Structure, Volume 23

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**

Crystal structures of the extracellular domain from PepT1 and PepT2 provide novel insights into mammalian peptide transport.

		i																		1 ọ							2	ò			
MousePepT1/1-709		.MGM	SK.															SI	٢G	CF	G.		ΥP	L	SI	F	FI	vν	NE	FC	EF
HsPepT1/1-708		.MGM	SK.															SI	IS	FΕ	G.		ΥP	\mathbf{L}	SI	F	FI	vv	NE	FC	EF
RatPepT1/1-710		.MGM	SK.															SI	٢G	CF	G.		ΥP	\mathbf{L}	SI	F	FI	vv	NE	FC	ΕF
DogPepT1/1-708		.MGM	SK.															S :	ζG	CF	G.		ΥP	\mathbf{L}	SI	F	FI	vv	NE	FC	ΕF
RabbitPepT1/1-707		.MGM	SK.															S]	S	CF	G.		ΥP	\mathbf{L}	SI	F	FI	vv	NE	FC	EF
RatPepT2/1-729	MNPF	QKNE	S K I	ΞTΙ	F	SΡ	VS	ΤE	ΞE	ΜI	LΡ	RE	? P	SI	ΡP	Κŀ	٢S	ΡŦ	ΡK	ΙF	GS	SS	ΥP	V	SI	A	FI.	vv	NE	FC	EF
HsPepT2/1-729	MNPF	QKNE	SK	ΞTΙ	F	SΡ	VS	ΙE	ΞE	VE	? P	RE	? P	SI	? P	Κŀ	٢P	SI	?Т	ΙC	GS	SΝ	ΥP	\mathbf{L}	SI	А	FI.	٧v	NE	FC	EF
																								-(
																													H1		

	зo	4 <u>0</u>	5 <u>0</u>	еò	7 <u>0</u>	вò
MousePepT1/1-709	FSYYGM <mark>R</mark> ALI	. V L Y F R N F L	GWDDDLSTA	IYHTFVALCYL	TPILGALIA	DSWLGKFKTI
HsPepT1/1-708	FSYYGMRAI	LILYFTNFI	SWDDNLSTA	IYHTFVA LCY L	TPILGALIA	DSWLGKFKTI
RatPepT1/1-710	FSYYGMRAL	LV LYF RN F L	GWDDDLSTA	IYHTFVA LCY L	TPILGALIA	DSWLGKFKTI
DogPepT1/1-708	FSYYGMRAL	LI LYF RR F I	GWDDNLSTA	I YH TFVA LCY L	TPILGA <mark>L</mark> IA	DSWLGKFKTI
RabbitPepT1/1-707	FSYYGMRAL	LILYFRNFI	GWDDNLSTV	IYHTFVA LCY L	TPILGALIA	DAWLGKFKTI
RatPepT2/1-729	FSYYGM <mark>K</mark> AV	LTLYFLYFL	HWNEDTSTS	V <mark>YH</mark> AFSS LCY F	TPILGA <mark>A</mark> IA	DSWLGKFKTI
HsPepT2/1-729	FSYYGM <mark>K</mark> AV	LILYFLYFL	HWNEDTSTS	IYHAFSSLCYF	TPILGA <mark>A</mark> IA	DSWLGKFKTI
	H1			H2	2	

MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729

 $\boldsymbol{\mathcal{C}}$

9 Q	100	110	120	130	140
VSLSIVYTI	GQAVISVSSI	NDLTDHDHNGS	PDSLPVHVAI	SMV <mark>GL</mark> A <mark>LIA</mark> I	GTGGIK
VS <mark>LS</mark> IVYTI	GQAVTSVSSI	N D L T D H N H D G T	PDSLPV <mark>H</mark> VVI	SLI <mark>GL</mark> A <mark>LIA</mark> I	GTGGIK
VSLSIVYTI	GQAVISVSSI	NDLTDHDHDGS	PNNLPL <mark>H</mark> VAI	SMI <mark>GL</mark> A <mark>LIA</mark> I	GTGGIK
VSLSIVYTI	GQAVTAVSSI	N D <mark>L</mark> T D Y N K D G T	PDNLSVHVAI	SMI <mark>GL</mark> A <mark>LIA</mark> I	GTGGIK
VWLSIVYTI	GQAVTSLSSV	N E <mark>L</mark> T D N N H D G T	PDSLPVHVAV	/CMI <mark>GL</mark> L <mark>LIA</mark>	GTGGIK
IY <mark>LS</mark> L <mark>VY</mark> VL	GHVFKSLGAI	PI <mark>L</mark> G	GKML <mark>H</mark> TII	SLV <mark>GL</mark> SLIAI	GTGGIK
IY <mark>LS</mark> L <mark>VY</mark> VL	GHVIKSLGAL	PI <mark>L</mark> G	GQVV <mark>H</mark> TVI	SLIGLSLIAL	GTGGIK
НЗ	•			H4	

	150	160	17 <u>0</u>	180	190	
MousePepT1/1-709	CVSAFGGDOFE	FGOEKORNRF	FSIFYLAINGGS	LLSTIITPI	LRVOOCGIHSO	OACY
HsPepT1/1-708	CVSAFGGDÕFE	EGÕEKÕRNRF	FSIFYLAINAGS	LLSTIITPM	LRVÕÕCGIHSŘ	OACY
RatPepT1/1-710	CVSAFGGDOFE	EGÕEKÕ <mark>R</mark> NRF	F S I F Y L A I N A G S	LLSTIITPI:	LRVÕÕCGIHSO	ÕACY
DogPepT1/1-708	CVSAFGGDOFE	EGOEKORNRF	FSIFYLAINAGS	LISTIVTPM	LRVHECGIYSO	KACY
RabbitPepT1/1-707	CVSAFGGDÕFE	EGÕEKÕ <mark>R</mark> NRF	F S I F Y L A I N A G S	LLSTIITPM	V <mark>R</mark> VOOCGIHVŔ	OACY
RatPepT2/1-729	CVAAFGGDOFE	EEHAEARTRY	FSVFYLAINAGS	LISTFITPM	L <mark>R</mark> GDVKCFG	ODCY
HsPep12/1-729	CVAAFGGDÕFE	E K H A E E <mark>R</mark> T <mark>R</mark> Y	F S V F Y L S I N A G S	LISTFITPM	L <mark>R</mark> GDVQCFG	E D C Y
			H5			

2 MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729

2 0	ò	210	220	230	240	250
Ρ	LAFGVPA	A <mark>lm</mark> ava <mark>l</mark> i	I VF VL <mark>GS</mark> GM <mark>Y</mark>	K <mark>K</mark> FQ <mark>PQGN</mark> IMGK	VAKCIGFA	I K <mark>NRF</mark> RH <mark>RS</mark> KAY
Ρ	LAFGVPA	ALMAVAL	I VF VL <mark>GS</mark> GM <mark>Y</mark>	K <mark>K</mark> FK <mark>PQGN</mark> IMGK	VAKCIGFA	I K <mark>NRF</mark> RH <mark>RS</mark> KAF
Ρ	LAFGVPA	A <mark>lm</mark> ava <mark>l</mark> i	I <mark>VF</mark> VL <mark>GS</mark> GM <mark>Y</mark>	K <mark>K</mark> FQ <mark>PQGN</mark> IMGK	VAKCIRFA	I K <mark>NRF</mark> RH <mark>RS</mark> KAF
Ρ	LAFGVPA	ALMAVSL	I VF VI <mark>GS</mark> GM <mark>Y</mark> I	K <mark>K</mark> FQ <mark>PQGN</mark> VMGK	VVKCIGFAI	LK <mark>NRF</mark> RH <mark>RS</mark> KQF
Ρ	LAFGIPA	ILMAVSL	I VFIIGS GM <mark>Y</mark>	K <mark>K</mark> F K P Q <mark>G N</mark> I L S K	VVKCICFA	I K <mark>NRF</mark> RH <mark>RS</mark> KQF
А	LAF GVPG	L <mark>LM</mark> VLA <mark>L</mark>	V VF AM <mark>GS</mark> KM <mark>Y</mark>	R <mark>K</mark> P P P E <mark>G N</mark> I V A Q	VIKCIWFAI	LC <mark>NRF</mark> RN <mark>RS</mark> GDL
А	LAFGVPG	L <mark>LM</mark> VIA <mark>L</mark>	V VF AM <mark>GS</mark> KI <mark>Y</mark>	NKPPPEGNIVAQ	VFKCIWFA	S <mark>NRF</mark> KN <mark>RS</mark> GDI

