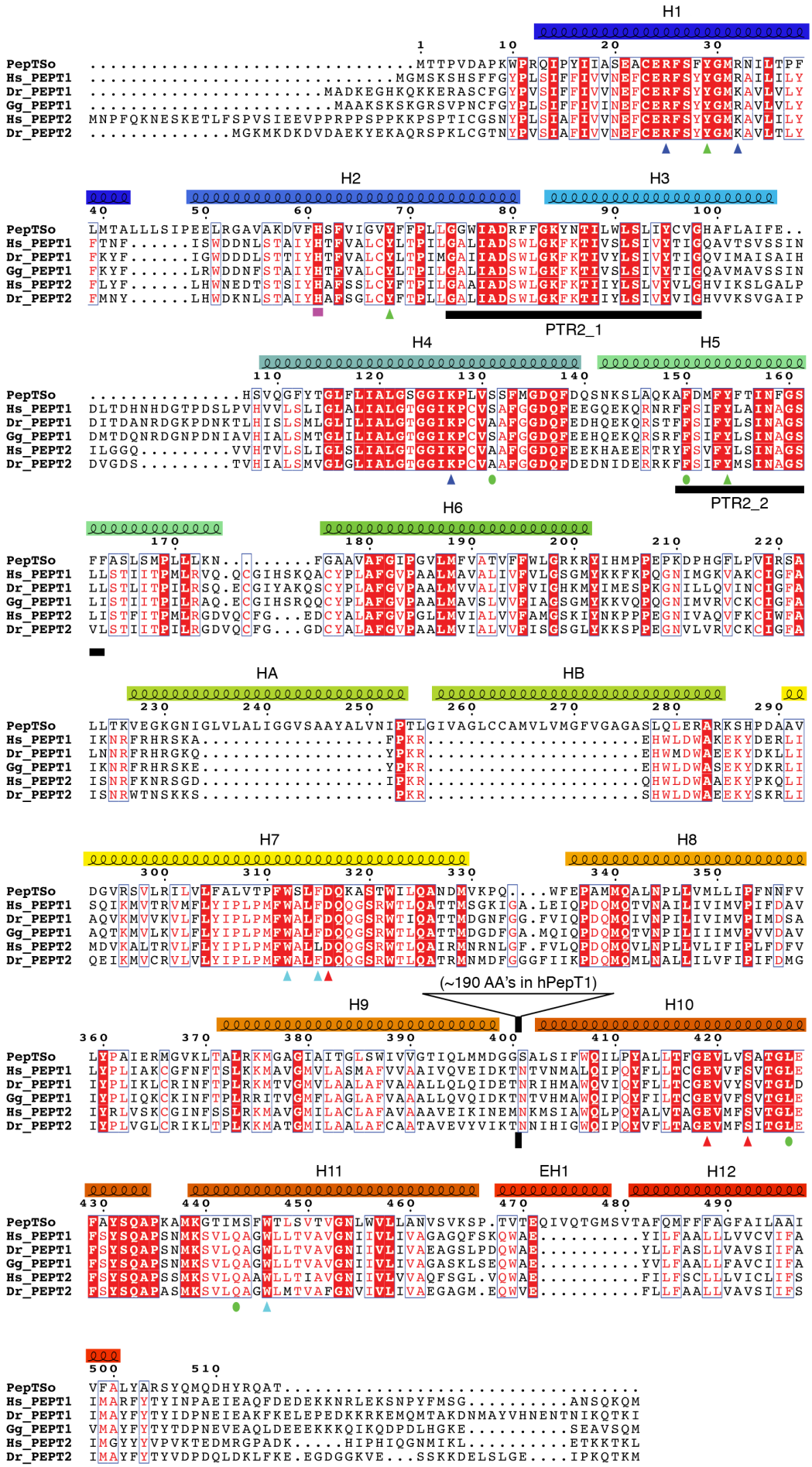
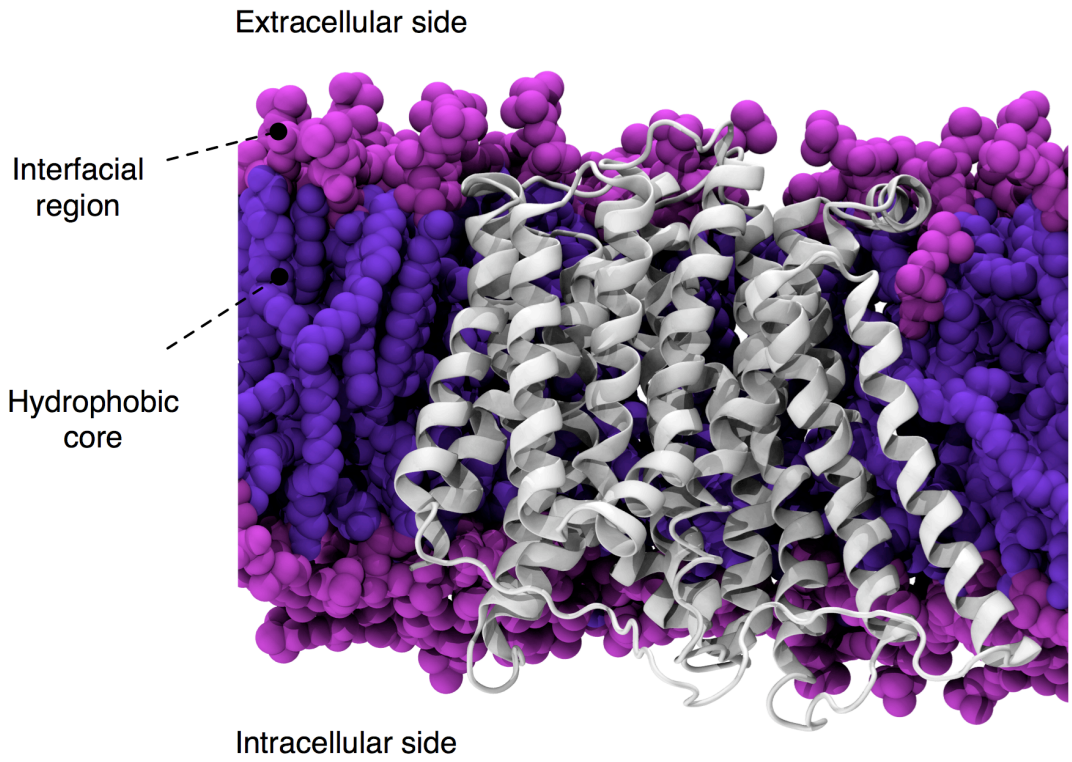
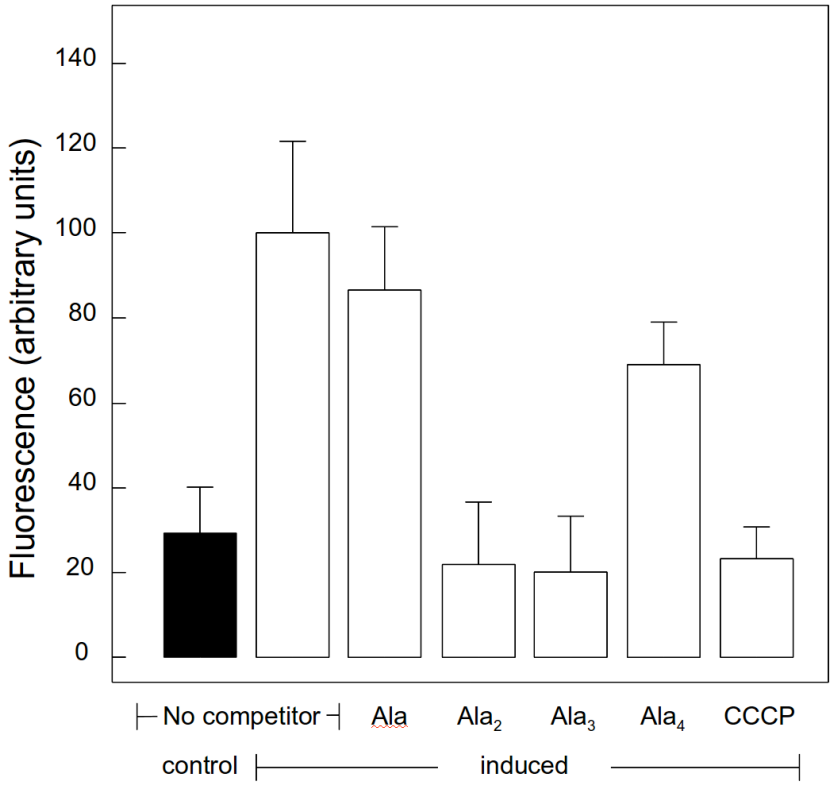
Supplementary Information


