4.9 Discussion

The optimisation of the *Mm*PepT1$^{\text{ECD}}$ crystals was serendipitous. The first batch of buffer bought from Emerald Bioscience produced the rod crystals (Figure 4.5A) which gave the 2 Å diffraction observed in N1, N2 and the Hg-SAD datasets. However, efforts to recreate this diffraction have proved unsuccessful. Only the box morphology crystals (3 Å diffraction), which have the same cell dimensions, have been grown since from either new bought buffer from Emerald Biosciences or made manually. Sparse matrix micro seeding experiments (described in § 2.7.5) using the box crystals as seeds, also did not produce any highly diffracting crystals. So it appears that the first batch of bought buffer was unique.

The N2 dataset highlights an error made during the data collection of a high resolution dataset; an incomplete dataset was collected during the 100 ° ϕ wedge. A subsequent wedge was collected so that a complete sampling of the reciprocal lattice could be made. However, this should not have been necessary, if the strategy produced by `iMosflm` (Powell, Johnson, and Leslie 2013) or `EDNA` (Incardona *et al.* 2009) had been followed.

Overall the structure of the ASU has been determined to 2.10 Å within an acceptable error. The data collection statistics (Table 4.2) are in keeping with current views on data reduction (Evans and Murshudov 2013). The model was built to an R_{work} and R_{free} of 0.193 and 0.242 respectively, indicating the model fits the density. 95% of the Phi/Psi angles are within favoured regions of the Ramachandran plot. The *Mm*PepT1$^{\text{ECD}}$ structure was in the 99th percentile of structures for a resolution of 2.10 Å. These two metrics are indicators of a correctly built structure within the current knowledge of all structures in the PDB.

The model of the *Mm*PepT1$^{\text{ECD}}$ indicates that the domain is predominantly structured. The ECD can be further divided into two sub-domains, lobes connected by a linker. It is not clear from examination of the ASU if the dimerisation observed in the ASU is a physiological arrangement the ECD forms between separate transporters in a biological membrane. The structure prompts several questions which are explored in Chapter 6 such as:

- Is the dimerisation observed in the ASU physiological?
- Are the two lobes free in solution or do they have a preferred orientation?

- How does the ECD sit in relation to the TM domain and does it have a functional role in transport?

5

Crystal structure of the *R. norvegicus* PepT2^{ECD}

5.1 Summary

In this chapter, the crystallisation and X-ray structure determination of the PepT2^{ECD} is explained. The organism and length of the PepT2^{ECD} required optimisation. A *Rn*PepT2^{ECD} 410 - 601 construct was found to crystallise. Trigonal crystals were grown using hanging drop vapour diffusion with a well condition of 21% v/v PEG 3350, 0.2 M $(NH_4)_2$ Citrate pH 5.8. The phases were solved using the SeMet SAD method and the structure was solved in space-group $P3_221$ to 2.84 Å resolution. These crystals were used as seeds to crystallise the protein in a different space-group, $P4_12_12$ that diffracted X-rays to 2.06 Å. The model shows that the PepT2^{ECD}, like PepT1^{ECD} is composed of two lobes connected by a flexible linker, although these lobes appear to be in a different conformation.

5.2 *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 spare matrix crystallisation experiments.

The *Mm*PepT2$^{\mathrm{ECD}}$ construct 410–609 was screened for crystallisation in sparse matrix 96-well screens. Protein was prepared prior to crystallisation by passing the sample through a Superdex S75 16/60 GF column to assess the mono-dispersity and purity of the sample (Figure 5.1). The *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 crystallisation concentration was screened in the PEG/Ion matrix at 5.0, 7.5, 10.0, 15.0, and 20.0 mg/mL. The optimum *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 concentration was found to be 7.5 mg/mL for PEG/Ion which showed precipitation in approximately 30% of drops. This concentration was then used for the other screens.

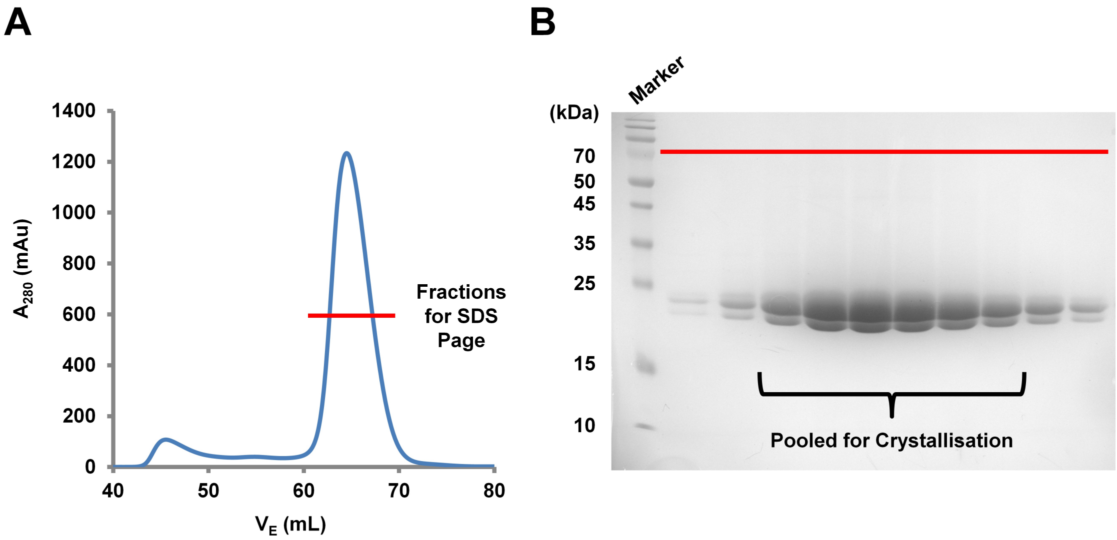


Figure 5.1: *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 GF trace and gel. (A) A Superdex S75 16/60 GF trace in 20 mM Tris pH 7.5, 150 mM NaCl. The red band indicates the fractions run on a reducing 15% Tris-Gly SDS-PAGE gel **(B)**. The double protein bands present on the gel were both sent for analysis by mass-spectroscopy and were both confirmed to be the full 410 – 609 construct.

Crystallisation screening experiments were prepared at the OPPF-UK using a Cartesian crystallisation robot in 96-well sitting drop Greiner plates (§ 2.7.1). Initially, the GF buffer used for preparation of the *Mm*PepT2$^{\mathrm{ECD}}$ 410–609 sample was 20 mM Tris pH 7.5, 150 mM NaCl. These crystallisation experiments did not yield any crystals. However, many conditions did produce drops which showed micro-crystals or crystalline-like precipitate. Table 5.1 shows the well conditions where crystalline precipitate was observed. 96-well optimisations (see Figure 4.4) were performed using the conditions in Table 5.1, but no improvement was observed. Predominantly the conditions which showed signs of crystallisation were buffered below pH 8.0. To improve the probability of crystallisation, the stability of the construct as a function of pH was assessed using FSC (Ericsson *et al.* 2006; Dupeux *et al.* 2011).

5.2.1 Fluorescence scanning calorimetry of the *Mm*PepT2$^{\mathbf{ECD}}$ 410 – 609 construct

FSC was first proposed by Pantoliano *et al.* (2001) as a HT method to assess the binding of compound libraries to a protein target. As a protein solution is heated, the protein will gradually start to unfold as the H-bond enthalpies are surmounted by the energy of the surrounding solution. As more heat is continually applied, the entire protein chain unfolds. Assuming a folded protein has a hydrophobic core, the process of complete protein unfolding exposes a proportional amount of hydrophobic residues. Unfolding is measured by the incorporation of a fluorescent dye, such as Sypro orange. Sypro orange's fluorescent excitation and emission wavelength is dependent upon the proximity of hydrophobic areas. Therefore the fluorescence of the Sypro orange increases proportionally as more hydrophobic residues are exposed to bulk solvent.

The protocol used for the determination of the protein melting temperature (T_m) was based upon Niesen, Berglund, and Vedadi (2007) (§ 2.8.1). The stability of *Mm*PepT2^{ECD} 410 – 609 construct was examined in the presence of four different buffers: PIPES pH 6.5, MOPS pH 7.0, HEPES pH 7.5 and HEPES pH 8.0. The raw data curves and calculated T_ms are shown in Figures 5.2A and 5.2B respectively. There was an observed drop in T_m from 50.8 to 45.8 °C over pH range 6.5 – 8.0. Dupeux *et al.* (2011) suggested that a T_m of > 45 °C indicates a good probability of crystallisation. The higher the T_m, the higher the stability and the greater the probability of crystallisation. Therefore *Mm*PepT2^{ECD} 410 – 409 was exchanged into 20 mM PIPES pH 6.5 and 20 mM Na Acetate pH 4.5, and sparse matrix crystallisation experiments were repeated. Unfortunately these experiments also yielded no crystals.

Screen	Well	Condition
JCSG+	B7	0.1 M Na Acetate pH 4.5 8% v/v PEG 4,000
	C7	0.2 M Zn Acetate 0.1 M Na Acetate pH 4.5 10% v/v PEG 3,000
	E8	0.1 M Na Acetate pH 4.5 1 M $(NH_4)_2HPO_4$
Wizard^{TM} I+II	E1	0.2 M Li Sulfate 0.1 M Na Acetate pH 4.5 10% v/v PEG 3,000
	E2	0.2 M Zn Acetate 0.1 M MES pH 6.0 35% v/v MPD
	F9	0.1 M Na Acetate pH 4.5 35% v/v MPD
	G6	0.2 M Zn Acetate 0.1 M Imidazole pH 8.0 20% v/v 1,4-Butandiol
	H11	0.2 M Zn Acetate 0.1 M Imidazole pH 8.0 2.5 M NaCl
Wizard^{TM} III+IV	E10	0.2 M $MgCl_2$ 0.1 M Cacodylate pH 6.5 50% v/v PEG 200
PEG/Ion	C5	0.2 M K Acetate 20% v/v PEG 3,350
	C5	4% w/v Tacsimate pH 7.0 12% v/v PEG 3,350

Table 5.1: A list of the conditions which grew *Mm*PepT2$^{\mathbf{ECD}}$ micro-crystals or crystalline like precipitate.

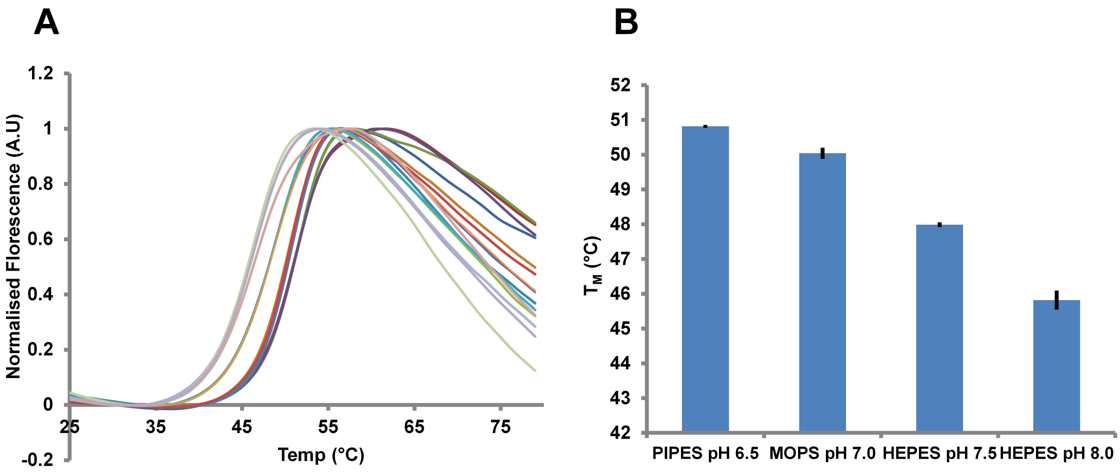


Figure 5.2: *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 FSC results. (A) Raw FSC data curves. **(B)** A bar graph of the averaged T_m of 4 samples with 95% confidence limits. The bar graph shows that the pH of the solution has a significant (to the .05 criterion) effect on the stability of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609.

5.2.2 Reductive methylation of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 construct

Increasing the stability of the protein through buffer optimisation did not appear to influence crystallisation favourably, so other techniques were pursued. Reductive methylation of exposed lysine residues reduces the conformational flexibility of the residues (Cooper *et al.* 2007) and changes the solvent exposed surface of the protein, increasing the propensity of crystallisation (Walter *et al.* 2006). The reductive methylation protocol (§ 2.7.3) used was essentially the same as Kim *et al.* (2008). Figure 5.3 shows the GF traces of the methylated 410 – 609 compared to the unmodified protein. The methylated protein shows a relative decrease in V_e compared to WT, indicating that the methylated protein is larger than WT and the methylation did not affect the mono-dispersity of the protein. The methylated protein was exchanged into 20 mM PIPES pH 6.5, 150 mM NaCl and used for sparse matrix crystallisation experiments. Again no crystals were observed in any of the experimental drops.

5.2.3 *In situ* limited proteolysis of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 construct

The *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 construct has a T_m at pH 6.5 of 50.8 °C, suggesting a stable folded protein. Diffracting crystals were not observed in approximately 13,000

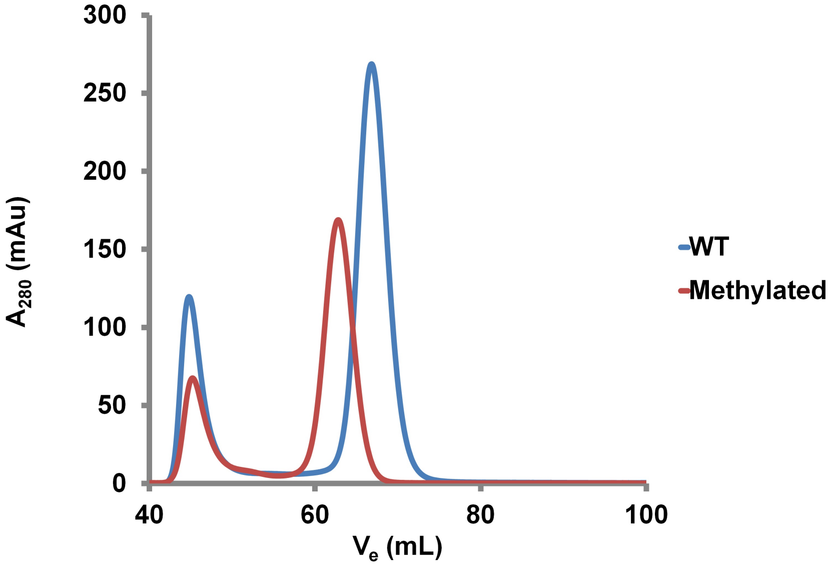


Figure 5.3: A comparison between the *Mm*PepT2 410 – 609 WT and methylated protein GF traces. The Superdex S75 16/60 GF column was equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl. The methylated *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 V_e decreased suggesting the protein has a larger volume compared to WT.

experimental drops, which suggests that a flexible region of the 410 – 609 construct might be preventing crystallisation. *In situ* limited proteolysis has been shown to be effective in aiding the crystallisation of bacterial and eukaryotic proteins with exposed flexible regions (Dong *et al.* 2007; Wernimont and Edwards 2009).

The Floppy-Choppy kit (Jena Biosciences) is composed of 4 proteases: α−chymotrypsin, trypsin, subtilisin A and papain. An initial cleavage assay was performed to see if discrete fragmentation was observed. Each protease was diluted into 3 stock solutions: 0.1, 0.01 and 0.0001 mg/mL. Each dilution was mixed 1:1 with 10 µL of 20 mg/mL 410 – 609 solution and incubated for 2 hr at RT. The reaction was quenched with reducing SDS-PAGE gel loading buffer and the samples run on a NuPAGE® 10% Bis-Tris Midi gel (Figure 5.4A).

The gel (Figure 5.4A) indicated that the 410 – 609 construct did fragment. A consistent fragment of 21 kDa was present in all of the protease digests. This suggested that a small fragment exists at either the N- or the C-terminus of the construct which was readily cleaved, and so is likely to be exposed and flexible. Figure 5.4B shows the *Mm*PepT2 410 – 609 sequence with the `PeptideCutter` (Gasteiger *et al.* 2005) predicted trypsin and α-chymotrypsin sites highlighted. The blue circle shows the probable cleavage point

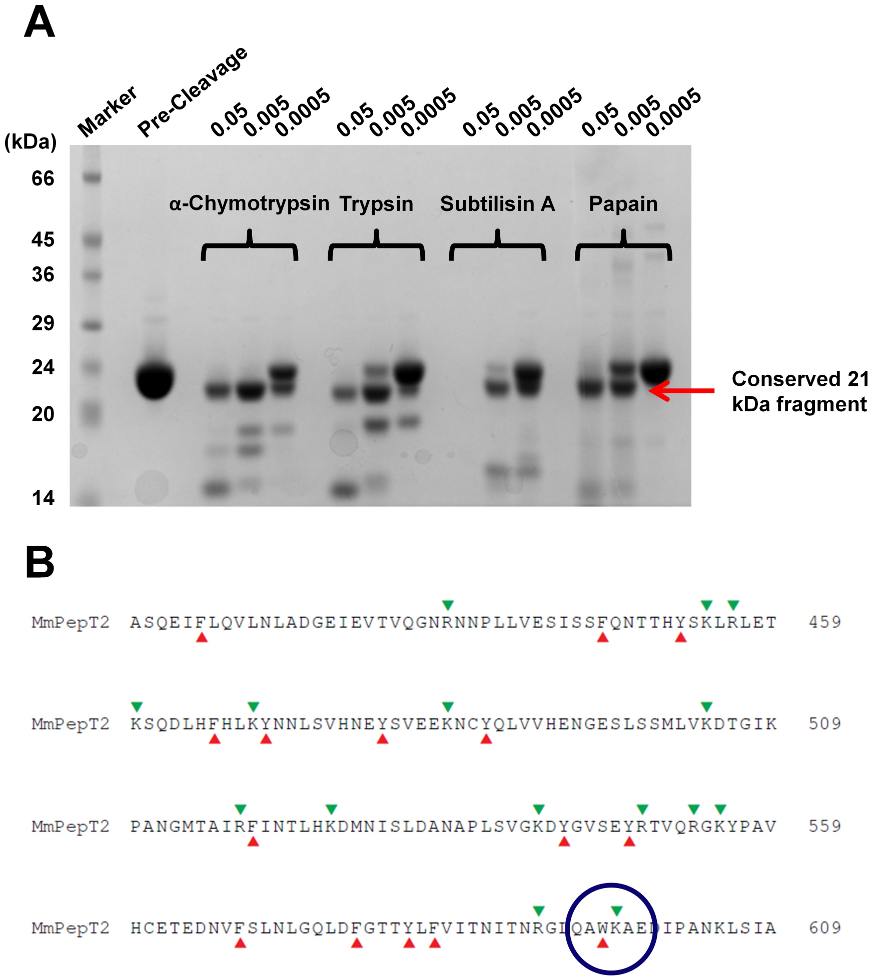


Figure 5.4: Results and analysis of the *Mm*PepT2$^{\text{ECD}}$ 410 – 609 limited proteolysis experiment. (A) A reducing NuPAGE® 10% Bis-Tris Midi gel showing the cleavage of *Mm*PepT2$^{\text{ECD}}$ 410–609 using α-chymotrypsin, trypsin, subtilisin A and papain. 0.05, 0.005 and 0.0005 corresponds to the final concentration of the proteases (mg/mL) in the experiment (the 0.05 mg/mL fraction for the subtilisin A did not show up on the gel). **(B)** The *Mm*PepT2$^{\text{ECD}}$ 410 – 609 amino-acid sequence. The `PeptideCutter` (Gasteiger *et al.* 2005) predicted cleavage residues for trypsin (▼) and α-chymotrypsin (▲) are shown. The blue circle indicates the likely cleavage point which resulted in the 21 kDa fragment.

which gave rise to the 21 kDa fragment observed in Figure 5.4A.

In situ crystallisation experiments were prepared using α-chymotrypsin and trypsin, as these two proteases seemed to give the clearest 21 kDa band. Prior to *in situ* proteolysis crystallisation experiments, 10 µL of 20 mg/mL protein solution was mixed with 10 µL of 40 µg/mL protease. Sparse matrix crystallisation experiments were prepared as described in § 2.7.1. Some potential hits were observed, but were discovered to be salt crystals.

5.2.4 Using NMR to investigate the structural flexibility in the 410 – 601 *Mm*PepT2$^{\mathbf{ECD}}$ construct

The crystallisation experiments of *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 were not yielding diffracting crystals. To assess the flexibility of the 410 – 609 construct, an Heteronuclear single quantum coherence (HSQC) experiment was performed (§ 2.8.2). Figure 5.5 shows an HSQC spectra of *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609. There are 7 – 8 very high intensity peaks in the centre of the spectrum (8 – 9 ^{1}H ppm, 116 – 126 ^{15}N ppm) which suggests backbone amide protons in a highly flexible state (Figure 5.5, red circle). The W596 indole ^{15}N peak (10 ^{1}H ppm, 129 ^{15}N ppm) also appears as a doublet, suggesting that the indole ring has two distinct conformations. It is not possible to assign the backbone amide peaks solely from the HSQC. However, given the conformational flexibility of the W596 observed in the HSQC spectra, and the results from the limited proteolysis experiment, the 7 – 8 highly flexible residues were assumed to be at the C-terminus.

5.3 *Mm*PepT2$^{\mathbf{ECD}}$ 410 – 601 sparse matrix crystallisation experiments.

The *in situ* proteolysis and HSQC experiment suggested that there was a flexible region in the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609. This flexible region may be preventing crystallisation. Based on the sequence alignments of the PepT2$^{\mathrm{ECD}}$ sequences (Figure 3.3), it was decided that this region was likely to be in the C-terminus of the ECD. Therefore a new truncated construct was designed, 410 – 601.

*Mm*PepT2$^{\mathrm{ECD}}$ 410 – 601 was expressed and purified as described in the § 3.5.3. The GF

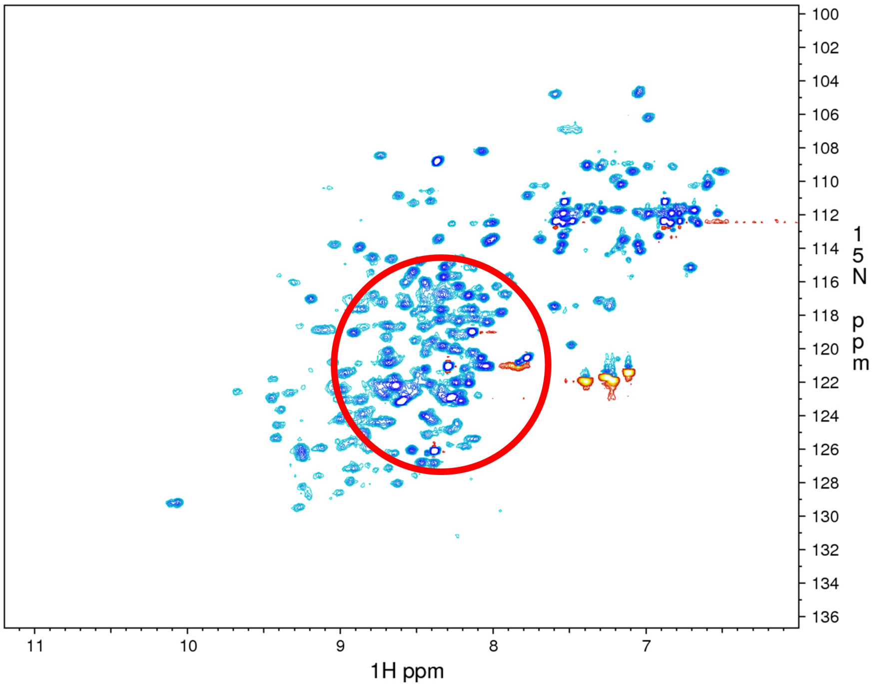


Figure 5.5: An HSQC spectrum of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 609 construct collected at pH 5.5. The red circle indicates the location of the strong amide proton peaks typical of flexible protein.

buffer used in the last step of the purification was 20 mM Na Acetate pH 5.5, 150 mM NaCl, and this buffer was also used as the crystallisation buffer. The 410 – 601 construct was subjected to sparse matrix crystallisation experiments (§ 2.7.1) at a concentration of 10 mg/mL, but once again no crystals were observed from these experiments.

5.3.1 HSQC experiment of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 601 construct

To ascertain whether the truncated 410 – 601 construct still contained areas of amide proton flexibility, an HSQC experiment was performed as described in § 2.8.2. Figure 5.6 shows an HSQC spectra of the *Mm*PepT2$^{\mathrm{ECD}}$ 410–601 construct. Compared to Figure 5.5, there is a reduction in the peak intensities in the 8 – 9 ^{1}H ppm, 116 – 126 ^{15}N ppm range. New peaks are also visible around the periphery of the spectrum at: 10 ^{1}H ppm–118 ^{15}N ppm and 7.4 ^{1}H ppm–102 ^{15}N ppm. The W596 indole ^{15}N peak appears to have shifted to 9.9 ^{1}H ppm–129 ^{15}N ppm and is now a singlet. These changes between the 410 – 609 and 410 – 601 spectra suggest that the flexible region was between residues 602 – 609, and that the 410 – 601 is more readily amenable to crystallisation.

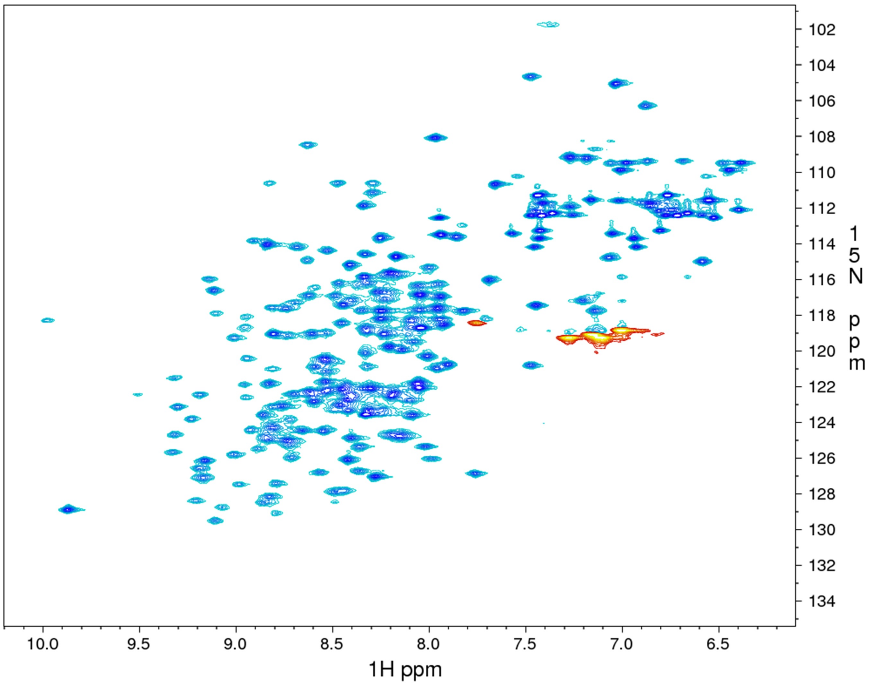


Figure 5.6: An HSQC spectrum of the *Mm*PepT2$^{\mathrm{ECD}}$ 410 – 601 construct collected at pH 5.5. The spectrum is more dispersed than the spectrum of the *Mm*PepT2$^{\mathrm{ECD}}$ 410–609 construct (see Figure 5.5) and the peak intensities in the backbone amide region appear less intense.

5.4 *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 crystallisation experiments

At this point, due to the lack of *Mm*PepT2$^{\mathrm{ECD}}$ crystals, more PepT2 genes were bought; *Rn*PepT2, *Xl*PepT2 and *Dr*PepT2 (Uniprot codes given in Table 2.1). The cloning and test expression of these genes in the pLou3 and pEHisTev vectors is described in § 3.5. The *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 construct could be stably expressed and purified using the same protocol as the *Mm*PepT2$^{\mathrm{ECD}}$ 410–601 (§ 3.5.3). The GF buffer used in the last step of the purification was 20 mM Na Acetate pH 5.5, 150 mM NaCl as this buffer increased the quality of the NMR spectrum. This buffer was also used as the crystallisation buffer.

5.4.1 Sparse matrix screening

The 410 – 601 construct was subjected to sparse matrix crystallisation experiments as described in § 2.7.1. The *Rn*PepT2$^{\mathrm{ECD}}$ 410–601 construct initially gave crystallisation hits in Tartrate and Citrate buffered conditions with 20% v/v PEG 3,350 in the matrix screens PEG/Ion, JCSG+, and Wizard$^{\mathrm{TM}}$ III+IV at 20 °C (Figure 5.7A). A list of the conditions

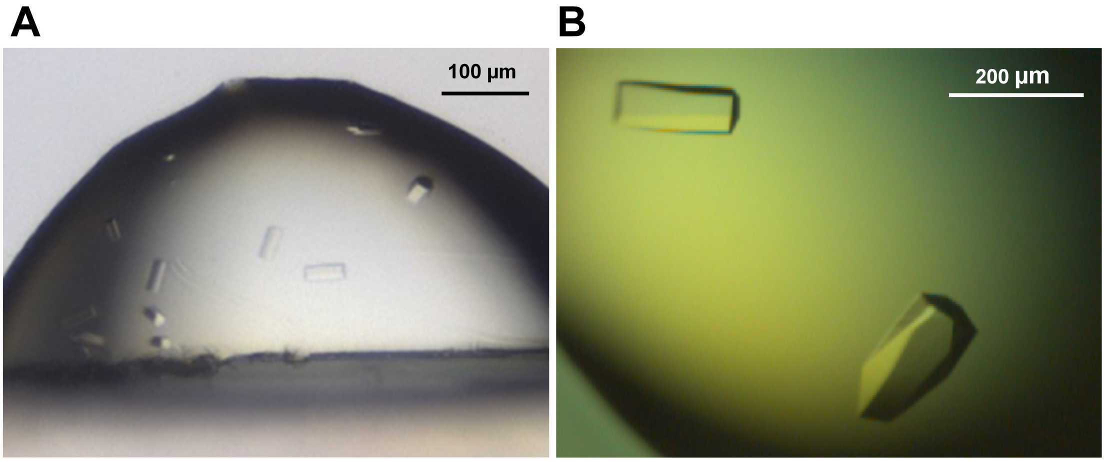


Figure 5.7: *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 crystals. (A) Initial crystal hit grown from a 96-well Wizard$^{\mathrm{TM}}$ III+IV screen, condition 1 (20% v/v PEG 3,350, 0.2 M $(NH_4)_2$ Citrate pH 5.8). **(B)** Optimised *Rn*PepT2$^{\mathrm{ECD}}$ crystals, grown in a 24-well hanging drop plate under the conditions.

which gave crystals is shown in Table 5.2. Crystal nucleation was generally observed after 3 hr and growth continued for approximately 4 days. Final crystal dimensions were approximately 15.0 × 15.0 × 75.0 µm with a hexagonal prism morphology.

5.4.2 Crystal optimisation

The crystal hits displayed in Table 5.2 could be repeated in the 96-well sparse matrix screens. To increase the size of the crystals, optimisation experiments were performed in 24-well hanging drop plates (described in § 2.7.2). The crystal hits outlined in Table 5.2 indicated that 20% v/v PEG 3,350 was important for crystallisation. However, it was unclear which Citrate or Tartrate salt, if any, were better for growing large crystals. Crystal optimisation (discussed in § 4.2.2) was performed in 24-well hanging drop plates with 6 different buffers; K_3 Citrate pH 8.3, Na_3 Citrate pH 8.3, Na_2 Tartrate pH 7.3, $(NH_4)_2$ Tartrate pH 6.6, $(NH_4)_2$ Citrate pH 5.1 or $(NH_4)_3$ Citrate pH 5.8, at either 0.1 or 0.2 M. The buffer which gave the clearest drop with the largest crystals (100.0×100.0×200.0 µm) was 0.2 M $(NH_4)_3$ Citrate pH 5.8 at either 20 or 21% v/v PEG 3,350. The crystals (Figure 5.7B) from these two conditions were used for the Native, Se-SAD and Se-pk datasets (Table 5.3).

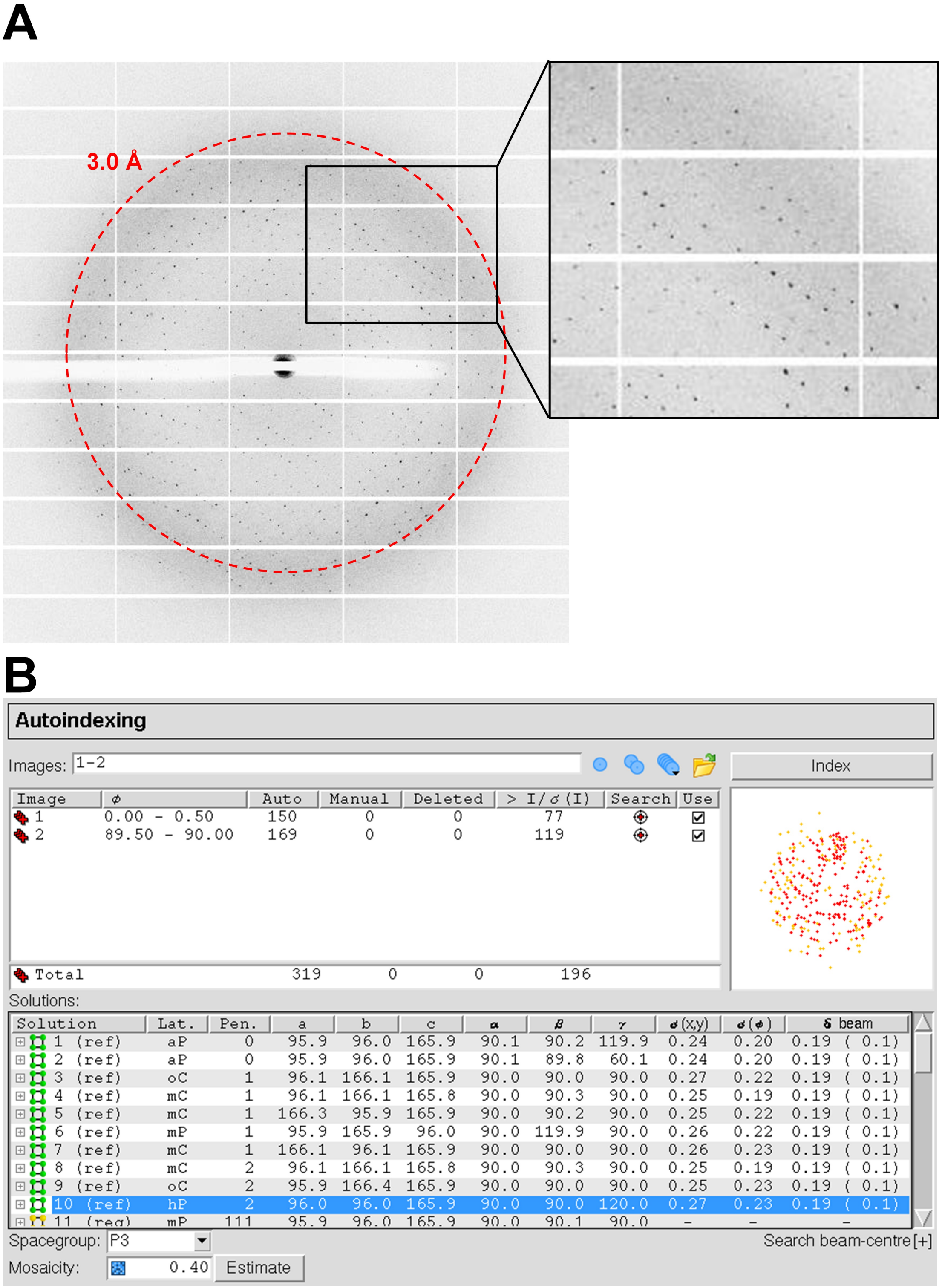
Screen	**Well**	**Condition**	**20 °C**	**4 °C**
PEG/Ion	C12	0.2 M Na_2Tartrate pH 7.3 20% v/v PEG 3,350	✓	
	D1	0.2 M KNaTartrate pH 7.4 20% v/v PEG 3,350	✓	
	D2	0.2 M $(NH_4)_2$Tartrate pH 6.6 20% v/v PEG 3,350	✓	
	D10	0.2 M Na_3Citrate pH 8.3 20% v/v PEG 3,350	✓	✓
	D11	0.2 M K_3Citrate pH 8.3 20% v/v PEG 3,350	✓	✓
	D12	0.2 M $(NH_4)_2$Citrate pH 5.1 20% v/v PEG 3,350	✓	
JCSG+	A3	0.2 M $(NH_4)_2$Citrate pH 5.0 20% v/v PEG 3,350	✓	
Wizard^{TM} III+IV	E10	0.2 M $(NH_4)_2$Citrate 20% v/v PEG 3,350	✓	

Table 5.2: A table showing the conditions from sparse matrix screens which gave *Rn*PepT2^{ECD} 410 – 601 crystallisation hits.

5.5 Data collection

Diffraction data was collected at the European Synchrotron Radiation Facility (ESRF) beamline ID23-2 and Diamond Light Source beamlines I04 and I03. Crystals were cryoprotected in mother liquor containing 25% v/v glycerol. The native *Rn*PepT2^{ECD} 410–601 crystals were screened at the ESRF and the Native dataset was collected. Figure 5.8A shows a diffraction image of the *Rn*PepT2^{ECD} crystal. `iMosflm` (Powell, Johnson, and Leslie 2013) and `EDNA` (Incardona *et al.* 2009) were used to calculate the data-collection strategy (described in § 4.3.3). Indexing suggested that the crystal had an trigonal primitive Bravais lattice with unit cell lengths a, b, and c, of 96.0, 96.0 and 165.9 Å respectively (Figure 5.8B).

The data collection parameters of the three datasets used to solve the *Rn*PepT2^{ECD} 410 – 601 X-ray crystal structure are given in Table 5.3. The Se-SAD dataset was used to solve the phase problem and the Se-pk dataset was the highest resolution trigonal dataset.



Autoindexing

Images: 1-2

Image	φ	Auto	Manual	Deleted	> I/σ(I)	Search	Use
1	0.00 - 0.50	150	0	0	77	⊕	☑
2	89.50 - 90.00	169	0	0	119	⊕	☑
Total		319	0	0	196		

Solutions:

Solution	Lat.	Pen.	a	b	c	α	β	γ	σ(x,y)	σ(φ)	δ beam
1 (ref)	aP	0	95.9	96.0	165.9	90.1	90.2	119.9	0.24	0.20	0.19 (0.1)
2 (ref)	aP	0	95.9	96.0	165.9	90.1	89.8	60.1	0.24	0.20	0.19 (0.1)
3 (ref)	oC	1	96.1	166.1	165.9	90.0	90.0	90.0	0.27	0.22	0.19 (0.1)
4 (ref)	mC	1	96.1	166.1	165.8	90.0	90.3	90.0	0.25	0.19	0.19 (0.1)
5 (ref)	mC	1	166.3	95.9	165.9	90.0	90.2	90.0	0.25	0.22	0.19 (0.1)
6 (ref)	mP	1	95.9	165.9	96.0	90.0	119.9	90.0	0.26	0.22	0.19 (0.1)
7 (ref)	mC	1	166.1	96.1	165.9	90.0	90.0	90.0	0.26	0.23	0.19 (0.1)
8 (ref)	mC	2	96.1	166.1	165.8	90.0	90.3	90.0	0.25	0.19	0.19 (0.1)
9 (ref)	oC	2	95.9	166.4	165.9	90.0	90.0	90.0	0.25	0.23	0.19 (0.1)
10 (ref)	hP	2	96.0	96.0	165.9	90.0	90.0	120.0	0.27	0.23	0.19 (0.1)
11 (reg)	mP	111	95.9	96.0	165.9	90.0	90.1	90.0	-	-	-

Spacegroup: P3

Search beam-centre [+]

Mosaicity: 0.40

Figure 5.8: Diffraction pattern and beamline processing of a *Rn*PepT2$^{\text{ECD}}$ 410 – 601 crystal. (A) An image of the *Rn*PepT2$^{\text{ECD}}$ crystal diffraction. **(B)** A screen shot of `iMosflm` (Powell, Johnson, and Leslie 2013) processing two images taken 90 ° a part with a 0.5 ° oscillation. The solution with highest symmetry and lowest penalty has been highlighted.

	Native	**Se-SAD**	**Se-pk**
Date	15/03/2012	06/05/2013	06/05/2013
Beamline	ID23-2	I04	I04
λ (Å)	0.8729	0.97942	0.97949
Temperature (°C)		−173	
ϕ (°)	0–180	81.4–531.4	135–315
$\Delta\phi$ (°)	0.5	0.15	0.15
Exposure time (s)	0.2	0.07	0.07
Transmission (%)	50	50	40
Detector Distance (mm)	345.200	536.439	545.151
Detector	MarMosaic 225	Pilatus 6M-F	Pilatus 6M-F

Table 5.3: Data collection parameters for the trigonal *Rn*PepT2$^{\mathbf{ECD}}$ 410 – 601 datasets.

5.6 Data reduction

5.6.1 Trigonal crystals

As it turned out, the Native crystals did not diffract as well as the SeMet derivatised *Rn*PepT2$^{\text{ECD}}$ 410 – 601 crystals. Therefore the Native dataset was only used in the initial processing to gain information about the unit cell, so it is only included here for completeness. The Native trigonal dataset was indexed and integrated in `iMosflm` (Powell, Johnson, and Leslie 2013), and scaled and merged in `Aimless` (Evans and Murshudov 2013). `Pointless` (Evans 2006) was used to estimate the space-group probability. Figure 5.9 shows the $I/\sigma(I)$ plotted against the c-axis indices. Regular absences for the 1st and 2nd indices along the c-axis suggests the likely presence of a 3-fold screw axis. The correct space-group is therefore likely to be either $P3_121$ or $P3_221$, so the trigonal crystals were merged into space-group $P3_121$ until the phases were calculated.

The data processing statistics are shown in Table 5.4. The Native dataset integrated poorly. The detector distance over the course of the dataset showed variation of greater than 2 mm. The resolution of the dataset was cut back to 3.50 Å to lower the R_{merge} and R_{meas}.

`xia2 -3daii` was used to process the Se-SAD dataset utilising `Distl`, `Labelit`, `XDS`, `Pointless`, `Aimless` and the CCP4 suite (Zhang *et al.* 2006; Sauter, Grosse-Kunstleve, and Adams 2004; Kabsch 2010; Evans 2006; Evans and Murshudov 2013; Winn *et al.* 2011). Se-SAD was used to calculate the phases by Se SAD method, and this determined

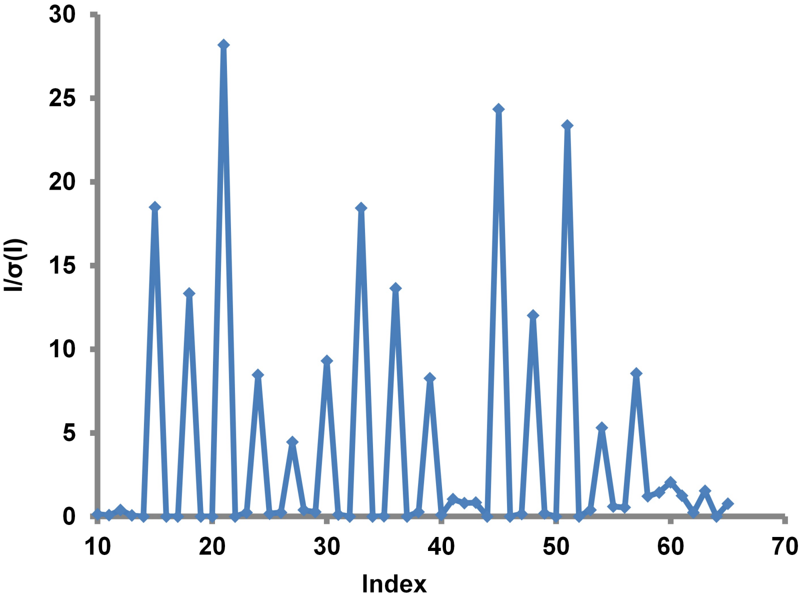


Figure 5.9: An $I/\sigma(I)$ against the c axis plot of the P3 *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 intensities. The figure was generated by `Pointless` (Evans 2006). A 3-fold skew axis would be broadly expected to show strong intensities every third index.

that the trigonal crystals were in space-group P$3_2$21. The Se-pk dataset was processed in this space-group using `XDS` and `Aimless` and was the highest resolution dataset for the trigonal crystal form.