H7

H6

	260	270	280	290	зоо	310
MousePepT1/1-709	PKREHWLDWA	KEKYDERLI	SQIKMVTKVME	LYIPLPMFW.	ALEDQQGSRW	TLOATTMN
HsPepT1/1-708	PKREHWLDWA	K <mark>EKY</mark> DER <mark>LI</mark>	SQIKMVTRVME	LYIPLPMFW.	ALFDQQGSRW	TLQATTMS
RatPepT1/1-710	PKR <mark>N</mark> HWLDW	K <mark>EKY</mark> DER <mark>LI</mark>	SQI <mark>K</mark> IM <mark>T</mark> KVMB	LYIPLPMFW.	AL <mark>F</mark> DQQGSRW	TLQATTMI
DogPepT1/1-708	PKREHWLDW	K <mark>EKY</mark> DER <mark>LI</mark>	SQI <mark>K</mark> MV <mark>T</mark> K <mark>V</mark> ME	LYIPLPMFW.	AL <mark>F</mark> DQQGSRW	TLQATAMS
RabbitPepT1/1-707	PKRAHWLDWA	K <mark>EKY</mark> DER <mark>LI</mark>	A Q I <mark>K</mark> M V <mark>T</mark> R <mark>V</mark> L B	LYIPLPMFW.	AL <mark>F</mark> DQQGSRW	TLQATTMS
RatPepT2/1-729	P K R Q H W L D W A	AEKYPKHLI	A D V <mark>K</mark> A L T R V L E	LYIPLPMFW.	AL <mark>L</mark> DQQGSRW	TLQANKMN
HsPepT2/1-729	P K R Q H W L D W A	AEKYPKQLI	MDVKALTRVLE	LYIPLPMFW.	AL <mark>LDQQGSRW</mark>	TLQAIRMN



