Cryo-EM structure of an atypical proton-coupled peptide transporter: Di- and tripeptide permease C

Maxime Killer[1,2], Giada Finocchio[1,2], Haydyn D.T. Mertens[2], Dmitri I. Svergun[2], Els Pardon[3,4], Jan Steyaert[3,4], Christian Löw[1,2,*]

[1] Centre for Structural Systems Biology (CSSB), Notkestrasse 85, D-22607 Hamburg, Germany.

[2] Molecular Biology Laboratory (EMBL), Hamburg Unit c/o Deutsches Elektronen Synchrotron (DESY), Notkestrasse 85, D-22607 Hamburg, Germany.

[3] Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels 1050, Belgium.

[4] VIB-VUB Center for Structural Biology, VIB, Brussels 1050, Belgium.

*** Correspondence:**
Corresponding Author
Christian Löw
European Molecular Biology Laboratory Hamburg, Notkestrasse 85, D-22607 Hamburg, Germany.
phone: +49 40 8998 87570
e-mail: christian.loew@embl-hamburg.de

Keywords: peptide transporter, SLC15, proton-dependent oligopeptide transporter (POT), nanobody, macrobody, single particle cryo-EM, DtpC, YjdL

Abstract

Proton-coupled Oligopeptide Transporters (POTs) of the Major Facilitator Superfamily (MFS) mediate the uptake of short di- and tripeptides in all phyla of life. POTs are thought to constitute the most promiscuous class of MFS transporters, with the potential to transport more than 8400 unique substrates. Over the past two decades, transport assays and biophysical studies have shown that various orthologues and paralogues display differences in substrate selectivity. The *E. coli* genome codes for four different POTs, known as Di- and Tripeptide permeases A-D (DtpA-D). DtpC was shown previously to favor positively charged peptides as substrates. In this study, we describe, how we determined the structure of the 53 kDa DtpC by cryogenic electron microscopy (cryo-EM), and provide structural insights into the ligand specificity of this atypical POT. We collected and analyzed data on the transporter fused to split superfolder GFP (split sfGFP), in complex with a 52 kDa macrobody and with a 13 kDa nanobody. The latter sample was more stable, rigid and a significant fraction dimeric, allowing us to reconstruct a 3D volume of DtpC at a resolution of 2.7 Å. This work provides a molecular explanation for the selectivity of DtpC, and highlights the value of small and rigid fiducial markers such as nanobodies for structure determination of low molecular weight integral membrane proteins lacking soluble domains.

1 Introduction

Membranes of cells compartmentalize metabolic processes and present a selective barrier for permeation. To preserve the characteristic intracellular milieu, membrane transporters with specialized functions have evolved to maintain the nutrient homeostasis of cells (Hediger et al., 2013; Zhang et al., 2019). Many of those are energized by an electrochemical proton gradient, providing a powerful driving force for transport and accumulation of nutrients above extracellular concentrations. Proton-dependent oligopeptide transporters (POTs) of the Solute Carrier 15 family (SLC15) are representatives of such secondary active transport systems and occur in all living organisms. They allow an efficient uptake of peptides and amino acids in bulk quantities (Daniel et al., 2006; Thwaites and Anderson, 2007). The best characterized members are the two mammalian PepT1 and PepT2 transporters which are known to play crucial roles in human health, being responsible for the uptake and distribution of nutrients such as di- and tripeptides (Brandsch et al., 2004; Smith et al., 2013; Spanier and Rohm, 2018; Viennois et al., 2018). They also play key roles in human diseases, and impact the pharmacokinetic profiles of orally administered drug molecules (Daniel, 2004; Brandsch, 2009; Ingersoll et al., 2012; Hillgren et al., 2013; Colas et al., 2017; Heinz et al., 2020). SLC15 transporters belong to the Major facilitator superfamily (MFS). MFS transporters share a well-characterized fold, consisting of twelve transmembrane helices organized in two six-helix bundles, expected to function according to the alternate access mechanisms (Jardetzky, 1966) where either side of the transporter is alternately exposed to one side of the membrane. Therefore, substantial conformational changes are required to complete an entire transport cycle with at least three postulated states: (i) inward-open, (ii) occluded, and (iii) outward-open (Yan, 2015; Drew and Boudker, 2016; Quistgaard et al., 2016; Bartels et al., 2021; Drew et al., 2021). POTs have been intensively studied on a structural and biochemical level over the last 30 years. More than 50 entries for this transporter class can be found in the protein data bank, representing ten different bacterial homologues and the mammalian PepT1 and PepT2 transporters, bound to a limited set of substrates and drugs (Newstead et al., 2011; Solcan et al., 2012; Doki et al., 2013; Guettou et al., 2013, 2014; Lyons et al., 2014; Zhao et al., 2014; Quistgaard et al., 2017; Martinez Molledo et al., 2018a, 2018b; Minhas et al., 2018; Nagamura et al., 2019; Ural-Blimke et al., 2019; Killer et al., 2021; Parker et al., 2021; Stauffer et al., 2022). Although bacterial and eukaryotic POTs share an overall conserved binding site, individual amino acids changes in or in close vicinity of the binding site are likely responsible for observed differences in affinities and selectivity for particular peptides and drugs among the studied POT homologues. Here, structural biology studies are particularly crucial to understand substrate promiscuity and drug coordination on a molecular level. While bacterial POT structures, determined by mainly X-ray crystallography, represent exclusively the inward-open or inward-open-partially occluded state, the mammalian PepT1 and PepT2 transporters were recently captured in various conformations by single particle cryo-EM, advancing the mechanistic understanding of the entire transport cycle (Killer et al., 2021; Parker et al., 2021). Despite their small size of typically only 50 kDa for an individual transporter unit, these systems become more and more accessible for single-particle cryo-EM approaches. Indeed, in 2021, more MFS transporter structures were determined by single-particle Cryo-EM (17 pdb entries; resolution range 3.0 – 4.2 Å) than X-ray crystallography (14 pdb entries; resolution range 1.8 – 3.6 Å).

Although known POT structures show a high level of similarity, various works have indicated that homologues can differ in their range of transported substrate and drug molecules (Lyons et al., 2014; Prabhala et al., 2014; Boggavarapu et al., 2015; Sharma et al., 2016; Martinez Molledo et al., 2018a). The *E. coli* genome codes for four different POTs named Di- and Tripeptide permease A-D (DtpA-D), also known as YdgR (=DtpA), YhiP (=DtpB), YjdL (=DtpC) and YbgH (=DtpD). They cluster in pairs, DtpA and B (sequence identity 51%), and DtpC and D (sequence identity 56%) with around 25% identity between them. DtpA and B exhibit a prototypical substrate preference similar to

the human PepT1 transporter (Chen et al., 2000; Harder et al., 2008; Foley et al., 2010; Prabhala et al., 2017, 2018), while DtpC and D have been classified as atypical POTs, because DtpC prefers di-peptides in particular those with a lysine residue in the second position. Although DtpC has been well characterized in terms of function over the last years (Ernst et al., 2009; Jensen et al., 2012c, 2012a, 2012b, 2014; Prabhala et al., 2014; Aduri et al., 2015), it has resisted structure determination by X-ray crystallography so far (Gabrielsen et al., 2011).

Here we describe the structure determination of the bacterial POT transporter DtpC by single particle cryo-EM. Considering that the transporter displays no characteristic cytoplasmic or periplasmic features which are helpful to drive the particle alignment, we applied different strategies previously described in the literature to increase the overall size of the transporter to overcome these limitations. We i) fused the transporter to split-sfGFP (Liu et al., 2020, 2022), ii) raised different nanobodies against DtpC (Pardon et al., 2014) and iii) extended the nanobody to a macrobody (Brunner et al., 2020; Botte et al., 2021). The various samples were subsequently imaged by cryo-EM and analysed. DtpC in complex with the conformation specific nanobody 26 turned out to be more rigid and a significant fraction of the sample dimeric, allowing us to reconstruct DtpC to 2.7 Å resolution. The DtpC structure now provides molecular insights into how selectivity within this transporter family is achieved.

2 Results and Discussion

Different fiducial marker strategies for structure determination

Since MFS transporters typically lack additional domains outside their transport unit, which is a major impediment for accurate particle alignment in single particle cryo-EM approaches, we assessed three fiducial marker strategies introducing additional density outside of detergent micelles containing DtpC, by analyzing the quality of 2D class averages (Fig 1). To obtain conformation specific binders against DtpC, we first immunized llamas with recombinant DtpC and selected nanobodies (Nbs) following standard procedures (Pardon et al., 2014). Three out of five selected binders (Nb17, Nb26, and Nb38) co-eluted with DtpC on gel filtration and increased the melting temperature of the respective DtpC-Nb complex by 20 °C, 16 °C, and 12 °C. (Fig 2 A,B). DtpC in complex with Nb17 and Nb26 yielded crystals in various conditions, but despite extensive optimization efforts, the crystals of the DtpC-Nb26 complex did not diffract X-rays better than 5 Å resolution. In a second step, we decided to increase the size of Nb26, which formed a tight complex with DtpC, by fusing one copy of the maltose binding protein (MBP) to its C-terminus as described previously (Botte et al., 2021)**.** This resulted in a 52 kDa macrobody, and we expected it to bind to the periplasmic side of the transporter as seen in other MFS transporter-Nb complexes (Fig 1). In a third approach, we fused the two self-assembling parts of split-sfGFP; with β1-6 on the N-terminus of DtpC, and β7-11 on the C-terminus. We named this construct split sfGFP-DtpC$_{FL}$. In order to minimize the mobility between the membrane protein and the split sfGFP fiducial, we also generated two additional constructs where the last five (split sfGFP-DtpC$_{1-475}$), or ten residues (split sfGFP-DtpC$_{1-470}$) of the transporter were deleted. We then assessed proper folding and complementation by monitoring the fluorescence of the chromophore on an HPLC system (Fig 2C). All constructs eluted at similar retention times and the fluorescence was highest in the non-truncated construct (split sfGFP-DtpC$_{FL}$) and lowest in the most truncated version (split sfGFP-DtpC$_{1-470}$). In order to extend this observation to other MFS transporters, we repeated this experiment with the human POT homologue PepT1, and noticed a similar trend upon shortening of the

termini. Yet, since the decrease of fluorescence was only minor in split sfGFP-DtpC$_{1-475}$ in comparison to split sfGFP-DtpC$_{FL}$, we proceeded to imaging with the shorter construct in the presence of Nb26.