		Native	**Se-SAD**	**Se-pk**
Unit Cell	a, b, c (Å)	95.8, 95.8, 165.4	96.03, 96.03, 166.60	95.75, 95.75, 163.93
Space-group $P3_221$	α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution Range		59.29 – 3.50 (3.83 – 3.50)	33.29 – 2.92 (3.00 – 2.92)	47.87 – 2.84 (3.00 – 2.84)
Measured Reflections		125716 (30133)	451479 (18569)	210301 (30663)
Unique Reflections		11967 (2807)	19910 (1438)	21355 (3090)
Multiplicity		10.5 (10.7)	22.7 (12.9)	9.8 (9.9)
Mosaicity (°)		0.390	0.116	0.090
Overall $\langle \lvert I \rvert / \sigma(I) \rangle$		7.2 (3.8)	21.0 (3.0)	13.2 (2.4)
Overall Completeness (%)		100.0 (100.0)	100.0 (100.0)	100.0 (99.8)
R_{merge}		0.400 (1.140)	0.127 (0.906)	0.130 (0.943)
R_{meas}		0.420 (1.198)	0.133 (0.981)	0.145 (1.049)
$\mathrm{CC}^{1}/_{2}$		0.980 (0.846)	0.998 (0.900)	0.998 (0.722)

Table 5.4: Data processing statistics of the *Rn*PepT2$^{\mathbf{ECD}}$ trigonal datasets. The values in parentheses are the highest resolution shell and complete crystallographic tables are shown in Table A.2.

Mol/ASU	V_m (Å^3/Da)	Solvent (%)	P|2.80 Å
1	10.27	88.03	0.00
2	5.13	76.05	0.00
3	3.42	64.08	0.10
4	2.57	52.11	0.53
5	2.05	40.14	0.35
6	1.71	26.16	0.01
7	1.47	16.19	0.00
8	1.28	4.22	0.00

Table 5.5: A table showing the V_m probability distribution for the P$3_2$21 Native dataset. Generated by the CCP4 program `Matthews_coef` (Winn *et al.* 2011).

5.7 Asymmetric unit analysis

Table 5.5 shows the V_m coefficients calculated per number of molecules in the ASU by the CCP4 program `Matthews_coef` (Winn *et al.* 2011). The $V_m = 2.57$ with a solvent content of 52.11% which indicates that there are likely to be four molecules in the ASU with a probability of 0.53. The $\chi = 180.0°$ SRF function (Figure 5.10) only shows the 3 strong peaks corresponding to the 3-fold axis. There were no other strong peaks in either the $\chi = 180.0°$ or $60.0°$ plots suggesting that the contents of the ASU are not related by a single NCS axis.

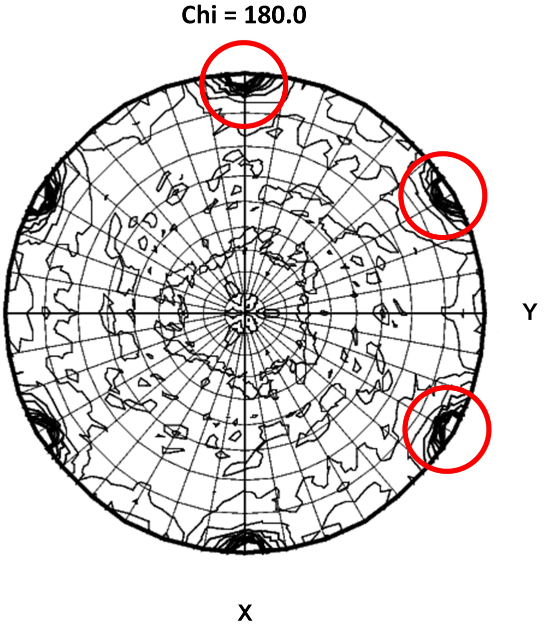


Figure 5.10: The SRF of a *Rn*PepT2^{ECD} trigonal crystal. This image was generated in `MOLREP` (Vagin and Teplyakov 1997). The red circles indicate the self rotation peaks which are consistent with a trigonal/hexagonal unit cell.

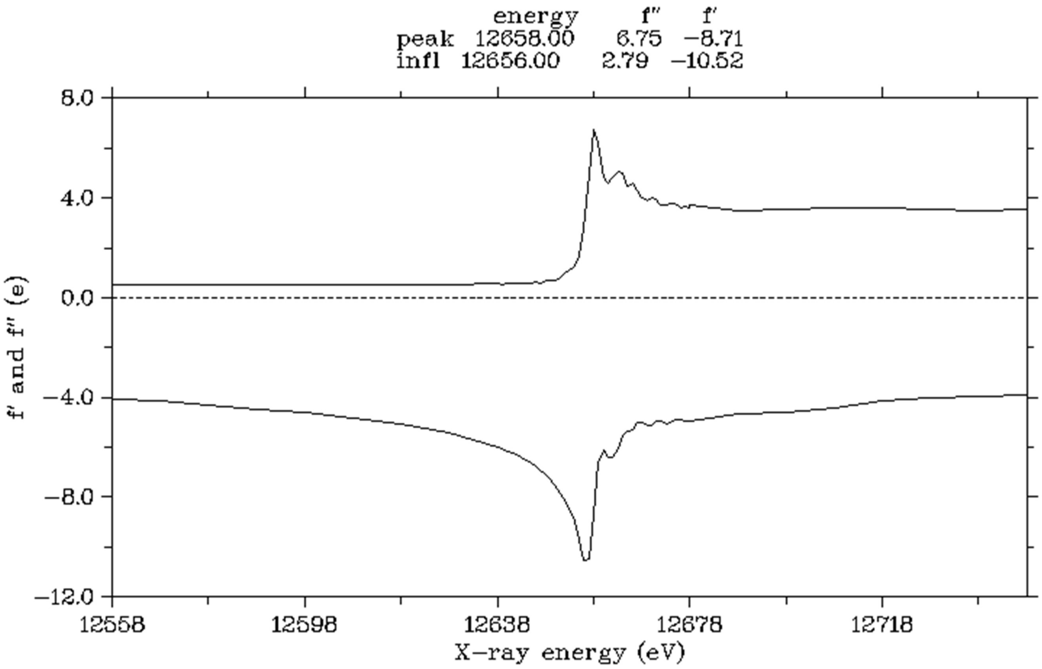


Figure 5.11: The peak wavelength dependent factors f' and f'' calculated by `Chooch` (Evans and Pettifer 2001) from a fluorescence scan collected at Diamond Light Source I04 over the Se absorptive edge on a SeMet derivatised *Rn*PepT2$^{\mathrm{ECD}}$ crystal.

5.8 Experimental phasing of the trigonal crystal

5.8.1 SeMet SAD phasing and model building

SeMet derivatised *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 was produced using an auto-inducing SeMet medium PASM-5052 as described in § 2.7.4. The derivatised protein was purified in the same manner as WT (§ 3.5.3). SeMet 410 – 601 was crystallised in 24-well hanging drop plates as described in § 5.4.2. Crystals were cryo-protected in mother liquor laced with 25% v/v glycerol and flashed cooled in liquid nitrogen. To assess the presence of Se, a fluorescence scan of the trigonal crystals was performed. Figure 5.11 shows the f' and f'' calculated by `Chooch` (Evans and Pettifer 2001) from a fluorescence scan of a 410 – 601 SeMet derivatised trigonal crystal. The scan gives a peak f' and f'' of −8.71 and 6.75 respectively and confirms the presence of Se.

The phases and a partial model of the ASU were calculated in `PHENIX autoSOL` (Terwilliger *et al.* 2009), an automated structure solution program utilising `HySS`, `Phaser`, `RESOLVE` and `Auto-Build` (Grosse-Kunstleve and Adams 2003; McCoy, Storoni, and

Atom	x	y	z	**Occupancy**	
1	-7.574	53.683	159.15	1.00	✓
2	-23.230	57.858	97.414	0.92	✓
3	-11.633	45.338	118.739	0.91	✓
4	-2.832	44.031	65.196	0.65	✓
5	-26.960	45.660	4.056	0.49	
6	0.544	50.314	138.982	0.47	
7	-23.534	57.337	115.890	0.20	
8	-6.159	24.784	117.445	0.19	

Table 5.6: HA positions found by `HySS` (Grosse-Kunstleve and Adams 2003). The top four positions have clearly defined difference density.

Read 2004; Terwilliger 2000; Terwilliger 2004; Terwilliger *et al.* 2008; Terwilliger *et al.* 2009; Read and McCoy 2011; Adams *et al.* 2010; Afonine *et al.* 2010). The Se-SAD .mtz with the *Rn*PepT2$^{\mathrm{ECD}}$ 410 – 601 sequence was input into `autoSOL`. Given the $V_m = 2.57$, 4 monomers were assumed to occupy the ASU with a solvent content of 52%.

`HySS` (Grosse-Kunstleve and Adams 2003) is a dual-space HA substructure solution and refinement program, similar to `SHELXD` (Usón and Sheldrick 1999) and `SnB` (Miller *et al.* 1994) (described in § 4.6.3). `HySS` differs from `SHELXD` and `SnB` by using the Fast translation function (FTF) (Navaza and Vernoslova 1995) to refine the 2 atom Patterson seeds. After real space refinement, the reciprocal space recycling is done in P1, and the FTF combines the resultant seeds back into the correct space-group. `HySS` is not as robust as `SHELXD` (Grosse-Kunstleve and Adams 2003). However, it is utilised by the `PHENIX` package and therefore is used by `PHENIX autoSOL`.

Table 5.6 shows the HA positions calculated by `HySS` (Grosse-Kunstleve and Adams 2003) with a CC of 0.570 which suggests a good solution. Well defined difference density is observed around the atoms 1 – 4. These positions were taken forward for $\mathbf{F}_{hkl}$ calculation and refinement by `Phaser`.

`Phaser` (McCoy, Storoni, and Read 2004) is a ML based phase refinement program which produces similar results to `SHARP` (Bricogne *et al.* 2003). The handedness of the solutions is determined by density modification in `RESOLVE`.

`RESOLVE` (Terwilliger 2000) is a ML reciprocal space solvent flattening program. The function should converge on a solution which has clearly defined regions of protein and solvent. `RESOLVE` also performs histogram matching (§ 4.6.4) and searches for NCS present in the ASU. The FOM for the $P3_221$ solution was 0.320 with an overall CC of 42.09 ± 11.10 and confirmed the space-group was $P3_221$. After density modification, a positive histogram skew of 0.13 and R-factor of 0.263 were calculated, suggesting a good solution with a reasonable map. The map showed good solvent flattening, and `Phenix Auto-Build` (Terwilliger *et al.* 2008) was able to build a disconnected Cα trace into the majority of the density with some correct residue assignment. The final R_{work} and R_{free} were 0.334 and 0.4095 respectively.

Examination of the maps and the `RESOLVE` log file indicated that solvent content was approximately 60% and that only 3 monomers occupied the ASU. The partial solution produced by `Phenix Auto-Build` was used as a template. The complete 410 – 601 sequence was built for one monomer in `Coot` (Emsley and Cowtan 2004). This complete thread was aligned onto the partial Cα-trace present for the other two monomers from the auto-build using `SSM` (Krissinel and Henrick 2004).

Examining the HA positions calculated by `AutoSOL`, it was clear that the N-terminal SeMet was giving coherent scattering (Figure 5.12A) and the SeMet in the third monomer had two anomalous peaks for the Se atom above 5.0 σ (Figure 5.12B). The 10 Se sites from the complete ASU model (Table 5.7) were input into `SHARP` to improve the phases. The phasing statistics are shown in Table 5.8. The complete model was used as a solvent mask to improve the solvent flattening. The map generated from the improved phases is shown in Figure 5.13.

5.9 Refinement

The refinement of the structure was initially done in `Buster` (Blanc *et al.* 2004) and finally in `Phenix.refine` (Afonine *et al.* 2010) against the higher resolution dataset Se-pk. After an initial round of refinement, protons were added to aid side chain positioning. Inspection of the monomeric *Rn*PepT2$^{\text{ECD}}$ showed it appeared to have a similar structure to PepT1 (Figure 5.14); two lobes connected by a flexible linker. The lobes were refined as TLS domains, and NCS was assumed between the monomers. Three areas of density were not well resolved but of reasonable size to suggest a non-protein compound. One was assigned as glycerol and the other two as citrate (Figure 5.14B and C). The molecules

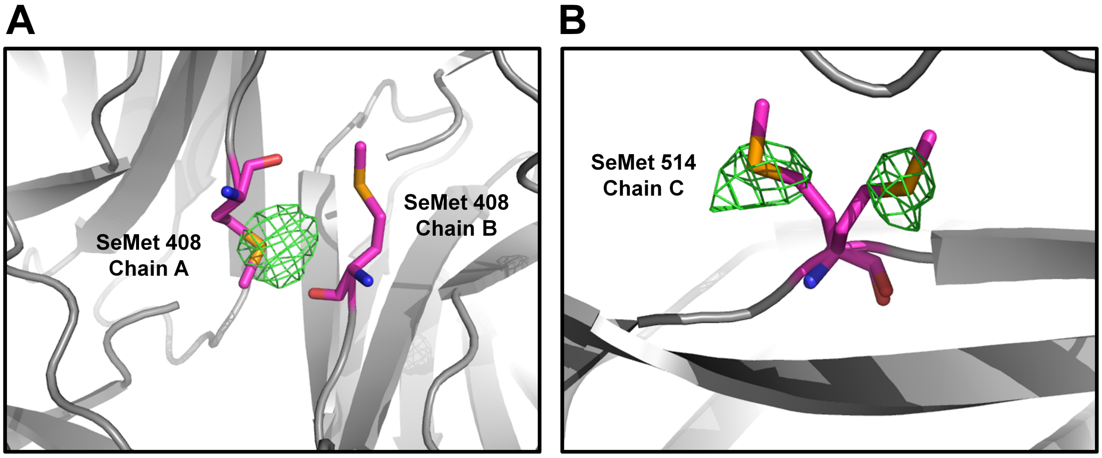


Figure 5.12: Examples of the Se density in the *Rn*PepT2^{ECD} P$3_2$21 crystal. The model SeMet built into the anomalous density calculated by `Phenic AutoSOl` (Terwilliger *et al.* 2009). **(A)** The N-terminal SeMet from chains A and B. **(B)** Chain C SeMet501 in 2 conformations.

Bin	D_{min}	D_{max}	φ–power	R_{cullis}	FOM
1	33.28	17.29	2.516	0.421	0.63745
2	17.29	13.14	2.765	0.415	0.62034
3	13.14	11.02	2.907	0.393	0.59047
4	11.02	9.68	2.569	0.464	0.52912
5	9.68	8.73	2.108	0.397	0.56320
6	8.73	8.02	2.314	0.404	0.56525
7	8.02	7.45	2.477	0.371	0.58475
8	7.45	6.99	2.058	0.419	0.56267
9	6.99	6.61	2.342	0.380	0.55005
10	6.61	6.28	2.074	0.417	0.56567
11	6.28	6.00	2.880	0.451	0.54417
12	6.00	5.75	2.685	0.516	0.50405
13	5.75	5.53	2.657	0.501	0.49453
14	5.53	5.34	2.484	0.547	0.49448
15	5.34	5.16	2.304	0.565	0.47714
16	5.16	5.00	2.344	0.557	0.47720

Table 5.7: The Se-SAD dataset acentric phasing statistics. The table was generated by `SHARP` (Bricogne *et al.* 2003). The φ–power, R_{cullis} and FOM are discussed in § 4.6.3.

Chain	Residue	x	y	z	Observed
A	408	4.467	58.090	54.294	✓
	501	23.167	58.010	69.181	✓
	514	-2.507	62.660	75.364	✓
B	408	-1.060	51.093	106.189	
	501	11.790	45.336	90.756	✓
	514	-11.419	58.672	92.969	✓
C	408	54.851	41.979	76.847	
	501	66.386	60.166	59.446	✓
	501	66.587	56.342	62.659	✓
	514	75.040	37.113	78.518	✓

Table 5.8: A list of the SeMet sites built into the density calculated by `AutoSOL` (Terwilliger *et al.* 2009). These site were input into `SHARP` (Bricogne *et al.* 2003) to improve the phases. The '✓' indicates whether anomalous density was observed from the refined phases.

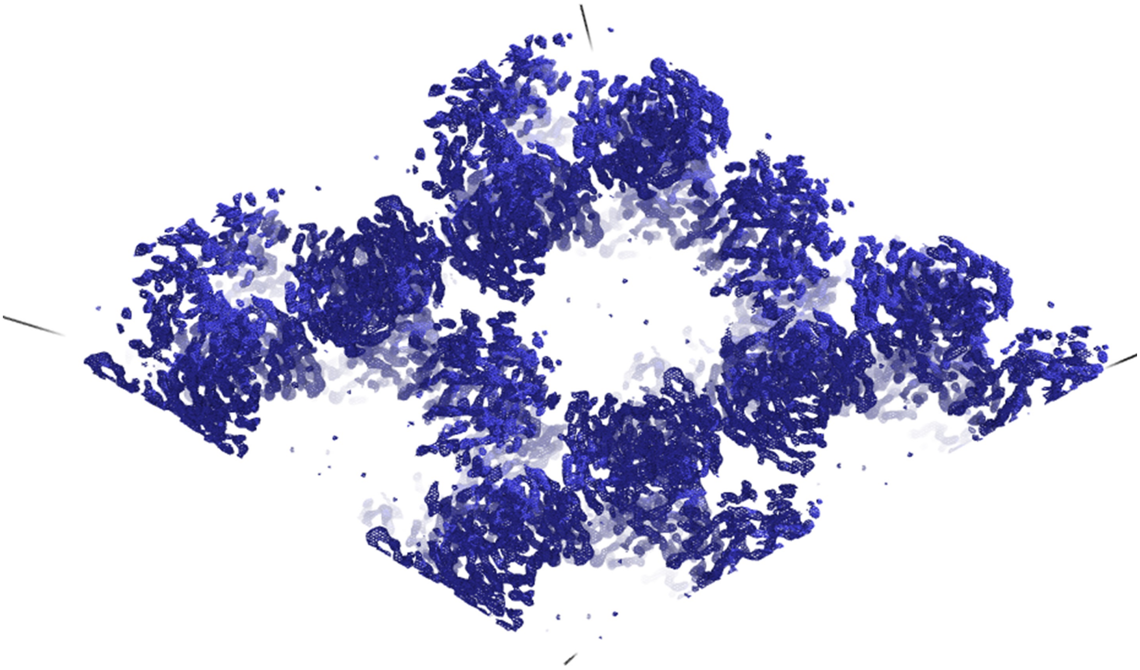


Figure 5.13: An electron density map of the trigonal *Rn*PepT2$^{\mathrm{ECD}}$ ASU generated from the Se-SAD $|\mathrm{F}_{hkl}|$ and `SHARP` refined φ_{hkl}.

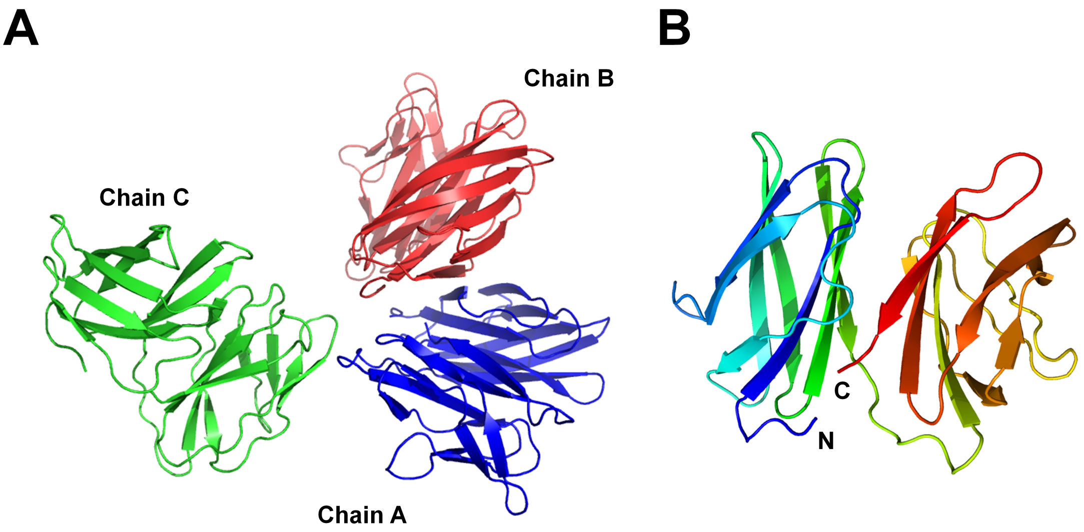


Figure 5.14: The *Rn*PepT2$^{\text{ECD}}$ trigonal crystal structure. **(A)** The complete model ASU. There are 3 monomers in the asymmetric unit coloured, red, blue and green. **(B)** The monomeric ECD structure, coloured in rainbow from the N- to the C-terminus.

fit the density and were refined. Water molecules were added and manually pruned, and `Molprobity` (Chen *et al.* 2010) was used to aid refinement and validate the structure. Table 5.9 shows the refinement statistics for the crystal structure.

5.10 Higher resolution *Rn*PepT2$^{\text{ECD}}$ 410 – 601 structure

The P$3_2$21 *Rn*PepT2$^{\text{ECD}}$ crystals resulted in a model of the ECD. However, the highest resolution trigonal crystal was 2.84 Å. To improve this resolution sparse matrix micro-seeding was performed.

5.10.1 Sparse matrix micro-seeding

Sparse matrix micro-seeding is an established method of crystal optimisation (Ireton and Stoddard 2004; D'Arcy, Villard, and Marsh 2007). The sparse matrix micro-seeding experiment works as an additive screen using a sparse matrix. Micro-seeds fall randomly on the crystallographic phase diagram (Figure 5.15), by-passing the nucleation zone. This allows crystal growth in conditions where the lack of nucleation would otherwise prevent crystals forming. To create the micro-seeds, the seed bead protocol from Hampton Research was followed (§ 2.7.5).

		$P3_221$
R_{work}		0.204
R_{free}		0.258
Refined Residues		576
Refined H_2O		93
B factor ($Å^2$)		65.1
RMSD Bond Length (Å)		0.00
RMSD Bond Angle (°)		0.75
Ramachandran Plot (%)	Favoured	94.1
	Allowed	5.89
	Disallowed	0.00
Rotamer Outliers (%)		2.14
Molprobity (Percentile)	Clashscore	100
	Score	99

Table 5.9: Refinement statistics for the *Rn*PepT2^{ECD} model from the trigonal crystal.

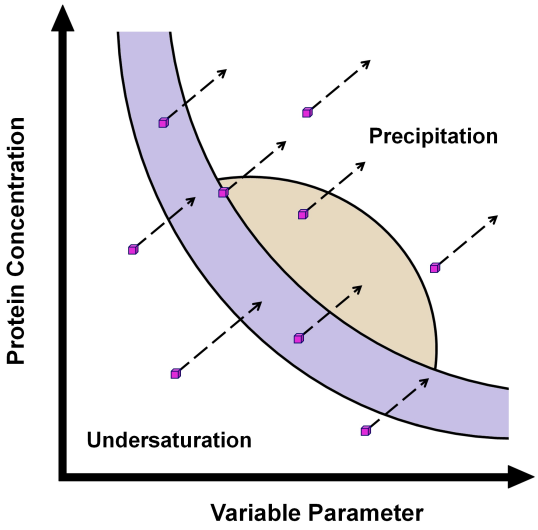


Figure 5.15: Sparse matrix seed phase diagram. The beige area denotes the crystal nucleation zone. The purple region is the metastable zone where growth is likely to occur. The purple crystals are seeds which are randomly dispersed. The phase diagram adapted from Chayen and Saridakis (2008).

Sparse matrix micro seeding was performed using the crystals grown from the $(NH_4)_3$ Citrate condition. PEG/Ion and JCSG+ were used as sparse-matrix screens. Hits were observed in the original conditions and in condition H9 from PEG/Ion (0.2 M $CsCl_2$, 15% v/v PEG 3,350). A crystal from this condition was used in collection of the Seed dataset (Table A.2).

5.10.2 Data collection and reduction

The Seed dataset was collected at Diamond Light Source beamline I03. Indexing of the Seed crystal dataset indicated that it had a different Bravais lattice and therefore space-group to the original trigonal *Rn*PepT2$^{\mathrm{ECD}}$ crystals. `iMosflm` (Powell, Johnson, and Leslie 2013) calculated the likely unit cell parameters, a, b, and c as 43.9, 43.9, 220.6 Å respectively in a tetragonal primitive lattice. The Seed tetragonal dataset was initially processed using `xia2 -3daii` in space-group $P4_12_12$ to a resolution of 2.13 Å. Once the phases were calculated by MR of the monomeric $P3_221$ structure (discussed in § 5.10.3), the dataset was reprocessed in `XDS` (Kabsch 2010) and `Aimless` (Evans and Murshudov 2013), to a resolution of 2.06 Å (Table A.2).

5.10.3 Molecular replacement

As stated in § 4.6.1, MR has become the standard phasing tool for macromolecular crystallography. The technique was first proposed by Rossmann and Blow (1962). Due to the conservation of protein folds, phase information from a known protein's $\mathbf{F}_{hkl}$ can be used to solve the phases of a homologous protein's $\mathbf{F}_{hkl}$. Assuming a similar structure is known, the task during MR is to calculate the NCS relating the contents of the unknown ASU (Rossmann 2001). The incorporation of ML methods into the search function has been instrumental in the success of the technique (Read 2001).

In `Phaser` (McCoy 2002; McCoy *et al.* 2007), MR is performed in 5 steps. The number of NCS related copies of the search model in the ASU is estimated using the V_{m}. Anisotropy present in the dataset is estimated using the Wilson plot (discussed in § 4.4.3) and minimised. A FTF of the search model is performed and the target function is refined. Throughout this process, the Log-likelihood gain (LLG) and Z-scores are maximised to converge on the correct solution.

The phasing of the $P4_12_12$ crystal was accomplished in `Phaser-MR` (McCoy *et al.* 2007) using the monomeric structure of the ECD from the $P3_221$ crystal. A Matthews coefficient of 2.37 Å^3/Da was calculated for a single monomer in this space-group at a solvent content of 48.15%. Both enantiomorphic space-groups ($P4_12_12$ and $P4_32_12$) were included in the search function. The top (and only) solution, with a refined LLG of 898 and a Translation function Z-score (TFZ) of 30.3, was in space-group $P4_12_12$ and given the LLG and TFZ, had a high probability of being correct.

5.10.4 Model building and refinement

`Phenix Auto-Build` (Terwilliger *et al.* 2008) was run to try to minimize any potential model errors introduced by the search model. The auto built model was refined in `Phenix.refine` (Afonine *et al.* 2010). Inspecting the density of the crystals in `Coot` (Emsley and Cowtan 2004), two difference map peaks were visible above 5 σ. An anomalous difference map was calculated using `FFT` (Winn *et al.* 2011), and two peaks had a σ of 11.23 and 8.81. This suggested that these two sites were occupied by a HA. $CsCl_2$ had been present in the seeding condition; therefore these sites were modelled as Cs coordination sites. The 8.81 site lay almost directly on the 2-fold axis and was coordinated by H451, E441 and their symmetry related mates (Figure 5.16). The 8.81 σ Cs atom and H451 had to be modelled in 2 conformations to allow for the Cs atom to be coordinated by either E441 or the 2-fold related E441. After the auto build the model refinement was done in `Buster` (Blanc *et al.* 2004) and finally in `Phenix.refine` to regularise the structure. The lobes were assigned as TLS domains and no density was observed for residues 589 – 592, so they were omitted from the structure. Protons were added to refine side-chain clashes, and the Ramachandran plot (Figure 5.17) was used to aid Phi/Psi angle refinement. Waters were added and manually pruned, and `Molprobity` (Chen *et al.* 2010) was used to validate the structure (Table A.2).

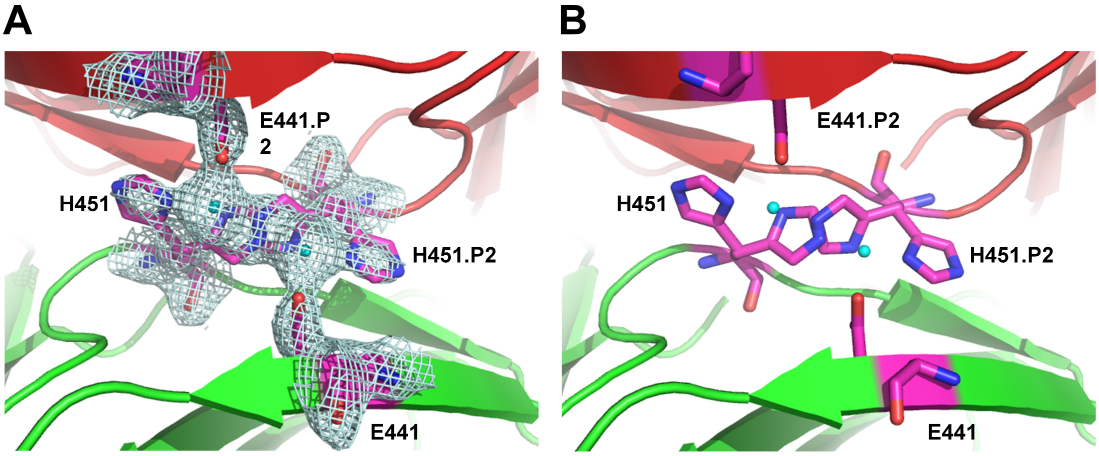


Figure 5.16: The dual histidine co-ordination of the Cs atom at the 2-fold crystallographic axis. (A) and **(B)** shows the region with and without $2F_0$-F_C density respectively. The suffix .P2 indicates a 2-fold relation.

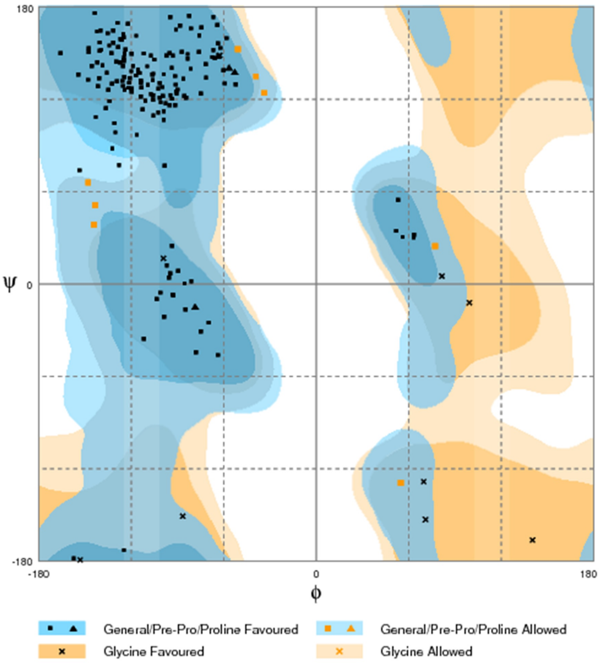


Figure 5.17: A Ramachandran plot of the $P4_12_12$ 410 – 601 ASU calculated using `Rampage` (Lovell *et al.* 2003).

5.11 The model of *Rn*PepT2$^{\mathrm{ECD}}$

Figure 5.18A shows the topology and connectivity of the *Rn*PepT2$^{\mathrm{ECD}}$. The ECD structure can be divided into two sub-domains connected by a flexible linker. These sub-domains are composed of a 4-sided β-sandwich fold like PepT1$^{\mathrm{ECD}}$. The P$3_2$21 crystal had three monomers in the ASU. Two monomers (A and B) are related by a 2-fold axis. The density for these monomers is relatively good. Monomer C is related to A and B through rotational and translational axes. Figure 5.18B shows an `SSM` alignment (Krissinel and Henrick 2004) of the 3 monomers with a Cα RMSD of ≈ 1.5 Å. The main source of variation between the monomers is the O-P loop which also has the highest B-factors associated with the structure (Figure 5.18C). It is this loop (residues 589 – 592) which could not be built in the P$4_1 2_1$2 crystal. The linker region between the lobes also shows variation. The density for monomer C is very poor in this linker with the β-sheets having no R-group definition.

The P$4_1 2_1$2 crystal structure does not show significant variation from any of the chains in the P$3_2$21 crystal, with a Cα RMSD of ≈ 1.0 Å. Figure 5.19 highlights the Cα differences between the P$4_1 2_1$2 ASU and chain A from the P$3_2$21 ASU. Variation is observed in loops C-D, J-K, M2-N1 and O-P, and the linker between the lobes. The reason for these variations is that although the B-factors indicate some of the loops have flexibility, they are engaged in different crystal contacts between the different crystals.

5.12 Discussion

The lack of crystallisation of the *Mm*PepT2$^{\mathrm{ECD}}$ in either the 410–609 or 410–601 construct was the major time limiting factor of this project. In hindsight, the flexibility of the C-terminus should have been discovered sooner. Although, the 410 – 609 was clearly folded and stable, with a T_m of 50.8 °C at pH 6.5, the HSQC of the 410 – 609 (Figure 5.5) suggests a very flexible region of the protein. Since the ECD is less than 25 kDa and ^{15}N labelling was straightforward in *E. coli*, an assessment of the ECD by NMR should have been performed earlier.

The discovery that the flexible region present in the 410 – 609 construct was located in the

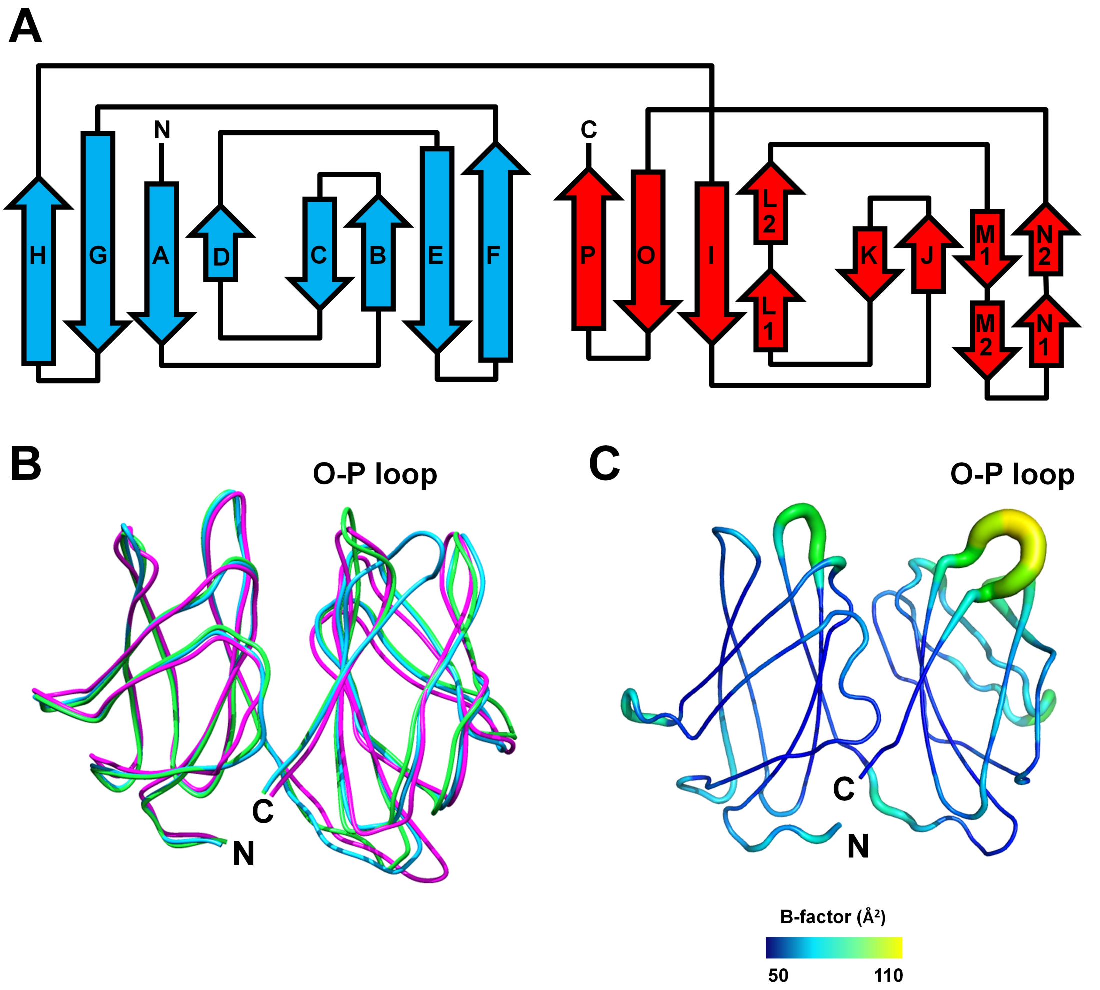


Figure 5.18: Analysis of the *Rn*PepT2$^{\text{ECD}}$ P$3_2$21 crystal crystal structure. **(A)** A topology diagram of *Rn*PepT2$^{\text{ECD}}$. **(B)** A Cα alignment of the 3 monomers in the asymmetric unit calculated by `SSM` (Krissinel and Henrick 2004). **(C)** Monomer A coloured and structured by B-factor.

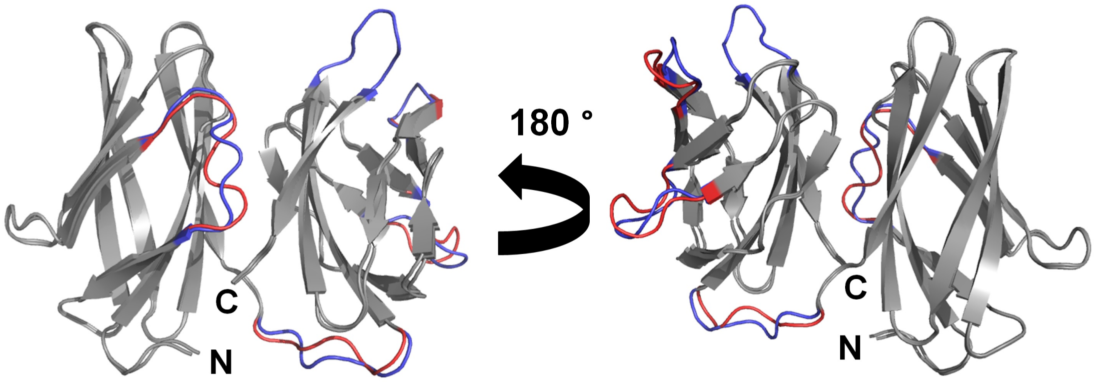


Figure 5.19: A Cα alignment of the P$4_1 2_1$2 ASU with chain A from the P$3_2$21 ASU. The loops which show different orientations between the P$4_1 2_1$2 ASU and P$3_2$21 chain A have been highlighted in red and blue respectively. The alignment was calculated by `SSM` (Krissinel and Henrick 2004).

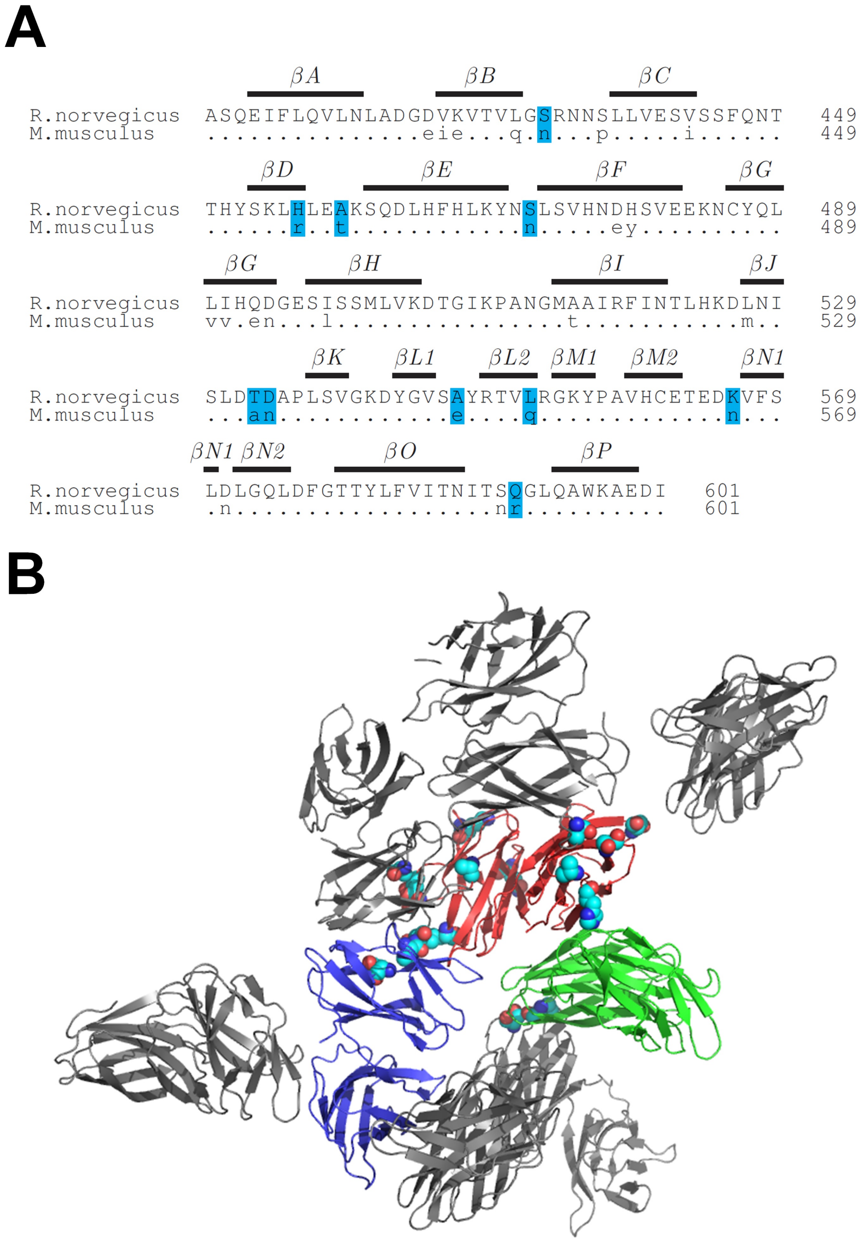
C-terminus was fortunate. A more thorough investigation of the fragments produced in a cleavage reaction (Figure 5.4) by mass-spectrometry would no doubt have indicated the C-terminus was removed by limited proteolysis and therefore likely to be flexible. However, as it was, the *Mm*PepT2 410 – 601 still did not crystallise.

Examining the crystal contacts present in the $P3_221$ ASU, it is possible to understand why the *Mm*PepT2 410 – 601 did not crystallise. Figure 5.20A shows an alignment between the the *Mm*PepT2 and *Rn*PepT2 410 – 601 sequences. The sequence similarity between the *M. musculus* and *R. norvegicus* 410 – 601 residues is 86%. Many of the different residues between the sequences are involved in crystal contacts (Figure 5.20B). These contacts presumably were essential for crystallisation; therefore the sequence alteration appears to be the reason why the *M. musculus* 410 – 601 did not crystallise.

The resolution of the $P3_221$ and $P4_12_12$ crystals were 2.84 and 2.06 Å respectively. The reason for this was the different solvent content of the crystals. Kantardjieff and Rupp (2003) noted a positive correlation between solvent content and maximum resolution. The final refined solvent content of the $P3_221$ and $P4_12_12$ crystals were 58.6 and 39.4% respectively. The reason for the new crystal form and reduced solvent content is due to the presence of the Cs in the seed condition. Chains A and B in the $P3_221$ ASU are related by a 2-fold axis. Figure 5.16 shows the coordination of a Cs atom between the 2-fold symmetry mates inside the $P4_12_12$ unit cell. This coordination changed the orientation of these symmetry mates compared to the $P3_221$ ASU. The $P4_12_12$ packing turned a 2-fold NCS axis into a 2-fold crystallographic axis. This tighter packing allowed a higher crystallographic symmetry, lower solvent content and higher resolution.