3	70 380	390	400	410	420
MousePepT1/1-709 HsPepT1/1-708	MAFVVAAIVQVI MAFVVAAIVQVI	EIDKTLPVFPGGN EIDKTLPVFPKGN	IQVQIK <mark>VLN</mark> IGI IEVQIK <mark>VLN</mark> IGI	NNNMTVHFPGN NNTMNISLPGE	
RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707		IDKTLPVFPSGN IDKTLPVFPKQN IDKTLPVFPKAN	VQVQIK VLN IGI VEVQIK <mark>VLN</mark> IGI VEVQIK <mark>VLN</mark> IGI	NNDMAVYFPGK NGAMNVSFPGA	VVTVAQ VVTVSQ
RatPepT2/1-729 HsPepT2/1-729	LAFAVAALVETE LAFAVAAAVEIE	(INGMIHPQPASC (INEMAPAQPGPC	2EIFLQ VLN LAI 2EVFLQ VLN LAI	DGDVKVTVLGSI DDEVKVTVLGSI	RNNSLLVESVSS ENNSLLIESIKS
	H9			β2	β3
	110		- <i>c</i> q	~ -	1-
1/1	430	440	450	460	470
MousePepT1/1-709 HsPepT1/1-708 BatPepT1/1-710	MSQTDTFMTFD MSQTNAFMTFD MSQTDTFMTFD	IDKLTS.INISSS /NKLTR.INISSE /DOLTS INVSSE	GSPGVTTVAHI GSP.VTAVTDI	OFEQGHRHTLL OFKQGQRHTLL FEPGHRHTLL	VWNPSQYRVV VWAPNHYQVV VWGPNL. YRVV
DogPepT1/1-708 RabbitPepT1/1-707	MSQSDGFMTFDV MSQTNEFMTFNF	/DKLTS.INISSI EDTLTS.INITS.	GSP.VIPVTY GSQ.VTMITP	NFEQGHRHTLL ^V SLEAGQRHTLL ^V	VWAPNNYRVV VWAPNNYRVV
RatPepT2/1-729 HsPepT2/1-729	FQNTTHYSKLHI FQKTPHYSKLHI	LEAKSQDLHFHLF LKTKSQDFHFHLF	YNSLSVHNDH: YHNLSLYTEH:	SVEEKNCYQLL: SVQEKNWYSLV:	I HQDGESISSML I REDGNSISSMM
	β4	β5	β 6	β7	β8
MousePepT1/1-709	480 KDGLNQKPEK <mark>G</mark> I	490 5 ENGI <mark>RF</mark> VNTLNEN	500 5: 1VTIK <mark>MS</mark> GKVYI	LOUTS.HNASG	20, 530 YQFFPS <mark>G</mark> EKQYT
HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708	KDGLNQKPEKG KDGLNQKPEKG	ENGIRFVNTFNEI ENGIRFVSTLNEN ENGIRFINSLNES	LITITMSGKVY/ 4ITIKMSGKVY/ SUNTTMGDKVY/	ANISS.YNAST Envts.hsasn Vnvts.hnase	YQFFPSGIKGFT YQFFPS <mark>G</mark> QKDYT YOFFSL G TKNTT
RabbitPepT1/1-707 RatPepT2/1-729	NDGLTQKSDKG VKDTGIKPANG	ENGI RF VNTYSQE 4AAI <mark>RF</mark> INTLHKI	PINVTMSGKVY DLNISLDTDAP	EHIAS.YNA <mark>S</mark> E LSVGKDYGV <mark>S</mark> A	YQFFTS <mark>G</mark> VKGFT YRTVLR <mark>G</mark> KYPAV
HsPepT2/1-729	VKDTESRTTNG	ATTV <mark>RF</mark> VNTLHKI	VNISLSTDIS	LNVGEDYGVSA	
		eta9	eta10	eta11	β 12 β 13
	540	5 5 ọ	560	57 <u>0</u>	5 8 Q
MousePepT1/1-709 HsPepT1/1-708	540 INT.TAVAPTCI ISS.TEIPPOC	550 LTDFKSSNLDFGS QPNFNTFYLFFG	560 SAYTYVIR.RAS	570 SDGCLEVKEFE NDSCPEVKVFE	580 DIPPNTVNMALO DISANTVNMALO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707	540 INT.TAVAPTCI ISS.TEIPPOCO INT.TEIAPNCO ISSTQQISQNCO VSS.AGISEQCE	550 JDFKSSNLDFG 2PNFNTFYLEFG SSDFKSSNLDFG KVLQSSNLEFG RRDFESPYLEFG	560 SAYTYVIR.RAS SAYTYVVQ.RKI SAYTYVIG.TQS SAYTYVIG.TQS SAYTYVIG.SQ	570 SDGCLEVKEFE NDSCPEVKVFE SDGCLEVKEFE STGCPELHMFE ATGCPQVTEFE	580 DIPPNTVNMALQ DISANTVNMALQ DIPNTVNMALQ DISPNTVNMALQ DISPNTWNMALQ
