per os* (i.e., oral administration) is the most common

route by which drugs are prescribed. Just as nutrients, the majority of drugs require specific uptake by transporters expressed at the brush boarder membrane of enterocytes, to enter the circulatory system. Transporters present in the epithelium of other organs (e.g. kidney, liver, brain), thereafter modulate the distribution of the drug, throughout the body (Zaïr *et al.*, 2008). As a matter of fact, PepT1 and PepT2 not only transport peptides, but also a large variety of clinically relevant, drugs. These include compounds, used in cancer treatments, inflammation, viral infections, bacterial infections, heart diseases, and diabetes (Daniel and Adibi, 1993; Ganapathy *et al.*, 1995; Sugawara *et al.*, 2000; Brodin *et al.*, 2002; Katsura and Inui, 2003; Luckner and Brandsch, 2005; Li *et al.*, 2006; Brandsch, Knütter and Bosse-Doenecke, 2008; Varma *et al.*, 2010; Yang and Smith, 2013; Sun *et al.*, 2018). The potential role of PepT2 in the brain was also clarified, with *in vivo* experiments in PepT2 knocked out mice. For instance, it was shown that PepT2 was responsible for the clearance of neurotoxic drugs from the cerebrospinal fluid (CSF), into choroid plexus epithelial cells (Smith, Johanson and Keep, 2004; Kamal, Keep and Smith, 2010; Chen *et al.*, 2018). In the nephron, PepT2 was also reported to reabsorb exogenous molecules, and increase their systemic levels (Kamal *et al.*, 2009). As for PepT1, targeted proteomics studies on human small intestines showed that it accounted for almost fifty percent of all known clinically relevant drug transporters (Drozdik *et al.*, 2014). As of today, it is widely accepted that PepT1 is the main route of entry in the human body, for orally administrated peptidomimetic drugs. Together with PepT2, they have a major impact on the absorption, and distribution of a wide portfolio of FDA approved compounds.

4.12 Role of SLC15 transporters in disease and inflammation

In addition to their predominant role in amino acid metabolism and drug disposition, transporters of the SLC15 family are involved in innate inflammatory responses (**Figure 9**). And while some of the underlying mechanisms were cross validated by many studies, investigations are actively ongoing to understand the full dynamic interplay between peptide transporters and other signalling macromolecules. The first way SLC15 transporters trigger innate immune response, is by transporting bacterial peptidoglycan components. Indeed, all four SLC15 paralogues can mediate cytosolic accumulation of inflammatory muramyl dipeptides (MDP), and diaminopimelic acid containing tripeptides (tri-DAP). (Swaan *et al.*, 2008; Charrière *et al.*, 2010; Dalmaso *et al.*, 2010, 2010; Nakamura *et al.*, 2014). This uptake occurs within epithelial cells largely expressing PepT1 or PepT2, and, within immune cells,

where PepT1, PepT2, PHT1, and PHT2 are all highly expressed as well (Dalmasso *et al.*, 2010; Ayyadurai *et al.*, 2013; Sun *et al.*, 2013; Wang *et al.*, 2014; Sung *et al.*, 2021). MDP and tri-DAP are pathogen associated molecular patterns (PAMPs). They are agonists of the cytosolic pattern recognition receptor (PRR) family of NOD like receptors (NLRs). Specifically, Tri-DAP is an activating ligand of NOD1 (Laroui *et al.*, 2011), while MDP of NOD2 (Grimes *et al.*, 2012, p. 2). Upon binding, NOD1 and NOD2 recruit and stimulate the receptor interacting protein kinase 2 (RIPK2, or RICK). In turn, RIPK2 regulates the activity of other signalling proteins such as NF- κ B (nuclear factor kappa light chain enhancer of activated B cells). NF- κ B is a transcription factor, which controls the expression of hundreds of genes involved in immune response, growth control, and protection against programmed cell death (e.g., apoptosis). Persistent stimulation from MDP and tri-DAP, lead to a constant expression of such genes, and severe health problems. For instance, many gastrointestinal disorders originate from chronic inflammation i.e., Crohn's disease (CD), ulcerative colitis (UC), and colitis associated colorectal cancers (CAC) (Shah and Itzkowitz, 2022). The incidence of inflammatory bowel diseases (IBD) such as CD and UC, is increasing in western populations, and constitutes a major burden on healthcare systems (Park *et al.*, 2020). It is also well established that chronic inflammation drives neoplastic progression, dysplasia, and cancerisation. Such progression is referred as colitis associated colorectal cancer, and is a leading cause of mortality worldwide (Shah and Itzkowitz, 2022). In all these conditions, abnormal expression of PepT1 in the late segments of the gastrointestinal tract, results in continuous uptake of inflammatory MDP and tri-DAP peptides, subsequent NOD1/2-NF- κ B mediated inflammation and cytokine production (Ingersoll *et al.*, 2012). Recruited innate immune cells further contribute to this process, through endocytosis, with PHT1/PHT2, or through regular uptake from the plasma membrane with PepT1 and PepT2. Interestingly, PepT1 and PHT1 are highly expressed in colorectal cancer cells (Lee *et al.*, 2009; Viennois *et al.*, 2016), and the former was successfully used as a targeted delivery system in mice affected with CAC (Viennois *et al.*, 2016).

The second way, by which SLC15 transporters promote inflammation, involves signalling pathways initiating from another class of PRR; the endosomal toll like receptors (TLRs). Endosomal TLR7, TLR8, and TLR9 recognise PAMPs and DAMPS (damaged associated molecular patterns, released from damaged or dying cells due to trauma or an infection by a pathogen) and activate downstream transcription factors such as NF- κ B, Myc, and IRF (interferon regulatory factors). Myc promotes the transcription of genes involved in cellular proliferation, and is considered as one of the major proto oncogenes. IRFs control the

production of interferon type I cytokines (IFN-1), which are potent pro-inflammatory molecules. Therefore, here as well, systematic exposure to PAMPS or DAMPS leads to a positive feedback loop and chronic inflammation. In various recent studies, PHT1 was showed to be essential for TLR-IRF mediated inflammation, contributing further to IBDs (Sasawatari *et al.*, 2011), but also being critical in autoimmune diseases such as systemic lupus erythematosus (SLE) (Elkon and Briggs, 2020; Heinz *et al.*, 2020; Rimann *et al.*, 2022). There, PHT1 is thought to (i) control the endo-lysosome physicochemical conditions *via* its intrinsic transporter activity, and (ii) serve as a receptor or scaffold protein for downstream activation of specific IRFs. Indeed, the histidine transport activity of PHT1 was shown to be important in controlling IFN-1 and antibody production. In this process, histidine efflux from lysosomes modulated the activity of mTOR, which in turn regulated the IRF7-IFN-I circuit (Kobayashi *et al.*, 2014). In addition, specific interaction between PHT1 and a protein - thereby named “TLR adaptor interacting with SLC15A4 on the lysosome”, or TASL – regulating IRFs (i.e., IRF5) was demonstrated to be critical in SLE disease (Elkon and Briggs, 2020; Heinz *et al.*, 2020). PHT1 is therefore an emerging drug target in inflammatory disorders.

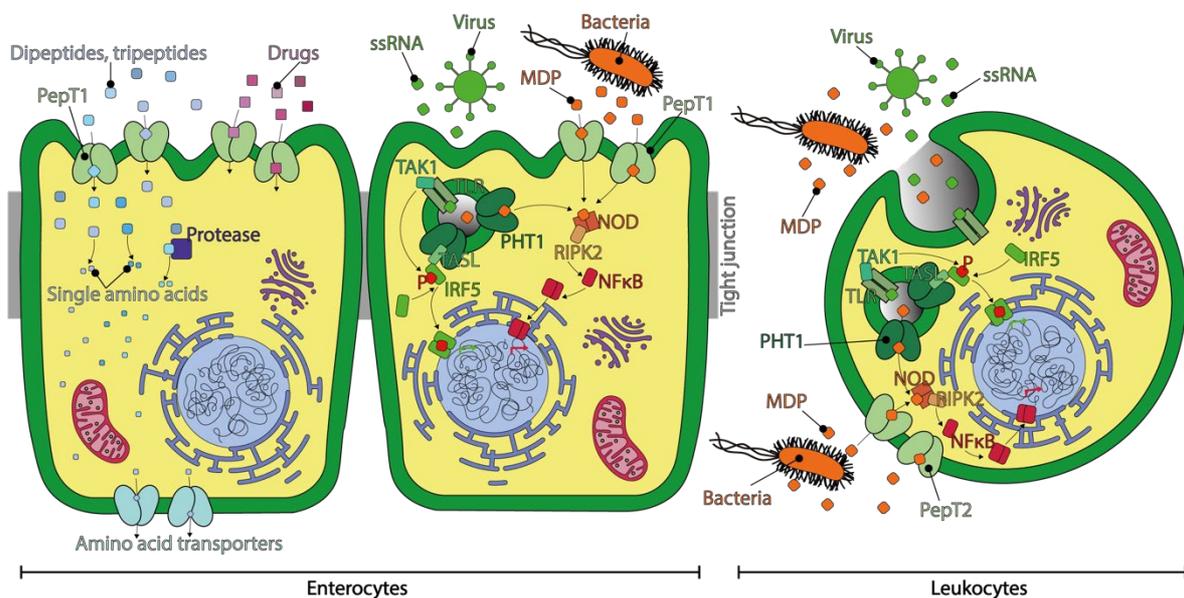


Figure 9. Role of the SLC15 transport gene family in metabolism and inflammation. PepT1 is highly expressed in the intestinal epithelium (enterocytes). There, it uptakes di- and tripeptides from the diet, which are then proteolyzed to single amino acids. These amino acids can be used by the enterocyte, or other cells - reached through the circulatory system - for various metabolic processes such as protein synthesis. As other prototypical POTs, PepT1 can be hijacked by various drugs, and increase their absorption in the body. In the case of an abnormal microbiota, or expression of PepT1 in the late segments of the intestine (i.e., colon),

bacterial peptides, such as MDP, are transported into the enterocytes through PepT1, and activate cytosolic NOD receptors, RIPK enzyme, and NF- κ B transcription factor, promoting expression of pro-inflammatory genes. The same process occurs on leukocytes, but this time involving mainly the paralogue PepT2. Viral infections can also trigger inflammation relying on SLC15A4 (PHT1). There, single stranded viral RNA (ssRNA) is recognized by endosomal/lysosomal Toll like receptors (TLR), and promote the phosphorylation and activation of the transcription factor IRF5, causing the production of type I interferon cytokines. The recruitment of IRF5 to the endosome/lysosome, is necessary for its phosphorylation and activation via TAK1. IRF5 is recruited to the endosome/lysosome after binding to the PHT1-TASL complex. Dysregulations in this pathway have been shown to be critical in rare inflammatory disease such as SLE.

4.13 Previous structural biology data, and open questions on POTs

Despite of the pharmacological and clinical importance of the SLC15 family, the SLC15 transporters had resisted structure determination. The absence of mammalian POT structures was predominantly due to the challenges described earlier: i.e., sample production, crystallisation, and limitations in SPA cryo-EM with such small, flexible, and featureless targets. Instead structural biologists have turned to more stable and crystallisable bacterial POTs, displaying a certain degree of homology with SLC15 transporters. Indeed, between 2011 and 2021, over 50 entries were deposited in the protein data bank (PDB), representing 11 bacterial homologues, bound to 8 unique natural peptides (Ala-Phe, Ala-Glu, Ala-Gln, Ala-Leu, Phe-Ala, Phe-Ala-Gln, Ala-Ala-Ala, alafosfalin) and the peptidomimetic drugs valaciclovir and valganciclovir (Newstead *et al.*, 2011; Solcan *et al.*, 2012; Doki *et al.*, 2013; Guettou *et al.*, 2013; Lyons *et al.*, 2014; Zhao *et al.*, 2014; Quistgaard, Martinez Molledo and Löw, 2017; Martinez Molledo *et al.*, 2018; Minhas *et al.*, 2018; Nagamura *et al.*, 2019; Ural-Blimke *et al.*, 2019; Stauffer *et al.*, 2022). Altogether, these crystal structures provided the basic principles underlying peptide binding in POTs. Peptide binding can be described as an electrostatic hooking mechanism between the constant part of peptides (termini and backbone) and several conserved residues in the transporter (arginine, lysine, glutamate, asparagine and tyrosine residues). In other words, great efforts have been carried out to improve our understanding of how POTs achieve promiscuity. Nonetheless, difficulties inherent to crystallization of IMPs (i.e., sample quantity, reproducibility, diffracting power of the crystals), have limited the portfolio of co-crystallized ligands, to a narrow stereo-chemical space, in comparison with the diversity of substrates, these proteins can transport. From a fundamental point of view, the plasticity occurring within the binding sites of POTs, and the conformational changes associated with unlike substrate recognitions, are therefore still poorly characterized,

and open to further investigation. From the more applied perspective of structure based drug design (SBDD), working with bacterial homologues is of course, not ideal. The majority of POTs, do share, a certain level of promiscuity (i.e., prototypical or canonical POTs), and critical residues for peptide binding, however, many residues involved in drug binding were already shown to differ among bacterial homologues (Minhas and Newstead, 2019; Ural-Blimke *et al.*, 2019). Troubleshooting and overcoming, the challenges associated with structure determination of human SLC15 transporters, would therefore open new possibilities for SBDD investigations.

As mentioned earlier, PHT1 and PHT2 can transport histidine single amino acids in addition to peptides and exogenous compounds, and a subclass of bacterial POTs, referred to as “atypical” or “non-canonical”, favour positively charged dipeptides (i.e., DtpC and DtpD from *E. coli*) (Ernst *et al.*, 2009; Jensen, Ismat, *et al.*, 2012; Jensen, Simonsen, *et al.*, 2012; Prabhala *et al.*, 2014; Aduri *et al.*, 2015). The structural determinants of these unusual ligand predilections are not yet understood.

All the POT crystal structures determined so far represent a particular 3D conformation, named inward facing (IF). In the IF state, the binding site of the transporter is exposed to the cytosolic side of the cell membrane, while the periplasmic side is closed. Again, MFS transporters probably share an overall alternate mechanism of transport, but important variations of the latter were reported in different sub-families, where structure determination of several states was achieved (Drew *et al.*, 2021). PepT1 and PepT2 possess an extracellular domain (ECD) between TM9 and TM10. This domain was isolated and crystalized in 2015 (Beale *et al.*, 2015). It displayed two immunoglobulin lobes, and further biophysical experiments suggested that it could interact with the intestinal trypsin protease, suggesting a role in clustering proteolytic activity to the site of peptide transport. The role of the ECD in transport activity is still not clear however. Therefore, the conformational transitions occurring during transport in POTs, were speculated, but not experimentally determined.

Finally, the recently reported interaction between PHT1 and the signalling protein TASL, constitutes an emerging drug target, for autoimmune diseases such as SLE (Heinz *et al.*, 2020). In addition, the regulatory role of PHT1 in controlling lysosomal inflammation, *via* its transport activity (Kobayashi *et al.*, 2014) encourages SBDD efforts to develop potent PHT1 inhibitors.

Here as well, establishing workflows allowing structural investigation of PHT1 seems elementary.

5 Aim of the work

The aim of this work is to provide new insights into the working principles of proton coupled oligopeptide transporters, including their transport cycle, substrate recognition mechanisms, and their possible interaction with signalling proteins. An additional aim is to provide a methodological basis supporting future structural studies on this transporter family. To these ends, the structures of human PepT1 (SLC15A1), human PepT2 (SLC15A2), PHT1 (SLC15A4), DtpB, and DtpC, were determined in various biochemical conditions, and analysed.

6 Results

6.1 Promiscuity in prototypical POTs

6.1.1 Structure determination of DtpB in complex with 14 peptides.

The diversity of substrates that prototypical POTs are able to recognise and transport, makes them some of the most promiscuous transporters to exist. The first structural rearrangements in POTs, in response to different substrates, were characterized with the POT PepT_{St} from *Streptococcus thermophilus* (Martinez Molledo *et al.*, 2018; Martinez Molledo, Quistgaard and Löw, 2018). There, residues Y68 and W427 (equivalent to Y64 and W420 in DtpB) were shown to modulate the architecture of the binding site. Notably, the movements of Y68 and W427 fine-tuned the pockets accommodating the residues in second (P2) and third (P3) positions in the peptide. The work described in this section, aims to continue exploring the promiscuous binding site of prototypical POTs. Previously our group, Dr. Jian Lei, grew well diffracting crystals of the prototypical di- and tripeptide permease B (DtpB), using the vapor diffusion method, and a conformation specific nanobody (named nanobody 132, or Nb132) as crystallisation chaperone (**Figure 10**)

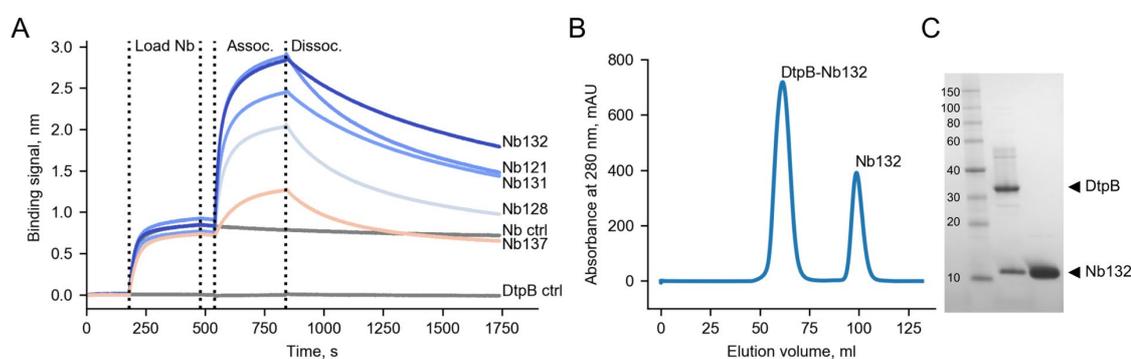


Figure 10. Characterization and complex formation of selected nanobodies with DtpB. (A) Bilayer interferometry was used to assess the binding of several alpaca nanobodies raised against DtpB. The two control conditions, where only a nanobody (Nb ctrl), or only DtpB (DtpB ctrl) were loaded on the sensor, are shown in grey. Vertical dashed lines indicate the steps of the assay: (i) loading of the nanobody, (ii) association of the complex, i.e., addition of DtpB, (iii) dissociation of the complex. (B) Representative steric exclusion chromatogram (SEC) of DtpB-Nb132 complex. (C) Representative sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of DtpB-Nb132 complex.

Nb132 clearly mediated protein-protein contacts within the crystal, and was crucial to obtain high resolution data. Dr Lei, determined the structures of the DtpB-Nb132 complex, bound to the dipeptide alanine-glutamine (Ala-Gln), and the tripeptide alanine-leucine-alanine (Ala-Leu-Ala), which both adopted the most common pose, so far reported. Following this work, we then reproduced the crystals successfully (**Figure 11**), and started a systematic co-crystallization campaign on an in-house library of 82 different di and tripeptides. This was done in collaboration with the group of Dr. Josan Marquez at EMBL Grenoble, and with co-workers in the group of Dr. Christian Loew, including Giada Finocchio, and myself.

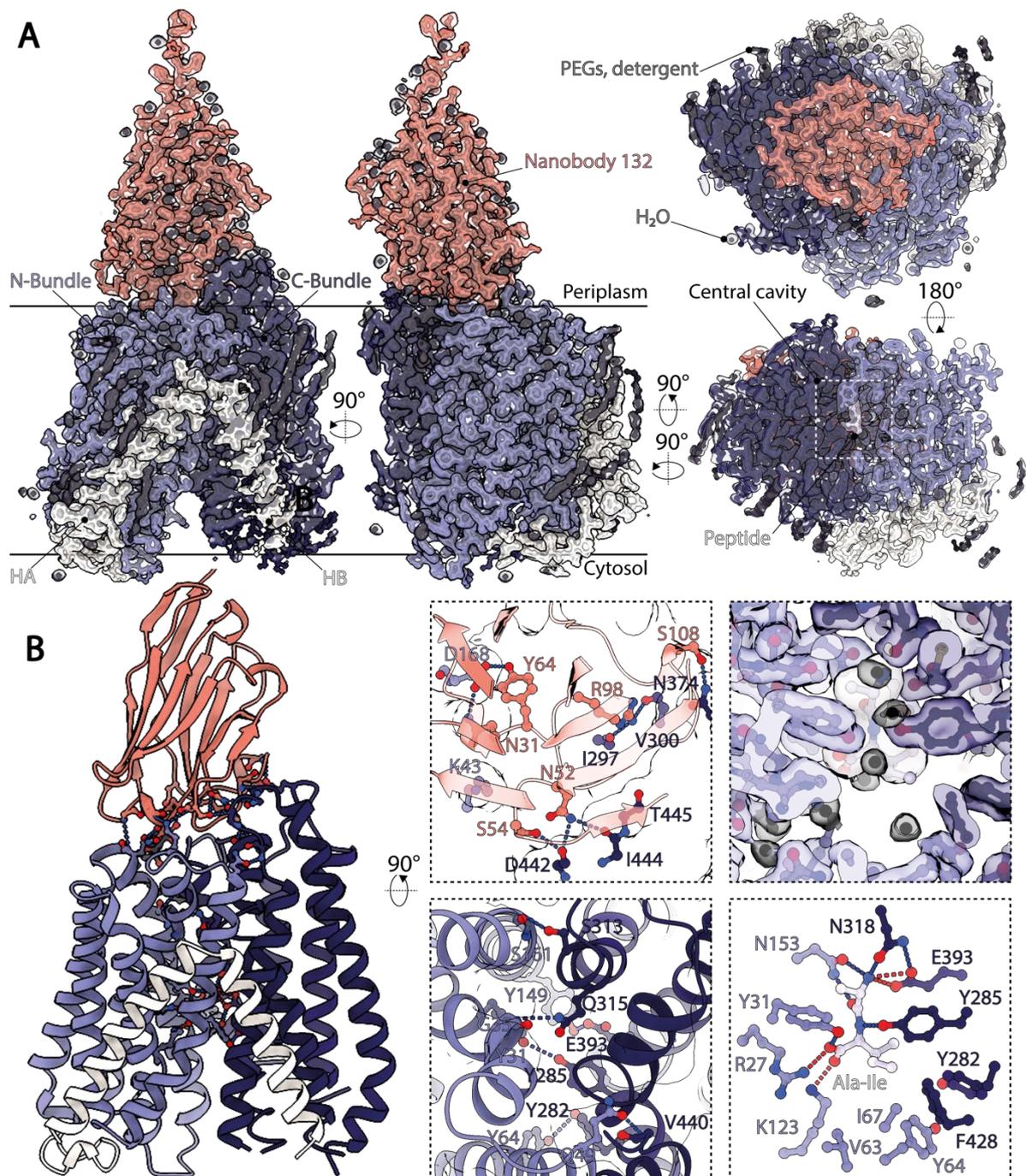


Figure 11. Structure of DtpB bound to Nb132, and the dipeptide Ala-Ile (A) The atomic model of DtpB-Nb132 fitted from the highest resolution dataset (Ala-Ile bound) is displayed as ball and sticks. The 2Fo-Fc map is shown as transparent surface (with $\sigma=1$). The different structural elements are labelled. (B) Residues stabilizing the observed conformation are displayed as ball and sticks and the secondary structural elements are shown as ribbons. Interactions between the transporter and the nanobody are shown in the top left close up view. The density of the binding site and the dipeptide are shown in the top right close up view, and the electrostatic interactions between the peptide and DtpB are show as dashes (red dashed lines=salt bridges; blue dashed lines=polar interactions).

The aim of this collaborative effort was to obtain a much broader vision of the plasticity of POTs binding site, in response to peptides possessing various chemical groups. Several batches of the DtpB-Nb132 complex were prepared, divided in aliquots, incubated with a given peptide, and dispensed robotically in 96 well screens containing three sets of crystallisation conditions. When the diffraction data was higher than 4 Å, the chemical screens were further refined around the best conditions. In this campaign, we measured X-ray diffraction of over 1500 crystals. For each dataset, diffracting higher than 3.5Å, the presence or absence of ligand was initially assessed by using a DtpB atomic model devoid of substrate, to calculate a difference electron density map (i.e., $F_{\text{obs}}-F_{\text{calc}}$) between the calculated structure factors amplitude, and the “observed” (i.e., experimental) ones. There, strong positive peaks within the central cavity indicated the presence of a missing density. The co-crystallized peptide was then modelled inside the positive peak, as a di- or trialanine moiety first, and then mutated to its original sequence, as the signal improved during the refinement steps. As a final validation, OMIT maps excluding the modelled ligands, were calculated for each structure (**Figure 12**).

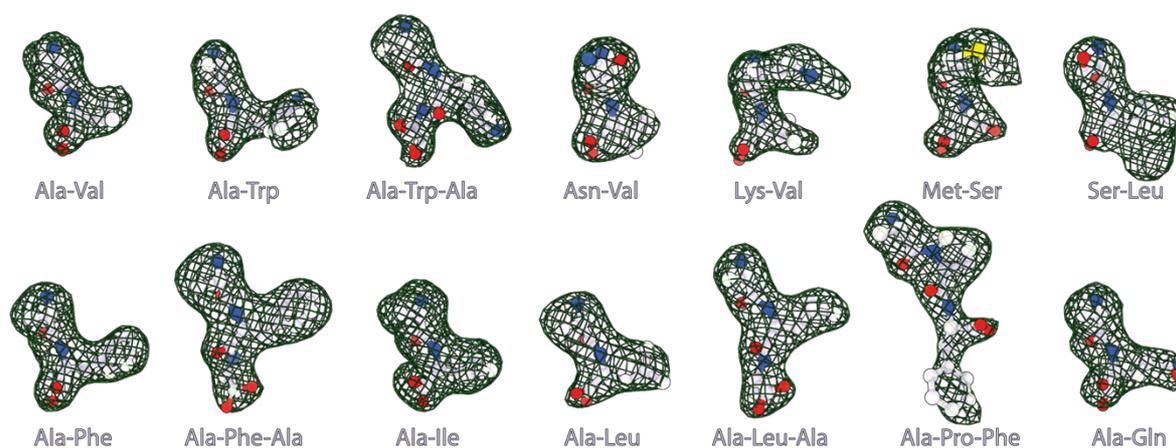


Figure 12. OMIT maps excluding the 14 modelled peptides. The positive peaks (shown as green mesh) indicate missing density, verifying the presence of the peptides.

Table 1. MX data collection and refinement statistics of DtpB-Nb132-peptide datasets.
 Statistics for the highest-resolution shell are shown in parentheses.

Ligand	AF	AFA	AI	AL	ALA	APF	AQ	AV	AW	AWA	KV	MS	NV	SL
Resolution range	49.9 - 2.3 (2.38 - 2.3)	70.3 - 2.45 (2.538 - 2.45)	51.69 - 2.15 (2.227 - 2.15)	51.95 - 2.8 (2.9 - 2.8)	51.79 - 2.56 (2.651 - 2.56)	62.58 - 2.3 (2.382 - 2.3)	58.74 - 2.47 (2.558 - 2.47)	84.47 - 2.35 (2.434 - 2.35)	59.16 - 2.5 (2.589 - 2.5)	58.76 - 2.5 (2.589 - 2.5)	69.95 - 2.6 (2.693 - 2.6)	50.95 - 2.55 (2.641 - 2.55)	54.14 - 2.8 (2.9 - 2.8)	51.16 - 2.67 (2.765 - 2.67)
Space group	P 2 2 1 21	P 2 2 1 21												
Unit cell	54.415 125.074 168.713 90.90.90	54.696 125.509 169.715 90.90.90	54.324 125.742 168.002 90.90.90	54.591 125.836 169.031 90.90.90	54.396 123.541 169.331 90.90.90	54.759 125.156 169.900 90.90.90	54.467 125.385 168.177 90.90.90	54.648 126.166 168.935 90.90.90	54.614 126.408 168.039 90.90.90	54.598 125.359 168.823 90.90.90	54.342 124.675 167.463 90.90.90	54.183 124.675 165.773 90.90.90	54.138 124.925 165.773 90.90.90	54.489 125.66 168.044 90.90.90
Total reflections	667300 (64221)	578325 (5600)	810303 (71235)	378285 (38108)	75320 (7396)	105697 (10434)	83033 (8176)	649577 (45053)	81675 (7972)	80967 (8074)	72749 (7106)	75797 (7414)	57042 (5622)	65106 (5161)
Unique reflections	52149 (5144)	43863 (4296)	63584 (6245)	29527 (2894)	37673 (3699)	52854 (5217)	42145 (4129)	49598 (4863)	40878 (4000)	40497 (4037)	36376 (3553)	37899 (3707)	28523 (2811)	32661 (2628)
Completeness (%)	99.93 (99.94)	99.90 (99.95)	99.94 (99.94)	99.88 (100.00)	99.89 (99.78)	99.93 (99.96)	99.71 (99.61)	99.88 (99.94)	99.15 (98.62)	98.65 (100.00)	99.92 (100.00)	99.92 (99.95)	99.77 (99.93)	96.74 (77.81)
Mean I/sigma(I)	16.79 (1.47)	16.45 (1.40)	18.87 (1.48)	16.30 (1.57)	16.74 (1.89)	13.91 (1.24)	14.56 (1.82)	15.18 (0.78)	12.36 (1.05)	14.29 (1.43)	13.06 (1.04)	9.44 (1.30)	12.01 (0.88)	12.66 (1.19)
Wilson B-factor	53.52	71.01	50.50	80.22	66.52	57.56	61.28	59.41	69.68	69.65	71.69	54.22	77.16	80.58
R merge	0.08998 (1.649)	0.09384 (2.31)	0.07757 (1.664)	0.1373 (2.112)	0.02248 (0.4068)	0.02381 (0.5902)	0.02899 (0.4735)	0.1113 (3.018)	0.01985 (0.6883)	0.02077 (0.4964)	0.02634 (0.6764)	0.04223 (0.5812)	0.03269 (0.8924)	0.02147 (0.675)
R meas	0.09399 (1.719)	0.09789 (2.405)	0.08102 (1.742)	0.1435 (2.198)	0.03179 (0.5752)	0.03368 (0.8347)	0.041 (0.6697)	0.1161 (3.199)	0.02807 (0.9734)	0.02937 (0.702)	0.03725 (0.9565)	0.05973 (0.822)	0.04623 (1.262)	0.03037 (0.9545)
R p _{rim}	0.02673 (0.4834)	0.02738 (0.6643)	0.02302 (0.511)	0.04087 (0.6025)	0.02248 (0.4068)	0.02381 (0.5902)	0.02899 (0.4735)	0.03238 (1.042)	0.01985 (0.6883)	0.02077 (0.4964)	0.02634 (0.6764)	0.04223 (0.5812)	0.03269 (0.8924)	0.02147 (0.675)
CC1/2	0.999 (0.648)	0.999 (0.522)	0.999 (0.591)	0.997 (0.531)	0.999 (0.694)	1 (0.609)	0.998 (0.64)	0.991 (0.285)	1 (0.432)	1 (0.643)	1 (0.488)	1 (0.545)	1 (0.367)	1 (0.35)
CC*	1 (0.887)	1 (0.825)	1 (0.862)	0.999 (0.833)	1 (0.905)	1 (0.87)	1 (0.883)	0.998 (0.666)	1 (0.777)	1 (0.885)	1 (0.81)	1 (0.84)	1 (0.733)	1 (0.72)
Reflections used in refinement	52127 (5142)	43856 (4293)	63564 (6244)	29508 (2894)	37667 (3699)	52830 (5216)	42133 (4122)	49579 (4862)	40846 (3993)	40479 (4037)	36360 (3553)	37882 (3703)	28489 (2809)	32590 (2577)
Reflections used for R-free	2589 (251)	2223 (236)	3113 (316)	2999 (306)	1892 (188)	2648 (249)	2078 (213)	2514 (252)	2015 (191)	2000 (193)	1807 (180)	1838 (200)	1382 (128)	1621 (122)
R work	0.2267 (0.3774)	0.2241 (0.3115)	0.2154 (0.3490)	0.2254 (0.3354)	0.2151 (0.2944)	0.2299 (0.3515)	0.2221 (0.3108)	0.2164 (0.2998)	0.2245 (0.3442)	0.2314 (0.3054)	0.2257 (0.3659)	0.2241 (0.3319)	0.2260 (0.3869)	0.2211 (0.3589)
R free	0.2536 (0.4101)	0.2673 (0.3354)	0.2390 (0.3854)	0.2633 (0.3708)	0.2410 (0.3293)	0.2575 (0.3866)	0.2669 (0.3498)	0.2363 (0.3244)	0.2481 (0.3548)	0.2647 (0.3278)	0.2602 (0.3766)	0.2574 (0.3686)	0.2503 (0.4129)	0.2636 (0.3964)
Number of non-hydrogen atoms	4718	4793	4787	4669	4781	4792	4786	4785	4786	4791	4787	4784	4789	4787
Protein residues	580	581	580	580	581	581	580	580	580	581	580	580	580	580
RMS(bonds)	0.011	0.011	0.010	0.012	0.012	0.010	0.011	0.010	0.010	0.012	0.012	0.011	0.013	0.010
RMS(angles)	1.32	1.30	1.08	1.44	1.45	1.32	1.32	1.10	1.32	1.32	1.37	1.30	1.56	1.19
Ramachandran favored (%)	98.42	97.72	98.06	96.83	98.42	98.42	98.42	98.06	98.06	97.54	97.89	97.18	97.89	97.36
Ramachandran allowed (%)	1.41	2.11	1.94	3.17	1.58	1.58	1.58	1.94	1.94	2.46	2.11	2.82	1.94	2.64
Ramachandran outliers (%)	0.18	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.00
Rotamer	0.42	1.06	1.06	1.91	0.64	0.64	0.64	1.06	1.70	0.85	0.64	0.85	0.64	1.06

In summary we obtained 14 unique peptide bound datasets in a resolution range between 2.0 and 2.8 Å with the following peptides: Ala-Phe ('AF'), Ala-Ile ('AI'), Ala-Leu ('AL'), Ala-Gln ('AQ'), Ala-Val ('AV'), Ala-Trp ('AW'), Lys-Val ('KV'), Met-Ser ('MS'), Asn-Val ('NV'), Ser-Leu ('SL'), Ala-Leu-Ala ('ALA'), Ala-Phe-Ala ('AFA'), Ala-Pro-Phe ('APF'), Ala-Trp-Ala ('AWA'). This represented a large portfolio of ligands, compared to the present literature data. With these structures, local rearrangements in the binding site of DtpB, can be analysed.

6.1.2 Plasticity of the binding site of DtpB

All DtpB complexes crystallized in the same space group and unit cell dimensions (**Table 1**). A superposition of the structures clearly shows that the N-termini of all dipeptides and tripeptides are anchored within a precise and invariable position. More precisely, the primary amine is steadily hooked by N153, N318, and E393 (**Figure 13**).

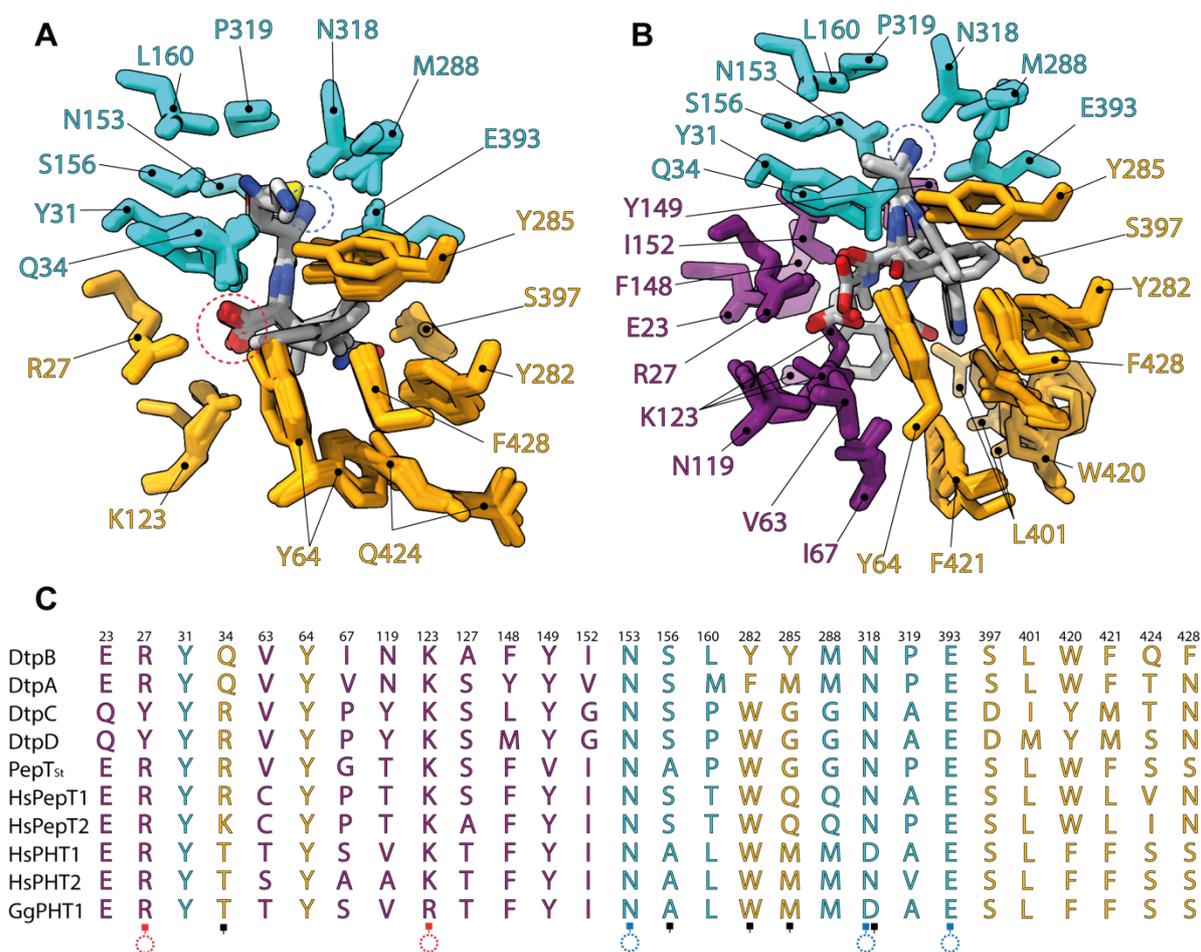


Figure 13. Definition of the binding pocket of DtpB. (A) Superimposition of the dipeptide co-crystal structures. (B) Superimposition of the tripeptide co-crystal structures. (C) Sequence alignment of the residues constituting the P1, P2, and P3 pockets in relevant proton coupled

oligopeptide transporters mentioned in this thesis. Residues constituting the P1, P2, and P3 pockets are respectively coloured in cyan, yellow, and purple. Blue and red dashed rings circle respectively the N-termini, and the C-termini. Blue and red squares indicate residues mediating electrostatic interactions with the termini of the co-crystallised peptides. Black squares indicate residues mediating polar interactions with the side chains of co-crystallised peptides. Note that the N-termini are all coordinated in the same manner, while the C-termini adopt different positions in tripeptides.

This triad of residues remains fixed in all structures, and is conserved in all prototypical POTs, with the exception of PHT1, for which N318 is an aspartate (**Figure 13**). Other residues of the binding site form a pocket around the residues in the N-terminal position of peptides (**Figure 14**). These include Y31, Q34, S156, S156, L160, M288, and P319. Together with the conserved N153, N318, E393 triad, they form the P1 pocket. The chemical diversity of N-terminal residues in the co-crystallized tripeptide data sets was poor (only alanine residues), but richer for dipeptides (i.e., alanine, lysine, serine, methionine, asparagine).

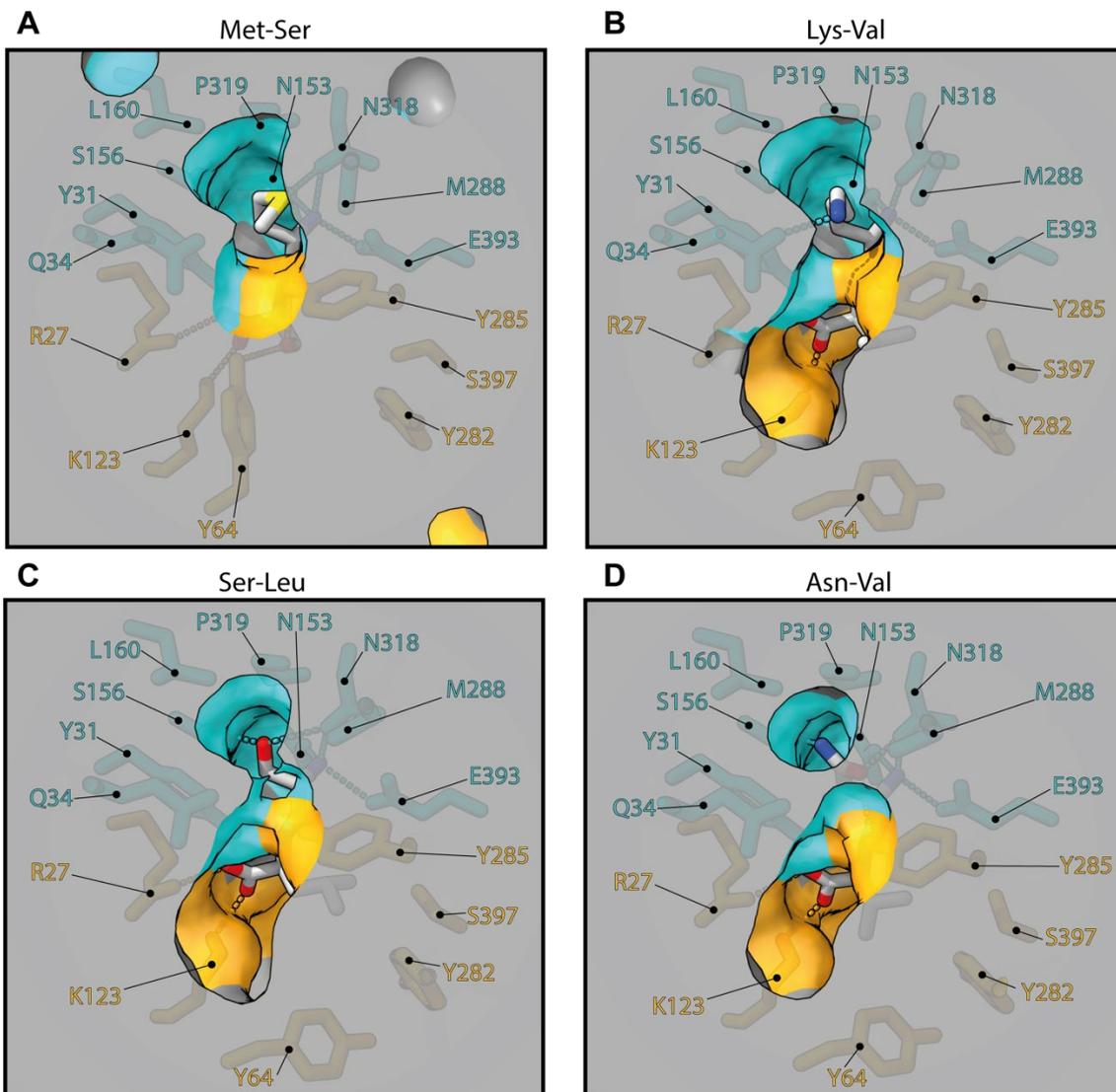


Figure 14. Fitting of side chains in the P1 pocket. (A) Coordination of the methionine side chain of MS peptide. (B) Coordination of the lysine side chain of the KV peptide. (C) Coordination of the serine side chain of the SL peptide. (D) Coordination of the asparagine side chain of the NV peptide. Note that P1 remains tight and stable in all these structures.