Overall the $P3_221$ and $P4_12_12$ crystals have produced two well refined X-ray crystal structures to 2.81 and 2.06 Å respectively, to acceptable degrees of error (Evans and Murshudov 2013). Although the $P3_221$ structure is of lower resolution, the density is complete for the entire 410 – 601 construct, therefore the $P3_221$ chain A is assumed to represent the complete structure of PepT2$^{\mathrm{ECD}}$. The $P4_12_12$ structure was still useful as it has shown that at a higher resolution, there is not a significant difference between the models.

The model of the *Rn*PepT2$^{\mathrm{ECD}}$ indicates that, like PepT1$^{\mathrm{ECD}}$, PepT2$^{\mathrm{ECD}}$ has a definite structure and therefore can actually be considered a domain. It is also clear from the examination of the PepT2 structure that there appears to be a strong similarity between



```
R.norvegicus   ASQEIFLQVLNLADGDVKVTVLGSRNNSLLVESVSSFQNT   449
M.musculus     ...............eie...q.n...p.....i......   449

R.norvegicus   THYSKLHLEAKSQDLHFHLKYNSLSVHNDHSVEEKNCYQL   489
M.musculus     ......r..t............n.....ey..........   489

R.norvegicus   LIHQDGESISSMLVKDTGIKPANGMAAIRFINTLHKDLNI   529
M.musculus     vv.en...l................t...........m..   529

R.norvegicus   SLDTDAPLSVGKDYGVSAYRTVLRGKYPAVHCETEDKVFS   569
M.musculus     ...an............e....q.............n...   569

R.norvegicus   LDLGQLDFGTTYLFVITNITSQGLQAWKAEDI   601
M.musculus     .n..................nr..........   601
```

Figure 5.20: An analysis of why the *Mm*PepT2$^{\mathrm{ECD}}$ construct failed to crystallise. (A) An alignment of the *R. norvegicus* and *M. musculus* PepT2$^{\mathrm{ECD}}$ 410 – 601 sequences. The secondary structure from the *R. norvegicus* crystal structure is shown and differences between the *M. musculus* and *R. norvegicus* sequences are shown in lower case. The blue indicates which of these residues which are involved in crystal contacts. **(B)** The *Rn*PepT2$^{\mathrm{ECD}}$ P$3_2$21 ASU highlighting the residue positions from **(A)** which differ between the *M. musculus* and *R. norvegicus* PepT2 sequences. The A, B and C chains of the P$3_2$21 ASU are shown in red, blue and green respectively and close symmetry related monomers are coloured in grey.

the ECDs of PepT1 and PepT2; the RMSD of the individual lobes between PepT1 and PepT2 is 0.5 Å, however, the lobes appear to have a different orientation. The similarities and differences between the PepT1 and PepT2 ECD are explored in Chapter 6.

6

Structural and biophysical analysis of the PepT1 and PepT2 ECD lobe orientation

6.1 Summary

This chapter details the analysis of the ECD structures from PepT1 and PepT2. The ECDs are structurally very similar; both are composed of two sub-domains (lobes) connected by a flexible linker. These lobes overlay with a Cα-RMSD of 0.5 Å, although the lobes in the crystal structures are in different orientations. A search of the PDB using the ECD structures discovered that the lobes share the same transthyretin-like fold and are structurally related to Transthyretin (TTR), a Thyroxine (T4) binding protein, with a Cα-RMSD of 2.5 Å. A sequence and structural comparison of the ECDs and TTR indicated that it was improbable that the ECDs share a similar function to

TTR, but suggested that the T4 binding residues have been conserved, and mediate the interface between the lobes in PepT1 and PepT2 ECD. The lobe interfaces were explored using Small-angle X-ray scattering (SAXS) and AUC to determine the probable physiological lobe conformation: open or closed. The closed conformation appears to be the preferred state of both PepT1 and PepT2 ECD. In PepT1$^{\mathbf{ECD}}$ this conformation is very stable, held together by two salt bridges. In PepT2$^{\mathbf{ECD}}$ the closed conformation appears to be more dynamic, as only one salt bridge is present. Ultimately these data suggest that the closed conformation of the lobes is the physiological state of both ECDs, and should be used when constructing the hybrid models of the native transporters (§ 7.3.1).

6.2 The structures of *Mm*PepT1 and *Rn*PepT2 ECD

Figure 6.1A shows the monomeric units of the *Mm*PepT1 and *Rn*PepT2 ECDs as determined by X-ray crystallography. The two ECDs appear to have similar structural features; both are composed of two sub-domains (lobes) connected by a flexible linker. Figures 6.1B and C show the similar topology and connectivity of the ECDs. *Mm*PepT1$^{\mathrm{ECD}}$ has an α-helix inserted into the D-E loop and a di-sulphide bond between C540 and C566 in the M-N and P-O loops respectively. These features are not in seen in *Rn*PepT2$^{\mathrm{ECD}}$. Interestingly, two cysteines are present in the *Rn*PepT2$^{\mathrm{ECD}}$ sequence, but the structure indicates that they do not form a di-sulphide bond.

Figure 6.2 shows an alignment of the *H. sapiens*, *M. musculus* and *R. norvegicus* PepT1 and PepT2 sequences with the secondary structure elements from the crystal structures mapped onto the alignment. Examining this alignment, the secondary structure generally matches between the two ECDs. That said, the sequence conservation is low, 19% identity between *H. sapiens* PepT1 and PepT2 ECD sequences. The sequence conservation is also different between the lobes. The majority of the sequence conservation is in lobe 2, 22% identity compared to 7% identity in lobe 1.

The structural alignment of a *Rn*PepT2$^{\mathrm{ECD}}$ monomer onto the ASU of the *Mm*PepT1$^{\mathrm{ECD}}$ crystal structure (Figure 6.3) proposed the question which is the basis of the chapter:

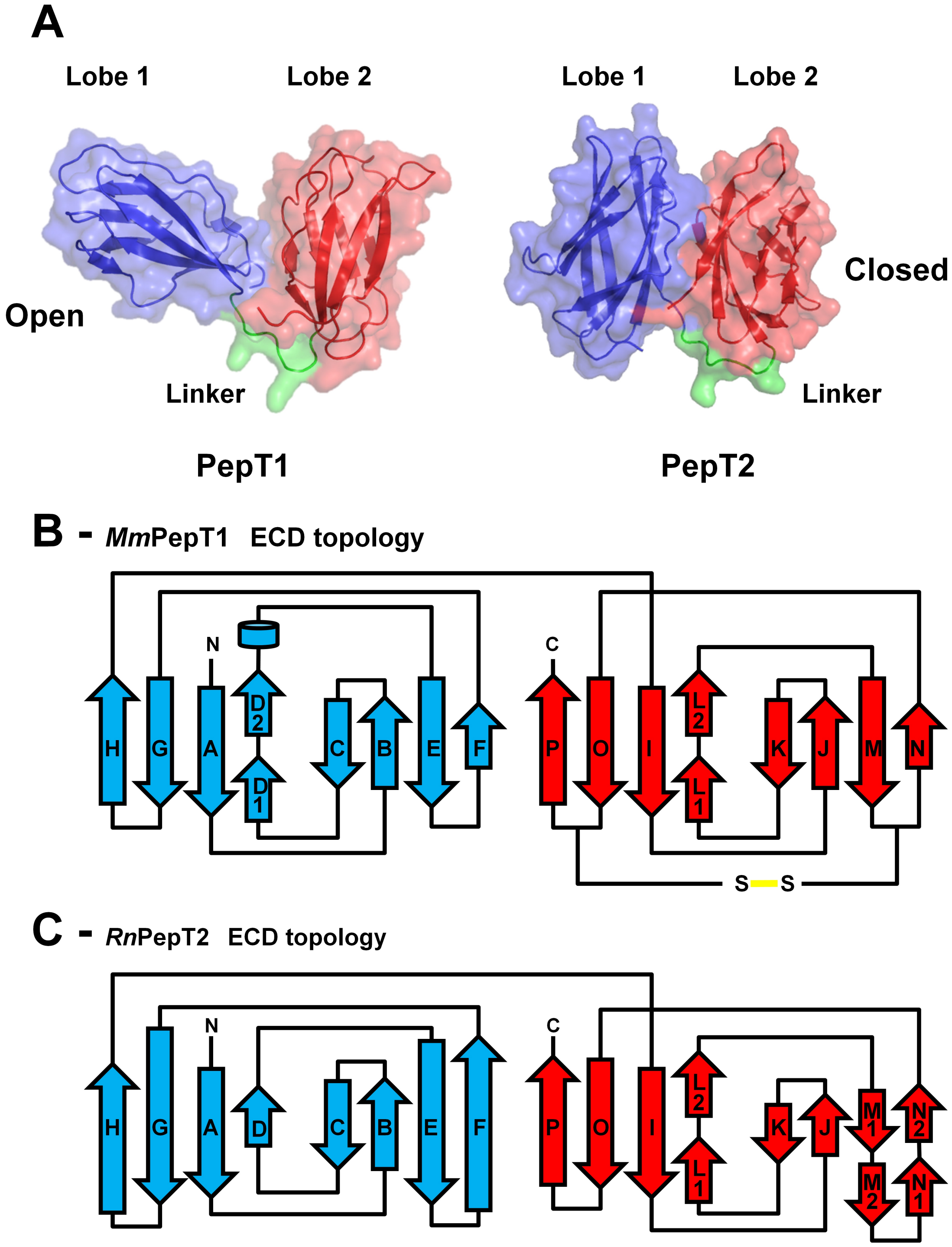


Figure 6.1: The structure and topology of *Mm*PepT1 and *Rn*PepT2 ECD. (A) The monomeric structures of the ECDs. The N- and C-terminal lobes of the ECDs have been coloured blue and red respectively and the linker in green. The lobes of *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ are arranged in different orientations with respect to each other in the crystal structures. *Mm*PepT1 and *Rn*PepT2 ECD present in open and closed conformations respectively. **(B)** and **(C)** show the topology diagrams of PepT1 and PepT2 ECD respectively.

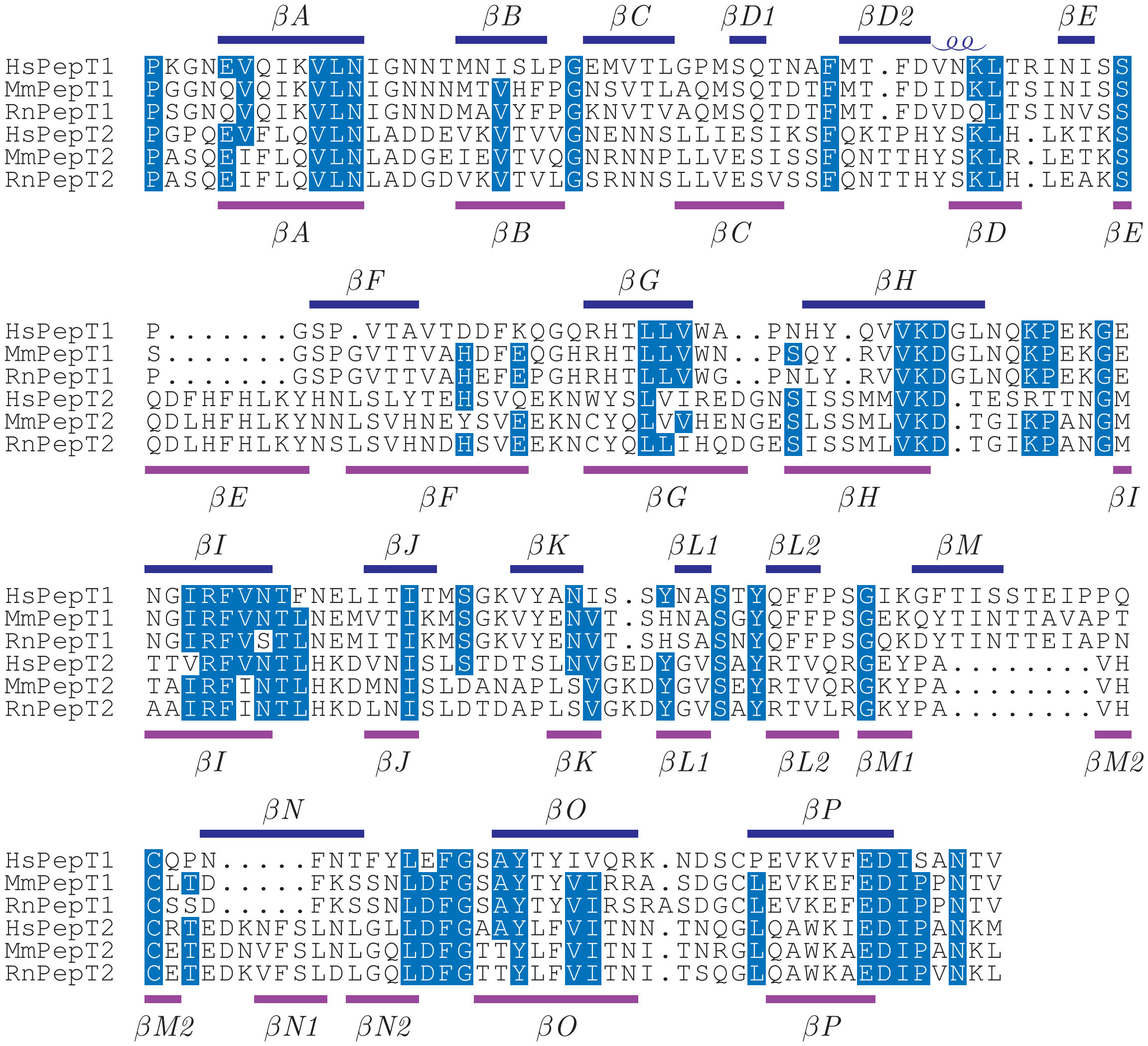


Figure 6.2: An alignment of *H. sapiens*, *M. musculus* and *R. norvegicus*, PepT1 and PepT2 ECD sequences. The secondary structure of the *Mm*PepT1 and *Rn*PepT2 ECD crystal structures has been mapped onto the top and bottom of the alignment respectively. The full names and Uniprot codes for each gene are listed in Table 2.1.

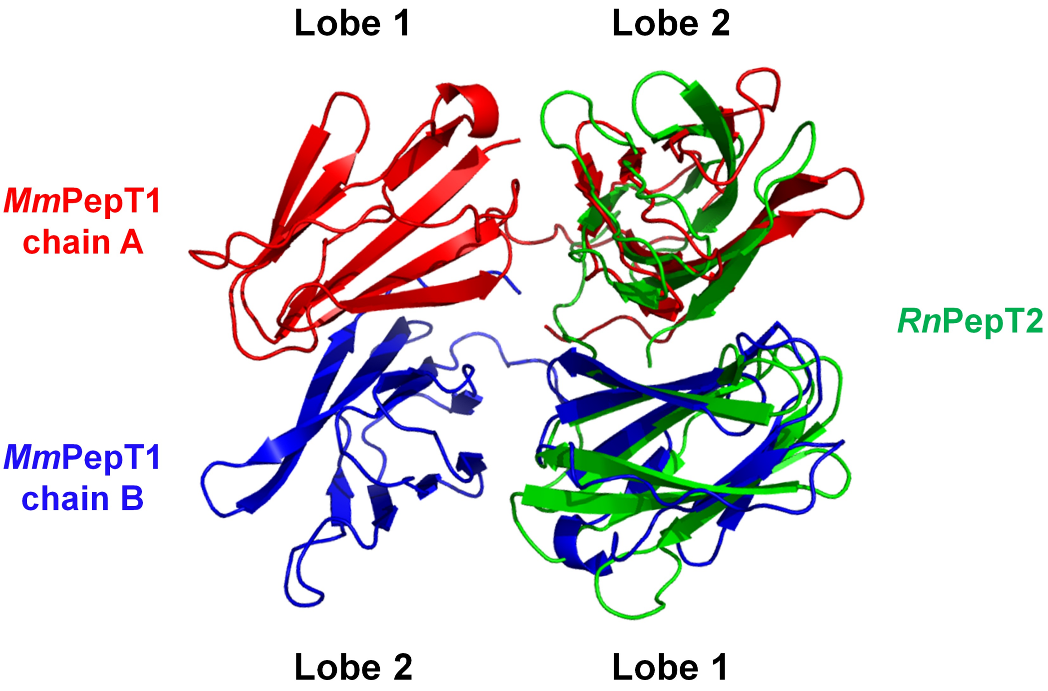


Figure 6.3: A Cα alignment of the monomeric *Rn*PepT2$^{\mathbf{ECD}}$ onto the ASU of *Mm*PepT1$^{\mathbf{ECD}}$. The Cα alignment was performed using `SSM` (Krissinel and Henrick 2004). *Mm*PepT1$^{\mathrm{ECD}}$ chain A and B are coloured in red and blue respectively, and the *Rn*PepT2$^{\mathrm{ECD}}$ monomer is coloured in green. The *Rn*PepT2$^{\mathrm{ECD}}$ monomer aligns onto lobe 1 and lobe 2 of *Mm*PepT1$^{\mathrm{ECD}}$ from the adjacent monomer chains in the ASU.

what are the lobe conformations of ECDs? The *Rn*PepT2$^{\mathrm{ECD}}$ structure aligns onto the *Mm*PepT1$^{\mathrm{ECD}}$ ASU with a Cα-RMSD of 1.5 Å, where the *Rn*PepT2$^{\mathrm{ECD}}$ structure overlays onto lobe 1 and lobe 2 from adjacent monomer chains (Figure 6.3). This indicated that the *Mm*PepT1$^{\mathrm{ECD}}$ may also be able to adopt the closed conformation like *Rn*PepT2$^{\mathrm{ECD}}$. It also therefore suggested, that *Rn*PepT2$^{\mathrm{ECD}}$ can form the open conformation like *Mm*PepT1$^{\mathrm{ECD}}$. To address the question of whether the ECD lobes are in an open, closed or other conformation in solution, a sequence and structure based analysis of the ECDs was performed to discover if the ECDs shared any features with known structures.

6.3 Alignment analyses of the ECDs

6.3.1 Sequence and structural database searches

Sequence based `BLAST` (Altschul *et al.* 1990) and `PSI-BLAST` (Altschul *et al.* 1997) searches of the ECDs against the Uniprot database suggests that PepT1 and PepT2

ECD have no close sequence homologues other than themselves. The structures of the ECDs were used to search for similar structures in the PDB using the `Dali` server (Holm and Rosenström 2010). This search indicated that there were two proteins; TTR and 5-hydroxyisourate (5-HIU) hydrolase which showed a similar topology and connectively to the ECDs, which align with a Cα-RMSD of 2.5 Å. No other close structural homologues to PepT1 or PepT2 ECD were observed.

Transthyretin-related protein family

TTR is a vertebrate plasma hormone binding protein with specific affinity for T4 (Hamilton and Benson 2001). TTR can also form a complex with retinol and retinol-binding protein, which enhances the stability of the complex (Monaco, Rizzi, and Coda 1995; Hamilton and Benson 2001). 5-HIUase functions in the purine catabolism pathway where it hydrolyses 5-HIU (Hennebry *et al.* 2006). Both of these proteins belong to the TTR superfamily (Fox, Brenner, and Chandonia 2014) and have a prealbumin-like fold.

The secondary and tertiary structures of 5-HIUase are identical to that of TTR. Figure 6.4A shows the TTR tetramer with bound T4. The tetramerization of the TTR monomers occurs through a dimerization of homo-dimers. The monomers in cis interact between anti-parallel β-sheets. Two of these dimers then come together, principally through interactions with a T4 molecule. The topology and connectivity of TTR and 5-HIUase are shown in Figure 6.4C. TTR shows high conservation in both sequence and function across vertebrate species. TTR-related protein (TRP) such as 5-HIUase, have predominantly been associated with purine metabolism (Eneqvist *et al.* 2003; Hennebry 2009) in both bacteria and eukaryae; however, many have an unknown function.

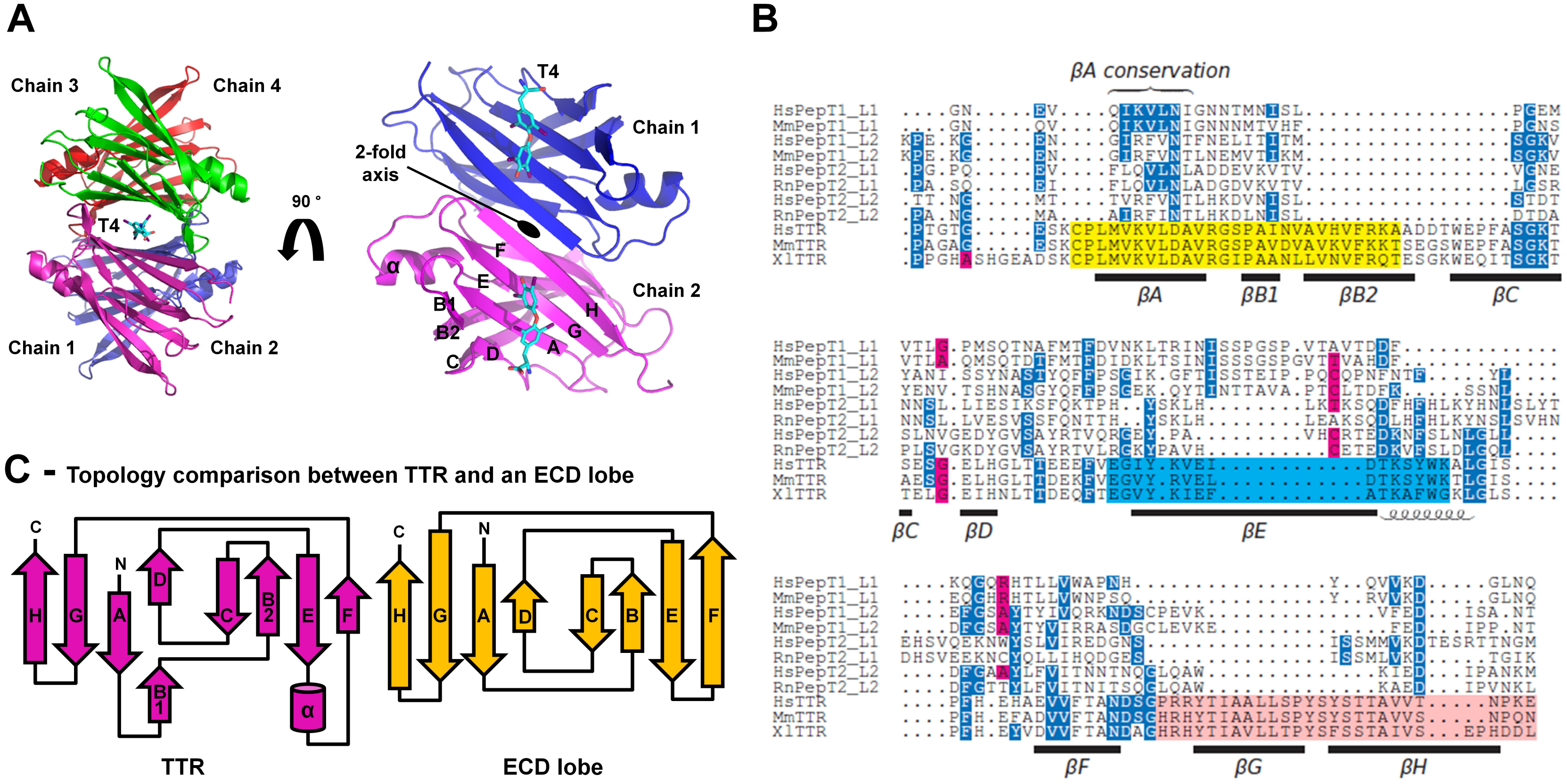


Figure 6.4: A sequence and structural comparison of TTR and the ECD lobes. **(A)** The multimerisation of TTR. The four chains which make up the TTR homotetramer are coloured purple, blue, green and red, with T4 in the central binding cavity. The planer view of the dimer shows that the monomers are related by a 2-fold rotational axis and that T4 contributes to the tetramisation interface. **(B)** A sequence alignment of the mammalian peptide transporter ECD lobes and TTR. Lobe 1 and lobe 2 are indicated by the suffix L1 and L2 receptively. The secondary structure of *Hs*TTR has been mapped below the alignment. The highlighted regions show three sequence motifs; B', C' and A' identified in Hennebry 2009. The full names and Uniprot codes for each gene are listed in Table 2.1. **(C)** A topology comparison between TTR and an ECD lobe.

6.3.2 Sequence analysis of the ECDs lobes and TTR

Figure 6.4B shows a sequence alignment between PepT1 and PepT2 lobe 1, 2, and TTR. The sequence identity of *Hs*PepT1 and *Hs*PepT2, lobe 1 and 2 between *Hs*TTR is 17.05, 18.18, 8.49 and 18.56% respectively. The secondary structure is relatively well conserved, as would be expected from the similar topology of the lobes and TTR (Figure 6.4C). Hennebry (2009) identified 3 sequence motifs which characterise TTR and TRPs, and these are highlighted in Figure 6.4C. These motifs do not appear to be conserved across the lobes of PepT1 or PepT2, so based upon sequence the lobes appear to be a novel TRPs.

The ECD lobes show sequence divergence from the TTR (Figure 6.4B). However, some sequence conservation is present on β-sheet A (Figure 6.4B βA). The K15-V16-L17 sequence from the TTRs also appears in PepT1$^{\mathrm{ECD}}$ lobe 1 and lobe 2, as K-V-L and R-F-V respectively. In PepT2$^{\mathrm{ECD}}$, the R-F-V is present in lobe 2. However, in lobe 1 the positive lysine or arginine becomes a conserved glutamine. The K15 and V17 residues in TTR are involved in T4 binding and oligomerisation. K15 has been shown to be essential for TTR stability (Choi *et al.* 2010), and this mutation increases the propensity of fibril formation in amyloidosis.

6.3.3 Sequence analysis of the ECD lobes

A sequence alignment of animal lobe 1 and lobe 2 sequences (Figure 6.5) shows that there are three regions of the lobe sequences that show conservation between PepT1 and PepT2 ECD. There is an absolutely conserved aspartate residue which is located on the end of βH and βP, of lobe 1 and 2 respectively. A conserved region is found on βA (Figure 6.5 βA motif) and in the F-G and N-O loops (Figure 6.5 Loop motif) of lobes 1 and 2 respectively. Figure 6.6A shows the consensus sequence of the βA and Loop motifs.

A relatively weak consensus sequence can be made for the Loop motif, LDFGxxYTLVI (Figure 6.6A). These residues in PepT1 and PepT2 lobe 1 (F-G Loop) are not well conserved, and the appearance of conservation is perhaps due to the alternating R-group characteristic of β-sheets (Figure 6.6B). This is not the case for the Loop motif in lobe 2. In the N-O Loop, the residues and the structure of the loop, are highly conserved

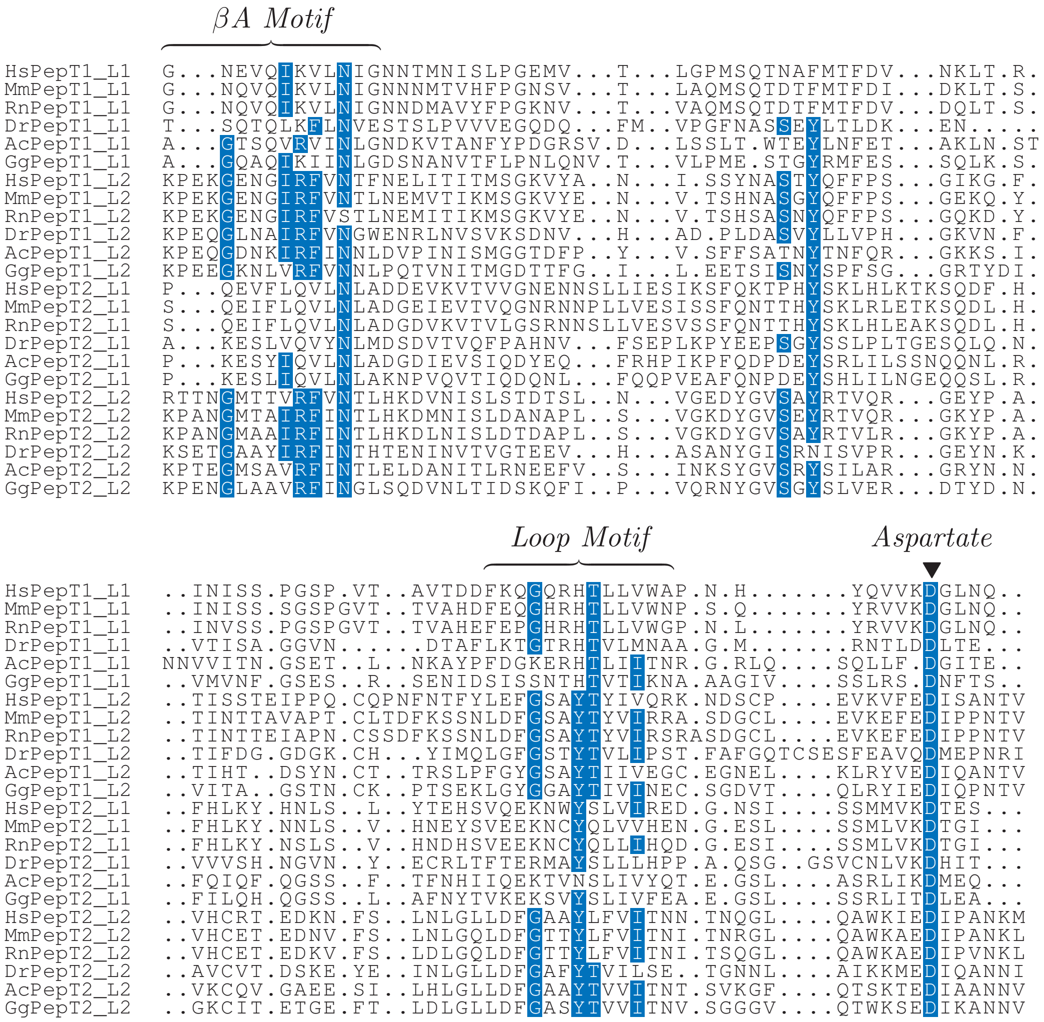


Figure 6.5: An alignment of vertebrate PepT1 and PepT2, lobe 1 and lobe 2 sequences. Lobe 1 or lobe 2 is indicated by the suffix L1 and L2 receptively. Two regions of the alignment which show sequence conservation are labelled βA, and Loop motif, and an absolutely conserved aspartate is highlighted by a black triangle (▼). The full names and Uniprot codes for each gene are listed in Table 2.1.

between PepT1 and PepT2 (Figure 6.6B). The reason for this conservation was not clear but is perhaps involved in the ECD function (discussed in § 7.5.1). Across both lobe 1 and 2, similar residues are observed towards the end of the Loop motif which extend into the interface. The valine position in the LDFGxxYTLVI sequence is almost exclusively a hydrophobic residue (either valine, leucine or isoleucine) which makes up a hydrophobic patch (discussed in § 6.4).

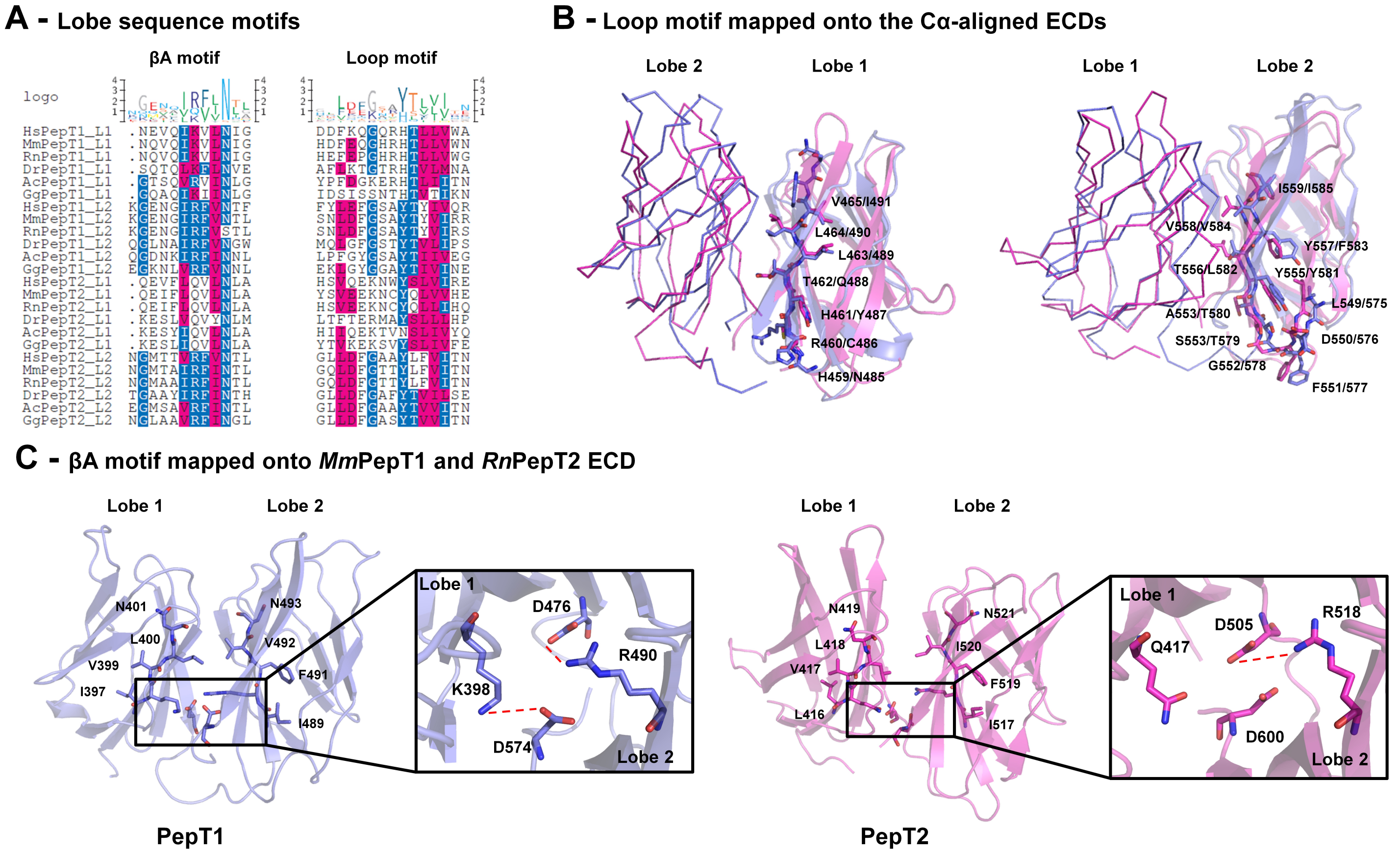


Figure 6.6: Structural analysis of the lobe motifs. **(A)** The consensus sequences of the two motifs highlighted in Figure 6.5: βA (left) and Loop (right). The βA and Loop motifs have been mapped on the ECDs in **(B)** and **(C)**. MmPepT1$^{\mathrm{ECD}}$ and RnPepT2$^{\mathrm{ECD}}$ are shown in blue and purple respectively and the lobes are orientated in the closed conformation. The residue numbers are labelled MmPepT1$^{\mathrm{ECD}}$/RnPepT2$^{\mathrm{ECD}}$. **(B)** A Cα alignment of the MmPepT1$^{\mathrm{ECD}}$ and RnPepT2$^{\mathrm{ECD}}$ structures with the Loop motif mapped onto lobe 1 (left) and lobe 2 (right). **(C)** The βA motif mapped onto the alignment structures of MmPepT1$^{\mathrm{ECD}}$ (left) and RnPepT2$^{\mathrm{ECD}}$ (right) with the absolutely conserved aspartate residue.

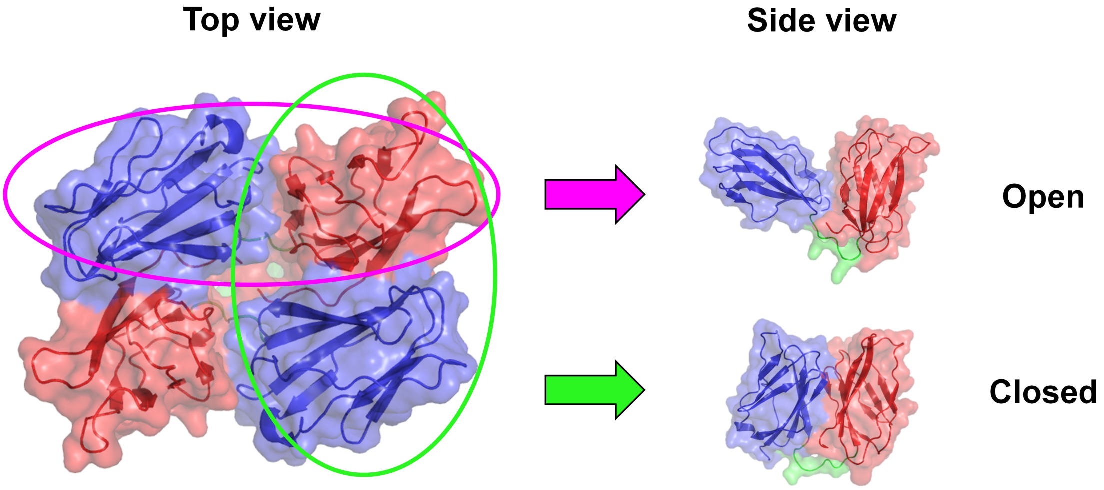


Figure 6.7: The two potential lobe conformations based on the *Mm*PepT1$^{\mathrm{ECD}}$ ASU. The open model was observed in the crystal structure. The closed model was constructed based on the lobe arrangement between inter-monomers in the ASU.

A stronger consensus is found for the βA Motif (Figure 6.6A) of IRFLN. This motif is related to the K-V-L sequence in TTR. Figure 6.6C shows the atomic positioning of the βA motif onto the crystal structures of *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ with the absolutely conserved aspartate residue. The βA motif exists in the lobe interface in PepT1 and PepT2 ECD. In *Mm*PepT1$^{\mathrm{ECD}}$, two salt bridges are formed between a K398/R490 and D476/D574 (lobe 1/lobe 2), but in *Rn*PepT2$^{\mathrm{ECD}}$ only a single salt bridge is observed between D505 and R518. The second bridge is not present as the K398 of *Mm*PepT1$^{\mathrm{ECD}}$ is a glutamine in *Rn*PepT2$^{\mathrm{ECD}}$. This glutamine appears to be absolutely conserved across PepT2 lobe 1 sequences (Figure 6.6A). The aspartate residues in PepT2 are also closer together in the PepT2$^{\mathrm{ECD}}$ structure compared to PepT1$^{\mathrm{ECD}}$, with distances of 3.7 and 5.6 Å respectively (discussed further in § 6.4). Therefore there is sequence and structural conservation in the lobe interface between PepT1 and PepT2 ECDs. However, there are also a distinct differences, in the removal of a salt bridge and the subsequent proximity of D505 and D600 in PepT2.

6.4 Bioinformatics analysis of the ECDs

Figure 6.7 shows the two possible lobe arrangements of the ECD based on the ASU of *Mm*PepT1$^{\mathrm{ECD}}$. The open conformation is observed in the monomeric *Mm*PepT1$^{\mathrm{ECD}}$. The

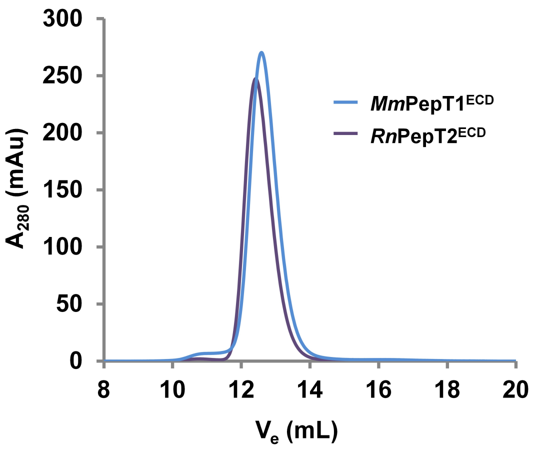


Figure 6.8: The S75 10/300 GF traces of *Mm*PepT1 and *Rn*PepT2 ECD. The column was equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl. The V_es of *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD}, 12.51 ± 0.02 and 12.41 ± 0.02 mL respectively (limits calculated by a T-test at the 0.05 level), equate to MWs of 21.0 and 22.0 kDa respectively (see globular protein equilibration curve § 2.8.3), suggest that both ECDs are monomeric in solution. There is a statistically significant difference in V_e between the ECDs with $n = 3$ at the 0.05 criterion.

closed conformation is observed in the *Rn*PepT2^{ECD} crystal structure and the *Mm*PepT1^{ECD} inter-monomer interface. The sequence and structural analysis in § 6.3.3 suggests the closed conformation is the more likely conformation of *Mm*PepT1^{ECD} as two salt bridges exist between the lobes in this arrangement. This is confirmed by the GF trace of *Mm*PepT1^{ECD} (Figure 6.8). If *Mm*PepT1^{ECD} is in the open conformation in solution, then the observed ASU dimer should also be observed in solution, the GF trace indicates that this is not the case.

For *Rn*PepT2^{ECD} it is less clear. The sequence and structural analysis (§ 6.3.3) showed that a salt bridge is present in the closed conformation lobe interface. The GF trace (Figure 6.8) suggests that like *Mm*PepT1^{ECD}, *Rn*PepT2^{ECD} is monomeric in solution. However, Figure 6.6C shows that D505 and D600 are within 3.7 Å of each other, which implies that the closed conformation may be an unfavourable conformation for *Rn*PepT2^{ECD}. The relative instability of the *Rn*PepT2^{ECD} closed conformation compared to *Mm*PepT1^{ECD} may be visible in the GF traces, as *Rn*PepT2^{ECD} elutes earlier than *Mm*PepT1^{ECD}.

