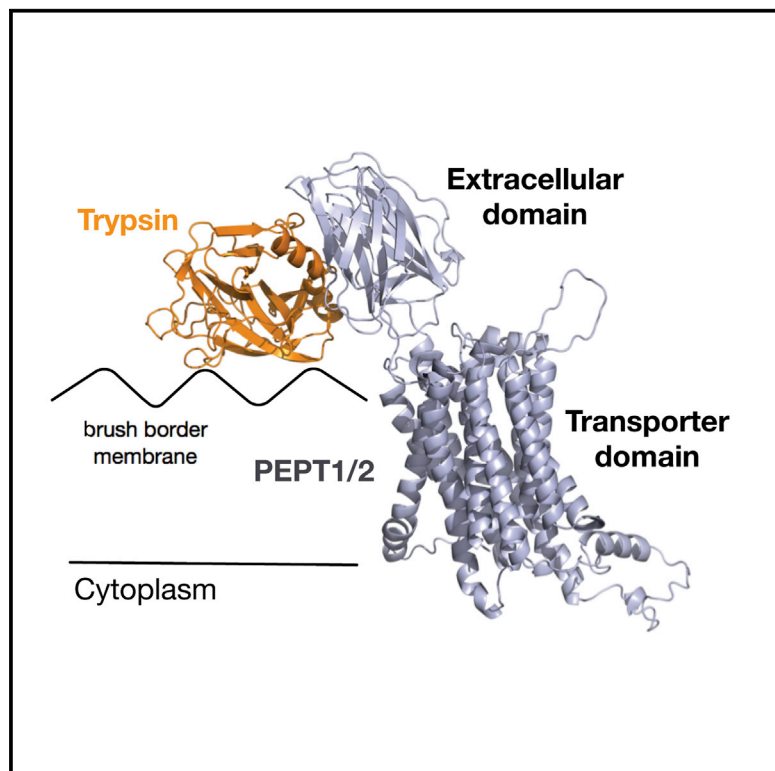# Structure

Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport

Graphical Abstract

Authors

John H. Beale, Joanne L. Parker, Firdaus Samsudin, ..., David Meredith, Philip W. Fowler, Simon Newstead

Correspondence

simon.newstead@bioch.ox.ac.uk

In Brief

The crystal structure of PepT1 and PepT2 reported by Beale et al. reveals two immunoglobulin-like domains connected in tandem inserted within the canonical major facilitator superfamily fold. Biophysical analyses reveal a specific interaction with trypsin, suggesting a role in clustering proteolytic activity to the site of peptide uptake across the membrane.

Highlights

- Crystal structure of the extracellular domains of PepT1 and PepT2
- Modular architecture for a mammalian MFS transporter
- Extracellular domains contain immunoglobulin-like fold and interact with trypsin

Accession Numbers

5A9D
5A9H
5A9I

CellPress

Structure

Article

CellPress

Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport

John H. Beale,[1,10] Joanne L. Parker,[1,10] Firdaus Samsudin,[1,10] Anne L. Barrett,[1,2] Anish Senan,[3] Louise E. Bird,[4,5] David Scott,[6,7,8] Raymond J. Owens,[4,5] Mark S.P. Sansom,[1,9] Stephen J. Tucker,[2,9] David Meredith,[4] Philip W. Fowler,[1] and Simon Newstead[1,9,*]

[1]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
[2]Clarendon Laboratory, Department of Physics, University of Oxford, Oxford OX1 3PU, UK
[3]Department of Biological Sciences, Oxford Brookes University, Oxford OX3 0BP, UK
[4]Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford OX3 0BP, UK
[5]OPPF-UK, Research Complex at Harwell, Harwell Oxford, Didcot, Oxfordshire OX11 0FA, UK
[6]Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0FA, UK
[7]ISIS Spallation Neutron and Muon Source, Rutherford Appleton Laboratory, Oxfordshire OX11 0FA, UK
[8]School of Biosciences, School of Biosciences, Sutton Bonington Campus, Leicestershire LE12 5RD, UK
[9]OXION Initiative in Ion Channels and Membrane Transport, University of Oxford OX1 3PU, UK
[10]Co-first author
*Correspondence: simon.newstead@bioch.ox.ac.uk
http://dx.doi.org/10.1016/j.str.2015.07.016

SUMMARY

Mammals obtain nitrogen via the uptake of di- and tri-peptides in the gastrointestinal tract through the action of PepT1 and PepT2, which are members of the POT family of proton-coupled oligopeptide transporters. PepT1 and PepT2 also play an important role in drug transport in the human body. Recent crystal structures of bacterial homologs revealed a conserved peptide-binding site and mechanism of transport. However, a key structural difference exists between bacterial and mammalian homologs with only the latter containing a large extracellular domain, the function of which is currently unknown. Here, we present the crystal structure of the extracellular domain from both PepT1 and PepT2 that reveal two immunoglobulin-like folds connected in tandem, providing structural insight into mammalian peptide transport. Functional and biophysical studies demonstrate that these domains interact with the intestinal protease trypsin, suggesting a role in clustering proteolytic activity to the site of peptide transport in eukaryotic cells.

INTRODUCTION

In mammals, the plasma membrane transporters PepT1 (SLC15A1) and PepT2 (SLC15A2) mediate the uptake and retention of dietary peptides (Adibi, 1997a; Matthews, 1991). PepT1 and PepT2 are proton-coupled symporters, recognizing di- and tri-peptides on the outside of the cell and utilizing the energy stored in the inwardly directed proton electrochemical gradient

($\Delta\mu H^+$) to drive their uptake into the cell (Daniel and Rubio-Aliaga, 2003; Fei et al., 1994). PepT1 and PepT2 also recognize and transport a number of important drug families, including β-lactam antibiotics and anti-cancer agents (reviewed in Brandsch, 2013), and are important targets in the ongoing attempts of the pharmaceutical industry to improve the pharmacokinetic properties of drug molecules (Brandsch, 2013; Smith et al., 2013).

PepT1 and PepT2 are members of the more widely distributed proton-dependent oligopeptide transporter, or POT, family (TC 2.A.17), which are evolutionarily well conserved from bacteria to man (Daniel et al., 2006). Structurally the POT family belongs to the major facilitator superfamily (MFS), with each member containing 12 transmembrane (TM)-spanning α helices arranged into two TM bundles of six that fold to resemble a V-shaped protein that resides within the inner membrane of bacteria and plasma membrane of eukaryotes (Figure 1) (Covitz et al., 1998; Fei et al., 1994; Yan, 2013). The MFS fold can be further subdivided; with each six-helix bundle being constructed from the inversion of two three-helix repeats (Radestock and Forrest, 2011). We recently proposed a structural framework for understanding the transport mechanism within the POT family based on the ability of the triple-helix repeats to work synergistically to alternate the central binding site to either side of the membrane (Fowler et al., 2015). Recent bioinformatics analyses, including sequence-based structure alignments supported by experimental structure validation, further strengthen the importance of the triple-helix repeats. These analyses show that functionally equivalent positions within the different MFS transporters crystallized to date superimpose in three dimensions. This observation has led to the suggestion that evolution within the MFS may have arisen through intragenic duplication and shuffling of these repeats (Madej et al., 2013; Madej and Kaback, 2013).

Crystal structures of bacterial POT family members have revealed a central peptide-binding site that is highly conserved with the mammalian homologs (Doki et al., 2013; Guettou
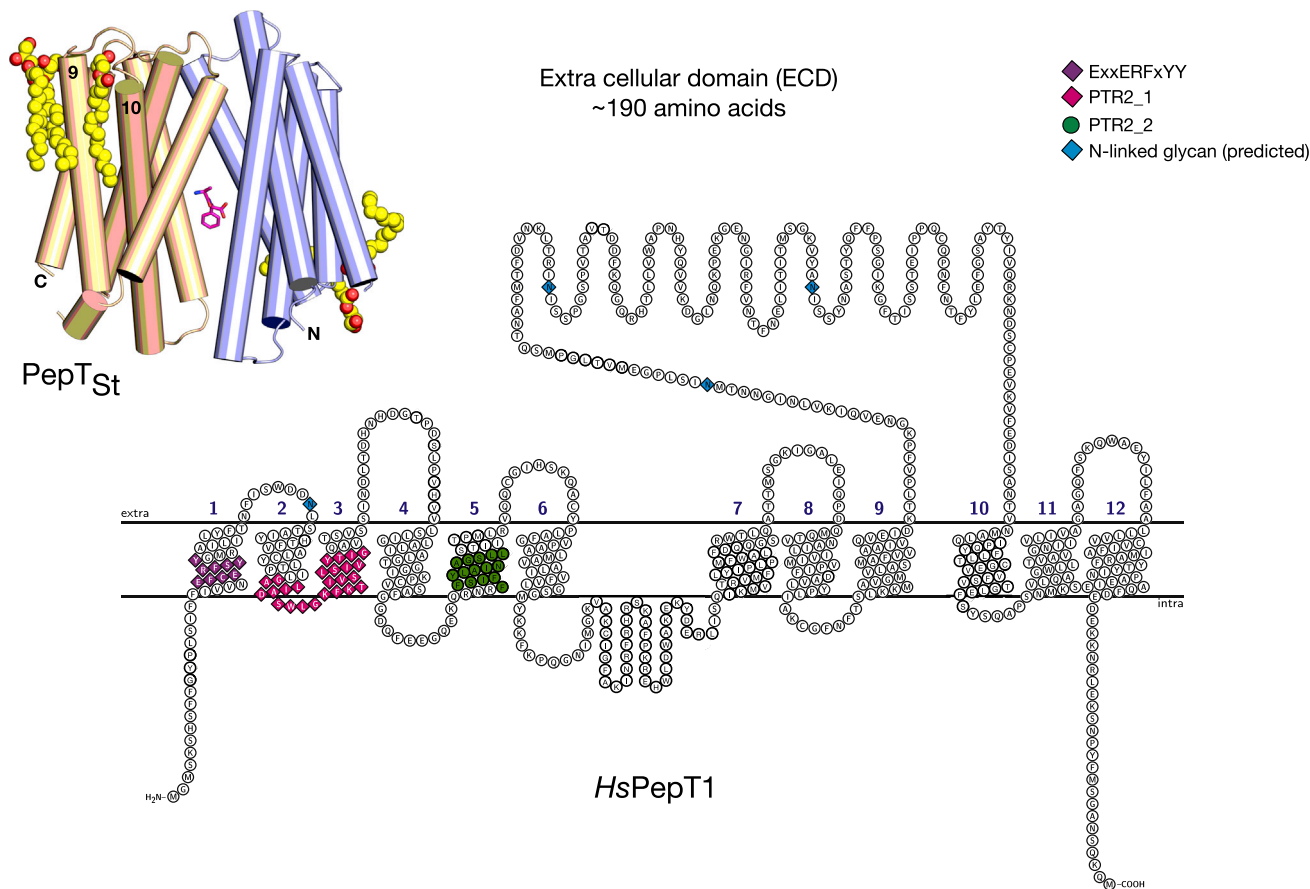
CellPress

Figure 1. Topology of Mammalian Peptide Transporters

Topology diagram of the human plasma membrane peptide transporter PepT1. Conserved PTR2/POT family signature motifs are indicated along with predicted N-linked glycosylation sites, three of which are in the extracellular domain. Inset: Crystal structure of the bacterial homolog PepT$_{St}$ (PDB: 4D2C). The N- (light blue) and C-terminal (wheat) domains are shown as cylinders, with the bound peptide indicating the location of the central peptide-binding site conserved between mammalian and bacterial proteins.

et al., 2014; Solcan et al., 2012). Both in vivo and in vitro assays have demonstrated that while a wide range of peptide substrates are transported by this family, there is conserved substrate specificity, with both bacterial and the mammalian proteins transporting hydrophobic peptides with approximate micromolar affinity and basic peptides with approximate millimolar affinity (Newstead, 2015). Between PepT1 and PepT2 there also exists a difference in overall substrate affinity, with PepT1 having a lower affinity for peptides and PepT2 a higher affinity (Smith et al., 2013). Recent crystal structures of peptide-bound complexes with a bacterial homolog of PepT1 have also suggested that peptides containing extended side chains, such as arginine and lysine, might adopt a less optimal position within the central peptide-binding site that could explain their lower affinity through less favorable interactions (Lyons et al., 2014).

Although the overall sequence identity between the mammalian and bacterial transporters is well conserved within their respective TM domains, epitope tagging analysis, supported by recent structure-based sequence alignments, reveal that in the mammalian PepT1 and PepT2 proteins there exists a significant portion of the protein that is positioned on the outside of the cell (Covitz et al., 1998; Newstead, 2015) (Figure 1). Intriguingly,

these domains are completely absent not only in the bacterial members of the POT family but also in the plant and fungal homologs (Parker and Newstead, 2014; Sun et al., 2014) (Figure S1). The functional role of this extracellular domain (ECD) is unknown; however, its presence would suggest the requirement for additional functionality to assist peptide uptake in mammals. Many eukaryotic channels and transporters have evolved to incorporate additional structural domains that extend, constrain, or regulate their function (Barabote et al., 2006). For example, eukaryotic CLC proteins contain intracellular CBS (cystathionine β-synthase) domains that regulate activity in response to nucleotides (Markovic and Dutzler, 2007; Meyer et al., 2007; Zifarelli and Pusch, 2009), whereas the voltage-sensing domains of potassium channels regulate channel opening following membrane depolarization (Pongs and Schwarz, 2010). To date, however, no equivalent domains have been described in detail for any member of the MFS, which forms the largest and most diverse family of secondary active transporters in biology (Reddy et al., 2012).