P1 does not undergo conformational changes in presence of these various residues, however, polar interactions stabilise certain substrates (**Figure 14**, **Figure 15**). Notably, Q34 interacts with the the ϵ -amino group of K1* in Lys-Val; N318 with the thioester group of M1* in Met-Ser; S156 and N318 with the hydroxymethyl group of S1* in Ser-Leu; and S156, Y285 and N318 with the carboxamide group of N1* in Asn-Val. The C-terminus of dipeptides adopted a constant position, stabilised by R27 and occasionally by K123 as well.

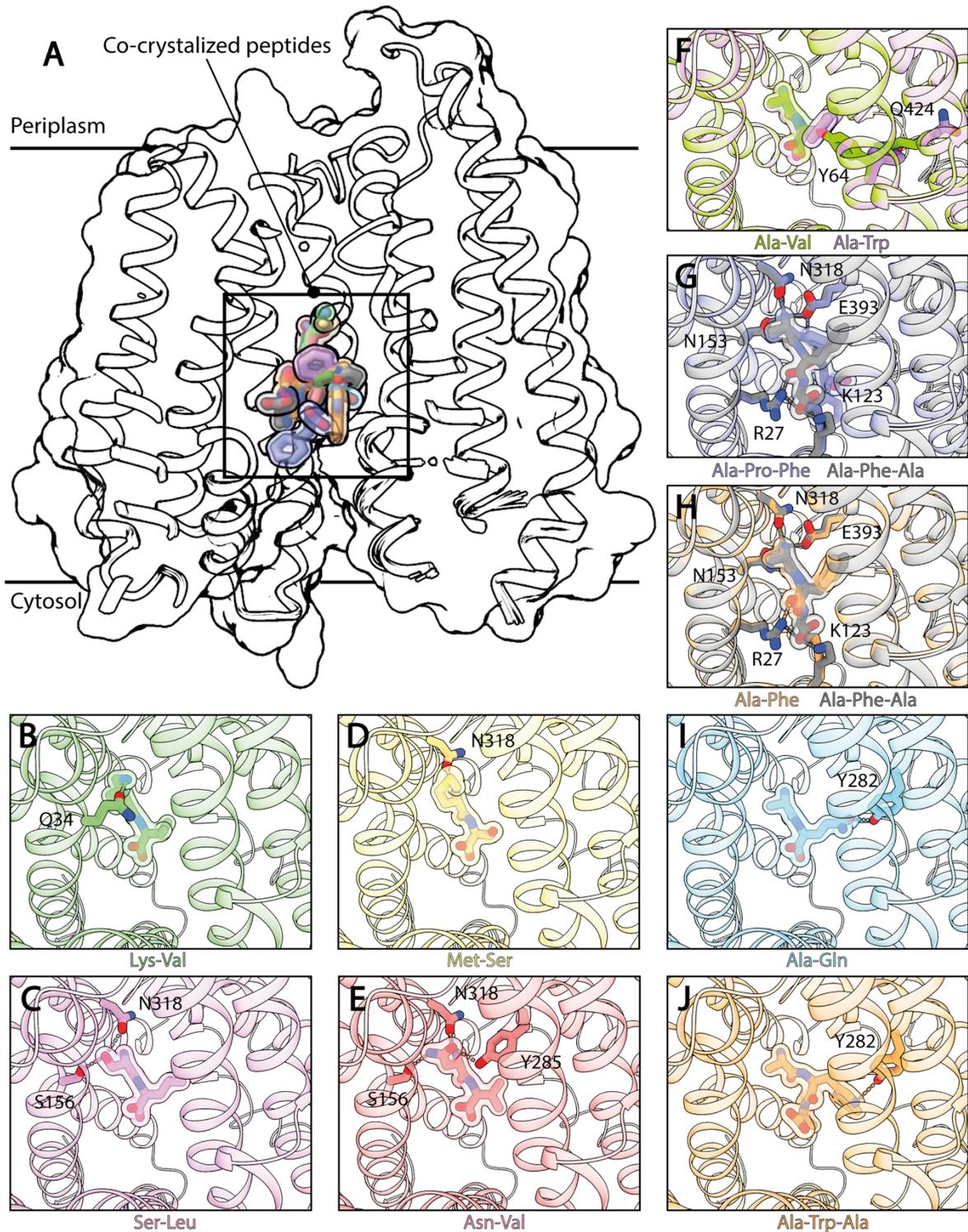


Figure 15. Structural determinants of multi-substrate recognition in DtpB (A) The 14 co-crystallized peptides occupy various poses within the binding site. The peptides were coloured differently and the secondary structure elements of the transporter are displayed, as well as the solvent excluded surface. (B, C, D, E) Polar interactions between the side chain of the indicated co-crystallised peptide, and the P1 pocket. (F) Y64 and Q424 rearranges in the presence of peptides with side chains of different bulkiness in the second position (e.g. Ala-Val vs Ala-Trp). (G, H) The position of the C-terminus can vary (e.g. in Ala-Pro-Phe vs Ala-Phe-Ala; or

Ala-Phe vs Ala-Phe-Ala), and is accompanied by movements of K123 while the N-terminus remains tightly anchored by the N153, N318, E393 triad. (I, J) Polar interactions between the side chain of the indicated co-crystallised peptide, and the P2 pocket.

To adapt the central cavity to the different sizes of side chains carried by the C-terminal residues of dipeptides, two rotamer conformations of Y64 and Q424 are possible (**Figure 16**), regulating the volume of the “upper region” of the P2 pocket. For instance, with the small dipeptide Ala-Val, the upper region of P2 is tightened by the conformation of Y64 and Q424, while it is widened in the presence of the bulky dipeptide Ala-Trp. Other residues such as Y285 and F428 fine-tune the upper region of P2. It should also be noted, that Y285 and Y64 delimits the P2 pocket from the P1 and P3 pockets respectively.

Unlike other side chains fitting in P2, the indole moiety of W2*, in the tripeptide Ala-Trp-Ala, extends further down, towards the cytosolic side of the C-bundle (**Figure 16**). This “lower region” of P2: i.e., L401, W420, and F421, is flexible, and closes up to stabilise W2*. Polar interactions also occur between P2 and the substrates. Y282 stabilises the indole ring of W2* in Ala-Trp-Ala, as well as the carboxamide group of Q2* Ala-Gln (**Figure 15 I, J**). In addition, the hydrophobic side chains in the second position (i.e., in the peptides Ala-Val, Ala-Leu, Ala-Ile, Ala-Phe, Ala-Trp, Ala-Leu-Ala, Ala-Phe-Ala, Ala-Trp-Ala) are increasingly stabilised as a function of their size, and through contraction of the “upper region” of P2, described earlier.

The backbone of tripeptides withstands different torsion angles around amide bounds. Since the primary amine of the N-terminus remains hooked in place between N153, N318, and E393, the carboxylic group of the C-terminus is subsequently shifted or rotated, resulting in different poses (**Figure 13 B, Figure 17**). For instance, the carboxylic group of Ala-Trp-Ala, coincides with the ones of dipeptides. This co-localization is achieved by kinking the backbone of the tripeptide. Ala-Phe-Ala, and Ala-Leu-Ala, are not kinked, but stretched. There, in comparison with the dipeptides Ala-Leu and Ala-Phe, a conformational change in K123 allows to create enough space for the carboxylic group of L3* and A3*, to fit in a position preserving the stabilizing salt bridges with R27 and K123 (**Figure 15 H, , Figure 17**). Finally, in the case of the tripeptide Ala-Pro-Phe, the proline residue restrains the backbone and evicts the C-terminus from the positively charged patch formed by R27 and K123, and the bulky phenyl group of F3* extends towards the cytosolic side of the N-bundle, in P3 (**Figure 13 B, Figure 17**).

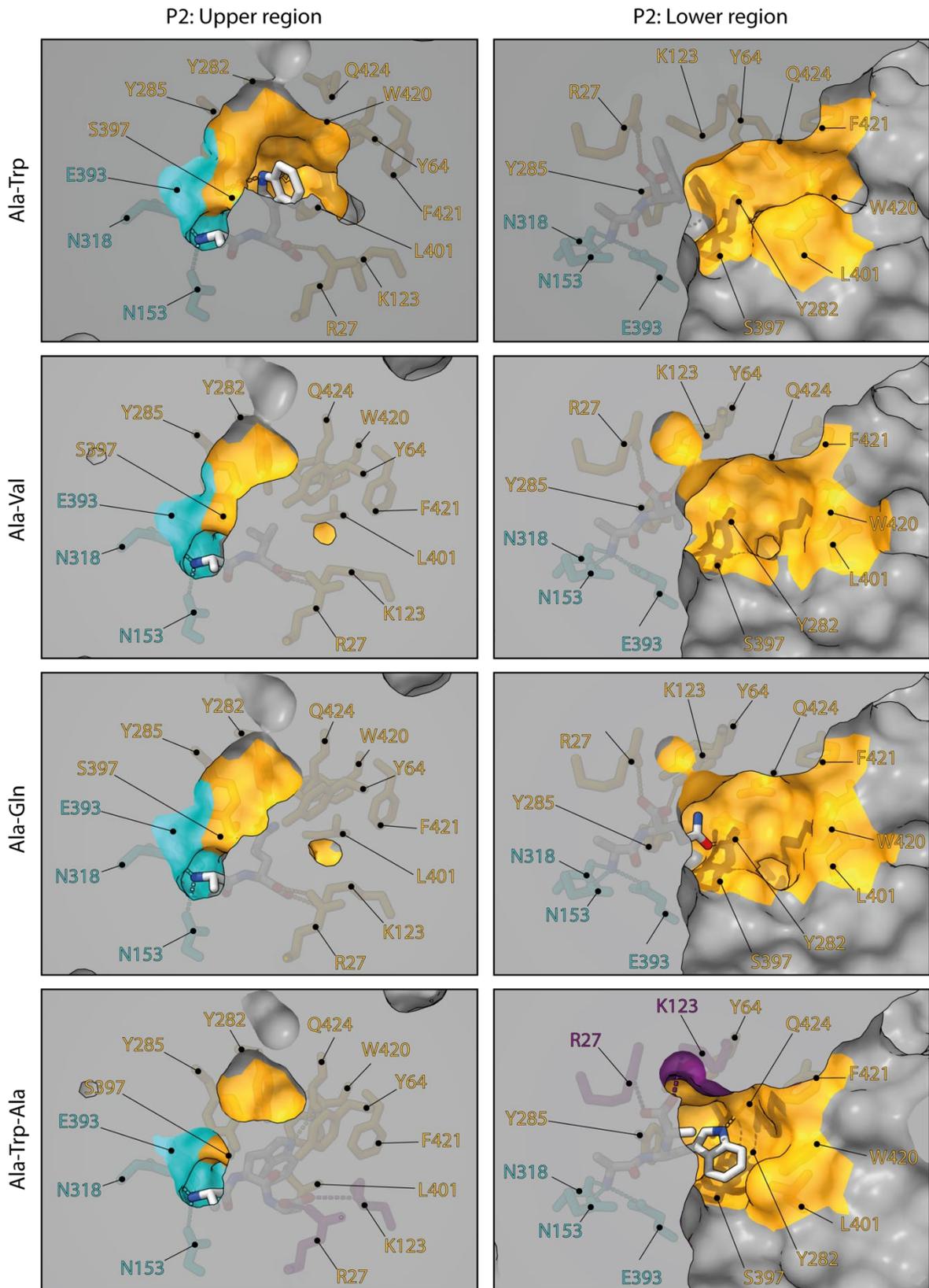


Figure 16. Versatility of the P2 pocket of DtpB. The upper region of P2 undergoes large rearrangements in the presence of various peptides, by switching the rotamer conformations of

Y64 and Q424 (left panels). The indole ring of W2*, in Ala-Trp-Ala, sits in the lower region of P2, where L401, and W420 can rearrange to cap the cavity.

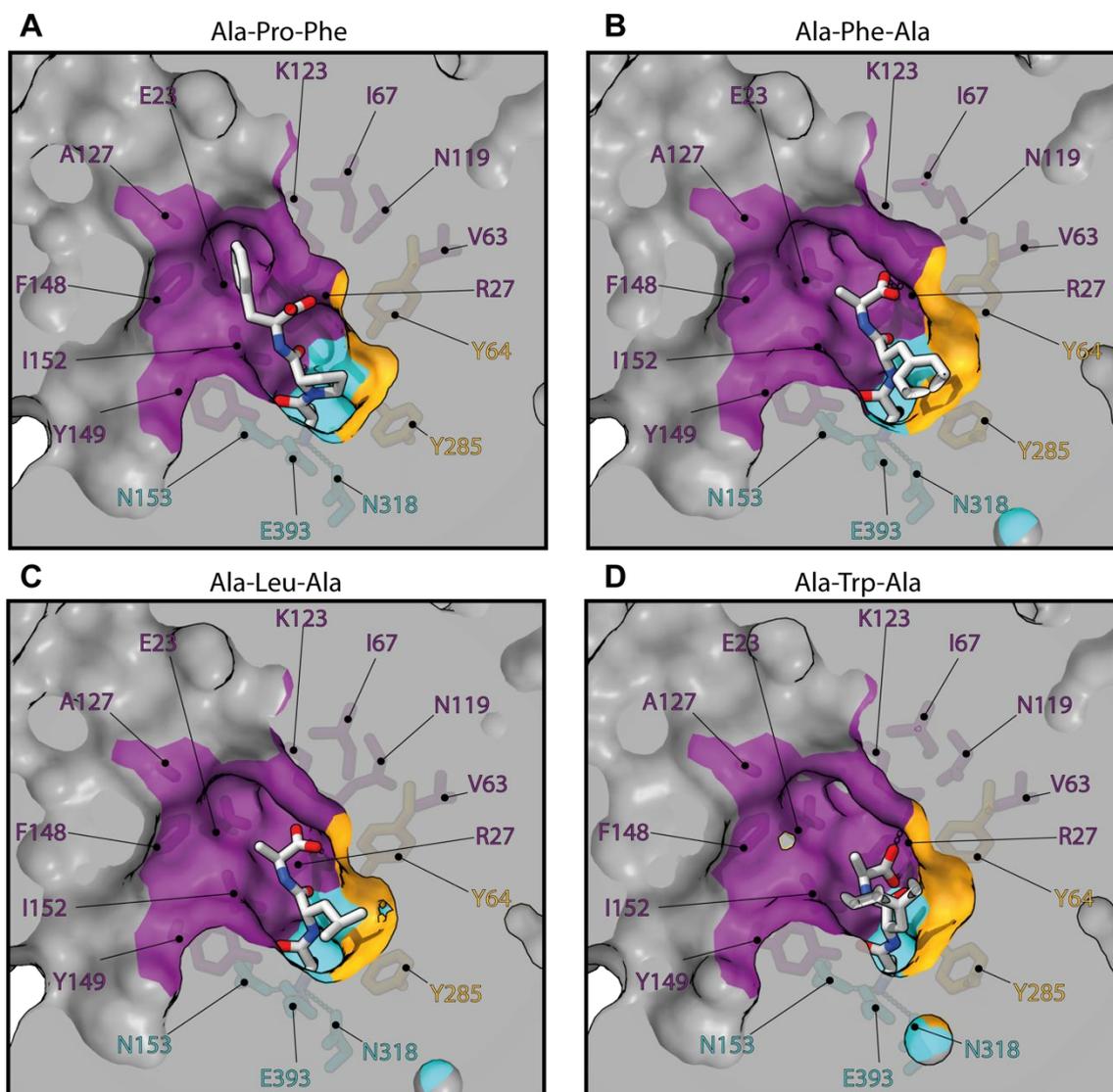


Figure 17. Substrate's C-termini, and side chains fitting in the P3 pocket. (A) APF tripeptide, (B) AFA tripeptide, (C) ALA tripeptide, (D) AWA tripeptide. Note that the backbones of the tripeptides can be more extended (A) or kinked (D), and that the C-termini adopt various positions as a result.

6.1.3 Discussion

In agreement with the literature, these results suggest that the N153, N318, E393 triad is a common anchor point of peptides N-termini. We find that the C-termini of peptides is often stabilised by R27 and K123, but the latter is not mandatory, contrary to what was previously thought (Newstead, 2017). Recent molecular dynamics studies (Parker *et al.*, 2021) on PepT2 and the tripeptide Ala-Ala-Ala, support this observation further, and suggest that peptides engage with the binding pocket via Ala1*, before being tightly locked in place by the triad

(N192, N348, E622 in HsPepT2). The simulation also suggests, that R27 and K64 later contributes to further stabilisation of the C-terminus, but the latter was not essential in DtpB, for the tripeptide Ala-Pro-Phe, and K64 is not conserved among POT homologues (Q34 in DtpB). It is likely that the presence of bulky side chains in the N-terminal position, would cause local changes in P1, but we failed to obtain such structures. The versatility of P2 was previously described with rearrangements of Y68 and W427 in PepT_{St} (Martinez Molledo *et al.*, 2018; Martinez Molledo, Quistgaard and Löw, 2018) (Y64 and W420 in DtpB). Here, these two residues contribute to adapt P2 to the various co-crystallized peptides, but other residues (Y282, Y285, L401, Q424, F428) are also involved. Except from the different rotamer conformations of K123 to adapt to the various positions of tripeptides, the P3 pocket was rather stable compared to P2. Although these observations indicate that the plasticity of POTs lies mostly in the P2 pocket, many more combinations of peptides are able to bind to DtpB and other prototypical POTs, and our findings remain limited to a small subset. Besides, not all residues involved in ligand promiscuity are conserved in POTs. This could explain the differences of substrate affinity reported in the literature, with different homologues. Finally, like other MFS transporters (Drew *et al.*, 2021), POTs must have to undergo large conformational changes to achieve alternate access of the binding site. During this process, the currently observed architectures of P1, P2, and P3 might be dramatically changed. Therefore, high resolution structures on the human homologues, particularly, in outward facing (OF) conformations, would be more useful for SBDD aiming at increasing oral bioavailability by hijacking the intestinal PepT1 shuttle system.

6.2 Transport in the SLC15 family

6.2.1 Crystallisation trials on HsPepT2

Former members in our lab successfully cloned and expressed the human gene of PepT2 (HsPepT2) in HEK293F (human embryonic kidney 293F) cells (Pieprzyk, Pazicky and Löw, 2018a). Briefly, HsPepT2 was fused to a EGFP (enhanced green fluorescent protein) reporter and a hexa-histidine (His) tag. The His-EGFP element was introduced to the N-terminal side of the transporters, upstream a TEV (tobacco etch virus) protease recognition sequence. Uptake assays in whole cells confirmed that recombinant HsPepT2 was able to transport various dipeptides, tripeptides, and drug molecules (**Figure 18**).

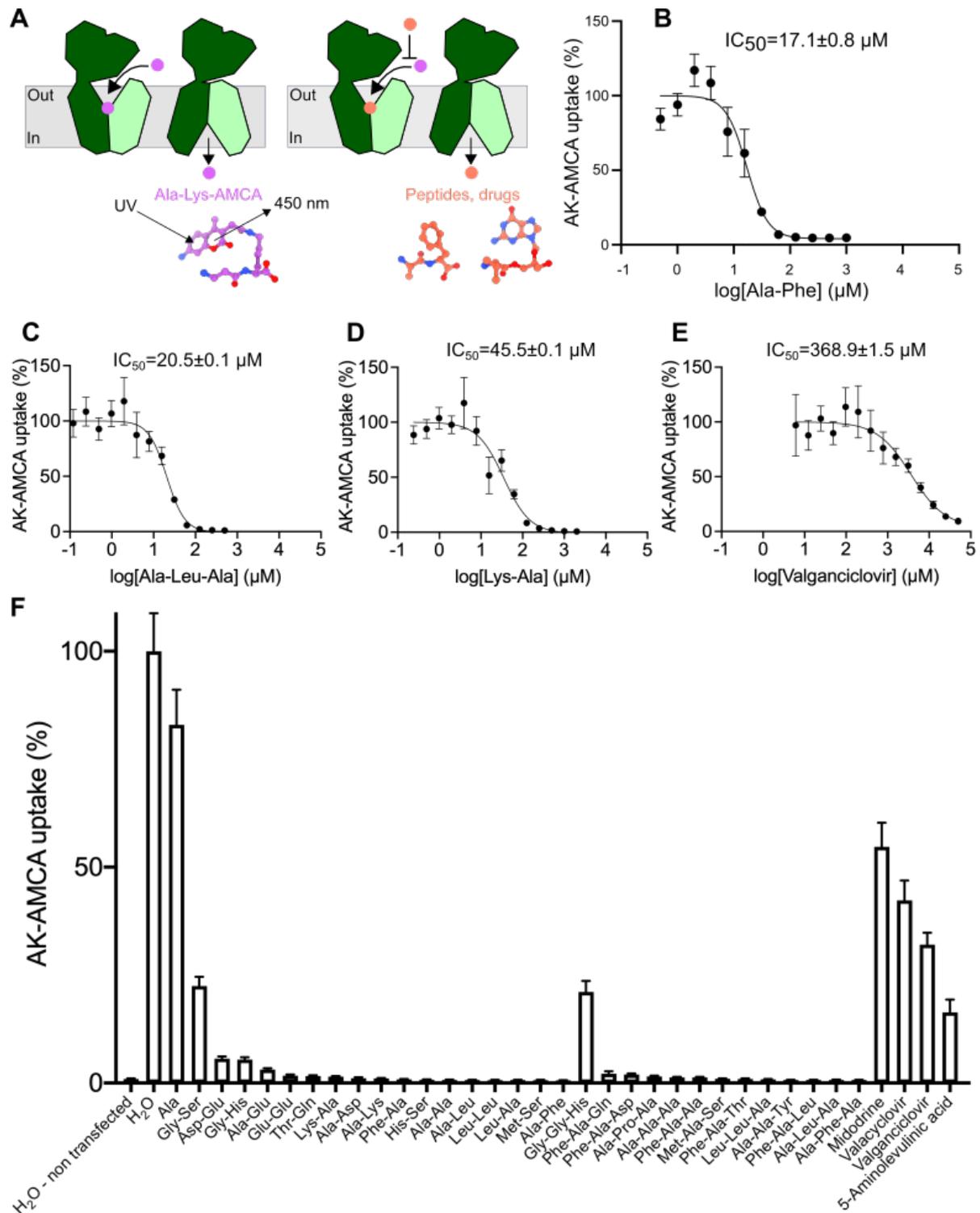


Figure 18. Transport assays in human embryonic kidney (HEK) cells. The β -Ala-Lys peptide coupled to the fluorescent reporter AMCA (AK-AMCA) is a known substrate of prototypical POTs. Here, its intracellular accumulation was measured in transfected HEK293F, cells in absence or presence of competing dipeptides, tripeptides, or drugs as schematized in (A). (B, C, D, E) Concentration dependent competition of the fluorescent reporter with the stated peptide or drug. (F) The assay was repeated on several substrates competing at a concentration of 5 mM. The average uptake value for each condition was calculated from three

independent measurements. The error bars correspond to the standard deviation from these independent measurements. This figure was adapted from (Killer *et al.*, 2021).

This recombinant construct is referred to, as His-EGFP-HsPepT2, and was well solubilised in N-dodecyl- β -D-maltopyranoside (DDM) detergent. When this work started in 2018, MFS transporters, including POTs, were not yet studied by cryo-EM, but rather by MX. Therefore, the biochemistry already in place (Pieprzyk, Pazicky and Löw, 2018a) was scaled up (**Figure 19**) to perform crystallization trials, with the following modifications: (i) the N528Q and N587Q mutations were introduced to remove two of the five predicted glycosylation sites, (ii) the lipid cholesteryl hemisuccinate (CHS), (iii) and the dipeptide alanine-phenylalanine (Ala-Phe) were added to the sample, to stabilise the transporter (**Figure 20**). All available commercial screens were tested multiple times, using the vapor diffusion method and in the lipidic cubic phase method, at various protein concentrations and temperatures. Mesophases of different lipid compositions were tested, to accommodate the soluble extracellular domain of HsPepT2 inside the cubic phase. Yet, no crystals appeared in these trials. From there, the purification protocol was modified, to yield a purer and more homogenous sample. A nanobody specific to EGFP (α EGFP-Nb) was coupled to affinity beads, and was used to increase the purity of the sample (**Figure 19 E-H**). The latter was successful, but led to a longer purification workflow. An improved approach was to modify the initial construct, by removing the His-EGFP element, and replacing it with a twin-streptavidin tag followed by a 3C protease cleavage site (referred to as the StrepII-HsPepT2 construct). This approach allowed to reduce the purification workflow duration, to less than a day, compared to two to three days for the previous ones. In addition, the StrepII approach addressed purity issues as effectively, as the α EGFP-Nb approach (**Figure 19 I-L**). Finally, in order to reduce the heterogeneity brought by the three remaining glycosylation sites, StrepII-HsPepT2 was expressed in a Glucose N-acetyltransferase 1 deficient cell line (HEK293S GnTI-), or treated enzymatically with the glycosidases Peptide:N-glycosidase F (PNGase F) or Endoglycosidase H (Endo H) (**Figure 21**). The crystallisation campaign was repeated on these three samples, but again failed to yield any crystal, neither by the vapor diffusion, nor the LCP method.

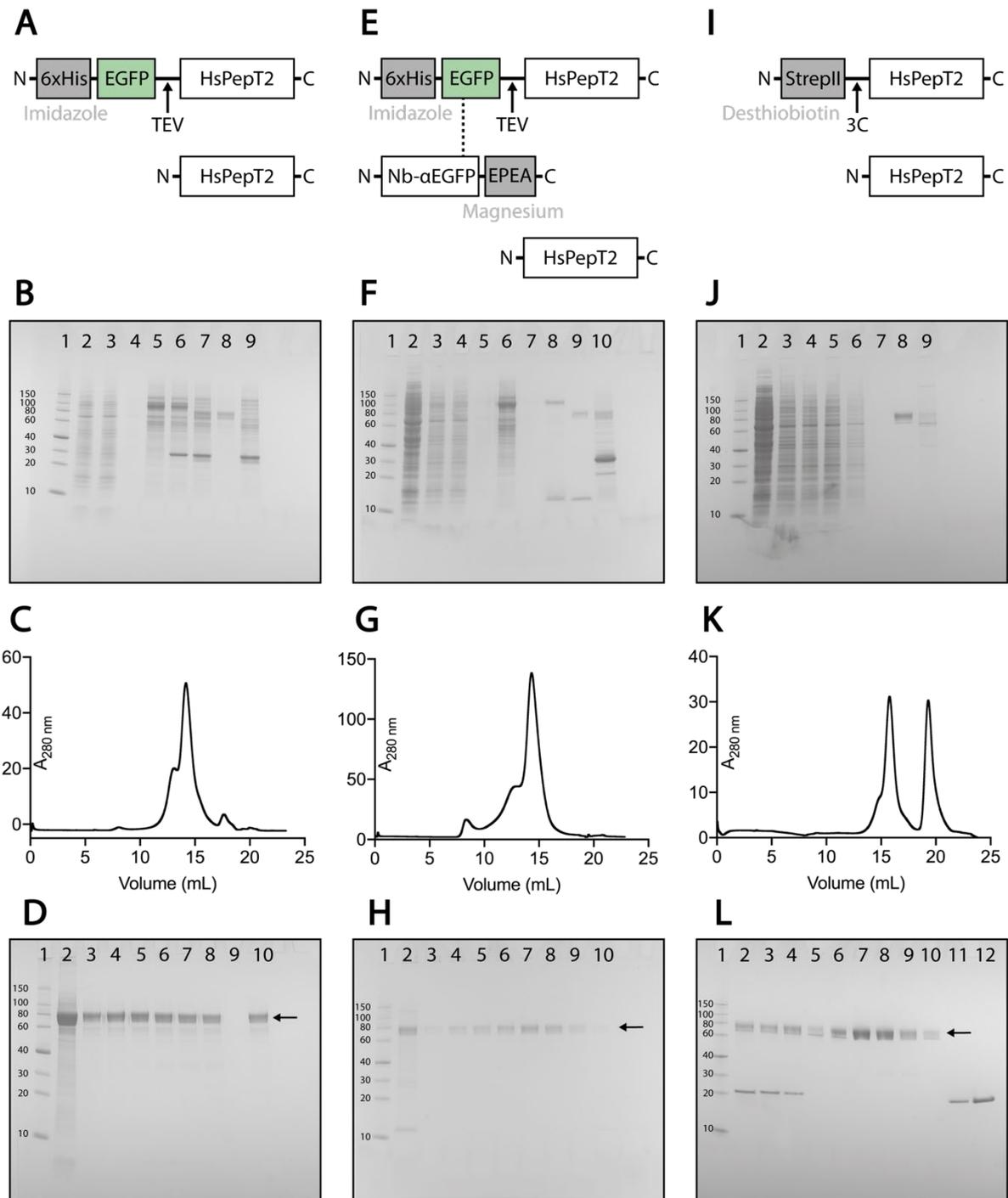


Figure 19. Purification strategies of HsPepT2. (A-D) The His-EGFP-HsPepT2 construct (A) was first purified by Ni-NTA affinity chromatography. (B) SDS-PAGE of the Ni-NTA purification: Molecular weight protein markers, 2: solubilised membranes, 3: Ni-NTA resin flow-through, 4: Ni-NTA resin wash, 5: Ni-NTA resin elution (250 mM imidazole), 6: addition of TEV protease time $t=0$, 7: time $t=16$ hours. 8: flow through from reverse Ni-NTA purification after TEV Cleavage of His-EGFP, 9: elution of the reverse Ni-NTA purification. (C) SEC chromatogram of fraction 8 from (B). (D) SDS-PAGE of the SEC. 1: Molecular weight protein markers, 2: injected sample, 3-8: fractions under the peak, 9: empty, 10: pooled fractions. (E-F) The His-EGFP-HsPepT2 construct (E) was purified by Ni-NTA and

CaptureSelect affinity chromatography. (F) SDS-PAGE of the Ni-NTA and CaptureSelect purification. 1: Molecular weight protein markers, 2: total cell lysate, 3: solubilised membranes, 4: Ni-NTA resin flow-through, 5: Ni-NTA resin wash, 6: Ni-NTA resin elution (250 mM imidazole), 7: flow through CaptureSelect resin, 8: elution CaptureSelect resin (2M MgCl₂), 9: flow through from reverse Ni-NTA purification after TEV Cleavage of His-EGFP-Nb α EGFP complex, 10: elution of the reverse Ni-NTA purification. (G) SEC chromatogram of fraction 9 from (F). (H) SDS-PAGE of the SEC. 1: Molecular weight protein markers, 2: injected sample, 3-10: fractions under the peak. (I-L) The StrepII-HsPepT2 construct (I) was purified using the strep-Tactin affinity chromatography. (J) SDS-PAGE of the strep-Tactin purification. 1: Molecular weight protein markers, 2: total cell lysate, 3: membrane fraction, 4: solubilised membranes, 5: StrepTactin resin flow-through, 6: StrepTactin resin first wash, 7: StrepTactin second wash, 8: StrepTactin resin elution (10 mM desthiobiotin), 9: protein leftovers on StrepTactin resin. (K) SEC chromatogram of fraction 8 from (J). (L) SDS-PAGE of the SEC. 1: Molecular weight protein markers, 2: purified StrepII-HsPepT2 supplemented with 3C protease, at time t=0. 3: purified StrepII-HsPepT2 supplemented with 3C protease, at time t=30 mins. The StrepII tag is already cleaved, as noticeable by the slight migration shift on the gel. 4: Injected sample. 5-10: fractions under the first peak (HsPepT2). 11-12: fractions under the second peak (3C protease). The target protein HsPepT2 is indicated by an arrow in the SEC SDS-PAGEs (D,H,L).

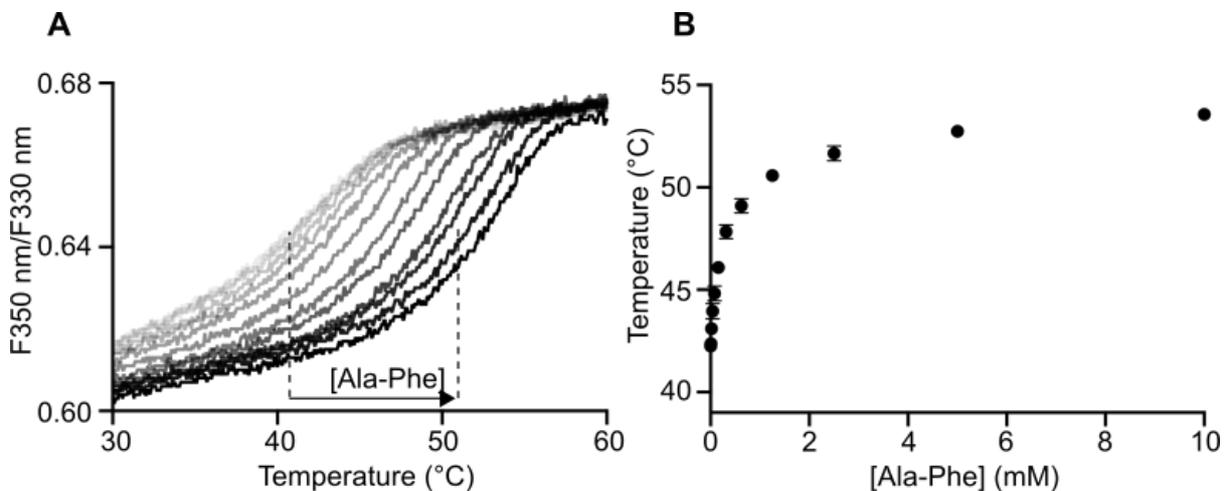


Figure 20. Stabilisation of purified HsPepT2 in presence of the dipeptide Ala-Phe. (A) Thermal stabilization upon substrate binding measured by nanoDSF at increasing concentrations of Ala-Phe. (B) The increase in melting temperature with increasing peptide concentration is shown. This figure was adapted from (Killer *et al.*, 2021)

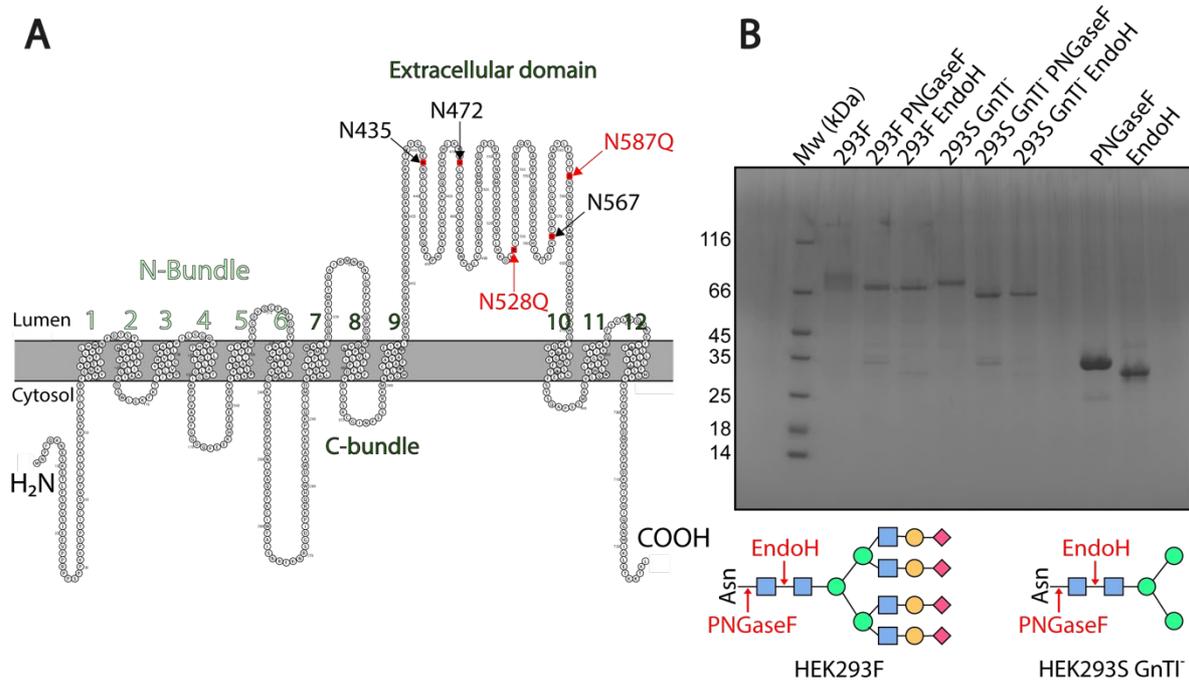


Figure 21. Strategies to tackle glycan heterogeneity in HsPepT2. (A) The double mutation N528Q and N587Q was introduced to remove two out of five N-linked glycan chains. (B) The purified double mutant was further treated by the glycosidases PNGaseF or EndoH, after expression in HEK293F or HEK293S lacking N-acetylglucosaminyltransferase I activity (GnTI⁻). The sites of hydrolysis for the two glycosidases are schematized in the cartoon in the lower panel.

6.2.2 Structure determination of HsPepT2

The first MFS transporter structure, determined by SPA cryo-EM, was deposited in the PDB in 2020, using a fab as fiducial marker to facilitate particle alignment (Li *et al.*, 2020). One year before that, a monomeric SLC transporter bound to a protein containing a rigid soluble domain, was already released (Yan *et al.*, 2019). These two example, although scarce, demonstrated that SPA cryo-EM could be an alternative to MX, for structure determination of monomeric SLCs and MFS transporters. This possibility was therefore explored with HsPepT2.

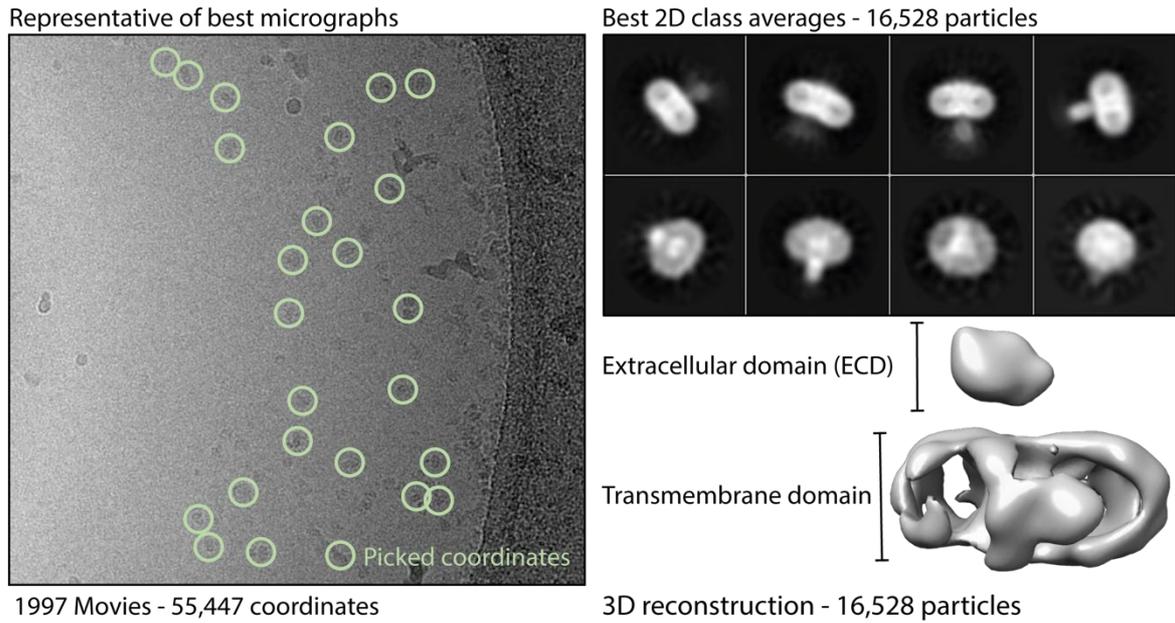


Figure 22. Initial SPA cryo-EM trial on HsPepT2 purified in DDM-CHS. The micrographs contain few coordinates, which reduces the amount of replicates collected within a certain time. The 2D class averages display some signal for the transporter, but the small amount of data does not allow to reconstruct a 3D volume with secondary structure elements.

The optimization of the sample preparation done throughout the crystallisation trials, allowed starting from a pure and homogenous sample (StrepII-HsPepT2). The first difficulties faced during grid preparation were the effective concentration of particles inside the vitrified meniscus, and the poor contrast between the protein and the solvent. 2D particle averaging only allowed to visualise some low resolution features, such as the shape of the extracellular domain, and the detergent micelle surrounding the transmembrane domain. (Figure 22).