Figure 6.9 shows the electrostatics of the conserved residues in the ECDs in the closed lobe conformation interface identified in § 6.3.3. These residues are grouped into two patches, hydrophobic (containing residues from the end of the βA and Loop motifs) and charged

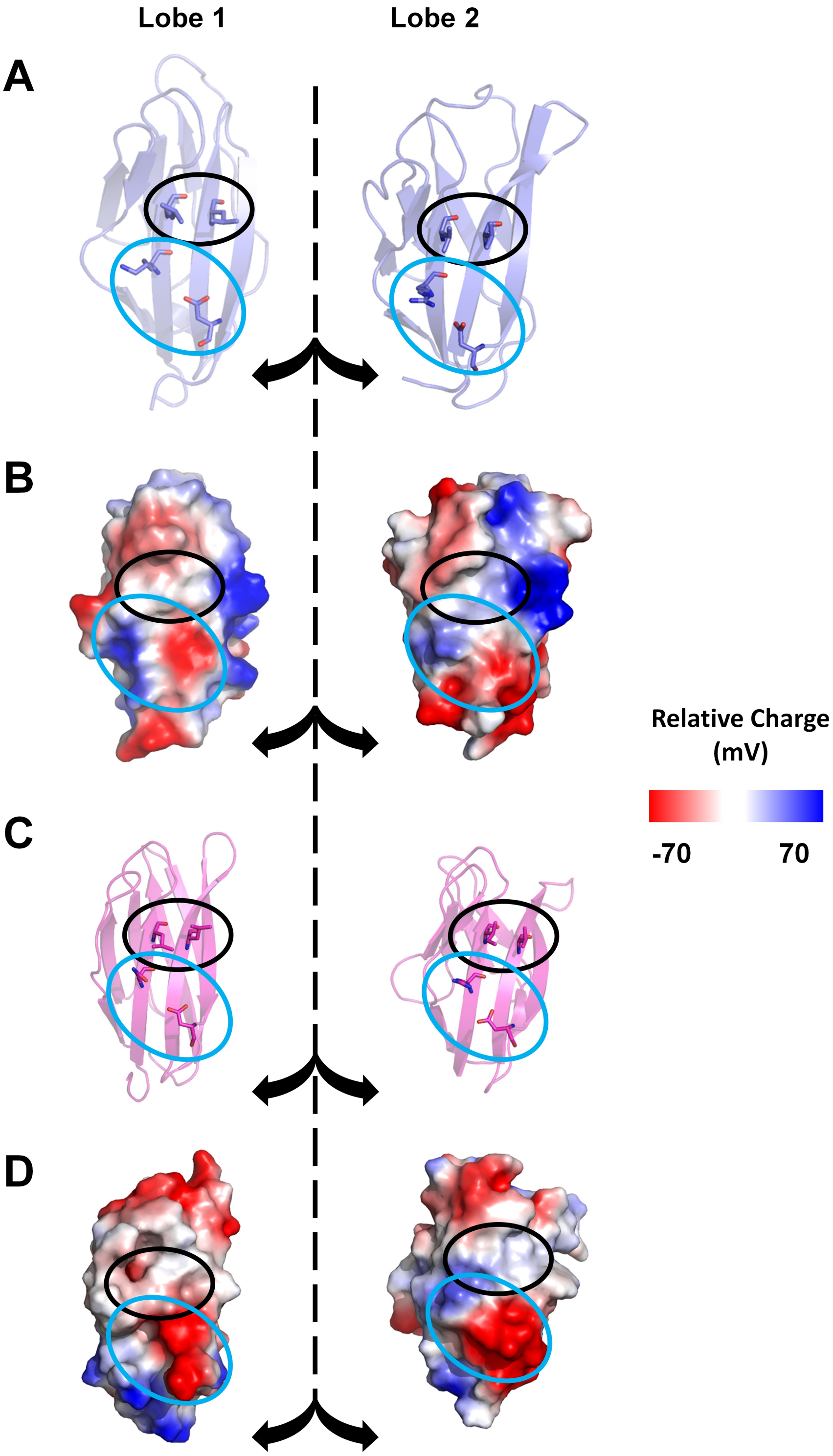


Figure 6.9: The four conserved residues which predominantly make up the closed interface between the lobes in the *Mm*PepT1 and *Rn*PepT2 crystal structure. **(A)** and **(C)** show the residues in PepT1 and PepT2 respectively. **(B)** and **(D)** show the electrostatics (Baker *et al.* 2001). The black oval indicates the hydrophobic patch. The blue oval indicates the charged patch.

(the salt bridges in Figure 6.6C). The salt bridges observed in *Mm*PepT1$^{\mathrm{ECD}}$ suggests a strong interaction in this conformation. The *Rn*PepT2$^{\mathrm{ECD}}$ closed conformation appears unlikely. The charged patch is predominantly negative in lobe 1 and 2 which indicates an unfavourable interaction when in close proximity to each other. A bioinformatic analysis was conducted to estimate the probability of the different lobe interfaces observed in the crystal structures of the ECDs.

6.4.1 `PISA` physiological interface predictions

The probability of a crystallographic interface being a physiological interface can be estimated using `PISA` (Krissinel and Henrick 2005). `PISA` estimates the solvent inaccessible area at a protein interface and estimates the ΔG_{Diss} based on the atoms present in the interface residues. The Complexation significance score (CSS) is a measure of how probable an interface is given the energy of solvation. The `PISA` predictions for the PepT1 and PepT2 ECD ASUs are shown in Tables 6.1A and B respectively, based on the lobe assignment in Figure 6.10.

The `PISA` prediction for the *Mm*PepT1 lobes indicates, that the adjacent monomer lobe interfaces, A-D and B-C (Figure 6.10A), in the ASU are the most probable physiologically, *i.e.* the closed lobe conformation. The *Rn*PepT2$^{\mathrm{ECD}}$ ASU presents slightly contradictory results. Two of the three monomers in the $P3_221$ ASU, E-F and G-H, have a CSS of 1.0, indicating that the interface is physiological, however, the I-J monomer has a CSS of 0.1, suggesting the interface is not physiological. The reason for this drop in the CSS is likely to be due to the drop in buried surface area in the I-J, which may be a result of the crystal packing, and the poorer electron density observed from chain C (I-J) compared to chains A (E-F) and B (G-H). Overall the `PISA` analysis supports the idea that the closed conformation is the preferred conformation of the lobes in both *Mm*PepT1 and *Rn*PepT2 ECD.

6.4.2 `PROPKA` analysis of the conserved aspartate residues in the ECDs

Given the relative positions of the conserved aspartate residues in PepT1 and PepT2 ECDs, an important consideration is the charge on the aspartate carboxyl side chain.

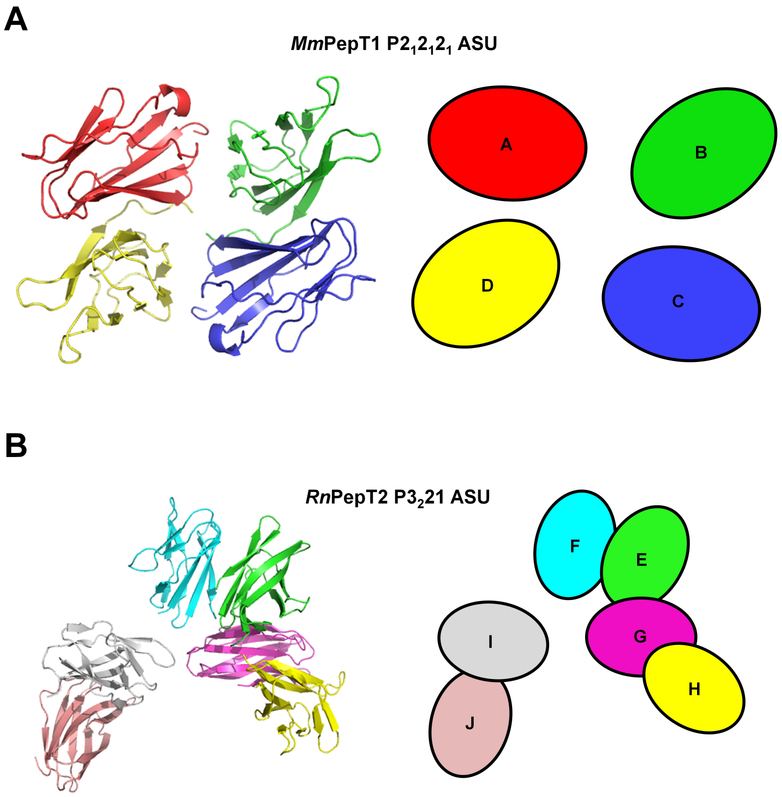


Figure 6.10: The lobe assignment used in the `PISA` (Krissinel and Henrick 2005) analysis presented in Table 6.1 for *Mm*PepT1 (A) and *Rn*PepT2 P$3_2$21 (B).

A. *Mm*PepT1$^{\mathrm{ECD}}$

Interface	Buried Surface Area (Å^2)	CSS
A–B	213.6	0.0
B–C	790.4	1.0
C–D	209.5	0.0
A–D	786.2	1.0

B. *Rn*PepT2$^{\mathrm{ECD}}$ P$3_2$21

Interface	Buried Surface Area (Å^2)	CSS
E–F	629.6	1.0
F–G	159.8	0.0
F–J	106.5	0.0
E–G	284.3	0.0
E–H	230.9	0.0
G–H	624.8	1.0
I–J	573.5	0.1

Table 6.1: The lobe interface `PISA` predictions of the *Mm*PepT1 and *Rn*PepT2 ASUs. Figure 6.10 shows the lobe assignment used in this Table. The buried surface area and interface probability (CSS) have been calculated using the `PISA` server (Krissinel and Henrick 2005).

		pK_a	Buried (%)
***Mm*PepT1^{ECD}**	**D476**	3.5	98.5
	D574	3.8	93.5
***Rn*PepT2^{ECD}**	**D505**	8.0 ± 1.7	87 ± 13
	D600	5.5 ± 1.0	83 ± 8

Table 6.2: The pK_as and percentage exposures of the conserved aspartate residues in the *Mm*PepT1 and *Rn*PepT2 ECDs. The pK_a and surface exposure values were calculated by `PROPKA` (Li, Robertson, and Jensen 2005; Bas, Rogers, and Jensen 2008; Olsson *et al.* 2011; Sondergaard *et al.* 2011). The values given are averages over the crystal structure ASUs with 95% confidence limits given where possible.

In the *Mm*PepT1^{ECD} closed conformation the charge on each aspartate is neutralised by an electrostatic interaction with a lysine or arginine. In *Rn*PepT2^{ECD}, only R518 (Figure 6.6C) is capable of stabilising the buried D505 and D600 carboxyl group. Therefore its likely that a H^+ is bound to either D505 or D600.

`PROPKA` (Li, Robertson, and Jensen 2005; Bas, Rogers, and Jensen 2008; Olsson *et al.* 2011; Sondergaard *et al.* 2011) was used to calculate the pK_a values of the ECD conserved aspartate residues (Table 6.2). The average pK_a value for a hydrated aspartate carboxyl group is 3.9. The pK_as calculated for *Mm*PepT1^{ECD} D476 and D574 show no deviation in the carboxyl group pK_a, even though the residues are almost 100% separated from solvent, no doubt due to the positive charge from K398 and R490.

The *Rn*PepT2^{ECD} aspartate residues, D505 and D600 however, do show significant deviation from a pK_a of 3.9 at the 0.05 criterion. Both D505 and D600 are approximately 85% buried. Due to the lack of a positively charged moiety near the D505, the pK_a is predicted to be 8.0 ± 1.7. This indicates that D505 is likely to have bound a H^+ to neutralise the negative charge. It also suggests that an energetically unfavourable action is occurring to satisfy the closed conformation in the *Rn*PepT2^{ECD} crystal structure. The increased stability of the *Mm*PepT2^{ECD} 410 – 609 construct at lower pH observed in FSC experiments (Figure 5.2B), is explained by this observation (discussed further in § 6.6).

6.5 Linker mutation study

The sequence (§ 6.3) and `PISA` (§ 6.4.1) analysis indicated that both PepT1 and PepT2 ECD are likely to be a closed conformation. To experimentally validate this, a 3C protease cut site was engineered into the flexible linker of the *Mm*PepT1 and *Rn*PepT2 ECD constructs (called 3CX). The link between the lobes was cleaved and the presence of a persistent interaction was assessed by GF (calibration curve shown in § 2.8.3). The cloning of the *Mm*PepT1$^{\text{ECD-3CX}}$ constructs is described in § 2.6.4. Unfortunately the *Rn*PepT2$^{\text{ECD-3CX}}$ construct was not stable, so SAXS and AUC were used to characterise the *Rn*PepT2$^{\text{ECD}}$ interface (described in § 6.6). Once the *Mm*PepT1$^{\text{ECD-3CX}}$ construct had been made, a D574A mutant variant was made to assay the salt bridges in determining the closed conformation stability. The D476A mutant was also made but did not stably express.

The *Mm*PepT1$^{\text{ECD-3CX}}$ and *Mm*PepT1$^{\text{ECD-3CX}}$-D574A constructs were transformed into C43 (DE3) cells (§ 2.6.2) and 2 L of *E. coli* were grown, and the cell lysates prepared as described in § 2.6.3. The purification protocols were the same as WT *Mm*PepT1$^{\text{ECD}}$ (§ 3.4.1) and the final GF traces and gels are shown in Figure 6.11. The *Mm*PepT1$^{\text{ECD-3CX}}$ construct expressed and purified at a similar level to WT (10 mg per 1 L of TB). *Mm*PepT1$^{\text{ECD-3CX}}$-D574A expressed well but appeared to be much more unstable; a significant proportion of the 3CX-D574A appeared in the V_0 of the GF column which suggests the mutant was more unstable than the WT-3CX.

The linker cleavage experiment was performed by taking 100 µL of 2.0 mg/mL 3CX (WT or D574A) ECD, adding 10 µL of 1.0 mg/mL 3C protease and incubating at 20 °C for 1 hr. The sample was then centrifuged at 30,000 × g/4 °C for 15 min to remove any aggregation. The supernatant was injected into an S75 10/300 GF column equilibrated in 20 mM Tris pH 7.5, 150 mM NaCl and the fractions collected (Figure 6.12A and C). The fractions were concentrated using a TCA precipitation reaction (§ 2.4.2) and run on a reducing 15% Tris-Gly SDS-PAGE gel (Figure 6.12B and D).

6.5.1 *Mm*PepT1$^{\text{ECD-3CX}}$ linker cleavage experiment results

The main elution peak at 12.55 mL in Figure 6.12A gives a MW of 20.7 kDa which is consistent with the WT uncleaved protein. There is also a trailing peak at approximately 13.75 mL giving a MW of 12.5 kDa which is likely to be the separate lobes. The SDS-PAGE gel (Figure 6.12B) confirms this hypothesis; both cleaved lobes present in the main peak in fractions 4, 5 and 6. Only the larger lobe appears to be fractions 7 and 8. Precipitation

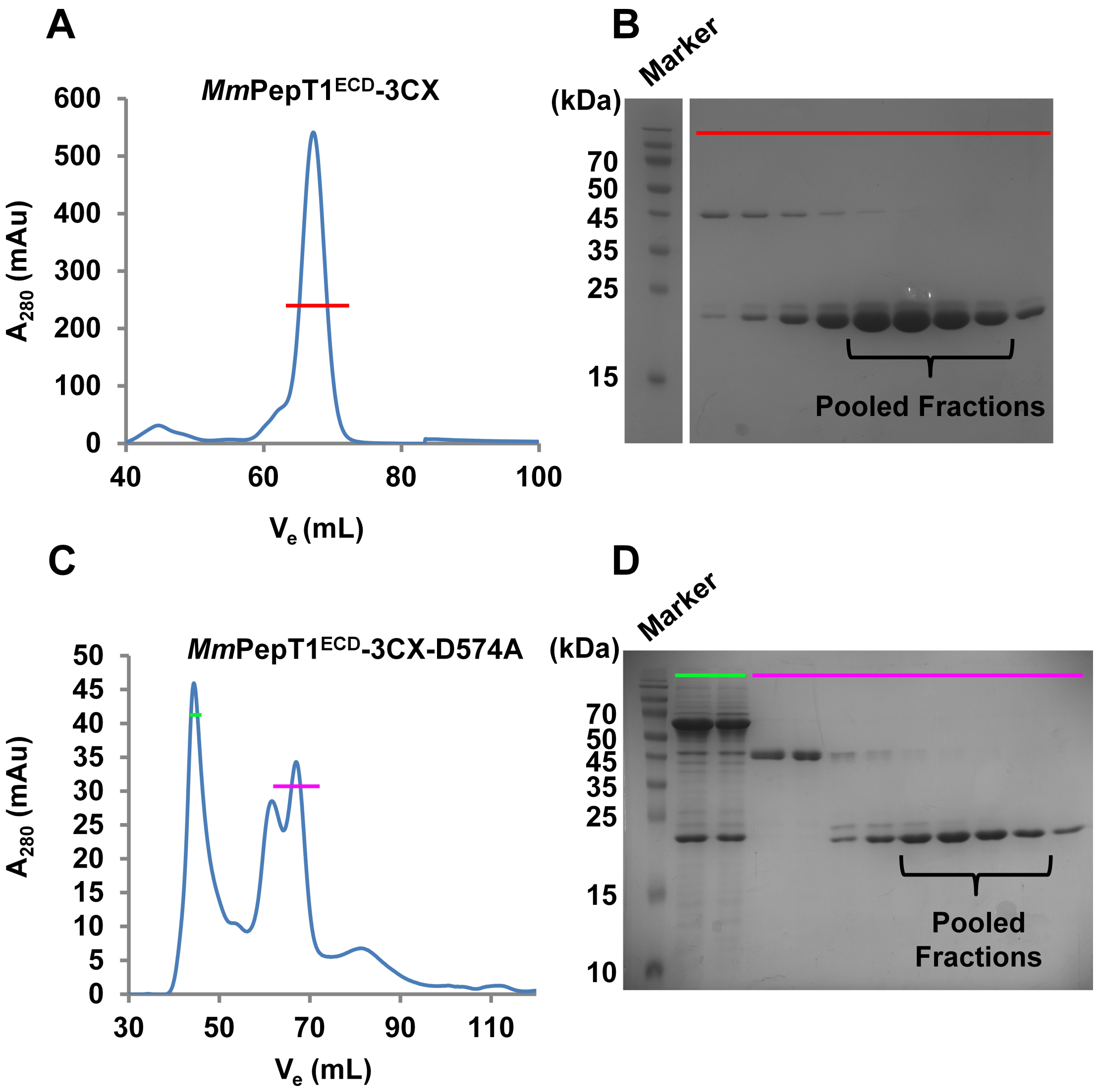


Figure 6.11: Final GF traces and reducing gels from the *Mm*PepT1$^{\text{ECD-3CX}}$ and *Mm*PepT1$^{\text{ECD-3CX}}$-D574A purifications. (**A**) and (**C**) are the S75 16/60 GF traces from the 3CX purifications with peak fractions run through reducing 15% Tris-Gly gels (**B**) and (**D**).

was observed post cleavage, and this was assumed to be the smaller lobe. This experiment suggests that, upon cleavage of the linker, the lobes do still interact strongly enough to remain associated down the GF column, although some dissociation is observed.

Comparing the GF traces in Figures 6.12A and C, it is clear that a significant shift in the elution profiles of the cleaved lobes has occurred. The V_e for the cleaved 3CX-D574A construct has shifted from 12.55 to 13.70 mL. The latter is consistent with a MW of 12.8 kDa, which suggests that the lobes are no longer bound together. This is clearly seen in the gel of the peak fractions from the GF (Figure 6.12D). These data strongly support the hypothesis that the *Mm*PepT1$^{\text{ECD}}$ lobes are bound in the closed conformation.

The other conclusion that can be drawn from the *Mm*PepT1$^{\text{ECD-3CX}}$ experiments is that the smaller cleaved lobe appears to be unstable. There is a hint of this in Figure 6.12B as only a single species is observable in the separate lobe peak, but this is clearly visible in Figure 6.12C. The smaller lobe precipitates out of solution post cleavage. This suggests that the lobe interaction is essential for the stability of the entire *Mm*PepT1$^{\text{ECD}}$.

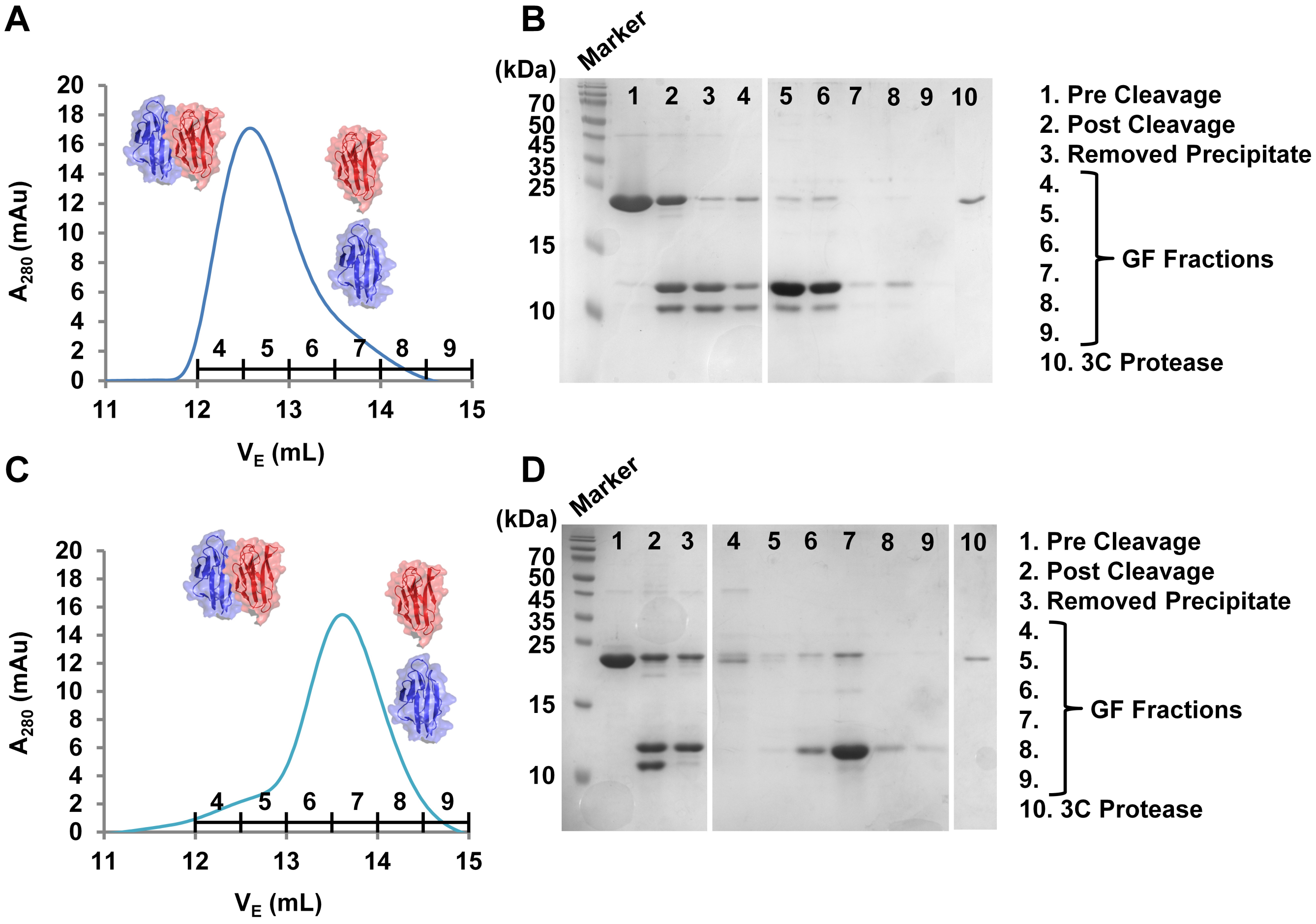


Figure 6.12: The results of the *Mm*PepT1$^{\text{ECD-3CX}}$ linker cleavage experiment. **(A)** and **(C)** show the S75 10/300 GF traces of the cleaved lobes. **(B)** and **(D)** show the reducing 15% Tris-Gly SDS-PAGE gels of the TCA concentrated fractions highlighted from the GF traces. The D574A mutant shows a shift in elution profile compared to the WT which is consistent with a breakdown in the interaction between the two lobes.

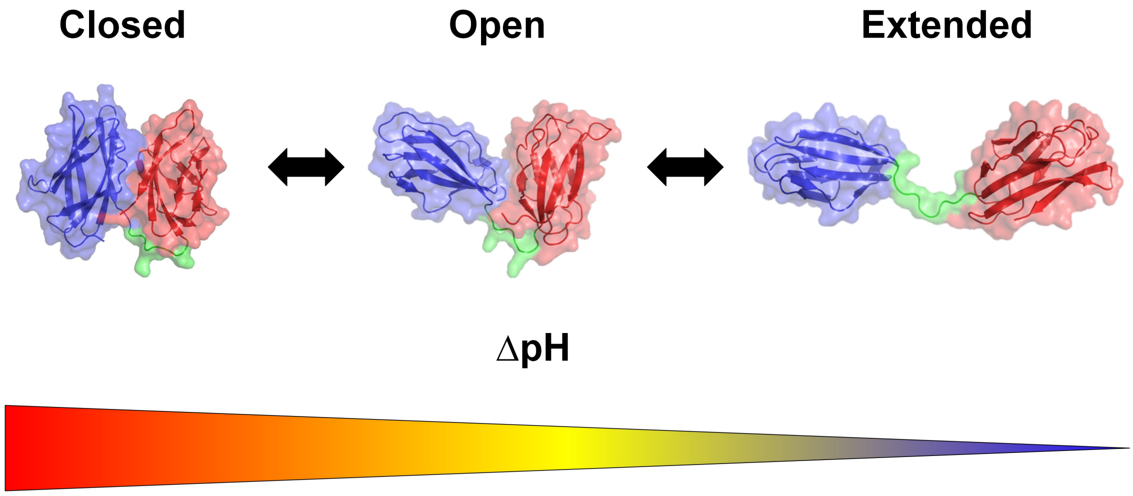


Figure 6.13: A diagram showing the hypothesised change in the PepT2$^{\mathrm{ECD}}$ lobe arrangement with an increase or decrease of H^+.

6.6 PepT2$^{\mathrm{ECD}}$ lobe dynamics

The *Rn*PepT2$^{\mathrm{ECD}}$ lobe interface could not be investigated by the the 3CX experiment as the *Rn*PepT2$^{\mathrm{ECD\text{-}3CX}}$ protein precipitated upon cleavage. So the decision was taken to visualise the lobes in the WT construct in solution. Given the closed conformation interface is present in the *Rn*PepT2$^{\mathrm{ECD}}$ crystal structure, and has been shown to be the *Mm*PepT1$^{\mathrm{ECD}}$ solution interface, it is a reasonable to assume that, if the PepT2$^{\mathrm{ECD}}$ lobes do interact, this interface mediates the interaction. This means that the absolutely conserved aspartate residues also mediates the PepT2 interface.

The `PROPKA` analysis (§ 6.4.2) identified that D505 and D600 are likely to have an increased pK_a to neutralise the charge of the D505. This means that in a low pH environment, a H^+ is more likely to be available to associate with the carboxyl group of D505 or D600. However, in a high pH environment, less H^+ are available and one of the aspartates is more likely to be deprotonated. In a deprotonated state, the interaction between the aspartates is highly unfavourable.

Therefore it was hypothesized that the interface between the lobes in *Rn*PepT2$^{\mathrm{ECD}}$ could be destabilised by increasing the pH (Figure 6.13), thus extending the structure. The difference in lobe conformation could then be detected using SAXS (§ 6.6.1) and AUC (§ 6.6.4). *Mm*PepT1$^{\mathrm{ECD}}$ was used as a control, as it should show no structural change with a change in pH.

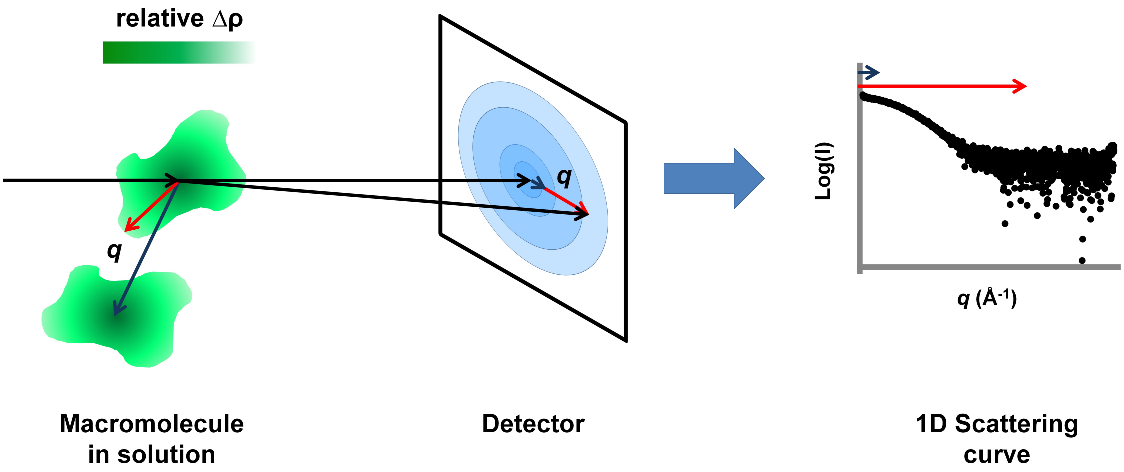


Figure 6.14: A diagram of SAXS data collection. The majority of the scattered X-ray $\boldsymbol{q}$ are from single macromolecules in random orientations (red arrow). Inter-particle scattering is observed at low scattering angles (blue arrow).

6.6.1 Small-angle X-ray scattering

The X-ray scattering observed from a solution of randomly dispersed, randomly orientated particles is similar to diffraction from a crystal, where the scattering vector $\boldsymbol{q}$ is proportional to the scattering angle and inversely proportional to the wavelength (Equation 6.1a) (Svergun and Koch 2003). However, large scattering angle information is not observed due to the lack of coherent discrete diffraction from a 3D crystal. The observed scattering from macromolecules in solution ($\rho(r)$) is given in Equation 6.1b, where r is the scattering length distribution of the macromolecule and s is the scattering from solvent.

$$\boldsymbol{q} = \frac{4\pi \sin 2\theta}{\lambda} \tag{6.1a}$$

$$\rho(r) = \rho(r)_s - \rho_s \tag{6.1b}$$

Figure 6.14 shows the scattering from a macromolecule in solution $\rho(r)_s$. The observed scattering intensity is a spherical average, proportional to the concentration of macromolecules. An assumption is made during SAXS data analysis that no inter-particle scattering is observed in the experiment. Inter-particle scattering can be observed at low scattering angles. The predominant observed scattering from $\rho(r)$ is from distances (r) between 10 and 200 Å. Therefore SAXS is a useful tool to extract low resolution structural information.

6.6.2 SAXS data collection

Samples were prepared for the SAXS pH screening experiment by S75 10/300 GF equilibrated in either: 20 mM Na Acetate pH 5.5, 20 mM Tris pH 7.5, 20 mM Tris pH 8.0 or 20 mM Bicine pH 9.5 (all with 150 mM NaCl). 100 µL of 20 mg/mL *Rn*PepT2 or *Mm*PepT1 ECD were loaded onto the equilibrated column and the peak fraction was collected and concentrated to approximately 5 mg/mL in a 100 µL volume using a 5 kDa MW centrifugal filter. The sample was then flash-frozen in liquid nitrogen. The buffer blank, required for an accurate measurement of the solvent (ρ_s) scattering, was taken from the void volume from the column.

SAXS data was collected at the DESY PETRA III P12 beamline. The beamline is fitted with a Pilatus 2M detector set to a sample distance of 3.1 m. Prior to data collection, the transported frozen samples were defrosted and passed through a 0.22 µm centrifugal filter. A serial dilution of the sample in the specific buffer blank was performed so that 6 protein concentrations for each sample were available. Data collection was designed so that each protein concentration illumination was flanked by buffer blank runs. In total, data from 9 different samples were collected; a BSA control, and *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$, equilibrated in the 4 buffers.

6.6.3 SAXS data processing

SAXS invariant processing

Protein samples illuminated on a SAXS beamline, like protein crystals (discussed in § 4.3.3) are subject to radiation damage (Hura *et al.* 2009; Meisburger *et al.* 2013). To overcome this the sample is continuously passed through the beam during data collection. The 20 curves generated from each sample concentration were each inspected in `PRIMUS` (Konarev *et al.* 2003) and curves which deviated from the norm were excluded from further analysis. The remaining curves were averaged and the buffer scattering subtracted. The resulting intensities represent the scattering of the $\rho(r)$. All further data processing was performed in `ScÅtter` (Rambo and Tainer 2013).

The intensity plots of each sample concentration were viewed to check for inter-particle

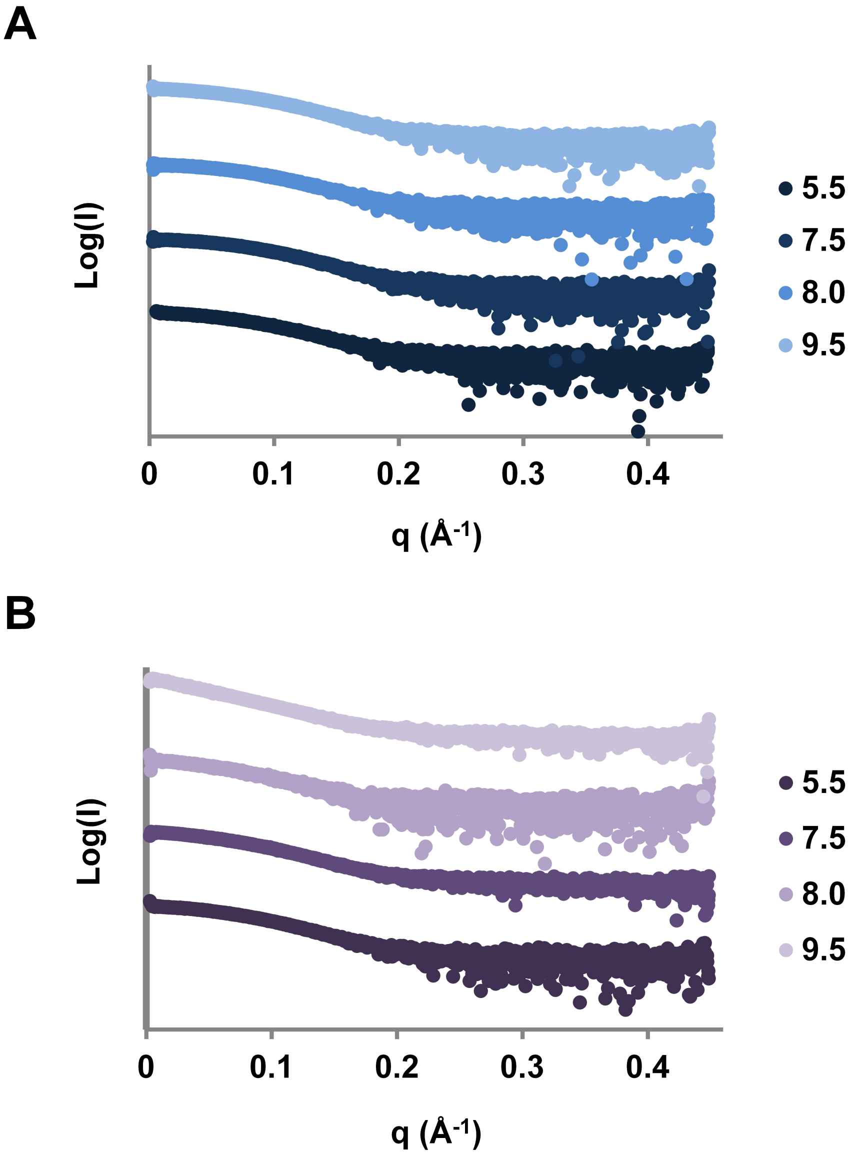


Figure 6.15: The stacked scattered intensities from the *Mm*PepT1 (A) and *Rn*PepT2 (B) ECD pH screens. The radial averaged intensities (I) were buffer subtracted and shown here plotted against the scattering vector length ($\boldsymbol{q}$).

scattering. As stated in § 6.6.1, inter-particle scattering is not taken into account during the processing of SAXS data, so it must be removed from the data if observed. No inter-particle effects were observed for any of the *Mm*PepT1$^{\mathrm{ECD}}$ samples. Some inter-particle effects were observed for *Rn*PepT2$^{\mathrm{ECD}}$ in the pH 8.0 and 9.5 samples and this was concentration dependent. In these two cases the low scattering angle portion of a low concentration curve, was merged with the high scatter angle intensities from a high concentration curve. Figure 6.15 shows the stacked scattering curves from the *Mm*PepT1 and *Rn*PepT2 ECD pH screens. All structural characteristics were inferred from these data.

Several particle parameters can be calculated from the scattered intensities. The $\mathrm{R_g}$ is a root mean squared distance between the centre of the particle mass and the particle circumference. The $\mathrm{R_g}$ therefore provides a measure of the particle size. The I(0) is the forward scattering from the particle and is dependent upon the size and concentration of the macromolecule. $\mathrm{V_c}$ was proposed by Rambo and Tainer (2013) as a new SAXS invariant, proportional to the I(0) and inversely proportional to the total scattered intensity. V_c reflects the flexibility of the sample, as an increase in particle conformations increases the I(0) without influencing the total scattered intensity. $\mathrm{V_p}$ is the particle volume and $\mathrm{D_{MAX}}$ is the maximum diameter of the scattering volume. The particle mass was calculated using the invariant $\mathrm{Q_R}$ method proposed in Rambo and Tainer (2013). Table 6.3 shows the SAXS invariants calculated *Mm*PepT1 and *Rn*PepT2 ECD.

	pH	Guinier R$_g$ (Å)	P(r) R$_g$ (Å)	Guinier I(0)	P(r) I(0)	V$_c$ (Å^2)	V$_p$ (nm^3)	Mass (kDa)	D$_{MAX}$ (Å)
***Mm*PepT1^{ECD}**	5.5	19.2	18.4	2200	1880	198.3	36.1	16.0	61
	7.5	18.3	18.1	2020	1960	197.0	34.5	17.2	57
	8.0	18.5	18.0	2120	1940	185.7	34.5	15.1	54
	9.5	19.0	18.1	2400	2190	202.2	39.2	17.5	54
***Rn*PepT2^{ECD}**	5.5	21.3	20.5	2880	2570	233.0	44.5	20.7	73
	7.5	25.2	22.5	3520	2960	256.5	54.0	21.2	73
	8.0	28.1	24.1	3430	2750	295.2	69.6	25.2	73
	9.5	46.7	43.2	7300	6150	490.6	119.0	41.9	154

Table 6.3: The SAXS parameters for *Mm*PepT1 and *Rn*PepT2 ECDs equilibrated at pH 5.5, 7.5, 8.0 and 9.5. These invariants were calculated using `ScÅtter` (Rambo and Tainer 2013).

The R_g and I(0) were calculated by the Guinier approximation (Guinier 1939) (Equation 6.2a) and inferred from the real space P(r) distribution. The Guinier approximation is only applicable if the low resolution scattered intensities regress to a straight line when plotted against the square of the scattering angle. The P(r) distribution is the integral of all the intra-molecular distance vectors between 0 and D_{MAX}. If there is a deviation between the real space and reciprocal space R_g and I(0), it suggests either, inter-particle scattering contributions, or poor buffer subtraction. This does not appear to be the case with the *Mm*PepT1 or *Rn*PepT2 ECD data, although there does appear to be an increasing difference in the *Rn*PepT2^{ECD} values as function of pH. The *Rn*PepT2 pH 9.5 data were excluding from further analysis due to the increase in R_g and D_{MAX} which suggested excessive inter-particle scattering.

$$\ln(\boldsymbol{q}) = \mathrm{I}(0) - \frac{\mathrm{R_g} \cdot \boldsymbol{q}^2}{3} \tag{6.2a}$$

$$\mathrm{V_p} = \frac{2\pi^2 \cdot \mathrm{I}(0)}{\int_0^\infty \boldsymbol{q}^2 \cdot I(\boldsymbol{q})\,\mathrm{d}q} \tag{6.2b}$$

Figure 6.16 shows the normalised P(r) distributions for the *Mm*PepT1 and *Rn*PepT2 ECD pH assays. *Mm*PepT1^{ECD} appears to be stable across the pH range. *Rn*PepT2^{ECD} however, shows a gradual decrease in the mid scattering distances and an increase in the large scattering distances. This change suggests that the *Rn*PepT2^{ECD} is increasing in size as a function of pH. This finding is also observable in the Kratky plot.

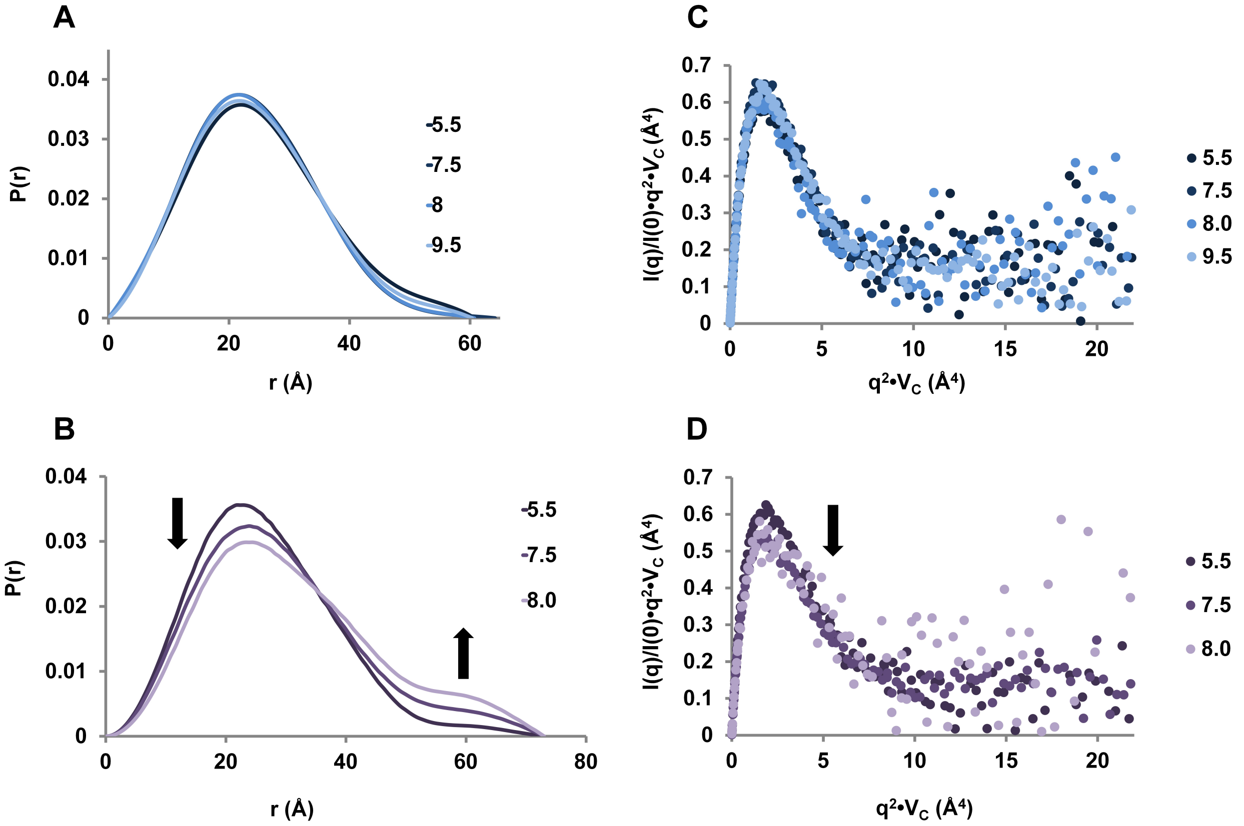


Figure 6.16: The normalised P(r) distribution functions and dimensionless Kratky plots for the *Mm*PepT1 and *Rn*PepT2 ECD SAXS pH curves. The curves were generated using `ScÅtter` (Rambo and Tainer 2013) and the *Rn*PepT2$^{\mathrm{ECD}}$ pH 9.5 data has been excluded because of sample aggregation. The P(r) distributions for *Mm*PepT1$^{\mathrm{ECD}}$ **(A)** and *Rn*PepT2$^{\mathrm{ECD}}$ **(B)** data were normalised against the sum of scattered distances. The P(r) distributions for *Mm*PepT1$^{\mathrm{ECD}}$ remain stable regardless of pH, whereas the *Rn*PepT2$^{\mathrm{ECD}}$ distributions show a marked decreased in scattering at 25 Å, accompanied by an increase at 60 Å, suggesting a more extended conformation. **(C)** and **(D)** show the dimensionless V_c Kratky plots for *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ respectively. Again the *Mm*PepT1$^{\mathrm{ECD}}$ plots remain constant, while the *Rn*PepT2$^{\mathrm{ECD}}$ plots show a decrease in the low scattering angles, indicating a more extended conformation.

	R_g (Å)	
Model	***Mm*PepT1**	***Rn*PepT2**
Extended	27.3	27.3
Open	20.0	20.0
Closed	17.5	18.3

Table 6.4: The calculated R_gs in `Hydropro` (Ortega, Amorós, and García de la Torre 2011) for the PepT1 and PepT2 ECD models shown in Figure 6.13.

The dimensionless Kratky plot was proposed by Durand *et al.* (2010) and developed by Rambo and Tainer (2011) and Receveur-Brechot and Durand (2012) as a measure of macromolecule flexibility. Figure 6.16 shows the V_c derived Kratky plot (Rambo and Tainer 2011) for *Mm*PepT1 and *Rn*PepT2 ECD. Over the pH range the *Mm*PepT1^{ECD} curves remain constant. For *Rn*PepT2^{ECD} there is a decrease in the intensities at 2.5 $q^2 \cdot V_c$ Å^4 as the pH is increased. Again, like the P(r) distribution, this suggests an increase in particle size.