Here, we reveal that a previously annotated extracellular "loop" in early topology models of the human PepT1 and PepT2 transporters in fact consists of two immunoglobulin-like domains connected in tandem. In vitro binding assays

Table 1. Data Collection and Refinement Statistics for *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD}

	*Mm*PepT1^{ECD}	*Mm*PepT1^{ECD}-Hg[a]	*Rn*PepT2^{ECD}-Se[a]	*Rn*PepT2^{ECD}
Space group	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2$_1$	P3$_2$21	P4$_1$2$_1$2$_1$
Cell dimensions				
a, b, c (Å)	53.48, 70.33, 111.22	53.55, 70.37, 111.16	95.75, 95.75, 165.93	43.1, 43.1, 220.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 90
Wavelength (λ)	0.968	1.006	0.979	0.976
Resolution (Å)	43–2.10 (2.19–2.10)	70–2.85 (3.05–2.85)	58–2.81 (2.96–2.81)	43–2.06 (2.12–2.06)
R_{merge}	5.0 (79.8)	17.2 (75.6)	15.4 (109)	6.6 (67.1)
MnI/σI	16.4 (2.7)	15.1 (4.1)	12.7 (2.3)	11.5 (2.3)
CC$_{1/2}$[b]	99.9 (48.0)	99.7 (91.0)	99.9 (69.7)	99.8 (63.9)
Completeness (%)	99.5 (99.0)	99.4 (99.7)	99.9 (99.9)	99.3 (99.5)
Redundancy	4.8 (4.8)	14.0 (14.5)	9.9 (10)	4.1 (4.3)
R_{cullis} (%)		69.2	45.1	
Phasing power[c]		1.492	2.468	
Resolution (Å)	43.6–2.10		58.6–2.81	40.1–2.06
No. of reflections	24, 975		22, 048	13, 635
R_{work}/R_{free}	19.7/23.8		19.7/24.5	19.9/24.0
Ramachandran favored	96.6		92.5	96.3
Ramachandran outliers	0.53		0.17	0
Rmsd				
Bond lengths (Å)	0.010		0.010	0.010
Bond angles (°)	1.18		1.31	1.25

[a]For details on derivatization, see Experimental Procedures.
[b]Mn(I) half-set correlation as reported by Aimless.
[c]Phasing power = rms($|F_H|/((F_H + F_P) − (F_{PH})))$.

demonstrate that the ECDs interact with the intestinal protease trypsin, potentially answering an interesting observation made in the late 1960s of a specific and tight interaction of trypsin with human intestinal mucosa (Goldberg et al., 1968, 1969a, 1969b). More surprising was the observation that these domains can be removed with no appreciable loss of transport function. These results provide the first structural and biochemical insights into the mammalian SLC15 family, and demonstrate modularity within the MFS that could have important implications for interpreting the function of these proteins in eukaryotic cells.

RESULTS

Crystal Structures of the Extracellular Domain of PepT1 and PepT2

Using the crystal structures from the bacterial POT family proteins PepT$_{So}$ (Newstead et al., 2011) and PepT$_{St}$ (Solcan et al., 2012) that share 31% and 21% identity, respectively, with their mammalian homologs, we identified the probable location of the ECD in the PepT1 transporters (Figure S1). We subsequently identified the ECD from *Mus musculus*, consisting of residues 391–580, as being stable and amenable to structural and biophysical analysis (Figure S2). This domain was crystallized and its structure determined using the single anomalous dispersion (SAD) method of phasing using mercury-derivatized protein. The structure was refined to a resolution of 2.1 Å with final R_{work} and R_{free} of 19.7% and 23.8%, respectively (Table 1). Following extensive screening, crystals from *Rattus norvegicus* PepT2,

residues 410–601, were also obtained (Figure S2). The phases were calculated from seleno-L-methionine incorporated protein using the SAD method in space group P3$_2$21 (Table 1). A higher-resolution structure in space group P4$_1$2$_1$2 was obtained using the P3$_2$21 crystals as seeds with a single monomer in the asymmetric unit. The final structure refined to a resolution of 2.06 Å with final R_{work} and R_{free} of 19.9% and 24.0%, respectively.

The crystallographic asymmetric unit of the PepT1^{ECD} crystal contained two monomers that formed a head-to-tail dimer, related by a two-fold non-crystallographic symmetry axis (Figure 2A), whereas the PepT2^{ECD} construct was crystallized in a monomeric state (Figure 2B). The overall structure of both ECDs consists of two compact β-sandwich immunoglobulin-like folds each comprising two four-stranded β sheets. The β sandwiches are composed of strands in the order 4-1-7-8 and 3-2-5-6, with a short connecting loop between the end of strand β-8 on lobe one and β-9 on lobe two. Despite having only 22% sequence identity the two structures adopt the same overall structure, superimposing with a root-mean-square deviation (rmsd) of 1.83 Å. Analytical ultracentrifugation on the purified ECDs show that both PepT1^{ECD} and PepT2^{ECD} are monomers in solution (Figure 2C). This led us to suspect that the head-to-tail dimer observed in the asymmetric unit was the result of crystallization and that the physiological state of the ECD in PepT1 is monomeric, as shown for the PepT2^{ECD}. The bilobal architecture of the ECDs suggested that the two immunoglobulin-like domains have the potential to be highly dynamic, which could represent an important structural and functional difference between PepT1 and pepT2.
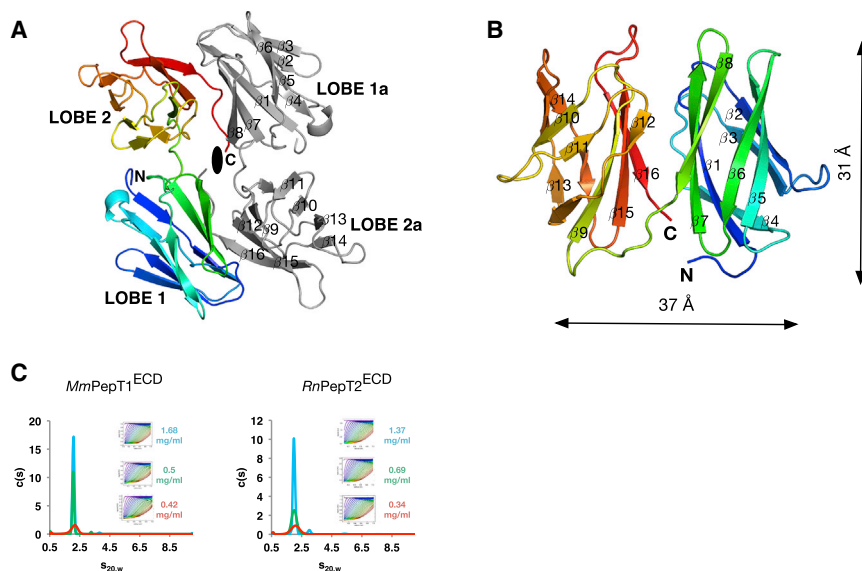
Figure 2. Crystal Structure of the Extracellular Domain from PepT1 and PepT2

(A) The asymmetric unit of MmPepT1^{ECD} containing two monomers related by a two-fold non-crystallographic symmetry axis (black oval). One monomer is rainbow colored from the N terminus to the C terminus; the second is shown in gray with the secondary structure labeled from β1 to β16.

(B) Structure of the RnPepT2^{ECD} colored from the N (blue) to the C terminus (red) and with the secondary structure components labeled as for (A).

(C) The $s^0_{20,W}$ values of MmPepT1^{ECD} and RnPepT2^{ECD} from the AUC analysis are 2.16 and 2.22, respectively, consistent with both proteins migrating as a 20-kDa monomer in solution. Inset: the Lamm equation fit profiles for MmPepT1^{ECD} and RnPepT2^{ECD}.

Salt Bridge Interactions Stabilize the Interface between the Immunoglobulin-like Domains in PepT1

The crystal structures revealed that in the PepT1^{ECD} there exist two conserved salt bridge interactions stabilizing the interface between lobe 1 and lobe 2, mediated by Asp574-Lys398 and Asp476-Arg490, whereas in PepT2^{ECD} only one equivalent salt bridge interaction is present, between Asp505 and Arg538 (Figures 3A and 3B). To investigate the nature of these interactions, we inserted a 3C protease site into the linker connecting lobe 1 and lobe 2 in PepT1^{ECD} (Figure 3C). Following cleavage by the protease 3C of the purified protein, we observed that the two lobes of PepT1^{ECD} remain associated down a size-exclusion chromatography column. Repeating this experiment with the Asp574Ala variant, however, resulted in the two lobes migrating independently, thus confirming the importance of the salt bridge in holding the two immunoglobulin-like domains together in the compact arrangement shown in Figure 3A.

The same experiment, however, could not be conducted with PepT2^{ECD} as the lobes proved too unstable after cleavage with 3C protease. Therefore, to understand the behavior of PepT2^{ECD} in solution we used small-angle X-ray scattering (SAXS), a method that allows the overall shape of a macromolecule to be modeled at low resolution. Guinier analysis of the scattering data in PRIMUS (Petoukhov et al., 2007) shows that PepT2^{ECD} has a larger radius of gyration (R_g) compared with PepT1^{ECD}, 23.1 versus 18.4 Å (Table 2), which suggests an increase in particle size. A shift to larger scattering distances can also be seen in the P(r) distribution and dimensionless V_c Krakty plot (Rambo and Tainer, 2011) (Figures S3A and S3B), indicating a more elongated structure for PepT2^{ECD}. Consistent with this analysis, 3D envelopes of the ECD generated using DAMMIF (Franke and Svergun, 2009) show that PepT1^{ECD} forms a compact shape ~48 Å in length whereas PepT2^{ECD} is more elongated, approximately 61 Å in length (Figures 3D and S3C). The larger envelope of PepT2^{ECD} indicates that the two immunoglobulin-like domains are structurally more dynamic than the PepT1^{ECD}, which is consistent with the loss of the second salt bridge and also the location of Asp505 on the unstructured loop connecting lobes 1 and 2.

PepT1 and PepT2 Contain a Functionally Independent Extracellular Domain

To date there have been no studies on the role of the ECDs in either of the mammalian peptide transporters. This is largely due to the ambiguity in identifying where this domain was located with respect to the TM helices. The crystal structure of the ECDs from both PepT1 and PepT2 now allow for structure-based homology models of the full-length human transporters to be built (Figure 4A). Molecular dynamics of the models in a palmitoyloleoyl phosphatidylglycerol membrane bilayer suggest that the domains are likely to adopt a vertical orientation sitting to one side of the transporter (Figure S4).

PepT1 and PepT2 are the first MFS transporters to date that have been shown to contain a folded structural domain inserted within the "core" MFS fold (Yan, 2013). This raises the important question of whether the ECDs play any role in transport or determine functional differences between PepT1 and PepT2. To investigate this question, we used the homology models to generate a number of different combinations of chimeric and mutant human PepT1 and PepT2 transporters that had their ECD domains removed (PepT1^{ΔECD}, PepT2^{ΔECD}) or swapped (PepT1^{ECD2}, PepT2^{ECD1}). The resulting constructs were expressed in Xenopus oocytes and their relative levels of expression examined (Figure S5). We observed that PepT1 was very sensitive to modification at or within the ECD, with only the wild-type (WT) showing stable levels of expression. In contrast, we observed high levels of expression for all of the PepT2 constructs, allowing us to investigate the role of this ECD. Peptide transporters are electrogenic carriers able to concentrate peptides inside the cell using the energy stored in the ΔμH$^+$ (Fei et al., 1994). We therefore examined their transport properties using two-electrode voltage-clamp (TEVC) recordings in the Xenopus laevis oocytes. This technique measures the inward movement of H$^+$ as a function of external peptide concentration. K_M values for transport of the non-hydrolyzable peptide glycyl-sarcosine (Gly-Sar) were determined for both WT PepT1 and PepT2. The calculated K_M values confirm the previously noted difference in peptide transport of 0.78 ± 0.09 mM and
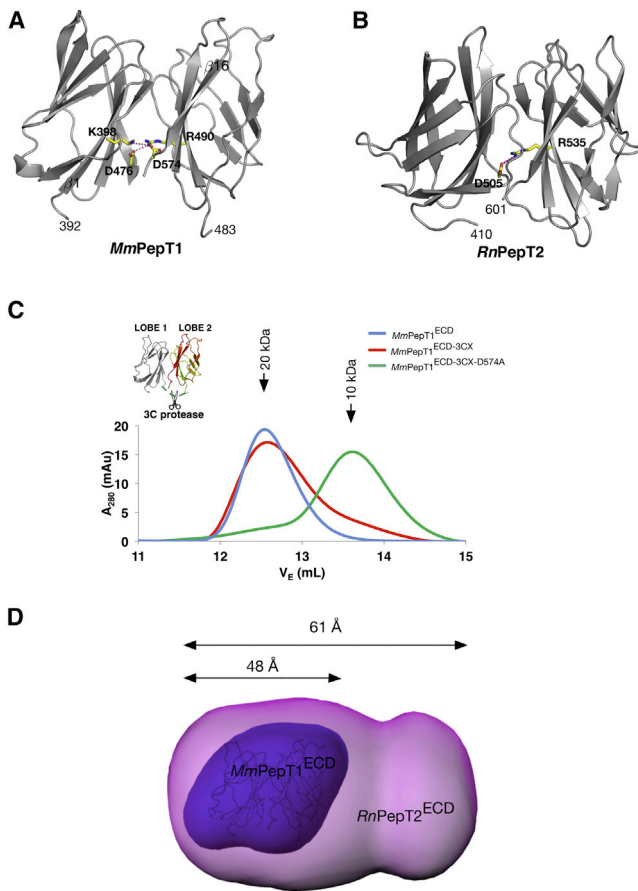
Figure 3. Salt Bridges Stabilize the Interface between the Two Immunoglobulin-like Domains in PepT1^{ECD} and PepT2^{ECD}

(A) Structure of PepT1^{ECD} illustrating the two salt bridges, K398 and D574 and R490 and D476, that form an interaction between the lobes.