H1
H2
H3
H4
H5
H6
HA
HB
H7
H8
H9
H10
H11
EH1
H12
PTR2_1
PTR2_2
(~190 AA's in hPepT1)
PepTSo
Hs_PEPT1
Dr_PEPT1
Gg_PEPT1
Hs_PEPT2
Dr_PEPT2

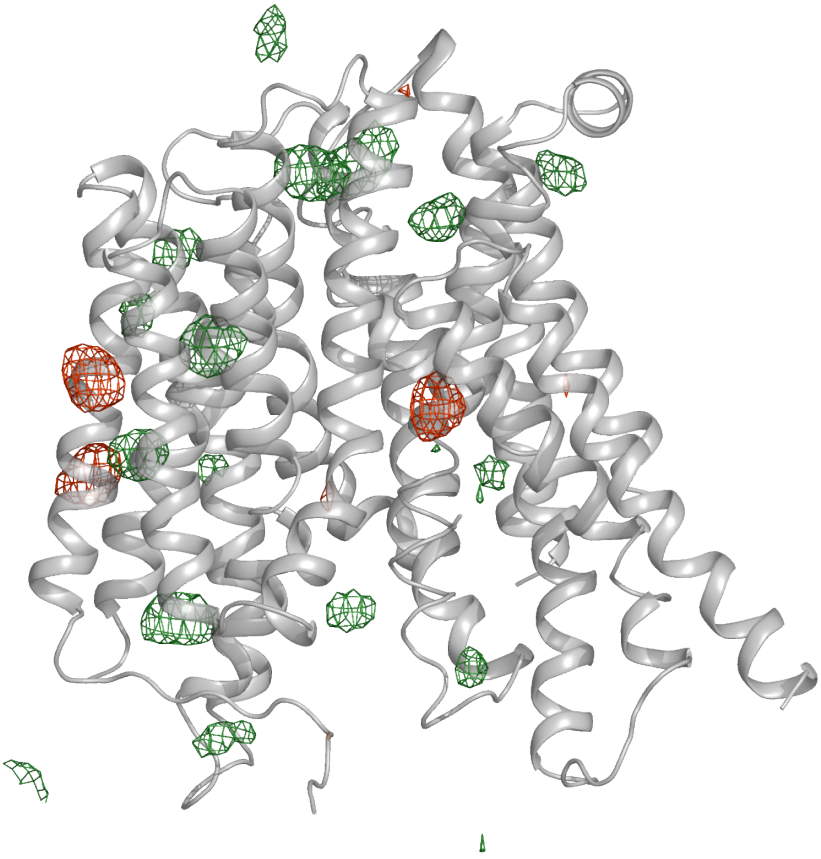
Supplementary Figure 1 Sequence alignment and secondary structure of $PepT_{So}$. Amino acid sequence alignment of *S. oneidensis* $PepT_{So}$ (UniProt: Q8EKT7) with the human PepT1 (B2CQT6) and PepT2 (Q16348), fish PepT1 (Q7SYE4) and PepT2 (Q2F800) and chicken PepT1 (Q90WH8) transporter homologues using ClustalW (*www.ebi.ac.uk/clustalw/)* with manual adjustment in JalView [31]. $PepT_{So}$ is ~30 % identical and ~45 % similar to human PepT1 and PepT2 as calculated using pairwise alignment within the transmembrane regions in JalView. Identical residues are highlighted in red. The α-helices in $PepT_{So}$ are depicted as coils above the sequence, colored blue to red respectively. The conserved PTR2 signature motifs are marked with black horizontal bars. Residues lining the central cavity are highlighted by triangles shown below the sequence and coloured according to type, Arg and Lys (blue), Glu and Ser (red), Tyr (green), Trp, Phe and Ile (cyan). Residues forming the intracellular gate are denoted by circles (green). Residue His61 is highlighted by a square (magenta). The vertical black bar and open triangle at the end of helix H9 represents the location of the additional extracellular domain present in the mammalian homologues but absent in prokaryotic peptide transporters.



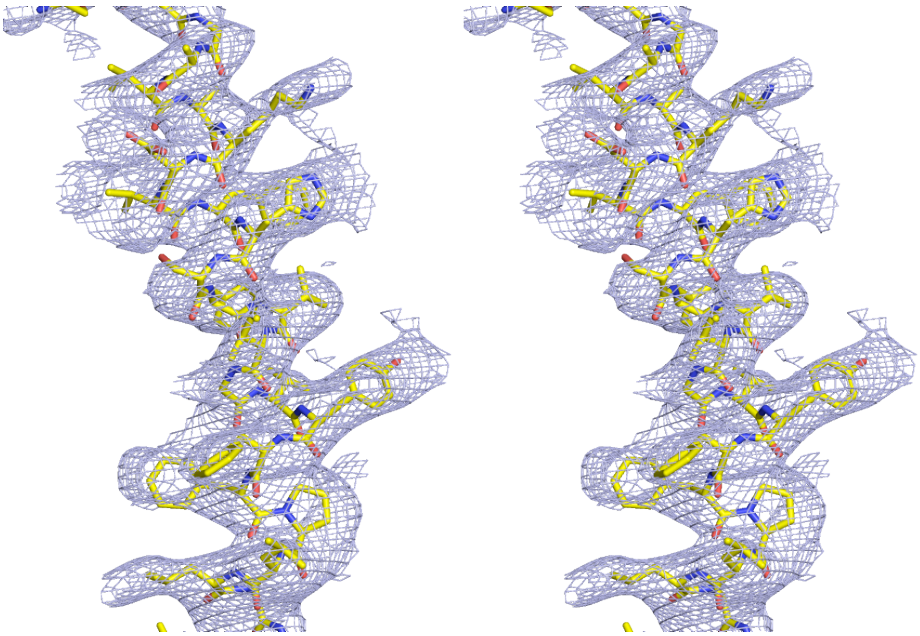
Supplementary Figure 2 Image taken of $PepT_{So}$ from one of the coarse grained bilayer self-assembly simulations. For clarity both the protein and lipids have been converted back to an atomistic representation (Supplementary Materials and Methods).