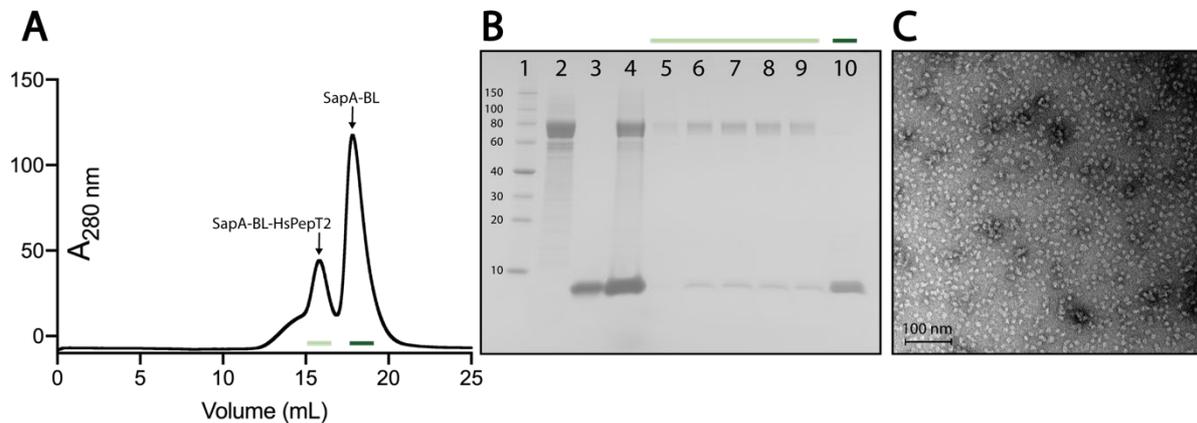


Figure 23. Reconstitution of HsPepT2 in Salipro. (A) SEC chromatogram after reconstitution of HsPepT2 in Saposin-Brain lipids nanoparticles. (B) SDS-PAGE of the reconstitution. 1: molecular weight protein markers, 2: pre-purified HsPepT2 solubilised in DDM-CHS, 3: purified Saposin-A lipid binding protein, 4: injected sample on the SEC column

(i.e., HsPepT2 mixed with saposin-A, brain lipids, and biobeads), 5-9: fractions under the first peak (HsPepT2-SapA-BL), 10: top fraction of the second peak (Saposin-A). (C) negative stain (uranyl acetate) electron micrograph of reconstituted HsPepT2-SapA-BL, diluted to 0.05 absorbance units at 280 nm.

To improve particle distribution, and contrast, HsPepT2 was reconstituted in Salipro. The rationale, being that detergents can (i) influence the surface tension properties of the solution in the meniscus of the grid holes (therefore impacting the dispersion of the protein), and (ii) scatter electrons, just like any atoms present in the sample (therefore reducing contrast between the solvent and the protein of interest). An initial screening of different molar ratios of HsPepT2:Lipid:Saposin was done with brain lipid, and soy lipid extracts. Judging from SEC profiles, the most homogenous and stable condition (i.e., in brain lipids) was chosen, scaled up for negative staining EM (**Figure 23**), and further up for SPA cryo-EM (**Figure 24**). The particle distribution and the contrast were both greatly improved by the Salipro reconstitution. However, the particles adopted a preferred orientation on the cryo-EM grid (mainly top views) resulting in a poorly interpretable volume.

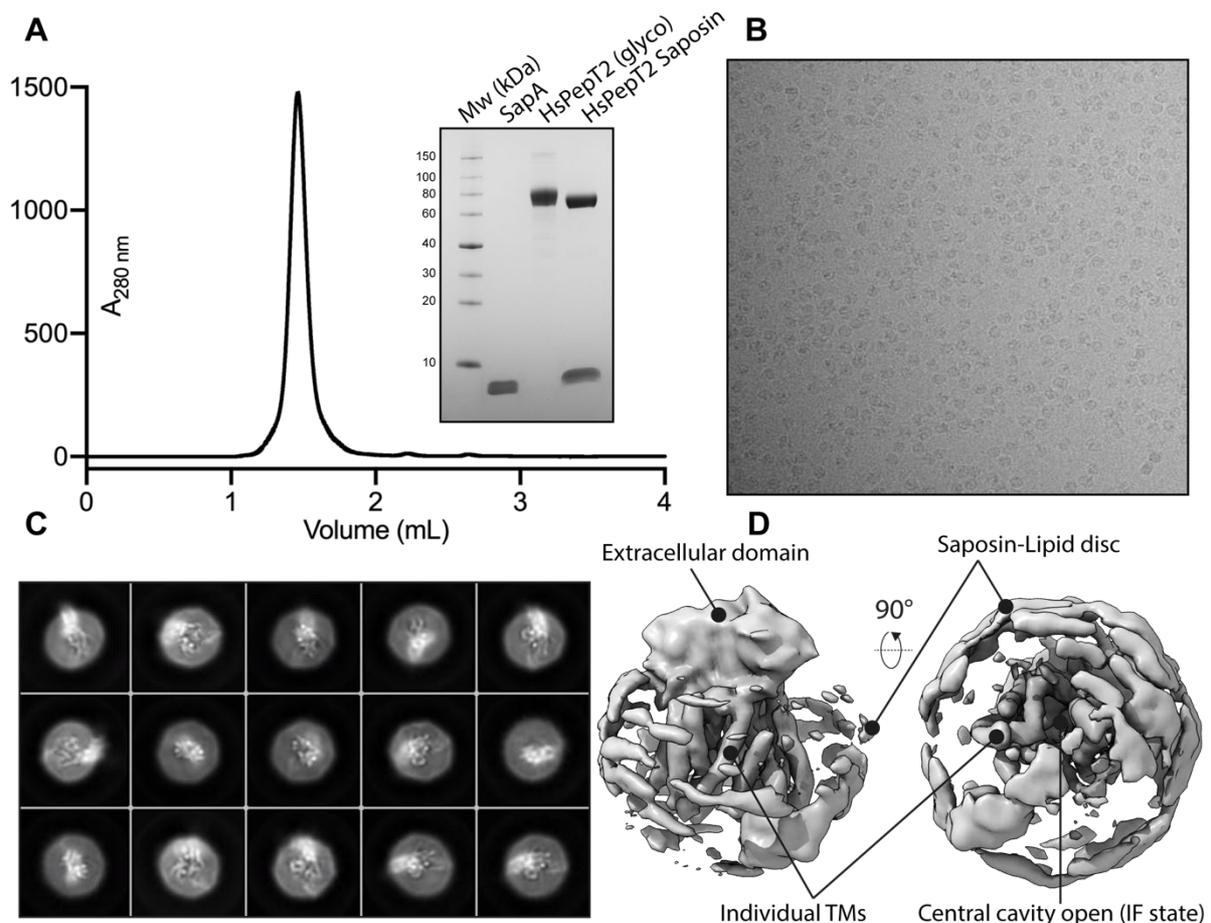


Figure 24. SPA cryo-EM on HsPepT2 in Salipro. (A) SEC chromatogram of reconstituted HsPepT2-Salipro, with corresponding SDS-PAGE. Note that the sample was de-glycosylated by the enzyme PNGase F. (B) Representative cryo-EM micrograph of the sample magnified 105,000 times, with an estimated defocus of 1.5 μm , showing near optimal particle distribution. (C) Representative 2D class averages of curated particles, showing protein secondary structure elements within the Salipro disc, all oriented parallel to the image plane. (D) 3D reconstruction of the selected particles, showing 12 transmembrane helices surrounded by the Salipro disc, and the extracellular domain. The resolution is low (i.e., $\sim 7\text{\AA}$), but allows to distinguish the conformational state of the transporter. This figure was adapted from (Killer *et al.*, 2021)

Next, HsPepT2 was solubilized in the low CMC detergent LMNG supplemented with CHS (**Figure 25 A-C**). Particle distribution was also improved in comparison to the DDM-CHS sample, however side views (i.e., transmembrane helices parallel to the image plane) were lacking, just as with the Salipro sample. Interestingly, solubilizing HsPepT2 in DDM-CHS, and exchanging the detergent to LMNG at the final gel filtration step improved the angular distribution of the particles and allowed to reach a 5 \AA reconstruction (**Figure 25 D-F**).

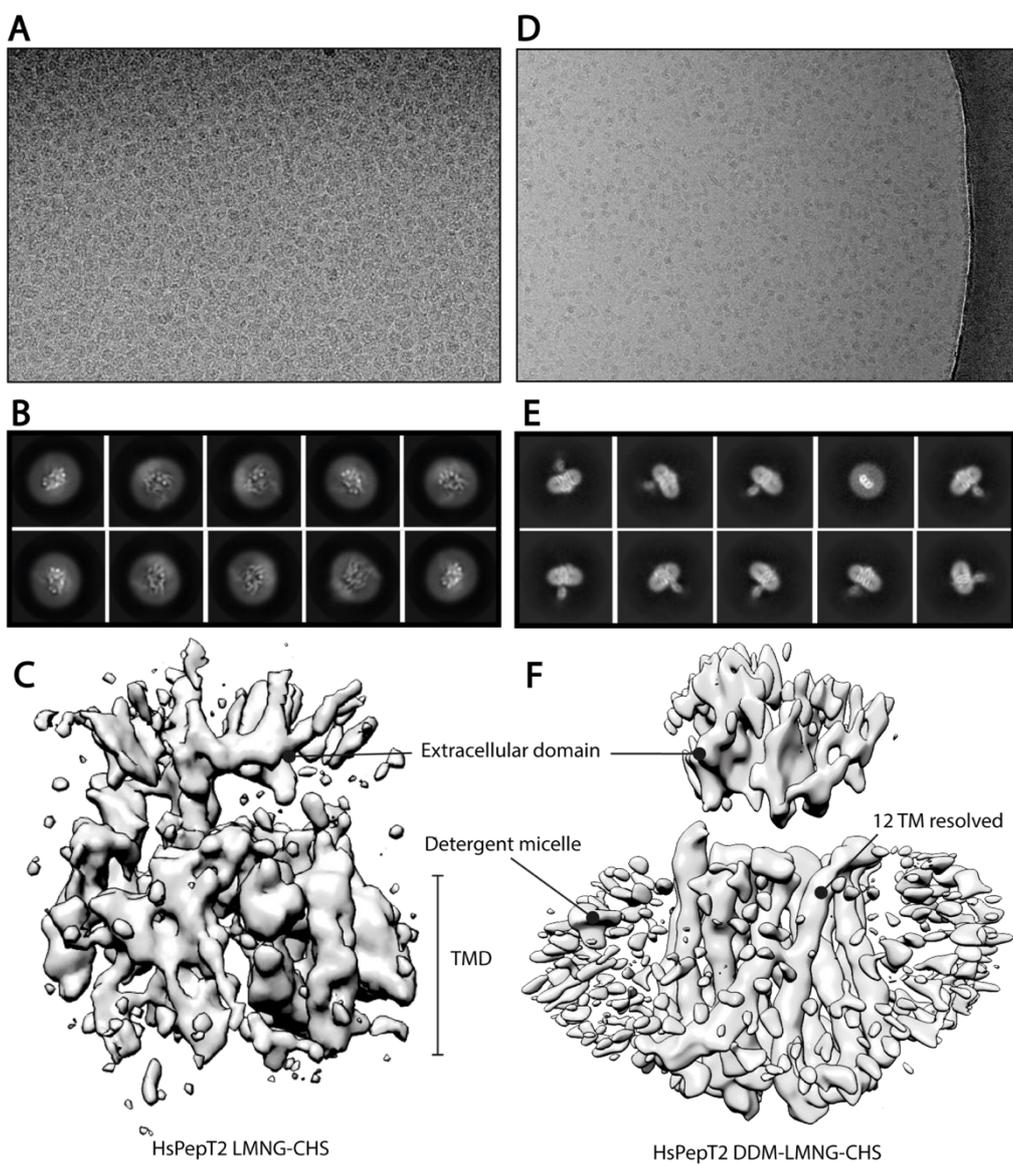


Figure 25. SPA cryo-EM on HsPepT2 with the detergent LMNG. (A-C) HsPepT2 solubilised and purified in LMNG-CHS was image (A) but showed a similar preferred orientation as the previous Salipro sample (B), resulting in a poorly interpretable 3D volume (C). (D-F) Solubilisation in DDM-CHS, followed by an exchange to LMNG-CHS, restored the views where the longitudinal plane of the micelle sits perpendicular to the image plane. (i.e., ‘side view’). The latter benefits to more accurate and complete 3D reconstruction (F) displaying all 12 transmembrane helices of the transporter, resolved at approximately 5Å.

HsPepT2 solubilized and purified in DDM-CHS was re-vitrified, this time with lower concentration of detergent (0.015% instead of 0.03%), salt (100 mM NaCl instead of 150 mM), and buffer (10 mM Hepes instead of 20 mM), and in the presence of 10 mM of the stabilising dipeptide Ala-Phe (**Figure 26**). Since HsPepT2 has a molecular weight of 80 kDa, and is surrounded by a micelle, concentrators with 100 kDa cutoff membranes can be used to increase the protein concentration in solution. This leads however to accumulation of empty micelles,

and increased detergent concentration. To remove these empty micelles, the sample was run on an analytical SEC column coupled to an HPLC system, and the highly concentrated top fraction of the elution peak was vitrified.

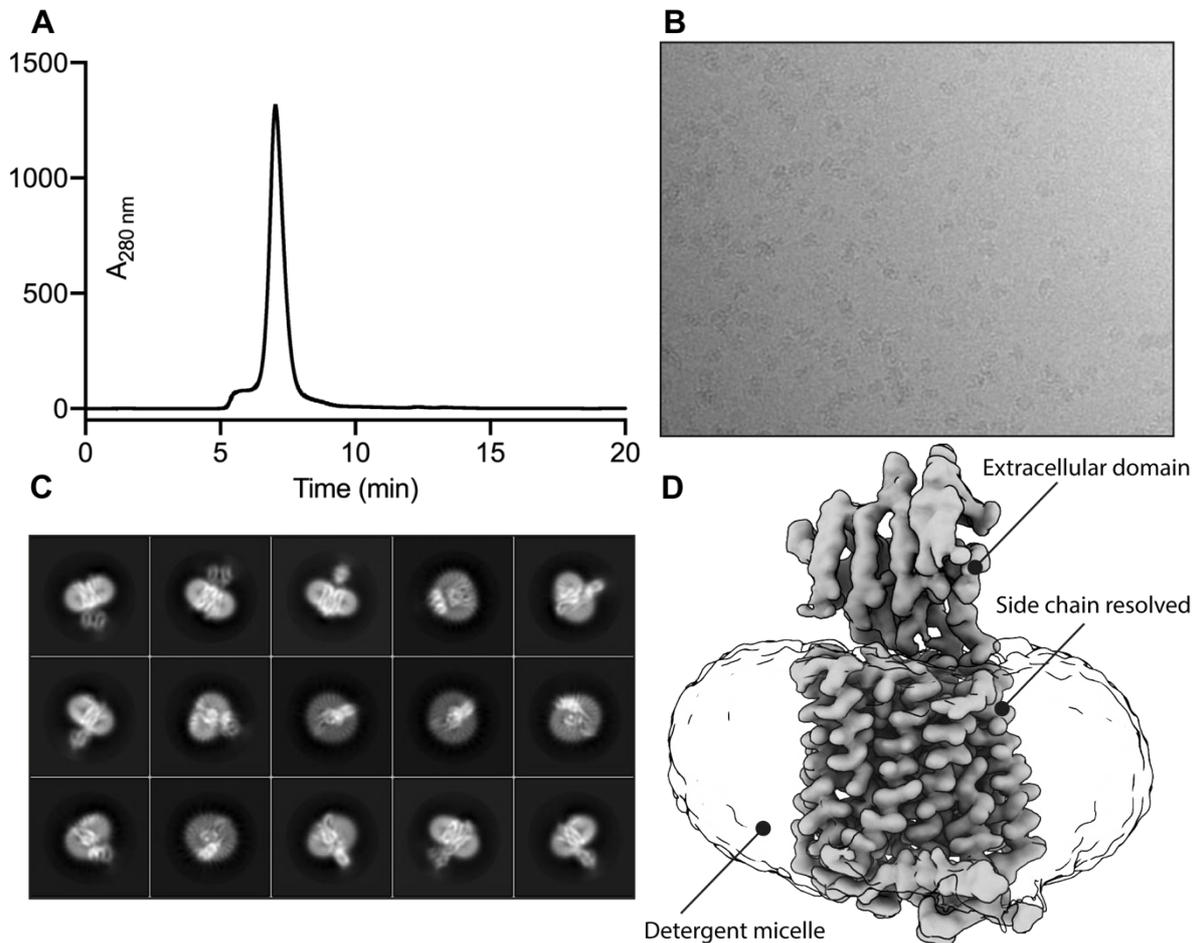


Figure 26. Final SPA cryo-EM trial on HsPepT2 with the detergent DDM-CHS. (A) SEC chromatogram. (B) Representative micrograph of the imaged sample (peak top fraction in (A)). (C) Representative 2D class averages, showing a diversity of orientations, and clear secondary structure elements. (D) 3D reconstruction of HsPepT2 at a nominal resolution of 3.8 Å. The side chains of the transport unit are clearly resolved. This figure was adapted from (Killer *et al.*, 2021)

To further improve the signal to noise ratio in the images, areas of thin ice were selected, at the cost of particle abundance per field of view. All these factors combined together led to a dataset of superior quality compared to the initial DDM-CHS dataset, as judged from the high contrast between the protein and the solvent. This also allowed to collect data with a low defocus range, while still being able to clearly visualise the particles. During single particle analysis, the extracellular domain of HsPepT2 displayed some degree of flexibility relative to the transporter unit. Iterative cycles of 3D clustering using the software Relion-3.1 allowed to obtain a more

homogenous population of particles, and reach higher resolution in the final 3D reconstruction, using the newly released non uniform refinement algorithm from cryoSPARCv3 (**Figure 27**).

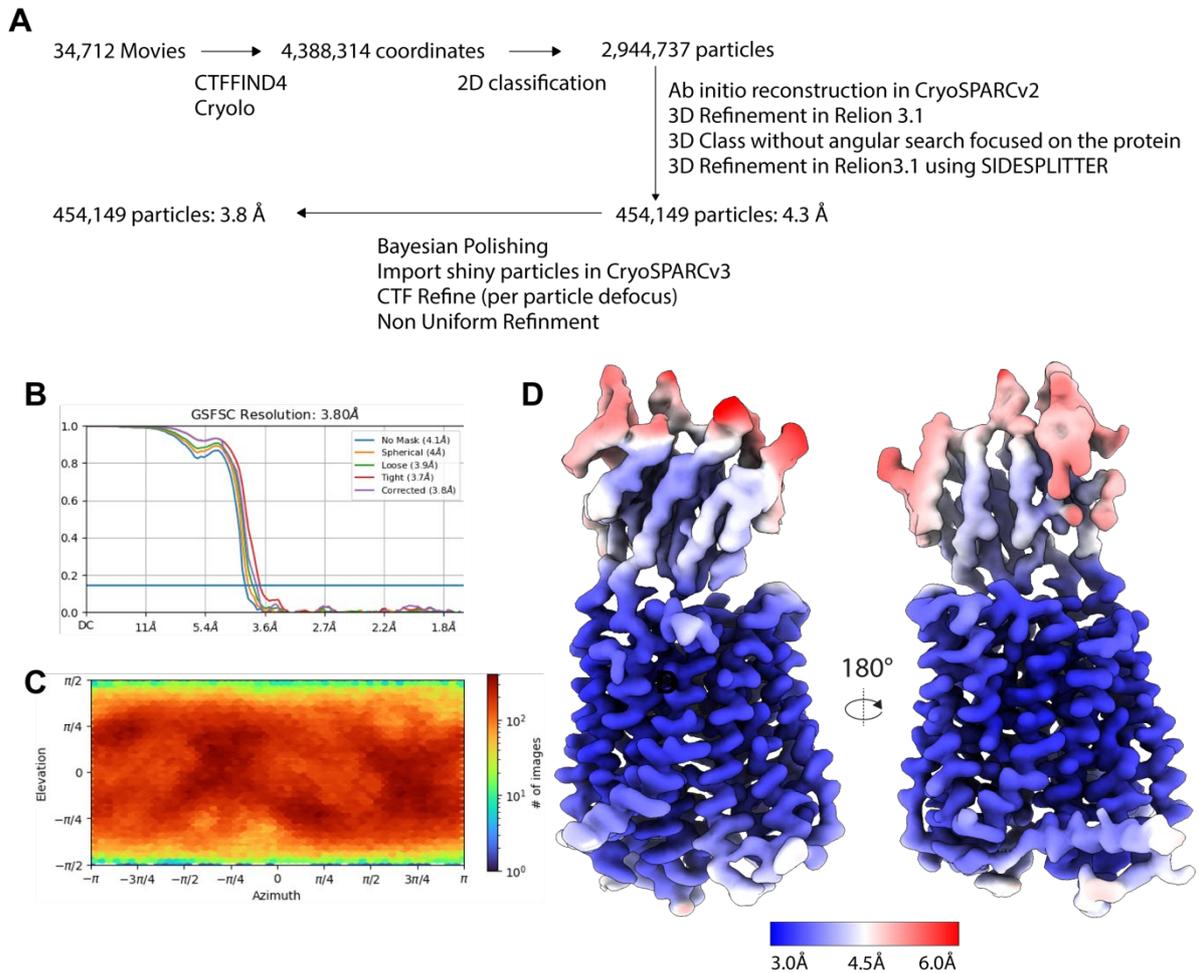


Figure 27. Cryo-EM data processing for inward facing partially occluded HsPepT2 bound to Ala-Phe. (A) Flowchart showing the image processing pipeline. Initial processing was performed in Relion-3.1. Particles were then transferred to cryoSPARCv3 for CTF-Refinement and Non Uniform refinement. The numbers of particles moving into each step are noted. (B) Final refinement from cryoSPARCv3 FSC curve. (C) angular distribution. (D) 3D volume coloured by local resolution, estimated in cryoSPARCv3 using the 0.5 as FSC threshold. This figure was adapted from (Killer *et al.*, 2021).

The nominal resolution of the final map was 3.8 Å and the estimated local resolution of the transporter unit reached 3.2 Å (**Figure 27 D**). This allowed to model the transmembrane region of HsPepT2 *de novo* (**Figure 28**).

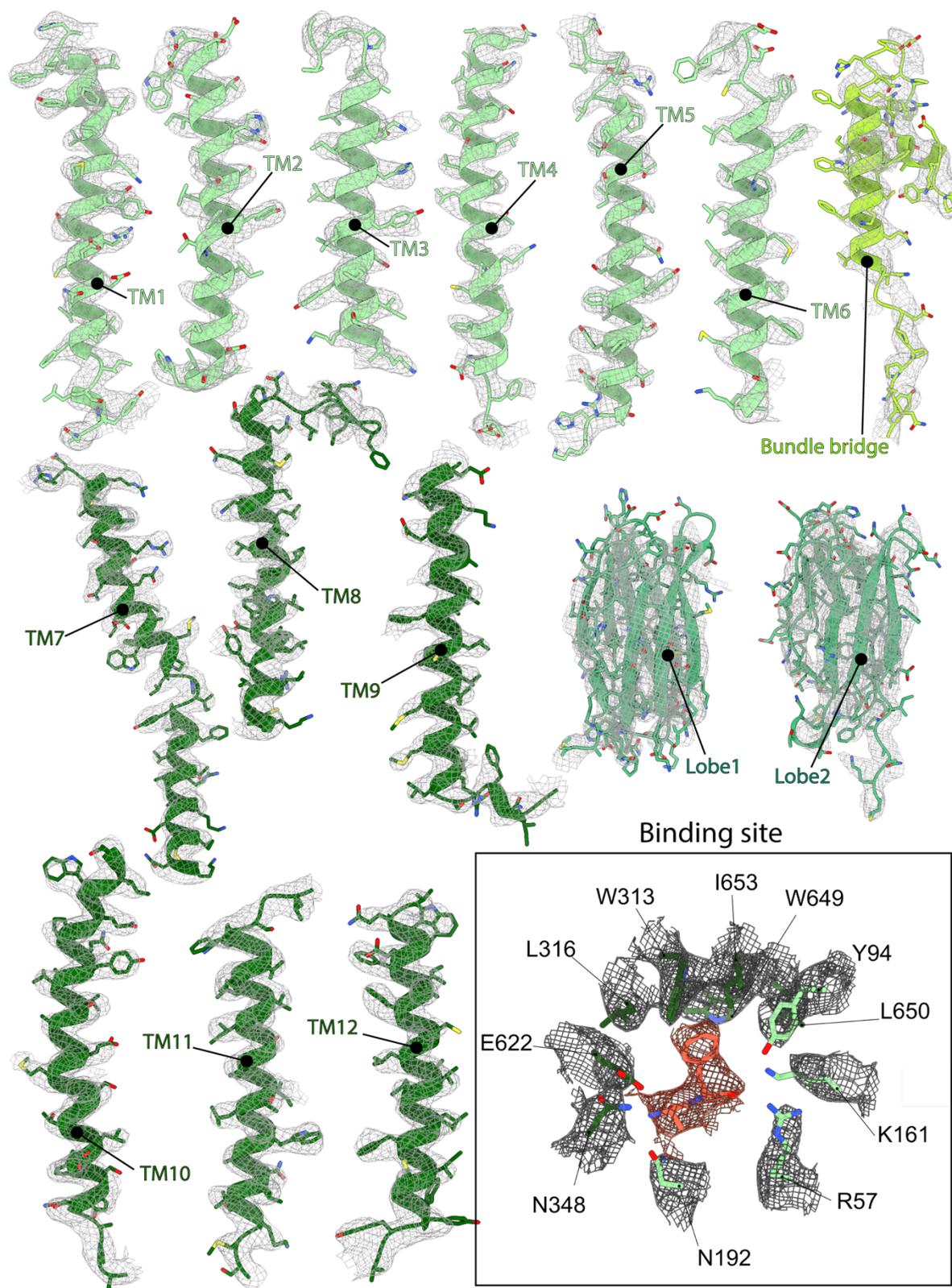


Figure 28. Modelling of HsPepT2 inside the density map. The density is shown as grey mesh for individual transmembrane helices, bundle bridge and the extracellular domain. The mesh depicts density within a 2.6 Å radius of any modelled atom. The close up view shows the density within the binding site. The dipeptide Ala-Phe is coloured in orange. This figure was adapted from (Killer *et al.*, 2021).

The transmembrane domain has the characteristic features of canonical MFS transporters: i.e., 12 TMs with both N and C termini facing the cytoplasm (**Figure 29**). No density was present for the first 40 residues at the N terminus and the last 32 C-terminal residues. These regions are also predicted to be disordered (McGuffin, Bryson and Jones, 2000), and were therefore not modelled. The 12 TMs arrange in two bundles, and connect *via* a linker between TM6 and TM7. This linker has two amphipathic α helices, which we named “bundle bridge” (**Figure 30**).

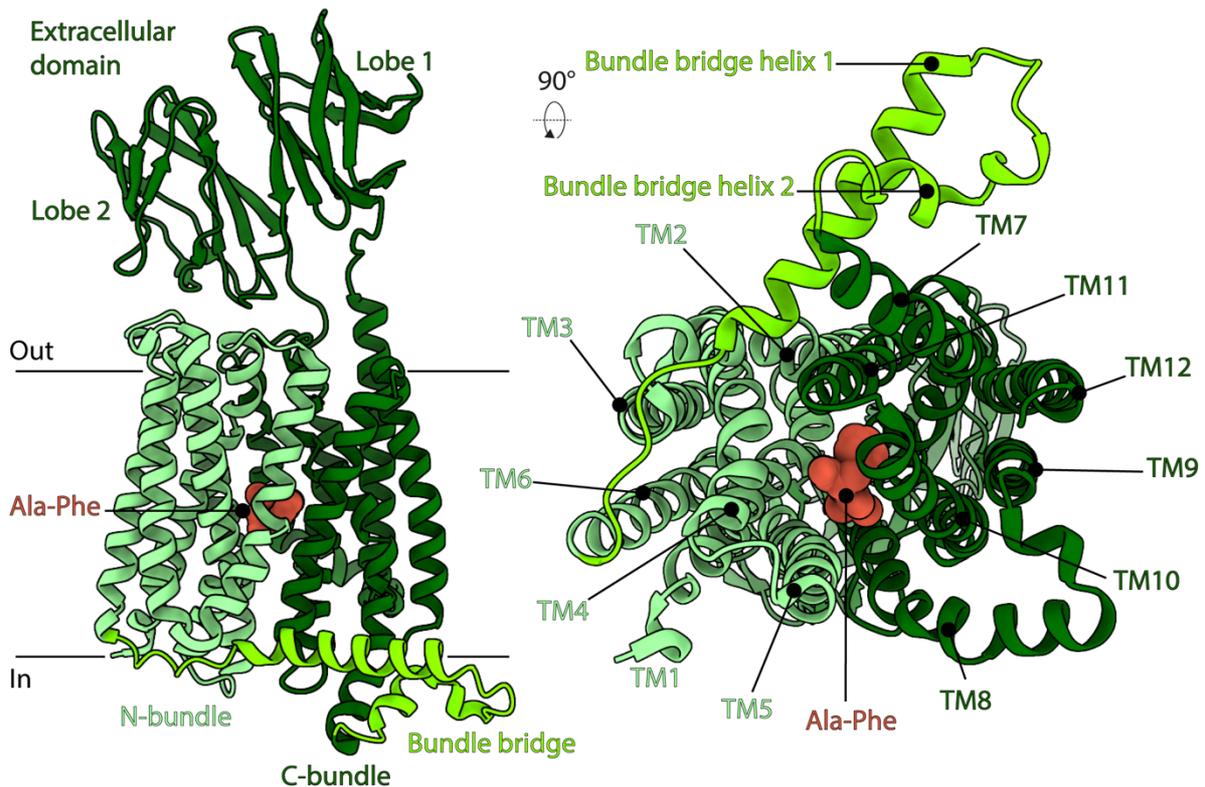


Figure 29. Atomic model IF partially occluded HsPepT2 bound to Ala-Phe, shown as ribbon representation. The different architectural elements are labelled. This figure was adapted from (Killer *et al.*, 2021).

The extracellular domain (ECD) of HsPepT2, is located between TM9 and TM10. A homologous ECD was isolated and crystallized in 2015 from the brown rat (RnPepT2^{ECD}), and revealed two immunoglobulin lobes interacting with each other (Beale *et al.*, 2015). This crystal structure was initially used, as template for flexible docking inside the density of HsPepT2, after mutating the residues to match the human sequence. During manuscript revision, the AlphaFold2 predictions became openly available (Jumper *et al.*, 2021). Using the prediction of the human ECD, for refinement, further improved the fit of the model into the density. Strong additional signal was present in the central cavity of HsPepT2. Since 10 mM of the dipeptide Ala-Phe was added to the sample before vitrification, and binding was verified

by thermal shift assays (**Figure 20**), it was rather certain that this density corresponded to the substrate. In addition, the peptide density matched the expected pose of similar ligands co-crystallized with prototypical bacterial POTs, in previous studies (Huang *et al.*, 2016; Martinez Molledo *et al.*, 2018) and in further experiments carried out during this thesis (i.e., work on DtpB).

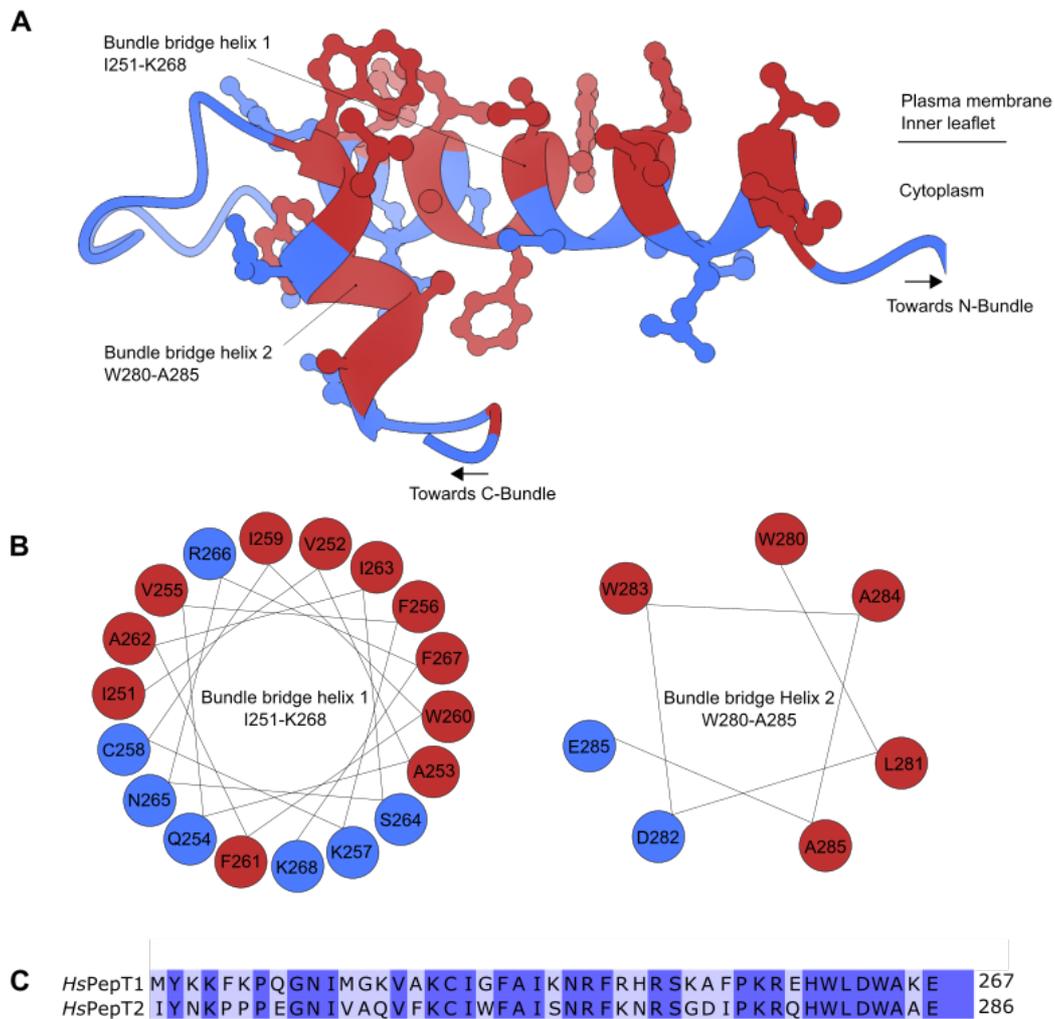


Figure 30. Amphipathic nature of the bundle bridge. (A) Ribbon representation of the HsPepT2 bundle bridge. Polar residues are coloured in blue, hydrophobic residues are coloured in red. (B) Edmundson wheel projection diagram of the bundle bridge helices 1 and 2 showing the concentration of hydrophobic residues facing the inner leaflet of the plasma membrane and the presence of polar residues facing the cytoplasm. (C) Sequence alignment of the bundle bridge from HsPepT1 and HsPepT2. This figure was directly taken from (Killer *et al.*, 2021)

6.2.3 Structure determination of HsPepT1

Previous work in the lab showed that as a His-EGFP-HsPepT1 construct led to very poor expression levels in HEK293F cells. After determining the structure of HsPepT2, the His-

EGFP element was replaced with a twin-streptavidin tag, followed by a 3C protease cleavage site (referred to as StrepII-HsPepT1 construct). This construct expressed to higher levels than His-EGFP-HsPepT1, and could be purified within a day, using the optimized workflow developed for StrepII-HsPepT2.

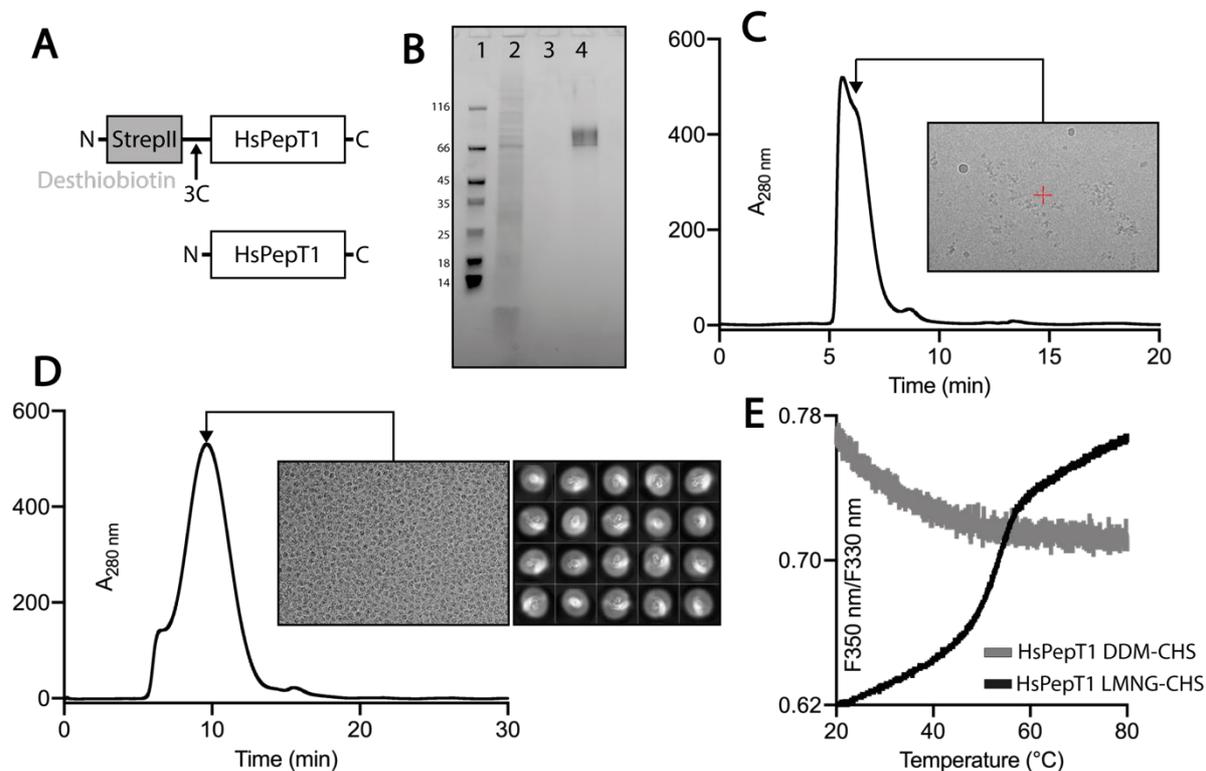


Figure 31. Initial SPA cryo-EM trials on HsPepT1. (A) Representation of the StrepII-HsPepT1 construct. (B) Affinity purification of HsPepT1 using Strep-Tactin resin. 1: Molecular weight protein markers, 2: Strep-Tactin resin flow through, 3: Strep-Tactin resin wash, 4: Strep-Tactin resin elution (10 mM desthiobiotin). (C) HsPepT1 in DDM-CHS, SEC chromatogram and cryo-EM micrograph of the indicated fraction, showing aggregated material. HsPepT1 in LMNG-CHS SEC chromatogram, cryo-EM micrograph, and 2D class averages, of the sample from the indicated fraction, showing soluble material, but preferred orientation. (E) The absence of a thermal unfolding transition in HsPepT1 DDM-CHS sample and the presence of one in the HsPepT1 LMNG-CHS sample, corroborates the respective aggregated and soluble states observed in the SEC experiments.

However, HsPepT1 aggregated after solubilisation in DDM-CHS (**Figure 31**). Solubilisation of the membranes in LMNG-CHS was performed and allowed to obtain a stable and folded sample. Images were collected, but showed clear preferred orientations, similarly as in the HsPepT2 LMNG-CHS sample. As an alternative, HsPepT1 was then extracted with LMNG-CHS, and DDM was progressively added during the affinity purification. This strategy allowed to obtain a folded sample, binding to the dipeptide Ala-Phe, and importantly, without preferred orientation on the EM grid. Separate datasets were collected, with, and without Ala-Phe. A 3.9

Å map was reconstructed for Apo-HsPepT1 (**Figure 32**, **Figure 33**, **Figure 34**). Two additional maps (3.5 Å, and 4.1 Å) were obtained in the presence of Ala-Phe (**Figure 32**, **Figure 35**, **Figure 36**). The three populations exhibited large rearrangements of the transmembrane domain, and the higher resolution map allowed to model the dipeptide and all coordinating residues with relative confidence.

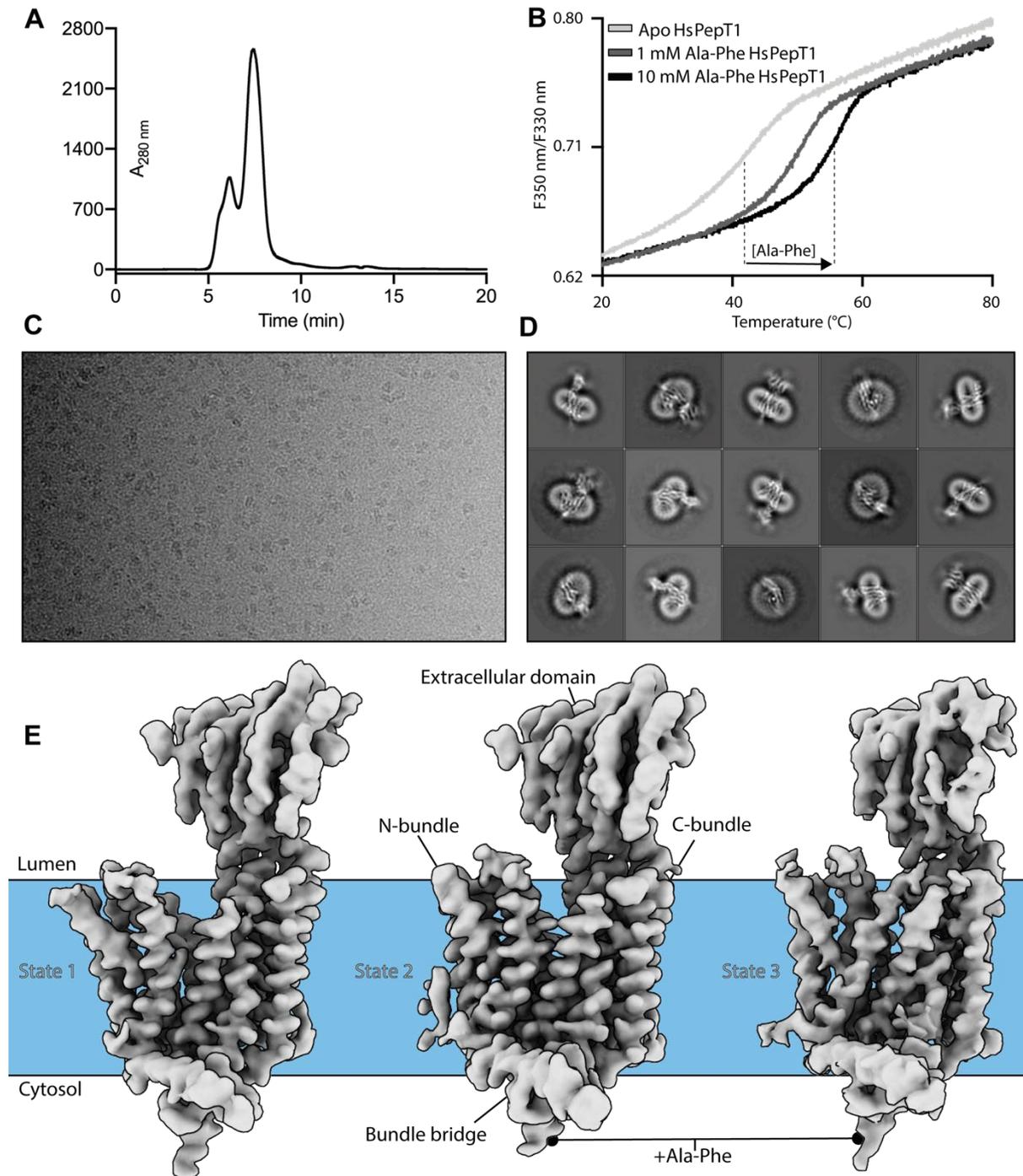


Figure 32. Cryo-EM structures of HsPepT1. (A) SEC chromatogram of HsPepT1 solubilised in a mixture of LMNG-DDM-CHS. (B) Thermal stabilization upon substrate binding measured by nanoDSF at increasing concentrations of Ala-Phe. (C) Representative cryo-EM micrograph,

of the sample. (D) Representative 2D class averages of HsPepT1. (E) 3D volumes of HsPepT1, reconstructed in the absence or in the presence of the dipeptide Ala-Phe. This figure was adapted from (Killer *et al.*, 2021).