(B) Comparative view in *Rn*PepT2^{ECD}, where a single salt bridge is observed between Asp505 and Arg518.

(C) Size-exclusion chromatography traces from the *Mm*PepT1^{ECD}-3CX experiment. The cleaved *Mm*PepT1^{ECD}-3CX constructs elute at the same volume as wild-type, showing that the lobes still interact in solution even after the two lobes are separated. The cleaved *Mm*PepT1^{ECD}-3CX-D574A construct, however, elutes in a larger volume consistent with disruption of the interaction.

(D) DAMMIN envelopes of *Mm*PepT1^{ECD} (dark purple) and *Rn*PepT2^{ECD} (light purple) calculated from the SAXS data, which show lengths of 48 and 61 Å, respectively, and illustrate the more dynamic behavior of PepT2^{ECD}. For scale, a black and white outline of A is overlaid on the *Mm*PepT1^{ECD} envelope.

0.32 ± 0.02 mM for PepT1 and PepT2, respectively (Figure 4B). However, swapping the ECD region from PepT1 to PepT2 (PepT2^{ECD1}) or even deleting it entirely from PepT2 (PepT2$^{\Delta ECD}$) resulted in no detectable change in the K_M for uptake of the Gly-Sar peptide, which remained ~0.3 mM. Additionally, removal of the salt bridge that stabilizes the interface between lobes 1 and 2 of PepT1 (PepT1^{D573A}) made no observable difference to the K_M of Gly-Sar transport (Figure 4C). We also tested whether these constructs interacted differently with the antibiotic cefaclor, a known substrate of PepT1 and PepT2 (Bretschneider et al., 1999; Luckner and Brandsch, 2005). We found that the K_i of drug uptake in *cis*-inhibition experiments was

also unchanged for all the PepT2 constructs, with a K_i of ~0.22 mM, compared with ~0.9 mM for WT PepT1 (Figure 4D). This similarity was also shown with a different non-hydrolyzable peptide, lysyl-lysine. These results indicate that the ECD is a structurally independent unit that has no obvious role in substrate specificity or intrinsic peptide transport function, in PepT2 at least. The immunoglobulin-like structure of the ECDs, however, strongly suggested a possible role in binding an external component that might be present in the intestinal lumen. Therefore, to further investigate the function of the ECDs we undertook a series of binding studies to investigate this possibility.

The Extracellular Domain Interacts with the Intestinal Protease Trypsin

Given the expression of both PepT1 and, to a lesser extent, PepT2 in the intestinal brush border membrane, we thought it possible that the interaction partners may be intestinal hormones, including cholecystokinin (CCK8) and thyroid hormone, which has been reported to decrease the surface expression of PepT1 in Caco-2 cells (Ashida et al., 2002). Another possibility was an interaction with the intestinal proteases that create the di- and tri-peptides that are subsequently recognized by PepT1 and PepT2 for transport into the cell. We therefore screened a panel of potential interaction candidates using surface plasmon resonance (SPR). From this panel of potential ligands we observed binding only for the intestinal protease trypsin with both mouse PepT1^{ECD} (K_D 80 ± 0.8 μM) and rat PepT2^{ECD} (K_D 165 ± 0.5 μM) (Figures 5A and 5B). However, we observed no significant interaction between PepT1^{ECD} and chymotrypsin or pepsin. It is possible that one function of the ECDs could be to interact with and accumulate peptides in the vicinity of the transporter, although this would be unlikely given the TEVC data. However, to test this we used microscale thermophoresis (MST), which can measure the interaction with small ligands with transporters in solution (Parker and Newstead, 2014). We did not observe an interaction with the peptide Gly-Sar, but this technique did confirm our previous SPR data showing a specific interaction with trypsin (Figure 5A). The K_D calculated using this technique, however, was tighter than that observed using SPR, being 8 ± 0.7 μM for PepT1^{ECD} and 6 ± 0.5 μM for PepT2^{ECD}. We interpret this discrepancy with the SPR values as being due to the freedom of interaction of a measurement in solution versus immobilization on a chip surface. To examine the nature of the interaction between trypsin and the ECDs, we repeated the MST binding assay with an increased salt concentration of 0.5 M. This resulted in abolition of the interaction, suggesting that the interaction between trypsin and the ECD is mediated through an electrostatic interface.

Trypsin Recognition Is Localized to a Conserved Di-aspartate Motif on the Extracellular Domain

To identify possible binding sites on the PepT1^{ECD} and PepT2^{ECD} molecules, we mapped the sequence conservation for six different mammalian homologs (Figure S1) onto the crystal structures. We identified two highly conserved charged residues, Asp550 and Glu573 in PepT1^{ECD} and Asp576 and Glu599 in PepT2^{ECD}, located on one face of the ECD structure (Figures 5D and 5E). Both of these residues are found in two conserved sequence motifs at the start of strand β15 and at the end of

CellPress

Table 2. SAXS Data Statistics for *Mm*PepT1^ECD and *Rn*PepT2^ECD

	R_g (Å)	$I(0)$/Conc (mg ml^{-1})	V_p (nm^3)	D_{max} (Å)	V_c (Å^2)	Mass (kDa)
*Mm*PepT1^ECD	18.4 ± 1.7	8.6 ± 0.0	33.4	64.3	197	17.1
*Rn*PepT2^ECD	23.0 ± 2.8	14.9 ± 0.0	45.4	73.1	238	20.1

The R_g, $I(0)$/Conc, V_p, and D_{max} were calculated in PRIMUS. The V_c and particle mass were calculated in ScÅtter, and show that *Rn*PepT2^ECD has a larger radius of gyration (R_g) in solution, indicating a more flexible arrangement for the two lobe domains.

strand β16 in lobe 2 of the mammalian proteins (Figure S6). Removal of these charges abolished the interaction with trypsin (Figures 5D and 5E, insets), further supporting the hypothesis that the function of the ECD is to recruit trypsin to the site of the peptide transport on the plasma membrane. The other face of the ECD, however, does not contain any conserved charged residues, suggesting that this face in unlikely to be important in an electrostatic interaction. Indeed, mutation of several surface residues in mouse PepT1^ECD resulted in no substantial impact on the K_D of trypsin binding compared with WT protein (Figure S7). The involvement of only two conserved residues in the trypsin interaction may also explain the fast binding kinetics observed in the SPR sensorgrams (Figures 5A and 5B, insets), which suggest that the interaction with trypsin is highly dynamic and likely transient, rather than forming a long-lived complex in the intestinal lumen.

DISCUSSION

The Modular Architecture of PepT1 and PepT2

Mammalian peptide transport is a physiologically important route to both assimilate dietary nitrogen in the form of small di- and tri-peptides from ingested protein and retain peptides in the body by selective reabsorption in the kidneys (Adibi, 1997b; Matthews, 1975; Matthews, 1991). Although originally identified in 1994, the 3D structures of PepT1 and PepT2 have remained elusive, and to date no crystal structures are available. However, recent crystal structures of several closely related bacterial homologs provide suitable templates to model the TM domain of the mammalian proteins (Newstead, 2015; Terada and Inui, 2012). Nevertheless, the identity of the ECD, originally identified from hydropathy analysis and epitope insertion studies (Covitz et al., 1998), has remained elusive. Our functional and structural analysis shows that the ECD is a fully independent module that has been incorporated by the mammalian members of the POT family to function in protein-protein interactions on the outside of the cell. Additional structural domains are found in many families of membrane transporters (Barabote et al., 2006). However, the locations of the ECDs within the PepT1 and PepT2 structures are particularly intriguing, in that they are inserted *within* the canonical 12-TM helix MFS fold. This is highly unusual; most additional domains in transporters are appended to either the N- or C-terminal end of the polypeptide chain (Markovic and Dutzler, 2007; Meyer et al., 2007; Warmuth et al., 2009). To our knowledge, this study reports the first structure of an additional domain inserted within the core architecture of a transporter. Interestingly the insertion site within PepT1 and PepT2 is the connection between TM9 and TM10, which represents the junction between the two triple-helix repeats that make up the C-terminal bundle within the MFS fold (Radestock

and Forrest, 2011). We recently suggested that these repeats operate in a scissor-like motion that control access to the central peptide-binding site during transport (Fowler et al., 2015). The insertion of the two immunoglobulin-like domains of the ECD, with minimal impact on the ability of PepT1 and PepT2 to function as proton-coupled peptide symporters, would appear to support our hypothesis that the two repeats operate in a coordinated but structurally independent manner.

However unusual it may be to observe the ECD inserted within the canonical MFS fold, our finding that both the PepT1^ECD and PepT2^ECD interact with trypsin presents a logical role for these domains in mammalian peptide import. It is interesting to note that as far back as 1968 both trypsin and chymotrypsin were observed to interact with and bind to the mucosa of human small intestine (Goldberg et al., 1968, 1969a, 1969b). The assay conditions used here to investigate the binding between the ECD and trypsin closely resemble the pH and ionic strength found in the small intestine (~150 mM NaCl and pH ~6.5) (Fallingborg, 1999; Fordtran et al., 1968), suggesting that the K_D values reported are likely to be in the physiological range. Furthermore, the K_D values reported in this study (μM range) are also consistent with the estimated concentration of trypsin in the small intestine, which was reported to be ~7 μM (Goldberg et al., 1969a). Taken together, these observations support the presence of a physiological interaction between the ECD of PepT1 and PepT2 and trypsin in the human body. Our data show that a conserved di-acidic motif on both PepT1^ECD and PepT2^ECD presents the most likely interaction site with trypsin. However, the fast binding kinetics show that the interaction is highly dynamic and likely to be transient in nature. We have combined all of the information presented to generate a working hypothesis for how trypsin might interact with PepT1 and PepT2 at the plasma membrane (Figure 6). Our data place trypsin on the opposite side of the ECD to the peptide translocation pathway (Newstead, 2015). In this configuration the binding of trypsin would not obstruct subsequent peptide transport, as the protease would be on the opposite side of the transporter from the peptide-binding site.