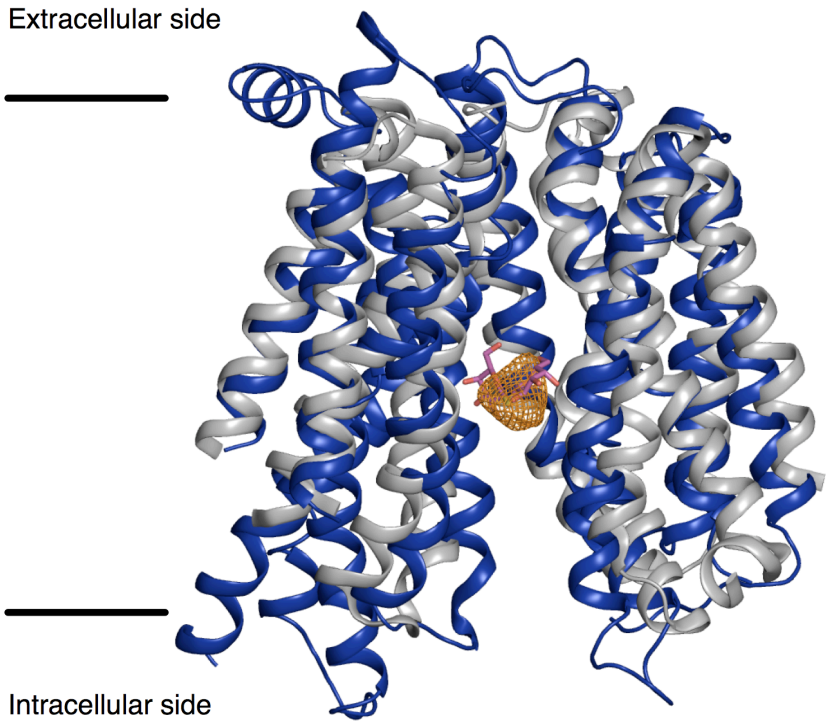
Supplementary Figure 3 Competition assay for the uptake of β-Ala-Lys-AMCA against a set of different length L-alanine peptides. Uptake of the fluorescent peptide conjugate β-Ala-Lys-AMCA (50 μM) was measured in *E. coli* cells harbouring the gene encoding $PepT_{So}$ before (control; filled bar) and after (open bars) induction of expression with isopropyl-β-D-1-thiogalactopyranoside. Uptake was measured in the absence or presence of the amino acid L-Alanine (Ala), the corresponding di- (Ala_2), tri- (Ala_3) and tetra- (Ala_4) peptides, each at a concentration of 10 mM, and of carbonyl cyanide *p*-chlorophenylhydrazone (CCCP), at a concentration of 10 μM. Results shown are means ± SD (n = 6).



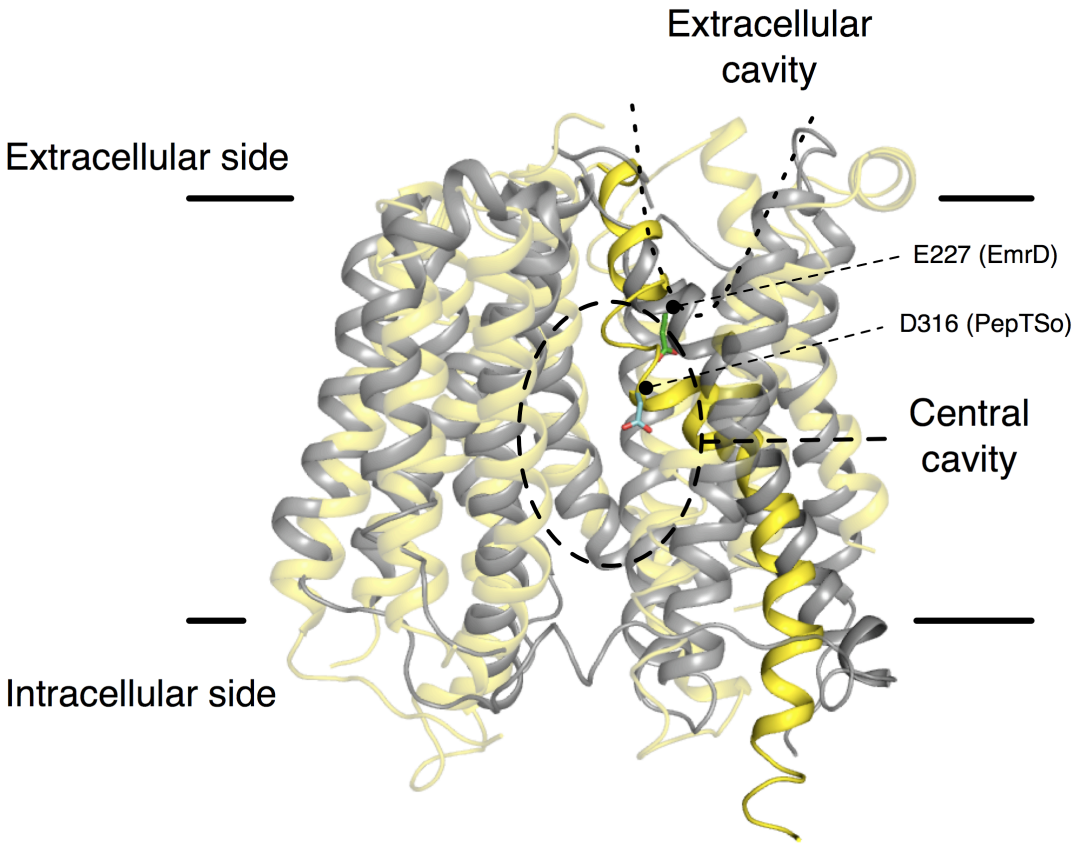
Supplementary Figure 4 Anomalous difference peaks for the Hg and Se atoms overlaid on the structure of $PepT_{So}$. Anomalous difference peaks for the bound Hg atoms calculated using MIR(AS), see supplementary text, are shown at 4 σ (red). Selenium peaks in the anomalous difference Fourier calculated using the Hg phases are shown at 4 σ (green).