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	540 INT.TAVAPTCI ISS.TEIPPQCC INT.TEIAPNCS ISSTQQISQNC VSS.AGISEQCI HCETEDK HCRTEDK	550 LTDFKSSNIDFG 2PMFNTFYLEFG SSDFKSSNIDFG RRDFESPILEFG RRDFESDILGG FSLNLGILPFG	560 SAYTYVIR.RA: SAYTYVIQ.RKI SAYTYVIRSRA: SAYTYVIG.TQ: SAYTYLIT.SQ: TYLFVIT.NI AAYLFVIT.NI	570 SDGCLEVKEFE NDSCPEVKVFE SDGCLEVKEFE STGCPELHMFE ATGCPQVFEE ISQGLQAWKATE INQGLQAWKIE	580 DIPPNTVMALO DISANTVMALO DIPPNTVMALO DIPPNTVMALO DIPNTVMALO DIPNKLSIAWO DIPANKMSIAWO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	540 INT.TAVAPTCI ISS.TEIPPQCC INT.TEIAPNCS ISSTQQISQNC VSS.AGISEQCI HCETEDK HCRTEDK	550 LTDFKSSNIDFG 2PNFNTFYLEFG SSDFKSSNIDFG RRDFESPYLEFG IFSLNLGLIDFG IFSLNLGLIDFG 244	560 SAYTYVIR.RAS SAYTYVIQ.RKI SAYTYVIG.TOS SAYTYVIG.TOS SAYTYVIG.TOS SAYTYLIT.SOS TYLEVIT.NI AAYLEVIT.NI	570 SDGCLEVKEFE NDSCPEVKVFE SDGCLEVKEFE STGCPELHMFE ISQGLQAWKATE INQGLQAWKATE	580 DIPPNTVMMALO DISANTVMMALO DIPPNTVMMALO DIPPNTVMALO DIPPNTKMMALO DIPNTKLSIAWO DIPANKMSIAWO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-708 RabbitPepT2/1-729 HsPepT2/1-729	549 INT.TAVAPTCI ISS.TEIPPQCC INT.TEIAPNCS ISSTQQISQNC VSS.AGISEQC HCETEDKY HCRTEDKY	550 LTDFKSSNLDFG PNFNTFYLEFG SSDFKSSNLDFG KVLQSSNLEFG KVLQSSNLEFG VFSLDLGGLDFG VFSLDLGGLDFG MFSLNLGLL β14	560. SAYTYVIR.RAS SAYTYVQ.RKI SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIT.NI SAYTYLIT.SQ TYLFVIT.NI β15	579 SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE STGCPELHMFE ATGCPQVTEFE ISQGLQAWKAE INQGLQAWKAE	580 DIPENTVIMALO DIPENTVIMALO DIPENTVIMALO DIPENTVIMALO DIPENTMIMANO DIPENTMIMANO DIPANKKSIAWO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	590 IPQYFLLTCGEX	550 LTDFKSSNIDFG PNFNTFYLEFG SSDFKSSNIDFG KVLQSSNIEFG KVLQSNIEFG VFSLDLGQLDFG VFSLNLGILDFG β 14 600 61 VFSVTGLEFSY	560 SAYTYVIR.RAS SAYTYVIR.RAS SAYTYVIRSRAS SAYTYVII.SQU TYLFVIT.NU AYLFVIT.NU β 15 LO 620 SQAPSNMKSVL	57 9 5D GC LE VKE FE ND SCP EVKVFE 5D GC LE VKE FE 5T GC PE LHM FE AT GC PQ VT EFE ISQGLQAWKAE INQGLQAWKAE β16 639 24 GWLLTVAVG	580 DIPPNTVIMALO DISANTVIMALO DISPNTVIMALO DIPPNTVIMALO DIPPNTVIMALO DIPVNKLSIAWO DIPVNKLSIAWO H10 H10 640 NIIVLIVACACH
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708	590 S90 S90 S90 S90 S90 S90 S90 S	550 LTDFKSSNIDFG SDFKSSNIDFG SDFKSSNIDFG SDFKSSNIDFG KOFESFUEFG VFSLNLGIDFG β14 600 61 VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS	560 SAYTYVIR.RAS SAYTYVIR.RAS SAYTYVIRSRAS SAYTYVIG.TCS SAYTYVIS.TCS SAYTYVIS.T	57 9 SDGCLEVKEFE NDSCPEVKVFE SDGCLEVKEFE STGCPELHMFE ISQGLQAWKAE INQGLQAWKAE β16 β16 β30 AGWLLTVAVG DAGWLLTVAVG DAGWLLTVAJG	580 DIPPNTVIMALO DISANTVIMALO DISANTVIMALO DIPPNTVIMALO DIPPNTVIMALO DIPVNKLSIAWO DIPVNKLSIAWO DIPANKMSIAWO H10 640 NIIVLIVAGAGG NIIVLIVAGAGG