Why would PepT1 and PepT2 have evolved to localize trypsin rather than other proteases, such as chymotrypsin? PepT1 and PepT2 play an important physiological role in absorbing small peptides arising from digestion of dietary proteins in the small intestine, as well as in reabsorbing filtered peptides generated from luminal peptidases in the kidney (Adibi, 1997a, 1997b). There is clearly an advantage in tethering a protease to the site of peptide uptake on the plasma membrane, in that the peptides will be locally concentrated at the site needed for their recognition and transport. The localization of trypsin, which recognizes and cleaves the peptide chain at arginine and lysine residues, would therefore increase the concentration of
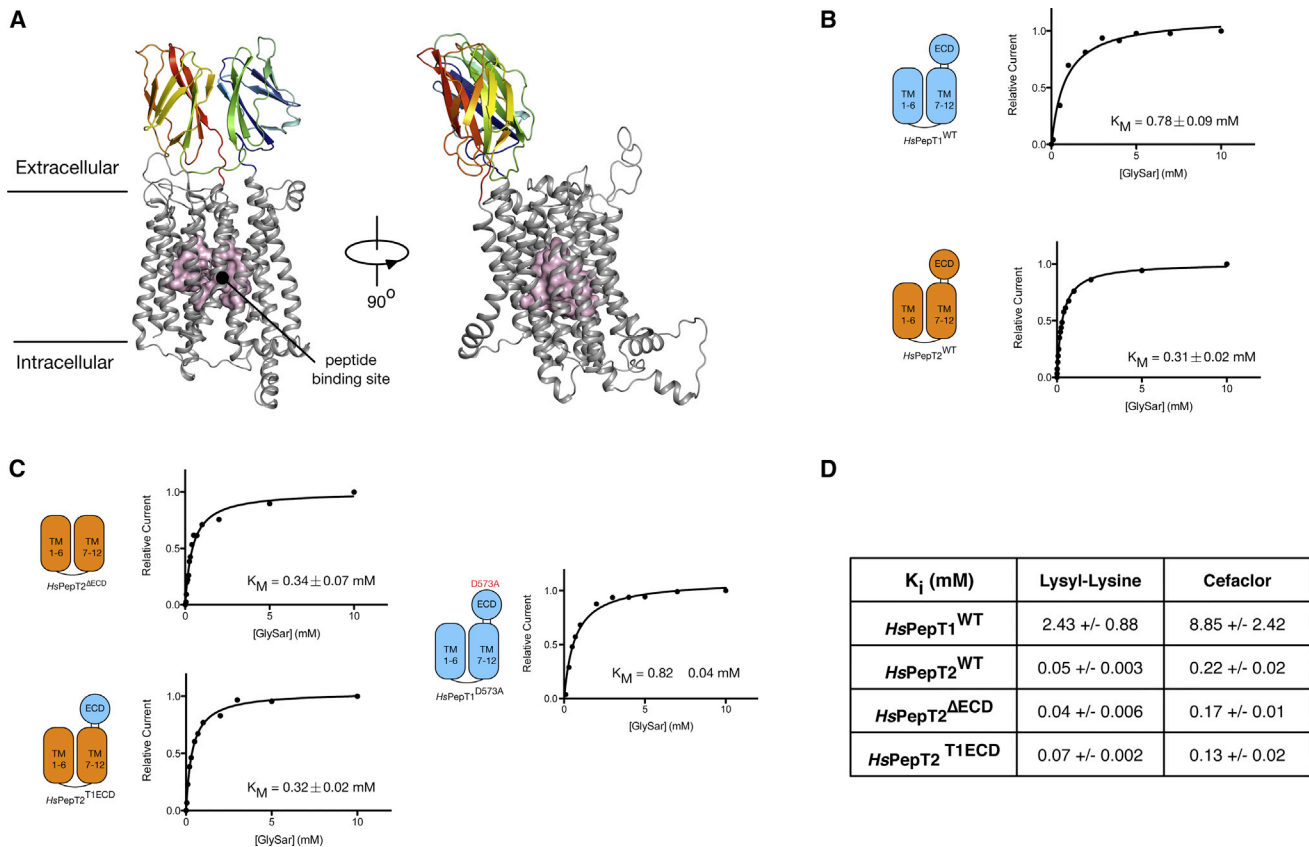
CellPress

Figure 4. PepT1 and PepT2 are Modular Proteins with Functionally Distinct Domains

(A) Homology model of the human PepT1 transporter generated using the crystal structure of MmPepT1^{ECD} (colored blue to red as in Figure 1B) and the recently determined bacterial homolog PepT$_{So}$ representing the transmembrane portion of the transporter (shown in gray). The peptide-binding site is highlighted (magenta).

(B) Kinetic analysis of Gly-Sar uptake in human PepT1 and PepT2 using the TEVC method.

(C) Kinetic analysis of Gly-Sar uptake in the PepT2$^{\Delta ECD}$, PepT2^{T1ECD}, and PepT1^{D573A} constructs.

(D) K_i values for the different constructs for lysyl-lysine and cefaclor are shown, indicating no effect of removing the ECD on peptide or drug uptake in PepT2.

K_i (mM)	Lysyl-Lysine	Cefaclor
HsPepT1^{WT}	2.43 +/- 0.88	8.85 +/- 2.42
HsPepT2^{WT}	0.05 +/- 0.003	0.22 +/- 0.02
HsPepT2$^{\Delta ECD}$	0.04 +/- 0.006	0.17 +/- 0.01
HsPepT2^{T1ECD}	0.07 +/- 0.002	0.13 +/- 0.02

peptides containing these side chains on the outside of the cell directly above the peptide-binding site, and would be expected to improve the efficiency of their uptake through the transporter domain. Indeed, clinical studies of peptide transport in the human body have indicated that the transport of arginine-containing peptides is less efficient (Steinhardt and Adibi, 1986), consistent with our findings from the bacterial homolog of PepT$_{St}$ (Solcan et al., 2012), suggesting that the localization of trypsin is an adaption to increase the concentration of peptides containing arginine and lysine, and therefore improve the transport of these peptides into the cell. We hasten to acknowledge, however, that our data do not unambiguously demonstrate a physiological requirement for trypsin in peptide transport and that our hypothesis will require further in vivo study for it to be supported or refuted.

To our knowledge, this study represents the first structural insight into MFS transporters that exist as multi-domain proteins. Interestingly, a similarly sized ECD is also observed in the SLC22 family of MFS transporters, again inserted between TM9 and TM10 (Kalliokoski and Niemi, 2009), suggesting that other eukaryotic MFS transporters have adopted similar mecha-

nisms to extend their functionality in the cell. The present study and homology models of the human PepT1 and PepT2 transporters therefore not only establish a framework for understanding mammalian peptide uptake, but also show how members of the MFS have evolved to incorporate additional structural domains that expand, augment, or constrain their function in eukaryotic cells.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification of the Extracellular Domains from PepT1 and PepT2

PepT1^{ECD} from $M. musculus$ (residues 391–580, UniProtKB: Q9JIP7) was cloned into an N-terminal maltose-binding protein (MBP) fusion expression vector, pOPINM (Berrow et al., 2007). PepT2^{ECD} from $R. norvegicus$ (residues 410–601, UniProtKB: Q63424) was cloned into a different MBP fusion expression vector, pLou3, a derivative of pMAL-c5 vector with a tobacco etch virus protease site, to remove the MBP and an N-terminal histidine tag. Recombinant protein was produced in $Escherichia coli$ strain BL21-DE3. Isopropyl β-D-1-thiogalactopyranoside was used to induce expression of the recombinant genes; cells were harvested following overnight induction at 25°C. WT and mutant proteins were purified to homogeneity using standard protocols for Ni-immobilized metal affinity chromatography- and amylose-based
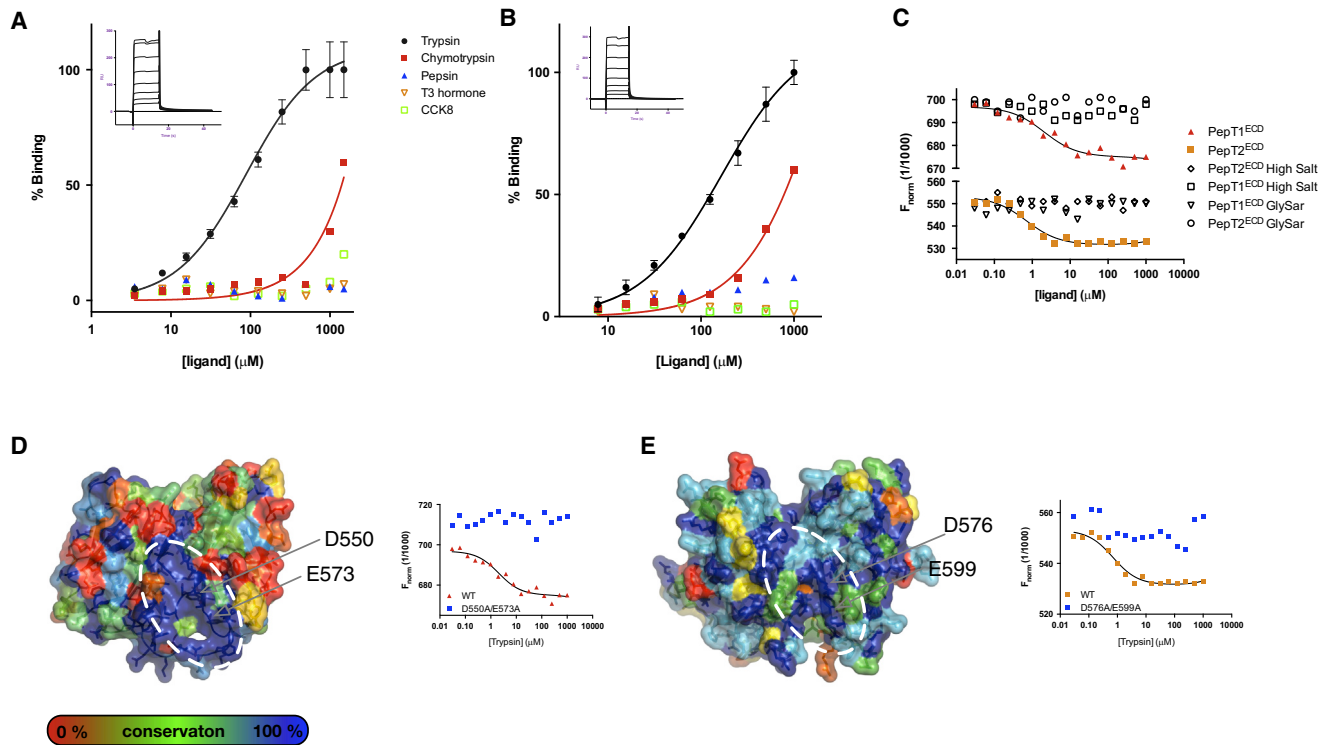
Figure 5. Trypsin Interacts with a Di-acidic Motif on the Extracellular Domain of PepT1 and PepT2

(A) SPR analysis of the MmPepT1^{ECD} interaction with trypsin. Inset: SPR sensorgram used to determine the binding constant. RU, response units. Error bars show the SEM (n = 3).

(B) The binding experiment in (A) was repeated with the RnPepT2^{ECD} protein.

(C) MST binding analysis reveals no interaction with the Gly-Sar peptide and abolition of trypsin interaction in the presence of high salt.

(D and E) Surface representation of (D) MmPepT1^{ECD} and (E) RnPepT2^{ECD} with the sequence conservation from cow, dog, chicken, human, mouse, and rat species mapped from blue to red. A highly conserved patch (indicated by the white dashed ellipse) was identified. Insets: MST binding analysis reveals an important role for D550 and E573 in MmPepT1^{ECD}, and D576 and E599 in RnPepT2^{ECD}, in mediating the electrostatic interaction with trypsin.

purification of the fusion proteins. Seleno-L-methionine-incorporated PepT2^{ECD} was produced using an auto-inducing medium, PASM-5052 (Studier, 2005).

Crystallization and Structure Determination

PepT1^{ECD} was crystallized in 20% polyethylene glycol (PEG) 6000, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) (pH 6.0), 0.2 M ammonium chloride, at 10 mg ml^{-1} and 4°C using sitting drop-vapor diffusion plates. All crystals were cryo-protected in mother liquor with 25% glycerol and cryo-cooled in liquid nitrogen for data collection. Diffraction data were collected on beamlines I24, I03, and I04 at Diamond Light Source, Harwell, UK. Initial data processing was carried out using the Xia2 pipeline (Winter et al., 2013) to XDS (Kabsch, 2010). Initial phases for PepT1^{ECD} were calculated using SAD with a single mercury-derivatized crystal. The space group was determined to be P2$_1$2$_1$2$_1$. Three mercury sites were initially located using SHELXC/D (Sheldrick, 2010), with their positions further refined and initial phases calculated using SHARP with solvent flattening in SOLOMON (Abrahams and Leslie, 1996). The complete primary structure was assigned in Coot and refined using BUSTER (Blanc et al., 2004) to a final resolution of 2.10 Å and R_{work}/R_{free} of 19.7/23.8 (Table 1).

RnPepT2^{ECD} was crystallized in 0.2 M (NH$_4$)$_3$ citrate (pH 5.8) and 21% PEG 3350, 10 mg ml^{-1} at 20°C. Initial phases were calculated using Se-SAD data using autoSOL from the PHENIX crystallography suite (Adams et al., 2010). A starting model was built using PHENIX Auto-Build showing three molecules in the asymmetric unit in space group P3$_2$21. The complete structure of the three molecules was built in Coot based on this initial map. Refinement of the structure was carried out in BUSTER to a final resolution of 2.81 Å with

an R_{work}/R_{free} of 19.7/24.5. To improve the resolution of the PepT2^{ECD} structure, we re-screened the original sparse matrix crystal screens using a seed stock generated from the P3$_2$21 crystals. Crystals grew in 0.2 M CsCl$_2$ and 15% PEG 3350. A dataset was collected on beamline I03 at Diamond Light Source and processed using XDS to a resolution of 2.06 Å in a new space group, P4$_1$2$_1$2. The phases for this new structure were calculated by molecular replacement of the monomeric PepT2^{ECD} structure in Phaser (McCoy et al., 2007). The model was refined in BUSTER to a final R_{work}/R_{free} of 19.9/24.0.

Homology Models of Human PepT1 and PepT2

Homology models of the human PepT1 (UniProt: P46059) and PepT2 (UniProt: Q16348) transporters were built consisting of the crystal structure of a bacterial peptide transporter, PepT$_{So}$ (Newstead et al., 2011) (PDB: 2XUT) as the template for the transmembrane region, and the crystal structures of the ECD regions inserted within the extracellular loop connecting TMH9 and TMH10. The two extra helices, HA and HB (residues 226–285), which are only present in a subset of the prokaryotic POT family transporters, were removed prior to the sequence alignment. Initial alignment was generated using Probcons (Do et al., 2005), and this was manually refined in Jalview (Waterhouse et al., 2009) to correctly align the functionally important residues identified previously through functional studies on the eukaryotic and mammalian homologs.