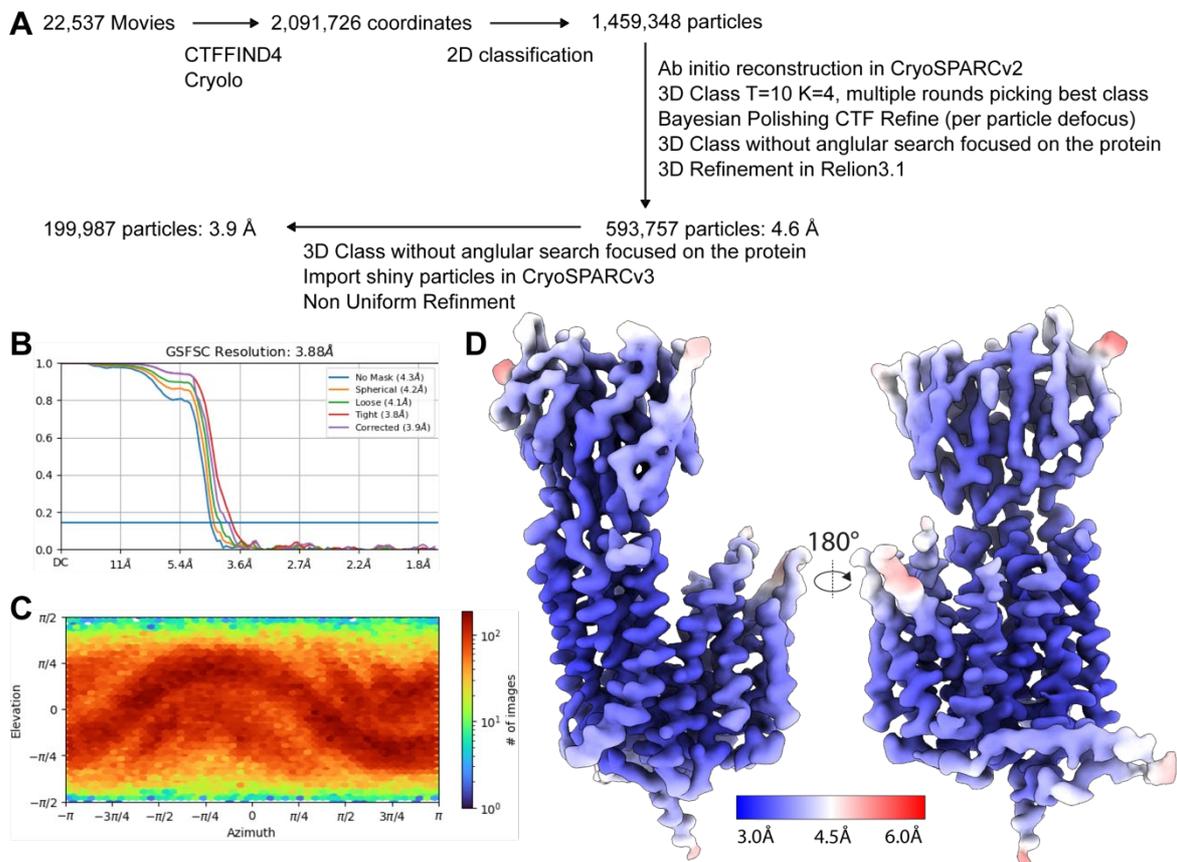


Figure 33. Cryo-EM data processing for outward facing, open, apo-HsPepT1 (OF, open, apo-HsPepT1). (A) Flowchart showing the image processing pipeline. Initial processing was performed in Relion-3.1. Particles were then transferred to cryoSPARCv3 for CTF-Refinement and Non Uniform refinement. The numbers of particles moving into each step are noted. (B) Final refinement from cryoSPARCv3 FSC curve. (C) Angular distribution. (D) 3D volume coloured by local resolution, estimated in cryoSPARCv3 using the 0.5 as FSC threshold. This figure was adapted from (Killer *et al.*, 2021).

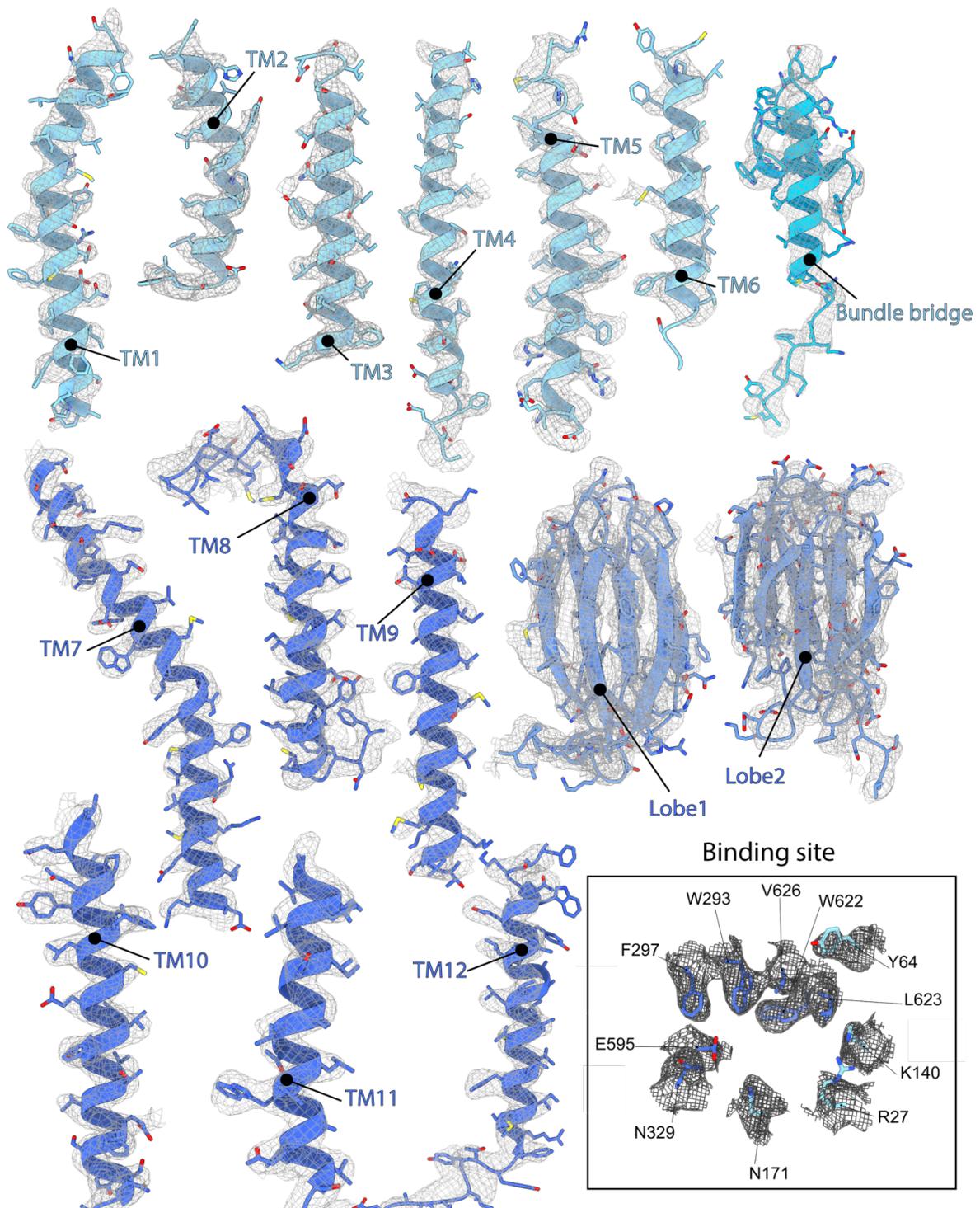


Figure 34. Modelling of OF, open apo-HsPepT1 inside the density map. The density is shown as grey mesh for individual transmembrane helices, bundle bridge and the extracellular domain. The mesh depicts density within a 2.6 Å radius of any modelled atom. The close up view shows the density within the binding site. This figure was adapted from (Killer *et al.*, 2021).

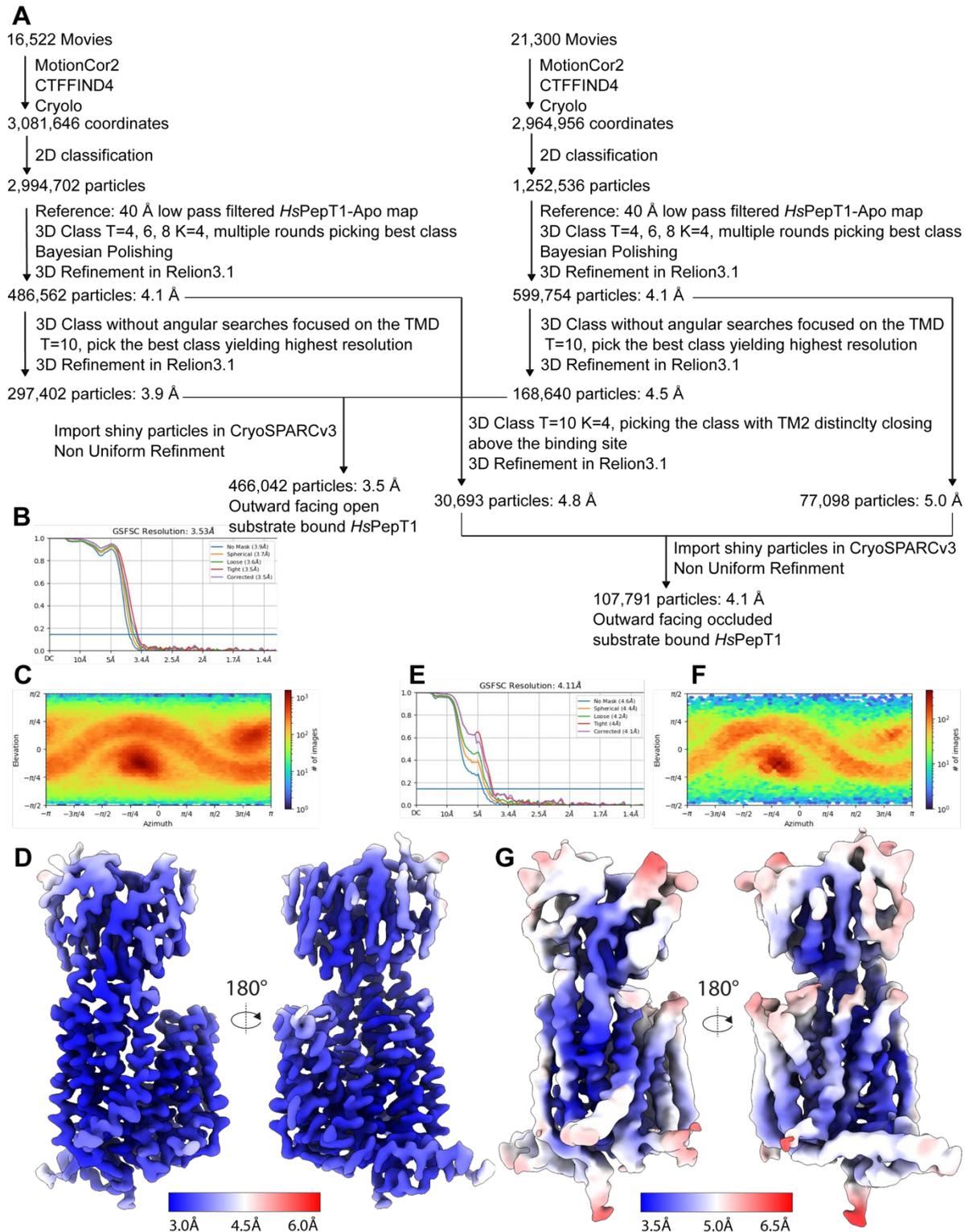


Figure 35. Cryo-EM data processing for outward facing, open and occluded, Ala-Phe bound-HsPepT1 (OF, open, and occluded Ala-Phe HsPepT1). (A) Flowchart showing the image processing pipeline. Initial processing was performed in Relion-3.1. Particles were then transferred to cryoSPARCv3 for CTF-Refinement and Non Uniform refinement. The numbers of particles moving into each step are noted. (B) Final refinement from cryoSPARCv3 FSC curve of OF, open, Ala-Phe HsPepT1 (C) angular distribution of OF, open, Ala-Phe HsPepT1.

(D) 3D volume of OF, open, Ala-Phe HsPepT1, coloured by local resolution, estimated in cryoSPARCv3 using the 0.5 as FSC threshold. (E) Final refinement from cryoSPARCv3 FSC curve of OF, occluded, Ala-Phe HsPepT1 (F) angular distribution of OF, occluded, Ala-Phe HsPepT1. (G) 3D volume of OF, occluded, Ala-Phe HsPepT1, coloured by local resolution, estimated in cryoSPARCv3 using the 0.5 as FSC threshold. This figure was adapted from (Killer *et al.*, 2021).

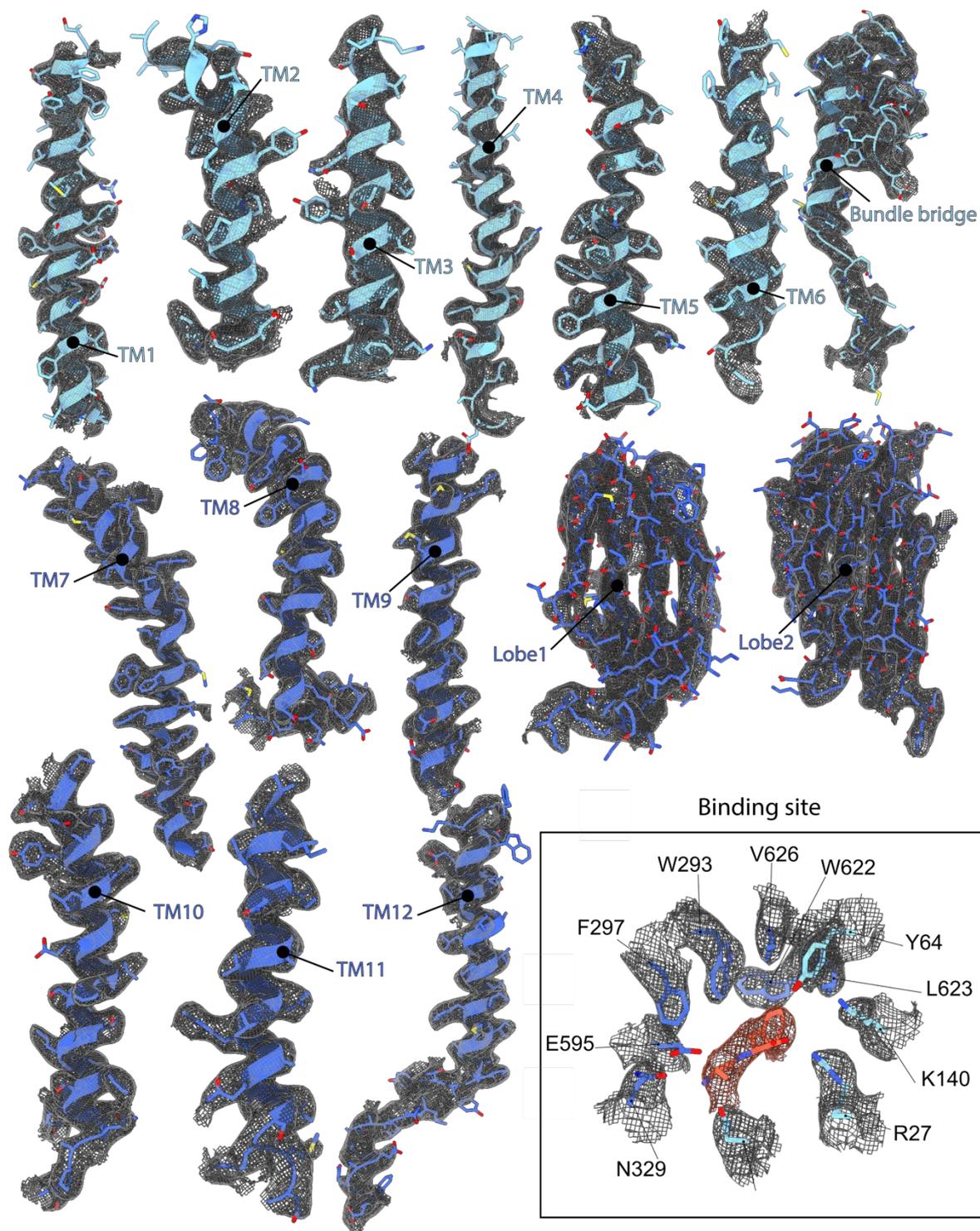


Figure 36. Modelling of OF, open Ala-Phe bound-HsPepT1 inside the density map. The density is shown as grey mesh for individual transmembrane helices, bundle bridge and the extracellular domain. The mesh depicts density within a 2.6 Å radius of any modelled atom. The close up view shows the density within the binding site. The dipeptide Ala-Phe is coloured in orange. This figure was adapted from (Killer *et al.*, 2021).

Architectural differences between mammalian and bacterial POTs

Human PepT1 and PepT2 exhibit striking architectural differences compared to their bacterial orthologues (**Figure 37**). The first, is the presence of the ECD, which is a unique feature present in mammalian SLC15A1 and SLC15A2 transporters.

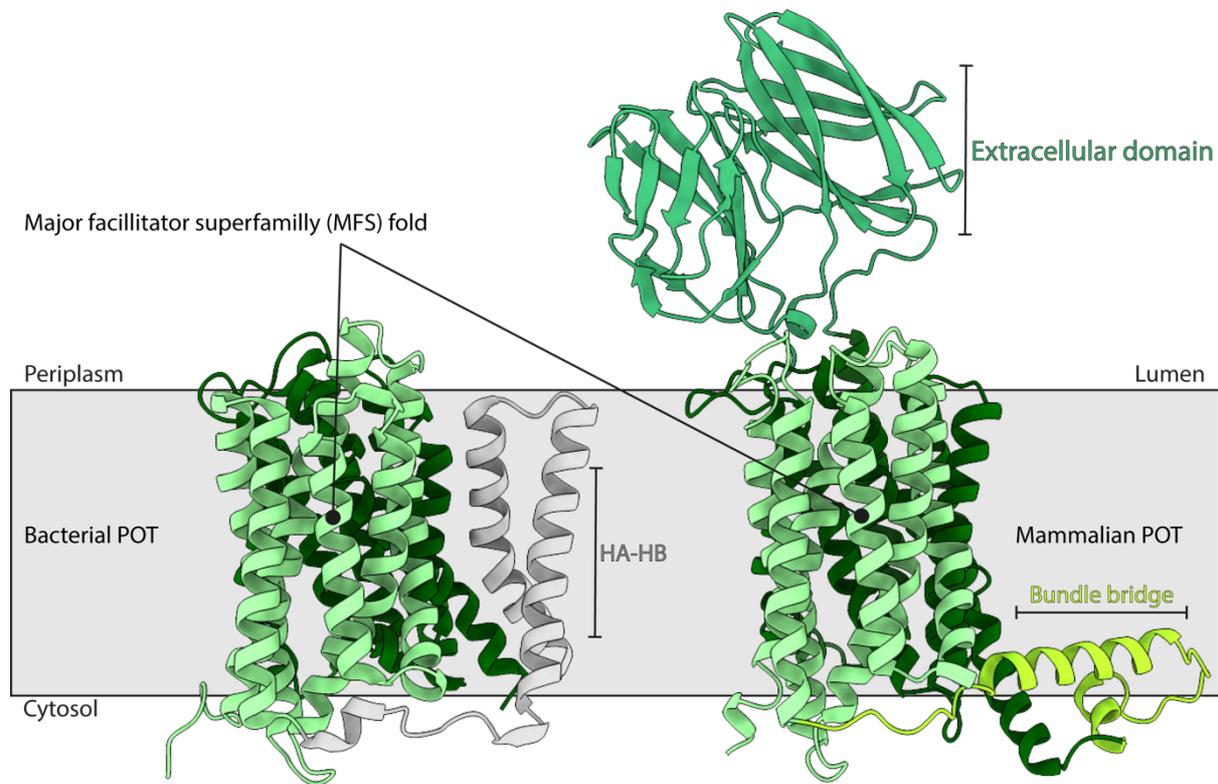


Figure 37. Architectural differences between bacterial and human POTs. While bacterial POTs are composed of 14 transmembrane helices, human homologues contain a transporter unit of 12 transmembrane helices, an extracellular domain and the connecting bundle bridge. This figure was adapted from (Killer *et al.*, 2021).

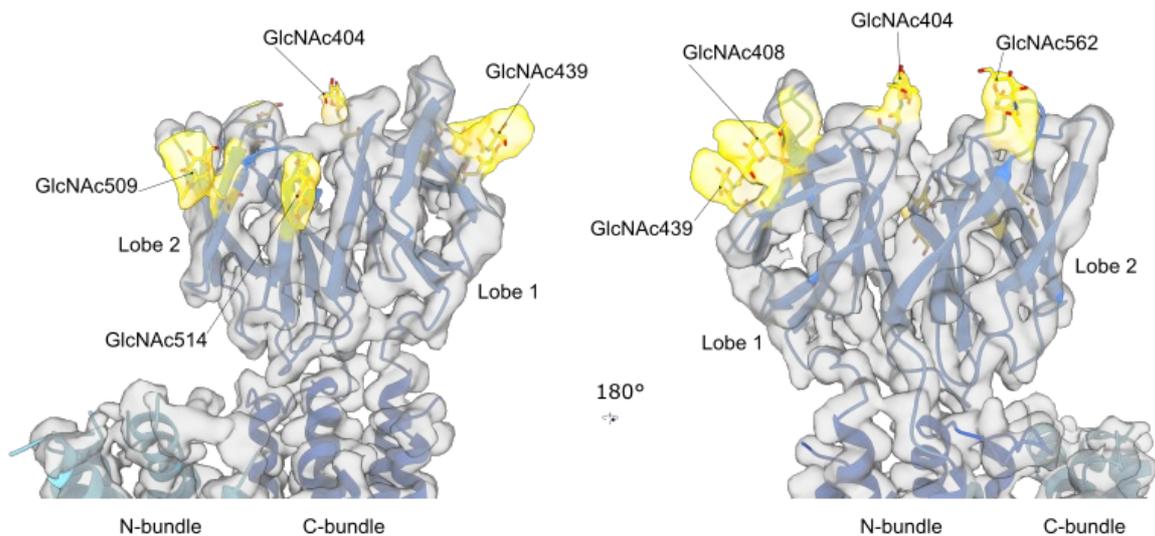


Figure 38. N-Glycans present on the extracellular domain of HsPepT1. Residues N404, N408, N439, N509, N514 and N562 are glycosylated. This figure was adapted from (Killer *et al.*, 2021).

The higher local resolution in HsPepT1-ECD allowed to identify six N-linked glycans (**Figure 38**), five of which were experimentally confirmed to be present on murine PepT1 and were previously shown to contribute to protection from proteolytic degradation (Stelzl *et al.*, 2017). The second striking difference, is the linker between the N- and C-terminal bundles, which folds in two amphipathic helices (i.e., the bundle bridge) in HsPepT1 and HsPepT2. In bacterial POTs however, the bundles are linked by two transmembrane helices, named HA and HB. The role of the bundle bridge, and of the HA-HB helices, were not studied yet.

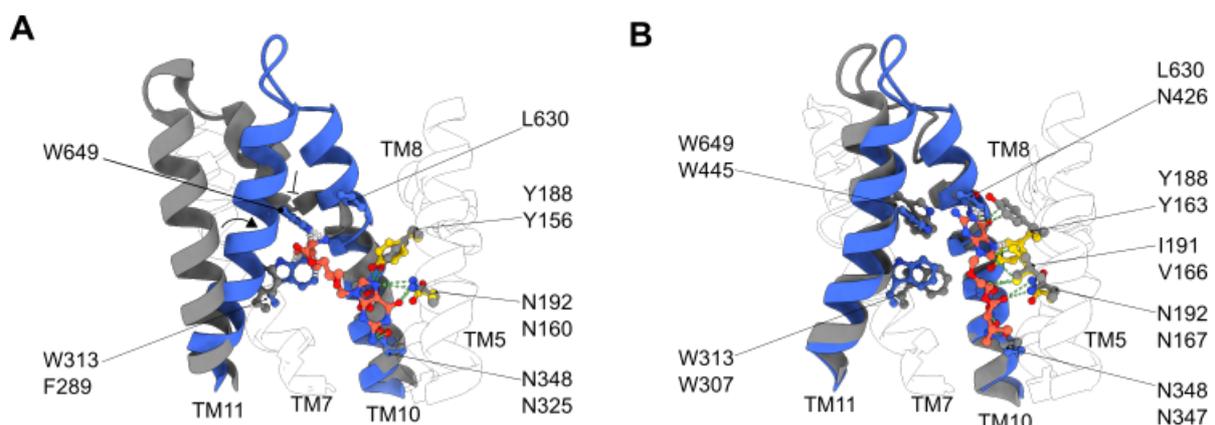


Figure 39. Superposition of HsPepT2 and the prokaryotic POTs DtpA and PepT_{sh} in grey bound to Valganciclovir and Valaciclovir. (A) The intra-helical loop in TM10 leads to a wider opening of TM11, important for Valganciclovir binding in DtpA. Residue F289 in DtpA is also involved in drug binding and is not conserved in HsPepT2 and in PepT_{sh}. (B) The drug Valaciclovir adopts a different position in PepT_{sh}. TM10 and TM11 superimpose well with HsPepT2 and residue W307 (F289 in DtpA) is conserved resulting in a better drug-position

compatibility with the HsPepT2 structure. Favourable contacts are shown as green dotted lines, severe clashes with atoms centre distances below 2 Å are shown as white dashes.

Within the binding site, the residues mediating stabilisation of peptide's termini are conserved, but as illustrated earlier (**Figure 13**), residues involved in side chain accommodation, within P1, P2, or P3, vary. This can be of particular relevance for *in silico* drug screening. For instance, in 2019, bacterial POT structures of DtpA from *E. coli* and PepT_{Sh} from *Staphylococcus hominis*, bound to the transported prodrug molecules Valganciclovir and Valaciclovir became available (Minhas and Newstead, 2019; Ural-Blimke *et al.*, 2019). These structures revealed different binding modes, for the two orthologues. A comparison to the HsPepT2 structure highlights structural differences to bacterial homologues in regions crucial for drug binding. The conformation of TM7, TM10 and TM11 in HsPepT2 would sterically clash with Valganciclovir binding as observed in DtpA, which is mainly caused by the presence of an intra-helical loop in DtpA, absent in HsPepT2 and PepT_{Sh}. A structural alignment of HsPepT2 and PepT_{Sh} illustrates that the position of Valaciclovir described in PepT_{Sh}, is likely to be more compatible in HsPepT2 (**Figure 39**).

6.2.4 Structural basis for peptide transport by PepT1 and PepT2

HsPepT1 and HsPepT2 share high sequence similarities. Therefore, with three distinct OF states of HsPepT1, and one IF state of HsPepT2 (**Figure 40**), the structures could be analysed, in order to shed light on the transport mechanism of POTs.

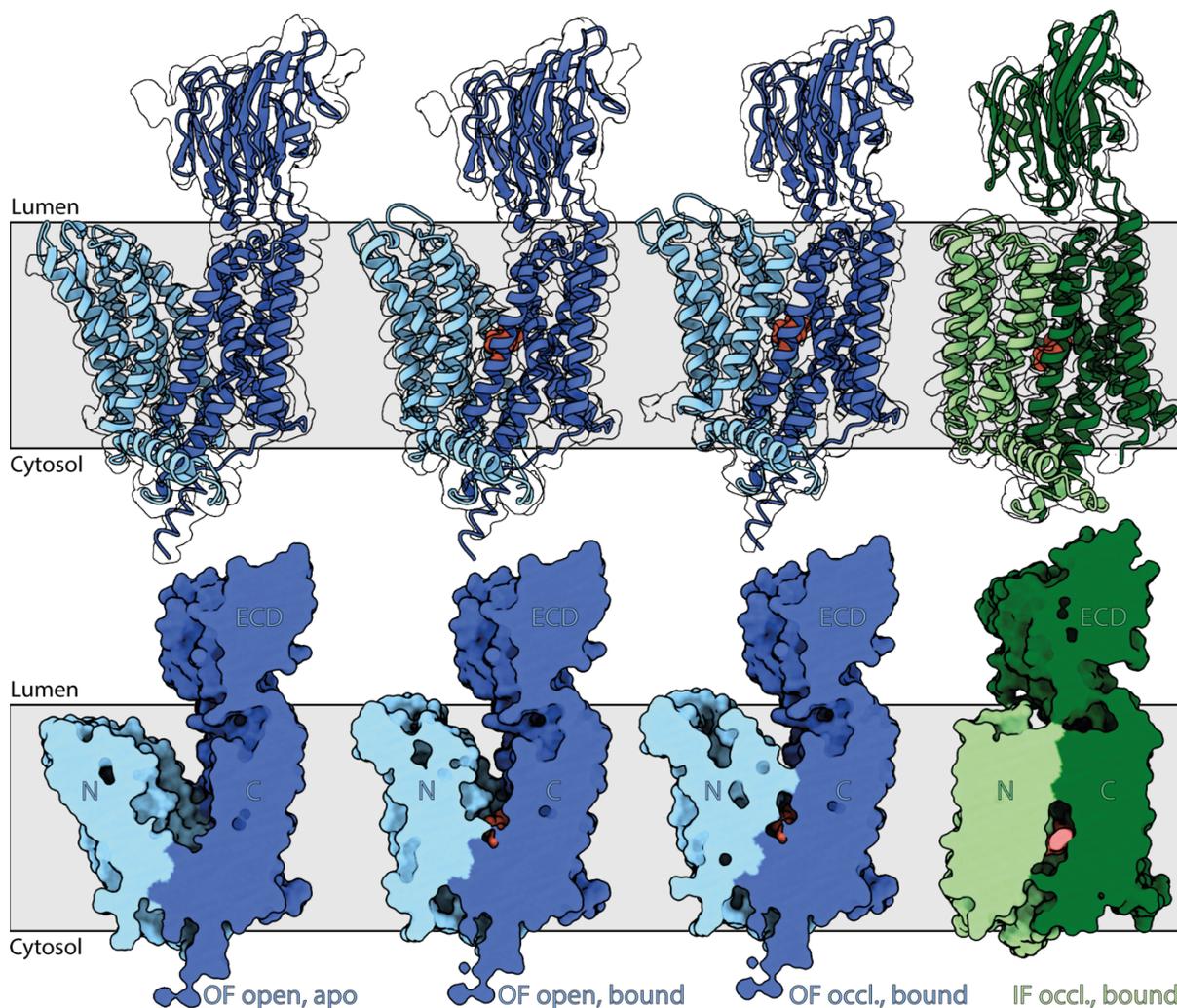


Figure 40. Side by side view of the experimentally determined structures of HsPepT1 and HsPepT2. In the upper panel, atomic models (ribbon representation) are fitted inside their respective cryo-EM density maps. In the lower panel, the models are represented as solvent excluded surfaces. The view cuts through the surface to display the central cavity widely open to the luminal side, and progressively closing, to partially open the binding site to the cytosol.

When aligning the models, on their transmembrane region, the sequence of events relating the captured snapshots appeared clearly, from a progressive closure of the extracellular side of the transporter, to a partial opening of the intracellular side. In this process, the N-bundle is undergoing the larger displacements. This is particularly obvious when comparing the two conformations with the highest root mean square deviation (rmsd) i.e., HsPepT1 OF, open, apo and HsPepT2 IF, occluded, bound (**Figure 41**). There, bending of individual TMs is pronounced in both bundles (TM1, TM2, TM4, TM7 and TM11), and large rocking motion is conspicuous for the whole N-bundle. In contrast, the C-bundle remains anchored within the plane of the membrane.

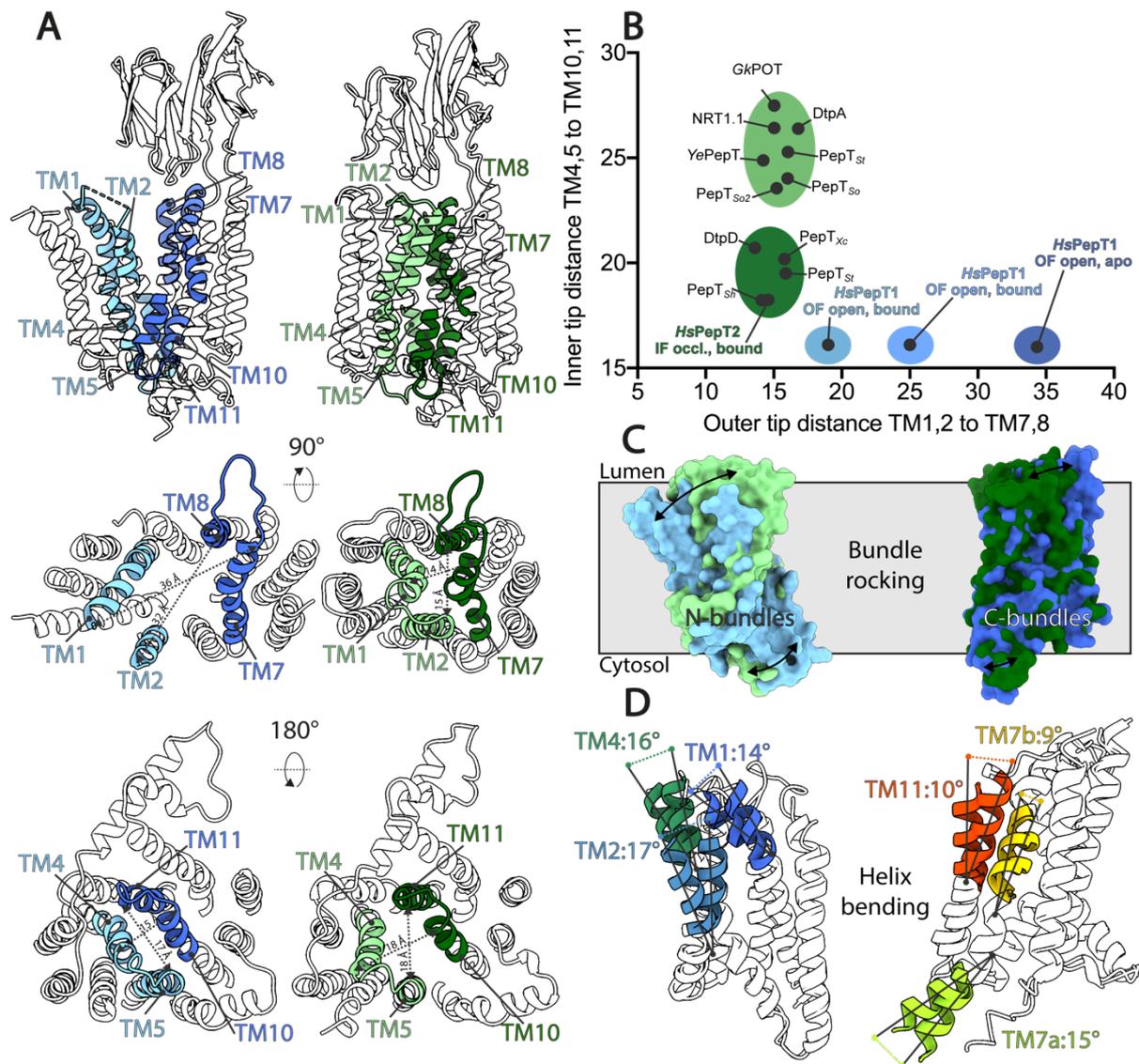


Figure 41. Structural comparison between the outward and inward facing states observed in apo HsPepT1 and substrate bound HsPepT2. (A) Opening and closing of the substrate binding site to the extracellular and intracellular milieu observed in HsPepT1 (blue) and HsPepT2 (green). (B) The distances between C- α atoms of the relevant pairs of helix tips from all bacterial POTs determined by X-ray crystallography were measured and compared to the human transporters. (C) Rocking motions of the N-bundle (HsPepT1: light blue, HsPepT2: light green) and C-bundle (HsPepT1: dark blue, HsPepT2: dark green) after structural alignment of both transporter units. (D) Bending of transmembrane helices with measured tilt angles observed in the N-bundle (left) and C-bundle (right) between HsPepT1 and HsPepT2. This figure was adapted from (Killer *et al.*, 2021).

Next, all transitions were further characterized. As mentioned earlier, most residues are conserved between HsPepT1 and HsPepT2. Therefore, for simplicity, the residue numbering in this section, will refer to HsPepT1, followed by the corresponding residue number in

HsPepT2, inside brackets. For instance, E585[622] corresponds to the glutamate residue numbered 585 in HsPepT1 and corresponding to the position 622 in the HsPepT2 sequence.

Starting outward-facing, open, apo (OF open, apo) state of HsPepT1 (**Figure 42 A, Figure 43**). In this conformation, the central cavity of the transporter is largely opened to the extracellular space, while tightly sealed from the cytoplasm. Sealing, is stabilized by two salt bridges (R159[180] in TM5 with E604[631] in TM10; and R161[182] in TM5 with D341[360] in TM8. This pair of salt bridges will be named “inner clamp”.

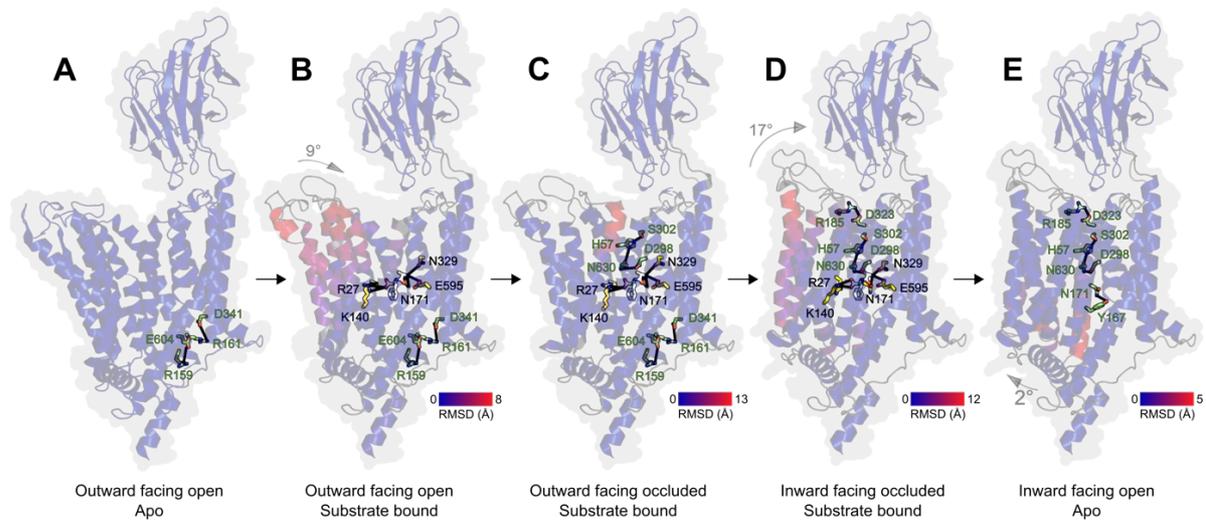


Figure 42. Mechanism for substrate recognition and transport in human POTs based on the presented structures. (A) In the first step of the transport cycle, the transporter is in an outward facing open state stabilized by two salt bridges between R159-E604 and R161-D341. (B) Upon peptide binding – accommodated in the charged central cavity via its N-terminus by N171, N329, E595 and optionally its C-terminus by R27 and K140 – the N-bundle helices follow bending and rigid body motions resulting in tightening of the central cavity. (C) Further bending of TM2 allows the interaction of H57, S302, N630, and D298 as a crucial step before (D) total sealing of the extracellular side and switching to the inward facing occluded state stabilized by the salt bridge R185-D323. (E) Finally, opening of the cytosolic side is achieved by TM4 and TM5 moving away from TM10 and TM11 resulting in the loss of the crucial interaction between the transporter and the peptide termini, allowing substrate release to the cytoplasm. The structures shown in (A), (B), (C), represent models of HsPepT1 based on the three different cryo-EM maps presented in this article. Missing loops have been added. The structure in (D) is derived from the experimental HsPepT2 structure. (E) corresponds to the AlphaFold2 structure prediction, available in the EMBL-EBI AlphaFold database. Numbering of residues illustrated in this model follow the HsPepT1 nomenclature. Conformational changes along the reaction cycle are coloured according to the RMSD between different state structures. This figure was directly taken from (Killer *et al.*, 2021)

Moving to the second OF, open, bound, state of HsPepT1 (**Figure 42 B, Figure 43**). The inner clamp remains in place, but here, the central cavity contracts around the bound dipeptide. Like in the DtpB, Ala-Phe is held from its N- and C-termini via electrostatic interactions with N171[192], N329[348], E595[622], and R27[57]. The salt bridge with K140[161], however is not present, as the second transmembrane helix is arranged differently in this state. The methyl group of the alanine residue (A1*) fits largely in the space formed by P1, and the phenyl group of the phenylalanine residue (F2*), fits more tightly in the hydrophobic groove of P2 (Y94, W313, L316, W649, L650, and I653). Contraction of the binding site occurs mainly by rigid motion of the N-bundle, and additional helix bending, while the C-bundle remains in place. This transition is likely driven by the substrate binding event.