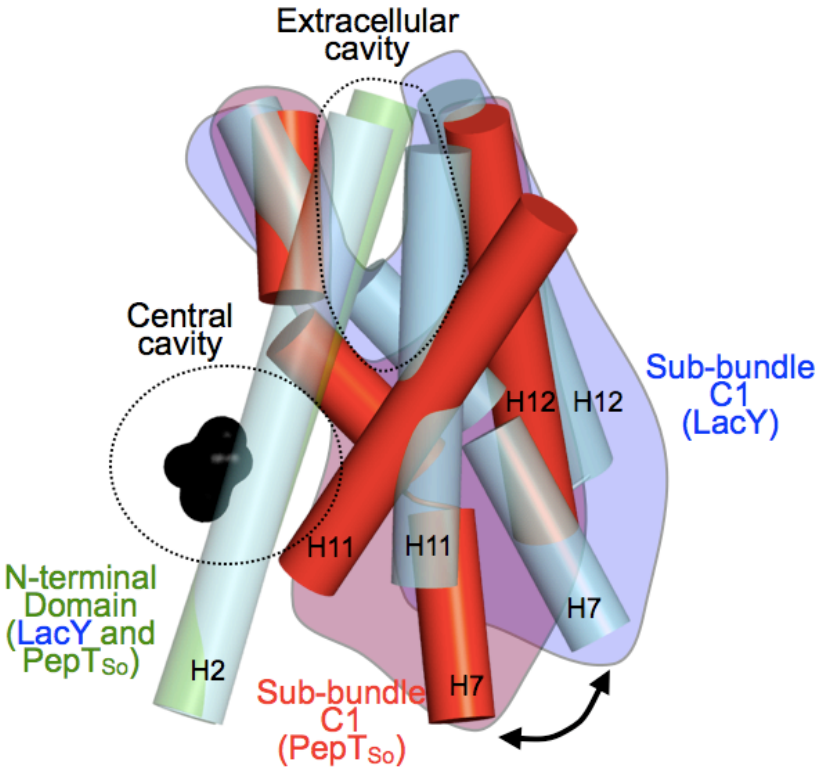
Supplementary Figure 5 Stereo view of the final 2mFo-DFc electron density map from refinement, contoured at 1 σ. Helix 2 from molecule B is shown.



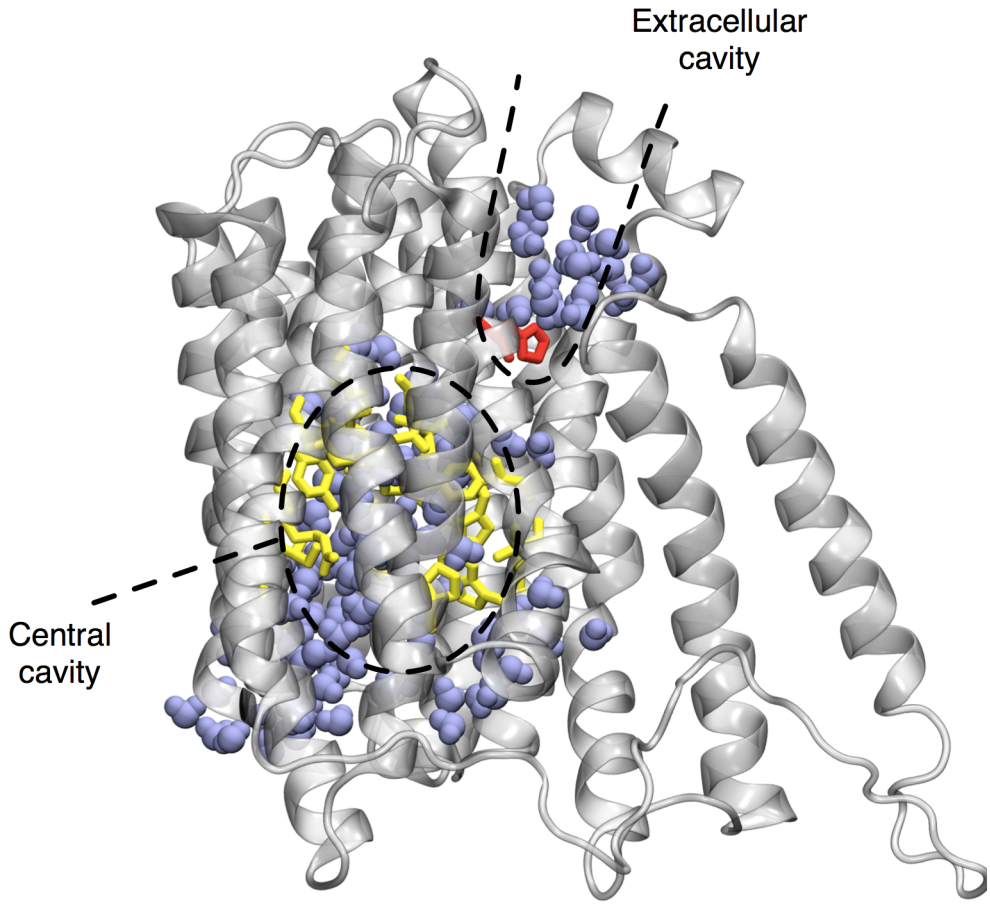
Supplementary Figure 6 Superposition of $PepT_{So}$ onto the substrate bound crystal structure of lactose permease, LacY. The coordinates for LacY (1PV7) (grey) were superimposed onto the structure of $PepT_{So}$ (blue) using secondary structure matching algorithm in the CCP4 program Superpose. The bound TDG substrate (magenta) in LacY is shown in sticks and the mFo-DFc difference density (orange), contoured at 4 σ, observed in the data from the $PepT_{So}$ crystals is shown overlaid.



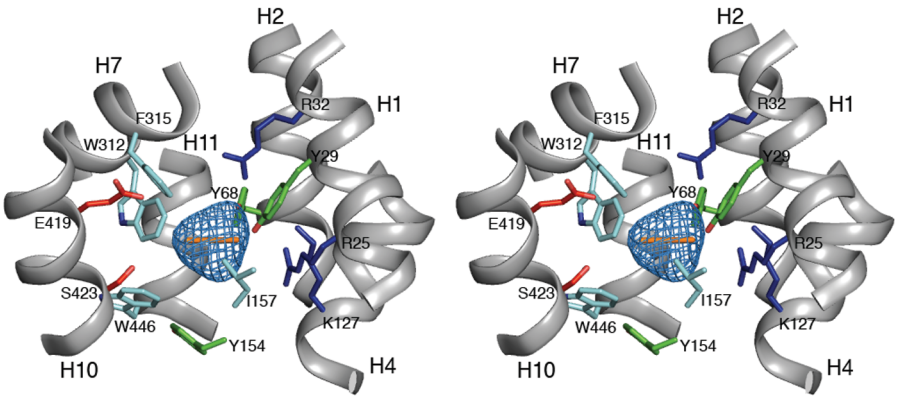
Supplementary Figure 7 Superposition of $PepT_{So}$ and EmrD. The structure of $PepT_{So}$ (yellow) was superposed onto the coordinates of EmrD (2GFP) (grey) using the secondary structure matching algorithm in the CCP4 program Superpose. The extracellular and central cavities identified in $PepT_{So}$ are highlighted. The conserved aspartic acid, D316 (green), of $PepT_{So}$ and glutamate, E227 (cyan), of EmrD are also shown. For clarity, helix H2 containing the conserved histidine, His61 in $PepT_{So}$ has been omitted.