Expression and Functional Characterization of Human PepT1 and PepT2 in *Xenopus* Oocytes

Open reading frames for the human PepT1 and PepT2 proteins were cloned into the pBF *Xenopus* oocyte expression vector, which adds the 5′ and 3′
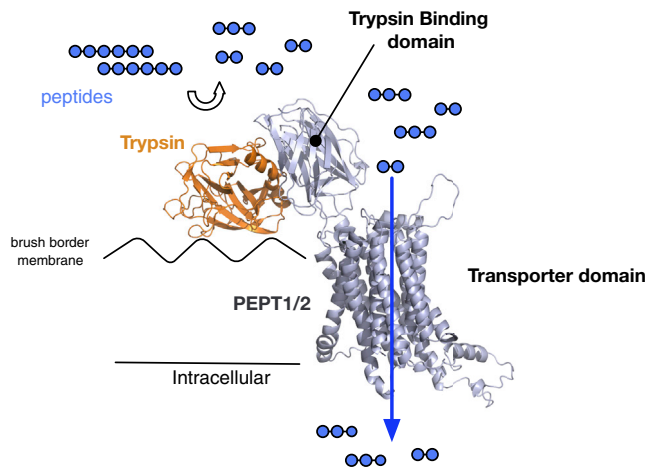
Figure 6. A Model for the Interaction between Trypsin and the Mammalian Peptide Transporters

During protein digestion in the small intestine, trypsin transiently docks onto the conserved di-acidic motif on the trypsin-binding domain, localizing the protease to the main site of peptide import on the brush border membrane. Localization would create an increase in the local concentration of arginine- and lysine-containing peptides (shown here as blue circles), which would be expected to increase the efficiency of their uptake into the cell.

UTRs of the *Xenopus* β-globin gene. ECD deletions and chimeras were generated using extension overlap PCR. A C-terminal FLAG tag epitope was engineered at the C terminus of these genes to aid detection and quantitation of expression by western blot. mRNA for injection was prepared by in vitro transcription using the AmpliCap SP6 High Yield Message Maker kit (Cellscript). *Xenopus* oocytes were injected with 50 ng of mRNA and incubated at 17°C for 3–4 days before recording. For TEVC recordings, microelectrodes were filled with 3 M KCl and had tip resistances of ~5 MΩ. Oocytes were voltage-clamped at −50 mV, and a voltage step protocol consisting of 300-ms long pulses from −160 to +60 mV in 10-mV increments was used to test oocyte stability and record currents. ND96 solution was used as the bath solution (95.4 mM NaCl, 2 mM KCl, 1.8 mM CaCl₂, 5 mM HEPES [pH 7.5] with NaOH). Test solutions consisted Gly-Sar at desired concentrations in ND96 solution. Transport-associated currents were estimated by subtracting currents recorded on ND96 solution only from those in the presence of substrate for each experiment. The currents at each Gly-Sar concentration were normalized to the maximal current (taken as current observed with 10 mM Gly-Sar) for each oocyte and the current at −120 mV at each Gly-Sar concentration used in non-linear regression analysis. Data were analyzed using Clampfit 10.4.0.36 (Axon Instruments) and non-linear regression analysis was performed on data from single oocytes using GraphPad Prism (version 6.04 for Mac, Graph-Pad Software). Acquired data were fitted to the Michaelis-Menten equation $I = (I_{max}[\text{Gly-Sar}])/(K_M + [\text{Gly-Sar}])$, where current was used in the place of velocity to calculate an apparent K_M for data from each oocyte. Average K_M was then taken as the mean of the K_M values from individual oocytes expressing the specific construct and displayed as mean ± SEM. Background currents were measured in uninjected oocytes 4 days after incubation. Inhibition experiments with the β-lactam antibiotic cefaclor and the dipeptide lysyl-lysine determined *cis*-inhibition of [³H]-D-Phe-L-Gln uptake by increasing concentrations of cefaclor or lysyl-lysine, using the method of Pieri et al. (2009). 10 ng of mRNA was injected into each oocyte.

Small-Angle X-Ray Scattering

SAXS data were collected at the P12 beamline (PETRA III, Hamburg) at a wavelength of 1.24 Å. Scattered images were collected on a PILATUS 2M detector (Dectris) at 297 K. Twenty images were taken while the sample was continually passed through a quartz capillary. Samples and buffers were prepared using the same protocol used for the area under the curve (AUC) analysis. A concentration series was collected for each sample from 5 to 0.15 mg ml⁻¹. Raw

images were individually examined using PRIMUS (Konarev et al., 2003), and images that showed radiation damage were excluded from the final averaged curves. No concentration-dependent scattering was observed in for PepT1ECD. Merged curves were created for PepT2ECD, as an increase in scattering was observed as a function of concentration in the low scattering angles. To interpret the scattering data, 20 3D models of each ECD were created using DAMMIF (Franke and Svergun, 2009), which were then aligned, clustered, and averaged using DAMAVER (Volkov and Svergun, 2003). DAMMIN (Svergun, 1999) was then used to compare the averaged model against the raw data to ensure a good fit as evaluated using reduced χ^2 values. Envelopes of the models were generated using Sculptor (Birmanns et al., 2011).

Surface Plasmon Resonance

SPR experiments were carried out using a BIAcore T200 instrument (GE Healthcare). Experiments were performed at 20°C in 25 mM MES (pH 6.5), 100 mM NaCl, 10 mM CaCl₂, 0.005% Tween 20, 2 mg ml⁻¹ dextran, and 1 mg ml⁻¹ salmon sperm DNA (Sigma-Aldrich). Either PepT1ECD or PepT2ECD were immobilized on a CM5 chip (GE Healthcare) by amine coupling (GE Healthcare kit) to a total of 1,000 response units. A concentration series of ligand (1000, 500, 250, 125, 62.5, 31.2, 15.6, and 7.81 μM) was injected over the ECD-coated chip for 45 s at 90 μl min⁻¹, followed by a 30-s dissociation time. The chip surface was then regenerated with 2 M NaCl for 30 s. Specific binding of trypsin was obtained by subtracting the response from a blank surface from that of the ECD-coated surface. The kinetic sensorgrams were fitted to a global 1:1 interaction model, allowing determination of the dissociation constant, K_D, using BIAevaluation software 1.0 (GE Healthcare).

Microscale Thermophoresis

MST was carried out using a Monolith NT.115 instrument (NanoTemper). Experiments were performed at 22°C in 25 mM MES (pH 6.5) and 100 mM NaCl. Both PepT1ECD and PepT2ECD were mutated to change a surface-exposed serine to a cysteine (PepT1ECD-S437C and PepT2ECD-S427C) and labeled with the blue maleimide labeling kit MO-L006 (NanoTemper). A range of concentrations of the required ligand (range 0.03–1,000 μM) was incubated with 1.5 μM of purified labeled protein. The sample was loaded into the Nano-Temper glass capillaries and microthermophoresis was carried out using 100% LED power and 80% MST. K_Ds were calculated using the mass action equation via the NanoTemper software from duplicate reads of triplicate experiments.

ACCESSION NUMBERS

The atomic coordinates have been deposited in the PDB under accession codes PDB: 5A9D (*Mm*PepT1ECD) and PDB: 5A9H and 5A9I (*Rn*PepT2ECD).

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.07.016.

AUTHOR CONTRIBUTIONS

S.N. conceived the study. J.H.B., J.L.P., L.E.B., R.J.O., and S.N. cloned and purified the protein. J.H.B. and S.N. crystallized the protein and collected the X-ray diffraction data, solved the structures, and built and refined the models. A.L.B. and S.J.T. designed, performed, and analyzed the TEVC experiments. A.S. and D.M. designed, performed, and analyzed the radioactive uptake assays. J.H.B. and D.S. designed, performed, and analyzed the AUC experiments. J.H.B. and J.L.P. designed, performed, and analyzed the biochemical and binding assays. F.S., M.S.P.S., and P.W.F. built the full-length PepT1 and PepT2 homology models and designed, performed, and analyzed the molecular dynamics experiments. J.L.P. and S.N. wrote the paper.

ACKNOWLEDGMENTS

CellPress

REFERENCES

Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D Biol. Crystallogr. 52, 30–42.

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

Adibi, S.A. (1997a). The oligopeptide transporter (Pept-1) in human intestine: biology and function. Gastroenterology 113, 332–340.

Adibi, S.A. (1997b). Renal assimilation of oligopeptides: physiological mechanisms and metabolic importance. Am. J. Physiol. 272, E723–E736.

Ashida, K., Katsura, T., Motohashi, H., Saito, H., and Inui, K. (2002). Thyroid hormone regulates the activity and expression of the peptide transporter PEPT1 in Caco-2 cells. Am. J. Physiol. Gastrointest. Liver Physiol. 282, G617–G623.

Barabote, R.D., Tamang, D.G., Abeywardena, S.N., Fallah, N.S., Fu, J.Y., Lio, J.K., Mirhosseini, P., Pezeshk, R., Podell, S., Salampessy, M.L., et al. (2006). Extra domains in secondary transport carriers and channel proteins. Biochim. Biophys. Acta 1758, 1557–1579.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. 35, e45.

Birmanns, S., Rusu, M., and Wriggers, W. (2011). Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. J. Struct. Biol. 173, 428–435.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr. D Biol. Crystallogr. 60, 2210–2221.

Brandsch, M. (2013). Drug transport via the intestinal peptide transporter PepT1. Curr. Opin. Pharmacol. 13, 881–887.

Bretschneider, B., Brandsch, M., and Neubert, R. (1999). Intestinal transport of beta-lactam antibiotics: analysis of the affinity at the H+/peptide symporter (PEPT1), the uptake into Caco-2 cell monolayers and the transepithelial flux. Pharm. Res. 16, 55–61.

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

Daniel, H., and Rubio-Aliaga, I. (2003). An update on renal peptide transporters. Am. J. Physiol. Ren. Physiol. 284, F885–F892.

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.

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

Do, C.B., Mahabhashyam, M.S., Brudno, M., and Batzoglou, S. (2005). ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. 15, 330–340.

Fallingborg, J. (1999). Intraluminal pH of the human gastrointestinal tract. Dan. Med. Bull. 46, 183–196.

Fei, Y.J., Kanai, Y., Nussberger, S., Ganapathy, V., Leibach, F.H., Romero, M.F., Singh, S.K., Boron, W.F., and Hediger, M.A. (1994). Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature 368, 563–566.

Fordtran, J.S., Rector, F.C., Jr., and Carter, N.W. (1968). The mechanisms of sodium absorption in the human small intestine. J. Clin. Invest. 47, 884–900.

Fowler, P.W., Orwick-Rydmark, M., Radestock, S., Solcan, N., Dijkman, P.M., Lyons, J.A., Kwok, J., Caffrey, M., Watts, A., Forrest, L.R., et al. (2015). Gating topology of the proton-coupled oligopeptide symporters. Structure 23, 290–301.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 42, 342–346.

Goldberg, D.M., Campbell, R., and Roy, A.D. (1968). Binding of trypsin and chymotrypsin by human intestinal mucosa. Biochim. Biophys. Acta 167, 613–615.

Goldberg, D.M., Campbell, R., and Roy, A.D. (1969a). Fate of trypsin and chymotrypsin in the human small intestine. Gut 10, 477–483.

Goldberg, D.M., Campbell, R., and Roy, A.D. (1969b). Studies on the binding of trypsin and chymotrypsin by human intestinal mucosa. Scand. J. Gastroenterol. 4, 217–226.

Guettou, F., Quistgaard, E.M., Raba, M., Moberg, P., Low, 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.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. 66, 125–132.

Kalliokoski, A., and Niemi, M. (2009). Impact of OATP transporters on pharmacokinetics. Br. J. Pharmacol. 158, 693–705.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. 36, 1277–1282.

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. Eur. J. Pharm. Biopharm. 59, 17–24.

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

Madej, M.G., and Kaback, H.R. (2013). Evolutionary mix-and-match with MFS transporters II. Proc. Natl. Acad. Sci. USA 110, E4831–E4838.

Madej, M.G., Dang, S., Yan, N., and Kaback, H.R. (2013). Evolutionary mix-and-match with MFS transporters. Proc. Natl. Acad. Sci. USA 110, 5870–5874.

Markovic, S., and Dutzler, R. (2007). The structure of the cytoplasmic domain of the chloride channel ClC-Ka reveals a conserved interaction interface. Structure 15, 715–725.

Matthews, D.M. (1975). Intestinal absorption of peptides. Physiol. Rev. 55, 537–608.

Matthews, D.M. (1991). Protein absorption: development and present state of the subject (Wiley-Liss).

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

Meyer, S., Savaresi, S., Forster, I.C., and Dutzler, R. (2007). Nucleotide recognition by the cytoplasmic domain of the human chloride transporter ClC-5. Nat. Struct. Mol. Biol. 14, 60–67.

Newstead, S. (2015). Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. Biochim. Biophys. Acta 1850, 488–499.

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

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

Petoukhov, M.V., Konarev, P.V., Kikhney, A.G., and Svergun, D.I. (2007). ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. J. Appl. Crystallogr. *40*, 223–228.

Pieri, M., Gan, C., Bailey, P., and Meredith, D. (2009). The transmembrane tyrosines Y56, Y91 and Y167 play important roles in determining the affinity and transport rate of the rabbit proton-coupled peptide transporter PepT1. Int. J. Biochem. Cell Biol. *41*, 2204–2213.