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729	590 S90 S90 S90 S90 S90 S90 S90 S	550 $TDFKSSNIDFG$ $PNFNTFYIEFG$ $SSDFKSSNIDFG$ $tVLQSSNIEFG$ $ROFESFYIEFG$ $ROFESFYIEFG$ $\beta 14$ 600 61 $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $FSVTGLEFSYS$ $FSVTGLEFSYS$	560. SAYTYVIR.RAS SAYTYVIR.RAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRSRAS SAYTYVIRS SAYTYVIRS SAYTYVIRS SAYTYVIRS β 15 β 15 β 15 β 15 β 20 SQAPSNMKSVLG	579. SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE STGCPELHMFE ATGCPQVTEFE ISQGLQAWKAE ISQGLQAWKAE ISQGLQAWKIE AGULTVAVG QAGWLLTVAVG QAGWLLTVAVG QAGWLLTVAVG QAGWLLTVAVG QAGWLLTVAVG	580 DIPPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTMIMAWO DIPNKKLSIAWO DIPANKKSIAWO H10 640 NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RabbitPepT2/1-729 HsPepT2/1-729 MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	590 S90 S90 S90 S90 S90 S90 S90 S	550 TDFKSSNIDFG PNFNTFYLEFG SSDFKSSNLDFG KVLQSSNIEFG KVLQSSNIEFG KFSLDLGQLDFG FSLDLGLDFG β 14 600 61 VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS	560 SAYTYVIR.RA: SAYTYVIRSR	579 SDGCLEVKEFE SDGCLEVKEFE STGCPEVKVFEI STGCPEVKVFEI STGCPQVTEFEI ISOGLQAWKAE INOGLQAWKAE INOGLQAWKAE AGWLITVAVG QAGWLITVAVG QAGWLITVAVG QAGWLITVAVG QAGWLITVAVG QAGWLITVAVG QAAWLITVAVG QAAWLITVAVG	580 DIPPNTVIMALO DISANTVIMALO DISANTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTKLSIAWO DIPNKLSIAWO H10 640 NIIVI VAGAGO NIIVI VAGAGO NIIVI VAGAGO NIIVI VAGAGO NIIVI VAGFSG
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-710 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	590 S90 S90 S90 S90 S90 S90 S90 S	550 STDFKSSNIDFG STDFKSSNIDFG SSDFKSSNIDFG SSDFKSSNIDFG SRDFESFUEFG K β14 600 61 VFSUTGLEFSYS	560 SAYTYVIR.RAS SAYTYVIRSRAS SAYTYVIG.RKI SAYTYVIG.TC SAYTYLIT.SQ TYLFVIT.NI AYLFVIT.NI β 15 β15 β15 β15 β15 β20 SQAPSNMKSVL SQAPSNMKSVL SQAPSNMKSVL SQAPSNMKSVL SQAPSNMKSVL SQAPSNMKSVL SQAPSNMKSVL	579 SDGCLEVKEFE SDGCLEVKEFE STGCPEVKVFE STGCPEVKVFE STGCPEVKEFE STGCPE	580 DIPPNTVIMALO DISANTVIMALO DISANTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMA H10 640 NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	590 INT. TAVAPTCI ISS. TEIPPQCC INT. TEIAPNCS ISSTQLISNC VSS. AGISEQCE HCETEDK HCRTEDK HCRTEDK IPQYFLITCGE IPQYFLLTCGE IPQYFLLTCGE IPQYFLITCGE IPQYFL	550 TDFKSSNIDFG SDFKSSNIDFG SDFKSSNIDFG KVLQSSNIEFG RADFESFYLEFG Ø14 600 61 VFSUTGLEFSYS VFSUTG	560. SANTYVIR.RAS SANTYVIR.RAS SANTYVIR.SRAS SANTYVIR.SRAS SANTYVIR.SRAS SANTYVIR.SRAS SANTYVIR.SRAS