SAXS invariant analysis

Table 6.4 shows the lobe model (Figure 6.13) R_gs calculated in `Hydropro` (Ortega, Amorós, and García de la Torre 2011). The R_g decreases as the lobes conformation becomes compacted in the closed conformation. The model R_g values, compared to the experimentally determined values (Table 6.3), suggest that *Mm*PepT1^{ECD} remains in the closed conformation regardless of the pH, which was expected given the linker mutation study (§ 6.5). However, there is a distinct trend of increasing particle size as a function of pH for *Rn*PepT2^{ECD} (Table 6.3). The real space R_g values increase from 20.5 to 24.1 Å from pH 5.5 to 8.0, which is consistent with a lobe conformation alteration from open to partially extended. The I(0), D_{MAX}, V_c and V_p also all increase as a function of pH while the mass remains reasonably constant, suggesting a structural rearrangement which increases the size of the ECD.

6.6.4 Analytical ultra-centrifugation

The SAXS data suggests that the *Rn*PepT2^{ECD} does increase in size as the pH is increased from pH 5.5 to 9.5, which was in keeping with the hypothesis that the *Rn*PepT2^{ECD} lobes, like the *Mm*PepT1$^{\mathrm{ECD}}$ lobes, are in the closed conformation, as the pH was able to disrupt the interaction between D505 and D600. To confirm that this was the case, a series of SV AUC experiments were performed using *Rn*PepT2$^{\mathrm{ECD}}$ D505A and D600A mutants. SV AUC, like SAXS, can measure the size of particles is solution. The SAXS showed that the WT *Rn*PepT2$^{\mathrm{ECD}}$ extends in a high pH environment, probably due to the proximity of D505 and D600. In the D505A and D600A mutants however, this extension should not be observed as there should not be an electrostatic potential to overcome.

*Rn*PepT2$^{\mathrm{ECD}}$ D505A and D600A mutants were created as described in § 2.6.6. The *Rn*PepT2$^{\mathrm{ECD}}$ mutant constructs were expressed in C43(DE3). 2 L of *E. coli* were grown and the cell lysates prepared as described in § 2.6.3 and the purification protocols for the mutants were the same as WT *Rn*PepT2$^{\mathrm{ECD}}$ (§ 3.5.3). Both mutants expressed at a similar level to WT, and the final GF traces and gels are shown in Figures 6.17.

SV experiments allow the calculation of the hydrodynamic properties of a macromolecule (Harding 1995; Cole *et al.* 2008; Erickson 2009). The movement of a macromolecules concentration boundary through a solution under a circular centrifugal force can be modelled by the Lamm equation (Equation 6.3a). The macromolecular sedimentation coefficient s is proportional to the buoyant mass ($\mathrm{M_b}$) and the translational diffusion coefficient (D) (Equation 6.3b) (Cole *et al.* 2008). Therefore the sedimentation coefficient contains information on the size and mass of a macromolecule.

$$\frac{\partial c}{\partial t} = D\left[\frac{\partial^2 c}{\partial r^2} + \frac{1}{r}\frac{\partial c}{\partial r}\right] - sw^2\left[r\frac{\partial c}{\partial r} + 2c\right] \tag{6.3a}$$

$$\frac{s}{D} = \frac{M_b}{RT} \tag{6.3b}$$

During the experiment, the concentration independent sedimentation coefficients ($\mathrm{s}^0_{20,w}$) from the WT, D505A and D600A *Rn*PepT2$^{\mathrm{ECD}}$ data were calculated from a concentration series of SV experiments (Harding 1995). The $\mathrm{s}^0_{20,w}$ sedimentation coefficient is normalised against sample concentration, temperature and buffer, therefore a change in the $\mathrm{s}^0_{20,w}$ coefficient represents a real change in the structure of the protein. The hypothesis of the experiment was that no change in the *Mm*PepT1$^{\mathrm{ECD}}$ $\mathrm{s}^0_{20,w}$ coefficients would be observed as a function of pH, whereas there would be an alteration in the $\mathrm{s}^0_{20,w}$ coefficients of the *Rn*PepT2$^{\mathrm{ECD}}$ WT as the pH was increased, and this change would be inhibited by the

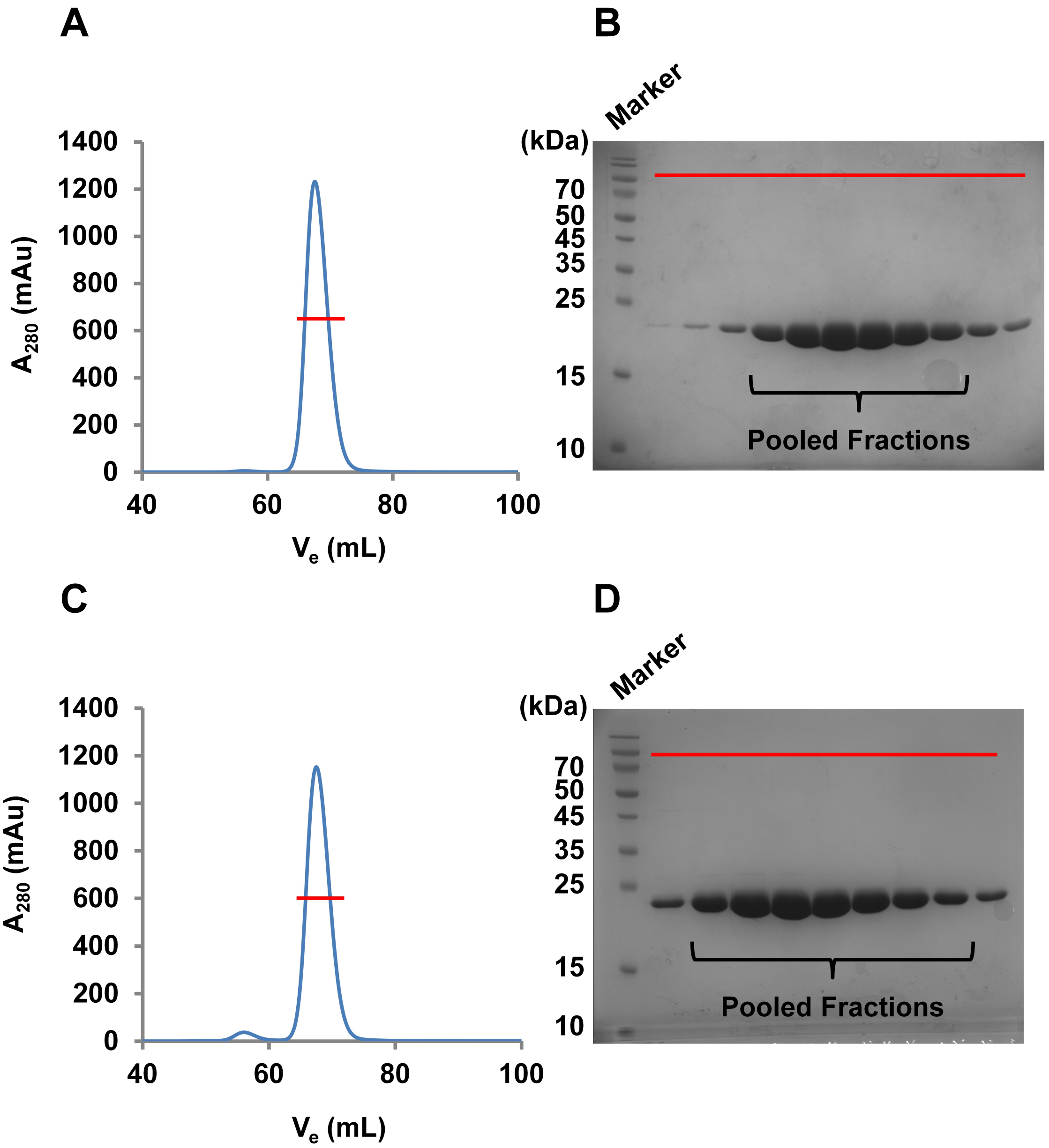


Figure 6.17: *Rn*PepT2$^{\text{ECD}}$ D505A and D600A mutant purifications. The *Rn*PepT2$^{\text{ECD}}$ D505A (**A** and **B**) and D600A (**C** and **D**) purifications. The final S75 16/60 GF traces (**A** and **C**) and peak fraction run through reducing 15% Tris-Gly SDS-PAGE (**B** and **D**).

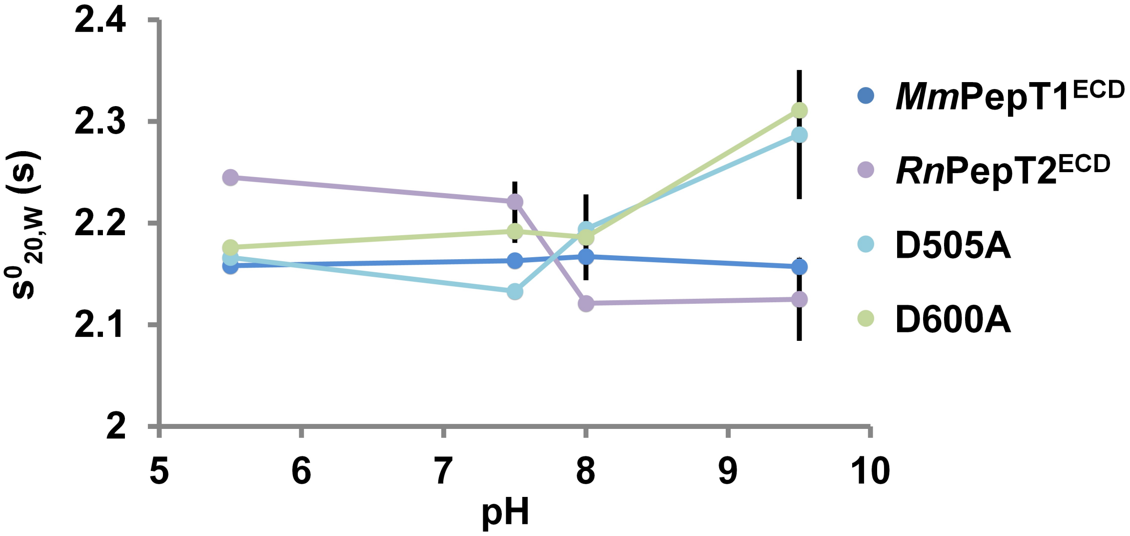


Figure 6.18: The $s^0_{20,w}$ plotted against pH for *Mm*PepT1^{ECD}, *Rn*PepT2^{ECD} WT, D505A and D600A. The error bars represent the regression error from the $s^0_{20,w}$ extrapolation. A change in $s^0_{20,w}$ represents a change in particle shape. The `Hydropro` (Ortega, Amorós, and García de la Torre 2011) predicted $s^0_{20,w}$ values in Table 6.5 suggest that an $s^0_{20,w}$ value of approximately 2.2 for the ECDs, represents a more closed lobe conformation. A reduction in $s^0_{20,w}$ to 1.9 represents a transition to a open lobe free state. *Mm*PepT1^{ECD} appears to remain stable as the pH is changed. *Rn*PepT2^{ECD} WT shows a decrease in $s^0_{20,w}$ as pH in increased, compared to the aspartate mutants which show an increase. These data suggest that the D505A and D600A mutants reverse the change from closed to more open state seen in the WT *Rn*PepT2^{ECD}.

D505A and D600A mutations.

AUC data analysis

The AUC data was collected and processed as described in § 2.8.4. Figure 6.18 shows the $s^0_{20,w}$ values for *Mm*PepT1^{ECD}, *Rn*PepT2^{ECD} WT, D505A and D600A plotted against pH. The *Mm*PepT1^{ECD} $s^0_{20,w}$ coefficients do not change as a function of pH. The *Rn*PepT2^{ECD} $s^0_{20,w}$ coefficient decreases from 2.22 to 2.1 *s* as the pH increases from pH 7.5 to 8.0. Both *Rn*PepT2^{ECD} mutants, D505A and D600A show a sharp increase in $s^0_{20,w}$ coefficient as pH is increased above 7.5.

`Hydropro` (Ortega, Amorós, and García de la Torre 2011) was used to calculate sedimentation coefficients (Table 6.5) for the 3 ECD model conformations (Figure 6.13). The *Mm*PepT1^{ECD} $s^0_{20,w}$ coefficients stayed at approximately 2.15 *s* regardless of the pH, suggesting that the ECD was in a closed conformation. The *Rn*PepT2^{ECD} WT $s^0_{20,w}$ coefficients shift as a function of pH. Given the trend of the values in Table 6.5 it is reasonable

	Sedimentation Coefficient (s)	
Model	***Mm*PepT1**	***Rn*PepT2**
Open	2.05	2.09
Closed	2.20	2.16
Extended	1.91	1.91

Table 6.5: The sedimentation coefficients calculated in `Hydropro` (Ortega, Amorós, and García de la Torre 2011) for the PepT1 and PepT2 ECD models shown in Figure 6.13.

to assume that the shift represents a transition to a more open conformation as observed in the SAXS experiment. The *Rn*PepT2$^{\text{ECD}}$ mutants, D505A and D600A also appear to reverse the change in $s^0_{20,w}$ coefficients seen in the WT. As the pH is increased the $s^0_{20,w}$ value also increases which suggests the ECD is becoming more compact. The fact that the aspartate mutants appear to prevent the change to the open conformation confirms the hypothesis that D505 and D600 mediate the flexibility observed in the ECD and that the *Rn*PepT2$^{\text{ECD}}$ crystal structure conformation (closed) is prevalent in solution.

6.7 Discussion

The ASUs of *Mm*PepT1$^{\text{ECD}}$ and *Rn*PepT2$^{\text{ECD}}$ presented an ambiguity in the ECD conformation which obstructed the aim to understand how this domain sits on the extra-cellular side of the TM domain. The *Mm*PepT1$^{\text{ECD}}$ crystallised as a TRP domain swapped dimer, where the TRP domains from adjacent monomers form a closed conformation, and the TRPs from the same monomer present in a more open conformation.

The *Rn*PepT2$^{\text{ECD}}$ crystallised in a conformation which brought two aspartate residues to within 3.5 Å of each other. This observation potentially explains pH dependent stability of the *Mm*PepT2$^{\text{ECD}}$ 410 – 609 construct observed in the FSC experiment (§ 5.2.1) and the low pH crystallisation condition; as the low pH was essential to stabilise the proximity of D505 and D600. The comparison of the lobe orientations between the crystal structures highlighted the two potential lobe conformations present in the *Mm*PepT1$^{\text{ECD}}$ ASU, and was the first indication that the favourable PepT1 conformation was between adjacent monomers (closed) like the *Rn*PepT2$^{\text{ECD}}$ monomer.

The `Dali` structural search indicated the lobes shared a common fold with TTR and

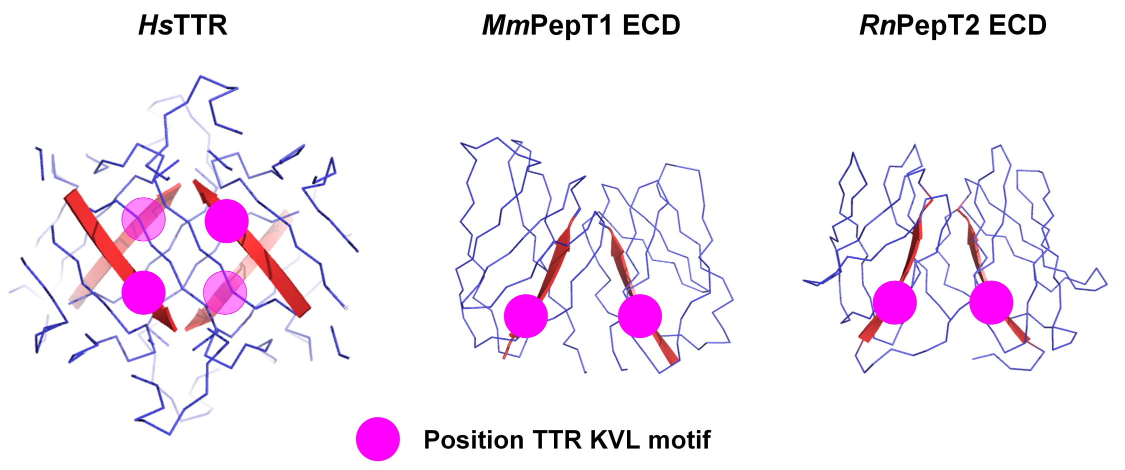


Figure 6.19: A comparison between the ECDs in the closed conformation and *Hs*TTR as a homo-tetramer. *Hs*TTR homo-tetramer was generated in `PISA` (Krissinel and Henrick 2005) using PDB: 1ROX (Wojtczak *et al.* 1996). The βA sheet has been highlighted in red to show the TRP domain orientations in respect to each other and the location of the TTR KVL motif is show as a purple spot.

TRPs. Speculation as to the function of the ECDs is discussed in Chapter 7, but a structural link between the ECD and a soluble hormone binding protein is interesting. It has been observed from other membrane proteins that the eukaryotic membrane proteome has borrowed structural elements extensively from the soluble proteome (Nam *et al.* 2013). In this case its possible that the peptide transporters have linked together two TRP domains, and inserted them into the middle of an MFS fold.

As shown in § 6.3 there is little sequence conservation between the TRPs and the ECD lobes (15% identity). Assuming that the ECD in the closed conformation represents a dimerisation of TRP domains, this dimerisation is novel in appearance. Figure 6.19 shows the *Hs*TTR and ECDs with the βA sheet highlighted in red. In the *Hs*TTR homo-tetramer each adjacent TRP domain runs anti-parallel with respect to each other. In the ECDs the TRP domains are parallel. However, it still appears that the same residues which are involved in T4 binding in TTR (KVL motif), and therefore its tetramerization, are also involved in mediating the closed conformation interface.

Ultimately the TRP analysis indicated that given the lack of residue conservation and the lobe conformation, it was unlikely that TTR and the lobe share a common function as well as a common fold. The comparison between the two ECD sequences did however highlight the common residues involved in the closed conformation interface and hence the closed conformation, as the likely lobe interface. The linker mutation study (§ 6.5) showed that the two salt bridges which are highly conserved PepT1, maintain the closed conformation in PepT1$^{\mathrm{ECD}}$.

The SAXS (§ 6.6.3) and AUC (§ 6.6.4) data confirmed the linker mutation study result in *Mm*PepT1$^{\mathrm{ECD}}$. The SAXS data also indicated that the *Rn*PepT2$^{\mathrm{ECD}}$ presented a closed conformation at low pH and became increasingly dynamic as a function of pH. This was also observed in the AUC data. The *Rn*PepT2$^{\mathrm{ECD}}$ D505A and D600A mutants prevented this increase in particle size, which indicated that the conserved aspartate residues in PepT2 mediate the pH dependent flexibility.

The reason for conducting this conformation analysis was ultimately to determine which lobe arrangement to use when building the hybrid models of the complete transporters. It should not be forgotten that the ECDs physiological environment is embedded into a membrane transporter, and this environment will perhaps influence the lobe conformation in unpredictable ways. However, the hybrid model usefulness will be diminished if the data used to construct it is poor. Based on the evidence presented in this chapter, it is highly probable that the closed conformation is exhibited by both PepT1 and PepT2 ECD.

7

Investigations into the function of the ECDs

7.1 Summary

It was not clear from the literature if the ECD has a role in the transport cycle. To help in the design of these experiments, hybrid models of *Hs*PepT1 and *Hs*PepT2 were constructed using the TM domain of $\mathrm{PepT_{So}}$ and the ECD closed conformation models. The hybrid models indicate that the ECDs sit above the TM domain of the transporter in such a way that the ECDs could influence the transport properties of PepT1 and PepT2. To discover if the ECD is essential for peptide transporter function, WT, ECD deleted, and ECD swapped mutants were assayed in *X. laevis* oocytes and the transport capacity was measured using the TEVC technique. The results indicated that $\mathrm{PepT2^{ECD}}$ was not required for transport, and that the transporter still functioned in a PepT2-like manner. No data could be gained from PepT1 ECD

mutants as these failed to express in the oocyte system. This could indicate that the ECD of PepT1 is required for its stability, but this has not been demonstrated conclusively. Due to the ECDs location above the TM domain, it was hypothesised that the ECD might act as a binding platform for an intestinal protease. SPR binding experiments were performed with *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} using two intestinal proteases, trypsin and α-chymotrypsin. *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} were both found to interact with trypsin with K_ds of 90 ± 20 and 170 ± 30 µM, significantly less, at the 0.05 criterion (n = 3), than the control protein MBP. A mutational analysis was conducted on *Mm*PepT1^{ECD} to definitively locate the trypsin binding site. Unfortunately this could not be comprehensively achieved, but this work has identified the probable location which could be conserved between PepT1 and PepT2.

7.2 Regulation of peptide transporter function

At the beginning of this project it was hoped that the solving of the ECD structures would suggest a probable function for the ECD. On solution of the crystal structures of *Mm*PepT1 and *Rn*PepT2 ECD it was found that both lobes of the ECD share the TRP fold. TTR is a T4 binding protein, so it was a reasonable hypothesis that the ECD also acts as a binding protein. The ECD is attached to the TM domain via TM 9 and 10. TM 10 has been shown to be involved in the closing of the intra-cellular gate in the bacterial transporters (Newstead 2014). So if the ECD does bind a ligand, the binding event could mediate the transport process (Figure 7.1). If this is the case, the peptide transporter would be the first documented ligand gated secondary active transporter.

Evidence has been found that a number of different molecules can down regulate either transporter expression or activity. Ashida *et al.* (2002) identified that T4 reduced PepT1 expression in Caco-2 cells. Phorbol esters were also found to inhibit GlySar transport in PepT1 transfected Caco cells (Brandsch *et al.* 1994). The inhibitory effect of phorbol esters was determined to be mediated by protein kinase C and could be prevented by using staurospourine, a PKC antagonist. Zn^{2+}, Fe^{2+} and Cu^{2+} were found to reduce the apparent GlySar affinity without changing the V_{MAX} indicating that the metal compounds were

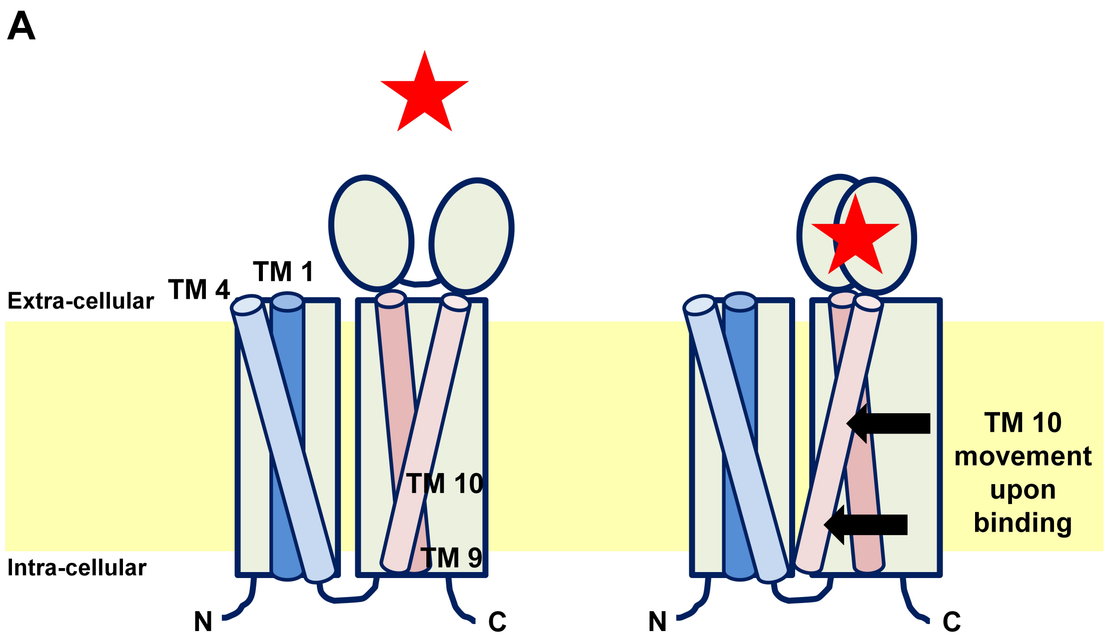


Figure 7.1: Hypothetical binding of a ligand to the ECD inducing a conformational change in TM 10 and blocking the pore.

acting like competitive inhibitors (Okamura *et al.* 2003). The opposite effect was found in kidney brush boarder membrane vesicles treated with Zn^{2+} solution (Daniel and Adibi 1995). It is not clear how these ions affect transporter expression or activity however, they suggest that higher level control of the peptide transporters is possible.

Several compounds also were found to increase peptide uptake in model systems. Insulin was found to increase the population of the PepT1 in the plasma membrane (Thamotharan *et al.* 1999). Colchicine inhibited the insulin effect, suggesting the increase was due to a draining of the cytoplasmic pool in the early endosomes. The same increase in plasma membrane PepT1 was observed for α2-adrenergic receptor agonists (Berlioz *et al.* 2000) and Leptin, a gastric hormone which stimulates digestion (Buyse *et al.* 2001). Flavonoids were also found to increase transport via the epidermal growth factor receptor (Wenzel, Kuntz, and Daniel 2001).

Zinc is the only compound which is documented to interact with the transporter directly, potentially through histidine residues in the extra-cellular loop between TM helices 3 and 4 (Okamura *et al.* 2003). However, the inhibitory and stimulatory effect which was observed in PepT1 and PepT2 respectively, may still be mediated by other factors in the experiment. No interaction could be determined between Zn^{2+} or T4 with either of the

	***Hs*PepT1 (%)**	***Hs*PepT2 (%)**
PepT$_{\mathrm{So}}$	30.81	30.81
PepT$_{\mathrm{St}}$	21.33	22.77
***Gk*POT**	22.98	21.17
PepT$_{\mathrm{So2}}$	16.25	16.63
***Ec*GlpT**	13.30	13.97
***Ec*LacY**	12.23	10.79

Table 7.1: Percentage sequence identity between the bacterial MFS structures and *H. sapiens* PepT1 and PepT2. The table shows that the bacterial peptide transporters share greater sequence identity with *H. sapiens* PepT1 and PepT2 than GlpT and LacY, and that PepT$_{\mathrm{So}}$ is the best structure to used when modelling the TM helices of PepT1 and PepT2. The sequence identity was calculated in `ClustalW` (Larkin *et al.* 2007). The full names and Uniprot codes for each gene are listed in Table 2.1.

ECDs, by Isothermal titration calorimetry (ITC) and SPR respectively. However, the idea that the peptide transporters do act as ligand mediated transporters is still possible and if they do act as 'transceptors' the ECD is the obvious place for a ligand to bind. The chimera studies (described in § 1.5.1) suggest that the ECD is not required for function, but this fact was not absolutely determined. To better understand the structure of the entire transporter, hybrid models of *H. sapiens* PepT1 and PepT2 were created.

7.3 Hybrid model construction

Table 7.1 shows the sequence identity between *H. sapiens* PepT1 and PepT2, and the bacterial MFS structures. PepT$_{\mathrm{So}}$ is the best current model of the TM helices of both PepT1 and PepT2. The X-ray crystal structures of PepT1 and PepT2 ECD from this thesis, for the first time provide the structural information necessary to construct relatively complete models of PepT1 and PepT2.

7.3.1 Method

The hybrid models of *H. sapiens* PepT1 (Uniprot: Q16348) and PepT2 (Uniprot: P46059) were built in conjunction between the author and Firdaus Samsudin (Structural Bioinformatics and Computational Biochemistry Unit) using the structure of PepT$_{\mathrm{So}}$ (Uniprot: Q8EHE6, PDB: 2XUT) as a template for the TM domain and the closed conformation of ECDs from the *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ crystal structures. Analysis of the bacterial peptide transporters indicated that there are a series of salt bridges which stabilise distinct states in the alternating access mechanism proposed for MFS solute transport (Law, Maloney, and Wang 2008; Solcan *et al.* 2012; Guettou *et al.* 2013; Newstead 2014; Lyons *et al.* 2014). Parker and Newstead (2014) and Sun *et al.* (2014) also found evidence for a similar mechanism in an *A. thaliana* nitrate transporter NRT1.1, which forms part of the PTR superfamily.

Figures 7.2A and B show the four salt bridges which line the substrate channel through $\mathrm{PepT_{So}}$ (previously identified in Figure 1.13). Figure 7.2C shows transport mutational data for $\mathrm{PepT_{So}}$ and $\mathrm{PepT_{St}}$ adapted from Solcan *et al.* (2012) and Solcan (2013). Apart from $\mathrm{PepT_{So}}$ NB2, which when mutated to alanine increased its transport activity, all other salt bridge alanine mutants either eliminate transport or significantly reduce transport activity. So it seemed sensible to identify these salt bridge residues in PepT1 and PepT2 to act as points of reference in the $\mathrm{PepT_{So}}$ alignment (Figure 7.3).

The initial alignment of PepT1 and PepT2 against $\mathrm{PepT_{So}}$ was generated using `Probcons` (Do *et al.* 2005) and manually refined in `Jalview` (Waterhouse *et al.* 2009). Figure 7.2D shows the salt bridge residues in $\mathrm{PepT_{So}}$ which were aligned to the comparable residues in PepT1 and PepT2. Based on the asymmetric distribution of charge and hydrophobic residues, Xu *et al.* (2010) predicted two amphipathic helices in the intra-cellular loop between TM helices 6 and 7. These helices were modelled by the secondary structure prediction in Xu *et al.* (2010) and the two extra TM helices observed in $\mathrm{PepT_{So}}$, HA and HB (residues 226 to 285) were removed.

Based on the alignment shown in Figure 7.3, `Modeller-9v5` (Sali and Blundell 1993) was used to generate 100 homology models of *Hs*PepT1 and *Hs*PepT2 and ranked according to their discrete optimised protein energy scores (Eramian *et al.* 2006). No structural information was available for the N-terminus, residues 1 – 14 and 1 – 42 in PepT1 and PepT2 respectively, and the C-terminus, residues 688 – 708 and 713 – 729 in PepT1 and PepT2 respectively, so these residues were removed from the models. Each model was assessed by Ramachandran analysis using `Rampage` (Lovell *et al.* 2003) (Figure 7.4) and validated in `Molprobity` (Chen *et al.* 2010) (Table 7.2). The overall `Molprobity` scores, shown in Table 7.2, indicate that in general the hybrid models of *H. sapiens* PepT1 and PepT2 are reasonable protein models (Figure 7.5).

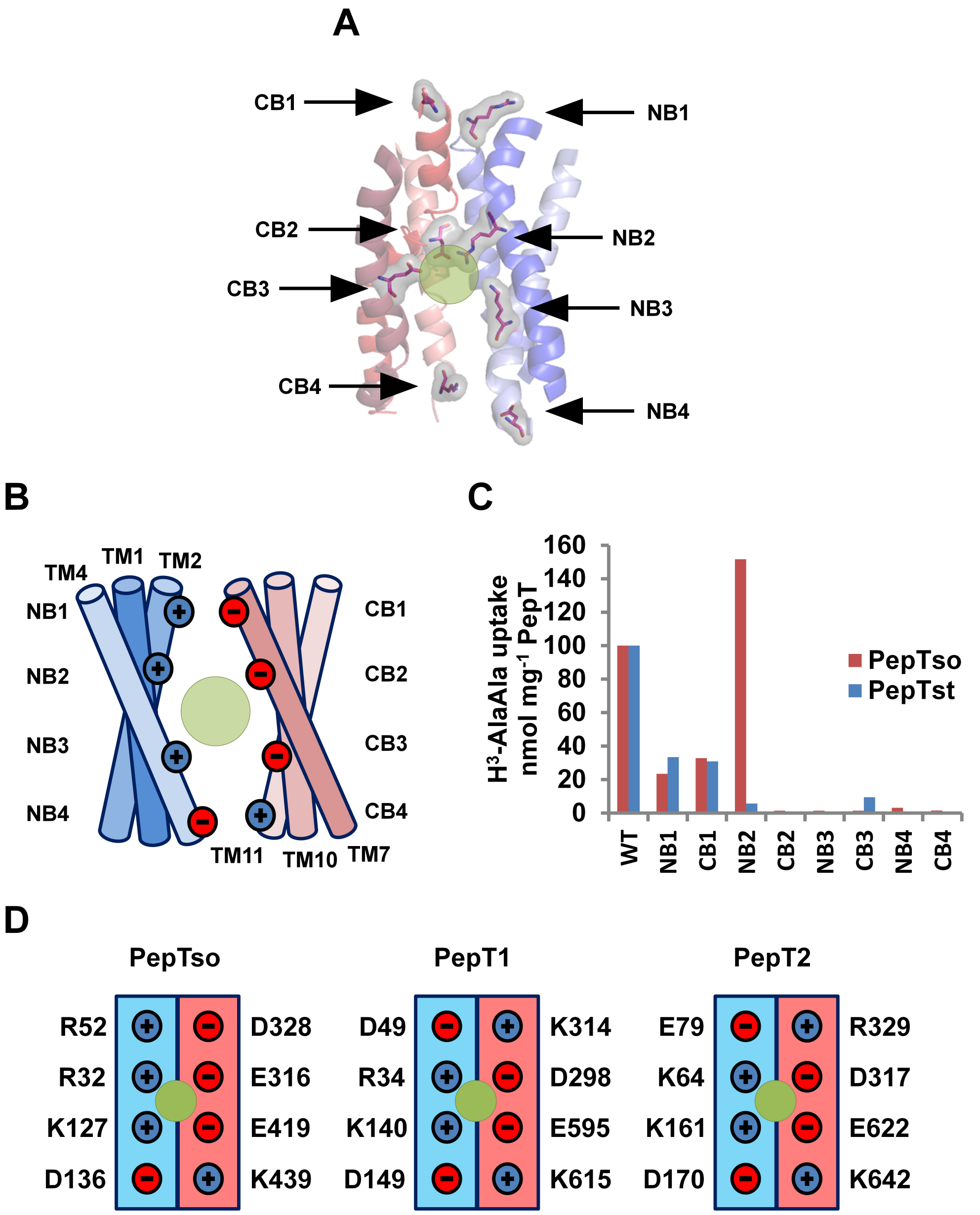


Figure 7.2: The conserved salt bridge residues from bacteria to humans. This figure has been adapted from Figure 1.13. The N- and C-terminal bundles of the transporters have been coloured blue and red, and the binding site is shown as a green circle. **(A)** and **(B)** show the four PepT$_{So}$ salt bridges. **(C)** The PepT$_{So}$ and PepT$_{St}$ liposome mutational data, adapted from Solcan *et al.* (2012) and Solcan (2013) using H^3-AlaAla as a substrate. **(D)** The relative salt bridge residues in *H. sapiens* PepT1 and PepT2 based on the alignment in Figure 7.3.

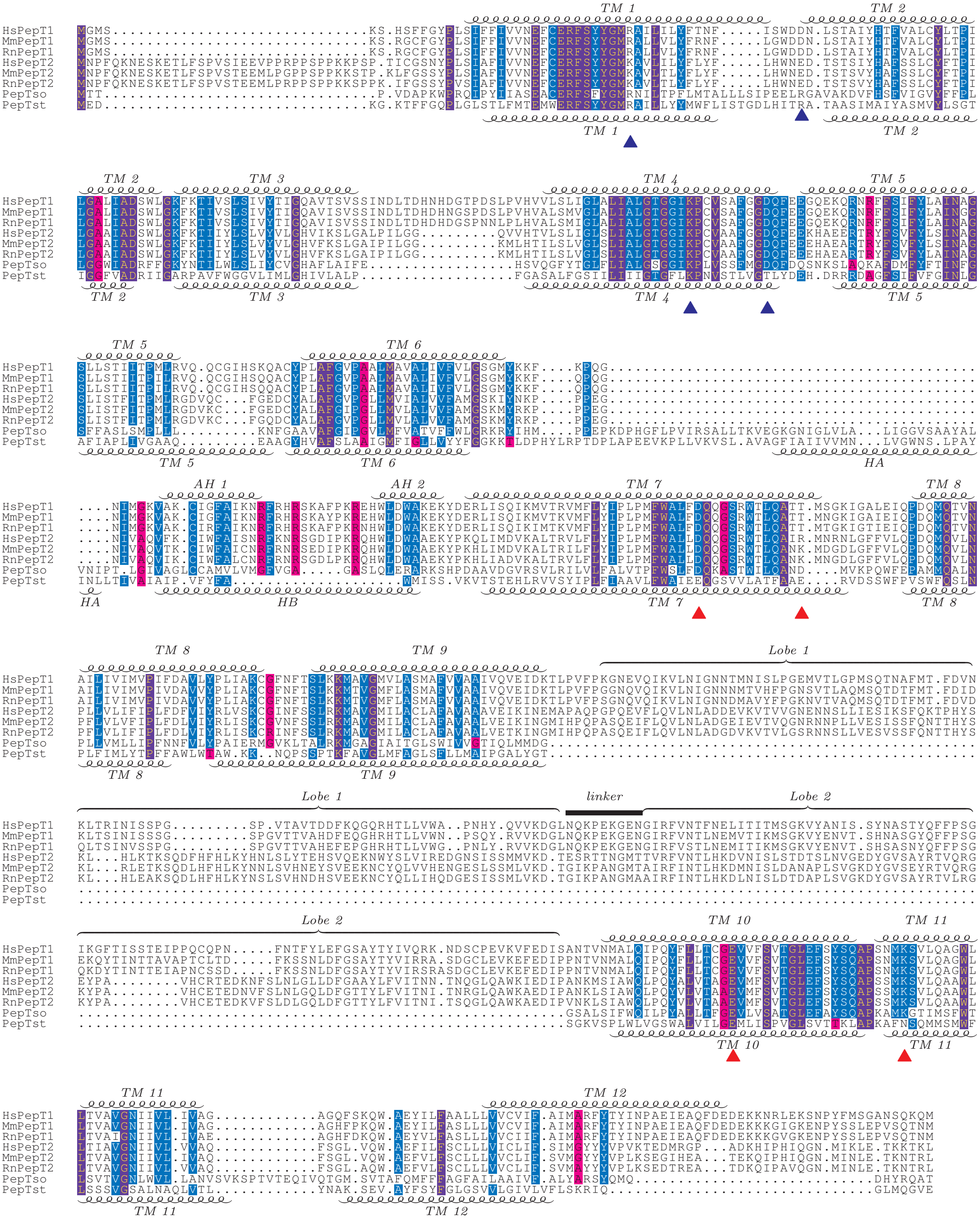


Figure 7.3: The alignment of the mammalian and bacterial peptide transporters used for the construction of the hybrid models. The full names and Uniprot codes for each gene are listed in Table 2.1. Sequence conservation is coloured from high to low in purple to blue. The secondary structure of the *Hs*PepT1 hybrid model and PepT$_{\mathrm{So}}$ is drawn on the top and bottom of the alignment respectively. The N- and C-Terminal salt bridge residues from Figure 7.2 are highlighted by blue (▲) and red (▲) triangles respectively.

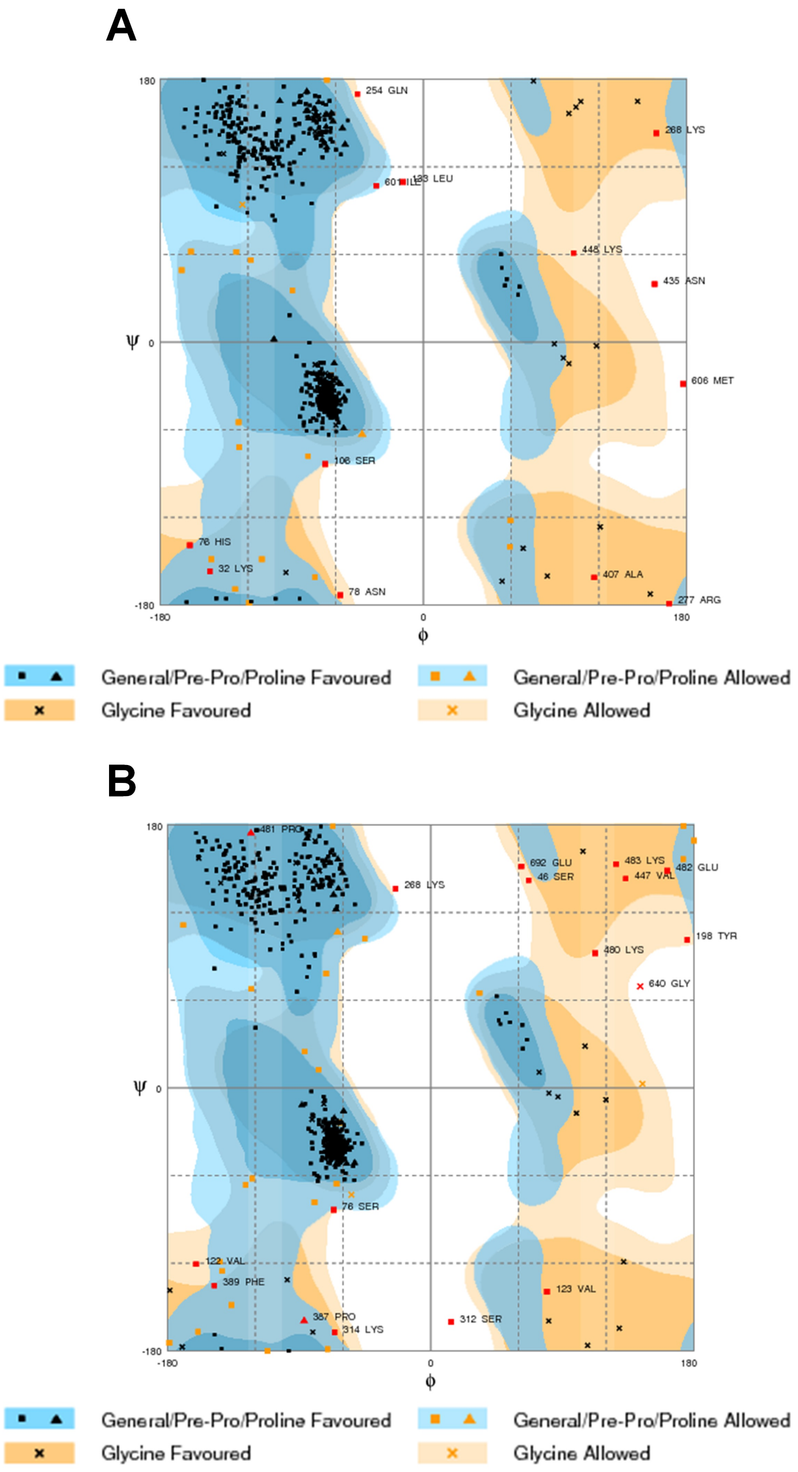
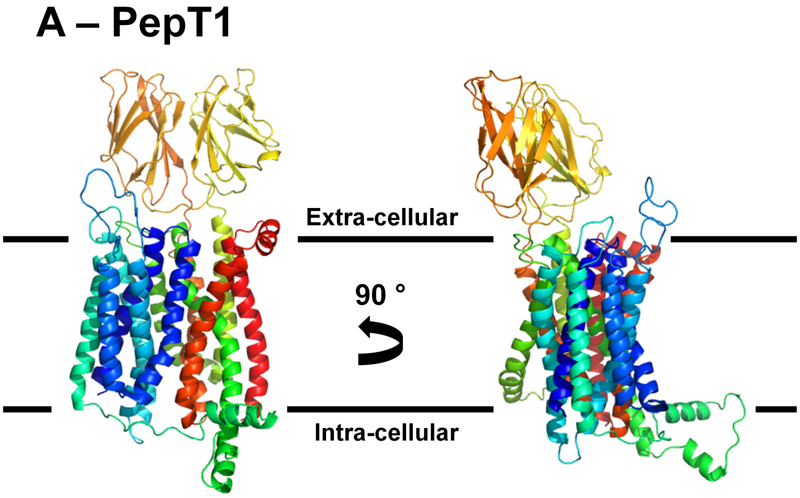


Figure 7.4: Ramachandran plots of the hybrid models. `Rampage` (Lovell *et al.* 2003) was used to generate the plots of the *H. sapiens* PepT1 **(A)** and PepT2 **(B)** structures.