Pongs, O., and Schwarz, J.R. (2010). Ancillary subunits associated with voltage-dependent K$^+$ channels. Physiol. Rev. *90*, 755–796.

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

Rambo, R.P., and Tainer, J.A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers *95*, 559–571.

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

Sheldrick, G.M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 479–485.

Smith, D.E., Clemencon, B., and Hediger, M.A. (2013). Proton-coupled oligo-peptide transporter family SLC15: physiological, pharmacological and patho-logical implications. Mol. Aspects Med. *34*, 323–336.

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

Steinhardt, H.J., and Adibi, S.A. (1986). Kinetics and characteristics of absorp-tion from an equimolar mixture of 12 glycyl-dipeptides in human jejunum. Gastroenterology *90*, 577–582.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

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

Svergun, D.I. (1999). Restoring low resolution structure of biological macro-molecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886.

Terada, T., and Inui, K.-I. (2012). Chapter 8. Recent advances in structural biology of peptide transporters. In Current Topics in Membranes, *Vol. 70*, M.O. Bevensee, ed. (Elsevier), pp. 257–274.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determi-nation in small-angle scattering. J. Appl. Cryst. *36*, 860–864.

Warmuth, S., Zimmermann, I., and Dutzler, R. (2009). X-ray structure of the C-terminal domain of a prokaryotic cation-chloride cotransporter. Structure *17*, 538–546.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Winter, G., Lobley, C.M.C., and Prince, S.M. (2013). Decision making in xia2. Acta Crystallogr. D Biol. Crystallogr. *69*, 1260–1273.

Yan, N. (2013). Structural advances for the major facilitator superfamily (MFS) transporters. Trends Biochemical Sciences *38*, 151–159.

Zifarelli, G., and Pusch, M. (2009). Intracellular regulation of human ClC-5 by adenine nucleotides. EMBO Rep. *10*, 1111–1116.

Supplemental Information

Crystal Structures of the Extracellular Domain

from PepT1 and PepT2 Provide Novel Insights

into Mammalian Peptide Transport

John H. Beale, Joanne L. Parker, Firdaus Samsudin, Anne L. Barrett, Anish Senan, Louise E. Bird, David Scott, Raymond J. Owens, Mark S.P. Sansom, Stephen J. Tucker, David Meredith, Philip W. Fowler, and Simon Newstead

Supplementary Figures

1

Crystal structures of the extracellular domain from PepT1 and PepT2

provide novel insights into mammalian peptide transport.

```
                      1                          10                  20
                      .                          .                   .
MousePepT1/1-709   .....MGMSK......................SRGCFG..YPLSIFFIVVNEFCER
HsPepT1/1-708      .....MGMSK......................SHSFFG..YPLSIFFIVVNEFCER
RatPepT1/1-710     .....MGMSK......................SRGCFG..YPLSIFFIVVNEFCER
DogPepT1/1-708     .....MGMSK......................SYGCFG..YPLSIFFIVVNEFCER
RabbitPepT1/1-707  .....MGMSK......................SLSCFG..YPLSIFFIVVNEFCER
RatPepT2/1-729     MNPFQKNESKETLFSPVSTEEMLPRPPSPPKKSPPKIFGSSYPVSIAFIVVNEFCER
HsPepT2/1-729      MNPFQKNESKETLFSPVSIEEVPPRPPSPPKKFSPTICGSNYPVSIAFIVVNEFCER
                                                          _____
                                                                  H1

              30         40        50         60        70        80
              .          .         .          .         .         .
MousePepT1/1-709   FSYYGMRALLVLYFRNFLGWDDDLSTAIYHTFVALCYLTPILGALIADSWLGKFKTI
HsPepT1/1-708      FSYYGMRAILILYFTNFISWDDNLSTAIYHTFVALCYLTPILGALIADSWLGKFKTI
RatPepT1/1-710     FSYYGMRALLVLYFRNFLGWDDDLSTAIYHTFVALCYLTPILGALIADSWLGKFKTI
DogPepT1/1-708     FSYYGMRALLILYFRRFIGWDDDLSTAIYHTFVALCYLTPILGALIADSWLGKFKTI
RabbitPepT1/1-707  FSYYGMRALLILYFRNFIGWDDNLSTVIYHTFVALCYLTPILGALIADAWLGKFKTI
RatPepT2/1-729     FSYYGMKAVLTLYFLYFLHWNEDTSTSVYHAFSSLCYFTPILGAAIADSWLGKFKTI
HsPepT2/1-729      FSYYGMKAVLILYFLYFLHWNEDTSTSIYHAFSSLCYFTPILGAAIADSWLGKFKTI
              _____        _____  ___
                    H1                          H2

              90         100       110        120       130       140
              .          .         .          .         .         .
MousePepT1/1-709   VSLSIVYTIGQAVISVSSINDLTDHDHNGSPDSLPVHVALSMVGLALIALGTGGIKP
HsPepT1/1-708      VSLSIVYTIGQAVTSVSSINDLTDHNHDGTPDSLPVHVVLSLIGLALIALGTGGIKP
RatPepT1/1-710     VSLSIVYTIGQAVISVSSINDLTDHDHDGSPNNLPLHVALSMIGLALIALGTGGIKP
DogPepT1/1-708     VSLSIVYTIGQAVTAVSSINDLTDYNKDGTPDNLSVHVALSMIGLALIALGTGGIKP
RabbitPepT1/1-707  VWLSIVYTIGQAVTSLSSVNELTDNNHDGTPDSLPVHVAVCMIGLLLIALGTGGIKP
RatPepT2/1-729     IYLSLVYVLGHVFKSLGAIPILG.........GKMLHTILSLVGLSLIALGTGGIKP
HsPepT2/1-729      IYLSLVYVLGHVIKSLGALPILG.........GQVVHTVLSLIGLSLIALGTGGIKP
              _____                      _____
                    H3                                  H4

              150        160       170        180       190
              .          .         .          .         .
MousePepT1/1-709   CVSAFGGDQFEEGQEKQRNRFFSIFYLAINGGSLLSTIITPILRVQQCGIHSQQACY
HsPepT1/1-708      CVSAFGGDQFEEGQEKQRNRFFSIFYLAINAGSLLSTIITPMLRVQQCGIHSKQACY
RatPepT1/1-710     CVSAFGGDQFEEGQEKQRNRFFSIFYLAINAGSLLSTIITPILRVQQCGIHSQQACY
DogPepT1/1-708     CVSAFGGDQFEEGQEKQRNRFFSIFYLAINAGSLISTIVTPMLRVHECGIYSQKACY
RabbitPepT1/1-707  CVSAFGGDQFEEGQEKQRNRFFSIFYLAINAGSLLSTIITPMVRVQQCGIHVKQACY
RatPepT2/1-729     CVAAFGGDQFEEEHAEARTRYFSVFYLAINAGSLISTFITPMLRGDVKCFG..QDCY
HsPepT2/1-729      CVAAFGGDQFEEKHAEERTRYFSVFYLSINAGSLISTFITPMLRGDVQCFG..EDCY
              _____       _____
                                          H5

              200        210       220        230       240       250
              .          .         .          .         .         .
MousePepT1/1-709   PLAFGVPAALMAVALIVFVLGSGMYKKFQPQGNIMGKVAKCIGFAIKNRFRHRSKAY
HsPepT1/1-708      PLAFGVPAALMAVALIVFVLGSGMYKKFKPQGNIMGKVAKCIGFAIKNRFRHRSKAF
RatPepT1/1-710     PLAFGVPAALMAVALIVFVLGSGMYKKFQPQGNIMGKVAKCIRFAIKNRFRHRSKAF
DogPepT1/1-708     PLAFGVPAALMAVSLIVFVIGSGMYKKFQPQGNVMGKVVKCIGFALKNRFRHRSKQF
RabbitPepT1/1-707  PLAFGIPAILMAVSLIVFIIGSGMYKKFKPQGNILSKVVKCICFAIKNRFRHRSKQF
RatPepT2/1-729     ALAFGVPGLLMVLALVVFAMGSKMYRKPPPEGNIVAQVIKCIWFALCNRFRNRSGDL
HsPepT2/1-729      ALAFGVPGLLMVIALVVFAMGSKIYNKPPPEGNIVAQVFKCIWFAISNRFKNRSGDI
              _____
                        H6

              260        270       280        290       300       310
              .          .         .          .         .         .
MousePepT1/1-709   PKREHWLDWAKEKYDERLISQIKMVTKVMFLYIPLPMFWALFDQQGSRWTLQATTMN
HsPepT1/1-708      PKREHWLDWAKEKYDERLISQIKMVTRVMFLYIPLPMFWALFDQQGSRWTLQATTMS
RatPepT1/1-710     PKRNHWLDWAKEKYDERLISQIKIMTKVMFLYIPLPMFWALFDQQGSRWTLQATTMT
DogPepT1/1-708     PKREHWLDWAKEKYDERLISQIKMVTKVMFLYIPLPMFWALFDQQGSRWTLQATAMS
RabbitPepT1/1-707  PKRAHWLDWAKEKYDERLIAQIKMVTRVLFLYIPLPMFWALFDQQGSRWTLQATTMS
RatPepT2/1-729     PKRQHWLDWAAEKYPKHLIADVKALTRVLFLYIPLPMFWALLDQQGSRWTLQANKMN
HsPepT2/1-729      PKRQHWLDWAAEKYPKQLIMDVKALTRVLFLYIPLPMFWALLDQQGSRWTLQAIRMN
              _____
                                      H7

              320        330       340        350       360
              .          .         .          .         .
MousePepT1/1-709   GKIGAIEIQPDQMQTVNAILIVIMVPIVDAVVYPLIAKCGFNFTSLKKMTVGMFLAS
HsPepT1/1-708      GKIGALEIQPDQMQTVNAILIVIMVPIFDAVLYPLIAKCGFNFTSLKKMAVGMVLAS
RatPepT1/1-710     GKIGTIEIQPDQMQTVNAILIVIMVPIVDAVVYPLIAKCGFNFTSLKKMTVGMFLAS
DogPepT1/1-708     GKIGLLEVQPDQMQTVNAILIVVMVPIMDAVVYPLIAKCGFNFTSLKKMTVGMFLAS
RabbitPepT1/1-707  GRIGILEIQPDQMQTVNTILIIILVPIMDAVVYPLIAKCGLNFTSLKKMTIGMFLAS
RatPepT2/1-729     GDLGFFVLQPDQMQVLNPFLVLIFIPLFDLVIYRLISKCRINFSSLRKMAVGMILAC
HsPepT2/1-729      RNLGFFVLQPDQMQVLNPLLVLIFIPLFDFVIYRLVSKCGINFSSLRKMAVGMILAC
              _____      _____
                          H8                              H9
```