Next, in the OF, occluded, bound state (**Figure 42 C**), a large tilt of TM2 (34°) occurs, which occludes the central cavity from the extracellular space. TM2 now forms a lid above the binding site, and is stabilized in this position by a network of hydrogen bonds between H57[87], and S302[321], N630[657], and D298[342]. Protonation of H57[87] is required for these interactions to occur, and is likely powered by the PMF. The inner clamp remains established, and sustains the sealing from the intracellular space. The resolution of this reconstruction does not allow to verify the pose of the peptide. It can be speculated that Ala-Phe remains coordinated in a similar fashion as in the preceding state, or as in the following.

In the IF, occluded, bound state, of HsPepT2 (**Figure 42 D, Figure 43**), TM2 remains in the same position as in the OF, occluded, bound state. The hydrogen bond network around H57[87] is maintained with the same set of conserved residues S302[321], N630[657], and D298[342]. Here, Ala-Phe remains held by its termini, and the salt bridge with K140[161] is established. However, the inner clamp is disrupted, possibly by protonation of the two acidic residues E604[631] and D341[360], and the N-bundle undergoes a large rocking motion of 17°, to close up the extracellular side of the central cavity, while opening the intracellular side. The binding site is now tightly sealed from the luminal solvent, and this is fastened by a new salt bridge interaction, this time called “outer clamp”, between R185[206] in TM5 with D323[342] in TM8. A prerequisite for the outer clamp to be established, is the deprotonation of D323[342].

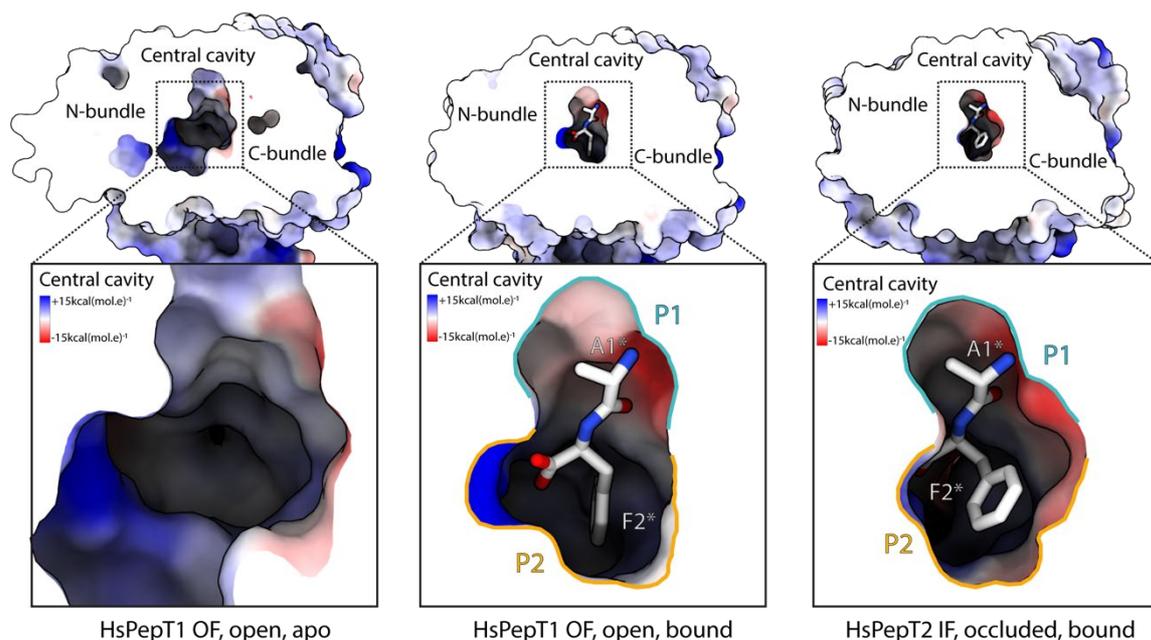


Figure 43. Development of the central cavity upon substrate binding and its transition to the inward facing state coloured by the electrostatic potential. (A) In the substrate free, outward facing conformation the central cavity has a dipole character and is widely open to the outside and can accommodate substrate of various chemical composition and molecular weights. (B) Upon binding of Ala-Phe, the N-bundle rearrangements lead to a tightening of the central cavity to fit the peptide and hooks it by its charged termini. (C) Upon switching to the inward facing state, the bi-polarity of the cavity is maintained to keep the substrate bound prior release to the cytoplasm. This figure was adapted from (Killer *et al.*, 2021).

Finally, the structure predictions of HsPepT1 and HsPepT2, by the Alphafold2 algorithm are naturally devoid of substrate, and are un-surprisingly, in IF states, as the POT database is composed exclusively of bacterial homologues adopting this conformation. The predicted IF open, apo, HsPepT1, and IF open, apo, HsPepT2, show a wider opening of the cytosolic side, through tilting of TM1, TM4, TM5, and TM10. In turn, several residues are out of distance for electrostatic stabilization of the dipeptide, including N171[192], R27[57], and K140[161] (**Figure 42 E**). Further destabilisation such as the protonation of E595[622], could lead to release of the substrate inside the cell. In order to enter a new transport cycle, H57[87], and E[595]622 would need to be deprotonated, as well as the acidic residues of the inner clamp E604[631] and D341[360], while the outer clamp would need to be disrupted by protonation of D323[342].

6.2.5 Discussion

Homologous structure of HsPepT2 (RnPepT2, rat) (Parker *et al.*, 2021) and of HsPepT1 (EcPepT1, horse) (Shen *et al.*, 2022) were determined also by other groups, using SPA cryo-

EM, in 2021 and 2022 respectively. For EcPepT1 structure determination, the protein was reconstituted in nanodiscs, and like in our studies, did not make use of fiducial markers. There, EcPepT1 was captured in the IF state, in the absence of a substrate. The structural similarity with the prediction from AlphaFold2 is striking (RMSD of 0.9 Å in the transporter domain). In comparison with HsPepT2 IF, partially occluded, bound structure, the positions of TM1, TM4, TM5, and TM10 are tilted in the EcPepT1 IF open, apo structure. As a consequence, the residues N171[192], R27[57], and K140[161], are out of distance for electrostatic stabilization of the dipeptide. This supports our overall interpretation of the transport cycle, using the AlphaFold2 predictions of the IF open, apo state, as complement. For RnPepT2, an inhibitory nanobody was first raised against the transporter, denuded of its ECD. Full length RnPepT2, in complex with the selected nanobody, was then reconstructed to an estimated resolution of 3.5 Å. In this work, the authors reported that it was crucial to immobilise the ECD in relation to the TMD (transmembrane domain), in order to obtain a 3D reconstruction. Principal component analysis (PCA) on the atomic model also suggested a high level of flexibility of the ECD. Although we observed some level of flexibility within the HsPepT2 and HsPepT1 datasets, the principal components explaining the variance between the particles indicate smaller movements of the ECD. The stretched conformation of the ECD proposed by Parker *et al.*, is stabilised by the presence of the nanobody, sandwiched between the N-bundle and the ECD. Of all the conformations we captured, the TMD of RnPepT2, resembles the TMD of HsPepT1 OF, open, bound. We note that out of 16,545,185 particles, 298,562 (1.8 %) made it to the final reconstruction of RnPepT2 in complex with the nanobody. 454,149 out of 4,388,314 (10.3%) were used in the reconstruction of HsPepT2. 199,987 of 2,091,726 (9.6%) for HsPepT1 in the apo state. 573,833 out of 6,046,602 (9.5 %) for HsPepT1 structures bound to Ala-Phe. And 438,001 out of 3,437,234 (12.7%) for EcPepT1. Most particles are discarded, because their position in the meniscus (thicker region, contact with carbon, air-water interface), makes their image improper to contribute to high resolution reconstructions. Yet, in this ‘discarding’ process, we also overlook the lesser stable states, coexisting within the whole conformation ensemble present on the grid. Therefore, the different conformations that were reported so far, are likely to coexist to some degree. Only, in a given experimental setting, certain states are more stable, dominant, and can be visualised. From all these captured conformations, it appears clear, that transport is largely driven by motions within the N-bundle, while the C-bundle remains rather anchored, as a scaffold. This seems energetically favourable, as the C-bundle carries the extracellular domain of PepT1 and PepT2. So far, other MFS

transporters (devoid of ECD) had shown a higher level of flexibility within their C-bundle (Guettou *et al.*, 2013; Lyons *et al.*, 2014; Quistgaard *et al.*, 2016b; Quistgaard, Martinez Molledo and Löw, 2017; Drew *et al.*, 2021; Wang *et al.*, 2021). Determining the structures of PHT1 and PHT2, preferably in the OF state, could help to find out whether this peculiar transport mechanism is a feature of the whole SLC15 family. The movements reported between the different structures require the proton motive force. As stated, protonation of several residues is likely to be crucial to transition from one conformation to another, but we cannot directly assess these events with the current methodology. For example, neutron diffraction, or MX and SPA cryo-EM at atomic resolution, could help to identify the protonation state of such residues, but seems extremely challenging for such targets. Another important limitation to consider regarding our work, is the artificiality of the environment in which we imaged PepT1 and PepT2. Proton gradients, membrane potential, curvature, lipid composition, viscosity; are all relevant factors, impacting the energy landscape. To this day, it appears rather difficult to measure useful and perceptive data on such small transporters, within a cellular environment. Complexity has to be decreased, to collect intelligible signals. A compromise, to visualise POTs in a more native environment, would be to prepare liposomes, or vesicles, enriched with the transporter, and induce chemical gradients before vitrification. This approach was shown successful, on the large AcrB homotrimer of 340 kDa from *E. coli* (Yao, Fan and Yan, 2020). Overall, this study, together with the work of others (Parker *et al.*, 2021; Shen *et al.*, 2022), established initial methodological guidelines to obtain 3D reconstructions of PepT1 and PepT2, allowing to visualize side chains and ligands to a resolution ranging between 3 and 4 Å. Several snapshots of the transport cycle could be captured, providing unprecedented insight into the molecular determinants of peptide translocation, across epithelial membranes. It seems like the use of fiducials is not crucial for SPA cryo-EM on PepT1 and PepT2 transporters. Their extracellular domains suffice to drive image alignment to a similar, if not better, precision than when bound to a nanobody. This opens new possibilities for SBDD studies, aiming at using PepT1 and PepT2 as targeted delivery systems. The promiscuity of PepT1 and PepT1 shuttle systems is far from being understood, which limits *in silico* screening accuracy, and effectiveness. Open access sets of structures, bound to chemically diverse ligands, could provide useful information, to guide computational predictions.

6.3 Structure determination of PHT1

6.3.1 SPA cryo-EM structure of PHT1 in complex with a synthetic nanobody

Chicken PHT1 (GgPHT1, SLC15A4), sharing high sequence similarity with its human orthologue, could previously be cloned, expressed and purified in our lab by Dr. Tania Custodio. I initially imaged the transporter extracted in DDM-CHS, but the projections failed to be precisely aligned due to the lack of structured elements outside of the micelle. Synthetic nanobodies (sybodies) were screened by Dr. Custodio following established protocols (Zimmermann *et al.*, 2018, 2020) against GgPHT1, and one binder named Sb27 was selected for SPA cryo-EM. Data acquisition on the GgPHT1-Sb27 complex, immediately led to better defined secondary structures in 2D averaged images, compared to the transporter alone (**Figure 44 A, B**).

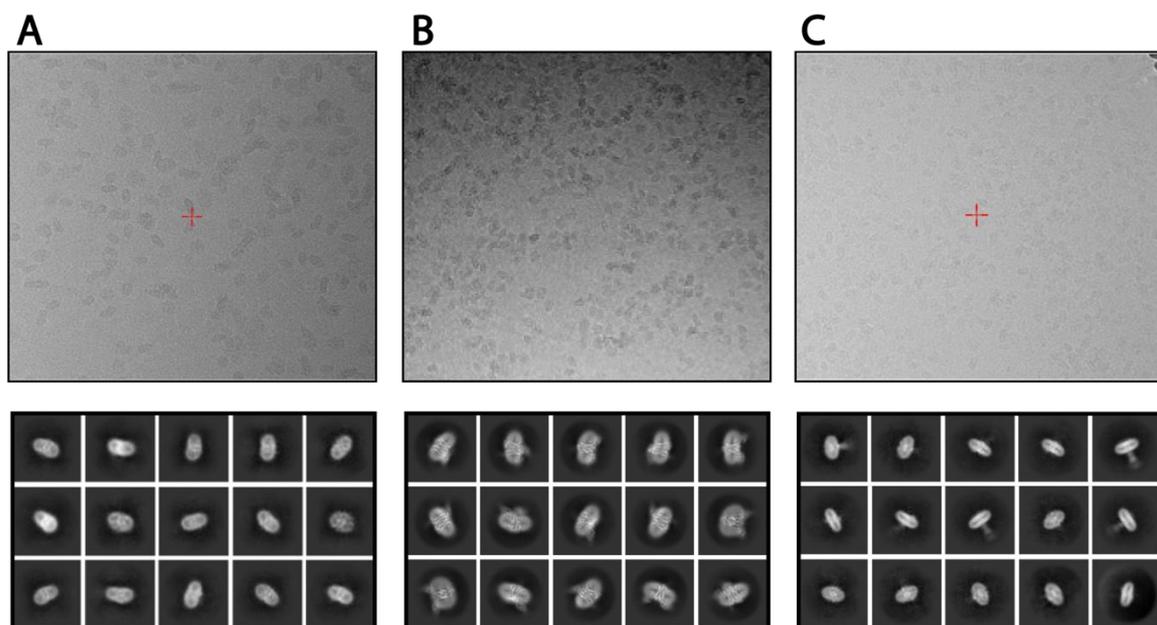


Figure 44. Side by side comparison of PHT1 samples imaged by cryo-EM. (A) PHT1 transporter alone. (B) PHT1 in complex with sybody 27 (Sb27). (C) PHT1 fused to a BRIL domain, bound to a α -BRIL-Fab, and TASL1-13 peptide.

PHT1 structure follows the MFS fold, with TM1-TM6 forming the N-bundle and TM7-TM12 forming the C-bundle. The bundle bridge connects the two bundles, similarly as in PepT1 and PepT2. PHT1 does not possess an immunoglobulin-like ECD between TM9 and TM10, but rather a beta sheet of two strands (**Figure 47B**)

Surprisingly, two extra-density features could be discerned at the periphery of the micelles. After sorting the projections contributing to high resolution reconstructions, a map resolved at 3.5 Å was obtained and confirmed the presence of two copies of Sb27 binding to two different epitopes on the cytosolic side of the GgPHT1 (**Figure 45**, **Figure 46**). The transporter was in a OF open, apo state, stabilised by the conserved salt bridges between the N and C bundles (i.e., inner clamp): R200-E482 (R159-E604 in HsPepT1). Additional inter-bundle polar interactions further stabilised this conformation; notably, D189-Q493, K192-Y485 and N204-K392 (**Figure 47**). These residues are not conserved in PepT1 and PepT2, but are present in both PHT1 and PHT2 transporters.

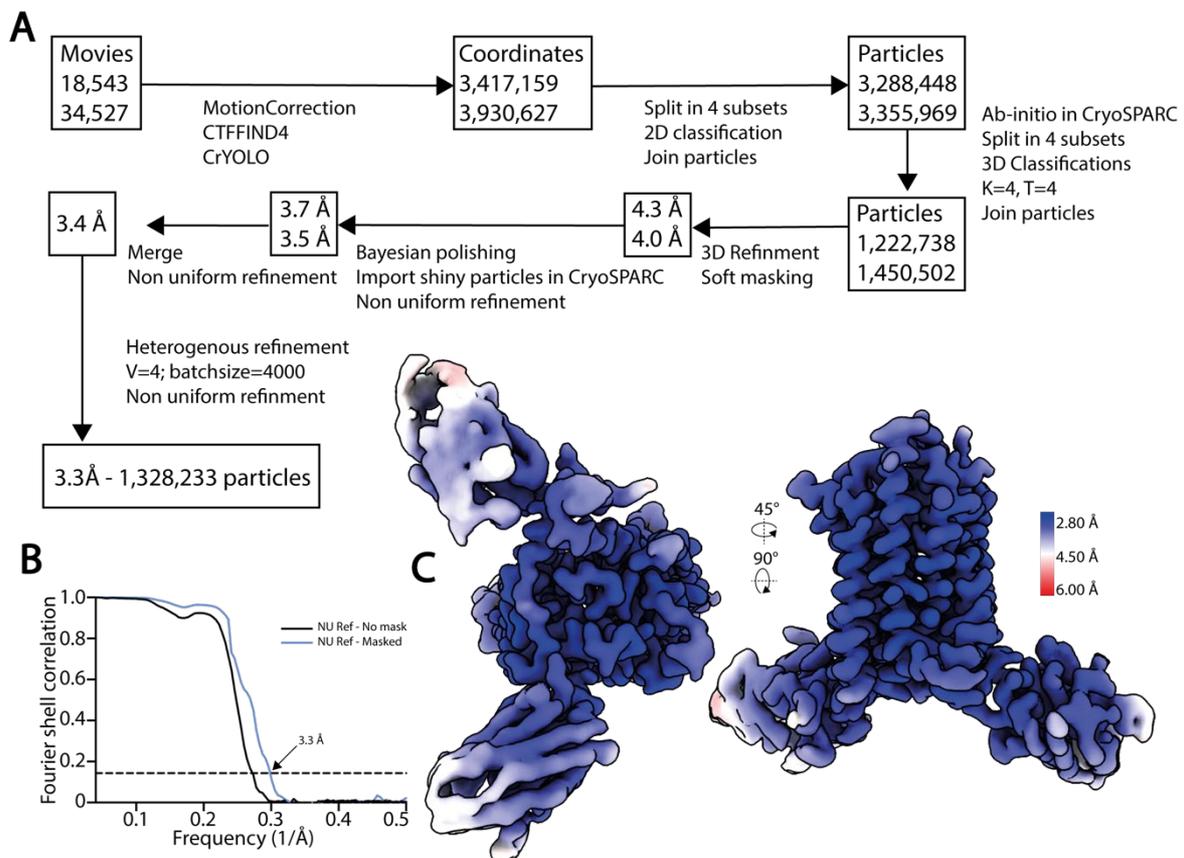


Figure 45. High resolution structure determination of PHT1-Sb27. (A) Summary of the image analysis workflow. (B) The Fourier transforms over different shells on frequency space, of two independent volumes (half maps) were compared (FSC) and plotted as a function of spatial frequency, to estimate the overall resolution using the 0.143 cutoff threshold. (C) 3D volume, coloured by local resolution, estimated in cryoSPARCv3 using the 0.5 as FSC threshold.

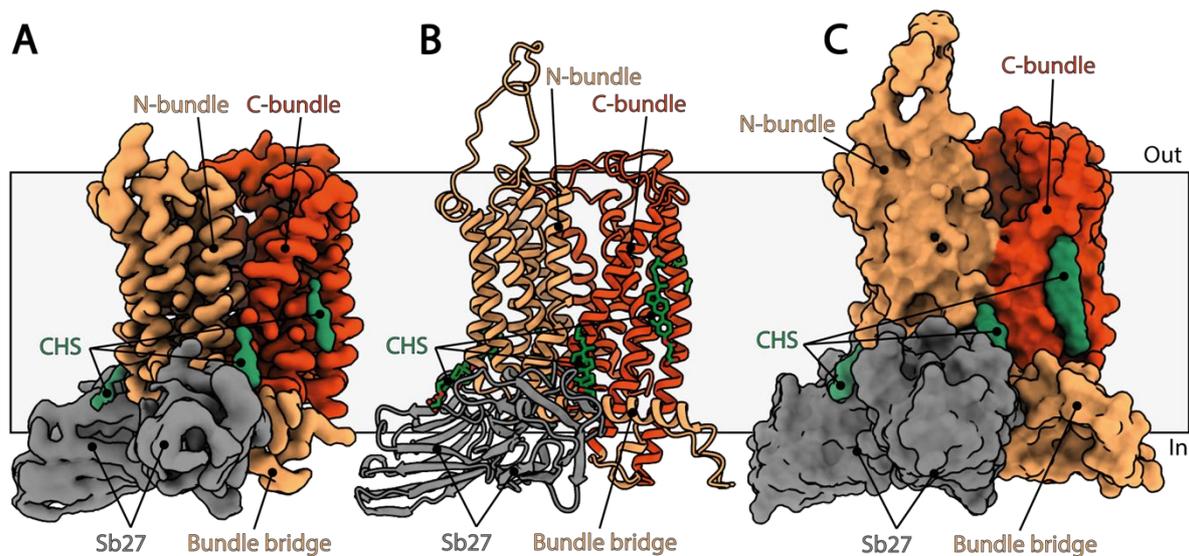


Figure 46. Cryo-EM structure of PHT1-Sb27. (A) 3D reconstruction. (B) Atomic model displayed as ribbon representation. (C) Atomic model displayed as surface excluded solvent representation. The different structural elements are labelled.

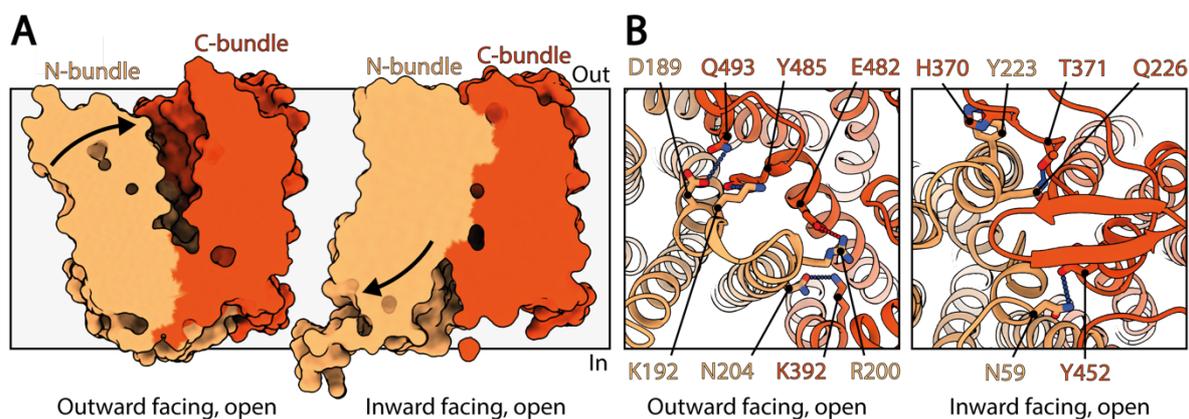


Figure 47. Stabilisation of the OF, open state in PHT1. (A) Experimental (left) and AlphaFold2 predicted (right) PHT1 are represented as solvent excluded surfaces. The view cuts through the surface to display the central cavity open to the luminal side (OF open) in the experimental structure, while opened to the cytosol in the prediction (IF open). (B) The electrostatic interactions stabilising the OF open, and IF open states are shown. Salt bridges are indicated as red dashed lines. Polar interactions are indicated as blue dashed lines.

Here again, without surprise, AlphaFold2 predicted an IF state for PHT1. The outer clamp does not involve a salt bridge in PHT1. Instead, other residues are predicted to mediate polar interactions on the luminal side of the transporter, between the N- and C-terminal bundles. I.e., Y223 (TM5) with H370 (TM8); Q226 (TM5) with T371 (TM8); and N59 (TM1) with Y452 (located on the extracellular beta sheet folding between TM9 and TM10) (**Figure 47**). As observed between the various states in PepT1 and PepT2, the C-bundle remains rigid, while the N-bundle undergoes large conformational changes to switch from the OF to the predicted

IF conformation. The outer tips of TM1, TM2, and TM4 are slightly less far apart from the C-bundle, than they are in the OF open states of PepT1. Nonetheless, the central cavity of the transporter is still largely accessible to the endo-lysosomal lumen (**Figure 46**, **Figure 47**, **Figure 48**).

Within the binding site, all critical residues present in prototypical POTs are conserved, except for the residue D381 (N329 in HsPepT1) (**Figure 13**, **Figure 48**). N329 contributes to stabilise the N-terminus of peptides, and was shown to be critical for transport in other prototypical POTs. The mutation to an aspartate was not studied so far, but could lead to further stabilisation of peptides from their N-terminus. In the case of single histidine transport for example, it might compensate for a poorer interaction of the carboxylic group with the distant R44 (R27 in HsPepT1) and K180 (K140 in HsPepT1) residues located in the N-bundle. In this case, the imidazole ring might also interact with D381. This is however hypothetical, and should be tested either biochemically, or with a high resolution histidine bound structure.

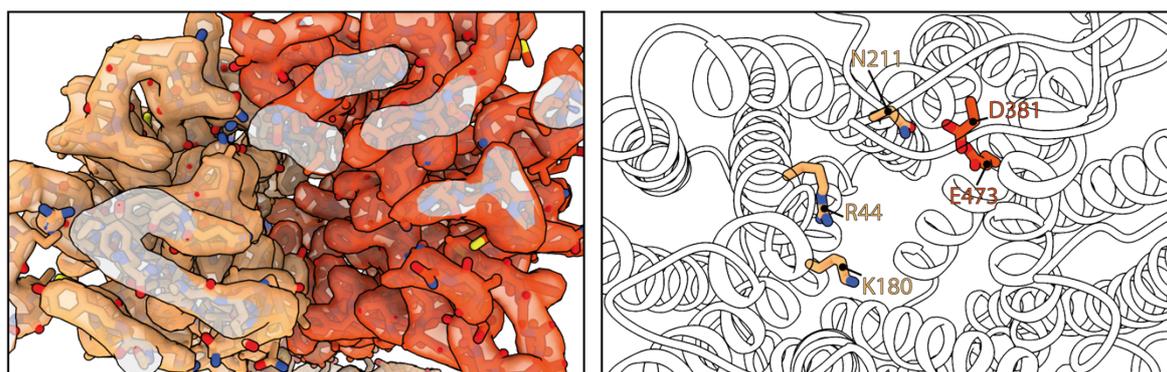


Figure 48. Overview of the binding site of PHT1. Left panel: the cryo-EM density map is displayed around the modelled atoms. Right panel: the conserved residues in prototypical POTs are shown. Notably, the triad of residues coordinating the N-termini of peptides (two asparagine residues, and one glutamate) is constituted by an aspartate, an asparagine, and a glutamate residue in PHT1.

6.3.2 Structural basis for Sb27 binding to PHT1

The binding of Sb27 to two distinct surface patches of the transporter, was quite unexpected. Both epitopes are located on the cytosolic side of the transporter. In both interfaces, a CHS molecule cements the transporter and the sybody together (**Figure 49**). The cholesterol moiety is sandwiched between hydrophobic regions of the transporter (either TM1, TM4, and TM6 for interface 1; or the bundle bridge for interface 2), and the CDR3 loop of Sb27 (W103). The succinic acid moiety penetrates a pocket formed by CDR1 (W32, Y33), and CDR2 (S52, V54, T55) loops, and binds through polar contacts. In interface 2, E29, present in the CDR1 loop of

Sb27, is in close vicinity to R316, at the bundle bridge of GgPHT1. However, the poor quality of the density in this region doesn't verify an interaction. In interface 1, E29 clearly forms a salt bridge with R194 at the cytosolic tip of TM4 (R186 in HsPHT1), and two additional CHS molecules further stabilise the docking of Sb27 to this site. This increased stability translates in a stronger signal in this region, and a higher local resolution (**Figure 45 C**). The residues involved in CHS and Sb27 binding to PHT1 are conserved between the chicken and human homologue (**Figure 13**).

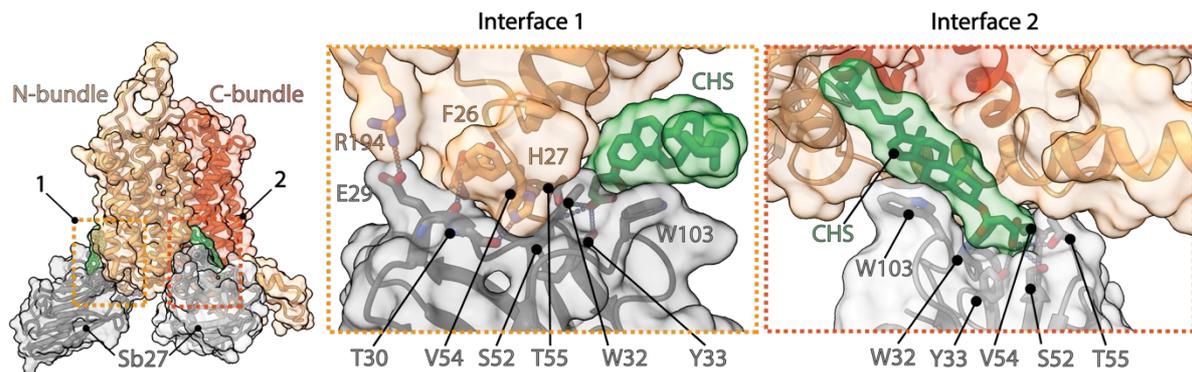


Figure 49. Binding interface between PHT1 and Sb27. Two Sb27 molecules bind to two distinct epitopes of PHT1. These binding interfaces both involve a CHS lipid molecule gluing the sybody and the membrane protein together. In addition, the most stable association (middle panel), involves a salt bridge between E29 in Sb27 and R194 in PHT1.

6.3.3 Putative binding of TASL to PHT1

With Sb27, as an effective tool to obtain high resolution reconstructions on GgPHT1, we then sought to study the interaction between the latter and TASL. TASL is an intrinsically disordered (IDP) protein, and its first 15 residues were identified to be important for binding to PHT1 (Heinz *et al.*, 2020). Since the expression and purification of TASL turned out challenging, a synthetic peptide of the N-terminus of TASL (13 residues length) was purchased and tested for binding to GgPHT1. Strong stabilisation and binding was measured, so a sample containing GgPHT1, TASL₁₋₁₃ and Sb27 was vitrified and imaged. The reconstructed volume reached a resolution of 3.5 Å, and was identical to the GgPHT1-Sb27 map.. TASL₁₋₁₃ being clearly not bound, the two datasets (TASL₁₋₁₃ free and in the presence of TASL₁₋₁₃) were merged to further increase the signal to noise ratio in the GgPHT1-Sb27 reconstruction. This led to an improved map, with a nominal resolution of 3.3 Å (**Figure 45, Table 10**). After AlphaFold2 multimer became more established, as a mean to predict quaternary structures of proteins, it seemed plausible, that TASL would bind specifically to the IF conformation of PHT1. Indeed, in the prediction, the first 15 residues of the N-terminal part of TASL folds into

an α helix, and penetrates inside PHT1 cavity formed between the N- and C-bundles, from the cytosolic side (**Figure 50**). The N-terminal amino group of TASL then reaches the strictly conserved E473 residue and forms a salt bridge with it. The fourth residue of TASL (E4) mimics the usual C-terminus of di- and tripeptides, and is involved in a salt bridge with R44. The rest of the helical peptide is braced between the two bundles, through hydrophobic contacts largely, and fewer polar interactions (i.e., R9-N204). *In vitro* and *in silico* data taken together, it seemed likely that Sb27 non-competitively inhibited the binding of TASL to PHT1, by locking the transporter in the OF state. Before trying to screen for potential competitive inhibitors of TASL, it was first important to validate the structural predictions, and therefore to image, the PHT1-TASL complex in the absence of Sb27. To compensate for the absence of Sb27, a similar strategy as previously employed for G-protein coupled receptors (GPCRs), was followed (Miyagi *et al.*, 2020; Mukherjee *et al.*, 2020; Tsutsumi *et al.*, 2020; Zhang *et al.*, 2022). The latter consists in fusing a thermostabilised version of the apocytochrome b562 (also called BRIL) to the target IMP. In addition, a fragment antigen binding (Fab) against the BRIL domain is used to enlarge the complex. Here, the BRIL domain was fused to GgPHT1 in its luminal side, between TM7 and TM8. With SPA cryo-EM, the signal could be detected for the BRIL-Fab, and seemed to improve the quality of the 2D averaged classes, compared to the naked GgPHT1 transporter, imaged previously. Yet, the majority of the particles showed a high level of flexibility between the transporter and the fiducial (**Figure 44 C**). The construct therefore requires further optimisation to allow high resolution 3D reconstruction.

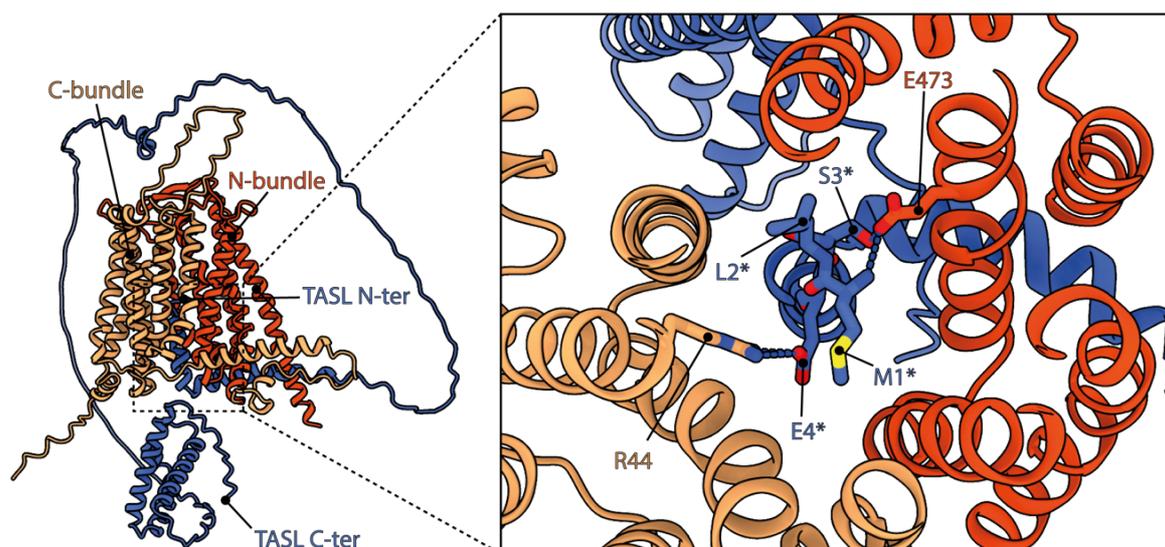


Figure 50. Putative interaction between TASL and PHT1. The AlphaFold2 Multimer quaternary structure prediction between TASL (displayed as blue model), and PHT1 (orange and red) is shown. The first 15 TASL residues (TASL₁₋₁₅) fold into an alpha helix, penetrating inside the central cavity, from the cytosolic side, and is hooked in the binding site by a similar

fashion as regular POT substrates. Importantly, a salt bridge between the N-terminus of TASL (M1*) and the conserved glutamate residue, is reminiscent of the observed interaction between all peptide bound structures reported in POTs. The glutamate side chain of the fourth residue (E4*) is also stabilised by a salt bridge, with R44, usually stabilising the C-terminus of di- and tripeptides.

6.3.4 Discussion

Over the last five years, an increasing number of evidence has indicated the predominant function of PHT1 in innate inflammatory response. On one hand, PHT1 can modify the lysosomal milieu which indirectly modulates inflammation. On the other hand, it is a direct and essential chain-link component of the TLR-TASL-IRF-NF- κ B pathway. (Heinz *et al.*, 2020) have recently shown that preventing the interaction between PHT1 and the protein TASL, blocked this pathway, and subsequently reduced symptoms associated with SLE disease. One way of blocking the interaction between PHT1 and TASL, is to develop competitive inhibitors. The 3.3 Å outward facing (OF) reconstruction we obtained for PHT1 could be a starting point for *in silico* compound screening, but experimental data on a PHT1-TASL complex, is likely to be more informative. The latter would allow to evaluate, the binding mode predicted by the AlphaFold2 Multimer algorithm. If these predictions would turn out to be accurate, the canonical di- and tripeptide binding site, perhaps in combination with the whole cytosolic cavity, could be targeted with small molecules. In addition, if TASL indeed, actually, requires PHT1, to be in an inward facing (IF) conformation, in order to be recruited to the lysosomal membrane, it then opens the route for the development of non-competitive inhibitors, blocking PHT1 in the OF state. Sb27 might constitute an interesting starting point in the design of such therapeutics. Very recently, Dr. Custodio successfully cloned, expressed, and purified the human homologue of PHT1 in DDM-CHS. Binding between Sb27 and HsPHT1 was confirmed *in vitro*. A point of concern for the relevance of Sb27 in SBDD, is that CHS molecules contribute to its binding to PHT1. CHS is a close derivative to cholesterol, but it is not guaranteed that the latter would bind to the same regions of the transporter, in a native context. Therefore, it appears crucial to evaluate the inhibitory effect of Sb27 to TASL binding, in relevant environments (i.e., *in cellulo* or *in vivo*). If so, Sb27, could be further engineered and matured, to increase its affinity to the outward facing conformation of PHT1. Lastly, other nanobodies or sybodies, could be raised against outward facing PHT1, and selected for direct binding to PHT1, independently of lipid molecules. A good target seems to be the short loops between TM4-TM5 and TM10-TM11. The structure presented in this work, shows that TM4, TM5, TM10 and TM11, are pinched together in the OF conformation, and tightly close the transporter, from the cytosolic side. The short loops between TM4-TM5 and TM10-TM11, could therefore be an attractive alternative epitope for OF specific binders.

6.4 Selectivity in atypical POTs

The atypical POT from *E. coli*, DtpC, prefers to transport Ala-Lys dipeptides. I wanted to determine the structure of this transporter, to look at the particularities of its binding pocket, and the possible structural determinants for peptide specificity.

6.4.1 Crystallisation trials on DtpC

Previous work in our group, allowed to obtain nanobodies specifically binding to DtpC (Nb17, Nb26, Nb38) (**Figure 51 A**). Crystallisation conditions were identified for the DtpC-Nb26 complex, solubilised and purified in DDM, using the vapour diffusion method. After several rounds of optimisations, the best crystals diffracted up to 5 Å, but did not improve further (**Figure 51 B**). To promote crystal contacts, DtpC-Nb26 was then purified in short acyl chain detergents, including DM, NM, and NG. As expected, this decreased the thermal stability of DtpC (**Figure 51 C**) but no crystal grew under these conditions. The LCP crystallisation method was then applied to DtpC and to the DtpC-Nb26 complex, with 9:9 MAG lipid. Small crystals, appeared, with an aspect reminiscent of many other small LCP IMP crystals (Ref Caffrey), and were therefore tested as well (**Figure 51 D**). Mesh scans of the harvested LCP bolus did not reveal any diffraction spots, however.

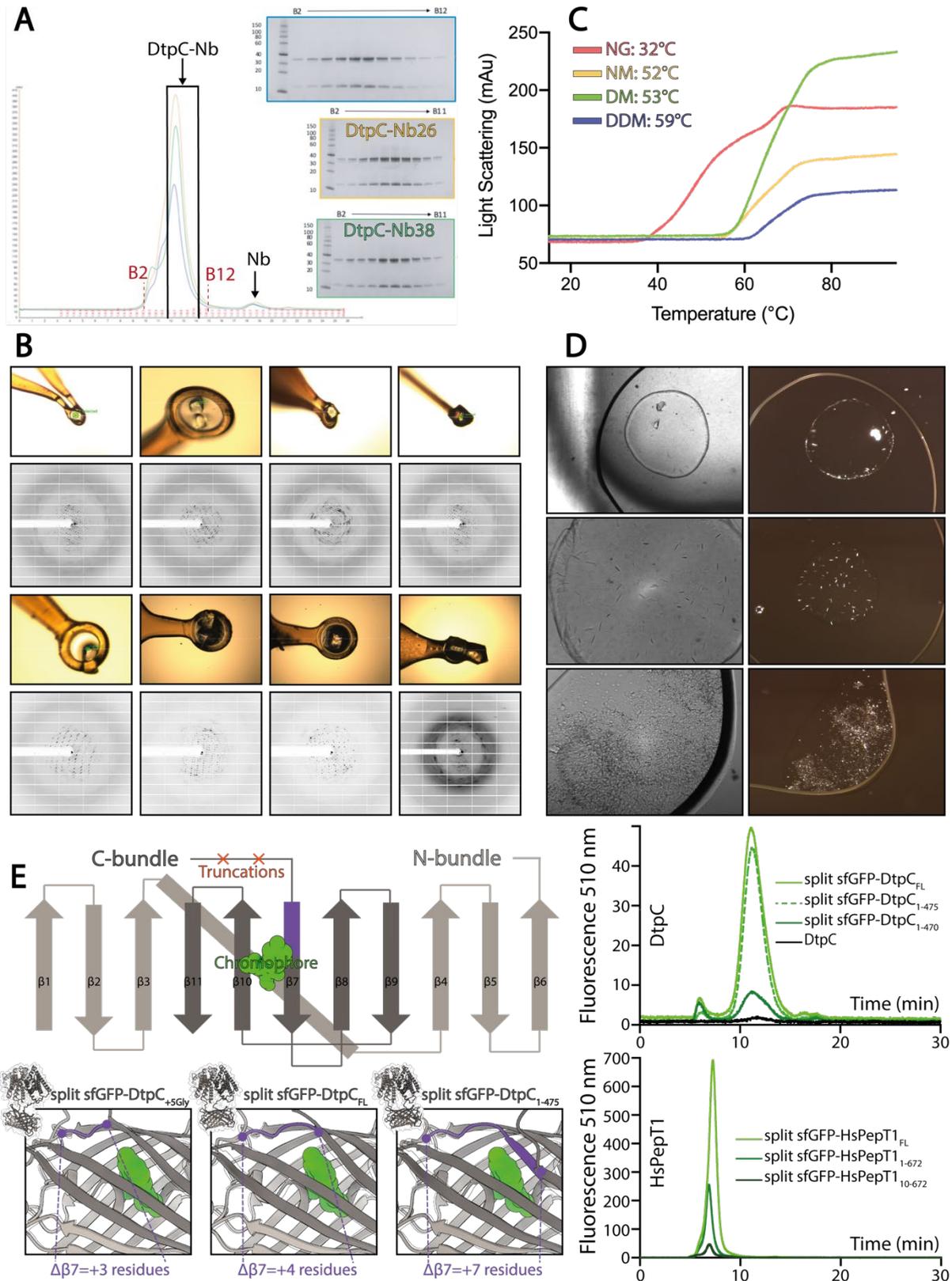


Figure 51. Crystallisation trials on DtpC. (A) DtpC co-eluted with three nanobodies raised in camelids. (B) DtpC-Nb26 was crystallized in many different conditions using the vapor diffusion method, but crystals did not diffract X-rays to higher resolutions than 5 Å. (C) DtpC-Nb26 was purified in short chain detergents, while still yielding folded samples, but failed to

produce crystals. (D) LCP crystals of DtpC-Nb26 were grown, but did not diffract X-rays. (E) The A schematic representation of the split sfGFP-DtpC architecture is shown on the top panel. Below, structure predictions were generated for split-sfGFP-DtpC^{+5Gly}, split-sfGFP-DtpC_{FL}, sfGFP-DtpC₁₋₄₇₅, and overlaid with sfGFP (PDB accession number 2B3P). The dark-violet colouring corresponds to the fraction of $\beta 7$ which is properly folded in sfGFP while unfolded in the restrained chimeric construct. The right panel shows HPLC chromatogram profiles monitoring the fluorescence of the chromophore of split sfGFP in the context of the indicated constructs, using 480 nm as excitation wavelength and recording at 510 nm the emitted light. This figure was adapted from (Killer *et al.*, 2022).