The particle density and distribution in the vitrified solution was similar in the three imaged samples. However, DtpC-Nb26 produced the best 2D class averages considering the sharpness of secondary structure elements inside the micelle, as judged by visual inspection (Fig 1, Fig S1). The Nb26-MBP (Mb26) fiducial was clearly visible in 2D class averages, but it adopted various positions in relation to the transporter, therefore making accurate alignment of the particles more difficult than in its shorter but more rigid and stable nanobody version (Fig 1, Fig 2 A,B, Fig S1). The split sfGFP-DtpC$_{(1-475)}$-Nb26 sample allowed clear visualization of the transmembrane helices after clustering a small subset of particles, but the majority of particles clustered in classes with blurry density for the split sfGFP fiducial, or with the two complementary parts β1-6 and β7-11 not assembled (Fig S1). Alphafold2 predictions on the imaged construct, as well as on the full length construct later suggested a destabilization of the beta-barrel upon increasing termini restrains, resulting in partial unfolding of β7 and exposure of the chromophore to solvent quenching. Interestingly, this effect could partially be reverted by adding a linker of five glycine residues between the C-bundle and β7 based on *in silico* data. We conclude that termini restraining using the split-sfGFP approach is a promising fiducial strategy for structural studies of MFS transporters, in addition to the previous demonstrated showcases on small membrane proteins (2, 4 and 6TMs) (Liu et al., 2020, 2022). However, the amount of restraining in larger membrane proteins such as MFS transporters where both termini are placed far from each other need to be optimized experimentally or *in silico*, to produce a stable and rigid fiducial; two crucial aspects for high resolution structure determination of MFS transporters by single particle cryo-EM.

As we obtained the best 2D class averages for DtpC with the fiducial marker Nb26, we proceeded to a large data collection (Table 1) and could cluster a subset of dimers within this data set. The presence of different oligomeric species was already expected based on the peak shape of the gel filtration run. The large mass of the dimer, and the stable and rigid signal of the Nb26 fiducial, allowed us to reconstruct the DtpC-Nb26 dimer to 3.0 Å resolution and model this assembly (Fig 3, 4, Fig S2). The quaternary structure consists of a non-symmetrical inverted dimer mediated by interactions through a large hydrophobic interface between the HA-HB helices of DtpC (Fig S2). Although other inverted dimers were reported in homologous POT structures (Quistgaard et al., 2017), the source of such arrangements is likely to be artificial. We also investigated the oligomer heterogeneity in solution with small angle X-ray scattering and obtained a good fit at low angles (corresponding to the overall shape of particles in solution) for the cryo-EM volume of the dimer (Fig S3). The fit to a monomeric cryo-EM volume was poor, indicating that in detergent solution a significant fraction of DtpC-Nb26 is dimeric. As for the interaction between the membrane protein and the fiducial marker, the CDR3 loop of Nb26 accounts for the strongest interactions with the periplasmic surface of the transporter with two salt bridges, while CDR1 and CDR2 contribute *via* hydrogen bounding (Fig 5). 3D variability analysis (Punjani and Fleet, 2021) revealed a small degree of flexibility between the two DtpC-Nb26 copies. Therefore, we performed a local refinement, focused on one copy of the membrane protein, which extended the resolution of the reconstruction to 2.7 Å and improved the accuracy of the atomic model for subsequent structural analysis (Fig 4).

Structural basis for ligand selectivity in DtpC

The DtpC structure revealed the expected and well-known MFS transporter fold, with twelve transmembrane helices (TMs) organized in two helical bundles and additional two TMs specific for the POT family (known as HA and HB domains). The peptide binding site of DtpC is exposed to the

cytoplasmic side (Fig 3, 4). Almost all bacterial POT structures described so far were determined by X-ray methods in a similar inward facing (IF) conformation. The extent to which the central cavity is open to the cytosol is regulated by a mechanism of occlusion mediated by TM4, TM5, TM10, and TM11, as supported by structures in IF occluded, partially occluded, and open states. In the case of the here described DtpC structure, the IF state is open (Fig 3).

Molecules from the periplasmic side, on the contrary, cannot enter the central cavity. Tight closure of both bundles above the binding site is mediated by a salt bridge between D43 (TM2, N-bundle) and R294 (TM7, C-bundle) and hydrogen bounds between H37 (TM1, N-bundle) and D293 (TM7) as well as R28 (TM1) and N421 (TM11, C-bundle) (Fig 6A). We also analyzed previously determined POT structures with clearly resolved side chain densities, to understand how the IF state is generally maintained in this transporter family. Except for human PepT2 and the POT transporter from *Shewanella oneidensis* ($PepT_{So}$), where the inter-bundle periplasmic salt bridge is formed between TM5 and TM7, the IF state is in all other analyzed structures stabilized by a salt bridge on the tip of TM2 and TM7 (Fig 6B). Additional hydrogen bonding networks as described in other studies, can occur, but vary greatly among different homologues. This analysis highlights that the alternate access mechanism in canonical and in so called 'atypical' POTs share similarities such as electrostatic clamping by formation and disruption of salt bridges. The differences in hydrogen bonding patterns however, could account for the various turnover rates seen among POT homologues.

Canonical POTs are characterized by i) the presence of the E_1XXE_2R motif on TM1 involved in proton coupling and ligand binding, and ii) the ability to accommodate dipeptides, tripeptides, and peptidomimetics, which relies on a set of conserved residues located in the central binding cavity. In DtpC, the E_1XXE_2R motif, has evolved to Q_1XXE_2Y (where Q_1=N17, E_2=E20, Y=Y21). In all high resolution X-ray structures of canonical POTs, R is in salt-bridge distance to E_2 and the C-terminus of substrate peptides. Mutation of either E_1 or E_2 in the conventional E_1XXE_2R motif to glutamine residues abolishes uptake (Aduri et al., 2015). A reverse mutation in DtpC, from Q_1XXE_2Y to E_1XXE_2Y or to E_1XXQ_2Y preserves high transport rates, while a Q_1XXQ_2Y motif significantly decreases it (Aduri et al., 2015). In addition, based on previous molecular dynamics experiments, a salt bridge switching mechanism from R-E_2 to R-E_1, upon protonation of E_2 in the E_1XXE_2R motif, was proposed (Aduri et al., 2015). This biochemical and *in silico* data strongly support a dual role of the E_1XXE_2R motif for both proton and peptide transport, where R can form a salt bridge interaction with the C-terminus of peptides or with E_1 when E_2 is protonated, and where the deprotonation event of the latter is required to disrupt the R-peptide interaction.

In DtpC, we now observe that the side chain pocket has a different architecture and characteristic in comparison with the one of canonical POTs. It displays an overall more acidic groove caused by the presence of the aspartate residue 392. Canonical POTs have a conserved serine residue instead, yielding a slightly changed hydrophobicity pattern in the binding site (Fig 7 A-D). A structural overlay of DtpC with a canonical POT structure bound to the dipeptide Ala-Phe allows us to position the peptide in the binding site. By replacing the phenyl group with a lysine side chain (generating the known DtpC dipeptide substrate Ala-Lys instead of Ala-Phe), we postulate a putative salt bridge between the carboxyl group of D392 and the ε-amino group of the lysine side chain. This observation, together with previous biochemical work (Jensen et al., 2012b; Aduri et al., 2015) allows us to hypothesize that the selectivity of DtpC for dipeptides with C-terminal lysine or arginine residues is caused by swapping a salt bridge between the recurrent carboxyl group of the peptide terminus and the transporter (R21Y mutation), to a side chain specific salt bridge with D392. Since the R-peptide interaction is lost in DtpC, there is no requirement for E1 to destabilize R-peptide for release, which would explain the presence of a Q_1XXE_2Y motif instead of E_1XXE_2R.

In summary, our work provides new insights into promiscuous *versus* selective substrate recognition in POTs and constitutes a step forward towards completing the family of *E. coli* POTs

structures. Lastly, it displays some of the challenges related to high resolution cryo-EM structure determination of MFS transporters devoid of soluble domains, and manifests once again, the benefit of fiducial markers in overcoming those.

3 Material and Methods

Expression and purification of membrane protein constructs: DtpC; split sfGFP-DtpC (full length split sfGFP-DtpC$_{FL}$, and truncated constructs split sfGFP-DtpC$_{1-475}$ and split sfGFP-DtpC$_{1-470}$); split sfGFP-*Hs*PepT1 (full length split sfGFP-*Hs*PepT1$_{FL}$, and truncated constructs split sfGFP-*Hs*PepT1$_{1-672}$ and split sfGFP-*Hs*PepT1$_{10-672}$)

The full-length cDNA of DtpC wild type (WT) was amplified from the *Escherichia coli* genome, and cloned into a pNIC-CTHF vector by ligation-independent cloning (LIC). This vector contains a C-terminal His-Tag and a Tobacco Etch virus (TEV) cleavage site and a kanamycin resistance gene as selectable marker. 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. We named this construct split sfGFP-DtpC$_{FL}$. Two additional constructs were cloned with truncations of 5 (split sfGFP-DtpC$_{1-475}$), and 10 residues (split sfGFP-DtpC$_{1-470}$), on the C-terminal side of DtpC.

*Hs*PepT1 was previously cloned into a pXLG vector containing an expression cassette composed of an N-terminal Twin-Streptavidin tag followed by the HRV-3C protease recognition sequence (Killer et al., 2021). Similarly, as for DtpC, the two self-assembling parts of split-sfGFP were first inserted into the N- and C-termini of the full-length version of *Hs*PepT1, and on two other versions with i) a C-terminal truncation of 36 residues (split sfGFP-HsPepT11-672), and ii) a C-terminal truncation of 36 residues and a N-terminal truncation of 10 residues (split sfGFP-HsPepT110-672) were cloned.