Supplementary Figure 8 Rotation of sub-bundle C1 between the occluded conformation of $PepT_{So}$ and inward open conformation of LacY. Superposed transmembrane helices of $PepT_{So}$ and LacY as viewed in the plane of the membrane. The N-terminal six helices of LacY were superposed onto the equivalent helices in $PepT_{So}$ using LSQMAN [Kleywegt et al, 1997]. The angle of rotation to bring the C-terminal six helices of $PepT_{So}$ onto LacY was then calculated. Helices comprising the C1 sub-bundle are colored red, those from LacY cyan. The central and extracellular cavities are highlighted and the modeled di-Alanine peptide is shown as a CPK model.



Supplementary Figure 9 Water molecules within the extracellular and central cavities in PepT$_{So}$. A snapshot from one of the atomistic simulations shows how water (light blue spheres) has penetrated the central binding cavity (yellow sticks). Water also solvates His61 (red sticks) at the base of the periplasmic cavity.



Supplementary Figure 10 The peptide binding site. Rotated stereo view of the central cavity in the plane of the membrane. Residues are colored according to side chain type, Arg and Lys (blue), Glu and Ser (red), Tyr (green) and Trp, Phe and Leu (cyan) as in Figure 4.

Supplementary Table I | Data collection and phasing statistics for Set 1.

	Native	**MMC-1**[a]	**HgAC**
Beamline	Diamond IO2	ESRF ID23eh1	Diamond IO3
Wavelength (Å)	0.979	1.006	1.005
Space Group	$P3_2$	$P3_2$	$P3_2$
Resolution (Å)	40 - 3.6 (3.73 – 3.60)	40 - 4.6 (4.76 – 4.60)	40 - 4.6 (4.76 – 4.60)
Cell dimensions *a, b, c* (Å)	159.4, 159.4 153.0	159.7, 159.7 153.9	159.7, 159.7 153.9
No. Unique Reflections	47,021	21,989	23,218
Completeness (%) [b]	93.8 (88.7)	90.3 (90.3)	96.4 (95.7)
Redundancy [b]	2.0 (1.8)	1.7 (1.7)	2.4 (2.2)
I/σ(I) [b]	9.0 (1.1)	7.1 (1.1)	8.5 (1.68)
R_{merge} (%) [b, c]	8.9 (67.5)	9.3 (82.4)	9.4 (50.2)
R_{cullis} (%) [d] *Isomorphous* *Anomalous*	 --- ---	 0.816 0.981	 0.737 0.968
Phasing Power [e] *Isomorphous* *Anomalous*	 --- ---	 0.488 0.476	 0.689 0.578

[a] For details on derivatisation see Supplementary Materials & Methods.
[b] Values in parentheses refer to data in the highest resolution shell.
[c] The last shell R_{merge} is high for some of the derivative data due to severe anisotropy in the diffraction images.
[d] $R_{cullis} = \sum |F_{PH} - |F_P + F_H|| / \sum |F_{PH} - F_P|$
[e] Phasing Power = rms $(|F_H| / ((F_H + F_P) - (F_{PH}))$

Supplementary Table II | Data collection and phasing statistics for Set 2.

	MMC-2[a]	**MMC-3**[a]	**Se**
Beamline	ESRF ID23eh1	ESRF ID23eh1	ESRF ID29
Wavelength (Å)	1.005	1.005	0.978
Space Group	$P3_2$	$P3_2$	$P3_2$
Resolution (Å)	40 - 4.0 (4.14 – 4.0)	40 - 4.5 (4.66 – 4.50)	40 - 5.0 (5.18 – 5.0)
Cell dimensions *a, b, c* (Å)	157.6, 157.6, 153.1	158.1, 158.1, 153.6	157.6, 157.6, 153.1
No. Unique Reflections	35,090	25,304	18,227
Completeness (%) [b]	97.7 (96.4)	99.7 (99.6)	99.4 (99.4)
Redundancy [b]	3.2 (2.8)	3.3 (3.0)	3.3 (3.1)
I/σ(I) [b]	8.8 (1.4)	7.7 (1.0)	7.5 (2.0)
R_{merge} (%) [b, c, d]	10.4 (76.1)	16.0 (82.3)	10.9 (50.0)
R_{cullis} (%) [e] *Isomorphous* *Anomalous*	--- 0.976	0.916 0.987	0.726 0.928
Phasing Power [f] *Isomorphous* *Anomalous*	--- 0.379	0.544 0.525	1.058 0.574

Supplementary Table III | Superimposition analysis between $PepT_{So}$ and lacY.

Protein domain*	Average Cα distances (Å)	Cα atom range $PepT_{So}$	Cα atom range lacY (1PV7)
Overall	3.62	15-499	10-400
N	2.79	15-195	10-184
N-1	2.66	H1 15-42; H5 (148-156 + 158 174); H6 176-195	H1 10-37; H5 (138-146 + 148-164); H6 165-184
N-2	2.95	H2 (49-62 + 70-78); H3 89-103; H4 117-137	H2 (41-54 + 63-71); H3 74-88; H4 116-136
C	4.39	299-499	219-400
C-1	4.72	H7 (299-309 + 310-317 + 319-326); H11 (447-450 + 451-466); H12 479-499	H7 (219-229 + 231-238 + 240-247); H11 (351-354 + 356-371); H12 380-400
C-2	4.09	H8 (336-352 + 353-358); H9 375-397; H10 406-434	H8 (260-276 + 278 -283); H9 287-309; H10 313-341

* Calculations were made in O using atoms from chain B of $PepT_{So}$ against atoms from chain A of LacY (1PV7).

Supplementary Table IV | Molprobity Summary Statistics.

All-Atom Contacts	Clashscore, all atoms:	16.42	97th percentile* (N=37, 3 Å – 999 Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers (%)	16.35	
	Ramachandran outliers (%)	0.89	
	Ramachandran favoured (%)	89.38	
	Cβ deviations >0.25 Å	60	
	MolProbity score**	3.20	82nd percentile* (N=342, 3.25 Å – 3.87 Å)
	Residues with bad bonds: (%)	0.00	
	Residues with bad angles: (%)	0.00	

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

** MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

MATERIALS AND METHODS

Amino acid sequence

The amino acid sequence of the $PepT_{So}$ protein produced for this study. The second and third amino acids were mutated from the original wild type sequence of MTTP during cloning. The C-terminus has the following extension after cleavage of the GFP during purification GSENLYFQ. These are highlighted as underlined residues in the sequence below.