Inspired by the termini-restrain methodology developed for small membrane proteins (constituted of 2, 4, and 6 TMs), (Liu *et al.*, 2020, 2022), the two self-assembling parts of split sfGFP were fused to the termini of DtpC. SfGFP $\beta 1-6$ was introduced at the N-terminus of DtpC, and $\beta 7-11$ on the C-terminus (**Figure 51 E**). This construct is named split sfGFP-DtpC_{FL}. Two additional constructs were also generated in order to minimize the flexibility between the membrane protein and the fiducial: either the last five (split sfGFP-DtpC₁₋₄₇₅ construct), or ten residues (split sfGFP-DtpC₁₋₄₇₀ construct) of DtpC were truncated. Folding and self-complementation was then assessed by measuring the fluorescence of the chromophore on an HPLC system. The three chimeric proteins eluted at similar retention times, and the fluorescence was the lowest for split sfGFP-DtpC₁₋₄₇₀ while highest for split sfGFP-DtpC_{FL}. This trend was also observed with HsPepT1, which was fused in a similar fashion to split sfGFP, with increasing termini deletions (Figure 51E). Anyhow, the decrease of fluorescence seemed insignificant in split sfGFP-DtpC₁₋₄₇₅ in comparison to split sfGFP-DtpC_{FL}, therefore, crystallisation trials were performed with the former. Neither vapor diffusion nor LCP crystallization, produced any crystal on this construct.

6.4.2 Structure determination of DtpC by SPA cryo-EM

Like PHT1, DtpC lacks cytoplasmic or periplasmic features, which, if stable, can be helpful for accurate particle alignment, and high resolution reconstruction. Here, the previously raised nanobody 26 (Nb26), was used as starting point for introducing extra-density outside of the detergent micelle. The size of Nb26 was also increased, by fusing the maltose binding protein (MBP) to its C-terminus, as described previously (Botte *et al.*, 2022). This resulted in a 52 kDa Promacrobody (Mb26). The three following samples were then vitrified and imaged: (i) split sfGFP-DtpC₁₋₄₇₅-Nb26, (ii) DtpC-Mb26, and (iii) DtpC-Nb26 (**Figure 52**). The potential of

these samples to reach highest resolution, was judged by the presence of secondary structure elements, in 2D averaged images (**Figure 52, Figure 53**).

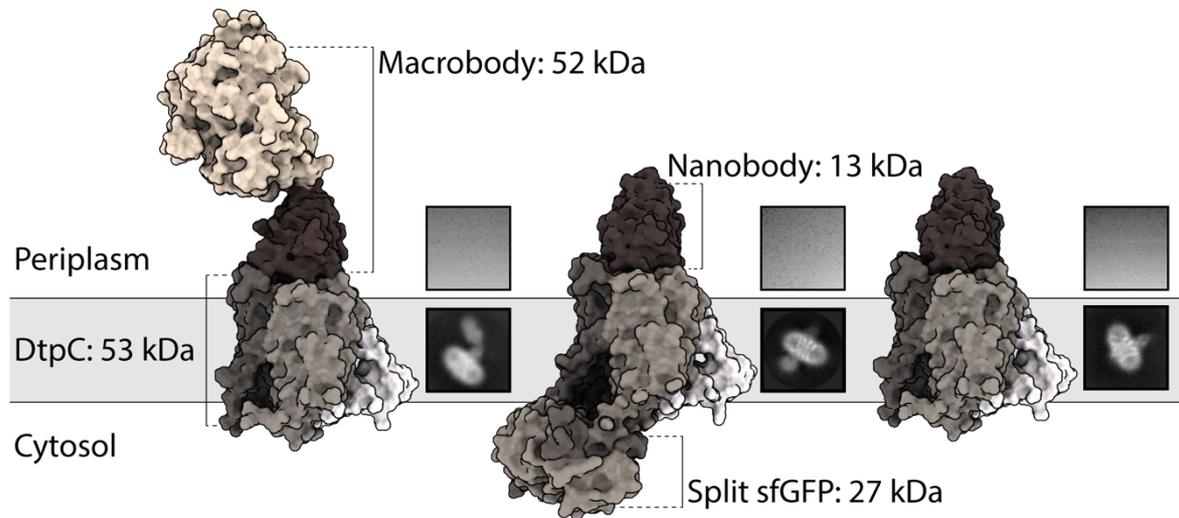


Figure 52. Utilization of different fiducial markers to improve particle alignment and 2D averaging from cryo-EM images. From left to right: DtpC-Mb26, split-sfGFP-DtpC₁₋₄₇₅-Nb26, and DtpC-Nb26 were purified, vitrified on grids and imaged. Single particles were identified, clustered and averaged. The best average from each sample is shown under a representative raw micrograph. This figure was taken from (Killer et al., 2022).

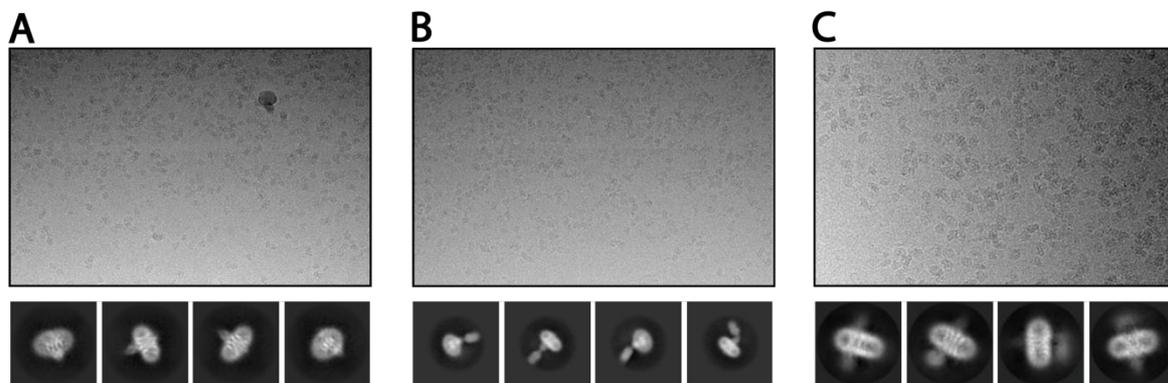


Figure 53. Enlarged representative micrographs and 2D class averages. (A) DtpC- Nb26, (B) DtpC-Mb26, (C) split sfGFP-DtpC₁₋₄₇₅-Nb26. This figure was taken from (Killer et al., 2022).

DtpC-Nb26 yielded 2D classes with sharper and better defined transmembrane helices within the micelle, which is, from experience, an encouraging sign, in subsequent steps of image processing. In addition, DtpC-Nb26 was more stable than split sfGFP-DtpC₁₋₄₇₅-Nb26 and DtpC-Mb26 (Figure 54). After the code of AlphaFold2 became publicly available, predictions on the constructs indicated a destabilisation of the self-assembling split sfGFP, as a function

of termini restrain (**Figure 51 E**). Decreasing the restrain by adding a linker between the termini of DtpC and the fiducial, reversed this effect.

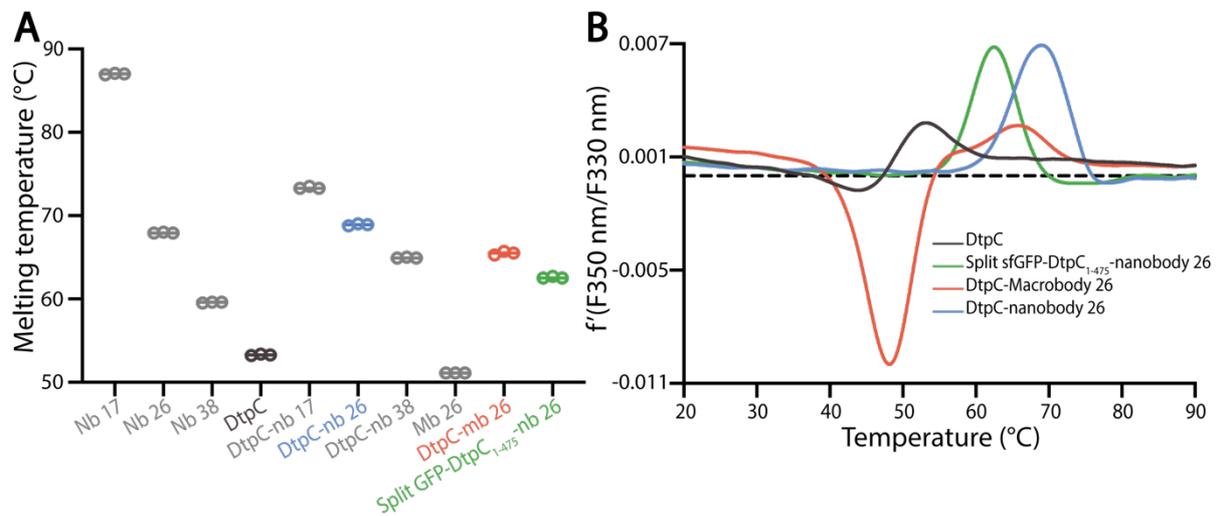


Figure 54. Characterization of the different fiducial markers. (A) The melting temperature of each fiducial and DtpC-fiducial complex was measured by nano-differential scanning fluorimetry (DSF) in triplicate measurements as shown as open circles. The average of the three values is marked by a line. (B) The first derivative of the summarized data in (A) is shown for DtpC and the three imaged samples together with the respective fiducial. This figure was adapted from (Killer et al., 2022).

Additional images were acquired on DtpC-Nb26, and about a third of the picked particles clustered in a population of dimers. This increase in mass, constituted a major advantage in image alignment, and immediately led to better resolved initial volumes, compared to the monomeric population. For the sake of efficacy and computation, the angles and shift relating monomeric projections, were not further calculated. Additional clustering and image correction, allowed to reconstruct the dimer at nominal resolution of 3.0 Å. Restricted angles and shifts refinements, focused on one copy of the transporter, increased the resolution to 2.7 Å (**Figure 55, Figure 56**).

The quality of the reconstructions permitted to build *de novo* inside the maps. In the dimeric assembly, two inverted DtpC molecules, contact each other through polar interactions and hydrophobic contacts, *via* the HA-HB helices (**Figure 57 A**). Similarly as for DtpB-Nb132, Nb26 binds DtpC from the periplasmic side (**Figure 57 B**).

DtpC was captured in an IF open state. This conformation is stabilised by the outer clamp formed by a salt bridge between D43 (TM2, N-bundle) and R294 (TM7, C-bundle), and other

polar interactions between H37 (TM1) and D293 (TM7), R28 (TM1) and N421 (TM11) (Figure 58 A).

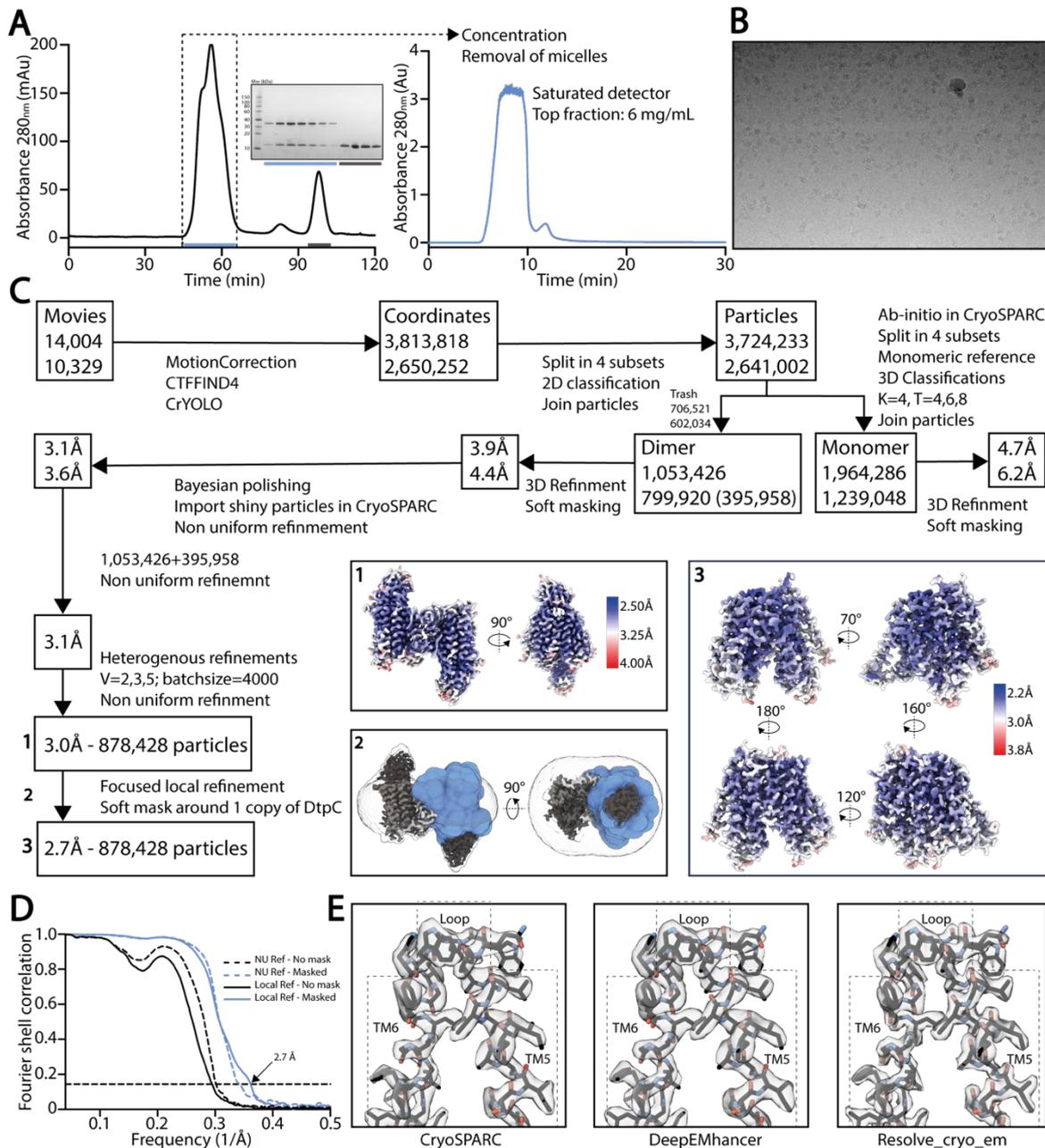


Figure 55. High resolution structure determination of DtpC-Nb26. (A) Gel filtration was performed on a preparative column (left) before concentrating the sample to 60 mg/ml and rerunning it on an analytic column on an HPLC system (right), in order to obtain a highly concentrated sample, free of empty detergent micelles. Peak shape already indicates a mixture of different oligomeric species. (B) Representative raw micrograph of the acquired dataset. The estimated applied defocus is $-1.5 \mu\text{m}$. (C) Summary of the image analysis. The angular assignments from the dimeric reconstruction were used as prior to perform a local focused refinement with reduced angular and translational searches on the masked region illustrated in

blue. (D) The Fourier transforms over different shells on frequency space, of two independent volumes (half maps) were compared (FSC) and plotted as a function of spatial frequency, to estimate the overall resolution using the 0.143 cutoff threshold. (E) The two half maps were used as inputs to assess various post-processing strategies. This figure was directly taken from (Killer et al., 2022).

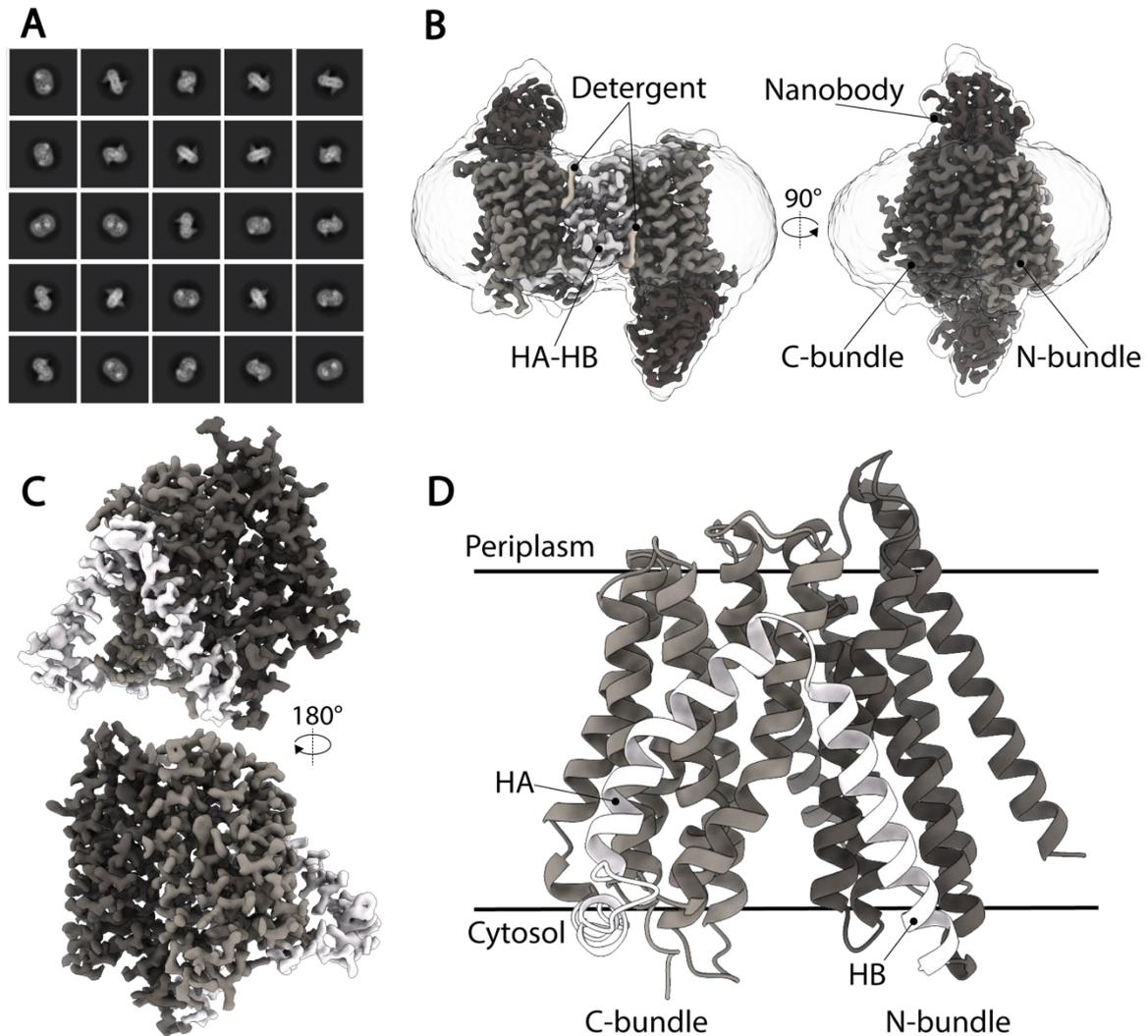


Figure 56. Cryo-EM structure of DtpC-Nb26. (A) Representative 2D class averages of the dimeric population. (B) 3D reconstruction of the DtpC-Nb26 inverted dimer used for local focused refinement on one copy of the transporter, shown in (C). (D) Atomic model of DtpC displayed as ribbon diagram. The different structural elements are labelled. This figure was directly taken from (Killer et al., 2022).

As mentioned earlier, the outer clamp in PepT1 and PepT2, is located between TM5 and TM8 (Figure 42). PHT1 and PHT2 do not have any salt bridge, but two compensatory stabilising hydrogen bonds between TM5 and TM8, and between TM1 and the short beta sheet located between TM9 and TM10 (Figure 47). Other than that, most bacterial POTs crystallized so far, have their outer clamp located between TM2 and TM7, similarly as DtpC (with the exception

of PepT_{So}, from *Shewanella oneidensis*, which also has it on TM5 and TM8) (Figure 58 B). Additional inter- and intra-bundle weak interactions, contribute to the stabilization of these transporters in particular conformations, but are less well conserved.

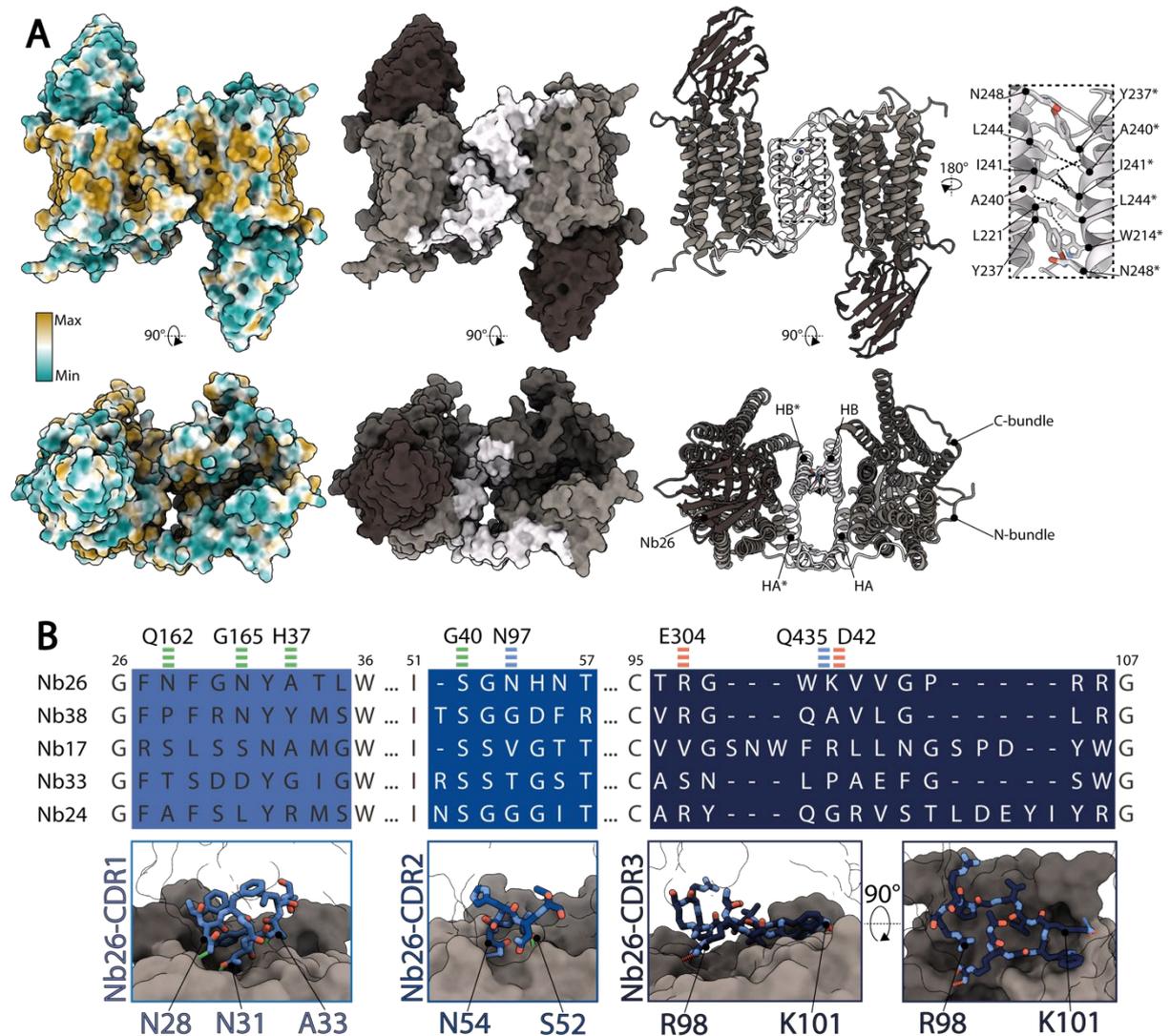


Figure 57. Quaternary structure of complex assembly. (A) Atomic model of the DtpC-Nb26 inverted dimer represented as surface colored by (left) molecular lipophilicity potential (MLP) and (right) by structural elements labelled on the ribbon diagram. The close up view on the right shows contacts at the dimer interface between atoms within a 3.8 Å distance. (B) Interactions between Nb26 and DtpC. The sequences of the five nanobodies representing five different families, obtained after selection, with their complementary determining regions (CDR) are shown. Interactions of Nb26 with DtpC are highlighted as green (hydrogen bonds involving the protein backbone), blue (hydrogen bonds involving side chains) and red dashes (salt bridges). These interactions are further displayed in 3D. CDR regions are depicted as sticks on the surface of DtpC where the N-terminal bundle is colored in grey, and the C-terminal bundle in dark grey. This figure was adapted from (Killer *et al.*, 2022).

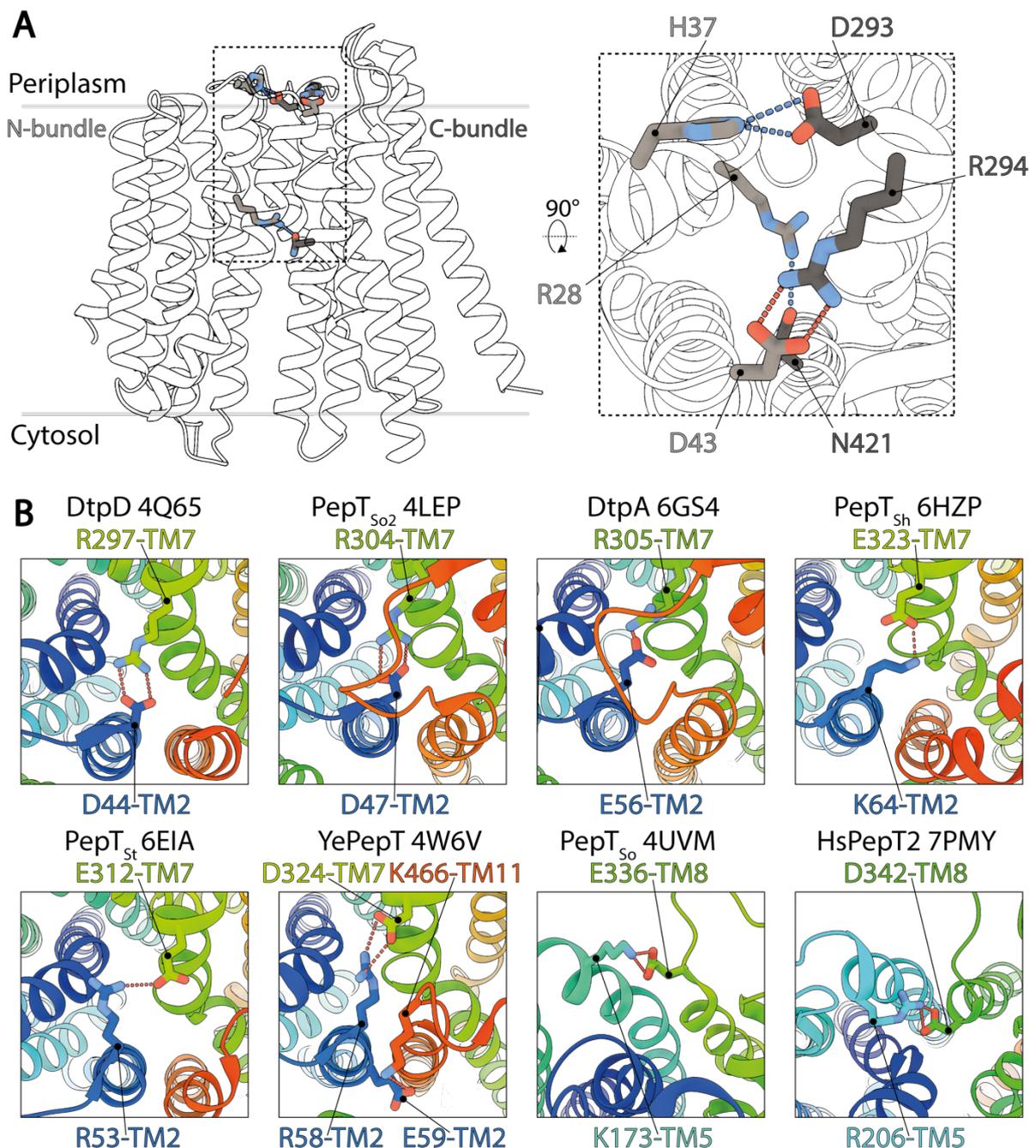


Figure 58. Structural basis for the stabilization of the inward facing state in DtpC and other POT homologues. (A) The salt bridge and hydrogen bonds favouring closure of the two bundles on the periplasmic side of DtpC are respectively shown as red and blue dashes. (B) The structures of homologous POTs from *Escherichia coli* (DtpD, DtpA), *Shewanella oneidensis* (PepT_{so2}, PepT_{so}) *Staphylococcus hominis* (PepT_{sh}), *Streptococcus thermophilus* (PepT_{st}), *Yersinia enterocolitica* (YePepT) and Homo sapiens (HsPepT2), were all previously captured in the IF state. Here they were analysed to identify the strongest interaction stabilizing their common conformation. The structures are coloured from blue to red, from their N- to C-termini, and the respective PDB accession numbers are indicated. Conserved salt bridges are

labelled and highlighted by red dashed lines. This figure was directly taken from (Killer et al., 2022).

These singularities, might contribute to the distinct turnover transport rates, reported in the literature, for different homologues. Still, either prototypical or atypical POTs share the principle of inter-bundle clamping mechanism (whether by salt bridging or hydrogen bounding), on both sides of the transporter, in order to mediate alternate access. DtpC's central cavity was then inspected. A striking difference with the binding site of prototypical POTs, is the presence of an aspartate residue (D392), acidifying the P2 pocket. Interestingly, this mutation seems to be one of the feature of atypical POTs (D395 in DtpD). Indeed, prototypical POTs, in contrast have a strictly conserved serine residue at this position (S599 in HsPepT1, S400 in DtpA, S397 in DtpB) (**Figure 59**). Previous biochemical studies showed that mutation of this residue, abolished transport of Ala-Lys-AMCA in both DtpC and DtpD (Jensen, Ismat, *et al.*, 2012). Docking of the dipeptide Ala-Lys, inside the binding site of DtpC corroborates a putative salt bridge between the lysine side chain of the dipeptide, and D392 (**Figure 59 D**). In prototypical POTs, the conserved E₁XXE₂R motif on TM1, was previously identified to be involved in proton coupling and peptide binding. Based on mutational studies and molecular dynamics simulations, it was proposed that E₁ (E23 in HsPepT1) and E₂ (E26 in HsPepT1) regulate the engagement of R (R27 in HsPepT1) with the C-terminus of dipeptides, through protonation-deprotonation events (Aduri *et al.*, 2015). In DtpC and DtpD, the E₁XXE₂R motif evolved to Q₁XXE₂Y. One of the constant salt bridge interaction between these transporters and the peptide is therefore lost. Based on the biochemical data available (Jensen, Ismat, *et al.*, 2012; Aduri *et al.*, 2015) and the structural analysis of DtpC, it is tempting to speculate that, the selectivity of atypical POTs for dipeptides with C-terminal lysine or arginine residues, is caused by the substitution of a salt bridge between the carboxyl group of the peptide C-terminus and the transporter (R21Y mutation in DtpC or R22Y in DtpD), to a side chain specific salt bridge with D392 (D395 in DtpD).

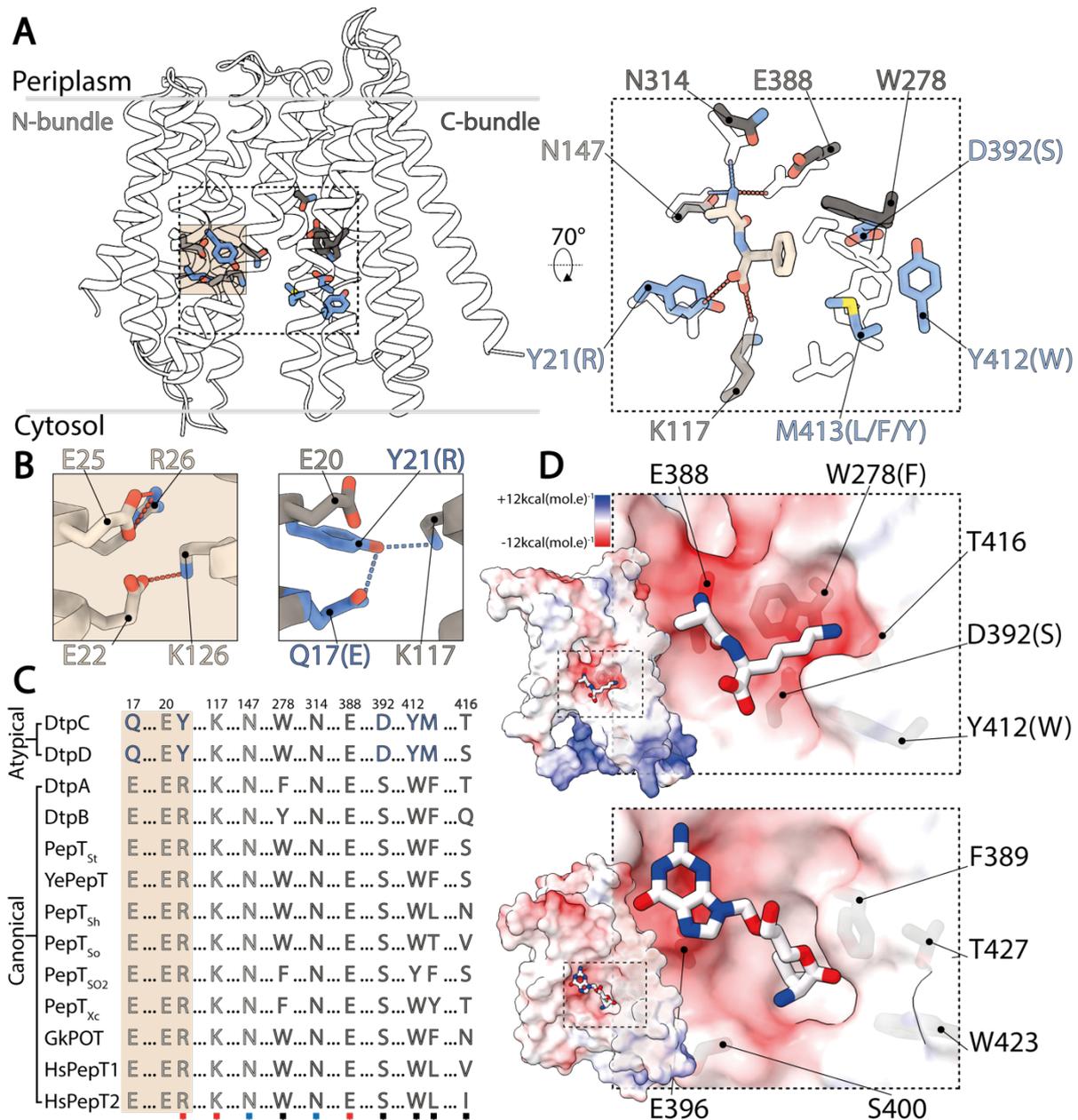


Figure 59. Structural basis for ligand selectivity in DtpC and atypical POTs. (A) Key residues involved in substrate binding are colored and shown as sticks. In the close up view, an overlay of HsPepT2 (transparent residues) bound to the dipeptide Ala-Phe (beige) with DtpC is shown. Residues colored in grey are conserved while blue residues are exclusive to atypical POTs. (B) The arrangement of the E₁XXE₂R motif from PepT_{St} is shown on the left panel, and the atypical Q₁XXE₂Y on the right. (C) The sequences of 13 POTs were aligned and residues involved in proton coupling and substrate binding are shown. The red squares mark residues strongly interacting with the charged termini of substrates peptides via salt bridges. The blue squares indicate two conserved asparagine residues stabilizing peptides through hydrogen bonds. The black squares point to residues constituting the side chain pocket of POTs, tuning ligand promiscuity or selectivity. (D) Surface representation colored by electrostatic potential, of the C-bundles of DtpC (top panel) and DtpA (bottom panel). A pose of the preferred substrate of DtpC, Ala-Lys, is proposed (top) and the co-crystallized

valganciclovir drug hijacking canonical POTs is shown in DtpA (bottom). PDB accession codes of previously published work: HsPepT2 bound to Ala-Phe: 7PMY; PepTSt: 5OXO; DtpA bound to valganciclovir: 6GS4. This figure was directly taken from (Killer *et al.*, 2022).

6.4.3 Discussion

The choice of the DtpC-Nb26 sample for larger data collection and high resolution structure determination, was guided by the quality of the 2D class averages, but certainly, does not signify that nanobodies are the best fiducials, for SPA cryo-EM on MFS transporters. Pro-macrobodies have been shown to be useful in the past (Botte *et al.*, 2022), and the split sfGFP strategy might only need further optimization. In fact, the Mb26 fiducial was clearly visible in 2D class averages. But, the MBP component of Mb26, adopted several configurations in relation to the nanobody (**Figure 53 B**). This challenge can be dealt with larger data collection, clustering of discrete conformations, and appropriate masking, but the more rigid DtpC-Nb26 sample seemed a simpler option. Regarding split sfGFP-DtpC₁₋₄₇₅-Nb26: although TMs could be visualised in 2D, the majority of classes displayed blurry density for the split sfGFP moiety. Non-self-complemented ends were also observed (**Figure 53 C**). The difference in fluorescence was not significant between sfGFP-DtpC₁₋₄₇₅ and sfGFP-DtpC_{FL}, compared to sfGFP-DtpC₁₋₄₇₀, but both of these samples might in fact have already too much restrained termini for stable self-complementation of sfGFP. Split sfGFP could be an interesting alternative to conformation specific fiducials (e.g., nanobodies, fabs, sybodies) for the study of MFS transporters by SPA cryo-EM, however, it would require more work to experimentally verify the optimal amount of restraint to introduce between the termini. Concerning the quaternary structure of the reconstructed protein: the allure of the steric exclusion chromatogram, already indicated the presence of higher oligomeric species. Small angle x-ray scattering (SAXS) was also performed on DtpC-Nb26, and showed a better fit for the dimeric volume, compared to the monomeric one. This suggests that a significant population of DtpC-Nb26 is dimeric in solution. However, this inverted dimer configuration, is unlikely to be biologically relevant. This highlights one of the disadvantage of highly concentrated *in vitro* systems, and artificial solubilisation methods, such as detergents, over *in situ* structural studies. Finally, one of the weakness of this study, is the lack of direct evidence for the binding of lysine, or arginine side chains to D392. Images were recently collected, in the presence of the dipeptide Ala-Lys, and will be analysed in the future.

7 Conclusion and outlook

Cells have evolved with transporters in order to absorb nutrients, fuel metabolic processes, and maintain homeostasis. Mammalian peptide transporters have resisted structure determination for many decades, until the recent technological advances of cryo-EM benefited to our work. I have demonstrated that this class of transporters uses an alternate access mechanism to transport substrates across membranes. Unlike other MFS transporters however, the task is performed with rearrangements of the N-terminal bundle, while the C-terminal bundle serves as a scaffold domain. I could also show how peptides are coordinated within the binding site while the transporter is in motion. There, the interaction between the transporter, and the charged termini of the peptide are critical, and allow to maintain the position of the latter, while alternate access takes place. Despite the great advances of cryo-EM, resolution and throughput remains limited for small and flexible targets, such as MFS transporters. To address the question of multi-substrate recognition, we used MX on the crystallizable bacterial homologue DtpB, displaying similar substrate promiscuity profiles as PepT1 and PepT2. With 14 high resolution co-crystal structures, I could produce a molecular movie of the binding site, changing and accommodating, to the presence of various peptides. It is striking how the N-termini of peptides are firmly hooked by conserved residues, while the side chains and C-termini can adopt various positions. The most important chemical feature which a prodrug should include to hijack prototypical POT shuttle systems, is therefore likely to be a positively charged group, such as a primary amine. In terms of methodological developments, the work on DtpC and PHT1, clearly showcased the usefulness of small and rigid fiducial markers, such as nanobodies and sybodies. The structure of DtpC allowed to provide a molecular explanation for substrate selectivity in atypical POTs, and the structure of PHT1 constitutes a starting point for *in silico* design of inhibitory drugs, competitively binding to the central cavity. In summary, this work constitutes a solid basis for understanding proton coupled oligopeptide transporters on a structural and molecular level. It also provides a methodological framework for structural studies on this transporter family, which could later benefit to drug design and discovery. If we now understand these individual transporters to a deeper level, their interactions with other signaling proteins is only starting to be considered as important in human health and disease. The field of research on SLC15 transporters, should therefore direct more efforts into characterizing such protein complexes, and focus on developing strategies to modulate their interactions.

8 Materials and methods

8.1 Materials

The materials used in this work is indicated in **Tables 2 to 7**.