Recombinant DtpC, and the three split sfGFP-DtpC constructs were expressed in *E. coli* C41(DE3) cells grown in terrific broth (TB) media supplemented with 30 µg/ml kanamycinaccording to established procedures (Löw et al., 2012, 2013). Cultures were grown at 37°C and protein expression was induced with 0.2 mM IPTG at an $OD_{600\ nm}$ of 0.6 - 0.8. After induction, culture growth continued at 18°C for 16-18 hours. Cells were harvested by centrifugation (10,000 g, 15 minutes, 4°C), and the pellet was 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 Avestin 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 (100,000g, 1 hour, 4°C using an Optima XE-90, Beckman Coulter centrifuge). Membranes were resuspended in lysis buffer supplemented with cOmplete EDTA-free protease inhibitors (Roche), and solubilized by adding 1% n-dodecyl-β-D-maltoside (DDM) detergent (Anatrace). The sample was centrifuged for 50 min at 90,000g, 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, 0.5mM 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 (Corning® Spin-X® UF concentrators) and run on an ÄKTA Pure system (GE Healthcare Life Sciences), using a HiLoad 16/ 600 Superdex 200 column for DtpC, and a Superdex 200 Increase 10/300 column for the split sfGFP-DtpC constructs. Fractions containing the protein were pooled, concentrated, flash frozen and stored at -80°C until further use.

For the split sfGFP-*Hs*PepT1 constructs , expression was done in mammalian cells as described previously (Pieprzyk et al., 2018; Killer et al., 2021). Briefly, HEK293F cells were collected 48 hours after transient transfection, and stored at −80°C until further use. Frozen cell pellets were resuspended in 300 mM NaCl, 20 mM NaPi (pH 7.5), 0.5 mM TCEP, and 5% glycerol, supplemented with cOmplete EDTA-free protease inhibitors, and were disrupted using an Avestin Emulsiflex homogenizer at 10,000-15,000 psi. The lysate was centrifuged for 10 min at 10,000g, 4°C, and the supernatant was centrifuged for 90 min at 100,000g, 4°C. The pellet containing the membrane fraction was solubilized in 1% N-dodecyl-β-D-maltopyranoside (DDM) and 0.1% cholesteryl hemisuccinate (CHS; Tris Salt, Anatrace) for 1 hour at 4°C. The sample was centrifuged for 50 min at 90,000g, and the supernatant was applied to Strep-TactinXT beads (IBA). After 20 min of incubation on a rotating wheel, the suspension was transferred to a gravity column. Following two wash steps with 300 mM NaCl, 20 mM HEPES (pH 7.5), 0.03% DDM, and 0.003% CHS, split sfGFP-*Hs*PepT1comnstructs were eluted with 0.03% DDM, 0.003% CHS, 150 mM NaCl, 20 mM HEPES (pH 7.5), and 10 mM desthiobiotin (Sigma-Aldrich).

Selection, expression and purification of nanobodies against DtpC

To generate DtpC specific nanobodies, two non-inbred llamas were injected six times at weekly intervals with a mixture of 94 different proteins including DtpC (50 µg of each antigen weekly). After six weeks of immunization, two separate phage display libraries were constructed, one from each animal, in the pMESy2 vector, which is a derivative of pMESy4 that contain a C-terminal EPEA-tag for affinity purification. After pooling both libraries, nanobodies were selected against individual antigens in two rounds of parallel panning in 96-well plates containing one immobilized antigen in each well. After two selection rounds on DtpC, 60 clones were picked for sequence analysis, 13 clones encoded antigen-specific nanobodies as tested in ELISA, grouping them in 5 different sequence families. A nanobody family is defined as a group of nanobodies with a high similarity in their CDR3 sequence (identical length and > 80% sequence identity). Nanobodies from the same family derive from the same B-cell lineage and likely bind to the same epitope on the target. Immunizations, library construction, selection by panning and nanobody characterization were performed according to standard procedures (Pardon et al., 2014). Five nanobodies were further characterized.

The nanobodies were expressed in *E. coli* WK6 cells and purified following standard procedures. Specifically, the cell pellet was resuspended in TES buffer (0.2 M TRIS, pH 8, 0.5 mM EDTA, 0.5 M sucrose) supplemented with one protease inhibitor tablet (Roche). 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 10,000 × g and additionally for 30 min at 100,000 × g. The supernatant was applied to CaptureSelect beads (Thermo Fisher Scientific), 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 $MgCl_2$. 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.

Expression and purification of macrobody 26

The nanobody 26 (Nb26) was first inserted into a pBXNPH3 vector containing a C-terminal penta-histidine tag preceded of a HRV-3C protease recognition sequence. The maltose binding protein (MBP) was then inserted 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., 2021). The resulting macrobody (Mb26) was expressed in *E. coli* WK6 cells as above. The cell pellet was resuspended in TES buffer (0.2 M TRIS, pH 8, 0.5 mM EDTA, 0.5 M sucrose) supplemented with one protease inhibitor tablet (Roche). 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 10,000 × g and additionally for 30 min at 142,000 × g. The supernatant was further 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 pooled, concentrated, flash frozen and stored at -80°C until further use.

Thermal stability measurements

The differential scanning fluorimetry method was used to follow the thermal unfolding event (Kotov et al., 2019) of Nb17, Nb26, Nb38, DtpC, DtpC-Nb17, DtpC-Nb26, DtpC-Nb38, Mb26, DtpC-Mb26, and split sfGFP-DtpC$_{1\text{-}475}$-Nb26 with a Prometheus NT.48 device (NanoTemper Technologies, Munich, Germany). The purified proteins were diluted to 16 µM, and the complexes were formed using a 1:1.5 molar ratio of membrane protein : fiducial. The fluorescence at 330 and 350 nm was recorded over a temperature gradient scan from 15° to 95°C and processed in GraphPad Prism 9.0 (GraphPad Software).

AlphaFold2 predictions

Structures with the following sequences were used as input for AlphaFold2 structure prediction (Jumper et al., 2021), and AMBER relaxation. The best ranked models were used for visualization.

> split sfGFP-DtpC$_{1\text{-}475}$
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNKTPSQPRAIYYIVAIQIWEYFSFYGMRALLILYLTHQLGFDD
NHAISLFSAYASLVYVTPILGGWLADRLLGNRTAVIAGALLMTLGHVVLGIDTNSTFSLYLA
LAIIICGYGLFKSNISCLLGELYDENDHRRDGGFSLLYAAGNIGSIAAPIACGLAAQWYGWHV
GFALAGGGMFIGLLIFLSGHRHFQSTRSMDKKALTSVKFALPVWSWLVVMLCLAPVFFTLL
LENDWSGYLLAIVCLIAAQIIARMMIKFPEHRRALWQIVLLMFVGTLFWVLAQQGGSTISLFI
DRFVNRQAFNIEVPTALFQSVNAIAVMLAGVVLAWLASPESRGNSTLRVWLKFAFGLLLMA
CGFMLLAFDARHAAADGQASMGVMISGLALMGFAELFIDPVAIAQITRLKMSGVLTGIYML
ATGAVANWLAGVVAQQTTESQISGMAIAAYQRFFSQMGEWTLACVAIIVVLAFATRFLFST

PNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQS
VLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

> split sfGFP-DtpC$_{FL}$
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNKTPSQPRAIYYIVAIQIWEYFSFYGMRALLILYLTHQLGFDD
NHAISLFSAYASLVYVTPILGGWLADRLLGNRTAVIAGALLMTLGHVVLGIDTNSTFSLYLA
LAIIICGYGLFKSNISCLLGELYDENDHRRDGGFSLLYAAGNIGSIAAPIACGLAAQWYGWHV
GFALAGGGMFIGLLIFLSGHRHFQSTRSMDKKALTSVKFALPVWSWLVVMLCLAPVFFTLL
LENDWSGYLLAIVCLIAAQIIARMMIKFPEHRRALWQIVLLMFVGTLFWVLAQQGGSTISLFI
DRFVNRQAFNIEVPTALFQSVNAIAVMLAGVVLAWLASPESRGNSTLRVWLKFAFGLLLMA
CGFMLLAFDARHAAADGQASMGVMISGLALMGFAELFIDPVAIAQITRLKMSGVLTGIYML
ATGAVANWLAGVVAQQTTESQISGMAIAAYQRFFSQMGEWTLACVAIIVVLAFATRFLFST
PTNMIQESNDNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

> split sfGFP-DtpC$_{+5Gly}$
MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNKTPSQPRAIYYIVAIQIWEYFSFYGMRALLILYLTHQLGFDD
NHAISLFSAYASLVYVTPILGGWLADRLLGNRTAVIAGALLMTLGHVVLGIDTNSTFSLYLA
LAIIICGYGLFKSNISCLLGELYDENDHRRDGGFSLLYAAGNIGSIAAPIACGLAAQWYGWHV
GFALAGGGMFIGLLIFLSGHRHFQSTRSMDKKALTSVKFALPVWSWLVVMLCLAPVFFTLL
LENDWSGYLLAIVCLIAAQIIARMMIKFPEHRRALWQIVLLMFVGTLFWVLAQQGGSTISLFI
DRFVNRQAFNIEVPTALFQSVNAIAVMLAGVVLAWLASPESRGNSTLRVWLKFAFGLLLMA
CGFMLLAFDARHAAADGQASMGVMISGLALMGFAELFIDPVAIAQITRLKMSGVLTGIYML
ATGAVANWLAGVVAQQTTESQISGMAIAAYQRFFSQMGEWTLACVAIIVVLAFATRFLFST
PTNMIQESNDGGGGGNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

Cryo-EM sample preparation, data collection, image analysis, and atomic modelling

One hour before vitrification, the purified protein complexes were thawed on ice and run on a Superdex Increase 200 5/150 column in 0.015% DDM, 100 mM NaCl, 10 mM HEPES (pH 7.5), 0.5 mM TCEP in order to remove the excess of empty detergent micelles earlier generated upon sample concentration. The top fraction reached a concentration ranging between 3 and 6 mg/ml, and for each sample, 3.6 µl were applied to glow-discharged gold holey carbon 2/1 300-mesh grids (Quantifoil). Grids were blotted for 4 s at 0 force and 1-s wait time before being vitrified in liquid propane using a Mark IV Vitrobot (Thermo Fisher Scientific). The blotting chamber was maintained at 4°C and 100% humidity during freezing.