A – PepT1

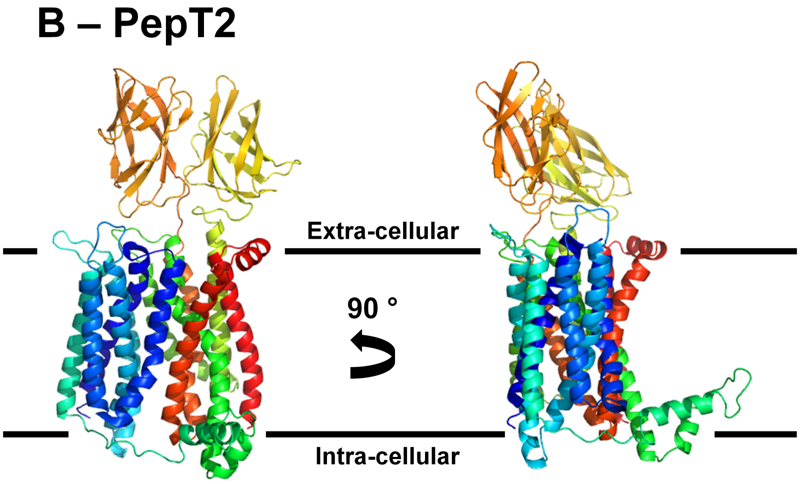

B – PepT2

Figure 7.5: The hybrid models of *H. sapiens* PepT1 (A) and PepT2 (B). The transporters have been coloured by rainbow from the N- to the C-terminus.

		Model	
		***Hs*PepT1**	***Hs*PepT2**
Ramachandran Plot (%)	Favoured	93.20	95.02
	Allowed	4.11	3.44
	Disallowed	2.69	1.51
Rotomers (%)		4.11	3.92
Cβ deviations (%)		2.27	1.31
Backbone bond deviations (%)		0.02	0.03
Backbone angle deviations (%)		1.79	1.48
Molprobity (Percentile)	Clashscore	0	0
	Overall	11	15

Table 7.2: `Molprobity` (Chen *et al.* 2010) analysis of the hybrid models.

	$\mathbf{T_m}$ (°C)
***Mm*PepT1$^{\mathbf{ECD}}$ WT**	58.8 ± 0.3
***Mm*PepT1$^{\mathbf{ECD}}$ D476A**	44.3 ± 0.3
***Mm*PepT1$^{\mathbf{ECD}}$ D574A**	38.0 ± 0.9
***Rn*PepT2$^{\mathbf{ECD}}$ WT**	44.4 ± 0.7

Table 7.3: The $\mathbf{T_m}$ of the WT and mutant ECDs at pH 7.5 as determined by CD thermal melt. The buffer used in the experiment was 10 mM $NaPO_4$ pH 7.5 and the methods are described in § 2.8.5. The limits shown are the 95% confidence limits in the fit of the curve to the data points.

7.3.2 Hybrid model validity

The areas of the structures which are likely to be incorrect are the areas for which there is no structural information *i.e.* the N- and C-terminal residues. In particular, the PepT2 N-terminus is approximately 40 residues which is highly conserved in mammalian species (Figure 7.3) and is therefore likely to have a functional relevance. A `Molprobity` analysis (Table 7.2) indicated that the models are not good in terms of their bond angles and distances, which is an indication that there are potentially some errors in the structures. That said, the models give the most complete interpretation of the structure of PepT1 and PepT2 and were used to help determine the role played by the ECD in the transporter cycle.

7.4 Investigating the role of the ECD in the transport cycle

The mutant and chimeric transporters constructed are shown in Figure 7.6A. The ECDΔ (ECD deleted) and ECD swapped mutants were designed to determine whether the specific deletion of the ECD from both PepT1 and PepT2 effected the apparent $\mathrm{K_m}$ compared to the native transporter. The instability observed in the *Mm*PepT1$^{\mathrm{ECD}}$ when D476 and D574 were mutated to alanine was also investigated. The alanine mutations, which remove the salt bridge interactions between the lobes were found to reduce the $\mathrm{T_m}$ of the *Mm*PepT1$^{\mathrm{ECD}}$ D476 and D574 mutants by approximately 15 and 20 °C respectively (Table 7.3). Therefore to ascertain if the stability of the ECD is essential for WT transporter function, two *Hs*PepT1 mutants were constructed; a single mutant *Hs*PepT1-D573A, and a double mutant *Hs*PepT1-D476A/D573A (D476 and D574 in the *M. musculus* sequence are D475 and D573 in *H. sapiens*). The WT and mutant transporters were assayed for function in *X. laevis* oocytes using GlySar as a substrate and transporter activity was determined using the TEVC method. All experiments and data processing were conducted by Anne Barrett (Department of Biochemistry, University of Oxford) in the lab of Dr Stephen Tucker (Department of Physics, University of Oxford) (see § 2.8.6 for methods).

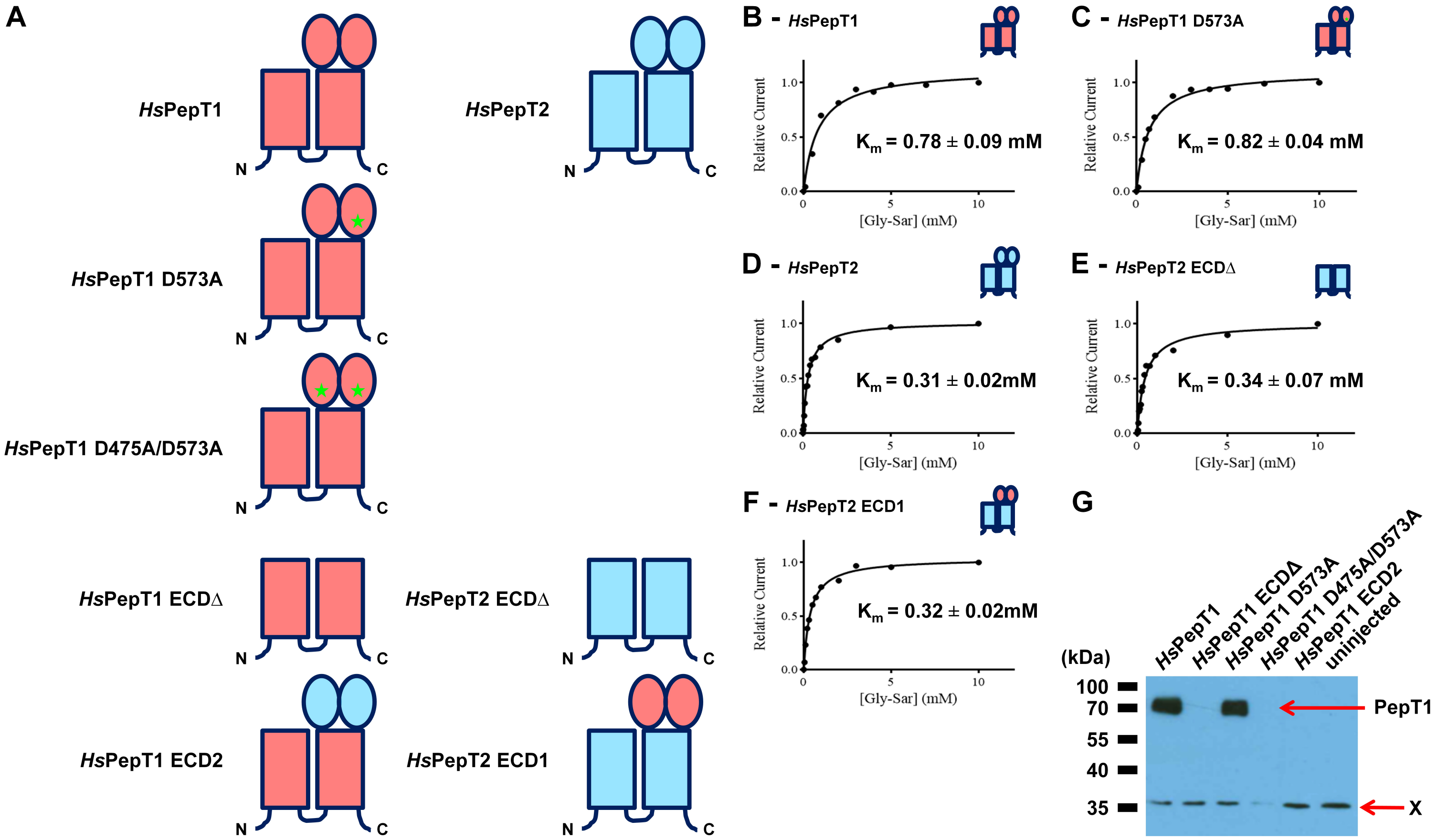


Figure 7.6: The composition and results of mutant and chimeric PepT1 and PepT2 transporters constructed to investigate the role of the ECDs in the transport cycle. Elements from PepT1 and PepT2 are highlighted in red and blue respectively. The green star annotate an aspartate to alanine mutation of the lobe salt bridge aspartate residue. **(B)-(F)** show the apparent K$_m$ curve fits for the respective WT, chimera and mutant transporters. The approximate K$_m$s of the WT transporters were assumed to indicate a PepT1 or PepT2 'like' property. The *Hs*PepT1-D573A mutant gives a similar K$_m$ to WT *Hs*PepT1, indicating that the aspartate mutant does not effect the transport properties of the PepT1. The *Hs*PepT2-ECDΔ and ECD1 mutants both give K$_m$s similar to WT, again suggesting the mutants do not effect transport. **(G)** Shows a reducing 10% Tris-Gly anti-FLAG western blot showing that expression was observed for the *Hs*PepT1 WT and D573A mutant transiently transfected oocytes. No expression was observed for the *Hs*PepT1 ECDΔ, ECD2 chimeras or the D475A/D573A mutant. **X**=unknown band.

Results

Figure 7.6B to F show the results of the *X. laevis* oocyte TEVC experiments. Figure 7.6G shows an anti-FLAG western blot of the PepT1 mutant constructs. Of the *Hs*PepT1 experiments attempted only the WT protein and D573A mutant expressed in the transiently transfected *X. laevis* oocytes. Figures 7.6B and C show that the *Hs*PepT1 WT and D573A mutant gave comparable apparent K_m values, indicating that the D573A mutant, does not effect transport.

Figures 7.6D, E and F show the results from the *Hs*PepT2 WT, ECDΔ and ECD1 chimeras experiments respectively. The graphs show that no change was observed in GlySar apparent K_m for the ECDΔ and ECD1 chimeras compared to WT. Given the stable membrane insertion of the *Hs*PepT2 ECD1 chimera, it is tempting to assume that the lack of *Hs*PepT1 ECD2 chimera expression in *X. laevis* oocytes, was due to protein instability. This would indicate that the PepT1^{ECD} is essential for protein stability, whereas in PepT2 the ECD is not.

It is also possible that the *Hs*PepT1 ECDΔ and ECD2 mutants, were constructed in such a way that prevented stable protein folding. PepT1^{ECD}, like PepT2^{ECD} may as the crystal structure suggests, be an entirely separate autonomous domain of the transporter, which is not required for transport. In some respects the *Hs*PepT1-D573A mutant result confirms this view, as the *Mm*PepT1^{ECD}-D574A mutant is unstable, so the ECD in the *Hs*PepT1-D573A construct is also likely to be unstable. Therefore if the WT *Hs*PepT1^{ECD} is required for transport, a reduction in the apparent K_m would have be expected from the *Hs*PepT1-D573A construct, but none was observed. It may be the ECD is required for the stable folding and insertion of PepT1 into the membrane. However, if this was only the case then the double mutant, D475A/D576A should also have produced folded protein, which it did not.

In summary, the *X. laevis* oocyte TEVC assays have demonstrated that PepT2^{ECD} is not required for the transport of GlySar by PepT2; furthermore that the ECD does not contribute to the PepT2 'like' transport properties of PepT2. For PepT1, based on this experiment, the story is still conflicted. There is an indication that the ECD is not required for transport, but that the ECD is important for the stability of the transporter in some way which, as yet, is still to be determined. Based on this work, it seemed reasonable to look for a primary function of both PepT1 and PepT2 ECD outside of the transport cycle.

7.5 Proteolytic cleavage of dietary protein

The digestion of dietary protein, introduced in § 1.2.1, is a highly regulated process. Upon entering the stomach, the protein component of food is immersed in a bath of acid (pH 2.0) and attached by pepsin, the first in a series of proteolytic enzymes (Berne and Levy 2008a). The H^+ concentration helps to unfold the dietary protein chains so that pepsin can cleave the peptide chains. Pepsin has a specificity for the large hydrophobic side chains, leucine, tryptophan, phenylalanine and tyrosine (Fruton and Bergmann 1939), and cleaves the downstream peptide bond. The partially digested peptide chains are retained in the stomach until non-solid (*chymus*); approximately 4 hr, depending on how well the food is homogenised (Hinder and Kelly 1977). The peptides are then passed to the small intestine. The length of these peptide chains is dependent upon the residue composition of the original protein. Based on the pepsinisation of myosin, the principle component of muscle tissue, 85% of amino-acids remain in peptides greater than 5 residues in length upon entry into the small intestine.

At the point of entry into the small intestine, the *chymus* is perfused with HCO_3^-, to neutralise the acidic pH, and 5 pancreatic proteases, trypsin, α-chymotrypsin, elastase, and carboxypeptidase A and B (Berne and Levy 2008b) are excreted from the pancreas. These peptidases are secreted as pro-enzymes which become activated upon proteolytic cleavage. Typsinogen, the trypsin pro-enzyme, is activated by enterokinase (Tarlow *et al.* 1970), a protease from the same family as α-chymotrypsin, which is anchored to the intestinal brush boarder membrane. Enterokinase cleaves an N-terminal fragment from the trypsinogen, activating trypsin, which in turn activates the other pancreatic proteases (Berne and Levy 2008b). The significant role that the pancreatic peptidases play in digestion cannot be understated. Individuals who possess an inactive form of enterokinase suffer from severe malnutrition and fits of vomiting (Tarlow *et al.* 1970).

The pancreatic enzymes work in concert to break down the peptide content of the *chymus* into di- and tri-peptides, and free amino-acids (Berne and Levy 2008b). Webb, Matthews, and DiRienzo (1992) estimated that 79% of amino acid absorption occurs in the form of peptides, which means that PepT1 is the principle route for the absorption of dietary protein. That said, PepT1 null (PepT1$^{-/-}$) mice are phenotypically WT under standard living conditions (Hu *et al.* 2008). This suggests that the dietary protein requirement can be met by complete peptide hydrolysis, and subsequent amino-acid absorption by lumenal amino-acid transporters. However, the PepT1$^{-/-}$ mice were impaired at absorbing protein compared to WT when given high protein diets (Nässl *et al.* 2011). It is probable that phenotypic deficiencies would be more apparent in animals which require a high protein diet. This can be seen in *C. elegans* PepT1$^{-/-}$ knock outs which show a marked reduction in growth and fertility (Meissner *et al.* 2004).

7.5.1 The ECD as a protease binding protein

Considering that di- and tri-peptides, the potential products of a peptide cleavage reaction, from trypsin and α-chymotrypsin, are the substrates of PepT1. It is not an unreasonable hypothesis that there may be an interaction between the proteases and the peptide transporter, which aids the presentation of substrate to the transporter. Some evidence for transporter and protease binding does already exist. Kowalczuk *et al.* (2008) found that the stability of the amino acid transporter, B_0AT1 (SLC6A19) was significantly increased by the co-expression of dipeptidyl carboxypeptidase I. If there was an interaction between a protease and PepT1, the ECD is the obvious place for the interaction to occur.

Figure 7.7 shows the electrostatic surfaces of the hybrid models of *H. sapiens* PepT1 and PepT2. It is not obvious from inspection of the models where a potential interaction site might reside. The sequence analysis of the ECDs (§ 6.3) indicated that the only area of the ECDs which shows significant sequence conservation is the lobe interface. It is possible that the ECD, in binding a protease acts like a clam shell, and captures the protease between the lobes. The SAXS and AUC analysis (§ 6.6) suggests this is more probable for PepT2 than PepT1, as PepT2 proved to have a more dynamic lobe interface. It may also be, that the ECD works with the TM domain to capture the protease. In PepT1, the unstructured extra-cellular loop between TM helices 3 and 4, which protrudes from the adjacent side of the transporter and may act as the other half of binding cleft.

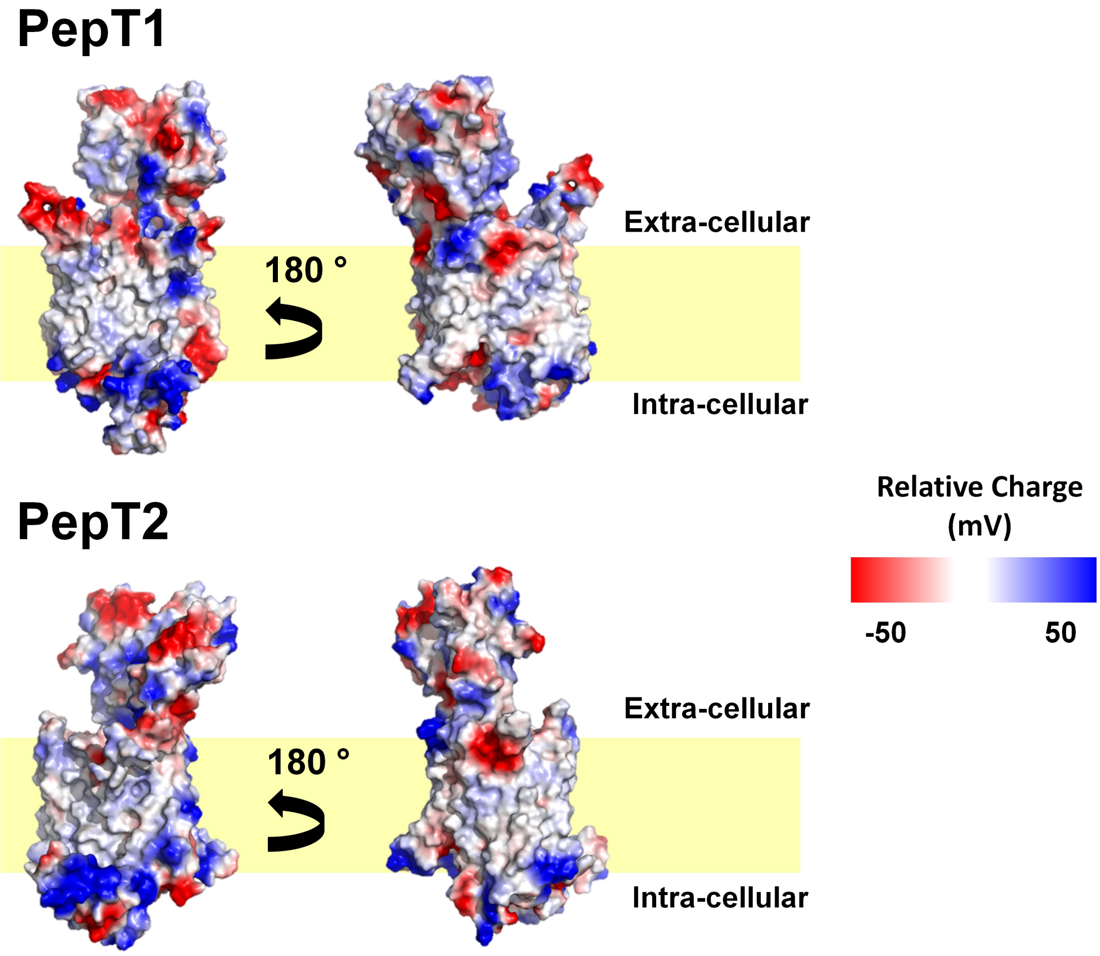


Figure 7.7: The electrostatics of *Hs*PepT1 and *Hs*PepT2. The electrostatic surfaces were generated using Baker *et al.* (2001).

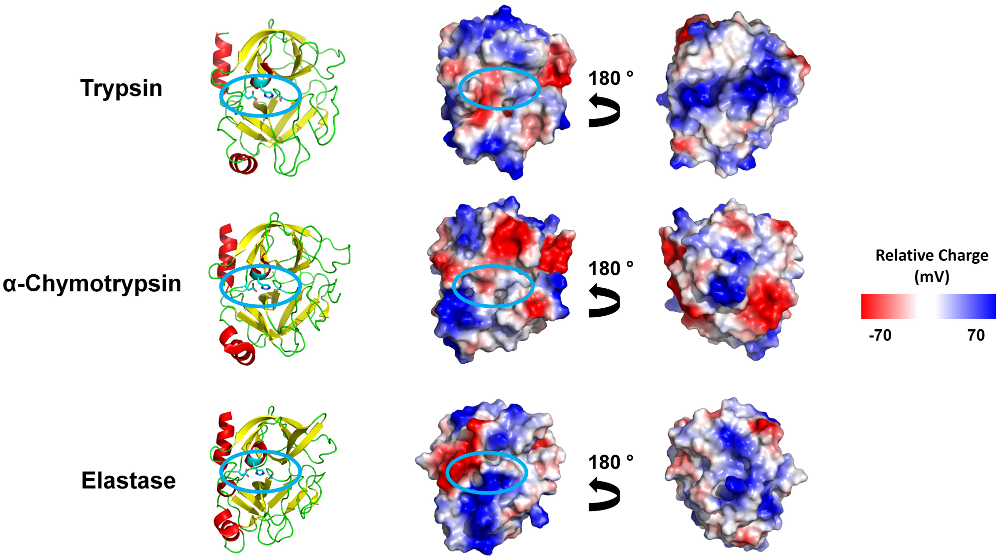


Figure 7.8: The structures of the pancreatic peptidases trypsin, α-chymotrypsin and elastase. The cartoon representations have the secondary structural elements and the catalytic triad highlighted. The cyan oval shows the location of the catalytic triad on the cartoon diagrams and the electrostatic surfaces of the peptidases (Baker *et al.* 2001).

7.5.2 Trypsin, α-chymotrypsin and elastase

Figure 7.8 shows the structures of trypsin, α-chymotrypsin and elastase. Despite structural and functional conservation, the proteases from the three families only share 40% sequence identity. Trypsin, α-chymotrypsin and elastase are serine proteases and have the same catalytic mechanism, the catalytic triad, which is highlighted in Figures 7.8. Inspection of the electrostatic surfaces of the enzymes shows there is significant sequence variation, even around the binding cleft of the proteases. That said, the other side of the proteases seems to have a relatively well conserved positive patch, which could be a binding interface.

7.6 ECD binding experiments

SPR was used to investigate the binding of proteases to *Mm*PepT1 and *Rn*PepT2 ECD. The materials and methods used are fully described in § 2.8.7. Briefly, the ligand (ECD) was immobilised to a Series S, CM5 chip (GE Healthcare) by amide coupling, aiming for a $\mathrm{RU}_{\mathrm{MAX}}$ of 1000 RUs. Blank flow cells were immobilised with ethanolamine. The digestive proteases trypsin, α-chymotrypsin, and pepsin were used as the analyte and passed over the active and blanked flow cells at concentrations of 1000, 500, 250, 125, 62.5, 31.25, 15.625 and 7.8125 µM. The limit of the assay appears to be 1000 µM. Higher analyte concentrations produce spurious sensograms, probably caused by the refractive index of the bulk analyte.

Other immobilisation strategies were attempted as the amide coupling would randomly orientate the immobilised ECD and therefore perhaps obscure the binding site in a particular orientation leading to inaccurate K_d calculation. AviTag^{TM} immobilised protein onto a Series S, CAP chip (GE Healthcare) was also attempted. Unfortunately trypsin showed high affinity for the single stranded DNA which make up the chip surface and the AviTag^{TM} appeared to obscure the trypsin binding site.

7.6.1 α-chymotrypsin and pepsin

At first, these SPR binding experiments were exploratory in nature, as the function of the ECDs were not known. § 7.6.2 shows that the ECDs have a specific interaction with trypsin and § 7.6.3 explores this interaction in terms of *Mm*PepT1^{ECD}. However, at the start α-chymotrypsin was also a possible interaction partner and pepsin was felt to be a good negative control; as pepsin should be inactive in the small intestine where PepT1 is expressed.

Figure 7.9 shows the blanked sensograms and binding curve fits for *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} with α-chymotrypsin and pepsin. α-chymotrypsin does show some affinity for the ECDs however, this interaction could not be saturated using this SPR protocol. Given the lack of saturation, there is a large error associated with the predicted K_ds of 1.3 ± 1.0 and 4.2 ± 5.0 mM for *Mm*PepT1^{ECD} and *Rn*PepT2^{ECD} respectively. Pepsin, did not give a positive sensogram for either *Mm*PepT1^{ECD} or *Rn*PepT2^{ECD}, which suggests that there is no interaction between the Pepsin and either ECD.

7.6.2 Trypsin

Initial binding experiments using an excess of bound ligand to the chip surface suggested a specific interaction between *Mm*PepT1^{ECD} and trypsin. An unsaturatable interaction was also observed between *Rn*PepT2^{ECD} and trypsin in this initial assay. To improve the accuracy of the calculated affinity constants, a reduced amount of ligand was immobilised.

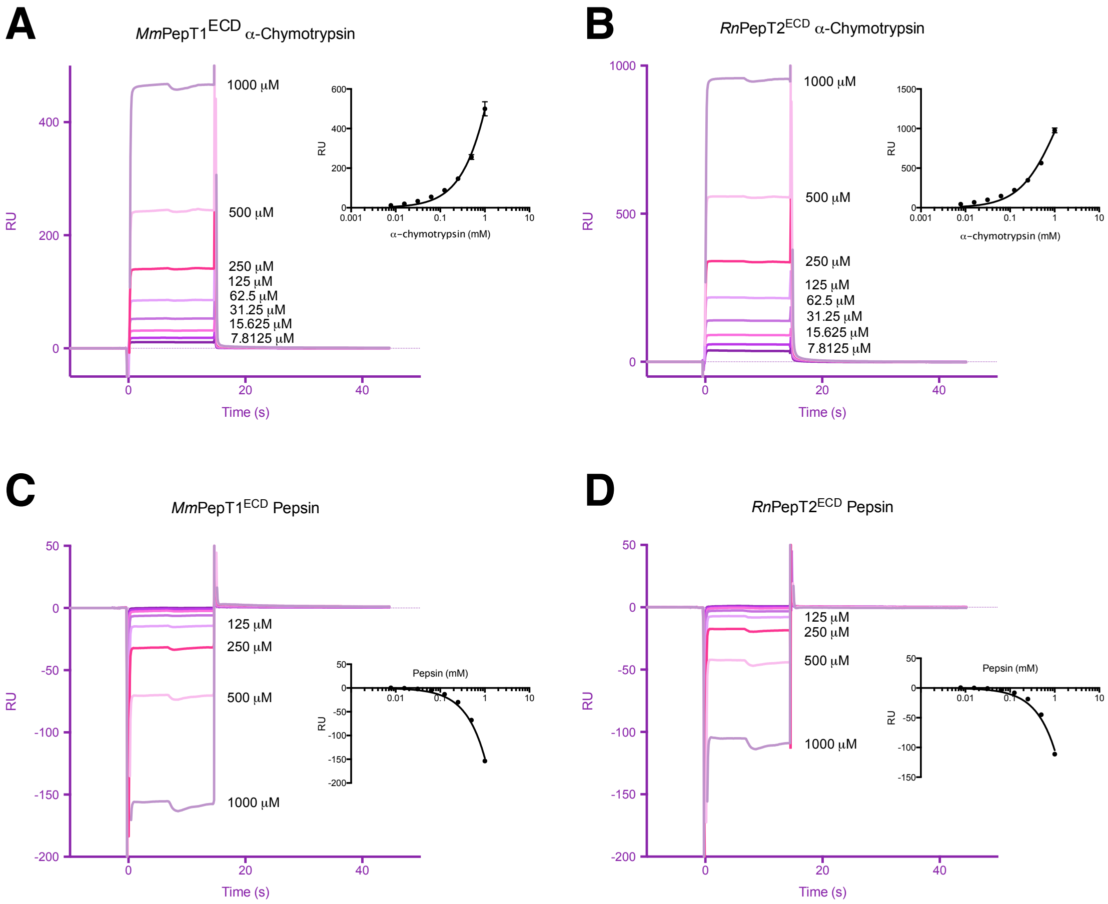


Figure 7.9: The blanked subtracted sensograms and equilibrium data fits for the ECDs with α-chymotrypsin and pepsin. Each plot **(A)** to **(D)** shows the blanked subtracted sensogram on the left and the equilibrium fit to the data on the right. Both ECDs show an interaction with α-chymotrypsin, with predicted K_ds for *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ of 1.3 ± 1.0 and 4.2 ± 5.0 mM (2.s.f) respectively at the 0.05 criterion and $n = 3$. Both of these values are extrapolations as the interaction could not be saturated in this assay format. Pepsin did not bind to either ECD immobilised cell greater than reference, suggesting no interaction.

Under these conditions the following results were obtained (Figure 7.10). MBP was used as a control protein as it became clear that trypsin has a general affinity for protein. MBP was chosen as the control protein as there is no physiological link between MBP and trypsin, therefore any affinity observed by trypsin for MBP was assumed to be a basal affinity which trypsin has for generic folded protein.

The results of the SPR binding experiments indicate that *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ have an affinity for trypsin which is significantly lower (at the 0.05 criterion) than the control protein, which suggests that the ECDs make a specific interaction with trypsin. The $\mathrm{K_d}$ of *Mm*PepT1$^{\mathrm{ECD}}$ is also significantly lower than *Rn*PepT2$^{\mathrm{ECD}}$ at 90 ± 20 and 170 ± 30 μM respectively. Since trypsin appeared to have a greater affinity for *Mm*PepT1$^{\mathrm{ECD}}$ than *Rn*PepT2$^{\mathrm{ECD}}$, primary focus to understand the interaction was put on *Mm*PepT1$^{\mathrm{ECD}}$.

7.6.3 Characterising the *Mm*PepT1$^{\mathrm{ECD}}$ trypsin binding site

Mutagenesis experiments were designed based on a PepT1$^{\mathrm{ECD}}$ sequence alignment (Figure 7.11A) and examination of the structure of *Mm*PepT1$^{\mathrm{ECD}}$ (Figure 7.11B). It was not clear by what mechanism the trypsin interacted with the ECD. It could be that the two lobes separate and bind trypsin in a clamshell-like manner, or the two lobes may remain locked together during the interaction. To test these two possible mechanisms, the two salt bridge aspartate residues (D476 and D574) were mutated plus a range of surface mutations were designed encompassing both conserved and non-conserved residues.

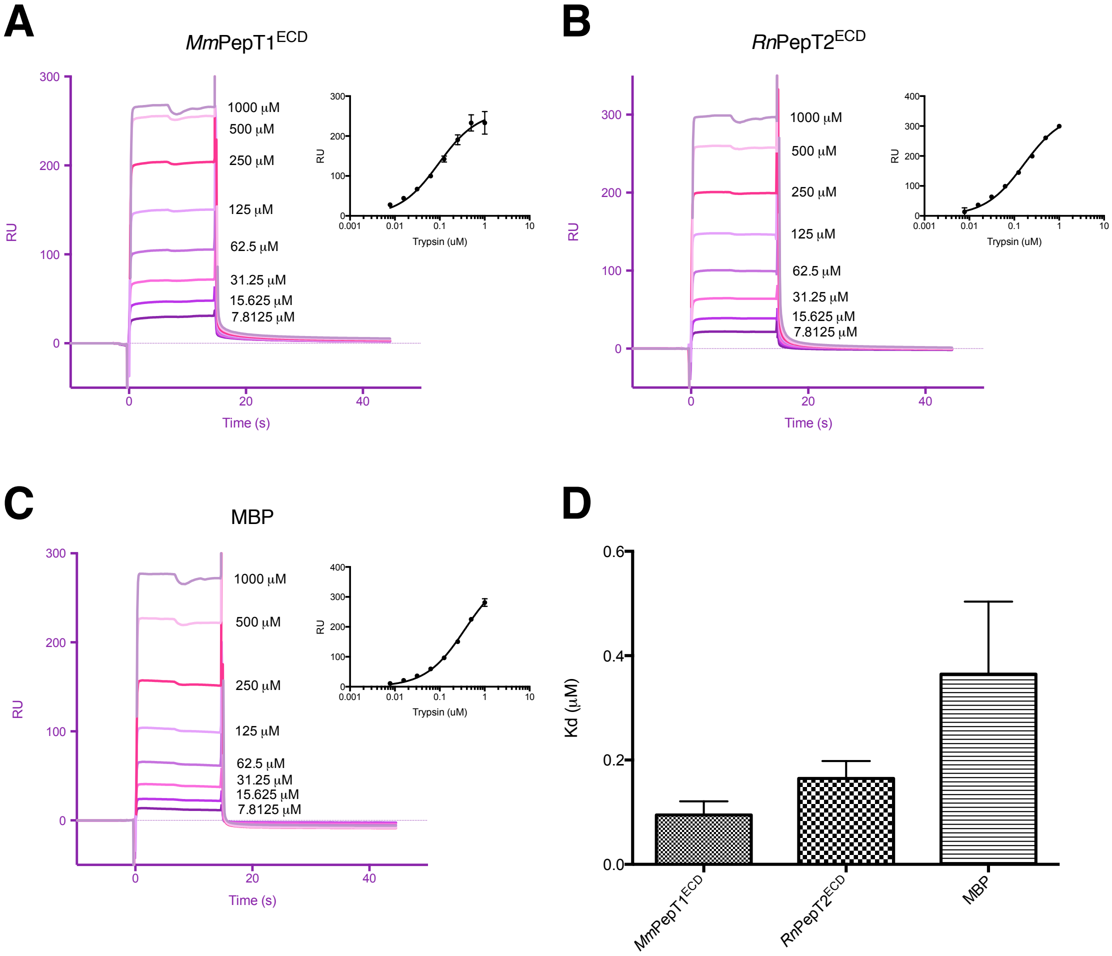


Figure 7.10: The blank subtracted sensograms, equilibrium data fit and binding constants for trypsin and *Mm*PepT1$^{\mathbf{ECD}}$, *Rn*PepT2$^{\mathbf{ECD}}$ and MBP. Each plot **(A)**, **(B)** and **(C)** shows the blanked subtracted sensogram on the left and the equilibrium fit to the data on the right. The calculated K_ds with confidence limits (0.05 criterion, n = 3) from the equilibrium fits are shown in **(D)**. The binding of *Mm*PepT1$^{\mathrm{ECD}}$ could be saturated in these conditions. The binding curves for both *Rn*PepT2$^{\mathrm{ECD}}$ and MBP are both showing a sign of saturation, but the reached the limits of the assay. The K_ds calculated for *Mm*PepT1$^{\mathrm{ECD}}$, *Rn*PepT2$^{\mathrm{ECD}}$ and MBP were 90 ± 20, 170 ± 30 and 360 ± 60 µM respectively.

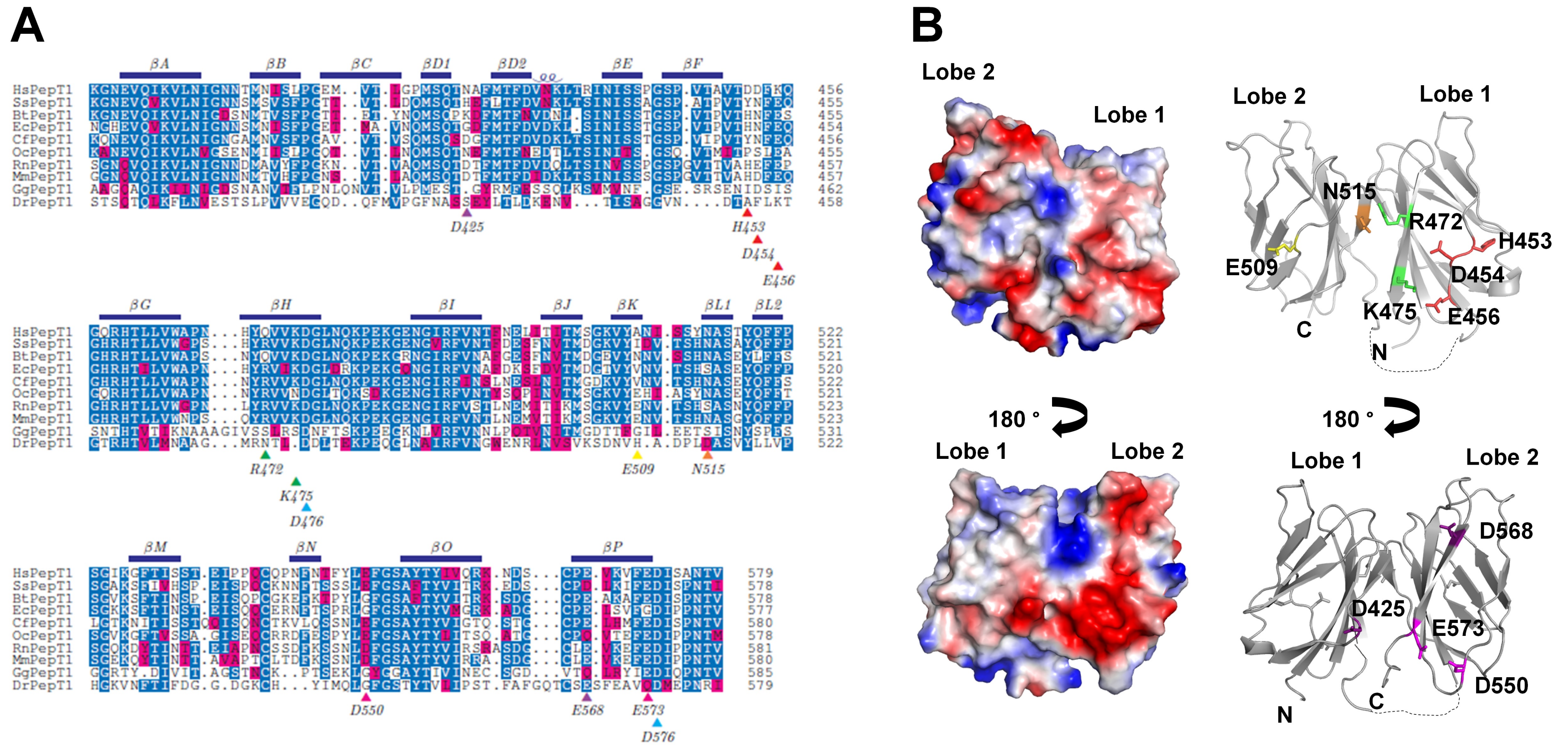


Figure 7.11: Design of *Mm*PepT1$^{\text{ECD}}$ mutants to localise trypsin binding site. **(A)** A sequence alignment of vertebrate PepT1$^{\text{ECD}}$ sequences, coloured blue and pink for identical and similar residues respectively. The secondary structure of *Mm*PepT1$^{\text{ECD}}$ has been mapped on top of the alignment and potentially residues involved in binding are shown below. **(B)** Two side views of the closed *Mm*PepT1$^{\text{ECD}}$ monomer showing the electrostatics (Baker *et al.* 2001) and ribbon of the structure. The residues highlighted are also shown on the alignment in **(A)**, apart from the two salt bridge aspartate residues which are marked with an blue triangle (▲).

It was hypothesised that D476A and D574A mutations would negatively effect the binding of *Mm*PepT1^{ECD} to trypsin. These resides were most likely residues to mediate the interaction in the event of a clamshell-like mechanism. They also maintain the closed conformation of the ECD (§ 6.5), so if trypsin did bind to the surface of *Mm*PepT1^{ECD}, the breakdown of the closed conformation was also likely to inhibit binding. The surface mutation were a good negative control for a clamshell mechanism, because if a single surface mutation inhibited binding, the binding face was unlikely to be located in the lobe interface.

Figure 7.12A shows a table of *Mm*PepT1^{ECD} mutant constructs designed to examine and characterise the binding of *Mm*PepT1^{ECD} to trypsin; the mutants shown in plain text were purified and tested for trypsin binding affinity using the SPR assay. Mutants were cloned as described in § 2.8.7. Figures 7.12B and 7.12C show an S75 10/300 GF trace and reducing 15% Tris-Gly SDS-PAGE gel of the purified mutants respectively. The GF traces show that the mutants all elute at approximately the same V_e as WT, indicating they are also folded and stable. The SDS-PAGE gel shows that there is still minor contaminations in many of the mutant samples, but the majority of the protein present is the ECD mutant. With the exception of the H453E/D454R/E456R mutant, where it is not clear the ECD band makes up the majority of the protein in solution.

Results

The *Mm*PepT1^{ECD} mutants were screened in the same manner as the WT protein as described in § 2.8.7. Figure 7.13 shows the blanked sensograms for the experimental mutants tested. Although protein had been produced for the D476A, D550A/E573A/D425A and D550A/E573A/D425A mutants, an error occurred with the immobilisation and the experiments were not be repeated. Good immobilisations and data were collected for the mutants shown in Figure 7.13. The only data which is questionable is from the H453E/D454R/E456R sample, as the gel from Figure 7.12C shows that the sample is not pure.

The K_d of each mutant is plotted with the 95% confidence limits (n = 3) with the WT *Mm*PepT1^{ECD} data in Figure 7.14. The R472A/K475A, R472E/K475E and E509R mutants, with calculated K_ds of 100 ± 10, 110 ± 10 and 70 ± 10 μM respectively, did not signif-

A

Mutant	Location
H453A/D454A/E456A	βF-βG loop
D453A	βF-βG loop
D454A	βF-βG loop
D456A	βF-βG loop
H453E/D454R/E456R	βF-βG loop
R472A/K475A	βH
R472E/K475E	βH
N515A	βL1
E509R	βK
D476A	βH
D574A	βP
D550A/E573A	βN-βO loop
D550A/E573A/D425A	βN-βO loop/βD
D550A/E573A/D568A	βN-βO loop/βP
D550A	βN-βO loop
D573A	βN-βO loop
D425A	βD
D568A	βP

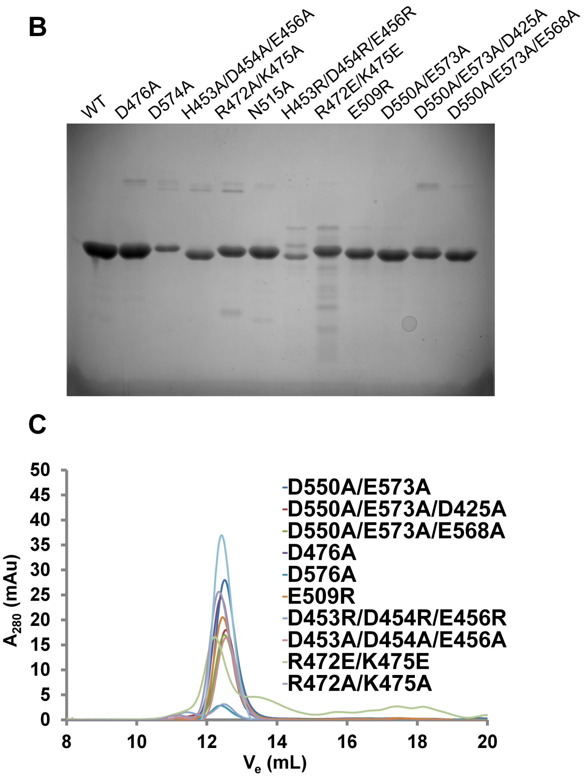


Figure 7.12: An assessment of the purity of the *Mm*PepT1$^{\text{ECD}}$ mutants. (A) A table of the *Mm*PepT1$^{\text{ECD}}$ mutants designed for the characterisation of the trypsin binding site. Each mutation is colour matched to Figure 7.11 and its position on the structure is given in the right hand table column. The mutants shown in italics have either yet to be purified, and/or data has not been collected. **(B)** and **(C)** show the S75 10/300 GF traces and a reducing 15% Tris-Gly SDS-PAGE gel of the mutants which could be purified.