Table 2. List of chemicals used in this work

Name	Supplier	Cat number
1,4-dithiothreitol (DTT), >99%	Roth	3483-12-3
10 x TBE	Roth	3061.2
2-Propanol	Roth	6752.3
Acetic acid	Roth	3738.4
Agarose for DNA electrophoresis	Serva	11404
Albumin, from Bovine Serum (BSA)	Sigma-Aldrich	A7906
Barium Chloride	Sigma-Aldrich	342920
Chloramphenicol	Roth	3886,3
cOmplete™ Protease Inhibitor Cocktail	Roche	5056489001
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	D8418
EDTA Tetrasodiumsalt 86-88%	Roth	3619.1
Ethanol >=99,8%	Roth	9065,3
Ethidium Bromide	Roth	HP471

Glycerol, Rotipuran >99,5%, p.a.	Roth	3783.1
HEPES, Pufferan >99,5%, p.a.	Roth	7365-45-9
Hydrochloric acid 32%	Roth	X896.1
Imidazole, >99%, p.a.	Roth	X998.4
Isopropyl- β -d-thiogalactopyranoside (IPTG)	Roth	2316.4
Kanamycin sulfate	Roth	T832.4
LB Agar (Lennox)	Roth	X965.1
LB Broth Low Salt Granulated	Melford	GL1703
Lithium sulphate	Sigma-Aldrich	203653
LDS Sample Buffer (4x), NuPAGE	Novex	NP0008
Magnesium chloride hexahydrate, >99%, ACS	Sigma-Aldrich	2189.1
MES, Pufferan >99%	Roth	4256.4
Nickel(II)chloride hexahydrate	Roth	4489.2
Potassium chloride, >99,5%, p.a., ACS, ISO	Roth	6781.1
PEG400	Sigma-Aldrich	06855
SDS Pellets	Roth	CN 30.3
Sekusept Plus	Ecolab	104372E
Sodium chloride, >99,5%, p.a., ACS, ISO	Roth	3957.2

Sodium hydroxide, pellets, >99%, p.a., ISO	Roth	6771
SuperSignal™ West Femto Maximum Sensitivity Substrate	ThermoFisher	34094
SuperSignal™ West Pico PLUS Chemiluminescent Substrate	ThermoFisher	34580
TB powder	Melford	T1510-1000.0
Tetracycline hydrochloride	Roth	HP63.2
Tris buffered saline with Tween® 20 (TBST)	Sigma-Aldrich	91414
TRIS hydrochloride, Pufferan, >99%, p.a.	Roth	9090.3
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP)	Soltec Bio Science	M115
TRIS, Pufferan, >99,9%, Ultra Qualitaet	Roth	5429.3
β-Ala-(L)-Lys-N-7-amino-4-methylcoumarin-3-acetic acid	Biotrend	BP0352
Peptide library	Bachem, Sigma-Aldrich, GL Biochem	
Brain total lipid extract (BL)	Avanti Polar Lipids	131101
Monoolein 9.9 MAG	Molecular Dimensions	MD2-67
Lauryl maltose neopentyl glycol (LMNG)	Anatrace	NG310
n-Decyl-β-D-maltoside (DM)	Anatrace	D322
n-Dodecyl-β-D-maltoside (DDM)	Anatrace	D310

n-Nonyl- β -D-glucoside (NG)	Anatrace	N324S
n-Nonyl- β -D-maltoside (NM)	Anatrace	N330
6x DNA loading dye	ThermoFisher	R0611
dNTP	New England Biolab	N0447S
Nuclease-free water	Qiagen	129115

Table 3. List of enzymes used in this work

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

Table 4. List of consumables used in this work

Name	Supplier	Cat number
Anti-Penta-HIS (HIS1K) biosensors	fortéBIO	18-5120
Bio-Beads™ SM-2 Resin	Biorad	152-8920
CapureSelect beads for EPEA-tag	ThermoFisher	194288010
Disposable PD 10 Desalting Columns	GE Healthcare	GE17-0851-01
Gene Ruler 1 kb DNA ladder	ThermoFisher	SM0311
Greiner CELLSTAR® 96 well plates flat bottom black polystyrene wells	Sigma	M0312
Greiner CELLSTAR® 96 well plates flat bottom clear wells	Sigma	M0812
illustra MicroSpin G-25 Columns	GE Healthcare	27532501
Instant Blue (coomassie based staining solution)	Expedeon	ISB
Loading Dye Purple	New England Biolab	B7024S
Mix & Go! <i>E. coli</i> Transformation Kit and Buffer Set	Zymo Research	T3001
Ni-NTA agarose	Invitrogen	R901-15
Novex NuPAGE LDS Sample Buffer	ThermoFisher	NP0008
Prometheus NT.48 Standard grade nanoDSF capillaries	Nanotemper	PK002
QIAquick Miniprep Kit	Qiagen	27104

QIAquick PCR Purification Kit	Qiagen	28106
Roti®-Mark 10-150	Roth	T850.1
RunBlue™ Bis-Tris Protein Gels 4-12% 12 well	Expedeon	NBT41212
RunBlue™ Bis-Tris Protein Gels 4-12% 17 well	Expedeon	NBT41227
SnakeSkin Dialysis Tubing MWCO 10 kDa	ThermoFisher	68100
Spin-X® UF 20 mL Centrifugal Concentrator, 10,000	Corning	431488
Spin-X® UF 20 mL/6mL Centrifugal Concentrator, 100,000 MWCO Membrane	Corning	431491/ 431486
Spin-X® UF 20 mL/6mL/0.5mL Centrifugal Concentrator, 5,000 MWCO Membrane	Corning	431487/ 431482/ 431477
Spin-X® UF 20 mL/6mL/0.5mL Centrifugal Concentrator, 50,000 MWCO Membrane	Corning	431490/ 431485/ 431480
Strep-Tactin® Sepharose® 50% suspension	iba Lifesciences	2-1201-010
Trans-Blot® Turbo™ Mini PVDF Transfer Packs	Biorad	1704156
EM grid for negative stain	Plano	S160-3
EM grids for cryo-EM	Plano	S174-8/S143-8/S343-8-UAUF
Tweezers to handle grids	Plano	5X.SA

Table 5. List of commercial crystallisation screens used in this work

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

Table 6. List of devices and instruments used in this work

Name	Supplier
Agilent 1260 with autosampler and fraction collector	Agilent technologies
ÄKTA Pure with F9C fraction collector	GE Healthcare
Analytical scale	Sartorius
Avanti JXB-26 Centrifuge	Beckmann-Coulter
CaptureSelect™ C-tagXL Pre-packed Column 1 mL	ThermoFisher
Centrifuge 5424 R	Eppendorf
Centrifuge 5810 R	Eppendorf
Electrophoresis chamber for agarose gels	NeoLab
Electrophoresis chamber for SDS gels	Invitrogen
EmulsiFlex-C3 homogenizer	Aventin
Glass homogenizer 55 mL	Wheaton
Freezer -20 °C	Liebherr

Freezer -80 °C	Eppendorf
Fridge 4 °C	Liebherr
Fume hood	Waldner
AnalytikJena UVP Chemstudio	AnalytikJena AG
Heatblock	Eppendorf
JLA 25.50 rotor for Avanti JXB-26 centrifuge	Beckmann-Coulter
JLA 8.1000 rotor for Avanti JXB-26 centrifuge	Beckmann-Coulter
Magnetic stirr plate	Roth
Photometer	Roth
Light Microscope	Nikon
Microwave	Severin
MilliQ machine	Millipore
MLA-130 rotor for Optima MAX-XP ultracentrifuge	Beckmann-Coulter
MonoQ HiTrap Q HP, 5 ml	GE Healthcare
Mosquito-LCP	ttplabtech
nanodrop 2000c	ThermoScientific
Nanotemper Prometheus NT.48	Nanotemper
New Brunswick™ Innova® 42 Incubator Shaker	Eppendorf
Octet RED96 System	Molecular Devices

Optima MAX-XP Benchtop Ultracentrifuge

Beckmann-Coulter

Optima XE-90 Ultracentrifuge

Beckmann-Coulter

PCR cycler

Eppendorf

Peristaltic pump

Medorex

pH meter

Mettler Toledo

Rotating wheel

Stuart

Rock imager

Formulatrix

Scales

Sartorius

Superdex® 200 HiLoad 16/600 pg

GE Healthcare

Superdex® 200 Increase 10/300 GL

GE Healthcare

Superdex® 75 HiLoad 16/600 pg

GE Healthcare

Shaking platform

Edward Bühler GmbH

Scorpion Screen Builder

ARI-arts Robbins Instruments

Sonicator waterbath

Elmasonic

TECAN Spark 20M multimode plate reader

TECAN

Ti 45 rotor for Optima XE-90 ultracentrifuge

Beckmann-Coulter

Vortex

Scientific Industries

Heating waterbath

VWR

MiniSpin® centrifuge	Eppendorf
MiniStar Microcentrifuge	VWR
Electron microscopes	ThermoScientific
High performance computing clusters (GPU and CPU nodes).	EMBL-Hamburg/EMBL-Heidelberg/DESY/CSSB Hamburg
Vitrobot	ThermoScientific

Table 7. List of cellular strains used in this work

Name	Organism	Use
BL21 (DE3)	<i>E. coli</i>	Production of soluble proteins (TEV, 3C protease)
C41 (DE3)	<i>E. coli</i>	Production of bacterial POTs (DtpB, DtpC)
DH5a	<i>E. coli</i>	Cloning and DNA production
Rosetta-gami 2 (DE3)	<i>E. coli</i>	Production of Saposin A
WK6	<i>E. coli</i>	Production of nanobodies (Nb132, Nb17, Nb26, Nb38)
HEK293F	<i>H. sapiens</i>	Production of HsPepT1, HsPepT2, GgPHT1
HEK293S GnTI-	<i>H. sapiens</i>	Production of HsPepT1, HsPepT2, GgPHT1
HEK293 Expi	<i>H. sapiens</i>	Production of HsPepT1, HsPepT2, GgPHT1

8.2 Methods

8.2.1 Transformation of bacteria with DNA:

50 μ L of chemically competent cells were mixed with 10 to 100 ng of plasmid DNA and incubated on ice for 30 min before a heat-shock at 42 °C for 30 s. Cells were grown in 900 μ L LB medium at 37 °C for 45 mins before spreading on LB agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight.

8.2.2 Polymerase chain reaction (PCR):

PCRs were done using single bacterial colonies, linearized inserts, or plasmid DNA (~10 ng) as templates. Typically, 10 mM of dNTPs, forward and reverse primers were added. 1% Vol of polymerase (Phusion or Taq polymerase), and appropriate enzyme reaction buffers were supplemented, according to New England Biolabs recommendations. The reactions steps (temperature and time) were set in appropriate values, depending on the polymerase, and the length of the template.

8.2.3 Gene sequencing

Samples were sequenced using the Sanger sequencing method by Microsynth AG, using appropriate primers.

8.2.4 Insertion of the genes of interest into DNA vectors

The genes of HsPepT2 was previously cloned in the lab in a pXLG vector with a N-terminal Hexa-histidine tag for affinity purification, a EGFP as expression reporter, and a TEV protease recognition sequence (Pieprzyk, Pazicky and Löw, 2018a), using the SLICE technique (Zhang, Werling and Edelman, 2014) . The N528Q-N587Q mutation was introduced by quick-change mutagenesis. The StrepII-HsPepT2 construct was designed by replacing the Hexa-histidine tag, EGFP, and TEV protease recognition sequence, for a twin-streptavidin tag, followed by a 3C protease recognition sequence. The gene of HsPepT1 was inserted in the latter vector, using the SLICE technique. The genes of DtpB and DtpC were previously cloned in the lab using the LIC technique in PNIC vectors (Celie, Parret and Perrakis, 2016). For SfGFP-DtpC_{FL}, the first 6 N-terminal beta strands of sfGFP were fused to the N-terminus of DtpC, and the beta strands 7 to 11 fused to the C-terminus. SfGFP-DtpC₁₋₄₇₅, and SfGFP-DtpC₁₋₄₇₀ were cloned with truncations of 5 and 10 residues respectively, on the C-terminal side of DtpC. The macrobody

26 (Mb26) was generated by first inserting the nanobody 26 (Nb26) into a pBXNPH3 vector containing a C-terminal penta-histidine tag preceded of a HRV-3C protease recognition sequence, and then, inserting the maltose binding protein (MBP) in frame with the 3' end of the nanobody, with two prolines as a linker between the two genes as described in (Botte *et al.*, 2022).

8.2.5 Expression and purification of sapsin A protein

Rosetta gami-2(DE3), transformed with the Sap-A plasmid, were grown at 37°C in TB medium supplemented with 30 µg/mL kanamycin, 30 µg/mL chloramphenicol and 10 µg/mL tetracycline. At an OD₆₀₀ nm of 0.8, the cells were induced with 1 mM IPTG and incubated for 4 h at 37°C. Cells were harvested by centrifugation (10,000xg, 15 min, 4 °C). Cell pellets were stored at -20 °C until further use. The SapA protein was purified by affinity chromatography (Ni-NTA), using standard protocols.

8.2.6 Expression and purification of nanobodies and macrobodies

The nanobody plasmids were transformed into *E. coli* strain WK6. The cells were grown at 37°C in TB medium supplemented with 100 µg/mL ampicillin. At an OD₆₀₀ nm of 0.8 the cells were induced with 1 mM IPTG and incubated for 16 h at 27°C. Cells were harvested (10,000xg, 15 min, 4 °C). Cell pellets were stored at -20 °C until further use. Cell pellets were resuspended in TES buffer (0.2 M TRIS, pH 8, 0.5 mM EDTA, 0.5 M sucrose) supplemented with protease inhibitor tablets. Osmotic shock was performed by the addition of diluted TES buffer to release the periplasmic proteins. The solution was first centrifuged for 20 min at 10,000xg and additionally for 30 min at 100,000xg. For the nanobodies, the supernatant was applied to CaptureSelect beads, which were equilibrated with wash buffer (20 mM NaPi, pH 7.5, 20 mM NaCl). After three column volumes of washing, the nanobody was eluted with 20 mM HEPES, pH 7.5, 1.5 M MgCl₂. The nanobodies were further purified on a HiLoad 16/600 Superdex 75 pg column in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 % glycerol, concentrated with a 5 kDa cut-off concentrator, flash-frozen and stored at -80 °C until further use. For Mb26, the supernatant was purified by immobilized-metal affinity chromatography (IMAC) on a gravity column. The beads were pre-equilibrated in 20 mM NaPi at pH 7.5, 300 mM NaCl, 5% glycerol, 15-30 mM imidazole, 0.5 mM TCEP and incubated. Loaded beads were washed with increasing imidazole concentrations (20 mM NaPi at pH 7.5, 300 mM NaCl, 5% glycerol, 15-30 mM imidazole, 0.5 mM TCEP, 0.03% DDM). The proteins were eluted from the column

with a buffer containing high imidazole concentration (20 mM NaPi at pH 7.5, 150 mM NaCl, 5% glycerol, 250 mM imidazole, 0.5 mM TCEP, 0.03% DDM) and combined with 1 mg of 3C protease to perform the His-tag cleavage. The cleaved protein was recovered by negative IMAC, concentrated to 0.5 ml using a 30 kDa concentrator (Corning® Spin-X® UF concentrators) and run on an ÄKTA Pure system (GE Healthcare Life Sciences), using a Superdex 75 Increase 10/300 column. Fractions containing the protein were also pooled, concentrated, flash frozen and stored at -80°C until further use.

8.2.7 Expression and purification of prokaryotic membrane proteins.

E. coli strain C41(DE3), transformed with the membrane protein genes, were grown in TB medium, supplemented with the appropriate antibiotic at 37°C. The cells were induced with 0.2 mM IPTG at OD₆₀₀ nm of 0.7. Then, cells were incubated further for 16 h at 18°C and harvested by centrifugation (10,000xg, 15 min, 4 °C). Cell pellets were stored at -20 °C until further use. Cell pellets were resuspended in lysis buffer (20 mM NaPi at pH 7.5, 300 mM NaCl, 5% (v/v) glycerol, 15 mM imidazole, with 3 ml of lysis buffer per gram of wet weight pellet), supplemented with lysozyme, DNase and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP). The cells were lysed by three cycles using an Emulsiflex homogenizer at 10,000-15,000 psi. Recovered material was centrifuged to remove non-lysed cells (10,000 × g, 15 minutes, 4°C) and the supernatant was subjected to ultracentrifugation to separate the membrane fraction (100,000 × g, 1 hour, 4°C). Membranes were resuspended in lysis buffer supplemented with cOmplete EDTA-free protease inhibitors, and solubilized by adding 1% n-Dodecyl-β-D-Maltoside (DDM) detergent (Anatrace). The sample was centrifuged for 50 min at 90,000 × g, and the supernatant was applied to Ni-NTA beads for immobilized-metal affinity chromatography (IMAC) on a gravity column. The beads were pre-equilibrated in lysis buffer and incubated with the solubilized membrane proteins for one hour at 4°C on a rotating wheel. Loaded beads were washed with buffer with increasing imidazole concentrations (20 mM NaPi at pH 7.5, 300 mM NaCl, 5% glycerol, 15-30 mM imidazole, 0.5 mM TCEP, 0.03% DDM). The proteins were eluted from the column with a buffer containing high imidazole concentration (20 mM NaPi at pH 7.5, 150 mM NaCl, 5% glycerol, 250 mM imidazole, 0.5 mM TCEP, 0.03% DDM) and combined with 1 mg of TEV protease to perform the His-tag cleavage during dialysis overnight at 4°C. The dialysis buffer contained 20 mM HEPES at pH 7.5, 150 mM NaCl, 5% glycerol, 0.5 mM TCEP, 0.03% DDM. The cleaved protein was recovered by negative IMAC, concentrated to 4 ml using a 50 kDa concentrator and run on an

ÄKTA Pure system, using a HiLoad 16/ 600 Superdex 200, Superdex 200 Increase 10/300 columns, or an Agilent HPLC system with Superdex 200 Increase 5/150, or home-packed superpose 6 columns. Fractions of interest were flash frozen and stored at -80°C until further use.

8.2.8 Expression and purification of mammalian membrane proteins.

Human strains (HEK293F, HEK293S GnTI-, or HEK293 Expi), were transiently transfected with the membrane protein genes. HEK293F cells were collected 48 h after transient transfection as previously (Pieprzyk, Pazicky and Löw, 2018b) by centrifugation (10,000xg, 15 min, 4 °C). Cell pellets were stored at -20 °C until further use. Cell pellets were resuspended in 300 mM NaCl, 20 mM NaPi pH 7.5, 0.5 mM TCEP, 5% glycerol, supplemented with cOmplete™ EDTA-free protease inhibitors. Cells were disrupted using an Emulsiflex homogenizer and the lysate was centrifuged 10 min at 10 000xg. and the supernatant centrifuged for 90 min at 95 000xg. The pellet containing the membrane fraction was solubilized in 1% detergent (LMNG or DDM) and 0.1% Cholesteryl Hemisuccinate for 1 h at 4°C. The sample was then centrifuged for 50 min at 70 000xg and the supernatant was applied to Strep-Tactin or Ni-NTA beads. After incubation, the suspension was transferred to a gravity column. With Strep-Tactin beads, two wash steps with 300 mM NaCl, 20 mM HEPES pH 7.5, 3xCMC of detergent, and 1:10 detergent:CHS weight ratio, were performed, and the target protein was eluted with 10 mM desthiobiotin. With Ni-NTA beads, two wash steps with 15 mM Imidazole, 300 mM NaCl, 20 mM HEPES pH 7.5, 3xCMC of detergent, and 1:10 detergent:CHS weight ratio, were performed, and the target protein was eluted with 250 mM Imidazole. 3C cleavage was performed in 30 mins and the protease was separated from the target protein by gel filtration using a Superose® 6 Increase 10/300 column. TEV cleavage was performed overnight, under dialysis, to remove the excess of Imidazole, and perform an additional reverse Ni-NTA purification step, before gel filtration. The best fractions were concentrated using a 100 kDa cutoff concentrator, and stored at -80°C until further use.

8.2.9 Reconstitution of HsPepT2 in saposin lipid nanoparticles (SapNP)

HsPepT2 solubilised and purified in DDM-CHS was reconstituted in SapNP, using a sapA:HsPepT2 molar ratio of 20, and a lipid:SapA molar ratio of 5. The brain lipid extract was solubilised to 5 mg/mL, in 0.28% DDM, using a sonication bath and several cycles of flash freezing and thawing. 750 µL of the lipid mix were then incubated at 37°C for 10 mins and

mixed with 1.1 mL of HsPepT2, concentrated at 5 mg/mL (45 μ M). The mixture was gently pipetted up and down 100 times, without making foam, and left 10 mins at room temperature. 2.65 mL of purified SapA at 3.2 mg/mL (360 μ M) was added and the mixture was homogenised again by pipetting up and down 100 times, before 30 mins incubation at room temperature. 500 μ L of buffer (20 mM Hepes pH 7.5 150 mM NaCl 0.5 mM TCEP) was added to the mix, and an additional incubation of 20 mins was done. 200 mg of biobeads, prewashed with methanol were added to the reaction mix, in order to remove the excess of detergent. The reaction was left on a rotating wheel, overnight, at 4°C. The next morning, the sample was recovered and separated from the bio beads using a thin SDS-PAGE loading tip. The sample was concentrated using a 50 kDa cut-off membrane, before injection in gel SEC column, to separate empty saposin-lipid nanoparticles from saposin-lipid-HsPepT2 ones.

8.2.10 Protein concentration determination

Protein concentrations were determined from the absorbance at 280 nm using the extinction coefficient at 280 nm estimated from the sequence.

8.2.11 Electrophoresis

For DNA electrophoresis, a 1 % agarose (w/v) gels, solubilised in running buffer, were prepared. Samples were mixed with Loading Dye Purple (NEB) and the gels were run at 100 mV for 60 min in running buffer. Gels were imaged using Ethidium bromide and a UV detector. For SDS-PAGE, 4-12% Bis-Tris pre-casted gel (Expedeon) were used. Samples were mixed with Novex NuPAGE LDS Sample Buffer (ThermoFisher) supplemented with 70 mM TCEP and the gels were run at 180 V for 45 min in running buffer. Gels were imaged using a visible light detector. For western blots, SDS-PAGE was performed and transferred on a PVDF membranes. 3% BSA in TBST, was applied to the membranes for saturation, 30 mins, at room temperature. The membranes were washed three times with TBST, and incubated with the appropriate antibody(ies). Membranes were imaged using AnalytikJena UVP Chemstudio system.

8.2.12 Transport assays with HsPepT2

This paragraph was adapted from (Killer *et al.*, 2021). HsPepT2 was expressed for 48 hours at 37°C, 8% CO₂ at 220 rpm. For competition assays, 4×10^6 cells/mL resuspended in PBS buffer at pH 6.0 supplemented with 5 mM glucose were incubated in 96 well-plates, with 50 μ M β -

Ala-Lys-AMCA in absence or presence of dipeptides, tripeptides, or drugs for 10 min at 37 °C. The reaction was stopped by adding 200 µL of ice-cold buffer, and the cells were then washed three times with the same buffer. Finally, the cells were resuspended in 200 µL of buffer, and the fluorescence was measured in a M1000 microplate reader (TECAN) with excitation at 350 nm and emission at 450 nm. All experiments were performed in triplicates. The results were normalized by the fluorescence value of the control (cells overexpressing HsPepT2 incubated with AK-AMCA in the absence of inhibitor) and plotted as AK-AMCA uptake rate percentage. For concentration dependent uptake experiments, IC50 values were processed in GraphPad Prism 9.0 (GraphPad Software) using sigmoidal four parameter curve fitting.

8.2.13 Thermal unfolding of proteins

The differential scanning fluorimetry method was used to follow the thermal unfolding event of proteins, with a Prometheus NT.48 device (NanoTemper Technologies, Munich, Germany). The purified proteins were diluted to 10 to 30 µM, and the fluorescence at 330 and 350 nm was recorded over a temperature gradient scan from 15° to 95°C, with an excitation power set between 15-25 %. Data was processed in GraphPad Prism 9.0 (GraphPad Software).

8.2.14 3D structure predictions

Amino-acid sequences were used as input for AlphaFold2 structure prediction (Jumper *et al.*, 2021), and AMBER relaxation. The best ranked models were used for visualization.

8.2.15 General crystallization trial workflow performed on bacterial and mammalian membrane proteins.

Mammalian membrane proteins (notably HsPepT2), were de-glycosylated using enzymatic treatments (PNGaseF or EndoH). Another strategy to reduce glycan heterogeneity, was to use the expression strain HEK293S GnTI-. The glycosidases were separated using gel filtration. The prokaryotic targets DtpB and DtpC did not contain any glycosylation sites, and were therefore not enzymatically treated. However, crystallisation chaperones (nanobodies) were generated following established protocols (Pardon *et al.*, 2014), to favour crystal contacts. The nanobodies were typically added in excess prior SEC, or 1h before crystallisation, in a 1:1.1 ratio. The membrane proteins, or membrane protein-nanobody complexes, were concentrated between 4 and 10 mg/mL for vapour diffusion crystallisation, and between 6 and 60 mg/mL

for lipid cubic phase (LCP) crystallisation, using concentrators with 50 kDa or 100 kDa membrane cut-off. Stabilising peptides were added in final concentrations ranging between 2 and 20 mM. To remove potential aggregates, samples were centrifuged 30 to 60 mins, at 14,000xg, and 4°C before setting up the plates with a Mosquito-LCP robot (TTP Labtech). For the vapour diffusion method, three ratios of protein-to-reservoir of 200:100 nL, 150:150 nL and 100:200 nL were typically used, and the plates were incubated at 19 °C or 4 °C, while being imaged by a RockImager system (Formulatrix). For the LCP method, the membrane protein (40 % v/v) was mixed, at room temperature, with pre-warmed lipid at 38-40 °C, such as Monoolein (60 % v/v), using two coupled Hamilton syringes. Care was taken, not to introduce any air in the system. When using other lipids, the ratio was adjusted according to the relevant literature, to reach the LCP (Caffrey, 2015). The mixing was stopped when the mixture turned optically clear. The plates were set-up using a Mosquito-LCP robot (TTP Labtech) dispensing 50-100 nL mesophase and 800-1000 nL reservoir buffer into 96-well plastic, or glass sandwich plates. All LCP plates were incubated at 19 °C, and were inspected manually, using a light microscope equipped with a polarizer. Crystals were manually mounted on fishing loops, under a light microscope, and rapidly cryo-cooled in liquid nitrogen. Their diffraction power was measured at the P13 and P14 beamlines operated by EMBL-Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). HsPepT2 did not produce any crystal. DtpC-Nb26 yielded many crystals but with low diffraction power. Only DtpB-Nb132 produced highly diffracting crystals.

8.2.16 Specifics of the co-crystallizations of DtpB-Nb132 with di- and tripeptides

DtpB and Nb132 were mixed in 1:1.1 molar ratio 1 h prior to crystallization and incubated at 4 °C. Crystallization plates were prepared with an automated liquid handler (Mosquito, TTP Labtech) using the sitting drop vapor diffusion technique with a final drop volume of 300 nL (at 1:1, 1:2 and 2:1 (v/v) ratios of protein to precipitant), as described above. Peptides were purchased from Sigma-Aldrich, Bachem and GL Biochem (Shanghai) and stocks were prepared by weighing the lyophilized powder using an analytical balance and diluting them in ultrapure water. The initial crystals of the Ala-Leu-Ala bound structure were obtained from the MemGold2 crystallization screen (Molecular Dimensions) at 19°C. After several rounds of optimization, several DtpB-Nb132 crystals grown in 100 mM Hepes pH 6.7, 30-40 % PEG 400 100-300 mM NaCl, 100-300 mM MgCl₂, 10 mM Ala-Leu-Ala, using protein:precipitant

volume ratios of 1:1, 1:2, and 2:1, diffracted X-rays to 2.5 Å. The structures bound to Ala-Phe, Ala-Ile, Ala-Leu, Ala-Gln, Ala-Val, Asn-Val, Ser-Leu, Ala-Gln were obtained from crystals grown in the same conditions as for Ala-Leu-Ala (100 mM HEPES pH 6.7, 30-40 % PEG 400 100-300 mM NaCl, 100-300 mM MgCl₂, and 10 mM of the respective peptide) . Structures bound to Lys-Val, Met-Ser were obtained from crystals grown in 100 mM MES pH 6.0, 30-40 % PEG 400 100-300 mM NaCl, 200 mM BaCl₂, and 20 mM of the respective peptides. Structures bound to Ala-Phe-Ala, Ala-Trp, Ala-Trp-Ala were obtained from crystals grown in 100 mM HEPES pH 7.0, 30-40 % PEG 400 100-300 mM NaCl, 200 mM Li₂SO₄ and 10 mM of the respective peptide. The Ala-Pro-phe structure was obtained from crystals grown in 100 mM HEPES pH 7.0, 30-40 % PEG 400 100-300 mM NaCl, 200 mM Li₂SO₄ and 4 mM Ala-Phe-Ala. Here, once the crystal growth seemed appropriate, the crystals were soaked with 25 mM Ala-Pro-Phe and incubated 1 h at 19°C before harvesting. For all structures, single-crystal monochromatic diffraction datasets of 3600 or 2400 frames were recorded by the rotation method on EIGER 6M detectors with respective oscillation angles of 0.1° and 0.15° at the P13 and P14 beamlines operated by EMBL-Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany). Collected reflections were indexed, integrated and scaled using the program XDS. The unit cell dimensions were roughly: (a,b,c; alpha, beta, gamma) 54, 124, 169; 90, 90, 90 and the space group was identified as P 2 21 21. Initial phases were obtained by molecular replacement using the atomic model of DtpA (6GS4) and a nanobody structure as search models in Phaser as part of CCP4i2 on the ala-leu-ala bound dataset. The model was further built manually in Coot and refined in PHENIX and Isolde in iterative cycles. Difference map and omit map positive peaks clearly indicated the presence of a ligand within the binding site. Fitting the tripeptide ala-leu-ala in the density led to better agreement between the experimental data and the atomic model. For the other ligand bound datasets, the latter was repeated with the appropriate peptides and also led to better agreement between the experimental data and the atomic models.

8.2.17 SPA cryo-EM on HsPepT1, HsPepT2, GgPHT1, and DtpC

For all samples, which reached ‘sidechain resolution’ (i.e., 2.7 Å to 3.9 Å), a final SEC run was performed on an HPLC system, just before vitrification. This was particularly useful, as it allowed to (i) check the homogeneity of the sample, before immobilising it, (ii) reduce the salt, buffer, detergent concentrations (especially accumulated empty micelles), as much as possible, (iii) reach a top fraction concentration ranging from 2 and 6 mg/mL, and select it for

vitrification, without further concentrating step. Gold holey Quantifoil 2/1 300 mesh, or Ultrafoil 1.2/1.3 300 mesh grids were glow discharged between 60 and 120 seconds on a GloCube device at -25 mA. The grids were blotted using a Mark IV Vitrobot (ThermoFisher) for 3 to 6 s with forces ranging from -7 to +7, or manually, depending on the state of the device. The blotting chamber was maintained as much as possible, at 4 °C and 100% humidity during freezing. Typically, EM grids were pre-screened on a Talos Arctica EM microscope, equipped with a Falcon III camera. The best grids were then selected for further measurements on a Titan Krios EM microscope, equipped with a K3 camera and a Gatan energy filter. Samples yielding interpretable volumes, were then selected for much larger data collection sessions (typically 15,000 to 50,000 movies). During data acquisition, care was taken to adjust the Gatan energy filter's slit to ~10 eV whenever it was possible, and image aberrations were estimated and minimized as much as possible using auto-correlation tools and stigmators, at least two times a day. The defocus range was set as low as possible, in order to obtain enough low spatial frequency information in the images, while minimizing envelope function, and delocalization effects on high spatial frequencies. Movies were collected at a nominal magnification of 105,000 \times , physical pixel size 0.85 Å, or at nominal magnification of 130,000 \times , and physical pixel size 0.67 Å, with a 70 μ m C2 aperture and 100 μ m objective aperture at a dose rates ranging between of 15 and 20 e-/pixel per second. The total doses ranged from 50 to 80 e-/Å², split in 40 to 60 frames. The images were analysed using both Relion (Scheres, 2012; Zivanov *et al.*, 2018) and CryoSPARC (Punjani *et al.*, 2017; Punjani, Zhang and Fleet, 2020) softwares. The combination of particle trajectories and cumulative beam damage correction in Relion (Bayesian Polishing), together with the Non-Uniform Refinement algorithm from CryoSPARC allowed to improve the quality and interpretability of 3D reconstructions. Calculations were performed using high performance computing clusters of GPUs and CPUs, provided by DESY, CSSB-Hamburg, EMBL-Heidelberg, and EMBL-Hamburg, depending on resource availabilities. For postprocessing, half maps were used as inputs to assess various post-processing strategies such as the CryoSPARC's sharpening tool, DeepEMhancer (Sanchez-Garcia *et al.*, 2020), and Resolve_cryo-em (Terwilliger *et al.*, 2020). Combining these different approaches helped in interpreting the volumes and facilitate model building in Coot and refinement in Isolde (Croll, 2018), and Phenix (Afonine *et al.*, 2018). AlphaFold2 (Jumper *et al.*, 2021) predictions further guided the interpretation of the experimental data. The characteristics of data acquisition, and analysis are detailed below, in **Table 8**, **Table 9**, and **Table 10**. More details are also available in (Killer *et al.*, 2021, 2022)

Table 8. Data collection and refinement statistics of the deposited HsPepT1 and HsPepT2 structures. Table adapted from (Killer et al., 2021)

Protein reconstructed	HsPepT2, Ala-Phe Inward facing partially occluded	HsPepT1, Apo Outward facing open	HsPepT1 Ala- Phe Outward facing open	HsPepT1 Ala- Phe Outward facing occluded
PDB accession code	7PMY	7PN1	7PMX	7PMW
EMDB accession code	EMD-13544	EMD-13545	EMD-13543	EMD-13542

Data acquisition				
Microscope/Detector	Titan Krios/Gatan K3	Titan Krios/Gatan K3	Titan Krios/Gatan K3	Titan Krios/Gatan K3
Imaging software	EPU	EPU	EPU	EPU
Magnification	105,000	105,000	130,000	130,000
Voltage (kV)	300	300	300	300
Electron exposure (e-/Å ²)	81	66	55	55
Dose rate (e-/pix/s)	19.5	16	15	15
Frame exposure (e-/Å ²)	1.8	1.3	1.4	1.4
Defocus range (µm)	-1.2 to -2.5	-0.75 to -2.5	-1.0 to -2.0	-1.0 to -2.0
Pixel size (Å)	0.85 (physical)	0.85 (physical)	0.67 (physical)	0.67 (physical)
Micrographs	34,712	22,537	37,822	37,822

Reconstruction				
Picked coordinates (cryolo)	4,388,314	2,091,726	6,046,602	6,046,602
Particles in 3D classification (RELION)	2,944,737	1,459,348	4,247,238	4,247,238
Particles in final refinement (CryoSPARC)	454,149	199,987	466,042	107,791
Symmetry imposed	C1	C1	C1	C1
Map sharpening method /Final map kurtosis	Phenix Autosharpen / 130.45	Phenix Autosharpen / 75.88	Phenix Autosharpen / 52.35	Phenix Autosharpen / 54.35
Map resolution, global FSC 0.143 (Å) unmasked/masked	3.8/3.8	3.8/4.0	3.5/3.5	4.1/4.1

Refinement				
Initial model used for ECD (PDB code)	AlphaFold2 <i>HsPepT2</i>	AlphaFold2 <i>HsPepT1</i>	AlphaFold2 <i>HsPepT1</i>	AlphaFold2 <i>HsPepT1</i>
Model resolution (Å)				
FSC 0.5, masked/unmasked	3.9/4.0	4.1/4.3	3.7/3.8	4.2/4.4
FSC 0.143, masked/unmasked	3.7/3.8	3.8/3.9	3.5/3.6	4.0/4.1
Model composition				
Non-hydrogen atoms	5224	5147	5164	5030
Protein residues	659	657	659	643
ADP B factor (Å ²) mean	64.06	106.64	110.68	69.49
R.m.s deviations				
Bond lengths (Å) (#>4σ)	0.002 (0)	0.002 (0)	0.003 (0)	0.002 (0)
Bond angles (°) (#>4σ)	0.470 (1)	0.568 (1)	0.536 (0)	0.440 (0)
Validation				
MolProbity score	1.40	1.48	1.79	1.57
Clashscore	4.86	6.17	6.24	7.69
Rotamer outliers (%)	1.40	0.00	2.13	0.91
Ramachadran plot				
Favored (%)	97.86	97.23	96.76	97.16
Allowed (%)	1.98	2.77	3.24	2.84
Outliers (%)	0.15	0.00	0.00	0.00

Table 9. Data collection and refinement statistics of the deposited DtpC structure. Table from (Killer et al., 2022)

Protein reconstructed	Di- and tripeptide permease C (DtpC)
PDB accession code	7ZC2
EMDB accession code	EMD-14618

Data acquisition	
Microscope/Detector	Titan Krios/Gatan K3
Imaging software	EPU
Magnification	105,000
Voltage (kV)	300
Electron exposure (e-/Å ²)	75
Dose rate (e-/pix/s)	19.5
Frame exposure (e-/Å ²)	1.5
Defocus range (μm)	-0.9 to -1.8
Physical pixel size (Å)	0.85
Micrographs	24,333

Reconstruction	
Picked coordinates (cryolo)	6,464,070
Particles in 3D classification (RELION)	6,365,235
Particles in final refinement (CryoSPARC)	878,428
Symmetry imposed	C1
Map sharpening method	Phenix Resolve cryo em
Map resolution, FSC _{half maps; 0.143} masked/unmasked (Å)	2.72/3.43

Refinement	
Initial model used for refinement	AlphaFold2 model, relaxed with Amber
Model resolution (Å)	
FSC 0.143, masked/unmasked	2.64/5.43
Model composition	
Non-hydrogen atoms	7334
Protein residues	471
ADP B factor (Å ²) mean	12.73
R.m.s deviations	
Bond lengths (Å) (#>4σ)	0.003 (0)
Bond angles (°) (#>4σ)	0.616 (0)
Validation	
MolProbity score	1.44
Clashscore	8.04
Rotamer outliers (%)	0.00
Ramachandran plot	
Favored (%)	98.29
Allowed (%)	1.71
Outliers (%)	0.00

Table 10. Data collection and refinement statistics of the PHT1-Sb27 structure.

Protein reconstructed	PHT1-Sb27
PDB accession code	NA
EMDB accession code	NA
Data acquisition	
Microscope/Detector	Titan Krios/Gatan K3
Imaging software	EPU
Magnification	105,000
Voltage (kV)	300
Electron exposure (e-/Å ²)	75
Dose rate (e-/pix/s)	19.5
Frame exposure (e-/Å ²)	1.5
Defocus range (µm)	-0.9 to -2
Physical pixel size (Å)	0.85
Micrographs	53,070
Reconstruction	
Picked coordinates (cryolo)	7,347,786
Particles in 3D classification (RELION)	6,644,417
Particles in final refinement (CryoSPARC)	1,328,233
Symmetry imposed	C1
Map sharpening method	CryoSPARCv3/DeepEMhancer tight target
Map resolution, FSC _{half maps; 0.143} masked/unmasked (Å)	3.3/3.4
Refinement	
Initial model used for refinement	AlphaFold2 model, relaxed with Amber
Model resolution (Å)	
FSC 0.143, masked/unmasked	3.3/3.3
Model composition	
Non-hydrogen atoms	6205
Protein residues	805
ADP B factor (Å ²) mean	89.11
R.m.s deviations	
Bond lengths (Å) (#>4σ)	0.003 (0)
Bond angles (°) (#>4σ)	0.421 (0)
Validation	
MolProbity score	1.43
Clashscore	4.92
Rotamer outliers (%)	0.00
Ramachandran plot	
Favored (%)	97.00
Allowed (%)	3.00
Outliers (%)	0.00

9 References

- Abramson, J. *et al.* (2003) 'Structure and mechanism of the lactose permease of *Escherichia coli*', *Science (New York, N.Y.)*, 301(5633), pp. 610–615. Available at: <https://doi.org/10.1126/science.1088196>.
- Adibi, S.A. *et al.* (1975) 'Evidence for two different modes of tripeptide disappearance in human intestine. Uptake by peptide carrier systems and hydrolysis by peptide hydrolases', *The Journal of Clinical Investigation*, 56(6), pp. 1355–1363. Available at: <https://doi.org/10.1172/JCI108215>.
- Adibi, S.A. (1976) 'Intestinal phase of protein assimilation in man', *The American Journal of Clinical Nutrition*, 29(2), pp. 205–215. Available at: <https://doi.org/10.1093/ajcn/29.2.205>.
- Adibi, S.A. and Morse, E.L. (1977) 'The number of glycine residues which limits intact absorption of glycine oligopeptides in human jejunum', *The Journal of Clinical Investigation*, 60(5), pp. 1008–1016. Available at: <https://doi.org/10.1172/JCI108851>.
- Aduri, N.G. *et al.* (2015) 'Salt Bridge Swapping in the EXXERFXYY Motif of Proton-coupled Oligopeptide Transporters', *The Journal of Biological Chemistry*, 290(50), pp. 29931–29940. Available at: <https://doi.org/10.1074/jbc.M115.675603>.
- Afonine, P.V. *et al.* (2018) 'Real-space refinement in PHENIX for cryo-EM and crystallography', *Acta Crystallographica. Section D, Structural Biology*, 74(Pt 6), pp. 531–544. Available at: <https://doi.org/10.1107/S2059798318006551>.
- Alberts, B. (2022) *Molecular biology of the cell*. Seventh edition. New York: W. W. Norton & Company.
- Anfinsen, C.B. (1973) 'Principles that govern the folding of protein chains', *Science (New York, N.Y.)*, 181(4096), pp. 223–230. Available at: <https://doi.org/10.1126/science.181.4096.223>.
- Ayyadurai, S. *et al.* (2013) 'PepT1 expressed in immune cells has an important role in promoting the immune response during experimentally induced colitis', *Laboratory Investigation*, 93(8), pp. 888–899. Available at: <https://doi.org/10.1038/labinvest.2013.77>.
- Baek, M. *et al.* (2021) 'Accurate prediction of protein structures and interactions using a three-track neural network', *Science*, 373(6557), pp. 871–876. Available at: <https://doi.org/10.1126/science.abj8754>.
- Bayburt, T.H. and Sligar, S.G. (2010) 'Membrane protein assembly into Nanodiscs', *FEBS letters*, 584(9), pp. 1721–1727. Available at: <https://doi.org/10.1016/j.febslet.2009.10.024>.
- Beale, J.H. *et al.* (2015) 'Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport', *Structure (London, England: 1993)*, 23(10), pp. 1889–1899. Available at: <https://doi.org/10.1016/j.str.2015.07.016>.

Berger, U.V. and Hediger, M.A. (1999) 'Distribution of peptide transporter PEPT2 mRNA in the rat nervous system', *Anatomy and Embryology*, 199(5), pp. 439–449. Available at: <https://doi.org/10.1007/s004290050242>.

Bosshart, P.D. and Fotiadis, D. (2019) 'Secondary Active Transporters', *Sub-Cellular Biochemistry*, 92, pp. 275–299. Available at: https://doi.org/10.1007/978-3-030-18768-2_9.

Botte, M. *et al.* (2022) 'Cryo-EM structures of a LptDE transporter in complex with Pro-macrobodies offer insight into lipopolysaccharide translocation', *Nature Communications*, 13(1), p. 1826. Available at: <https://doi.org/10.1038/s41467-022-29459-2>.

Bragg, L. (1956) 'The diffraction of x-rays', *The British Journal of Radiology*, 29(339), pp. 121–126. Available at: <https://doi.org/10.1259/0007-1285-29-339-121>.

Brandsch, M. (2009) 'Transport of drugs by proton-coupled peptide transporters: pearls and pitfalls', *Expert Opinion on Drug Metabolism & Toxicology*, 5(8), pp. 887–905. Available at: <https://doi.org/10.1517/17425250903042292>.