All movies were collected using a Titan Krios (Thermo Fisher Scientific) outfitted with a K3 camera and BioQuantum energy filter (Gatan) set to 10 eV. Automated data acquisitions were set using EPU (Thermo Fisher Scientific). The applied defocus ranged between -0.9 µm and -1.8 µm in all datasets.

For DtpC-Nb26 and DtpC-Mb26, movies were collected at a nominal magnification of ×105,000 and a physical pixel size of 0.85 Å, with a 70-µm C2 aperture and 100-µm objective aperture at a dose rate of 19.5 e−/pixel per second. A total dose of 75 e−/Å^2 was used with 2.8 s exposure time,

fractionated in 50 frames. For split sfGFP-DtpC$_{1-475}$-Nb26, movies were collected at a nominal magnification of ×130,000 and a physical pixel size of 0.67 Å, with a 50-µm C2 aperture and 100-µm objective aperture at a dose rate of 19.0 e−/pixel per second. A total dose of 57 e−/Å^2 was used with 3 s exposure time fractionated in 40 frames.

All movies were motion-corrected using Relion-3.1 (Scheres, 2012; Zivanov et al., 2018) own implementation of MotionCor2 (Zheng et al., 2017). Contrast transfer function parameters were calculated using CTFFIND4 (Rohou and Grigorieff, 2015). and putative particle coordinates were initially defined using CrYOLO (Wagner et al., 2019).

For DtpC-Mb26, 13257 movies were collected, 3,062,337 coordinates were picked and used for 2D averaging and clustering. For split sfGFP-DtpC$_{1-475}$-Nb26, 7602 movies were collected, 1,049,399 coordinates were picked and used for 2D averaging and clustering. For DtpC-Nb26, 24,333 movies were collected, 6,464,070 coordinates were picked and used for 2D averaging and clustering, and 878,428 particles were used in the final 3D reconstruction. Briefly, DtpC-Nb26 dimeric population was clustered using 3D class averaging in Relion3.1 (Scheres, 2012). Particle trajectories and cumulative beam damage were further corrected by Bayesian polishing in Relion3.1 (Zivanov et al., 2019), and the resulting shiny particles were exported to cryoSPARCv3 (Punjani et al., 2017) for further 3D clustering *via* successive heterogeneous refinement cycles using "bad" and "good" volumes as references to denoise the dataset. Non uniform refinement (Punjani et al., 2020), followed by a local refinement using a soft mask around one transporter unit resulted in a 2.7 Å reconstruction of DtpC. The overall resolution was estimated in CryoSPARCv3 using the FSC = 0.143 cutoff. Local resolution estimations were also calculated in CryoSPARCv3 using the 0.5 FSC cutoff. The two 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). The latter led to a slightly better defined contour of the atoms, and was subsequently used for the last atomic-model refinement of DtpC. The initial models of DtpC and Nb26 were generated using AlphaFold2, and refined against the experimental maps; first in Isolde (Croll, 2018), and last in Phenix (Afonine et al., 2018), principally to refine atomic displacement parameters (B-factors) and perform a slight energy minimization while keeping restrains from Isolde's reference model. Half-maps, and postprocessed maps of the dimeric arrangement and of the focused refinement, as well as the atomic model of DtpC were deposited in the PDB and EMDB as deposition numbers 7ZC2, and EMD-14618. The atomic model of the dimeric DtpC-Nb26 is available upon request.

Small-angle X-ray scattering data collection and analysis

Synchrotron SAXS data from solutions of DtpC-Nb26 in β-DDM micelles (SEC-SAXS) were collected on the EMBL P12 (Blanchet et al., 2015) beamline at the PETRA III storage ring (Hamburg, Germany), in a buffer consisting of 0.015% DDM, 100 mM NaCl, 10 mM HEPES (pH 7.5), and 0.5 mM TCEP. Sample (10 mg/ml) was injected onto a Superdex Increase 200 10/300 column (Cytiva) and run at 0.5 ml/min at 20°C. 3000 successive 1 second frames were collected using a Pilatus 2M detector at a sample-detector distance of 3.1 m and at a wavelength of $\lambda = 0.124$ nm (I(s) vs s, where $s = 4\pi\sin\theta/\lambda$, and 2θ is the scattering angle). The data were normalized to the intensity of the transmitted beam and radially averaged; the scattering of the solvent-blank was subtracted using CHROMIXS (Panjkovich and Svergun, 2018). Cryo-EM volume maps of DtpC-Nb26 were fit to the scattering data across the low-angle range (shape region only) using EM2DAM (Franke et al., 2017) at a density threshold of 0.1.

Data visualization

Graphs were generated using GraphPad Prism 9.0 (GraphPad Software). Molecular graphics and analyses performed with UCSF ChimeraX-1.2.5 (Pettersen et al., 2021). Figures were prepared in Adobe Illustrator 2021.

4 References

Aduri, N. G., Prabhala, B. K., Ernst, H. A., Jørgensen, F. S., Olsen, L., and Mirza, O. (2015). Salt Bridge Swapping in the EXXERFXYY Motif of Proton-coupled Oligopeptide Transporters. *J Biol Chem* 290, 29931–29940. doi:10.1074/jbc.M115.675603.

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., et al. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* 74, 531–544. doi:10.1107/S2059798318006551.

Bartels, K., Lasitza-Male, T., Hofmann, H., and Löw, C. (2021). Single-Molecule FRET of Membrane Transport Proteins. *Chembiochem*. doi:10.1002/cbic.202100106.

Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., et al. (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J Appl Crystallogr* 48, 431–443. doi:10.1107/S160057671500254X.

Boggavarapu, R., Jeckelmann, J.-M., Harder, D., Ucurum, Z., and Fotiadis, D. (2015). Role of electrostatic interactions for ligand recognition and specificity of peptide transporters. *BMC Biol* 13, 58. doi:10.1186/s12915-015-0167-8.

Botte, M., Ni, D., Schenck, S., Zimmermann, I., Chami, M., Bocquet, N., et al. (2021). Insight into Lipopolysaccharide Translocation by Cryo-EM structures of a LptDE Transporter in Complex with Pro-Macrobodies. Biochemistry doi:10.1101/2021.03.23.436624.

Brandsch, M. (2009). Transport of drugs by proton-coupled peptide transporters: pearls and pitfalls. *Expert Opin Drug Metab Toxicol* 5, 887–905. doi:10.1517/17425250903042292.

Brandsch, M., Knütter, I., and Leibach, F. H. (2004). The intestinal H+/peptide symporter PEPT1: structure-affinity relationships. *Eur J Pharm Sci* 21, 53–60. doi:10.1016/s0928-0987(03)00142-8.

Brunner, J. D., Jakob, R. P., Schulze, T., Neldner, Y., Moroni, A., Thiel, G., et al. (2020). Structural basis for ion selectivity in TMEM175 K+ channels. *Elife* 9, e53683. doi:10.7554/eLife.53683.

Chen, X. Z., Steel, A., and Hediger, M. A. (2000). Functional roles of histidine and tyrosine residues in the H(+)-peptide transporter PepT1. *Biochem Biophys Res Commun* 272, 726–730. doi:10.1006/bbrc.2000.2851.

Colas, C., Masuda, M., Sugio, K., Miyauchi, S., Hu, Y., Smith, D. E., et al. (2017). Chemical Modulation of the Human Oligopeptide Transporter 1, hPepT1. *Mol Pharm* 14, 4685–4693. doi:10.1021/acs.molpharmaceut.7b00775.

Croll, T. I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* 74, 519–530. doi:10.1107/S2059798318002425.

Daniel, H. (2004). Molecular and integrative physiology of intestinal peptide transport. *Annu Rev Physiol* 66, 361–384. doi:10.1146/annurev.physiol.66.032102.144149.

Daniel, H., Spanier, B., Kottra, G., and Weitz, D. (2006). From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology (Bethesda)* 21, 93–102. doi:10.1152/physiol.00054.2005.

Doki, S., Kato, H. E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., et al. (2013). Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc Natl Acad Sci U S A* 110, 11343–11348. doi:10.1073/pnas.1301079110.

Drew, D., and Boudker, O. (2016). Shared Molecular Mechanisms of Membrane Transporters. *Annu Rev Biochem* 85, 543–572. doi:10.1146/annurev-biochem-060815-014520.

Drew, D., North, R. A., Nagarathinam, K., and Tanabe, M. (2021). Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS). *Chem Rev* 121, 5289–5335. doi:10.1021/acs.chemrev.0c00983.

Ernst, H. A., Pham, A., Hald, H., Kastrup, J. S., Rahman, M., and Mirza, O. (2009). Ligand binding analyses of the putative peptide transporter YjdL from E. coli display a significant selectivity towards dipeptides. *Biochem Biophys Res Commun* 389, 112–116. doi:10.1016/j.bbrc.2009.08.098.

Foley, D. W., Rajamanickam, J., Bailey, P. D., and Meredith, D. (2010). Bioavailability through PepT1: the role of computer modelling in intelligent drug design. *Curr Comput Aided Drug Des* 6, 68–78. doi:10.2174/157340910790980133.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., et al. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* 50, 1212–1225. doi:10.1107/S1600576717007786.

Gabrielsen, M., Kroner, F., Black, I., Isaacs, N. W., Roe, A. J., and McLuskey, K. (2011). High-throughput identification of purification conditions leads to preliminary crystallization conditions for three inner membrane proteins. *Mol Membr Biol* 28, 445–453. doi:10.3109/09687688.2011.628954.

Guettou, F., Quistgaard, E. M., Raba, M., Moberg, P., Löw, C., and Nordlund, P. (2014). Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat Struct Mol Biol* 21, 728–731. doi:10.1038/nsmb.2860.