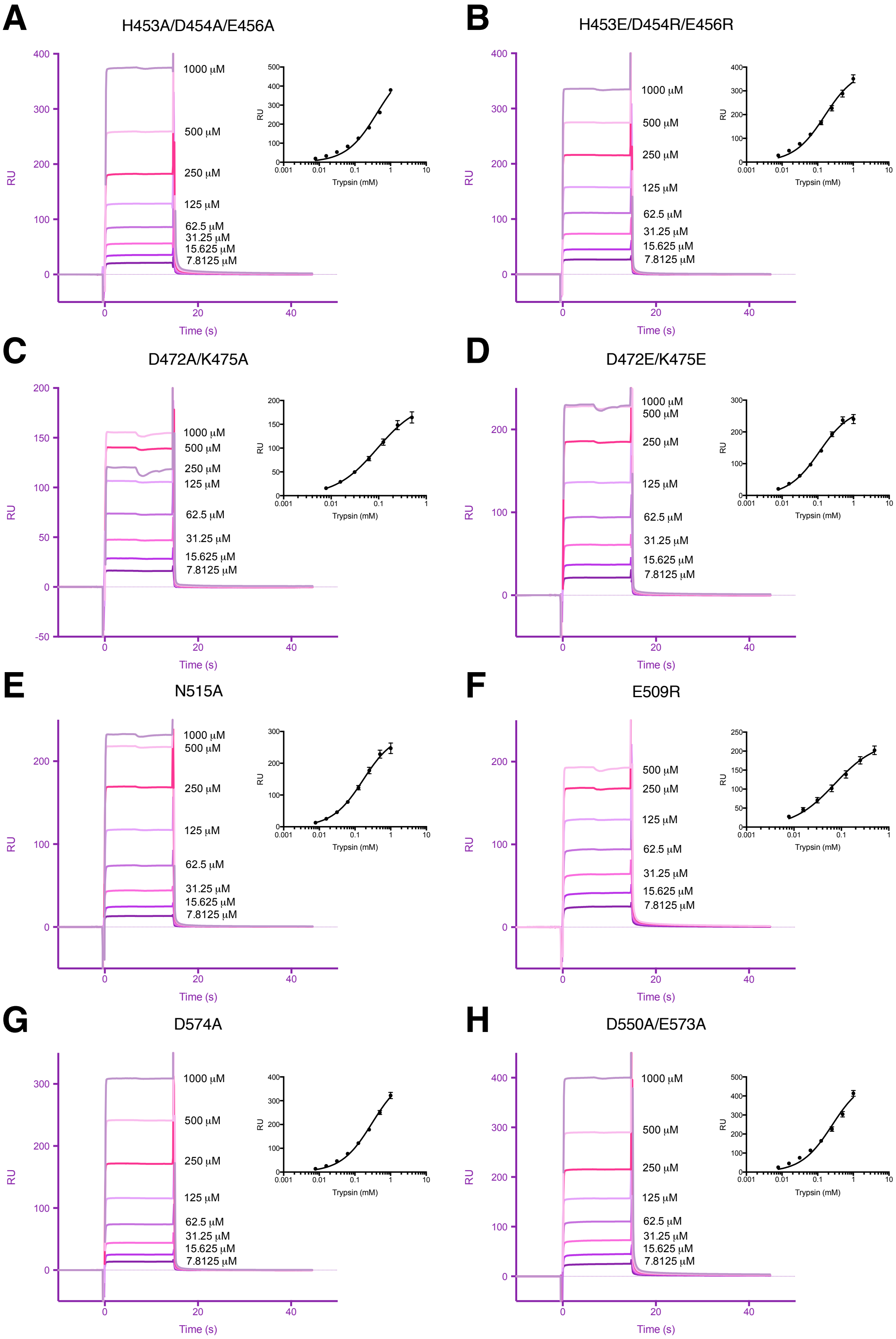


Figure 7.13: The blanked subtracted sensograms and equilibrium data fits for the immobilised *Mm*PepT1$^{\mathbf{ECD}}$ mutants and trypsin. Each plot, **(A)** to **(H)**, shows the mutant blanked subtracted sensogram and equilibrium binding fit on the left and right hand side respectively.

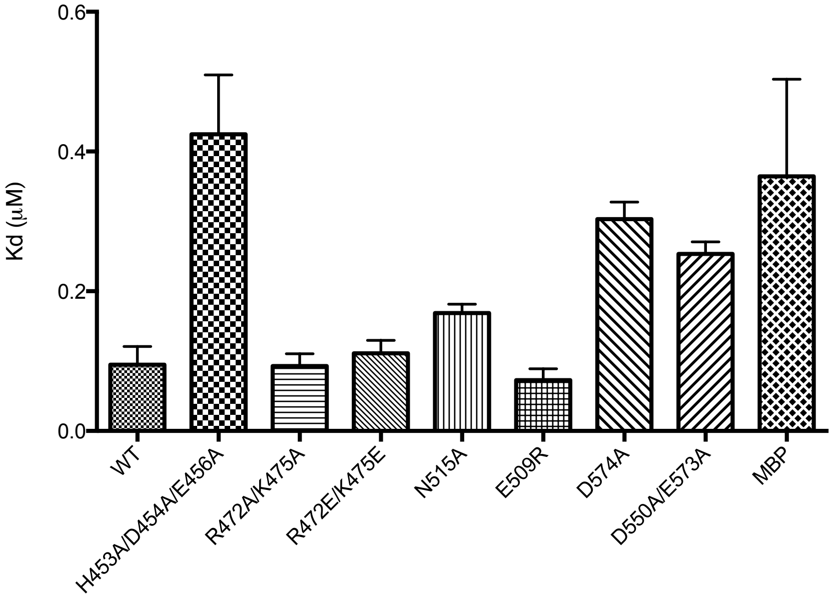


Figure 7.14: A plot of the *Mm*PepT1$^{\mathrm{ECD}}$ mutant K$_{\mathrm{d}}$s calculated from equilibrium data analysis. Each K$_{\mathrm{d}}$ is plotted with the 95% confidence limits at n = 3. The trypsin K$_{\mathrm{d}}$ and confidence limits for MBP are also plotted to show the general affinity trypsin has for globular protein.

icantly differ from *Mm*PepT1$^{\mathrm{ECD}}$ WT at the 0.05 criterion. The H453A/D454A/E456A, N515A, D574A and D550A/E573A mutants K$_{\mathrm{d}}$s of 420 ± 40, 170 ± 10, 300 ± 10 and 250 ± 10 μM respectively, were significantly lower compared to WT at the 0.05 criterion. The K$_{\mathrm{d}}$s of H453A/D454A/E456A, D574A and D550A/E573A mutants, fall with the range of MBP, which suggests that the mutation has removed the specific affinity which trypsin had for *Mm*PepT1$^{\mathrm{ECD}}$.

Analysis

The results from § 7.6.3 show that the affinity between *Mm*PepT1$^{\mathrm{ECD}}$ and trypsin can be modulated by mutating specific residues in *Mm*PepT1$^{\mathrm{ECD}}$, which is consistent with a specific interaction between two proteins. The N515A mutant did significantly effect binding compared to WT, but not significantly enough to be within the binding range of MBP. Three mutants, the H453A/D454A/E456A, D574A and D550A/E573A all reduced the K$_{\mathrm{d}}$ of the *Mm*PepT1$^{\mathrm{ECD}}$ trypsin interaction to within the range of the control protein.

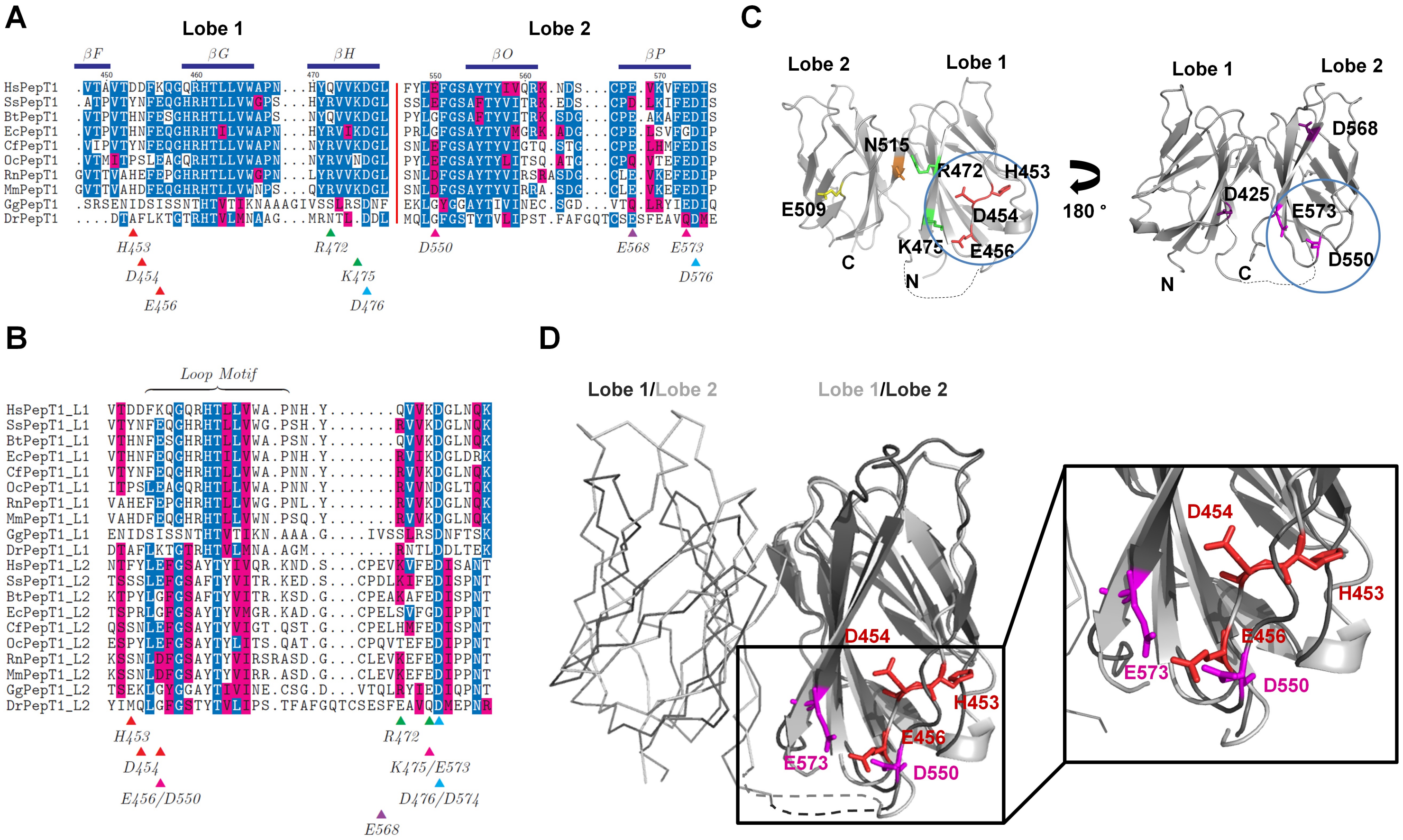


Figure 7.15: Analysis of the *Mm*PepT1$^{\text{ECD}}$ sequence and structure in the region of the H453A/D454A/E456A and D550A/E573A mutants. Throughout this figure the colour coding of the mutants is consistent with the Figures 7.11 and 7.12A. **(A)** A sequence alignment of the areas of *Mm*PepT1$^{\text{ECD}}$ lobe 1 and lobe 2 which contain the H453A/D454A/E456A and D550A/E573A mutants. The other mutants which also fall in these regions have also been shown. **(B)** An alignment of lobe 1 and lobe 2 *Mm*PepT1$^{\text{ECD}}$ sequences from **(A)**, showing the mutant positions, and the Loop motif and conserved aspartate identified in § 6.3.3. **(C)** Maps the mutants locations to the *Mm*PepT1$^{\text{ECD}}$ structure in the closed conformation. The blue circles show the surface mutant locations which have a negative impact on trypsin binding. **(D)** An `SSM` (Krissinel and Henrick 2004) Cα alignment of two *Mm*PepT1$^{\text{ECD}}$ monomers which superimposes lobe 1 and lobe 2. The H453A/D454A/E456A and D550A/E573A mutants have been highlighted in red and magenta respectively.

Figure 7.15 shows a sequence and structural analysis of the binding data. Based on the results it is still not clear exactly how *Mm*PepT1$^{\mathrm{ECD}}$ and trypsin interact as there are several ways in which the data can interpreted. The mutants were designed to narrow down the binding interface to either a surface interaction or a clamshell-like mechanism between the two lobes. The decreased in trypsin binding observed for the H453A/D454A/E456A and D550A/E573A mutants suggest that the trypsin binding site is located on the surface of *Mm*PepT1$^{\mathrm{ECD}}$. Figure 7.15C shows that the relative locations of H453/D454/E456 and D550/E573 are on opposite sides (blue circles) of *Mm*PepT1$^{\mathrm{ECD}}$ in the closed conformation. So there could be two locations on the surface of the PepT1$^{\mathrm{ECD}}$ which interact with trypsin.

There is some evidence for theory from the the sequence (Figure 7.15A and B) and structural alignment (Figure 7.15D). Figure 7.15A shows that of the residues in the H453/D454/E456 mutant, only E456 shows conservation; in the D550/E573 mutant, both are well conserved. The alignment of lobe 1 and lobe 2 (Figure 7.15B) shows that E456 and D550 align together. This can also be seen in the structural alignment of both lobes (Figure 7.15D). It is unlikely that only one residue mediates the interaction between *Mm*PepT1$^{\mathrm{ECD}}$ and trypsin, however this analysis shows that due to the fact that the ECD is composed of two repeating units related by a 2-fold rotation, it could be possible for trypsin to have affinity for two locations on either side of the protein. Another explanation of these data is that, either mutant could have a destabilising or allosteric effect on the entire ECD, which inhibits binding at the other location.

The decrease in affinity from the D574A mutant suggests that the lobes do need to remain in the closed conformation to allow trypsin to bind. The N515A mutant to a certain extent supports this conclusion as it is located at the edge of the binding interface and may also affect the stability of the closed conformation. However, if the residues which line the interface can affect the closed conformation stability, it might also be expected that the R472A/K475A and R472E/K475E mutants also negatively effect binding, and this was not the case. Also, since the location of trypsin binding could not be definitively pinpointed to an *Mm*PepT1$^{\mathrm{ECD}}$ surface location, it could still be that binding occurs in a clamshell-like mechanism, and the negative affect observed in the H453A/D454A/E456A and D550A/E573A mutants is again related to an allosteric effect.

7.7 Discussion

The final aim of this thesis was to investigate potential functions of the peptide transporter ECD. Three possible functions were hypothesised; that the ECD binds some other factor which regulates or aids the function of the transporter, that the ECD was an essential component of the transport cycle or that the ECD is necessary in some capacity for the stability of the transporter. A literature review into potential ligands and regulatory factors of PepT1 and PepT2 did not identify any ligands which could be shown to have an interaction with either *Mm*PepT1 or *Rn*PepT2 ECD.

To further investigate these ideas, hybrid models of *H. sapiens* PepT1 and PepT2 were constructed. The hybrid models are the first relatively complete models of *H. sapiens* PepT1 and PepT2. The principle missing pieces of structure are located in the N- and C-Termini and extra-cellular loops. The `Molprobity` and Ramachadran analysis indicate that the models are reasonable in terms of there overall bond angles, distances and internal clashes. The structures represent the best current models of the mammalian peptide transporters and so should be used as a basis for future experiments, but with a degree of caution. From a practical point of view, the models potentially could be used to solve the phases of a native protein crystal by molecular replacement.

The hybrid models indicate that the ECDs of the vertebrate peptide transporters are located above the TM domain of the transporter. This location of the ECD at the end of intra-cellular gate pore helix, suggested that the ECD could function as a mediator of the transporters function. To examine this hypothesis, the ECDΔ and ECD swapped mutants were created. The results indicated that PepT2$^{\mathrm{ECD}}$ is not required for transport, and that PepT1$^{\mathrm{ECD}}$ may have a stabilising role. In support of this latter hypothesis, the PepT1-D574A mutant gave WT-like transport properties, even though the mutation alters the ECD structure, changing it from a closed lobe conformation to open. The mutation also significantly reduces the stability of the ECD in solution, however, it may be that the presence of the TM domain silenced this effect.

The SPR experiments indicated that both *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ bind trypsin with affinities of 90 ± 20 and 170 ± 30 μM respectively. This interaction, certainly for PepT1 is physiologically reasonable as the concentration of trypsin in the small intestine

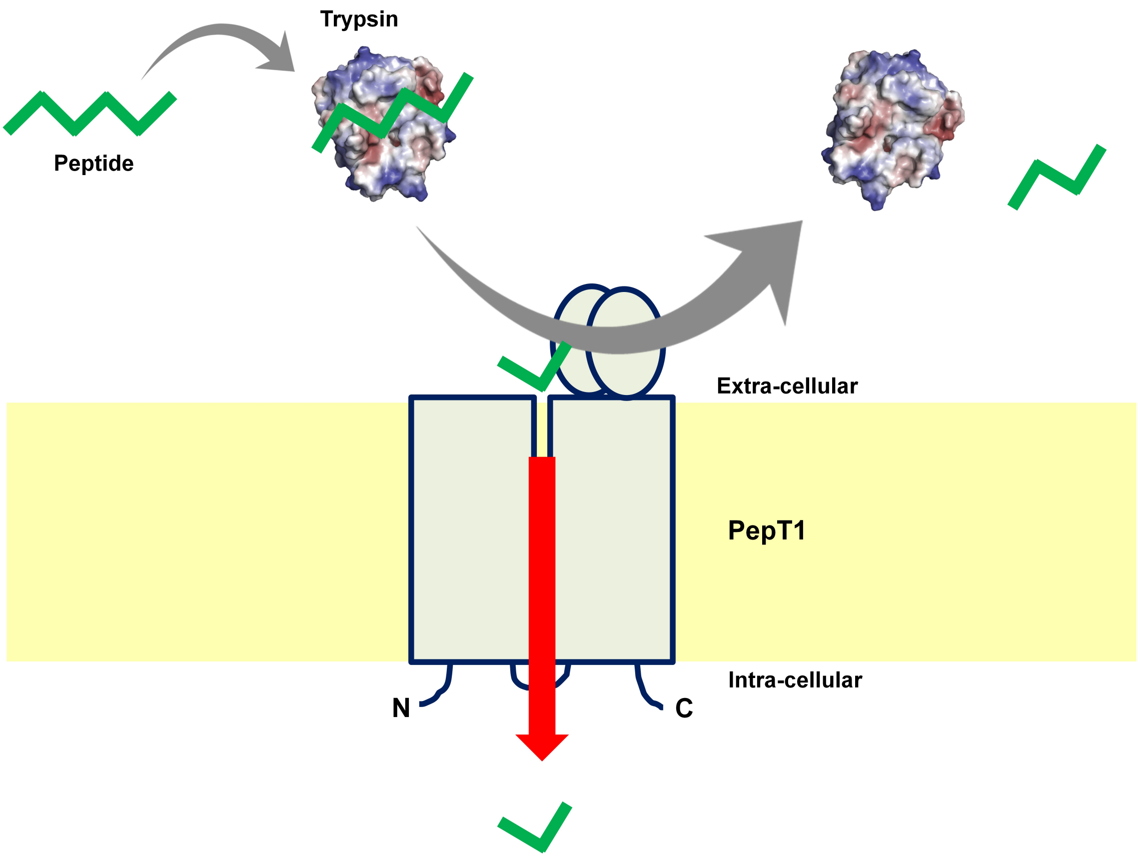


Figure 7.16: A representation of how an interaction between PepT1^{ECD} and trypsin aids the uptake of di- and tri-peptides in the small intestine. Trypsin in the unstirred layer, interacts with a peptide. As peptide bond hydrolysis occurs, the predisposition of trypsin for PepT1^{ECD} localises the tri-peptide product to the transporter binding site, and transport ensues. The trypsin then dissociates to diffuse and find more substrate.

has been estimated at 1 – 40 μM (Kong *et al.* 1997). Unfortunately due to speed of the on and off rates, SPR could not be used to estimate the on and off rates of binding. However, inspection of the sensograms indicated that association and dissociation are instantaneous at pH 6.5. This weak interaction is therefore exceptional transient, and would localise the trypsin in the vicinity of the transporter, but still allow the protease to diffuse throughout the unstirred layer finding substrate. Figure 7.16 gives a representation of how this interaction could aid the transport of di- and tri-peptides into the gut lumen.

The interaction between PepT1 and PepT2, and trypsin is also of importance outside the environment of the small intestine. Trypsin mRNA has been detected in the epithelial tissues of the kidney, liver and CNS, and in white blood cells (Koshikawa *et al.* 1998), suggesting that trypsin's role in the body is not confined to the intestine. Expression of trypsin in these tissues suggests that trypsin would be present in the blood, and hence the need for plasma trypsin inhibitors (Stockley 2014). If trypsin is in the blood it is possible that it plays a protein cleavage role in the proximal loop of the kidney; cleaving small plasma peptides and helping the re-absorption of nitrogen lost during kidney blood filtration.

The localisation of the *Mm*PepT1$^{\mathrm{ECD}}$ trypsin binding site was also investigated using site-directed mutagenesis and the SPR assay. Unfortunately a definitive binding mechanism could not be elucidated. Based on the results presented in this chapter, the clamshell-like mechanism is still possible as the surface mutants may have affected an interface centred binding site but, in the authors opinion this is unlikely. The linker mutation studied in § 6.5 demonstrated that the lobes in *Mm*PepT1$^{\mathrm{ECD}}$ primarily remain associated together along the closed conformation lobe interface. It is improbable that such a transient interaction, as the trypsin association is, could be based upon an interface which is not readily available. It is more likely that residues on the surface of the *Mm*PepT1$^{\mathrm{ECD}}$ mediate such a fast interaction.

Even given the topological similarity highlighted in Figure 7.15, it also seems unlikely that two separate areas of the protein are involved in binding trypsin. It is more likely that one or other of the mutations has effected the other allosterically, producing a comparable negative effect on binding. The complete set of mutations proposed in Figure 7.12A would allow for a greater characterisation of the single residue contributions to the binding face. In particular, the single E456A and D550A mutants, as these two appear to show sequence and structural conservation.

There is sequence and structural evidence to support both potential surface binding sites. The lobe 1 binding face, faces the top of the TM domain in the hybrid models, but both orientations are possible. This orientation was selected, by observation, as the residues in the linkers between the lobes, and the ECD and the transporter fit in a more logical way *e.g.* the ECD-TM domain linkers did not cross; this of course could be incorrect.

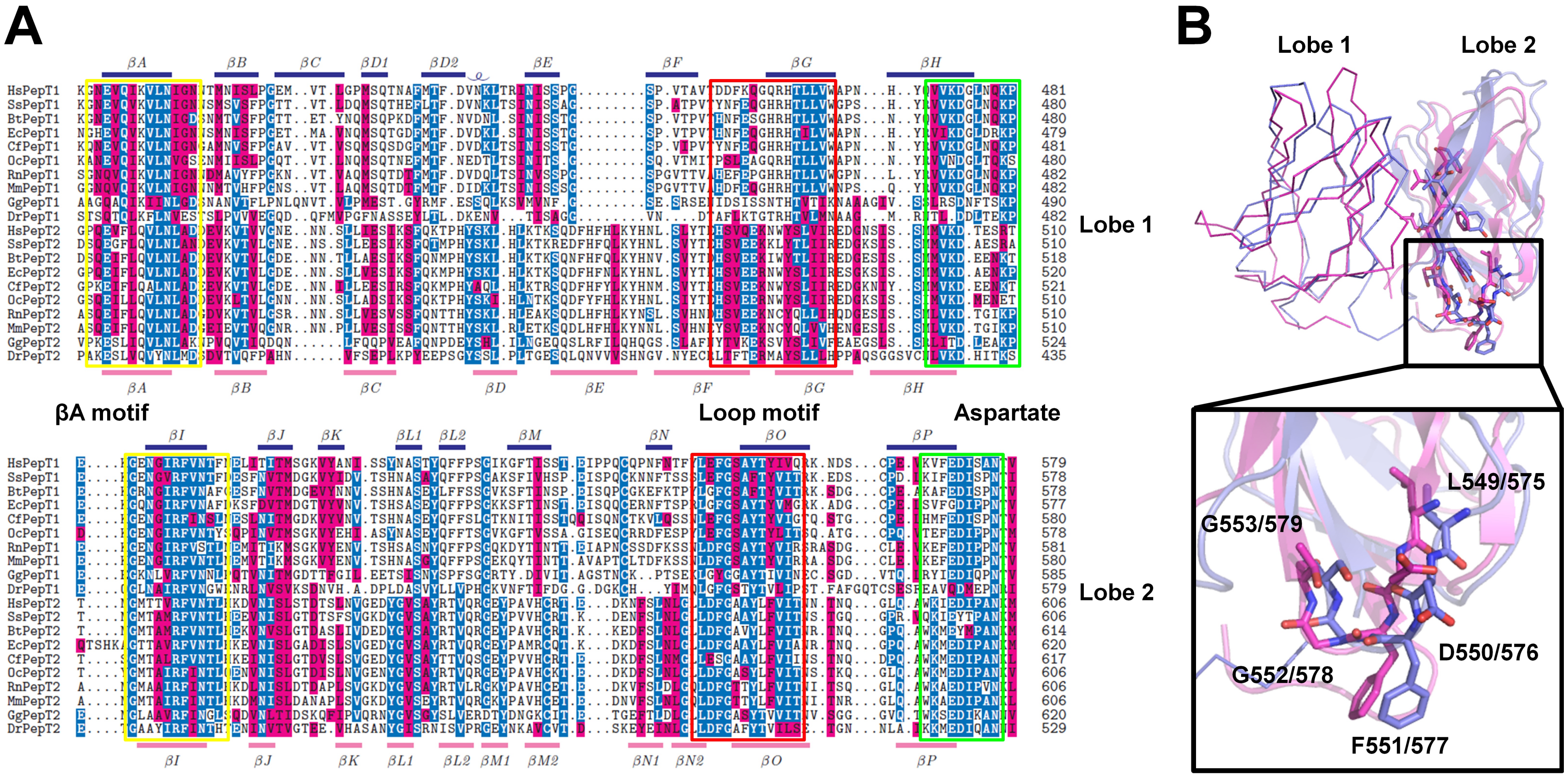


Figure 7.17: Possible conservation of trypsin binding interface between PepT1 and PepT2 ECD. (A) Shows a sequence alignment of vertebrate PepT1 and PepT2 ECD sequences. The secondary structure of *Mm*PepT1$^{\mathrm{ECD}}$ and *Rn*PepT2$^{\mathrm{ECD}}$ is shown above and below respectively. The βA, Loop and aspartate motifs (identified in § 6.3.3) are shown in the lobe 1 and lobe 2 regions by yellow, red and green boxes respectively. The βA motif and the region around the conserved aspartate are well conserved between PepT1 and PepT2. The Loop motif is only well conserved between PepT1 and PepT2 in the lobe 2. **(B)** An SSM (Krissinel and Henrick 2004) Cα alignment of *Mm*PepT1$^{\mathrm{ECD}}$ (blue) and *Rn*PepT2$^{\mathrm{ECD}}$ (purple). The surface exposed residues of the Loop motif in Lobe 2 have been highlighted (*Mm*PepT1$^{\mathrm{ECD}}$/*Rn*PepT2$^{\mathrm{ECD}}$). The sequence conservation observed in the alignment translates well when mapped onto the structure.

Compared to lobe 1, the lobe 2 binding face, faces away from the TM domain in the hybrid models, but appears to be conserved between PepT1 and PepT2. Figure 7.17A shows a vertebrate alignment of PepT1 and PepT2 ECD sequences with the motifs identified in § 6.3.3 highlighted. The Loop motif in lobe 2 shows much stronger conservation in sequence and structure (Figure 7.17B) between PepT1 and PepT2, and this is where D550 from *Mm*PepT1$^{\mathrm{ECD}}$ is located. Since it is likely that PepT1 and PepT2 bind trypsin in the same manner, other than the βA and aspartate regions of the sequence, this is the only section which shows conservation. An investigation into the binding residues in PepT2$^{\mathrm{ECD}}$ was not attempted in this thesis, but should be completed to characterise the interaction and understand why there is a difference in trypsin affinity between PepT1 and PepT2.

Ultimately, this chapter has completed, and partially completely two aims of thesis (§ 1.7); to build homology models of the two transporters and to discover a possible function for the ECD. Work has been performed to characterise the *Mm*PepT1$^{\mathrm{ECD}}$ trypsin binding site, and advances in understanding have been made, but not completed. The PepT2$^{\mathrm{ECD}}$ trypsin binding site has not been investigated but based on the results of PepT1, a study into the potential conservation of the Loop motif as a trypsin interface should be tested.

8

Final discussion

8.1 Final analysis

At the beginning of this work no structural and functional data had been collected on the vertebrate peptide transporter ECD. It was not even known whether this region of the protein, which had always been annotated as a ‘loop’, was folded. The crystal structures of the bacterial peptide transporter homologues, PepT_{So} and PepT_{St} (Newstead *et al.* 2011; Solcan *et al.* 2012), gave the first insight into the structure of the TM domain of the transporter. Since the TM domain of the bacterial peptide transporters is a self-contained functional unit and shares 30% sequence identity with the human transporters, it was considered likely that the loop was also a self-contained functional unit. Bioinformatic analysis (§ 1.6) suggested that a significant proportion of the loop region was ordered and contained secondary structural elements. It was considered possible that the loop’s tertiary structure was dependent upon the TM domain, but it was sensible to explore the recombinant expression of the loop region.

The OPPF-UK pipeline was used to screen the initial mutants which proved successful. The *Mm*PepT1$^{\mathrm{ECD}}$ construct proved to be very stable and expressed in large quantities. Of the other PepT1 genes examined, only the *O. cuniculus* and *B. taurus* ECDs expressed and purified, but not to the same level as the *M. musculus* construct. The *O. cuniculus* and *B. taurus* PepT1$^{\mathrm{ECD}}$s were also less stable than the *M. musculus* construct. The purification of PepT1$^{\mathrm{ECD}}$ demonstrated the importance of multiple gene screening in situations where no structural data is present.

The crystallisation of *Mm*PepT1$^{\mathrm{ECD}}$ was perhaps the most fortunate event of the whole project. The high resolution rod-like crystals have still never been reproduced despite extensive efforts to so. Crystallisation appears to have been completely dependent upon the well condition bought from Emerald Biosciences. Even other batches of Emerald Bioscience well condition did not yield the rod-like crystals. Examination of the ASU of the crystal possibly yields the answer. The biophysical analysis, performed in Chapter 6, strongly suggests that the lobes of PepT1 remain in the closed conformation, regardless of the external environment. This in turn suggests that the ASU conformation is not a thermodynamically stable state under normal conditions. Therefore, during nucleation, it follows that the well condition must have encouraged this unfavourable interaction and, based on the difficulty in recreating these crystals, this condition must be quite unique.

PepT2$^{\mathrm{ECD}}$ presented a very different challenge. The time taken from initial construct cloning to the solution of the structure was approximately 2.5 years. The majority of that time was spent working with the *Mm*PepT2 410 – 609 construct. Even when the construct was truncated to residues 410 – 601, the *Mm*PepT2$^{\mathrm{ECD}}$ failed to crystallise. However, the change to the *R. norvegicus* gene produced crystals in a matter of hours. This story is probably familiar to many crystallographers; hopefully this experience will be beneficial for future projects.

It was not predicted that the domains would share any structural features, and it is amazing how similar in structure they are, given they share so little sequence identity. It was difficult to analyse a single ECD. When the structures of both were known, headway in probing their properties was made. The areas of sequence conservation were not obvious until it was clear that the ECDs shared a common fold. The principle difficulty was that it was assumed that the solution of the structures would give insight into the function of the ECDs, and then this function could be probed biophysically. The discovery that both domains were structurally related to TTR was of limited usefulness, given that the sequences was not well conserved. The βA motif appeared to be the only area which showed conservation which was relatively weak.

The sequence and structural analysis did highlight the lobe interface and the conservation of the aspartate residues, D476 and D574 in *Mm*PepT1$^{\text{ECD}}$, and D505 and D600 in *Rn*PepT2$^{\text{ECD}}$. Mutating the aspartate residues produces predictable effects on the stability and conformation of the ECDs. The stability of *Mm*PepT1$^{\text{ECD}}$ was dependent upon the salt bridges, which also reduced the domain's affinity for trypsin. In *Rn*PepT2$^{\text{ECD}}$, the D505A and D600A mutation produced the opposite effect, increasing the stability of the ECD. The reason for the proximity of the two aspartate residues in *Rn*PepT2$^{\text{ECD}}$ was not elucidated. Probing of the entire transporter function will hopefully give an answer to this question in due course.

A limitation to studying the entire transporter was that the full length PepT1 and PepT2 protein could not be stably expressed in either *E. coli* or *S. cerevisiae*; therefore the project was dependent upon *X. laevis* oocyte assays. A large amount of work has been performed using heterologous organism based assays to probe the function of residues in the peptide transporters. When analysing the hybrid models, it became clear that mutational data into the function of the transporters had been misinterpreted due to the absence of a good model of PepT1 and PepT2 and to the limitations of assay using transiently expressed protein. Conclusions based on the mutational data and chimeric studies presented in § 7.4 were made with caution due to the limitation of the assay. That said, the results did suggest that the ECDs were not essential for transport; therefore the idea of the ECDs acting as a binding platform was explored.

The protease binding studies presented in § 7.6 suggest that both *Mm*PepT1$^{\text{ECD}}$ and *Rn*PepT2$^{\text{ECD}}$ bind trypsin in a specific manner. The interaction between trypsin and *Mm*PepT1$^{\text{ECD}}$ was explored further and the possible location of the binding site was identified. The full study remains to be completed as the final mutants could not be performed in the time left for this thesis. The *Rn*PepT2$^{\text{ECD}}$ binding needs also to be explored. The results and analysis of the *Mm*PepT1$^{\text{ECD}}$ data do suggest a possible conservation in the binding site between the two domains, and it would be interesting to see if this is the case.

The binding of PepT1$^{\mathrm{ECD}}$ to trypsin was relatively weak, 90 ± 20 μM, suggesting that under physiological conditions trypsin would exhibit a predisposition for binding PepT1, and thereby would increase the local concentration of di- and tri-peptides in the vicinity of the transporter. However, it is possible that the trypsin interacts with more of the transporter than just the ECD. The large extra-cellular loop between TM 3 and 4 sits on the opposite side on the transporter to the ECD and given the size of the loop, it is very possible that it acts as part of a binding cavity.

The functional binding of PepT2$^{\mathrm{ECD}}$ to trypsin has yet to be explained. It may be that the domain binds another trypsin-like protease. Given the number of proteases which exist in the body, this is quite an attractive hypothesis. The binding assay performed in § 7.6 only examined intestinal proteases, an organ which does not express PepT2. The tissue distribution of PepT2 in the lung and CNS suggests that PepT2 has a more diverse role to play in the bodily peptide transport than PepT1. One of these roles in the lung has been identified as part of the innate immune system (Swaan *et al.* 2008). Like many tissue proteases perform a homostatic role in the lung; as a potential site of bacterial infection neutrophil proteases, such as neutrophil elastase, are involved in the immune response to pathogenesis (Greene and McElvaney 2009) in the lung epithelium. A immune protease may act in breaking down immunogenic proteins and then aid there transport by PepT2 by also have a predisposition for binding PepT2$^{\mathrm{ECD}}$.

Ultimately, the recombinant expression, purification, crystallisation and biophysical analysis of the ECDs has given insight into a hitherto unstudied region of a highly important transport system. The aims of this thesis (§ 1.7) were thus for the most part completed.

8.2 Future work

The functional studies of the ECD and proteases was not completed. The SPR assay established an interaction between both ECDs and trypsin. The binding site characterisation of *Mm*PepT1$^{\mathrm{ECD}}$ needs to be completed and the started in *Rn*PepT2$^{\mathrm{ECD}}$. It would also be valuable to see these results confirmed using another biophysical assay. Preliminary work using ITC indicated that this technique could not be used to study the interaction as the enthalpy of binding is not high enough to be detected by the instrument. However, the interaction could be observed by microscale thermophoresis, so it would be interesting to see if the same results were obtained using this technique.

It would also be sensible to establish an experiment which looked to identify trypsin localisation to the WT transporters. The stable expression of PepT1 has been demonstrated in a number of tissues. The fluorescent labelling of Pefabloc® inactivated trypsin should, given the interaction with the ECD, localise to the plasma membrane of cell expressing the WT transporter. It would be interesting to discover what effect the binding of trypsin has on the function of the transport. A reasonable hypothesis would be that transport is increased, but this might not be the case as the binding of trypsin could obscure the substrate channel running through the transporter. Given the importance of PepT1 and PepT2, as drug delivery targets, any data which suggests a positive or negative effect on substrate transport in the peptide transporters is of great importance to the administration of drugs which utilise the the peptide transporters.

Finally, the functional expression of WT transporters is a clear future aim. All of the functional work on the WT peptide transporters presented in this report was done using transiently expressed protein in a host cell. This system, although informative, makes it hard to draw solid inferences on the exact factors which effect function, and creates an environment that can give misleading results. With the solution of the bacterial X-ray crystal structures and functional studies (Newstead *et al.* 2011; Solcan *et al.* 2012; Doki *et al.* 2013), a great leap in understanding of the transporters has been made. It is now important to build on that progress and discover the protocols for large scale expression of the WT peptide transporters.

The solution of the ECD X-ray crystal structures and the building of the hybrid models has increased the ability to do this. Prior to this work, it was unclear if the ECD was a discrete folded domain, folded into the transporter or a molten globule. Expression constructs can now be designed based on the hybrid models in order to facilitate expression. *E. coli* expression is unlikely; however, *S. cerevisiae* still needs to be adequately explored. The recent structure of NRT1.1 (Parker and Newstead 2014) utilised a *S. cerevisiae* expression system which would indicate that the PTR fold as at least amenable to it. Insert cell infect baculovirus and immortalise HEK293 cell expression may also be feasible for the right construct.

Once large scale expression of either, or hopefully both, mammalian transporters has been achieved, the significant questions regarding the function of PepT1 and PepT2 can be better addressed. For the author of this work, this would primarily comprise of further studies into how the ECD interacts with proteases. It is also not clear from the hybrid models how, if at all, the ECD interacts with the TM domain. Liposome based functional studies would be a step forward in correctly inferring function from residue mutations. Ultimately, large scale expression would allow for crystallographic studies of the transporters which could experimentally validate the hybrid models presented in this thesis.

A

Complete X-ray crystallographic data statistics

The complete crystallographic data tables for *Mm*PepT1$^{\mathrm{ECD}}$ (Table A.1) and *Rn*PepT2$^{\mathrm{ECD}}$ (Table A.2).

		Datasets	
	N1	**Hg-SAD**	**N2**
Data Collection			
Experiment type	Native (low)	Hg (SAD)	Native (high)
Date	06/10/2011	04/11/2011	21/03/2012
Beamline	I24	I04	I24
λ (Å)	0.968	1.0069	0.9686
Temperature (°C)	-173	-173	-173
ϕ (°)	0 – 120	320 – 319	0 – 30, 80 – 180
$\Delta\phi$ (°)	0.2	1.0	0.2
Exposure time (s)	0.2	1.0	0.2
Transmission (%)	40	50	40
Detector Distance (mm)	369.65	281.71	298.49
Detector	Pilatus 6M-F	Quantum 315	Pilatus 6M-F
Data processing			
Unit Cell (a, b, c)	53.07, 70.41, 111.30	53.55, 70.37, 111.60	53.45, 70.12, 111.19
(α, β, γ)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution Range	70.41 – 2.37 (2.47 – 2.37)	70.37 – 2.85 (3.05 – 2.85)	43.62 – 2.10 (2.16 – 2.10)
Measured Reflections	74130 (8384)	144862 (26394)	119801 (9797)
Unique Reflections	17446 (1929)	10365 (1823)	25021 (2027)
Multiplicity	4.2 (4.3)	14.0 (14.5)	4.8 (4.8)
Mosaicity (°)	0.114	0.14	0.09
Average B (Å^2)	56.9	61.43	43.84
Overall $\langle\lvert I\rvert/\sigma(I)\rangle$	12.8 (2.2)	15.1 (4.1)	15.9 (2.7)
Overall Completeness (%)	99.4 (99.7)	100.0 (100.0)	99.5 (99.0)
R_{merge}	0.072 (0.736)	0.172 (0.756)	0.050 (0.798)
R_{meas}	0.092 (0.946)	0.178 (0.782)	0.062 (0.988)
$\mathrm{CC}_{1/2}$	0.998 (0.737)	0.997 (0.910)	0.999 (0.480)
Model building			
R_{work}			0.193
R_{free}			0.242
Refined Residues			376
Refined H_2O			150
B factor (Å^2)			41.52
RMSD Bond Length (Å)			0.01
RMSD Bond Angle (°)			1.01
Ramachandran Plot (%)		Favoured	97.38
		Allowed	2.62
		Disallowed	0.00
Rotamer Outliers (%)			0.88
`Molprobity` (Percentile)		Clashscore	98
		Overall	99

Table A.1: Complete X-ray crystallographic data statistics for the *Mm*PepT1$^{\mathrm{ECD}}$ datasets. The values in parentheses are the highest resolution shell.

		Datasets	
	Se-SAD	**Se-pk**	**Seed**
Data collection			
Date	06/05/2013	06/05/2013	24/06/2013
Beamline	I04	I04	I03
λ (Å)	0.97942	0.97949	0.97625
Temperature (°C)		−173	
ϕ (°)	81.4−531.4	135−315	154−214
$\Delta\phi$ (°)	0.15	0.15	0.2
Exposure time (s)	0.07	0.07	0.2
Transmission (%)	50	40	20
Detector Distance (mm)	536.439	545.151	317.552
Detector	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F
Data processing			
Unit Cell (a, b, c)	96.03, 96.03, 166.60	95.75, 95.75, 163.93	43.07, 43.07, 220.53
(α, β, γ)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution Range	33.29 – 2.92 (3.00 – 2.92)	47.87 – 2.84 (3.00 – 2.84)	30.46 – 2.06 (2.12 – 2.06)
Measured Reflections	451479 (18569)	210301 (30663)	150949 (11076)
Unique Reflections	19910 (1438)	21355 (3090)	13809 (1003)
Multiplicity	22.7 (12.9)	9.8 (9.9)	10.9 (11.0)
Mosaicity (°)	0.116	0.090	0.080
Overall $\langle \lvert I \rvert / \sigma(I) \rangle$	21.0 (3.0)	13.2 (2.4)	18.9 (2.8)
Overall Completeness (%)	100.0 (100.0)	100.0 (99.8)	99.9 (99.5)
R_{merge}	0.127 (0.906)	0.130 (0.943)	0.073 (0.892)
R_{meas}	0.133 (0.981)	0.145 (1.049)	0.080 (0.977)
$CC^{1}/_{2}$	0.998 (0.900)	0.998 (0.722)	0.999 (99.5)
Model building			
R_{work}		0.204	0.197
R_{free}		0.258	0.240
Refined Residues		576	189
Refined H_2O		93	94
B factor (Å^2)		65.1	39.1
RMSD Bond Length (Å)		0.00	0.01
RMSD Bond Angle (°)		0.75	1.29
Ramachandran Plot (%)	Favoured	94.1	96.3
	Allowed	5.89	3.70
	Disallowed	0.00	0.00
Rotamer Outliers (%)		2.14	0.50
`Molprobity` (Percentile)	Clashscore	100	100
	Overall	99	94

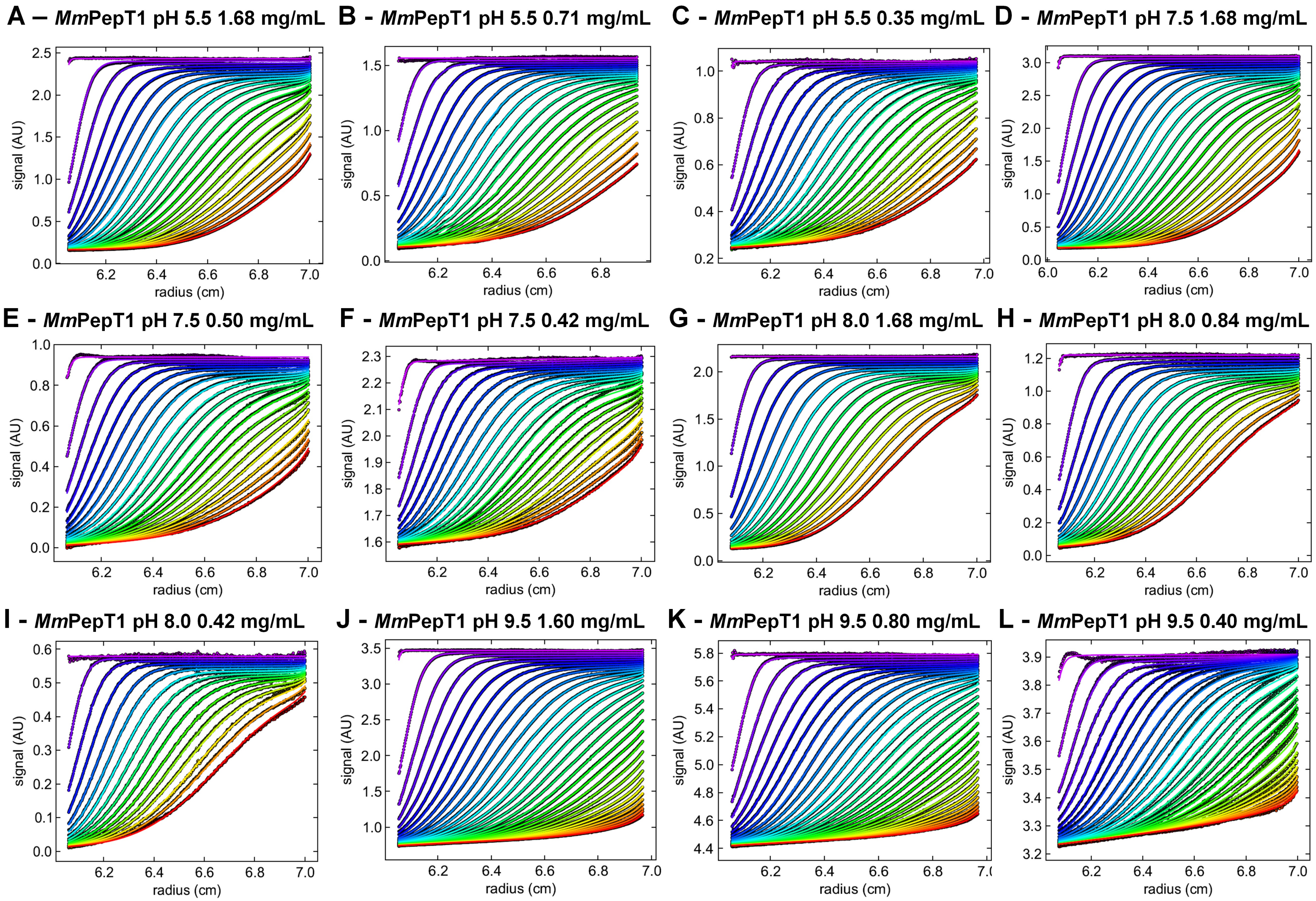
Table A.2: Complete X-ray crystallographic data statistics for the *Rn*PepT2$^{\mathbf{ECD}}$ datasets. The values in parentheses are the highest resolution shell.