2

Sequence alignment (residues 370–700) for MousePepT1/1-709, HsPepT1/1-708, RatPepT1/1-710, DogPepT1/1-708, RabbitPepT1/1-707, RatPepT2/1-729, HsPepT2/1-729.

```
                  370        380        390        400        410        420
MousePepT1/1-709  MAFVVAAIVQVEIDKTLPVFPGGNQVQIKVLNIGNNNMTVHFPGN......SVTLAQ
HsPepT1/1-708     MAFVVAAIVQVEIDKTLPVFPKGNEVQIKVLNIGNNTMNISLPGE......MVTLGP
RatPepT1/1-710    MAFVVAAIVQVEIDKTLPVFPSGNQVQIKVLNIGNNDMAVYFPGK......NVTVAQ
DogPepT1/1-708    MAFVMAAIVQLEIDKTLPVFPKQNEVQIKVLNIGNGAMNVSFPGA......VVTVSQ
RabbitPepT1/1-707 MAFVAAAILQVEIDKTLPVFPKANEVQIKVLNVGSENMIISLPGQ......TVTLNQ
RatPepT2/1-729    LAFAVAALVETKINGMIHPQPASQEIFLQVLNLADGDVKVTVLGSRNNSLLVESVSS
HsPepT2/1-729     LAFAVAAVEIKINEMAPAQPGPQEVFLQVLNLADDEVKVTVVGNENNSLLIESIKS
                         H9                    β1        β2        β3

                  430        440        450        460        470
MousePepT1/1-709  MSQTDTFMTFDIDKLTS.INISSSGSPGVTTVAHDFEQGHRHTLLVWNPSQ..YRVV
HsPepT1/1-708     MSQTNAFMTFDVNKLTR.INISSPGSP.VTAVTDDFKQGQRHTLLVWAPNH..YQVV
RatPepT1/1-710    MSQTDTFMTFDVDQLTS.INVSSPGSPGVTTVAHEFEPGHRHTLLVWGPNL..YRVV
DogPepT1/1-708    MSQSDGFMTFDVDKLTS.INISSTGSP.VIPVTYNFEQGHRHTLLVWAPNN..YRVV
RabbitPepT1/1-707 MSQTNEFMTFNEDTLTS.INITS.GSQ.VTMITPSLEAGQRHTLLVWAPNN..YRVV
RatPepT2/1-729    FQNTTHYSKLHLEAKSQDLHFHLKYNSLSVHNDHSVEEKNCYQLLIHQDGESISSML
HsPepT2/1-729     FQKTPHYSKLHLKTKSQDFHFHLKYHNLSLYTEHSVQEKNWYSLVIREDGNSISSMM
                    β4          β5        β6        β7        β8

                  480        490        500        510        520        530
MousePepT1/1-709  KDGLNQKPEKGENGIRFVNTLNEMVTIKMSGKVYENVTS.HNASGYQFFPSGEKQYT
HsPepT1/1-708     KDGLNQKPEKGENGIRFVNTFNELITITMSGKVYANISS.YNASTYQFFPSGIKGFT
RatPepT1/1-710    KDGLNQKPEKGENGIRFVSTLNEMITIKMSGKVYENVTS.HSASNYQFFPSGQKDYT
DogPepT1/1-708    KDGLNQKPEKGENGIRFINSLNESLNITMGDKVYVNVTS.HNASEYQFFSLGTKNIT
RabbitPepT1/1-707 NDGLTQKSDKGENGIRFVNTYSQPINVTMSGKVYEHIAS.YNASEYQFFTSGVKGFT
RatPepT2/1-729    VKDTGIKPANGMAAIRFINTLHKDLNISLDTDAPLSVGKDYGVSAYRTVLRGKYPAV
HsPepT2/1-729     VKDTESRTTNGMTTVRFVNTLHKDVNISLSTDTSLNVGEDYGVSAYRTVQRGEYPAV
                         β9        β10       β11       β12       β13

                  540        550        560        570        580
MousePepT1/1-709  INT.TAVAPTCLTDFKSSNLDFGSAYTYVIR.RASDGCLEVKEFEDIPPNTVNMALQ
HsPepT1/1-708     ISS.TEIPPQCQPNFNTFYLEFGSAYTYIVQ.RKNDSCPEVKVFEDISANTVNMALQ
RatPepT1/1-710    INT.TEIAPNCSSDFKSSNLDFGSAYTYVIRSRASDGCLEVKEFEDIPPNTVNMALQ
DogPepT1/1-708    ISSTQQISQNCTKVLQSSNLEFGSAYTYVIG.TQSTGCPELHMFEDISPNTVNMALQ
RabbitPepT1/1-707 VSS.AGISEQCRRDFESPYLEFGSAYTYLIT.SQATGCPQVTEFEDIPPNTMNMAWQ
RatPepT2/1-729    HCE....TEDKVFSLDLGQLDFGTTYLFVIT.NITSQGLQAWKAEDIPVNKLSIAWQ
HsPepT2/1-729     HCR....TEDKNFSLNLGLLDFGAAYLFVIT.NNTNQGLQAWKIEDIPANKMSIAWQ
                           β14         β15       β16        H10

                  590        600        610        620        630        640
MousePepT1/1-709  IPQYFLLTCGEVVFSVTGLEFSYSQAPSNMKSVLQAGWLLTVAVGNIIVLIVAGAGH
HsPepT1/1-708     IPQYFLLTCGEVVFSVTGLEFSYSQAPSNMKSVLQAGWLLTVAVGNIIVLIVAGAGQ
RatPepT1/1-710    IPQYFLLTCGEVVFSVTGLEFSYSQAPSNMKSVLQAGWLLTVAIGNIIVLIVAEAGH
DogPepT1/1-708    IPQYFLITCGEVVFSVTGLEFSYSQAPSNMKSVLQAGWLLTVAVGNIIVLIVAGAGQ
RabbitPepT1/1-707 IPQYFLITSGEVVFSITGLEFSYSQAPSNMKSVLQAGWLLTVAVGNIIVLIVAGAGQ
RatPepT2/1-729    LPQYVLVTAAEVMFSVTGLEFSYSQAPSSMKSVLQAAWLLTVAVGNIIVLVVAQFSG
HsPepT2/1-729     LPQYALVTAGEVMFSVTGLEFSYSQAPSSMKSVLQAAWLLTIAVGNIIVLVVAQFSG
                         H10                  H11

                  650        660        670        680        690
MousePepT1/1-709  FPKQWAEYILFASLLLVVCVIFAIMARFYTYINPAEIEAQFDEDEKKKGIGKENPYS
HsPepT1/1-708     FSKQWAEYILFAALLLVVCVIFAIMARFYTYINPAEIEAQFDEDEKKNRLEKSNPYF
RatPepT1/1-710    FDKQWAEYVLFASLLLVVCIIFAIMARFYTYINPAEIEAQFDEDEKKKGVGKENPYS
DogPepT1/1-708    FSEQWAEYILFAALLLVVCVIFAIMARFYTYVNPAEIEAQFDDDEKKN.LEKMNVYS
RabbitPepT1/1-707 INKQWAEYILFAALLLVVCVIFAIMARFYTYVNPAEIEAQFEEDEKKKNPEKNDLYP
RatPepT2/1-729    L.AQWAEFVLFSCLLLVVCLIFSVMAYYYVPLKSEDTREATD....KQIPAVQGNMI
HsPepT2/1-729     L.VQWAEFILFSCLLLVICLIFSIMGYYYVPVKTEDMRGPAD....KHIPHIQGNMI
                             H12

                  700
MousePepT1/1-709  SLEPVSQTNM
HsPepT1/1-708     MSGANSQKQM
RatPepT1/1-710    SLEPVSQTNM
DogPepT1/1-708    TVTPVSQTQM
RabbitPepT1/1-707 SLAPVSQTQM
RatPepT2/1-729    NLETKNTRL.
HsPepT2/1-729     KLETKKTKL.
```
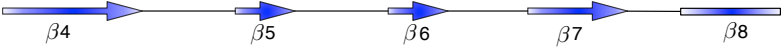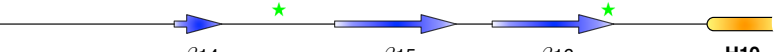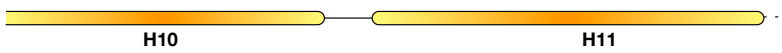
3

Figure S1, related to Figure 1 - Sequence alignment and secondary structure of the mammalian POT family peptide transporters.

An amino-acid sequence alignment between the mammalian peptide transporters; mouse PepT1 (Q9J1P7), human PepT1 (P46059), Rat PepT1 (Q75YE4), Dog PepT1 (Q8WMX5), Rabbit PepT1 (P36836), Rat PepT2 (Q63424) and human PepT2 (Q16348). The sequences were aligned using CLUSTALW as implemented in Jalview (Clamp et al., 2004; Waterhouse et al., 2009) and the figure produced using ESPript (Robert and Gouet, 2014). Identical residues are highlighted in red. The locations of the trans-membrane α-helices were determined following sequence alignments with the bacterial POT family transporter PepTSo (Newstead et al., 2011) and are shown as orange tubes below the alignment. The secondary structure of the extra-cellular domain region, as reported here, is highlighted as arrows. Green stars highlight the two residues identified as interacting with trypsin.
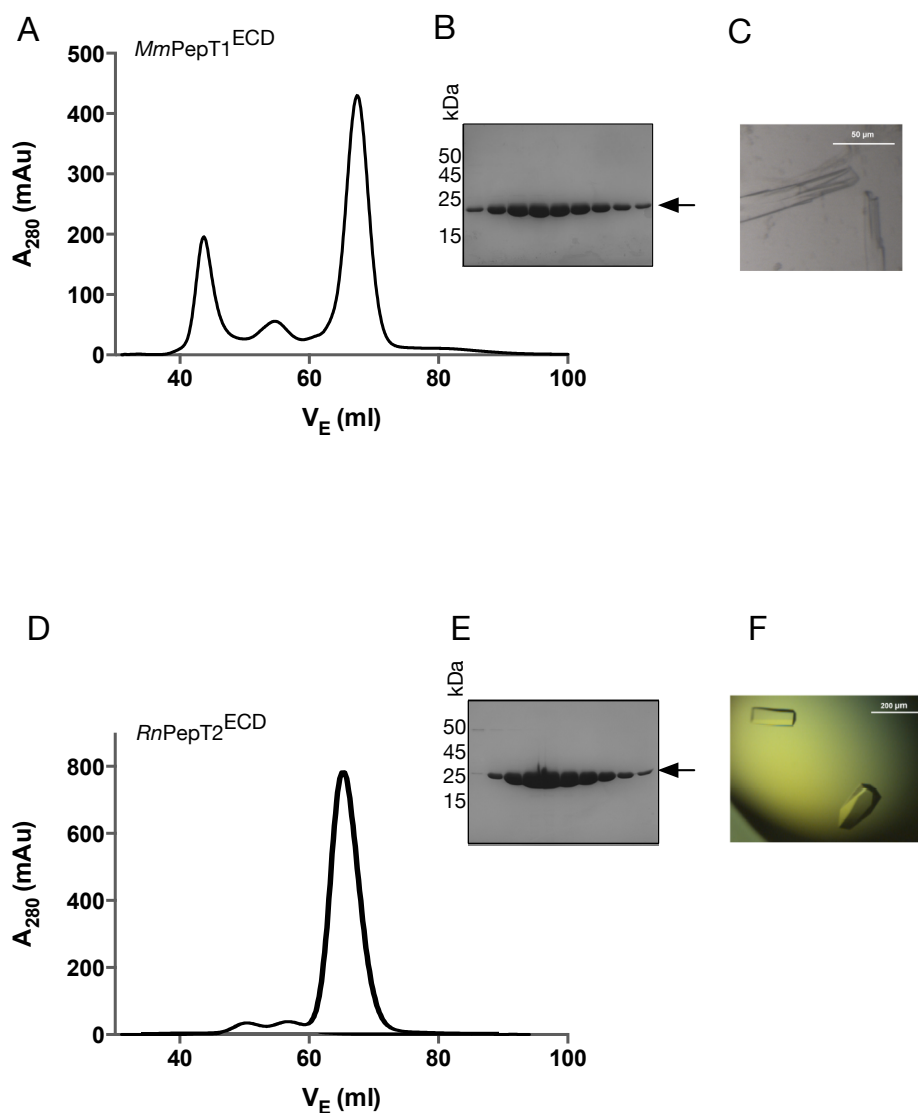
Figure S2, related to Figure 2 - Purification and crystallization of *Mm*PepT1ECD & *Rn*PepT2ECD.

A-C. The final step in the *Mm*PepT1ECD purification showing the UV trace from the S75 16/60 sixe exclusion column. A 15 % Tris-Gly SDS PAGE showing the peak fractions and a representative crystal obstained from the hanging drop vapor diffusion plates. **D-F.** The same data for *Rn*PepT2ECD.
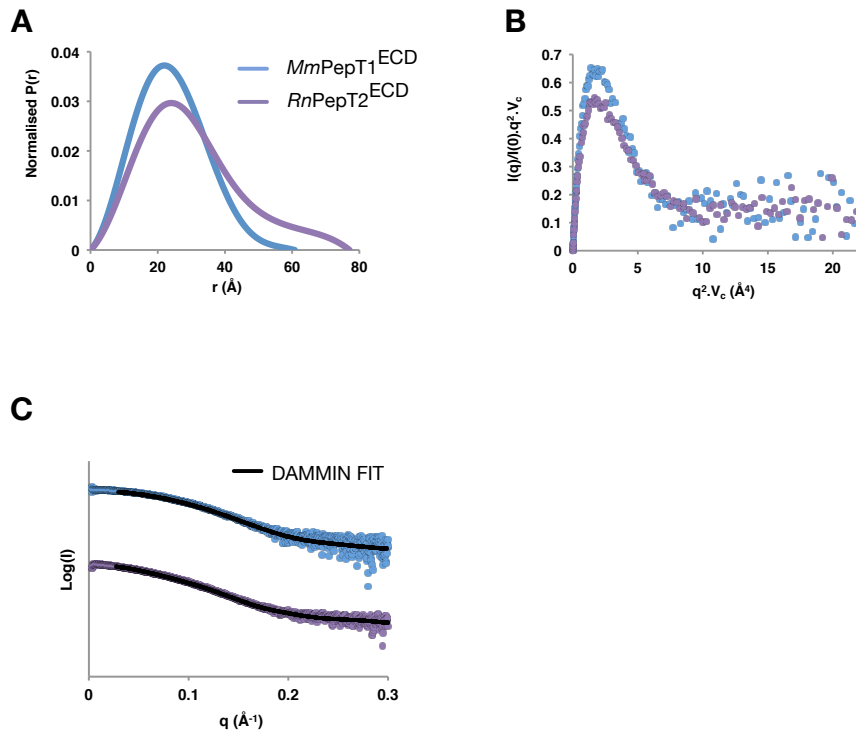
Figure S3, related to Figure 3 – SAXS data analysis of *Mm*PepT1[ECD] and *Rn*PepT2[ECD] (shown in blue and purple respectively).