Guettou, F., Quistgaard, E. M., Trésaugues, L., Moberg, P., Jegerschöld, C., Zhu, L., et al. (2013). Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO Rep* 14, 804–810. doi:10.1038/embor.2013.107.

Harder, D., Stolz, J., Casagrande, F., Obrdlik, P., Weitz, D., Fotiadis, D., et al. (2008). DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of Escherichia coli. *FEBS J* 275, 3290–3298. doi:10.1111/j.1742-4658.2008.06477.x.

Hediger, M. A., Clémençon, B., Burrier, R. E., and Bruford, E. A. (2013). The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med* 34, 95–107. doi:10.1016/j.mam.2012.12.009.

Heinz, L. X., Lee, J., Kapoor, U., Kartnig, F., Sedlyarov, V., Papakostas, K., et al. (2020). TASL is the SLC15A4-associated adaptor for IRF5 activation by TLR7-9. *Nature* 581, 316–322. doi:10.1038/s41586-020-2282-0.

Hillgren, K. M., Keppler, D., Zur, A. A., Giacomini, K. M., Stieger, B., Cass, C. E., et al. (2013). Emerging transporters of clinical importance: an update from the International Transporter Consortium. *Clin Pharmacol Ther* 94, 52–63. doi:10.1038/clpt.2013.74.

Ingersoll, S. A., Ayyadurai, S., Charania, M. A., Laroui, H., Yan, Y., and Merlin, D. (2012). The role and pathophysiological relevance of membrane transporter PepT1 in intestinal inflammation and inflammatory bowel disease. *Am J Physiol Gastrointest Liver Physiol* 302, G484-492. doi:10.1152/ajpgi.00477.2011.

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* 211, 969–970. doi:10.1038/211969a0.

Jensen, J. M., Aduri, N. G., Prabhala, B. K., Jahnsen, R., Franzyk, H., and Mirza, O. (2014). Critical role of a conserved transmembrane lysine in substrate recognition by the proton-coupled oligopeptide transporter YjdL. *Int J Biochem Cell Biol* 55, 311–317. doi:10.1016/j.biocel.2014.09.016.

Jensen, J. M., Ernst, H. A., Wang, X., Hald, H., Ditta, A. C., Ismat, F., et al. (2012a). Functional investigation of conserved membrane-embedded glutamate residues in the proton-coupled peptide transporter YjdL. *Protein Pept Lett* 19, 282–287. doi:10.2174/092986612799363109.

Jensen, J. M., Ismat, F., Szakonyi, G., Rahman, M., and Mirza, O. (2012b). Probing the putative active site of YjdL: an unusual proton-coupled oligopeptide transporter from E. coli. *PLoS One* 7, e47780. doi:10.1371/journal.pone.0047780.

Jensen, J. M., Simonsen, F. C., Mastali, A., Hald, H., Lillebro, I., Diness, F., et al. (2012c). Biophysical characterization of the proton-coupled oligopeptide transporter YjdL. *Peptides* 38, 89–93. doi:10.1016/j.peptides.2012.08.012.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. doi:10.1038/s41586-021-03819-2.

Killer, M., Wald, J., Pieprzyk, J., Marlovits, T. C., and Löw, C. (2021). Structural snapshots of human PepT1 and PepT2 reveal mechanistic insights into substrate and drug transport across epithelial membranes. *Sci Adv* 7, eabk3259. doi:10.1126/sciadv.abk3259.

Kotov, V., Bartels, K., Veith, K., Josts, I., Subhramanyam, U. K. T., Günther, C., et al. (2019). High-throughput stability screening for detergent-solubilized membrane proteins. *Sci Rep* 9, 10379. doi:10.1038/s41598-019-46686-8.

Liu, S., Li, S., Krezel, A. M., and Li, W. (2022). Stabilization and structure determination of integral membrane proteins by termini restraining. *Nat Protoc* 17, 540–565. doi:10.1038/s41596-021-00656-5.

Liu, S., Li, S., Yang, Y., and Li, W. (2020). Termini restraining of small membrane proteins enables structure determination at near-atomic resolution. *Sci Adv* 6, eabe3717. doi:10.1126/sciadv.abe3717.

Löw, C., Jegerschöld, C., Kovermann, M., Moberg, P., and Nordlund, P. (2012). Optimisation of over-expression in E. coli and biophysical characterisation of human membrane protein synaptogyrin 1. *PLoS One* 7, e38244. doi:10.1371/journal.pone.0038244.

Löw, C., Moberg, P., Quistgaard, E. M., Hedrén, M., Guettou, F., Frauenfeld, J., et al. (2013). High-throughput analytical gel filtration screening of integral membrane proteins for structural studies. *Biochim Biophys Acta* 1830, 3497–3508. doi:10.1016/j.bbagen.2013.02.001.

Lyons, J. A., Parker, J. L., Solcan, N., Brinth, A., Li, D., Shah, S. T. A., et al. (2014). Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO Rep* 15, 886–893. doi:10.15252/embr.201338403.

Martinez Molledo, M., Quistgaard, E. M., Flayhan, A., Pieprzyk, J., and Löw, C. (2018a). Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* 26, 467-476.e4. doi:10.1016/j.str.2018.01.005.

Martinez Molledo, M., Quistgaard, E. M., and Löw, C. (2018b). Tripeptide binding in a proton-dependent oligopeptide transporter. *FEBS Lett* 592, 3239–3247. doi:10.1002/1873-3468.13246.

Minhas, G. S., Bawdon, D., Herman, R., Rudden, M., Stone, A. P., James, A. G., et al. (2018). Structural basis of malodour precursor transport in the human axilla. *Elife* 7, e34995. doi:10.7554/eLife.34995.

Nagamura, R., Fukuda, M., Kawamoto, A., Matoba, K., Dohmae, N., Ishitani, R., et al. (2019). Structural basis for oligomerization of the prokaryotic peptide transporter PepTSo2. *Acta Crystallogr F Struct Biol Commun* 75, 348–358. doi:10.1107/S2053230X19003546.

Newstead, S., Drew, D., Cameron, A. D., Postis, V. L. G., Xia, X., Fowler, P. W., et al. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J* 30, 417–426. doi:10.1038/emboj.2010.309.

Panjkovich, A., and Svergun, D. I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* 34, 1944–1946. doi:10.1093/bioinformatics/btx846.

593 Pardon, E., Laeremans, T., Triest, S., Rasmussen, S. G. F., Wohlkönig, A., Ruf, A., et al. (2014). A
594 general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* 9, 674–
595 693. doi:10.1038/nprot.2014.039.

596 Parker, J. L., Deme, J. C., Wu, Z., Kuteyi, G., Huo, J., Owens, R. J., et al. (2021). Cryo-EM structure
597 of PepT2 reveals structural basis for proton-coupled peptide and prodrug transport in
598 mammals. *Sci Adv* 7, eabh3355. doi:10.1126/sciadv.abh3355.

599 Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021).
600 UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein*
601 *Sci* 30, 70–82. doi:10.1002/pro.3943.

602 Pieprzyk, J., Pazicky, S., and Löw, C. (2018). Transient Expression of Recombinant Membrane-
603 eGFP Fusion Proteins in HEK293 Cells. *Methods Mol Biol* 1850, 17–31. doi:10.1007/978-1-
604 4939-8730-6_2.

605 Prabhala, B. K., Aduri, N. G., Iqbal, M., Rahman, M., Gajhede, M., Hansen, P. R., et al. (2017).
606 Several hPepT1-transported drugs are substrates of the Escherichia coli proton-coupled
607 oligopeptide transporter YdgR. *Res Microbiol* 168, 443–449.
608 doi:10.1016/j.resmic.2017.01.005.

609 Prabhala, B. K., Aduri, N. G., Jensen, J. M., Ernst, H. A., Iram, N., Rahman, M., et al. (2014). New
610 insights into the substrate specificities of proton-coupled oligopeptide transporters from E.
611 coli by a pH sensitive assay. *FEBS Lett* 588, 560–565. doi:10.1016/j.febslet.2014.01.004.

612 Prabhala, B. K., Aduri, N. G., Sharma, N., Shaheen, A., Sharma, A., Iqbal, M., et al. (2018). The
613 prototypical proton-coupled oligopeptide transporter YdgR from Escherichia coli facilitates
614 chloramphenicol uptake into bacterial cells. *J Biol Chem* 293, 1007–1017.
615 doi:10.1074/jbc.M117.805960.

616 Punjani, A., and Fleet, D. J. (2021). 3D variability analysis: Resolving continuous flexibility and
617 discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* 213,
618 107702. doi:10.1016/j.jsb.2021.107702.

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

622 Punjani, A., Zhang, H., and Fleet, D. J. (2020). Non-uniform refinement: adaptive regularization
623 improves single-particle cryo-EM reconstruction. *Nat Methods* 17, 1214–1221.
624 doi:10.1038/s41592-020-00990-8.

625 Quistgaard, E. M., Löw, C., Guettou, F., and Nordlund, P. (2016). Understanding transport by the
626 major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol* 17, 123–
627 132. doi:10.1038/nrm.2015.25.

628 Quistgaard, E. M., Martinez Molledo, M., and Löw, C. (2017). Structure determination of a major
629 facilitator peptide transporter: Inward facing PepTSt from Streptococcus thermophilus

crystallized in space group P3121. *PLoS One* 12, e0173126. doi:10.1371/journal.pone.0173126.

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

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J., Sorzano, C., and Vargas, J. (2020). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Bioinformatics doi:10.1101/2020.06.12.148296.

Scheres, S. H. W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180, 519–530. doi:10.1016/j.jsb.2012.09.006.

Sharma, N., Aduri, N. G., Iqbal, A., Prabhala, B. K., and Mirza, O. (2016). Peptide Selectivity of the Proton-Coupled Oligopeptide Transporter from Neisseria meningitidis. *J Mol Microbiol Biotechnol* 26, 312–319. doi:10.1159/000447129.