MNSPVDAPKWPRQIPYIIASEACERFSFYGMRNILTPFLMTALLLSIPEELRGAV
AKDVFHSFVIGVYFFPLLGGWIADRFFGKYNTILWLSLIYCVGHAFLAIFEHSV
QGFYTGLFLIALGSGGIKPLVSSFMGDQFDQSNKSLAQKAFDMFYFTINFGSFF
ASLSMPLLLKNFGAAVAFGIPGVLMFVATVFFWLGRKRYIHMPPEPKDPHGFL
PVIRSALLTKVEGKGNIGLVLALIGGVSAAYALVNIPTLGIVAGLCCAMVLVMG
FVGAGASLQLERARKSHPDAAVDGVRSVLRILVLFALVTPFWSLFDQKASTWI
LQANDMVKPQWFEPAMMQALNPLLVMLLIPFNNFVLYPAIERMGVKLTALRK
MGAGIAITGLSWIVVGTIQLMMDGGSALSIFWQILPYALLTFGEVLVSATGLEF
AYSQAPKAMKGTIMSFWTLSVTVGNLWVLLANVSVKSPTVTEQIVQTGMSV
TAFQMFFFAGFAILAAIVFALYARSYQMQDHYRQATGSENLYFQ

Protein Expression and Purification

The gene encoding $PepT_{So}$ (SO_0002, Uniprot identifier Q8EKT7) was amplified from a previously constructed expression plasmid pMPSIL0079A and cloned into the pWaldo-GFPe plasmid [(Drew et al, 2001)] using primer pairs 5'-GGAATTCCTCGAGATGACTACACCTGTTGATG-3' and 5'-CGAGGAAGATCTTGTCGC TTGCCGATAATGATC-3' encoding *XhoI* and *BglII* restriction sites, respectively. Expression of the $PepT_{So}$-GFP fusion was carried out in *E. coli* C43 (DE3) cells [(Miroux & Walker, 1996)] and was measured using whole cell and in gel fluorescence as described in Drew *et al*. [(Drew et al, 2006)]. Large volume cultures were carried out using batch fed bioreactor vessels. *Escherichia coli* C43 (DE3) cells freshly transformed with the $PepT_{So}$-pWaldo-GFPe vector were used to inoculate a starter culture consisting of 20 mL of MDG media [(Studier, 2005)]. After overnight culturing at 37 °C the starter culture was added to either a 5 or 15 L bioreactor prepared with TB media in a 1/1000 dilution. Culturing and induction conditions were as described above for the small volume cultures. Following addition of isopropyl-β-D-thiogalactopyranoside (IPTG) the aeration and impeller speeds were set to the maximum allowed limits to support optimal biomass production. Harvested cells were resuspended in phosphate buffered saline (PBS) and frozen at -80 °C. Samples were thawed and the cells disrupted using a TS series continuous cell disruptor (Constant Systems, UK) operating at 35 kpsi at 4 °C. Following removal of cellular debris at 14,000 g, whole cell membranes were isolated using centrifugation at 130,000 g for 2 hours at 4 °C. Membrane were resuspended in ice cold PBS and flash frozen in 50 mL falcon tubes for storage. Protein purification was carried out essentially as described in Drew *et al*. [(Drew et al, 2008)]. Briefly, thawed membranes were

diluted to a final total protein concentration of 3-5 $mg.ml^{-1}$ in 1 x PBS, 10 mM imidazole (pH 8.0), 150 mM NaCl, 10 % glycerol (wt/vol) and 1 % n-Dodecyl-β-D-maltoside (DDM) (Anatrace, USA) to a final volume of 360 mL. Membranes were solubilized for 1 hour at 4 °C with mild stirring and non-solubilized membranes were removed by centrifugation at ~100,000 g for 45 minutes. Nickel-NTA Superflow resin (Qiagen, DE) was then added to the sample, using a ratio of 1 mL of resin per mg of GFP present. The resin was incubated with the solubilized membranes for 1 hour at 4 °C with mild stirring. The resin was then packed into a glass econo-column (BioRad, USA) under gravity. The resin was washed with ~ 30 column volumes (CVs) of wash buffer (1 x PBS, 30 mM imidazole (pH 8.0), 150 mM NaCl, 10 % glycerol (wt/vol) and 0.1 % DDM). The resin was further washed with 10 CVs of 40 mM imidazole. The bound $PepT_{So}$-GFP fusion protein was the eluted using 250 mM imidazole. Eluted protein was then combined with an equal amount of hexa histidine tagged TEV protease and dialyzed overnight against 20 mM Tris-HCl, pH 7.50, 150 mM NaCl, and 0.03 % DDM at 4 °C. Following overnight dialysis and cleavage of the fusion protein, the sample was recovered and subjected to reverse purification. The sample was first filtered using a 0.22 μm syringe filter (Millipore, USA) and then passed through a 5 mL HisTrap column (GE Healthcare, USA) using a P-1 peristaltic pump. Contaminants co-purifying with the $PepT_{So}$ fusion from the first affinity purification step were subsequently captured by the HisTrap resin, along with the hexa histidine tagged TEV protease and octa histadine tagged GFP. Pure $PepT_{So}$, as judged by SDS-PAGE, was then collected in the flow through from the HisTrap column. $PepT_{So}$ was concentrated to a final volume of 500 μL and applied to a Superdex 200 10/300 gel filtration column (GE Healthcare, USA), pre-equilibrated in 20 mM Tris-HCl, pH 7.50, 150 mM NaCl, and 0.03 % DDM using a flow rate of 0.4

$ml.min^{-1}$. Fractions containing the purified protein were pooled and concentrated to a final volume of 12-15 $mg.ml^{-1}$ for crystallization.

Crystallisation

A number of potential crystal conditions were identified using the MemGold crystallization screen [(Newstead et al, 2008)] (Molecular Dimensions, UK). Promising crystals were observed in 33 % PEG 300, 0.05 M Glycine, pH 9.5, and 0.1 M NaCl (MemGold condition 2.12) at 4 °C. Following optimization around this condition, the best diffracting crystals were obtained in 30 % PEG 300, 0.1 M MES pH 6.50 and 0.1 M NaCl using the hanging drop vapor diffusion technique at 4 °C. To form drops, 1 mL of protein sample was mixed with an equal volume of reservoir solution containing 26 % PEG 300, 0.1 M MES pH 6.50 and 0.1 M NaCl. Crystals appeared after 2-3 days and could be harvested up to several weeks after. For cryoprotection the crystals were transferred to a solution containing 36-40 % PEG 300, 0.1 M MES pH 6.50, 0.1 M NaCl and 0.03% DDM, before being flash vitrified in liquid nitrogen. The crystals always showed strong anisotropic diffraction, with the best crystals diffracting between 3.6 - 3.8 Å in the best direction.

Mercury derivatives

Mercury derivatives were prepared in two ways, either prior to crystallization or through soaking of native crystals. Data sets MMC-1, HgAc, and MMC-3 were obtained from crystals grown using protein pre-incubated with mercury compounds, whilst data set MMC-2 was obtained through soaking a seleno-L-methionine

incorporated crystal. Modification of free cysteines prior to crystal growth was carried out following the reverse purification step and prior to gel filtration. The protein was incubated with a three times molar excess of either methyl mercury acetate (MMC) or mercury acetate (HgAc) and incubated at 19 ºC for one hour. The sample was then applied to the Superdex 200 10/300 gel filtration column as described for the native protein. For soaking, the native or seleno-L-methionine crystals were transferred to a fresh drop containing 35 % PEG 300, 0.1 MES pH 6.50, 0.1 M NaCl and 0.03 % DDM with either 2 mM MMC or HgAc and soaked overnight suspended over a well containing the same concentration of PEG 300.