A. Scattering pair distribution functions of the PepT1[ECD] and PepT2[ECD] calculated using ScÅtter and normalized to the sum of the paired distances. The distributions shown an increase in the PepT2 D_{MAX} and a shift anyway from the spherical shape observed for PepT1[ECD] to a wider, more elongated shape. **B.** The dimensionless V_c based Kratky plot (Rambo and Tainer, 2011, 2013) of the ECDs data curves. The plot clearly shows a reduction in the main peak height for *Rn*PepT2[ECD] compared to *Mm*PepT1[ECD] indicating an increase the surface area to volume ratio and therefore a large particle. **C.** Shows the stacked scattering curves of the ECDs with the plotted DAMMIN fits of the averaged spherical harmonic models shown in Figure 3C. The DAMMIN χ^2 fits for PepT1[ECD] and PepT2[ECD] were 1.09 and 1.27 respectively.
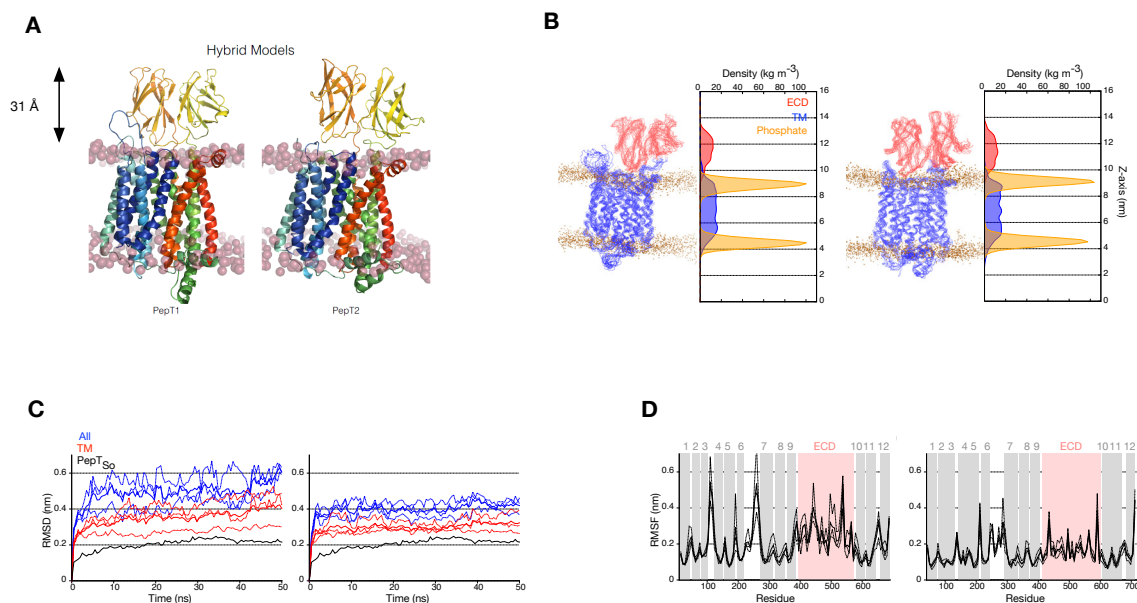
Figure S4, related to Figure 4 - Homology models of human PepT1 and PepT2. **A.** Cartoon representation of the homology models of the human transporters. The trans membrane domains are coloured blue to red, with the extracellular domains shown in yellow. In each subsequent panel PepT1 is on the right and PepT2 the left. **B.** Each hybrid model was inserted into an equilibrated 381-molecule POPC bilayer using GROMACS g_membed protocol. The simulations were run for 50 ns and clearly showed the ECD remaining in the upright position, away from the membrane and the phospholipid head groups**.** **C.** Simulations were analyzed using g_rmsd and g_rmsf tools in GROMACS and showed the models were stable. As a control the crystal structure of PepTSo is shown for reference (black line). **D.** The RMSF analysis shows that the ECD structure is stable and has similar backbone fluctuations to the homology model of the transmembrane domain.
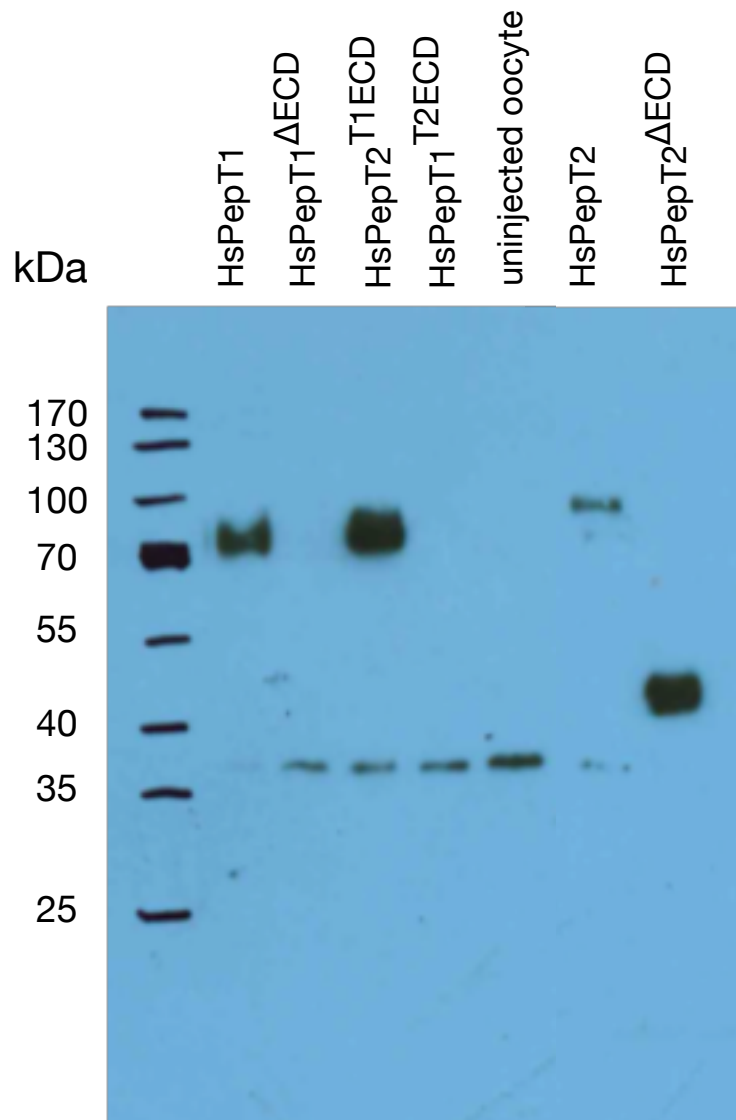
Figure S5, related to Figure 4 - Relative expression levels of *Hs*PepT1 and *Hs*PepT2 constructs in *Xenopus leavis* oocytes. Western blot of a 10% Tris-Gycine SDS-PAGE gel probed using an anti-FLAG antibody. Each lane contains 5 Xenopus eggs after 4 days incubation with either injected mRNA or water control. A cross-reactive band at ~38 kDa was observed in some of the eggs. *Hs*PepT2$^{\Delta ECD}$ runs ~ 10 kDa smaller than the expressed protein, most likely due to known faster migration of transmembrane proteins in SDS-PAGE gels.
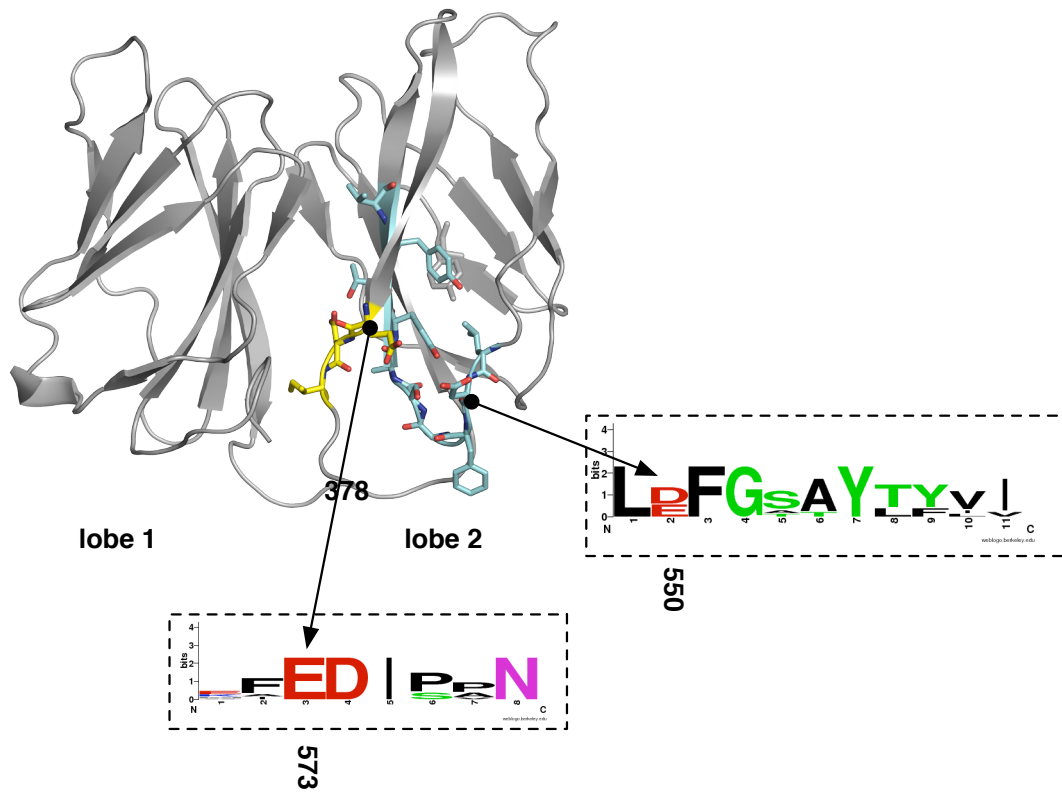
Figure S6, related to Figure 5 – Conserved motifs in the extracellular domains of the mammalian PepT1 and PepT2 proteins. The residues identified as playing an important role in binding to the trypsin protease are located in very conserved regions of the primary structure. Illustrated here the crystal structure of the mouse PepT1^{ECD} as it is thought to exist in solution. Highlighted are the two conserved regions of the sequence, in cyan the region containing the first acidic residue D550 and in yellow the region containing the second acidic residue, E573. Of note is that the second residue in this motif, D574, form part of the salt bridge network identified as playing a role in stabilizing the interface between lobe1 and lobe 2 of the ECD in solution.
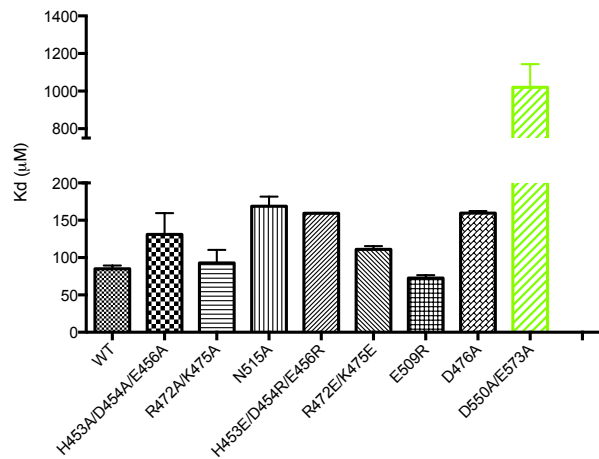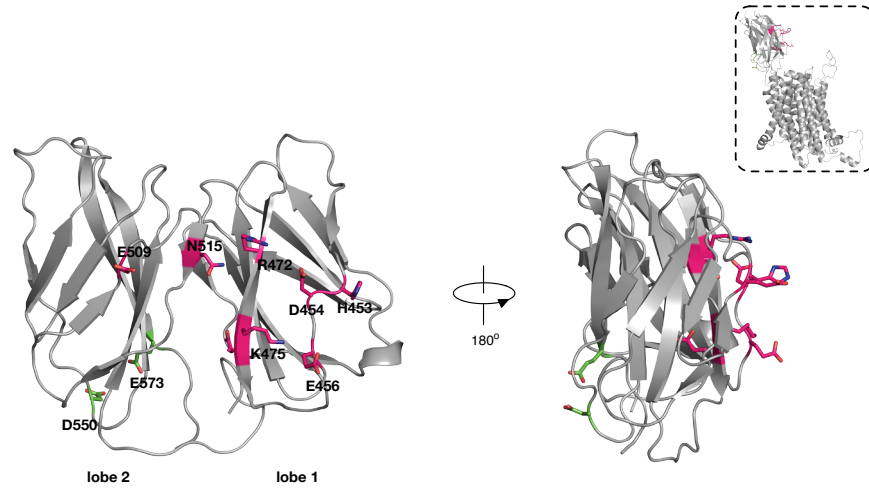
Figure S7, related to Figure 5 – Analysis of binding sites for trypsin on mouse PepT1 ECD. Residues on both the front (magenta) and back (green) face of the mouse PepT1 ECD were analyzed for their interaction with bovine trypsin using SPR. Only residues on the back face of the ECD (D550/E573) resulted in a significant reduction of binding affinity compared to WT. D476A forms part of the salt bridge network stabilizing the interface between lobe1 and lobe 2. Top right - homology model of full length mouse PepT1 showing the location of the residues analyzed with respect to the trans membrane helices and peptide translocation pathway.

Supplementary References:

Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. (2004). The Jalview Java alignment editor. Bioinformatics (Oxford, England) *20*, 426-427.

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

Rambo, R.P., and Tainer, J.A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers *95*, 559-571.

Rambo, R.P., and Tainer, J.A. (2013). Super-resolution in solution X-ray scattering and its applications to structural systems biology. Annu Rev Biophys *42*, 415-441.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320-324.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.