Smith, D. E., Clémençon, B., and Hediger, M. A. (2013). Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol Aspects Med* 34, 323–336. doi:10.1016/j.mam.2012.11.003.

Solcan, N., Kwok, J., Fowler, P. W., Cameron, A. D., Drew, D., Iwata, S., et al. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J* 31, 3411–3421. doi: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. *Compr Physiol* 8, 843–869. doi:10.1002/cphy.c170038.

Stauffer, M., Jeckelmann, J.-M., Ilgü, H., Ucurum, Z., Boggavarapu, R., and Fotiadis, D. (2022). Peptide transporter structure reveals binding and action mechanism of a potent PEPT1 and PEPT2 inhibitor. *Commun Chem* 5, 23. doi:10.1038/s42004-022-00636-0.

Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D., and Afonine, P. V. (2020). Improvement of cryo-EM maps by density modification. *Nat Methods* 17, 923–927. doi:10.1038/s41592-020-0914-9.

Thwaites, D. T., and Anderson, C. M. H. (2007). H+-coupled nutrient, micronutrient and drug transporters in the mammalian small intestine. *Exp Physiol* 92, 603–619. doi:10.1113/expphysiol.2005.029959.

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

Viennois, E., Pujada, A., Zen, J., and Merlin, D. (2018). Function, Regulation, and Pathophysiological Relevance of the POT Superfamily, Specifically PepT1 in Inflammatory Bowel Disease. *Compr Physiol* 8, 731–760. doi:10.1002/cphy.c170032.

668 Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., et al. (2019). SPHIRE-
669 crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* 2,
670 218. doi:10.1038/s42003-019-0437-z.

671 Yan, N. (2015). Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev*
672 *Biophys* 44, 257–283. doi:10.1146/annurev-biophys-060414-033901.

673 Zhang, Y., Zhang, Y., Sun, K., Meng, Z., and Chen, L. (2019). The SLC transporter in nutrient and
674 metabolic sensing, regulation, and drug development. *J Mol Cell Biol* 11, 1–13.
675 doi:10.1093/jmcb/mjy052.

676 Zhao, Y., Mao, G., Liu, M., Zhang, L., Wang, X., and Zhang, X. C. (2014). Crystal structure of the
677 E. coli peptide transporter YbgH. *Structure* 22, 1152–1160. doi:10.1016/j.str.2014.06.008.

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

681 Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., et al. (2018). New
682 tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7,
683 e42166. doi:10.7554/eLife.42166.

684 Zivanov, J., Nakane, T., and Scheres, S. H. W. (2019). A Bayesian approach to beam-induced motion
685 correction in cryo-EM single-particle analysis. *IUCrJ* 6, 5–17.
686 doi:10.1107/S205225251801463X.

687 **5 Conflict of Interest**

688 The authors declare that the research was conducted in the absence of any commercial or financial
689 relationships that could be construed as a potential conflict of interest.

690 **6 Author Contributions**

691 Conceptualization: MK, CL

692 Methodology: MK, GF, HDTM, EP

693 Investigation: MK, GF, HDTM, EP

694 Visualization: MK, HDTM

695 Funding acquisition: DIS, JS, CL

696 Project administration: CL

697 Supervision: DIS, JS, CL

698 Writing—original draft: MK, CL

699 Writing—review & editing: MK, GF, HDTM, DIS, EP, JS, CL

7 Funding

This work was supported by a grant from the BMBF (grant number: 05K18YEA). Part of this work was performed at the CryoEM Facility at CSSB, supported by the UHH and DFG grant numbers (INST 152/772-1|152/774-1|152/775-1|152/776-1|152/777-1 FUGG).

8 Acknowledgements

We thank the Sample Preparation and Characterization facility of EMBL Hamburg for support in this project and the beamlines P13 and P14 at EMBL Hamburg for regular access. We acknowledge Instruct-ERIC and the FWO for their support to the Nb discovery and Saif Saifuzzaman for the technical assistance during Nb discovery. All past and current group members are acknowledged for their input to this manuscript and their efforts to crystallize DtpC over the years.

9 Data Availability statement

Half-maps, and post processed maps of the dimeric arrangement and of the focused refinement, as well as the atomic model of DtpC were deposited in the PDB and EMDB as deposition numbers 7ZC2, and EMD-14618.

10 Figures

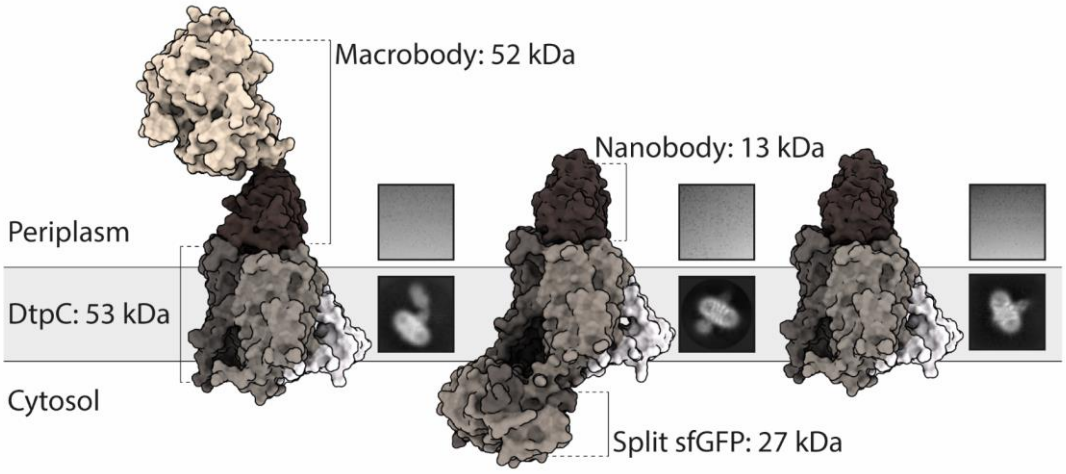


Figure 1: Utilization of different fiducial markers to improve particle alignment and 2D averaging from cryo-EM images. From left to right: DtpC-Mb26, split-sfGFP-DtpC$_{1\text{-}475}$-Nb26, and DtpC-Nb26 were purified, vitrified on grids and imaged. Single particles were identified, clustered and averaged. The best average from each sample is shown under a representative raw micrograph.

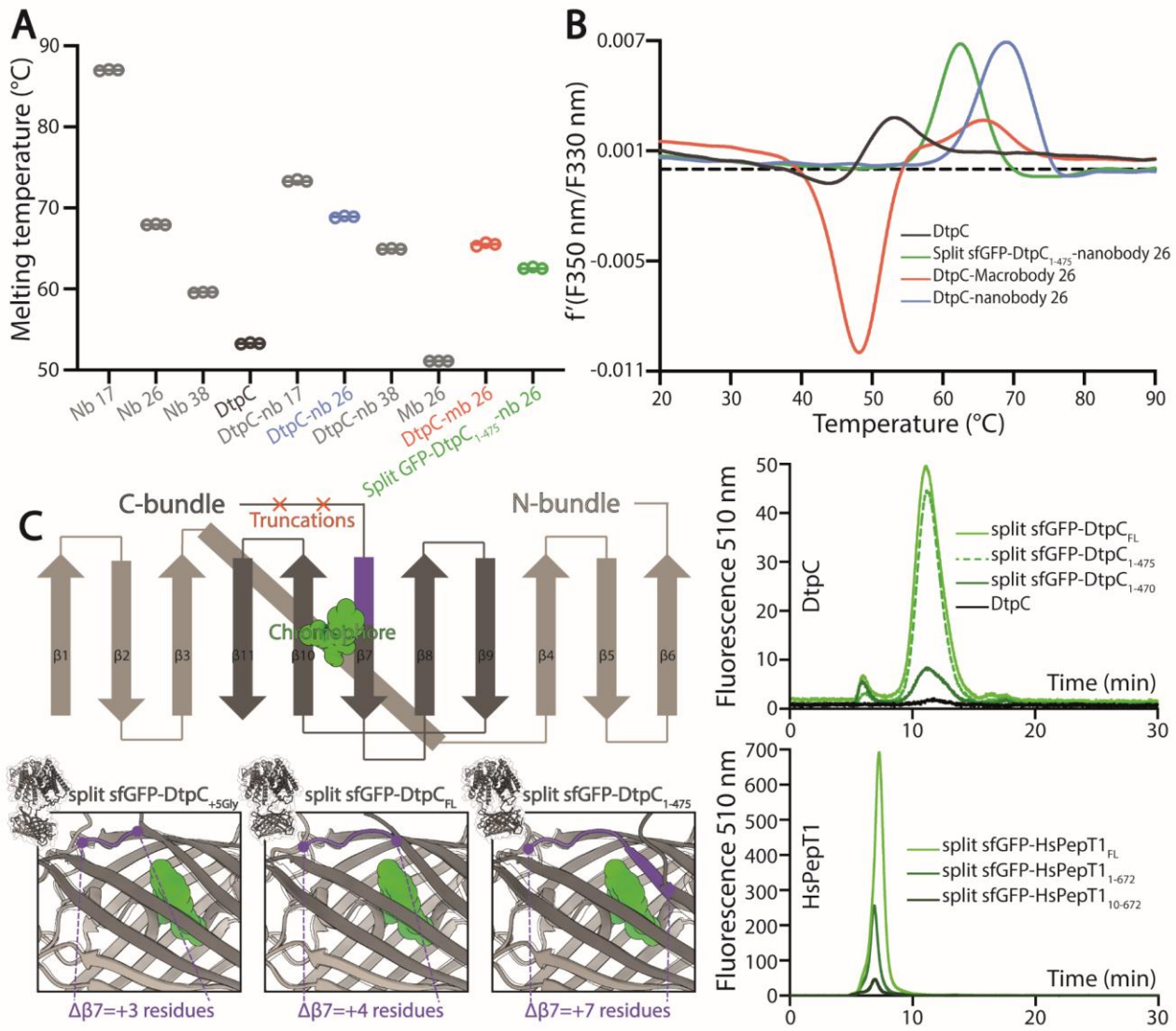


Figure 2: Characterization of the different fiducial markers. (A) The melting temperature of each fiducial and DtpC-fiducial complex was measured by nano-differential scanning fluorimetry (DSF) in triplicate measurements as shown as open circles. The average of the three values is marked by a line. (B) The first derivative of the summarized data in (A) is shown for DtpC and the three imaged samples together with the respective fiducial. (C) A schematic representation of the split sfGFP-DtpC architecture is shown on the top left panel. Below, structure predictions were generated for split-sfGFP-DtpC$_{+5Gly}$, split-sfGFP-DtpC$_{FL}$, sfGFP-DtpC$_{1-475}$, and overlaid with sfGFP (PDB accession number 2B3P). The dark-violet coloring corresponds to the fraction of ß7 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.