Seleno-L-methionine incorporation

Incorporation of seleno-L-methionine was carried out using *E. coli* C43 (DE3) cells. Overnight inoculum in MDG media was diluted into PASM-5052 media [Studier, 2005]. Induction and culturing were performed as described for the native protein.

Data collection and phasing

Data were collected either on beamlines ID23eh1 or ID29 at the European Synchrotron Radiation Facility (ESRF) or on IO2 and IO3 at the Diamond Light Source Ltd., UK. Data were processed and scaled using the HKL suite of programs [Otwinowski & Minor, 1997]. Further processing was carried out using programs from the CCP4 package (Supplementary Tables I and II) [CCP4, 1994]. The space group was determined to be $P3_2$ with three molecules in the asymmetric unit. Two mercury sites were initially located in each of the three molecules in the asymmetric unit. These were

found manually using RSPS [Knight, 2000] in derivative difference Patterson maps calculated in FFT. The mercury positions were refined and initial phases calculated using SHARP [de La Fortelle & Bricogne, 1997]. The resulting phases were used to locate a third mercury site, giving 9 sites in total. Phases from the mercury sites were used to locate the selenium sites in difference Fourier maps and these were added to the 9 mercury sites to improve the phase information.

Over 30 datasets were collected during the structure determination from > 1000 crystals screened at the synchrotron. These datasets were grouped according to isomorphism and anomalous signal strength, as judged by Xtriage [Adams et al, 2002]. Each set of isomorphous data was used to calculate phases in SHARP and the calculated maps analyzed. This process resulted in two sets of non-isomorphous data that gave good initial maps that showed clear solvent boundaries around the three molecules. Set1 contained the high-resolution native and two mercury derivative data sets, MMC-1 and HgAc. Set2 contained two mercury derivative data sets, MMC-2, which was a seleno-L-methionine incorporated crystal soaked overnight in 2 mM MMC and MMC-3. Set2 also contained a separate seleno-L-methionine crystal collected at the anomalous edge for selenium. The non-crystallographic symmetry present in the packing of the crystal was used to improve the phase information. A rough mask was calculated around one of the three molecules using O [Jones & Kjeldgaard, 1997] and the NCS operators calculated and improved using programs from the RAVE [Jones & Kjeldgaard, 1997] package. These were then used in DM [Cowtan, 1994] to average over the three molecules in the asymmetric unit. The two sets of phases were also combined using cross crystal averaging in DMMmulti [Cowtan, 1994]. The resulting maps were then of sufficient quality to see all 14 helices from each of the three molecules. The optimal solvent content for density modification was found to be 79 %.

Model building and refinement

An initial Cα model was built into the density from all three sources of phases described above, from SHARP, DM and DMMulti, using O [(Jones & Kjeldgaard, 1997)]. Reflections for the R_{free} calculation were selected in resolution shells using SFTOOLS. The partial models were further cycled back into phase calculation in SHARP to improve the initial solvent envelope used for the solvent flipping procedure. This cycle was iterated many times until a reasonably complete model could be built. The amino acid side chains were then built into the partial model using the selenium and mercury sites to determine the correct register. The heavy atom sites were broadly distributed within the primary structure allowing some confidence in the assignment of the register at this resolution. The resulting map allowed the model to be traced from residue 13 to 500, with two breaks, the first in the cytoplasmic loop connecting the N- and C-terminal halves (Pro206 – Lys226) and the second at the end of H8 (Leu359 – Thr371). Refinement of the model was carried out in BUSTER [(BUSTER-TNT 2.X)] against the highest resolution dataset. Experimental phase information was included in the form of Hendrickson-Lattman coefficients throughout. Refinement was dramatically improved by anisotropic truncation of the structure factors [(Strong et al, 2006)]. The resolution along the A and B axes was truncated to 4.3 Å, whilst the C axis was kept at 3.6 Å. Refinement was carried out using strict NCS restraints with the –autoncs_noprune flag. Inclusion of TLS parameterization also improved the quality of the model and subsequent maps. Each molecule was modeled as a single TLS group using the TLSbasic macro in BUSTER. To increase the contribution of the high-resolution terms in the resulting 2mFo-DFc electron

density maps, a B-factor sharpening term was introduced during map calculation in FFT of between -50 and -80 $Å^2$. Model validation was carried out using the Molprobity server (Davis et al, 2007) (Supplementary Table 4). The quality of the model compares favorably with structures of a similar resolution in the Protein Data Bank. Images were prepared using PyMol (The PyMOL Molecular Graphics System) and VMD (Humphrey et al, 1996).

***In vivo* [^{3}H]-Glycylsarcosine transport assay.**

Peptide transport was assayed using *E. coli* strain BL21-gold (DE3) (Stratagene) cells harbouring the pTTQ18-RGSH6-UraA-derived (Rahman et al, 2007) plasmid construct pMPSIL0079A, which encodes $PepT_{So}$ bearing a C-terminal $RGSH_6$ tag. Bacteria were cultured at 37 °C in M9 minimal medium containing 0.2% glycerol and 0.2% casamino acids until an A_{600nm} of 0.8 had been reached, and then incubated for a further 1 h with 0.2 mM isopropyl-β-D-thiogalactoside (IPTG) to induce transporter expression. Cells were harvested by centrifugation and resuspended in transport buffer (5 mM MES buffer, pH 6.6, containing 150 mM KCl and 20 mM glycerol), to yield an A_{680nm} of 2.75. Uptake of the radioactive dipeptide [^{3}H]-Glycylsarcosine (Gly-Sar) (Moravek Biochemicals Inc.) was then measured at 37 °C. In brief, samples of bacteria (75 µl) were mixed with 25 µl of transport buffer containing the appropriate concentration of Gly-Sar. After an uptake period of 15 s, employed to approximate initial velocities of transport, cells were filtered to terminate transport and washed twice with ice-cold transport buffer (Henderson & Macpherson, 1986). For estimation of the apparent V_{max} and K_m values for transport, uptake was measured over a substrate concentration range of 10 µM to 5 mM. $PepT_{So}$-mediated transport was calculated by subtraction of rates seen in induced cells harboring a control expression

vector encoding the *E. coli* nucleoside transporter NupG. The resultant rates were expressed per mg of $PepT_{So}$ protein, as measured by quantitative immunoblotting. Data were fitted to the Michaelis-Menten equation using the non-linear curve-fitting program KaleidaGraph (version 4.0, Synergy software) in order to estimate kinetic parameters