Brandsch, M., Knütter, I. and Bosse-Doenecke, E. (2008) 'Pharmaceutical and pharmacological importance of peptide transporters', *The Journal of Pharmacy and Pharmacology*, 60(5), pp. 543–585. Available at: <https://doi.org/10.1211/jpp.60.5.0002>.

Brandsch, M., Knütter, I. and Leibach, F.H. (2004) 'The intestinal H⁺/peptide symporter PEPT1: structure-affinity relationships', *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, 21(1), pp. 53–60. Available at: [https://doi.org/10.1016/s0928-0987\(03\)00142-8](https://doi.org/10.1016/s0928-0987(03)00142-8).

Brodin, B. *et al.* (2002) 'Transport of peptidomimetic drugs by the intestinal Di/tri-peptide transporter, PepT1', *Pharmacology & Toxicology*, 90(6), pp. 285–296. Available at: <https://doi.org/10.1034/j.1600-0773.2002.900601.x>.

Bröer, S. (2009) 'The role of the neutral amino acid transporter B⁰ AT1 (SLC6A19) in Hartnup disorder and protein nutrition', *IUBMB Life*, 61(6), pp. 591–599. Available at: <https://doi.org/10.1002/iub.210>.

Bröer, S., Cavanaugh, J.A. and Rasko, J.E.J. (2005) 'Neutral amino acid transport in epithelial cells and its malfunction in Hartnup disorder', *Biochemical Society Transactions*, 33(Pt 1), pp. 233–236. Available at: <https://doi.org/10.1042/BST0330233>.

Bröer, S. and Fairweather, S.J. (2018) 'Amino Acid Transport Across the Mammalian Intestine', in R. Terjung (ed.) *Comprehensive Physiology*. 1st edn. Wiley, pp. 343–373. Available at: <https://doi.org/10.1002/cphy.c170041>.

Caffrey, M. (2015) 'A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes', *Acta Crystallographica. Section F, Structural Biology Communications*, 71(Pt 1), pp. 3–18. Available at: <https://doi.org/10.1107/S2053230X14026843>.

Casares, D., Escribá, P.V. and Rosselló, C.A. (2019) 'Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic

Avenues', *International Journal of Molecular Sciences*, 20(9), p. E2167. Available at: <https://doi.org/10.3390/ijms20092167>.

Celie, P.H., Parret, A.H. and Perrakis, A. (2016) 'Recombinant cloning strategies for protein expression', *Current Opinion in Structural Biology*, 38, pp. 145–154. Available at: <https://doi.org/10.1016/j.sbi.2016.06.010>.

Charrière, G.M. *et al.* (2010) 'Identification of *Drosophila* Yin and PEPT2 as Evolutionarily Conserved Phagosome-associated Muramyl Dipeptide Transporters', *Journal of Biological Chemistry*, 285(26), pp. 20147–20154. Available at: <https://doi.org/10.1074/jbc.M110.115584>.

Chen, X. *et al.* (2018) 'Influence of Peptide Transporter 2 (PEPT2) on the Distribution of Cefadroxil in Mouse Brain: A Microdialysis Study', p. 23.

Choy, B.C. *et al.* (2021) 'A 10-year meta-analysis of membrane protein structural biology: Detergents, membrane mimetics, and structure determination techniques', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1863(3), p. 183533. Available at: <https://doi.org/10.1016/j.bbamem.2020.183533>.

Croll, T.I. (2018) 'ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps', *Acta Crystallographica. Section D, Structural Biology*, 74(Pt 6), pp. 519–530. Available at: <https://doi.org/10.1107/S2059798318002425>.

Dalmasso, G. *et al.* (2010) 'PepT1 mediates transport of the proinflammatory bacterial tripeptide L-Ala- γ -D-Glu- meso -DAP in intestinal epithelial cells', *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 299(3), pp. G687–G696. Available at: <https://doi.org/10.1152/ajpgi.00527.2009>.

Daniel, H. and Adibi, S.A. (1993) 'Transport of beta-lactam antibiotics in kidney brush border membrane. Determinants of their affinity for the oligopeptide/H⁺ symporter', *The Journal of Clinical Investigation*, 92(5), pp. 2215–2223. Available at: <https://doi.org/10.1172/JCI116824>.

Daniel, H. and Kottra, G. (2004) 'The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology', *Pflugers Archiv: European Journal of Physiology*, 447(5), pp. 610–618. Available at: <https://doi.org/10.1007/s00424-003-1101-4>.

Dessau, M.A. and Modis, Y. (2011) 'Protein crystallization for X-ray crystallography', *Journal of Visualized Experiments: JoVE*, (47), p. 2285. Available at: <https://doi.org/10.3791/2285>.

Di Resta, C. and Becchetti, A. (2010) 'Introduction to ion channels', *Advances in Experimental Medicine and Biology*, 674, pp. 9–21. Available at: https://doi.org/10.1007/978-1-4419-6066-5_2.

Dobson, C.M. (2003) 'Protein folding and misfolding', *Nature*, 426(6968), pp. 884–890. Available at: <https://doi.org/10.1038/nature02261>.

Doki, S. *et al.* (2013) 'Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT', *Proceedings of the National Academy of Sciences of the United*

- States of America*, 110(28), pp. 11343–11348. Available at: <https://doi.org/10.1073/pnas.1301079110>.
- Drew, D. *et al.* (2021) ‘Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS)’, *Chemical Reviews*, 121(9), pp. 5289–5335. Available at: <https://doi.org/10.1021/acs.chemrev.0c00983>.
- Drew, D. and Boudker, O. (2016) ‘Shared Molecular Mechanisms of Membrane Transporters’, *Annual Review of Biochemistry*, 85, pp. 543–572. Available at: <https://doi.org/10.1146/annurev-biochem-060815-014520>.
- Drozdziak, M. *et al.* (2014) ‘Protein abundance of clinically relevant multidrug transporters along the entire length of the human intestine’, *Molecular Pharmaceutics*, 11(10), pp. 3547–3555. Available at: <https://doi.org/10.1021/mp500330y>.
- Dubochet, J. *et al.* (1988) ‘Cryo-electron microscopy of vitrified specimens’, *Quarterly Reviews of Biophysics*, 21(2), pp. 129–228. Available at: <https://doi.org/10.1017/s0033583500004297>.
- Dyla, M. *et al.* (2020) ‘Structure and Mechanism of P-Type ATPase Ion Pumps’, *Annual Review of Biochemistry*, 89, pp. 583–603. Available at: <https://doi.org/10.1146/annurev-biochem-010611-112801>.
- Elkon, K.B. and Briggs, T.A. (2020) ‘The (Orf)ull truth about IRF5 and type I interferons in SLE’, *Nature Reviews Rheumatology*, 16(10), pp. 543–544. Available at: <https://doi.org/10.1038/s41584-020-0472-7>.
- Engl, E. and Attwell, D. (2015) ‘Non-signalling energy use in the brain: Non-signalling energy use in the brain’, *The Journal of Physiology*, 593(16), pp. 3417–3429. Available at: <https://doi.org/10.1113/jphysiol.2014.282517>.
- Ernst, H.A. *et al.* (2009) ‘Ligand binding analyses of the putative peptide transporter YjdL from *E. coli* display a significant selectivity towards dipeptides’, *Biochemical and Biophysical Research Communications*, 389(1), pp. 112–116. Available at: <https://doi.org/10.1016/j.bbrc.2009.08.098>.
- Fei, Y.J. *et al.* (1994) ‘Expression cloning of a mammalian proton-coupled oligopeptide transporter’, *Nature*, 368(6471), pp. 563–566. Available at: <https://doi.org/10.1038/368563a0>.
- Frauenfeld, J. *et al.* (2016) ‘A saposin-lipoprotein nanoparticle system for membrane proteins’, *Nature Methods*, 13(4), pp. 345–351. Available at: <https://doi.org/10.1038/nmeth.3801>.
- Ganapathy, M.E. *et al.* (1995) ‘Differential recognition of beta -lactam antibiotics by intestinal and renal peptide transporters, PEPT 1 and PEPT 2’, *The Journal of Biological Chemistry*, 270(43), pp. 25672–25677. Available at: <https://doi.org/10.1074/jbc.270.43.25672>.
- Ganapathy, V. and Leibach, F.H. (1983) ‘Role of pH gradient and membrane potential in dipeptide transport in intestinal and renal brush-border membrane vesicles from the rabbit.

Studies with L-carnosine and glycyl-L-proline.’, *Journal of Biological Chemistry*, 258(23), pp. 14189–14192. Available at: [https://doi.org/10.1016/S0021-9258\(17\)43843-9](https://doi.org/10.1016/S0021-9258(17)43843-9).

Ganapathy, V. and Leibach, F.H. (1985) ‘Is intestinal peptide transport energized by a proton gradient?’, *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 249(2), pp. G153–G160. Available at: <https://doi.org/10.1152/ajpgi.1985.249.2.G153>.

Ganapathy, V. and Leibach, F.H. (1986) ‘Carrier-mediated reabsorption of small peptides in renal proximal tubule’, *The American Journal of Physiology*, 251(6 Pt 2), pp. F945–953. Available at: <https://doi.org/10.1152/ajprenal.1986.251.6.F945>.

Grimes, C.L. *et al.* (2012) ‘The Innate Immune Protein Nod2 Binds Directly to MDP, a Bacterial Cell Wall Fragment’, *Journal of the American Chemical Society*, 134(33), pp. 13535–13537. Available at: <https://doi.org/10.1021/ja303883c>.

Grotjohann, I. and Fromme, P. (2005) ‘Structure of cyanobacterial photosystem I’, *Photosynthesis Research*, 85(1), pp. 51–72. Available at: <https://doi.org/10.1007/s11120-005-1440-4>.

Guettou, F. *et al.* (2013) ‘Structural insights into substrate recognition in proton-dependent oligopeptide transporters’, *EMBO reports*, 14(9), pp. 804–810. Available at: <https://doi.org/10.1038/embor.2013.107>.

Harder, D. *et al.* (2008) ‘DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of *Escherichia coli*’, *The FEBS journal*, 275(13), pp. 3290–3298. Available at: <https://doi.org/10.1111/j.1742-4658.2008.06477.x>.

Hashmi, M.S. and Gupta, V. (2022) ‘Hartnup Disease’, in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK559076/> (Accessed: 24 September 2022).

Heinz, L.X. *et al.* (2020) ‘TASL is the SLC15A4-associated adaptor for IRF5 activation by TLR7-9’, *Nature*, 581(7808), pp. 316–322. Available at: <https://doi.org/10.1038/s41586-020-2282-0>.

Hellier, M.D. *et al.* (1972) ‘Dipeptide absorption in man’, *Gut*, 13(12), pp. 965–969. Available at: <https://doi.org/10.1136/gut.13.12.965>.

Henderson, R. (1995) ‘The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules’, *Quarterly Reviews of Biophysics*, 28(2), pp. 171–193. Available at: <https://doi.org/10.1017/s003358350000305x>.

Hu, Y. *et al.* (2008) ‘Targeted disruption of peptide transporter Pept1 gene in mice significantly reduces dipeptide absorption in intestine’, *Molecular Pharmaceutics*, 5(6), pp. 1122–1130. Available at: <https://doi.org/10.1021/mp8001655>.

Huang, C.Y. *et al.* (2016) ‘In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures’, *Acta Crystallographica. Section D, Structural Biology*, 72(Pt 1), pp. 93–112. Available at: <https://doi.org/10.1107/S2059798315021683>.

- Ilari, A. and Savino, C. (2008) ‘Protein structure determination by x-ray crystallography’, *Methods in Molecular Biology (Clifton, N.J.)*, 452, pp. 63–87. Available at: https://doi.org/10.1007/978-1-60327-159-2_3.
- Ingersoll, S.A. *et al.* (2012) ‘The role and pathophysiological relevance of membrane transporter PepT1 in intestinal inflammation and inflammatory bowel disease’, *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 302(5), pp. G484-492. Available at: <https://doi.org/10.1152/ajpgi.00477.2011>.
- Ingólfsson, H.I. *et al.* (2014) ‘Lipid Organization of the Plasma Membrane’, *Journal of the American Chemical Society*, 136(41), pp. 14554–14559. Available at: <https://doi.org/10.1021/ja507832e>.
- Jensen, J.M., Simonsen, F.C., *et al.* (2012) ‘Biophysical characterization of the proton-coupled oligopeptide transporter YjdL’, *Peptides*, 38(1), pp. 89–93. Available at: <https://doi.org/10.1016/j.peptides.2012.08.012>.
- Jensen, J.M., Ismat, F., *et al.* (2012) ‘Probing the putative active site of YjdL: an unusual proton-coupled oligopeptide transporter from *E. coli*’, *PloS One*, 7(10), p. e47780. Available at: <https://doi.org/10.1371/journal.pone.0047780>.
- Jumper, J. *et al.* (2021) ‘Highly accurate protein structure prediction with AlphaFold’, *Nature* [Preprint]. Available at: <https://doi.org/10.1038/s41586-021-03819-2>.
- Kamal, M.A. *et al.* (2009) ‘Influence of genetic knockout of Pept2 on the in vivo disposition of endogenous and exogenous carnosine in wild-type and Pept2 null mice’, 296, p. 6.
- Kamal, M.A., Keep, R.F. and Smith, D.E. (2010) ‘Role and Relevance of PEPT2 in Drug Disposition, Dynamics, and Toxicity’, p. 12.
- Karnovsky, M.J. *et al.* (1982) ‘The concept of lipid domains in membranes’, *The Journal of Cell Biology*, 94(1), pp. 1–6. Available at: <https://doi.org/10.1083/jcb.94.1.1>.
- Katsura, T. and Inui, K.-I. (2003) ‘Intestinal absorption of drugs mediated by drug transporters: mechanisms and regulation’, *Drug Metabolism and Pharmacokinetics*, 18(1), pp. 1–15. Available at: <https://doi.org/10.2133/dmpk.18.1>.
- Kendrew, J.C. *et al.* (1958) ‘A three-dimensional model of the myoglobin molecule obtained by x-ray analysis’, *Nature*, 181(4610), pp. 662–666. Available at: <https://doi.org/10.1038/181662a0>.
- Kermani, A.A. (2021) ‘A guide to membrane protein X-ray crystallography’, *The FEBS journal*, 288(20), pp. 5788–5804. Available at: <https://doi.org/10.1111/febs.15676>.
- Killer, M. *et al.* (2021) ‘Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes’, *Science Advances*, 7(45), p. eabk3259. Available at: <https://doi.org/10.1126/sciadv.abk3259>.
- Killer, M. *et al.* (2022) ‘Cryo-EM Structure of an Atypical Proton-Coupled Peptide Transporter: Di- and Tripeptide Permease C’, *Frontiers in Molecular Biosciences*, 9, p. 917725. Available at: <https://doi.org/10.3389/fmolb.2022.917725>.

- Kobayashi, T. *et al.* (2014) ‘The histidine transporter SLC15A4 coordinates mTOR-dependent inflammatory responses and pathogenic antibody production’, *Immunity*, 41(3), pp. 375–388. Available at: <https://doi.org/10.1016/j.immuni.2014.08.011>.
- Kühlbrandt, W. (2019) ‘Structure and Mechanisms of F-Type ATP Synthases’, *Annual Review of Biochemistry*, 88, pp. 515–549. Available at: <https://doi.org/10.1146/annurev-biochem-013118-110903>.
- Laroui, H. *et al.* (2011) ‘l-Ala- γ -d-Glu-meso-diaminopimelic Acid (DAP) Interacts Directly with Leucine-rich Region Domain of Nucleotide-binding Oligomerization Domain 1, Increasing Phosphorylation Activity of Receptor-interacting Serine/Threonine-protein Kinase 2 and Its Interaction with Nucleotide-binding Oligomerization Domain 1’, *Journal of Biological Chemistry*, 286(35), pp. 31003–31013. Available at: <https://doi.org/10.1074/jbc.M111.257501>.
- Lee, J. *et al.* (2009) ‘pH-dependent Internalization of Muramyl Peptides from Early Endosomes Enables Nod1 and Nod2 Signaling’, *Journal of Biological Chemistry*, 284(35), pp. 23818–23829. Available at: <https://doi.org/10.1074/jbc.M109.033670>.
- Li, F. *et al.* (2020) ‘Ion transport and regulation in a synaptic vesicle glutamate transporter’, *Science*, 368(6493), pp. 893–897. Available at: <https://doi.org/10.1126/science.aba9202>.
- Li, M. *et al.* (2006) ‘Interactions of amoxicillin and cefaclor with human renal organic anion and peptide transporters’, *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, 34(4), pp. 547–555. Available at: <https://doi.org/10.1124/dmd.105.006791>.
- Liu, S. *et al.* (2020) ‘Termini restraining of small membrane proteins enables structure determination at near-atomic resolution’, *Science Advances*, 6(51), p. eabe3717. Available at: <https://doi.org/10.1126/sciadv.abe3717>.
- Liu, S. *et al.* (2022) ‘Stabilization and structure determination of integral membrane proteins by termini restraining’, *Nature Protocols*, 17(2), pp. 540–565. Available at: <https://doi.org/10.1038/s41596-021-00656-5>.
- Liu, W. *et al.* (1995) ‘Molecular cloning of PEPT 2, a new member of the H⁺/peptide cotransporter family, from human kidney’, *Biochimica Et Biophysica Acta*, 1235(2), pp. 461–466. Available at: [https://doi.org/10.1016/0005-2736\(95\)80036-f](https://doi.org/10.1016/0005-2736(95)80036-f).
- Liu, X. (2019) ‘SLC Family Transporters’, *Advances in Experimental Medicine and Biology*, 1141, pp. 101–202. Available at: https://doi.org/10.1007/978-981-13-7647-4_3.
- Luckner, P. and Brandsch, M. (2005) ‘Interaction of 31 beta-lactam antibiotics with the H⁺/peptide symporter PEPT2: analysis of affinity constants and comparison with PEPT1’, *European Journal of Pharmaceutics and Biopharmaceutics: Official Journal of Arbeitsgemeinschaft Fur Pharmazeutische Verfahrenstechnik e.V.*, 59(1), pp. 17–24. Available at: <https://doi.org/10.1016/j.ejpb.2004.07.008>.
- Lyons, J.A. *et al.* (2014) ‘Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters’, *EMBO reports*, 15(8), pp. 886–893. Available at: <https://doi.org/10.15252/embr.201338403>.

- Martinez Molledo, M. *et al.* (2018) ‘Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter’, *Structure (London, England: 1993)*, 26(3), pp. 467–476.e4. Available at: <https://doi.org/10.1016/j.str.2018.01.005>.
- Martinez Molledo, M., Quistgaard, E.M. and Löw, C. (2018) ‘Tripeptide binding in a proton-dependent oligopeptide transporter’, *FEBS letters*, 592(19), pp. 3239–3247. Available at: <https://doi.org/10.1002/1873-3468.13246>.
- Mathews, D.M. and Adibi, S.A. (1976) ‘Peptide absorption’, *Gastroenterology*, 71(1), pp. 151–161.
- Matthews, D.M. (1975) ‘Intestinal absorption of peptides’, *Physiological Reviews*, 55(4), pp. 537–608. Available at: <https://doi.org/10.1152/physrev.1975.55.4.537>.
- McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) ‘The PSIPRED protein structure prediction server’, *Bioinformatics (Oxford, England)*, 16(4), pp. 404–405. Available at: <https://doi.org/10.1093/bioinformatics/16.4.404>.
- McPherson, A. and Cudney, B. (2014) ‘Optimization of crystallization conditions for biological macromolecules’, *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 11), pp. 1445–1467. Available at: <https://doi.org/10.1107/S2053230X14019670>.
- Minhas, G.S. *et al.* (2018) ‘Structural basis of malodour precursor transport in the human axilla’, *eLife*, 7, p. e34995. Available at: <https://doi.org/10.7554/eLife.34995>.
- Minhas, G.S. and Newstead, S. (2019) ‘Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(3), pp. 804–809. Available at: <https://doi.org/10.1073/pnas.1813715116>.
- Miyagi, H. *et al.* (2020) ‘The discovery of a new antibody for BRIL-fused GPCR structure determination’, *Scientific Reports*, 10(1), p. 11669. Available at: <https://doi.org/10.1038/s41598-020-68355-x>.
- Mueckler, M. and Thorens, B. (2013) ‘The SLC2 (GLUT) family of membrane transporters’, *Molecular Aspects of Medicine*, 34(2–3), pp. 121–138. Available at: <https://doi.org/10.1016/j.mam.2012.07.001>.
- Mukherjee, S. *et al.* (2020) ‘Synthetic antibodies against BRIL as universal fiducial marks for single-particle cryoEM structure determination of membrane proteins’, *Nature Communications*, 11(1), p. 1598. Available at: <https://doi.org/10.1038/s41467-020-15363-0>.
- Nagamura, R. *et al.* (2019) ‘Structural basis for oligomerization of the prokaryotic peptide transporter PepTSo2’, *Acta Crystallographica. Section F, Structural Biology Communications*, 75(Pt 5), pp. 348–358. Available at: <https://doi.org/10.1107/S2053230X19003546>.
- Nakamura, N. *et al.* (2014) ‘Endosomes are specialized platforms for bacterial sensing and NOD2 signalling’, *Nature*, 509(7499), pp. 240–244. Available at: <https://doi.org/10.1038/nature13133>.

- Nässl, A.-M. *et al.* (2011) ‘Amino acid absorption and homeostasis in mice lacking the intestinal peptide transporter PEPT1’, *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 301(1), pp. G128–G137. Available at: <https://doi.org/10.1152/ajpgi.00017.2011>.
- Newstead, S. *et al.* (2011) ‘Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2’, *The EMBO journal*, 30(2), pp. 417–426. Available at: <https://doi.org/10.1038/emboj.2010.309>.
- Newstead, S. (2017) ‘Recent advances in understanding proton coupled peptide transport via the POT family’, *Current Opinion in Structural Biology*, 45, pp. 17–24. Available at: <https://doi.org/10.1016/j.sbi.2016.10.018>.
- Nygaard, R., Kim, J. and Mancina, F. (2020) ‘Cryo-electron microscopy analysis of small membrane proteins’, *Current Opinion in Structural Biology*, 64, pp. 26–33. Available at: <https://doi.org/10.1016/j.sbi.2020.05.009>.
- Pardon, E. *et al.* (2014) ‘A general protocol for the generation of Nanobodies for structural biology’, *Nature Protocols*, 9(3), pp. 674–693. Available at: <https://doi.org/10.1038/nprot.2014.039>.
- Park, K.T. *et al.* (2020) ‘The Cost of Inflammatory Bowel Disease: An Initiative From the Crohn’s & Colitis Foundation’, *Inflammatory Bowel Diseases*, 26(1), pp. 1–10. Available at: <https://doi.org/10.1093/ibd/izz104>.
- Parker, J.L. *et al.* (2021) ‘Cryo-EM structure of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in mammals’, *Science Advances*, 7(35), p. eabh3355. Available at: <https://doi.org/10.1126/sciadv.abh3355>.
- Penczek, P.A. (2010) ‘Fundamentals of three-dimensional reconstruction from projections’, *Methods in Enzymology*, 482, pp. 1–33. Available at: [https://doi.org/10.1016/S0076-6879\(10\)82001-4](https://doi.org/10.1016/S0076-6879(10)82001-4).
- Pieprzyk, J., Pazicky, S. and Löw, C. (2018a) ‘Transient Expression of Recombinant Membrane-eGFP Fusion Proteins in HEK293 Cells’, *Methods in Molecular Biology (Clifton, N.J.)*, 1850, pp. 17–31. Available at: https://doi.org/10.1007/978-1-4939-8730-6_2.
- Pieprzyk, J., Pazicky, S. and Löw, C. (2018b) ‘Transient Expression of Recombinant Membrane-eGFP Fusion Proteins in HEK293 Cells’, *Methods in Molecular Biology (Clifton, N.J.)*, 1850, pp. 17–31. Available at: https://doi.org/10.1007/978-1-4939-8730-6_2.
- Powell, H.R. (2017) ‘X-ray data processing’, *Bioscience Reports*, 37(5), p. BSR20170227. Available at: <https://doi.org/10.1042/BSR20170227>.
- Powell, H.R. (2021) ‘A beginner’s guide to X-ray data processing’, *The Biochemist*, 43(3), pp. 46–50. Available at: https://doi.org/10.1042/bio_2021_124.
- Prabhala, B.K. *et al.* (2014) ‘New insights into the substrate specificities of proton-coupled oligopeptide transporters from *E. coli* by a pH sensitive assay’, *FEBS letters*, 588(4), pp. 560–565. Available at: <https://doi.org/10.1016/j.febslet.2014.01.004>.

- Prabhala, B.K. *et al.* (2017) ‘Several hPepT1-transported drugs are substrates of the Escherichia coli proton-coupled oligopeptide transporter YdgR’, *Research in Microbiology*, 168(5), pp. 443–449. Available at: <https://doi.org/10.1016/j.resmic.2017.01.005>.
- Prabhala, B.K. *et al.* (2018) ‘The prototypical proton-coupled oligopeptide transporter YdgR from Escherichia coli facilitates chloramphenicol uptake into bacterial cells’, *The Journal of Biological Chemistry*, 293(3), pp. 1007–1017. Available at: <https://doi.org/10.1074/jbc.M117.805960>.
- Punjani, A. *et al.* (2017) ‘cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination’, *Nature Methods*, 14(3), pp. 290–296. Available at: <https://doi.org/10.1038/nmeth.4169>.
- Punjani, A., Zhang, H. and Fleet, D.J. (2020) ‘Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction’, *Nature Methods*, 17(12), pp. 1214–1221. Available at: <https://doi.org/10.1038/s41592-020-00990-8>.
- Quistgaard, E.M. *et al.* (2016a) ‘Understanding transport by the major facilitator superfamily (MFS): structures pave the way’, *Nature Reviews. Molecular Cell Biology*, 17(2), pp. 123–132. Available at: <https://doi.org/10.1038/nrm.2015.25>.
- Quistgaard, E.M. *et al.* (2016b) ‘Understanding transport by the major facilitator superfamily (MFS): structures pave the way’, *Nature Reviews. Molecular Cell Biology*, 17(2), pp. 123–132. Available at: <https://doi.org/10.1038/nrm.2015.25>.
- Quistgaard, E.M., Martinez Molledo, M. and Löw, C. (2017) ‘Structure determination of a major facilitator peptide transporter: Inward facing PepTSt from Streptococcus thermophilus crystallized in space group P3121’, *PloS One*, 12(3), p. e0173126. Available at: <https://doi.org/10.1371/journal.pone.0173126>.
- Ramlaul, K. *et al.* (2020) ‘Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER’, *Journal of Structural Biology*, 211(2), p. 107545. Available at: <https://doi.org/10.1016/j.jsb.2020.107545>.
- Rimann, I. *et al.* (2022) ‘The solute carrier SLC15A4 is required for optimal trafficking of nucleic acid-sensing TLRs and ligands to endolysosomes’, *Proceedings of the National Academy of Sciences*, 119(14), p. e2200544119. Available at: <https://doi.org/10.1073/pnas.2200544119>.
- Rubio-Aliaga, I. *et al.* (2003) ‘Targeted disruption of the peptide transporter Pept2 gene in mice defines its physiological role in the kidney’, *Molecular and Cellular Biology*, 23(9), pp. 3247–3252. Available at: <https://doi.org/10.1128/MCB.23.9.3247-3252.2003>.
- Rubio-Aliaga, I. and Daniel, H. (2008) ‘Peptide transporters and their roles in physiological processes and drug disposition’, *Xenobiotica; the Fate of Foreign Compounds in Biological Systems*, 38(7–8), pp. 1022–1042. Available at: <https://doi.org/10.1080/00498250701875254>.
- Sanchez-Garcia, R. *et al.* (2020) *DeepEMhancer: a deep learning solution for cryo-EM volume post-processing*. preprint. Bioinformatics. Available at: <https://doi.org/10.1101/2020.06.12.148296>.

- Sasawatari, S. *et al.* (2011) ‘The solute carrier family 15A4 regulates TLR9 and NOD1 functions in the innate immune system and promotes colitis in mice’, *Gastroenterology*, 140(5), pp. 1513–1525. Available at: <https://doi.org/10.1053/j.gastro.2011.01.041>.
- Scheres, S.H.W. (2012) ‘RELION: implementation of a Bayesian approach to cryo-EM structure determination’, *Journal of Structural Biology*, 180(3), pp. 519–530. Available at: <https://doi.org/10.1016/j.jsb.2012.09.006>.
- Sezgin, E. *et al.* (2017) ‘The mystery of membrane organization: composition, regulation and roles of lipid rafts’, *Nature Reviews Molecular Cell Biology*, 18(6), pp. 361–374. Available at: <https://doi.org/10.1038/nrm.2017.16>.
- Sgro, G.G. and Costa, T.R.D. (2018) ‘Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis’, *Frontiers in Molecular Biosciences*, 5, p. 74. Available at: <https://doi.org/10.3389/fmolb.2018.00074>.
- Shah, S.C. and Itzkowitz, S.H. (2022) ‘Colorectal Cancer in Inflammatory Bowel Disease: Mechanisms and Management’, *Gastroenterology*, 162(3), pp. 715–730.e3. Available at: <https://doi.org/10.1053/j.gastro.2021.10.035>.
- Shen, H. *et al.* (1999) ‘Localization of PEPT1 and PEPT2 proton-coupled oligopeptide transporter mRNA and protein in rat kidney’, *The American Journal of Physiology*, 276(5), pp. F658–665. Available at: <https://doi.org/10.1152/ajprenal.1999.276.5.F658>.
- Shen, H. *et al.* (2004) ‘Immunolocalization of the proton-coupled oligopeptide transporter PEPT2 in developing rat brain’, *Molecular Pharmaceutics*, 1(4), pp. 248–256. Available at: <https://doi.org/10.1021/mp049944b>.
- Shen, J. *et al.* (2022) ‘Extracellular domain of PepT1 interacts with TM1 to facilitate substrate transport’, *Structure (London, England: 1993)*, pp. S0969-2126(22)00138–1. Available at: <https://doi.org/10.1016/j.str.2022.04.011>.
- Sigworth, F.J. (2016) ‘Principles of cryo-EM single-particle image processing’, *Microscopy (Oxford, England)*, 65(1), pp. 57–67. Available at: <https://doi.org/10.1093/jmicro/dfv370>.
- Smith, D.E., Johanson, C.E. and Keep, R.F. (2004) ‘Peptide and peptide analog transport systems at the blood–CSF barrier’, *Advanced Drug Delivery Reviews*, p. 27.
- Solcan, N. *et al.* (2012) ‘Alternating access mechanism in the POT family of oligopeptide transporters’, *The EMBO journal*, 31(16), pp. 3411–3421. Available at: <https://doi.org/10.1038/emboj.2012.157>.
- Spanier, B. and Rohm, F. (2018) ‘Proton Coupled Oligopeptide Transporter 1 (PepT1) Function, Regulation, and Influence on the Intestinal Homeostasis’, *Comprehensive Physiology*, 8(2), pp. 843–869. Available at: <https://doi.org/10.1002/cphy.c170038>.
- Stauffer, M. *et al.* (2022) ‘Peptide transporter structure reveals binding and action mechanism of a potent PEPT1 and PEPT2 inhibitor’, *Communications Chemistry*, 5(1), p. 23. Available at: <https://doi.org/10.1038/s42004-022-00636-0>.

Stelzl, T. *et al.* (2017) ‘Glycans in the intestinal peptide transporter PEPT1 contribute to function and protect from proteolysis’, *American Journal of Physiology. Gastrointestinal and Liver Physiology*, 312(6), pp. G580–G591. Available at: <https://doi.org/10.1152/ajpgi.00343.2016>.

Sugawara, M. *et al.* (2000) ‘Transport of valganciclovir, a ganciclovir prodrug, via peptide transporters PEPT1 and PEPT2’, *Journal of Pharmaceutical Sciences*, 89(6), pp. 781–789. Available at: [https://doi.org/10.1002/\(SICI\)1520-6017\(200006\)89:6<781::AID-JPS10>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1520-6017(200006)89:6<781::AID-JPS10>3.0.CO;2-7).

Sun, D. *et al.* (2013) ‘Functional and Molecular Expression of the Proton-Coupled Oligopeptide Transporters in Spleen and Macrophages from Mouse and Human’, *Molecular Pharmaceutics*, 10(4), pp. 1409–1416. Available at: <https://doi.org/10.1021/mp300700p>.

Sun, Y. *et al.* (2018) ‘PEPT1-mediated prodrug strategy for oral delivery of peramivir’, *Asian Journal of Pharmaceutical Sciences*, 13(6), pp. 555–565. Available at: <https://doi.org/10.1016/j.ajps.2018.05.008>.

Sung, J. *et al.* (2021) ‘PepT1-knockout mice harbor a protective metabolome beneficial for intestinal wound healing’, *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 320(5), pp. G888–G896. Available at: <https://doi.org/10.1152/ajpgi.00299.2020>.

Swaan, P.W. *et al.* (2008) ‘Bacterial Peptide Recognition and Immune Activation Facilitated by Human Peptide Transporter PEPT2’, *American Journal of Respiratory Cell and Molecular Biology*, 39(5), pp. 536–542. Available at: <https://doi.org/10.1165/rcmb.2008-0059OC>.

Terwilliger, T.C. *et al.* (2020) ‘Improvement of cryo-EM maps by density modification’, *Nature Methods*, 17(9), pp. 923–927. Available at: <https://doi.org/10.1038/s41592-020-0914-9>.

Tsutsumi, N. *et al.* (2020) ‘Structure of human Frizzled5 by fiducial-assisted cryo-EM supports a heterodimeric mechanism of canonical Wnt signaling’, *eLife*, 9, p. e58464. Available at: <https://doi.org/10.7554/eLife.58464>.

Ural-Blimke, Y. *et al.* (2019) ‘Structure of Prototypic Peptide Transporter DtpA from *E. coli* in Complex with Valganciclovir Provides Insights into Drug Binding of Human PepT1’, *Journal of the American Chemical Society*, 141(6), pp. 2404–2412. Available at: <https://doi.org/10.1021/jacs.8b11343>.

Varma, M.V. *et al.* (2010) ‘Targeting intestinal transporters for optimizing oral drug absorption’, *Current Drug Metabolism*, 11(9), pp. 730–742. Available at: <https://doi.org/10.2174/138920010794328850>.

Viennois, E. *et al.* (2016) ‘Critical role of PepT1 in promoting colitis-associated cancer and therapeutic benefits of the anti-inflammatory PepT1-mediated tripeptide KPV in a murine model’, *Cellular and Molecular Gastroenterology and Hepatology*, 2(3), pp. 340–357. Available at: <https://doi.org/10.1016/j.jcmgh.2016.01.006>.

Waghray, D. and Zhang, Q. (2018) 'Inhibit or Evade Multidrug Resistance P-Glycoprotein in Cancer Treatment', *Journal of Medicinal Chemistry*, 61(12), pp. 5108–5121. Available at: <https://doi.org/10.1021/acs.jmedchem.7b01457>.

Walian, P., Cross, T.A. and Jap, B.K. (2004) 'Structural genomics of membrane proteins', *Genome Biology*, 5(4), p. 215. Available at: <https://doi.org/10.1186/gb-2004-5-4-215>.

Wang, H.-W. and Wang, J.-W. (2017) 'How cryo-electron microscopy and X-ray crystallography complement each other', *Protein Science: A Publication of the Protein Society*, 26(1), pp. 32–39. Available at: <https://doi.org/10.1002/pro.3022>.

Wang, N. *et al.* (2021) 'Structural basis of human monocarboxylate transporter 1 inhibition by anti-cancer drug candidates', *Cell*, 184(2), pp. 370–383.e13. Available at: <https://doi.org/10.1016/j.cell.2020.11.043>.

Wang, X.-X. *et al.* (2017) 'A novel role for PHT1 in the disposition of l-histidine in brain: In vitro slice and in vivo pharmacokinetic studies in wildtype and Pht1 null mice', *Biochemical Pharmacology*, 124, pp. 94–102. Available at: <https://doi.org/10.1016/j.bcp.2016.11.012>.

Wang, Y. *et al.* (2014) 'Expression and Regulation of the Proton-Coupled Oligopeptide Transporter PhT2 by LPS in Macrophages and Mouse Spleen', *Molecular Pharmaceutics*, 11(6), pp. 1880–1888. Available at: <https://doi.org/10.1021/mp500014r>.

Weissenberger, G., Henderikx, R.J.M. and Peters, P.J. (2021) 'Understanding the invisible hands of sample preparation for cryo-EM', *Nature Methods*, 18(5), pp. 463–471. Available at: <https://doi.org/10.1038/s41592-021-01130-6>.

Weitz, D. *et al.* (2007) 'Functional and structural characterization of a prokaryotic peptide transporter with features similar to mammalian PEPT1', *The Journal of Biological Chemistry*, 282(5), pp. 2832–2839. Available at: <https://doi.org/10.1074/jbc.M604866200>.

Wlodawer, A. (2021) 'Protein crystallography: alive and well', *The FEBS journal*, 288(20), pp. 5786–5787. Available at: <https://doi.org/10.1111/febs.15822>.

Yan, R. *et al.* (2019) 'Structure of the human LAT1–4F2hc heteromeric amino acid transporter complex', *Nature*, 568(7750), pp. 127–130. Available at: <https://doi.org/10.1038/s41586-019-1011-z>.

Yang, B. and Smith, D.E. (2013) 'Significance of peptide transporter 1 in the intestinal permeability of valacyclovir in wild-type and PepT1 knockout mice', *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, 41(3), pp. 608–614. Available at: <https://doi.org/10.1124/dmd.112.049239>.

Yao, X., Fan, X. and Yan, N. (2020) 'Cryo-EM analysis of a membrane protein embedded in the liposome', *Proceedings of the National Academy of Sciences*, 117(31), pp. 18497–18503. Available at: <https://doi.org/10.1073/pnas.2009385117>.

Zair, Z.M. *et al.* (2008) 'Pharmacogenetics of OATP (*SLC21* / *SLCO*), OAT and OCT (*SLC22*) and PEPT (*SLC15*) transporters in the intestine, liver and kidney', *Pharmacogenomics*, 9(5), pp. 597–624. Available at: <https://doi.org/10.2217/14622416.9.5.597>.

Zhang, K. *et al.* (2022) ‘Fusion protein strategies for cryo-EM study of G protein-coupled receptors’, *Nature Communications*, 13(1), p. 4366. Available at: <https://doi.org/10.1038/s41467-022-32125-2>.

Zhang, Q. and Cherezov, V. (2019) ‘Chemical tools for membrane protein structural biology’, *Current Opinion in Structural Biology*, 58, pp. 278–285. Available at: <https://doi.org/10.1016/j.sbi.2019.06.002>.

Zhang, Y., Werling, U. and Edelman, W. (2014) ‘Seamless Ligation Cloning Extract (SLiCE) cloning method’, *Methods in Molecular Biology (Clifton, N.J.)*, 1116, pp. 235–244. Available at: https://doi.org/10.1007/978-1-62703-764-8_16.

Zhao, Y. *et al.* (2014) ‘Crystal structure of the E. coli peptide transporter YbgH’, *Structure (London, England: 1993)*, 22(8), pp. 1152–1160. Available at: <https://doi.org/10.1016/j.str.2014.06.008>.

Zimmermann, I. *et al.* (2018) ‘Synthetic single domain antibodies for the conformational trapping of membrane proteins’, *eLife*, 7, p. e34317. Available at: <https://doi.org/10.7554/eLife.34317>.

Zimmermann, I. *et al.* (2020) ‘Generation of synthetic nanobodies against delicate proteins’, *Nature Protocols*, 15(5), pp. 1707–1741. Available at: <https://doi.org/10.1038/s41596-020-0304-x>.

Zivanov, J. *et al.* (2018) ‘New tools for automated high-resolution cryo-EM structure determination in RELION-3’, *eLife*, 7, p. e42166. Available at: <https://doi.org/10.7554/eLife.42166>.

10 Contributions

Bourenkov Gleb: P13 and P14 MX beamline set-up and assistance.

Cazey Cornelia: Cryo-EM training, microscope operation assistance.

Custodio Tania: PHT1 project main investigator. PHT1 protein production and characterization. Selection of Sb27. Assistance in designing Mb26 construct.

Dupeux Florine: Preparation of crystallisation plates of DtpB-Nb132-peptides

Finocchio Giada: Cloning of Mb26, split sfGFP-DtpC constructs, split sfGFP-HsPepT1 constructs. Production, purification, and co-crystallisation of DtpB-Nb132-peptides.

Preparation of the crystallisation plates, crystal harvesting, data collection, and analysis.

Crystallisation trials on DtpC and split sfGFP-DtpC constructs.

Guedez Gabriela: Preparation of cell pellets of HsPepT1.

Guenther Christian: Training and assistance in Vapour diffusion and LCP crystallisation methods.

Humm Anne-Sophie: Preparation of crystallisation plates of DtpB-Nb132-peptides

Lei Jian: Selection of Nb132. Initial structure determination of DtpB-Nb132, in complex with Ala-Leu-Ala. Production, purification, and co-crystallisation of DtpB-Nb132-peptides.

Löw Christian: Principal investigator and supervisor, in all projects presented in this thesis.

Lugmayr Wolfgang: Training and assistance in SPA cryo-EM image analysis.

Marquez Jose: Supervision in DtpB-Nb132-peptides co-crystallisation trials.

Marlovits Thomas : Supervision in HsPepT2 structure determination.

Mertens Haydyn: SAXS on DtpC-Nb26

Pardon Els: Generation of nanobodies raised against DtpB and DtpC

Pieprzyk Joanna: Cloning of HsPepT2, preparation of cell pellets of HsPepT2, cellular uptake assays on HsPepT2. Supervision on HsPepT2 production and purification.

Pompidor Guillaume: P13 MX beamline set-up and assistance.

Seuring Carolin: Cryo-EM training, microscope operation assistance.

Steyaert Jan: Generation of nanobodies raised against DtpB and DtpC

Svergun Dimitri: SAXS on DtpC-Nb26

Wald Jiri: Supervision in HsPepT2 structure determination.

Weis Felix: Cryo-EM training, microscope operation assistance.

11 Acknowledgements

I thank Christian Löw, for allowing me to work on this project, for his availability, supervision, and sympathy.

I express my gratitude to the numerous people listed in the previous section, who contributed to the presented results, for their time and efforts. I also thank past and present members of the group of Christian Löw, for their kindness and advices.

I thank the members of the Thesis Advisory Committee and the Thesis Defence Committee of this dissertation: Britta Brügger, Thomas Schneider, Wojciech Galej, Christian Löw, Sebastian Eustermann and Irmgard Sinning ; for their time, guidance and support.

I thank the EMBL international PhD program, as well as the graduate office, for funding and support, during this work.

Finally, I thank my family and friends, for their moral support.