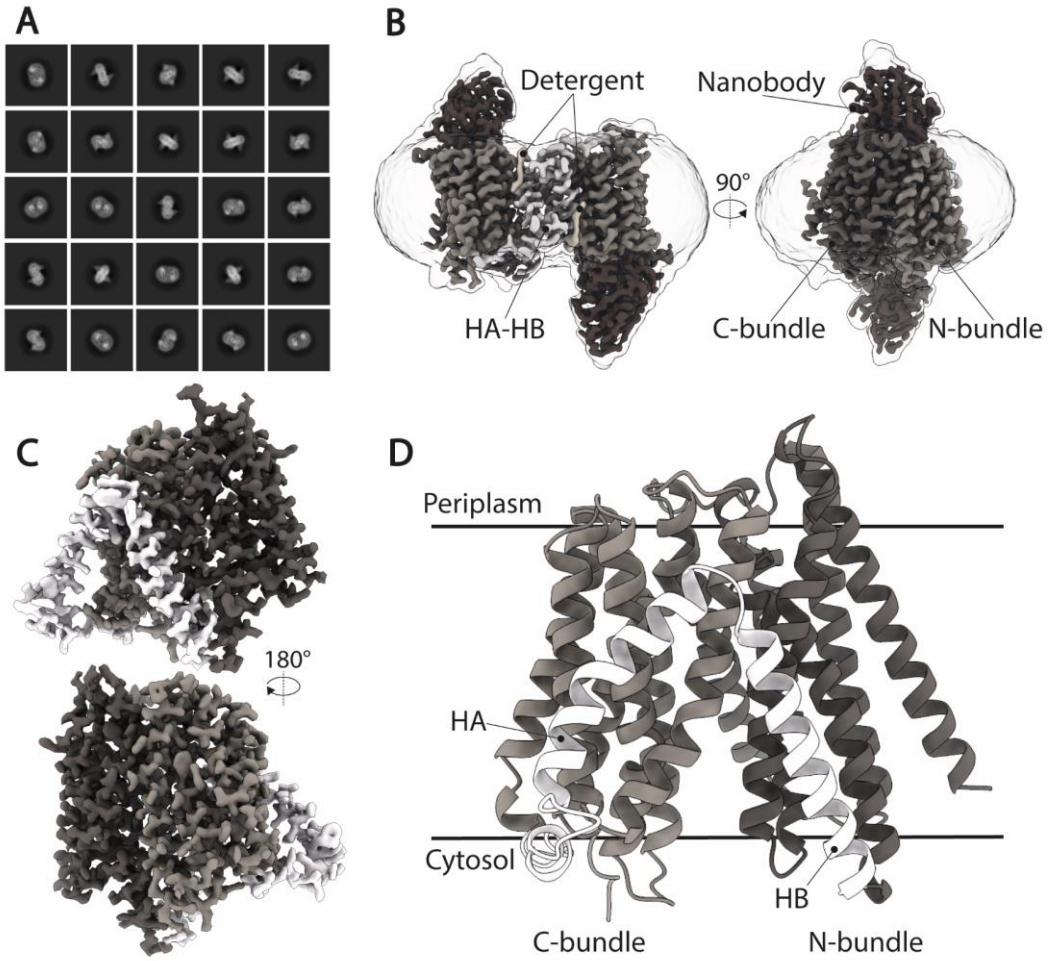


Figure 3: Cryo-EM structure of DtpC-Nb26. (A) Representative 2D class averages of the dimeric population. (B) 3D reconstruction of the DtpC-Nb26 inverted dimer used for local focused refinement on one copy of the transporter, shown in (C). (D) Atomic model of DtpC displayed as ribbon diagram. The different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

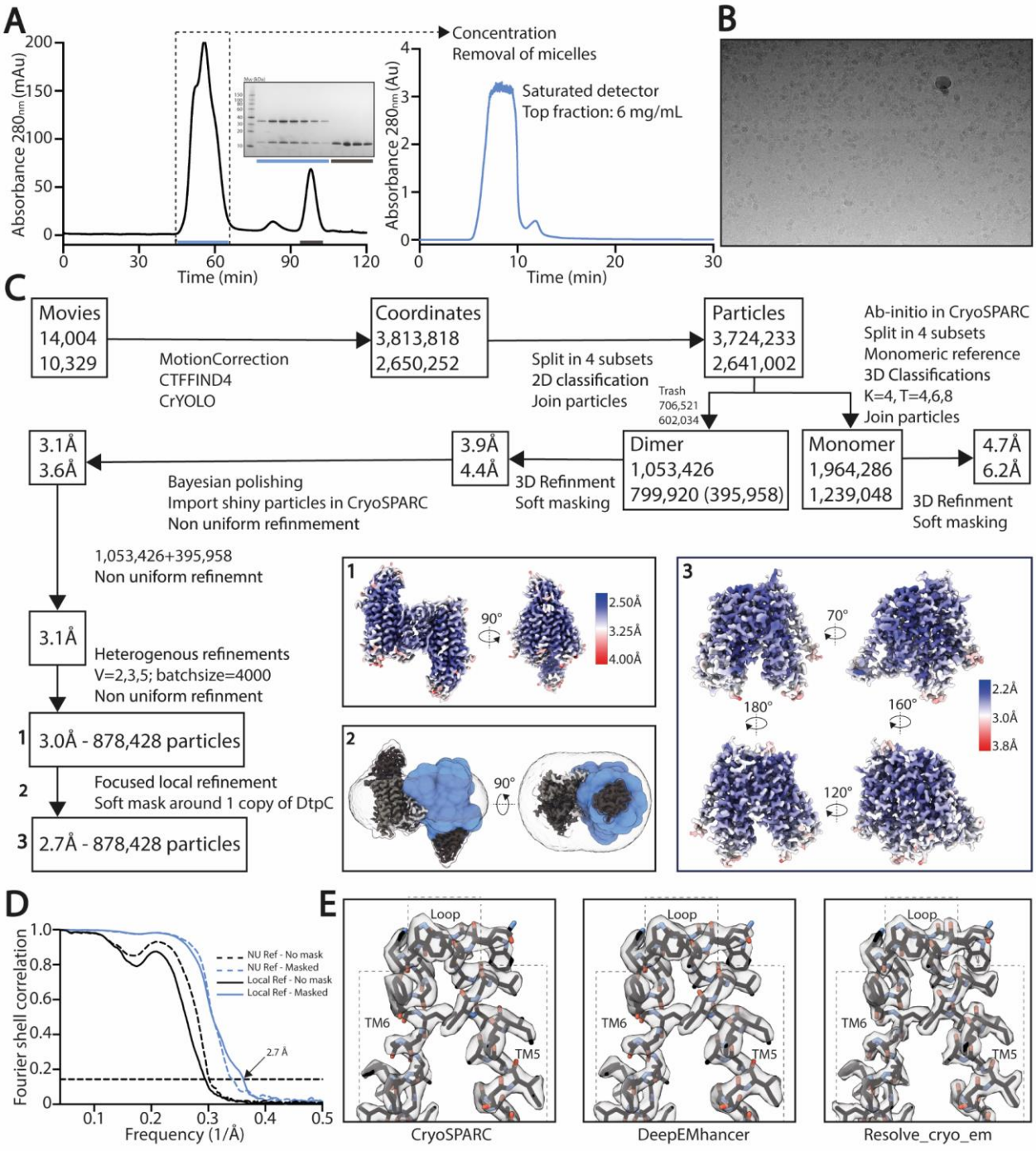


Figure 4: High resolution structure determination of DtpC-Nb26. (A) Gel filtration was performed on a preparative column (left) before concentrating the sample to 60 mg/mL and rerunning it on an analytic column on an HPLC system (right), in order to obtain a highly concentrated sample, free of empty detergent micelles. Peak shape already indicates a mixture of different oligomeric species. (B) Representative raw micrograph of the acquired dataset. The applied defocus is -1.5 µm. (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,

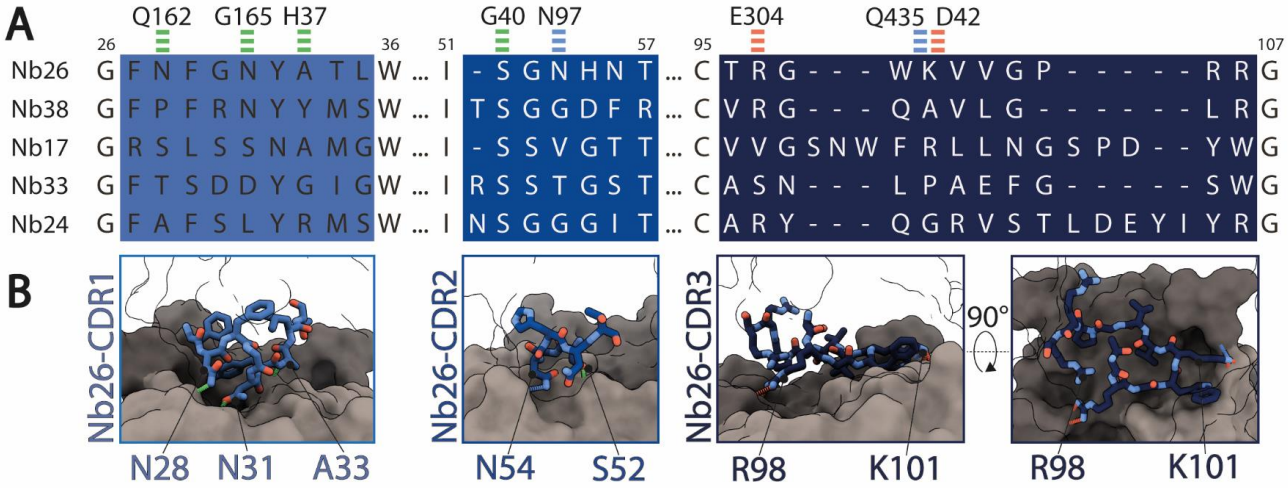
763 to estimate the overall resolution using the 0.143 cutoff threshold. (E) The two half maps were used as
764 inputs to assess various post-processing strategies.