***In vivo* β-Ala-Lys-AMCA competition transport assay.**

Peptide transport was assayed using *E. coli* strain BL21-gold (DE3) (Stratagene, USA) cells harboring the pTTQ18-RGSH6-UraA-derived [(Rahman et al, 2007)] plasmid construct pMPSIL0079A, which encodes $PepT_{So}$ bearing a C-terminal $RGSH_6$ tag. Bacteria were cultured at 37 °C in M9 minimal medium containing 0.2% glycerol and 0.2% casamino acids until an A_{600nm} of 0.6 had been reached, and then incubated for a further 1 h with or without 0.5 mM IPTG to induce transporter expression. Cells were harvested by centrifugation and resuspended in transport buffer (5 mM MES buffer, pH 6.6, containing 150 mM KCl and 20 mM glycerol), to yield an A_{680nm} of 1.5. Uptake of the fluorescent dipeptide β-Ala-Lys-N_e-7-amino-4-methylcoumarin-3-acetic acid (β-Ala-Lys-AMCA; Biotrend, Cologne, Germany) was then measured at 37 °C essentially as described by Weitz *et al*. [(Weitz et al, 2007)]. In brief, samples of bacteria (40 μl) were mixed with 50 μl of transport buffer with or without competitive inhibitors plus 10 μl of 500 μM β-Ala-Lys-AMCA, yielding a final substrate concentration of 50 μM. After 15 min, washing the cells twice with ice-cold transport buffer terminated uptake. The cells were then lysed by incubation for 10 min at 25 °C in 20 μl lysis buffer (50 mM HEPES buffer, pH 8.0, 5 mM $MgCl_2$, 1% Triton X-100, 10 U ml^{-1} OmniCleave endonuclease (Epicentre Biotechnologies) and 0.1 mg ml^{-1} lysozyme). The lysates were transferred to a 384-well plate and fluorescence

measured in a Fluostar Optima plate reader (BMG labtech) using an excitation wavelength of 340nm and an emission filter with a wavelength of 460 nm.

Molecular Dynamics

Plausible conformations of the missing loops in the structure were generated using Modeller 9v7 (Sali & Blundell, 1993). The structure was then converted to a coarse-grained (CG) representation (Bond & Sansom, 2006; Marrink et al, 2004). To maintain the tertiary structure an elastic network model was applied to the protein Cα particles with a distance cutoff of 7 Å and a force constant of 10 kJ mol $^{-1}$ Å $^{-2}$. The resulting CG structure was placed in the center of a box containing water, counter ions and either POPC or DPPC lipids as described by Scott *et al*. (Scott et al, 2008). Coarse-grained molecular dynamics was run for 0.2 μs using GROMACS3.3.1 (Van Der Spoel et al, 2005). Electrostatic and van der Waals interactions were only calculated between atoms separated by less than 12 Å with a switching function applied to the latter after 9 Å. The temperature was maintained at 323 K using a Berendsen thermostat with a coupling constant of 1.0 ps (Berendsen et al, 1984). The pressure was also held at 1.0 bar by a Berendsen barostat (Berendsen et al, 1984); this was applied semi-isotropically with a coupling constant of 40 ps and a compressibility of 1 x 10 $^{-5}$ bar. The integration timestep was 20 fs. Ten simulations were run. The box and, therefore, the number of lipids were different in each simulation. In all ten cases, the bilayer self-assembled and the protein inserted into the bilayer as expected. The protein and lipids from the final structure of each of these coarse-grained simulations were then converted to an atomistic representation as described by Stansfeld *et al*. (Stansfeld et al, 2009). The psfgen,

solvate and autoionize packages of VMD1.8.6 were used to add hydrogens, waters and neutralising chloride ions. Three atomistic simulation unit cells of slightly different sizes were constructed in this way. In each, the lipid bilayer was made up of POPC. Fully-atomistic molecular dynamics were then run for 50 ns using NAMD2.7b1 (Phillips et al, 2005) and the CHARMM27 forcefield with the CMAP correction (Mackerell, 2004). The lengths of all bonds in the protein and lipids or water that involve hydrogen were restrained using the SHAKE (Mackerell, 2004) and SETTLE (Mackerell, 2004) algorithms, respectively, permitting an integration timesteps of 2 fs. Electrostatic forces were calculated by the PME method (Mackerell, 2004) using a real-space cutoff of 12 A° . Van der Waals forces were cutoff at 12 Å and a switching function was applied from 10 Å. The pressure was held at 1.0 bar using a Berendsen barostat (Berendsen et al, 1984) with a coupling constant of 0.2 ps and a compressibility of 4.46 x 10^{-5} bar^{-1}. The temperature was kept at 310 K using Langevin dynamics with a damping coefficient of 1 ps^{-1}. Figures were prepared using VMD (Humphrey et al, 1996; Kleywegt et al, 1997).

Competing Financial Interests

No competing financial interests.

REFERENCES

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**(Pt 5): 760-763

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11): 1948-1954

Berendsen HJ, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **81**(8): 3684-3690

Bond PJ, Sansom MS (2006) Insertion and assembly of membrane proteins via simulation. *J Am Chem Soc* **128**(8): 2697-2704

BUSTER-TNT 2.X GPL, Sheraton House, Cambridge CB3 0AX, UK

Cowtan K (1994) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. **31**: 34-38

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity:

all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**(Web Server issue): W375-383

de La Fortelle E, Bricogne G (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology* **276**: 472

Drew D, Lerch M, Kunji E, Slotboom DJ, de Gier JW (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* **3**(4): 303-313

Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, Iwata S (2008) GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat Protoc* **3**(5): 784-798

Drew DE, von Heijne G, Nordlund P, de Gier JW (2001) Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett* **507**(2): 220-224

Henderson PJF, Macpherson AJ (1986) Assay, genetics, proteins and reconstitution of proton-linked galactose, arabinose and xylose transport systems of Escherichia coli. *Methods in Enzymology* **125**: 387-429

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**(1): 33-38, 27-38

Jones TA, Kjeldgaard M (1997) Electron-density map interpretation. *Methods in Enzymology* **277**: 173-208

Kleywegt GJ, Zou JY, Kjeldgaard M, Jones TA (1997) Crystallography of Biological Macromolecules. In *International Tables for Crystallography*, Rossman MG, Arnold E (eds), Vol. F, pp 353-356 & 366-367. Dordrecht: Kluwer Academic Publishers

Knight SD (2000) RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr D Biol Crystallogr* **56**(Pt 1): 42-47

Mackerell AD, Jr. (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* **25**(13): 1584-1604

Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained semiquantitative lipid simulations. *Journal of Physical Chemistry B* **108**(2): 750-760

Miroux B, Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**(3): 289-298

Newstead S, Ferrandon S, Iwata S (2008) Rationalizing alpha-helical membrane protein crystallization. *Protein Sci* **17**(3): 466-472

Otwinowski Z, Minor W (1997) Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**: 307

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**(16): 1781-1802

Rahman M, Ismat F, McPherson MJ, Baldwin SA (2007) Topology-informed strategies for the overexpression and purification of membrane proteins. *Mol Membr Biol* **24**(5-6): 407-418

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**(3): 779-815

Scott KA, Bond PJ, Ivetac A, Chetwynd AP, Khalid S, Sansom MS (2008) Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16**(4): 621-630

Stansfeld PJ, Hopkinson R, Ashcroft FM, Sansom MS (2009) PIP(2)-binding site in Kir channels: definition by multiscale biomolecular simulations. *Biochemistry* **48**(46): 10926-10933

Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**(21): 8060-8065

Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**(1): 207-234

The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* **26**(16): 1701-1718

Weitz D, Harder D, Casagrande F, Fotiadis D, Obrdlik P, Kelety B, Daniel H (2007) Functional and structural characterization of a prokaryotic peptide transporter with features similar to mammalian PEPT1. *J Biol Chem* **282**(5): 2832-2839