SANTYVIR.SRAS BANKSVIS SANTYVIT.NNS β15 LO 620 SANTSVIS SANT	579 SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE STGCPEVHMPE ATGCPQVTEFE ISQGLQAWKAE (NQGLQAWKAE (NQGLQAWKAE (NQGLQAWKAE) (NQGLQAWKAE (NQGLQAWKAE) (NQGLQAWLTVAYGC) (NQGLQAWLTVAYGC) (NQGULTVAYGC) (NQGULTVAYGC) (NQGWLTVAYGC) (NQGWLTVAYGC) (NQGWLTVAYGC) (NQGWLTVAYGC) (NQGWLTVAYGC) (NQGULTVAYGC) (NQGULTVAYGC) (NQGWL	580 DIPPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMA H10 640 NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO NIIVIVAGAGO
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT2/1-729	540 INT.TAVAPTCI ISS.TEIPPQCC INT.TEIAPNCS ISSTQQISQNC VSS.AGISEQCE HCETEDKY HCRTEDKY HCRTEDKY IPQYFLLTCGEY IPQYFLLTCGEY IPQYFLLTCGEY IPQYFLITCGEY IPQYFLITCGEY IPQYFLITCGEY IPQYFLITSGEY H 650 FPKQWAEYILF	550 $TDFKSSNIDFG$ $PNFNTFYLEFG$ $SSDFKSSNIDFG$ $KVLQSSNIEFG$ $RDFESPILEFG$ $FSLDLGUDFG$ $FSLDLGUDFG$ $\beta 14$ $600 61$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $MFSVTGLEFSYS$	560 SAYTYVIR.RA: SAYTYVIR.SRA: SAYTYVIR:	579 SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE STGCPELHMFE ATGCPQVTEFE ISQGLQAWKAE INQGLQAWKAE INQGLQAWKEE β16 9 639 0 64 0 639 0 64 0 639 0 64 0 7 0 7 0 7 0 7 0 7 0 7 0 7 0 7	580 DIPPNTVIMALO DISANTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMA H10 640 NIVLTVAGAGO NIVLTVAGAGO NIVLTVAGAGO NIVLIVAGAGO NIVVIVAQFSG NIVVIVAQFSG NIVVIVAQFSG NIVVIVAQFSG NIVVIVAQFSG NIVVIVAQFSG
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-700 DogPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT2/1-729 HsPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708	590 IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLLTCGET IPQYFLITSGE	550 $TDFKSSNIDFG$ $PNFNTFYLEFG$ $SSDFKSSNIDFG$ $KVLQSSNIEFG$ $RDFESFUEFG$ $FSLNLGIDFGF$ $FSLNLGIDFGF$ $\beta14$ $600 61$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $MFSVTGLEFSYS$ $MFSVT$	560 SAYTYVIR.RA: SAYTYVIR.RA: SAYTYVIG.TC: SAYTYVIG.TC: SAYTYVIT.SQ: TYLFVIT.NI AYLFVIT.NI AYLFVIT.NI β 15 10 620 SQAPSNMKSVL SQAPSNM	579 SDGCLEVKEFE SDGCLEVKEFE STGCPEVKVFEI STGCPEVKVFEI STGCPUTEFEI ISOGLQAWKAE INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIVAUG INOGLQAWLIE INOGLQAWLIE INOGLQAWLIE INOGLQAWLIE INOGLQAWLIE INOGLQAWLIE INOGLQAWLIE INOGLQAWE INOGLQA	580 DIPPNTVIMALO DISANTVIMALO DISANTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALSIAWO H10 640 H10 640 H10 640 NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGAGO NIIVLIVAGASG NIIVLIVAGFSG NIIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVLVAOFSG NIVVAOFSG NIVLVAOFSG NIVVAOF