765

766

767

768



769

770 **Figure 5: Interactions between Nb26 and DtpC.** (A) The sequences of the five nanobodies
771 representing five different families, obtained after selection, with their complementary determining
772 regions (CDR) are shown. Interactions of Nb26 with DtpC are highlighted as green (hydrogen bounds
773 involving the protein backbone), blue (hydrogen bounds involving side chains) and red dashes (salt
774 bridges). (B) These interactions are further displayed in 3D. CDR regions are depicted as sticks on the
775 surface of DtpC where the N-terminal bundle is colored in grey, and the C-terminal bundle in dark
776 grey.

777

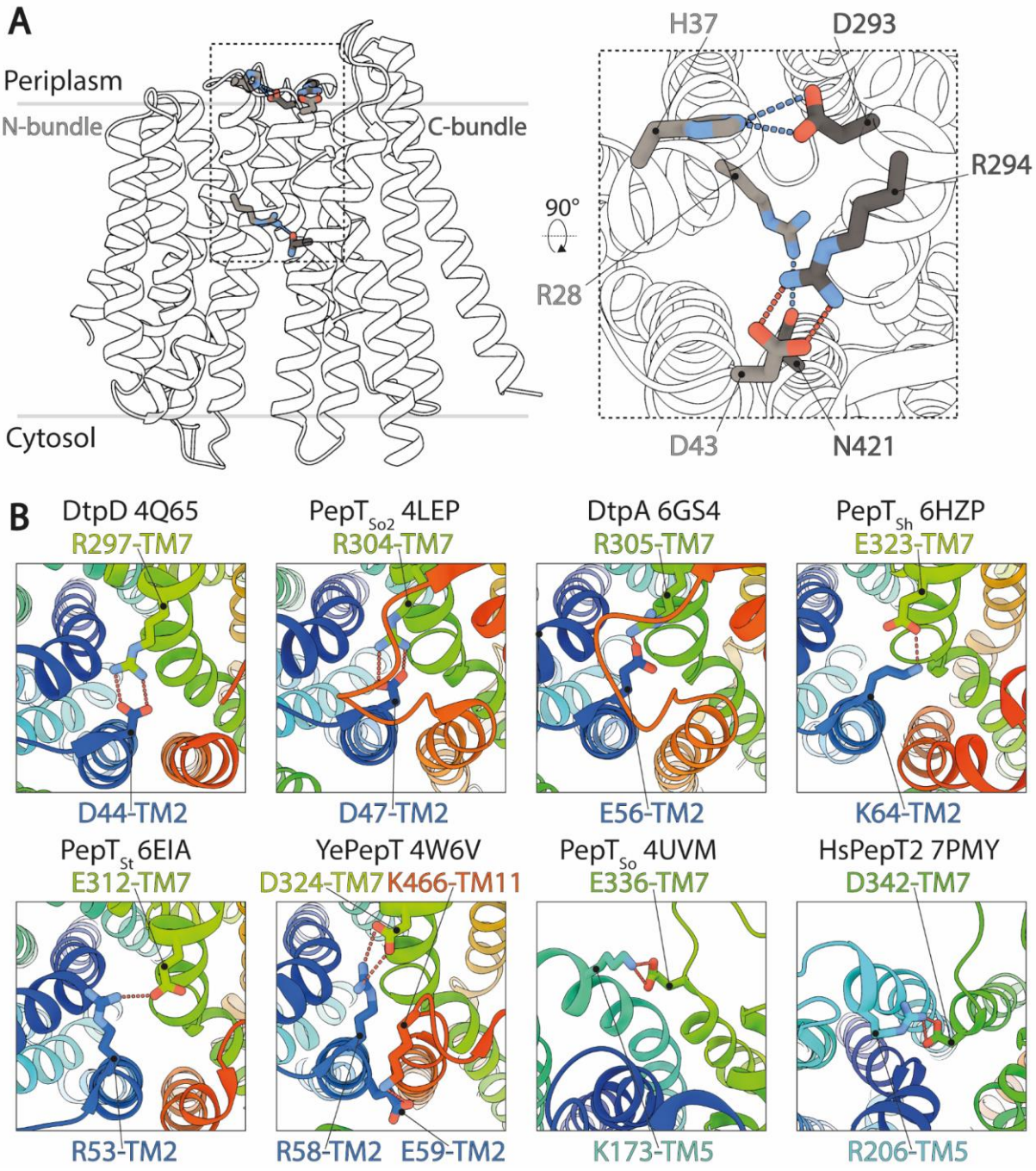


Figure 6: Structural basis for the stabilization of the inward facing state in DtpC and other POT homologues. (A) The salt bridge and hydrogen bounds favoring closure of the two bundles on the periplasmic side of DtpC are respectively shown as red and blue dashes. (B) The structures of homologous POTs 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 analyzed to identify the strongest interaction stabilizing their common conformation. The structures are colored 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.

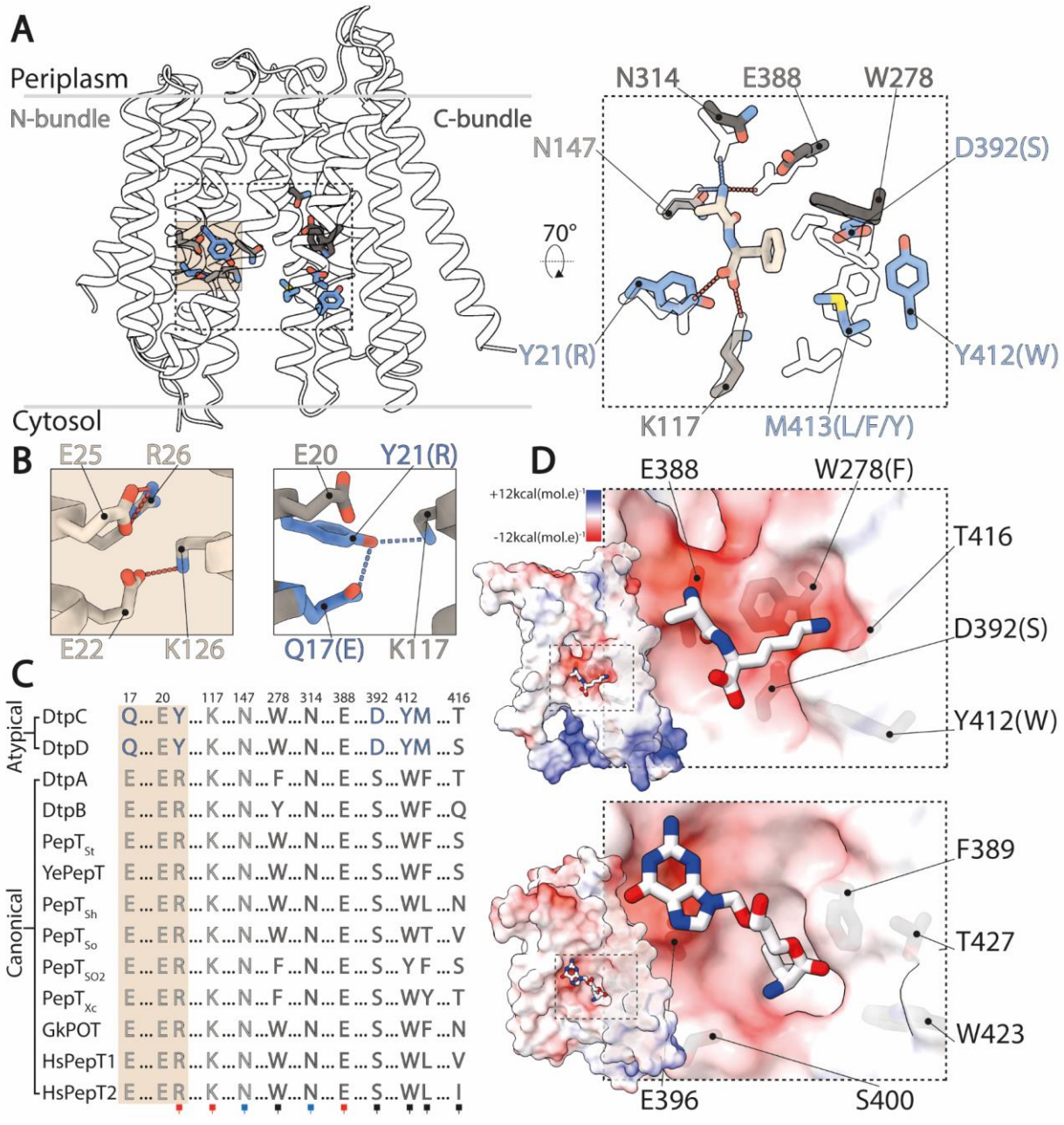


Figure 7: Structural basis for ligand selectivity in DtpC and atypical POTs. (A) Key residues involved in substrate binding are colored and shown as sticks. In the close up view, an overlay of HsPepT2 (transparent residues) bound to the dipeptide Ala-Phe (beige) with DtpC is shown. Residues colored in 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 bounds. 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; $PepT_{St}$: 5OXO; DtpA bound to valganciclovir: 6GS4.

11 Tables

Table 1: Data collection and refinement statistics of the deposited DtpC structure

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, $FSC_{\text{half maps; 0.143}}$ 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

807