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT2/1-729	590 INT.TAVAPTCI ISS.TEIPPQCC INT.TEIAPNCS VSS.AGISEQCE HCETEDKY HCRTECKY HCRTEC	550 $LTDFKSSNIDFG$ $PNFNTFYLEFG$ $SSDFKSSNIDFG$ $KVLQSSNIEFG$ $RVLQSSNIEFG$ $RVLGSNIEFG$ $RVFSVTGLEFSY$ $JFSLNLGLIFFY$ $\beta 14$ 600 610 $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $VFSVTGLEFSYS$ $MFSVTGLEFSYS$	560 SANTYVIR.RA: SANTYVIR.RA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: SANTYVIR.SRA: BANKSVI: SANTYVIT.NN: BANKSVI: SANTYVIR.SNA:	579 SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE ISOGLQAWKAE	580 DIPPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTMIMAWO OIPNTMIMAWO H10 640 NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIIVAGAGO NIIVIVIVAGESG NIIVIVVAGESG NIIVIVVAGESG NIIVIVVAGESG NIIVIVVAGESG NIIVIVAGACO NIIVIVAGACO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVAGASO NIIVIVASO NIIVIVAGASO NIIVI NAGASO NAGASO NIIVI NAGASO NIIVI NAGASO NIIVI NAGASO NAGASO NIIVI NAGASO NAGASO NAGASO NIIVI NAGASO NIVI NAGASO NAGASO NIVI NAGASO NIVI NAGASO NI NAGASO NIVI NAGASO NI NI NI NI NI NI NI NI NI NI NI NI NI
MousePepT1/1-709 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RabbitPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729 HsPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-708 RatPepT1/1-707 RatPepT2/1-729 HsPepT2/1-729	590 S90 S90 S90 S90 S90 S90 S90 S	550 TDFKSSNIDFG SSPFKSSNIDFG SSPFKSSNIDFG KVLQSSNIEFG KVLQSSNIEFG KSLDLGQLDFG FSLDLGQLDFG FSLDLGUDFG FSLTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS VFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS MFSVTGLEFSYS SLLVVCVTFAT ALLLVCVTFAT ALLLVCTFAT ALLLVCTFAT	560 SAYTYVIR.RA: SAYTYVIRSR	579. SDGCLEVKEFE SDGCLEVKEFE SDGCLEVKEFE SDGCPELHMFE ATGCPQVTEFE ISQGLQAWKAE	580 DIPPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTVIMALO DIPNTMIMANO HIVISTANO HIVITVAGAG NIVITVAGAGQ NIVITVAGAGQ NIVVIVAGAGQ NIVVIVAGAGQ NIVVIVAQFSG NIVVIVA

	700									
MousePepT1/1-709	S	L	E	Ρ	V	S	Q	Т	Ν	М
HsPepT1/1-708	М	S	G	A	Ν	S	Q	K	Q	М
RatPepT1/1-710	S	\mathbf{L}	Е	Ρ	V	S	Q	Т	Ν	М
DogPepT1/1-708	Т	V	Т	Ρ	V	S	Q	Т	Q	М
RabbitPepT1/1-70	7 S	\mathbf{L}	A	Ρ	V	S	Q	Т	Q	М
RatPepT2/1-729	N	\mathbf{L}	Е	Т	K	Ν	Т	R	L	•
HsPepT2/1-729	K	L	Е	Т	K	K	Τ	K	L	

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*PepT1^{ECD} & *Rn*PepT2^{ECD}.

A-C. The final step in the $MmPepT1^{ECD}$ 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 $RnPepT2^{ECD}$.



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 peek 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^{Δ 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.