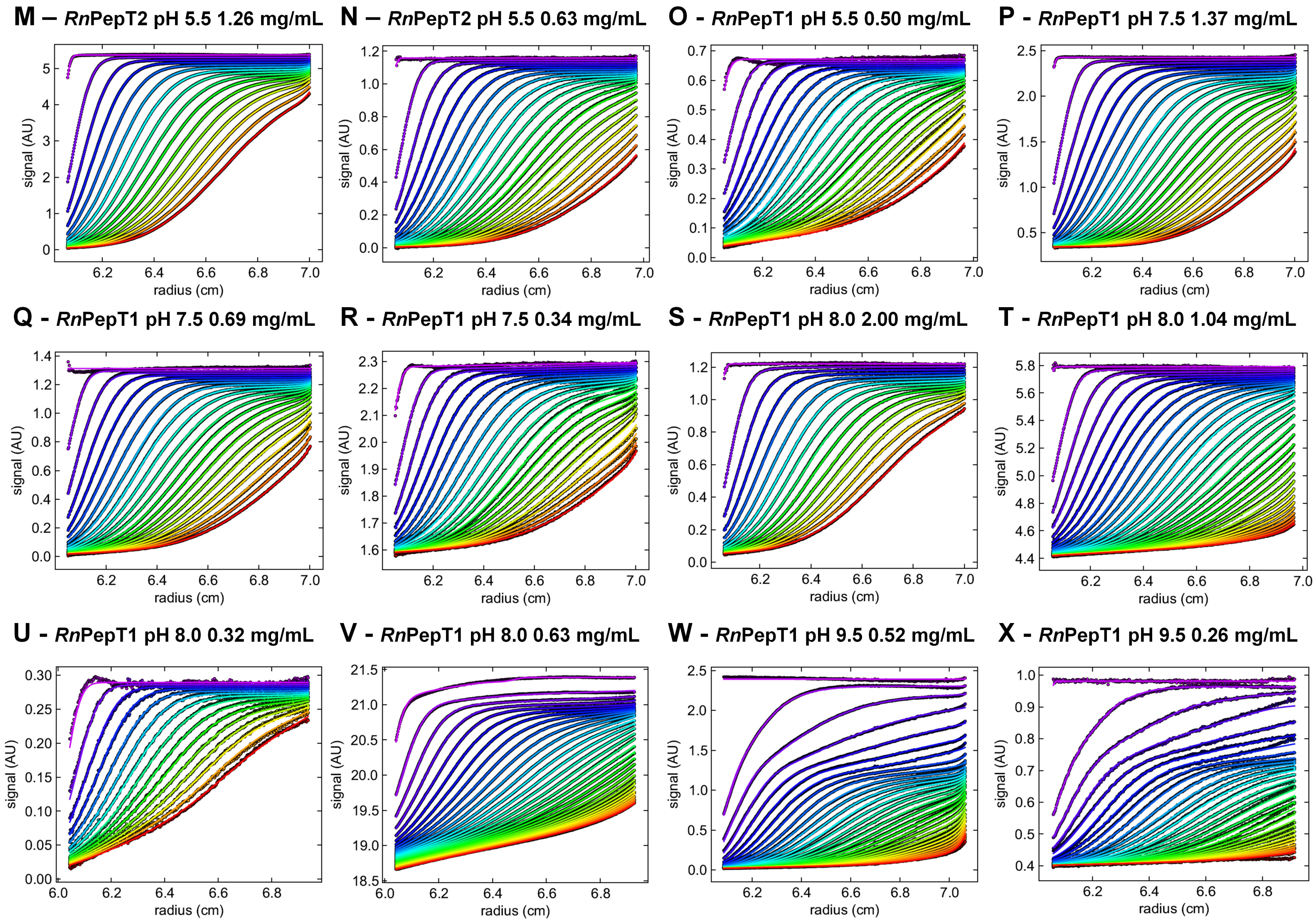
B

AUC data

B.1 Raw data Lamm equation fits

The sedimentation profile fits for *Mm*PepT1$^{\mathrm{ECD}}$ (**A**-**L**), *Rn*PepT2$^{\mathrm{ECD}}$ WT (**M**-**X**), *Rn*PepT2$^{\mathrm{ECD}}$-D505A (**Y**-**AJ**) and *Rn*PepT2$^{\mathrm{ECD}}$-D600A (**AK**-**AV**), equilibrated in either: 20 mM Na Acetate pH 5.5, 20 mM Tris pH 7.5, 20 mM Tris pH 8.0 or 20 mM Bicine pH 9.5 (all with 150 mM NaCl). The fits were calculated using `Sedfit` (Schuck 2000; Brown and Schuck 2006) and plotted using `GUSSI`.


A – MmPepT1 pH 5.5 1.68 mg/mL
B - MmPepT1 pH 5.5 0.71 mg/mL
C - MmPepT1 pH 5.5 0.35 mg/mL
D - MmPepT1 pH 7.5 1.68 mg/mL
E - MmPepT1 pH 7.5 0.50 mg/mL
F - MmPepT1 pH 7.5 0.42 mg/mL
G - MmPepT1 pH 8.0 1.68 mg/mL
H - MmPepT1 pH 8.0 0.84 mg/mL
I - MmPepT1 pH 8.0 0.42 mg/mL
J - MmPepT1 pH 9.5 1.60 mg/mL
K - MmPepT1 pH 9.5 0.80 mg/mL
L - MmPepT1 pH 9.5 0.40 mg/mL
signal (AU)
radius (cm)
6.0
6.2
6.4
6.6
6.8
7.0



M – *Rn*PepT2 pH 5.5 1.26 mg/mL

N – *Rn*PepT2 pH 5.5 0.63 mg/mL

O - *Rn*PepT1 pH 5.5 0.50 mg/mL

P - *Rn*PepT1 pH 7.5 1.37 mg/mL

Q - *Rn*PepT1 pH 7.5 0.69 mg/mL

R - *Rn*PepT1 pH 7.5 0.34 mg/mL

S - *Rn*PepT1 pH 8.0 2.00 mg/mL

T - *Rn*PepT1 pH 8.0 1.04 mg/mL

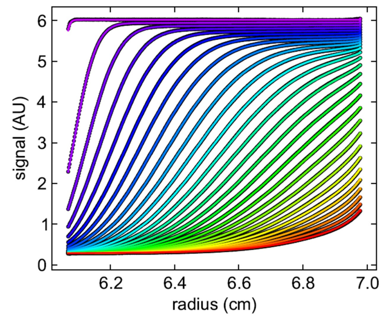
U - *Rn*PepT1 pH 8.0 0.32 mg/mL

V - *Rn*PepT1 pH 8.0 0.63 mg/mL

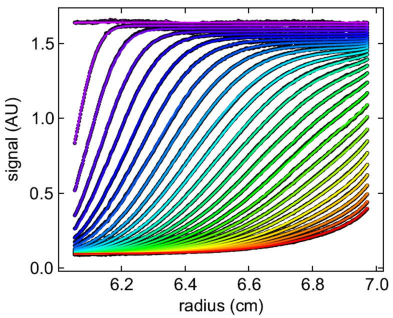
W - *Rn*PepT1 pH 9.5 0.52 mg/mL

X - *Rn*PepT1 pH 9.5 0.26 mg/mL

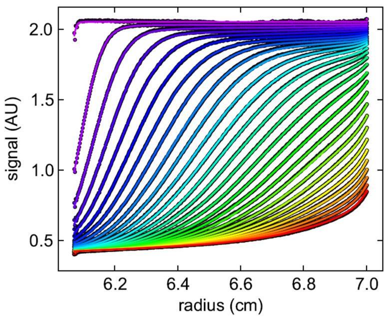
Y - **D505A pH 5.5 2.00 mg/mL**



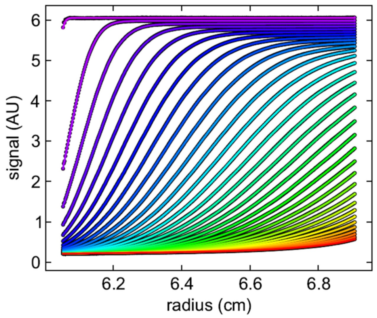
Z - **D505A pH 5.5 1.00 mg/mL**



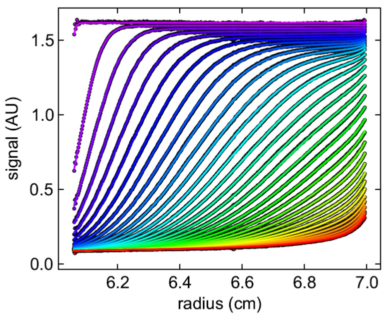
AA - **D505A pH 5.5 0.50 mg/mL**



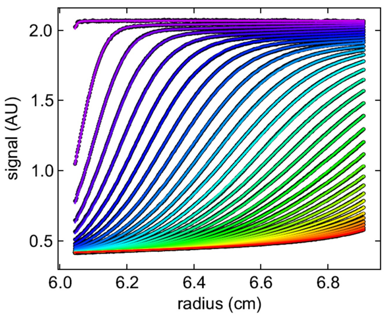
AB - **D505A pH 7.5 2.00 mg/mL**



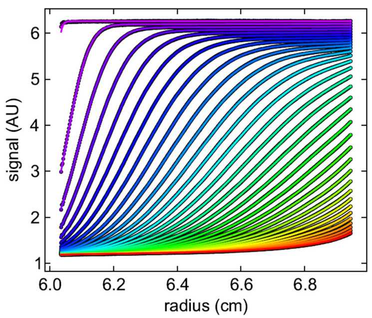
AC - **D505A pH 7.5 1.00 mg/mL**



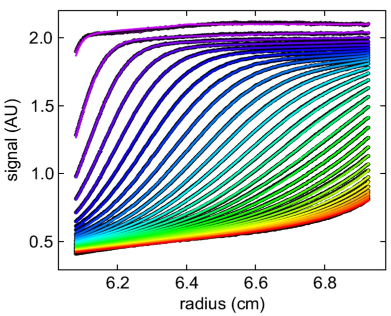
AD - **D505A pH 7.5 0.50 mg/mL**



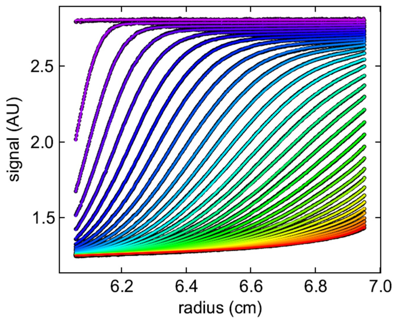
AE - **D505A pH 8.0 2.00 mg/mL**



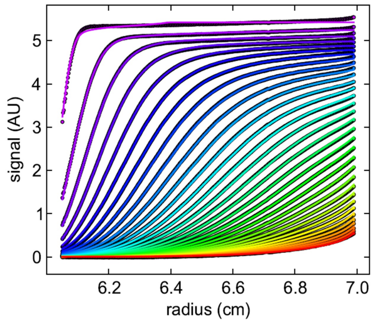
AF - **D505A pH 8.0 1.00 mg/mL**



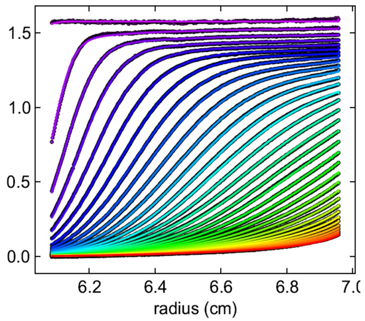
AG - **D505A pH 8.0 0.50 mg/mL**



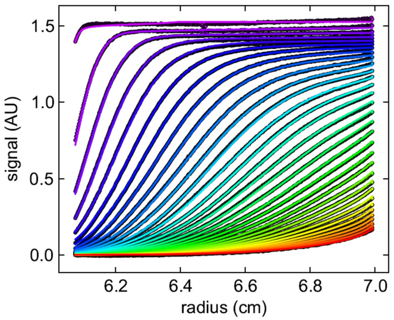
AH - **D505A pH 9.5 2.00 mg/mL**



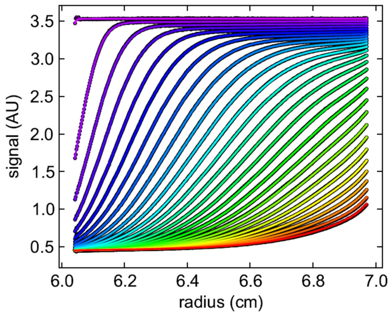
AI - **D505A pH 9.5 1.00 mg/mL**



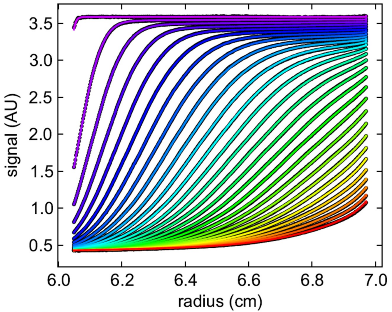
AJ - **D505A pH 9.5 0.50 mg/mL**



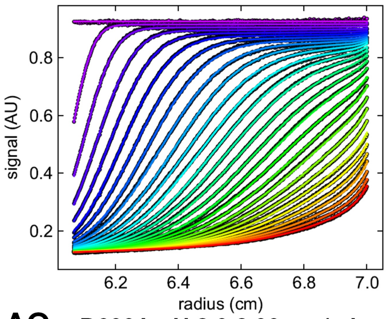
AK - D600A pH 5.5 2.00 mg/mL



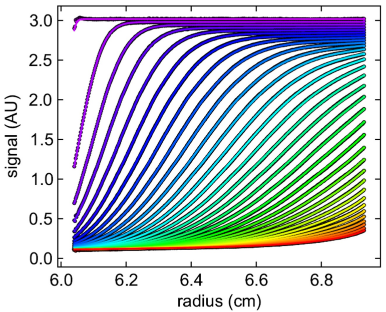
AL - D600A pH 5.5 1.00 mg/mL



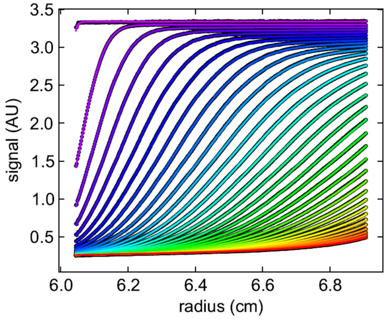
AM - D600A pH 5.5 0.50 mg/mL



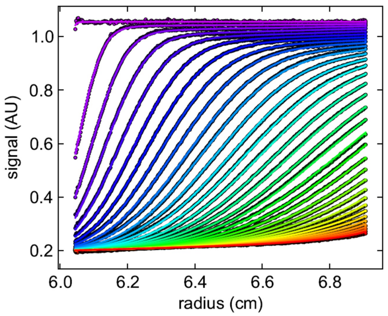
AN - D600A pH 7.5 2.00 mg/mL



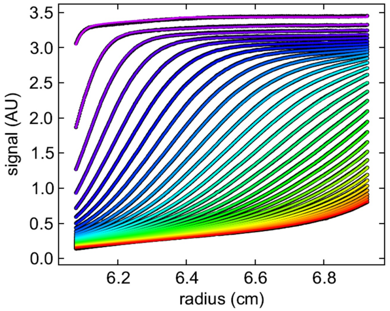
AO - D600A pH 7.5 1.00 mg/mL



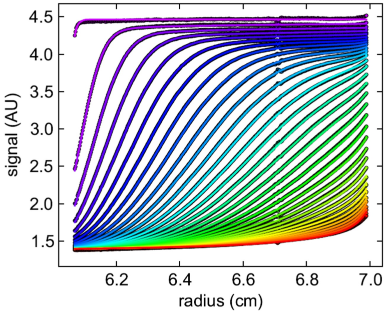
AP - D600A pH 7.5 0.50 mg/mL



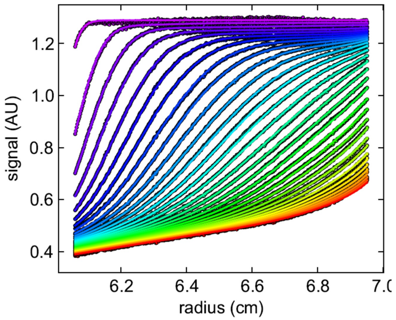
AQ - D600A pH 8.0 2.00 mg/mL



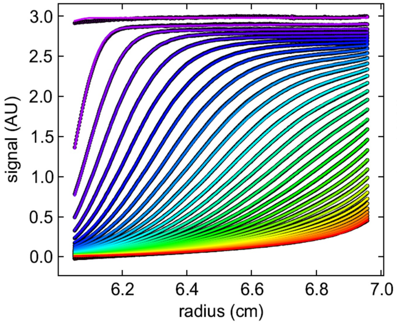
AR - D600A pH 8.0 1.00 mg/mL



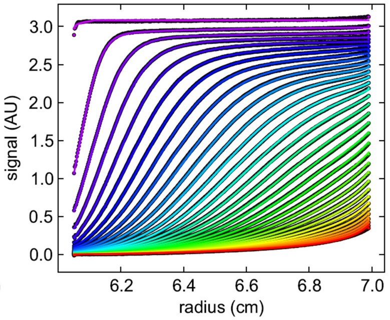
AS - D600A pH 8.0 0.50 mg/mL



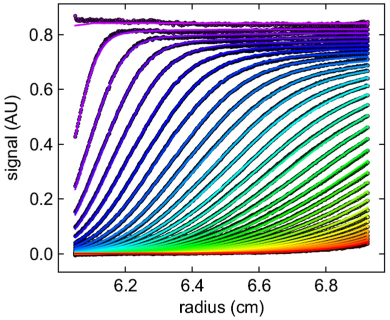
AT - D600A pH 9.5 2.00 mg/mL



AU - D600A pH 9.5 1.00 mg/mL

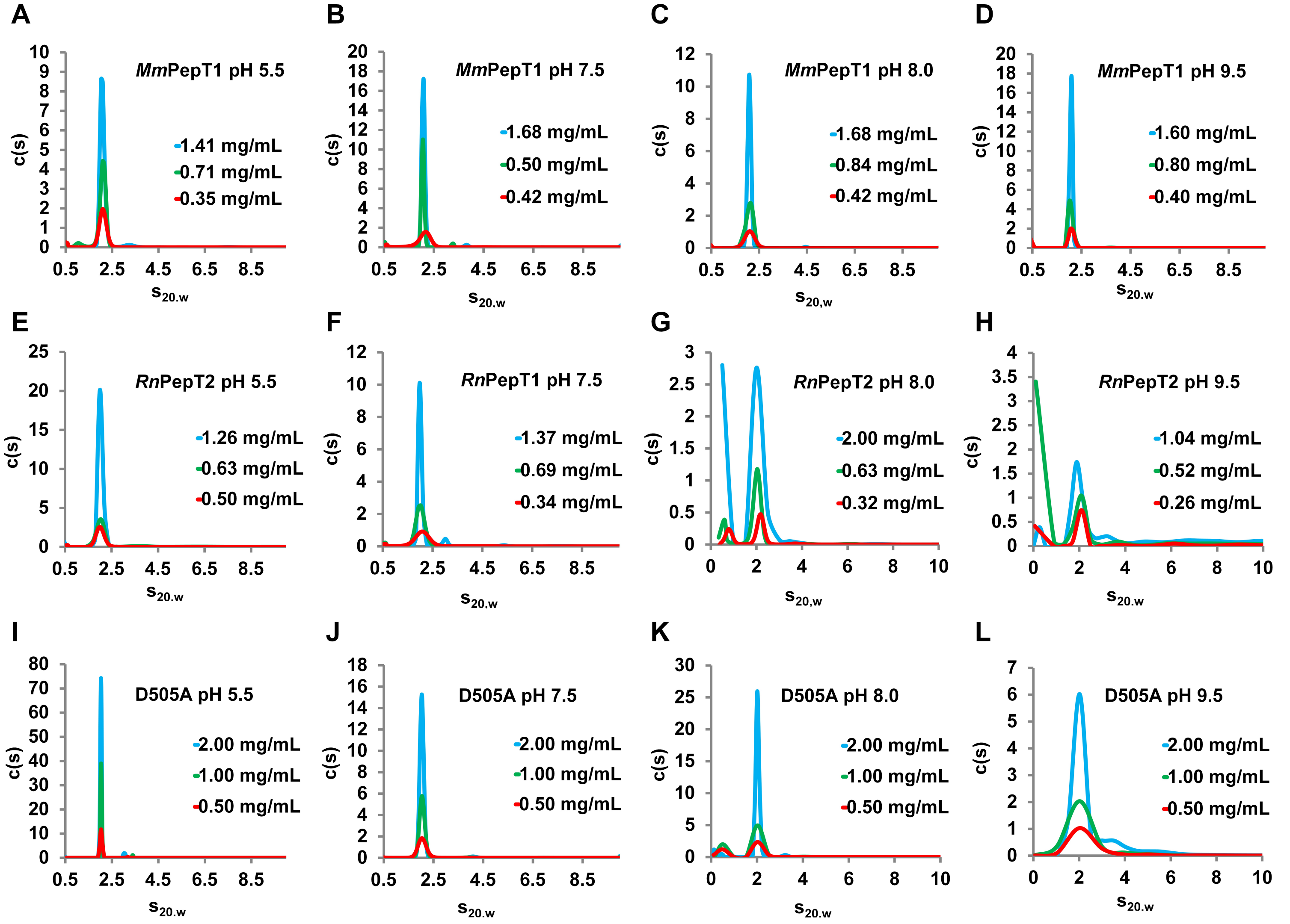


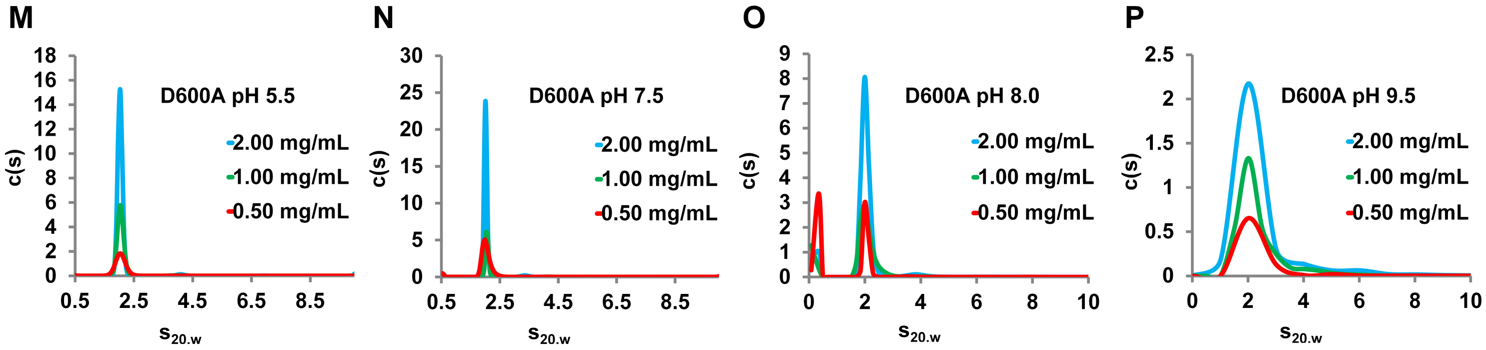
AV - D600A pH 9.5 0.50 mg/mL



B.2 c(s) distributions

The c(s) distributions for *Mm*PepT1$^{\mathrm{ECD}}$ (**A**-**D**), *Rn*PepT2$^{\mathrm{ECD}}$ WT (**E**-**H**), *Rn*PepT2$^{\mathrm{ECD}}$-D505A (**I**-**L**) and *Rn*PepT2$^{\mathrm{ECD}}$-D600A (**M**-**P**) equilibrated in either: 20 mM Na Acetate pH 5.5, 20 mM Tris pH 7.5, 20 mM Tris pH 8.0 or 20 mM Bicine pH 9.5 (all with 150 mM NaCl) calculated using `Sedfit` (Schuck 2000; Brown and Schuck 2006).


A
MmPepT1 pH 5.5
1.41 mg/mL
0.71 mg/mL
0.35 mg/mL
c(s)
0 1 2 3 4 5 6 7 8 9 10
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
B
MmPepT1 pH 7.5
1.68 mg/mL
0.50 mg/mL
0.42 mg/mL
c(s)
0 2 4 6 8 10 12 14 16 18 20
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
C
MmPepT1 pH 8.0
1.68 mg/mL
0.84 mg/mL
0.42 mg/mL
c(s)
0 2 4 6 8 10 12
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
D
MmPepT1 pH 9.5
1.60 mg/mL
0.80 mg/mL
0.40 mg/mL
c(s)
0 2 4 6 8 10 12 14 16 18 20
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
E
RnPepT2 pH 5.5
1.26 mg/mL
0.63 mg/mL
0.50 mg/mL
c(s)
0 5 10 15 20 25
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
F
RnPepT1 pH 7.5
1.37 mg/mL
0.69 mg/mL
0.34 mg/mL
c(s)
0 2 4 6 8 10 12
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
G
RnPepT2 pH 8.0
2.00 mg/mL
0.63 mg/mL
0.32 mg/mL
c(s)
0 0.5 1 1.5 2 2.5 3
0 2 4 6 8 10
$s_{20,w}$
H
RnPepT2 pH 9.5
1.04 mg/mL
0.52 mg/mL
0.26 mg/mL
c(s)
0 0.5 1 1.5 2 2.5 3 3.5 4
0 2 4 6 8 10
$s_{20,w}$
I
D505A pH 5.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0 10 20 30 40 50 60 70 80
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
J
D505A pH 7.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0 2 4 6 8 10 12 14 16 18
0.5 2.5 4.5 6.5 8.5
$s_{20,w}$
K
D505A pH 8.0
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0 5 10 15 20 25 30
0 2 4 6 8 10
$s_{20,w}$
L
D505A pH 9.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0 1 2 3 4 5 6 7
0 2 4 6 8 10
$s_{20,w}$


M
D600A pH 5.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0
2
4
6
8
10
12
14
16
18
0.5
2.5
4.5
6.5
8.5
$s_{20.w}$
N
D600A pH 7.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0
5
10
15
20
25
30
0.5
2.5
4.5
6.5
8.5
$s_{20.w}$
O
D600A pH 8.0
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0
1
2
3
4
5
6
7
8
9
0
2
4
6
8
10
$s_{20.w}$
P
D600A pH 9.5
2.00 mg/mL
1.00 mg/mL
0.50 mg/mL
c(s)
0
0.5
1
1.5
2
2.5
0
2
4
6
8
10
$s_{20.w}$

Bibliography

Abouhamad, W. N. and Manson, M. D. (1994) The dipeptide permease of Escherichia coli closely resembles other bacterial transport systems and shows growth-phase-dependent expression. *Molecular microbiology.* 14 (5). p.1077–1092.

Abrahams, J. P. and Leslie, A. G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta crystallographica. Section D, Biological crystallography.* 52 (1). p.30–42.

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H., and Iwata, S. (2003) Structure and mechanism of the lactose permease of Escherichia coli. *Science.* 301 (5633). p.610–615.

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., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography.* 66 (2). p.213–221.

Addison, J. M., Burston, D, Payne, J. W., Wilkinson, S, and Matthews, D. M. (1975) Evidence for active transport of tripeptides by hamster jejunum in vitro. *Clinical science and molecular medicine.* 49 (4). p.305–312.

Adibi, S. A., Krzysik, B. A., and Drash, A. L. (1977) Metabolism of intravenously administered dipeptides in rats: effects on amino acid pools, glucose concentration and insulin and glucagon secretion. *Clinical Science and Molecular Medicine.* 52 (2). p.193–204.

Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography.* 66 (11). p.1153–1163.

Agar, W. T., Hird, F. J., and Sidhu, G. S. (1953) The active absorption of amino-acids by the intestine. *The Journal of physiology.* 121 (2). p.255–263.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of molecular biology.* 215 (3). p.403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research.* 25 (17). p.3389–3402.

Arndt, U. W., Crowther, R. A., and Mallett, J. F. (1968) A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. *Journal of scientific instruments.* 1 (5). p.510–516.

Ashida, K., Katsura, T., Motohashi, H., Saito, H., and Inui, K. I. (2002) Thyroid hormone regulates the activity and expression of the peptide transporter PEPT1 in Caco-2 cells. *American journal of physiology. Gastrointestinal and liver physiology.* 282 (4). p.617–623.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences.* 98 (18). p.37–41.

Barak, Z. and Gilbarg, C. (1975) Specialized peptide transport system in Escherichia coli. *Journal of bacteriology.* 122 (3). p.1200–1207.

Barak, Z. and Gilvarg, C. (1974) Triornithine-resistant strains of Escherichia coli. Isolation, definition, and genetic studies. *The Journal of biological chemistry.* 249 (1). p.143–148.

Barry, R. J., Dikstein, S., Matthews, J., Smyth, D. H., and Wright, E. M. (1964) Electrical Potentials associated with intestinal sugar transfer. *The Journal of physiology.* 171. p.316–338.

Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins.* 73 (3). p.765–783.

Beitz, E. (2000a) TEXshade: shading and labeling of multiple sequence alignments using LATEX2 epsilon. *Bioinformatics.* 16 (2). p.135–139.

Beitz, E. (2000b) T(E)Xtopo: shaded membrane protein topology plots in LAT(E)X2epsilon. *Bioinformatics.* 16 (11). p.1050–1051.

Berger, U. V. and Hediger, M. A. (1999) Distribution of peptide transporter PEPT2 mRNA in the rat nervous system. *Anatomy and embryology.* 199 (5). p.439–449.

Berlioz, F., Maoret, J. J., Paris, H., Laburthe, M., Farinotti, R., and Rozé, C. (2000) alpha(2)-adrenergic receptors stimulate oligopeptide transport in a human intestinal cell line. *The Journal of pharmacology and experimental therapeutics.* 294 (2). p.466–472.

Berne, R. and Levy, M. (2008a) The gastric phase of the intergated reponse to a meal. In: *Physiology.* Ed. by B. Koeppen and B. Stanton. 6th ed. St. Louis: Mosby. Chap. 5. p.505–515.

Berne, R. and Levy, M. (2008b) The small intestinal phase of the integrated response to a meal. In: *Physiology.* Ed. by B. Koeppen and B. Stanton. 6th ed. St. Louis: Mosby. Chap. 5. p.516–532.

Bird, L. E. (2011) High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods.* 55 (1). p.29–37.

Bishop, W. R. and Bell, R. M. (1988) Assembly of phospholipids into cellular membranes: biosynthesis, transmembrane movement and intracellular translocation. *Annual review of cell biology.* 4. p.579–610.

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 crystallographica. Section D, Biological crystallography.* 60 (12). p.2210–2221.

Bolger, M. B., Haworth, I. S., Yeung, A. K., Ann, D., von Grafenstein, H., Hamm-Alvarez, S., Okamoto, C. T., Kim, K. J., Basu, S. K., Wu, S, and Lee, V. H. (1998) Structure, function, and molecular modeling approaches to the study of the intestinal dipeptide transporter PepT1. *Journal of pharmaceutical sciences.* 87 (11). p.1286–1291.

Boll, M., Herget, M., Wagener, M., Weber, W. M., Markovich, D., Biber, J., Clauss, W., Murer, H., and Daniel, H. (1996) Expression cloning and functional characterization of the kidney cortex high-affinity proton-coupled peptide transporter. *Proceedings of the National Academy of Sciences.* 93 (1). p.284–289.

Bourenkov, G. P. and Popov, A. N. (2010) Optimization of data collection taking radiation damage into account. *Acta crystallographica. Section D, Biological crystallography.* 66 (4). p.409–419.

Boyd, C. A. and Ward, M. R. (1982) A micro-electrode study of oligopeptide absorption by the small intestinal epithelium of Necturus maculosus. *Journal of physiology.* 324 (1). p.411–428.

Brandsch, M., Miyamoto, Y., Ganapathy, V., and Leibach, F. H. (1994) Expression and protein kinase C-dependent regulation of peptide/H+ co-transport system in the Caco-2 human colon carcinoma cell line. *The Biochemical journal.* 299 (1). p.253–260.

Brandsch, M., Knütter, I., and Bosse-Doenecke, E. (2008) Pharmaceutical and pharmacological importance of peptide transporters. *The Journal of pharmacy and pharmacology.* 60 (5). p.543–585.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003) Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta crystallographica. Section D, Biological crystallography.* 59 (11). p.2023–2030.

Brown, P. H. and Schuck, P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophysical journal.* 90 (12). p.4651–4661.

Brünger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature.* 355 (6359). p.472–475.

Buyse, M., Berlioz, F., Guilmeau, S., Tsocas, A., Voisin, T., Péranzi, G., Merlin, D., Laburthe, M., Lewin, M. J., Rozé, C., and Bado, A. (2001) PepT1-mediated epithelial transport of dipeptides and cephalexin is enhanced by luminal leptin in the small intestine. *The Journal of clinical investigation.* 108 (10). p.1483–1494.

Chang, A. B., Lin, R., Keith Studley, W., Tran, C. V., and Saier, M. H. (2009) Phylogeny as a guide to structure and function of membrane transport proteins. *Molecular membrane biology.* 21 (3). p.171–181.

Chayen, N. E. and Saridakis, E. (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nature methods.* 5 (2). p.147–153.

Cheeseman, C. I., Newey, H., and Smyth, D. H. (1971) Effect of amino acids on the hydrolysis of dipeptides by rat small intestine. *The Journal of physiology.* 212 (2). p.25–26.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography.* 66 (1). p.12–21.

Choi, S., Connelly, S., Reixach, N., Wilson, I. A., and Kelly, J. W. (2010) Chemoselective small molecules that covalently modify one lysine in a non-enzyme protein in plasma. *Nature chemical biology.* 6 (2). p.133–139.

Christopher, G., Phipps, A., and Gray, R. (1998) Temperature-dependent solubility of selected proteins. *Journal of Crystal Growth.* 191 (4). p.820–826.

Cole, C., Barber, J. D., and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic acids research.* 36. p.197–201.

Cole, J. L., Lary, J. W., P Moody, T., and Laue, T. M. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods in cell biology.* 84. p.143–179.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007) Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta crystallographica. Section D, Biological crystallography.* 63 (5). p.636–645.

Cowtan, K. (1994) Dm: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography.* 31. p.34–38.

Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta crystallographica. Section D, Biological crystallography.* 62 (9). p.1002–1011.

Cowtan, K. (2012) Completion of autobuilt protein models using a database of protein fragments. *Acta crystallographica. Section D, Biological crystallography.* 68 (4). p.328–335.

Crane, R. K., Forstner, G., and Eichholz, A. (1965) Studies on the mechanism of the intestinal absorption of sugars X. An effect of Na+ concentration on the apparent michaelis constants for intestinal sugar transport, in vitro. *Biochimica et Biophysica Acta (BBA) - Biophysics including Photosynthesis.* 109 (2). p.467–477.

Dalmasso, G., Nguyen, H. T. T., Charrier-Hisamuddin, L., Yan, Y., Laroui, H., Demoulin, B., Sitaraman, S. V., and Merlin, D. (2010) PepT1 mediates transport of the proinflammatory bacterial tripeptide L-Ala-{gamma}-D-Glu-meso-DAP in intestinal epithelial cells. *American journal of physiology. Gastrointestinal and liver physiology.* 299 (3). p.687–696.

Daniel, H. and Adibi, S. A. (1995) Selective effect of zinc on uphill transport of oligopeptides into kidney brush border membrane vesicles. *FASEB Journal.* 9 (11). p.1112–1117.

Daniel, H., Spanier, B., Kottra, G., and Weitz, D. (2006) From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology.* 21 (2). p.93–102.

D'Arcy, A. (1994) Crystallizing proteins - a rational approach? *Acta crystallographica. Section D, Biological crystallography.* 50 (Pt 4). p.469–471.

D'Arcy, A., Villard, F., and Marsh, M. (2007) An automated microseed matrix-screening method for protein crystallization. *Acta crystallographica. Section D, Biological crystallography.* 63 (4). p.550–554.

Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta crystallographica. Section D, Biological crystallography.* 56 (8). p.232–237.

de La Fortelle, E. and Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology.* Methods in Enzymology 276. p.472–494.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR.* 6 (3). p.277–293.

Diederichs, K. and Karplus, P. A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature Structural Bology.* 4 (4). p.269–275.

Do, C. B., Mahabhashyam, M. S. P., Brudno, M., and Batzoglou, S. (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Research.* 15 (2). p.330–340.

Doki, S., Kato, H. E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N., Tsukazaki, T., Sugita, Y., Kandori, H., Newstead, S., Ishitani, R., and Nureki, O. (2013) Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proceedings of the National Academy of Sciences of the United States of America.* 110 (28). p.11343–11348.

Dong, A., Xu, X., Edwards, A. M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., Filippova, E., Gu, J., Guthrie, J., Ignatchenko, A., Joachimiak, A., Klostermann, N., Kim, Y., Korniyenko, Y., Minor, W., Que, Q., Savchenko, A., Skarina, T., Tan, K., Yakunin, A., Yee, A., Yim, V., Zhang, R., Zheng, H., Akutsu, M., Arrowsmith, C., Avvakumov, G. V., Bochkarev, A., Dahlgren, L. G., Dhe-Paganon, S., Dimov, S., Dombrovski, L., Finerty, P., Flodin, S., Flores, A., Gräslund, S., Hammerström, M., Herman, M. D., Hong, B. S., Hui, R., Johansson, I., Liu, Y., Nilsson, M., Nedyalkova, L., Nordlund, P., Nyman, T., Min, J., Ouyang, H., Park, H. W., Qi, C., Rabeh, W., Shen, L., Shen, Y., Sukumard, D., Tempel, W., Tong, Y., Tresagues, L., Vedadi, M., Walker, J. R., Weigelt, J., Welin, M., Wu, H., Xiao, T., Zeng, H., and Zhu, H. (2007) In situ proteolysis for protein crystallization and structure determination. *Nature methods.* 4 (12). p.1019–1021.

Döring, F., Dorn, D., Bachfischer, U., Amasheh, S., Herget, M., and Daniel, H. (1996) Functional analysis of a chimeric mammalian peptide transporter derived from the intestinal and renal isoforms. *The Journal of physiology.* 497 (3). p.773–779.

Döring, F., Martini, C., Walter, J., and Daniel, H. (2002) Importance of a small N-terminal region in mammalian peptide transporters for substrate affinity and function. *The Journal of membrane biology.* 186 (2). p.55–62.

Dringen, R., Hamprecht, B., and Bröer, S. (1998) The peptide transporter PepT2 mediates the uptake of the glutathione precursor CysGly in astroglia-rich primary cultures. *Journal of neurochemistry.* 71 (1). p.388–393.

Dunn, M. S. and Lewis, H. B. (1921) A comparative study of the hydrolysis of caesin and deaminized caesin by proteolytic enzymes. *Journal of Biological Chemistry.* 49 (2). p.343–350.

Dupeux, F., Röwer, M., Seroul, G., Blot, D., and Márquez, J. A. (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta crystallographica. Section D, Biological crystallography.* 67 (11). p.915–919.

Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *Journal of structural biology.* 169 (1). p.45–53.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography.* 60 (12). p.2126–2132.

Eneqvist, T., Lundberg, E., Nilsson, L., Abagyan, R., and Sauer-Eriksson, A. E. (2003) The transthyretin-related protein family. *European Journal of Biochemistry.* 270 (3). p.518–532.

Eramian, D., Shen, M. Y., Devos, D., Melo, F., Sali, A., and Marti-Renom, M. A. (2006) A composite score for predicting errors in protein structure models. *Protein Science.* 15 (7). p.1653–1666.

Erickson, H. P. (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological procedures online.* 11. p.32–51.

Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N., and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry.* 357 (2). p.289–298.

Evans, G. and Pettifer, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of Applied Crystallography.* 34 (34). p.82–86.

Evans, P. (2006) Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography.* 62 (1). p.72–82.

Evans, P. R. and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography.* 69 (7). p.1204–1214.

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 (6471). p.563–566.

Fei, Y. J., Liu, J. C., Fujita, T., Liang, R., Ganapathy, V., and Leibach, F. H. (1998) Identification of a potential substrate binding domain in the mammalian peptide transporters

PEPT1 and PEPT2 using PEPT1-PEPT2 and PEPT2-PEPT1 chimeras. *Biochemical and biophysical research communications.* 246 (1). p.39–44.

Fisher, R. B. (1954) *Protein metabolism.* 1st ed. London: Methuen and Co., Ltd., London.

Fogh, R. H., Boucher, W., Vranken, W. F., Pajon, A., Stevens, T. J., Bhat, T. N., Westbrook, J., Ionides, J. M. C., and Laue, E. D. (2005) A framework for scientific data modeling and automated software development. *Bioinformatics.* 21 (8). p.1678–1684.

Forrest, L. R., Krämer, R., and Ziegler, C. (2011) The structural basis of secondary active transport mechanisms. *Biochimica et biophysica acta.* 1807 (2). p.167–188.

Forstner, G., Sabesin, S., and Isselbac, K. (1968) Rat intestinal microvillus membranes. Purification and biochemical characterization. *Biochemical Journal.* 106 (2). p.381–390.

Fox, N. K., Brenner, S. E., and Chandonia, J. M. (2014) SCOPe: Structural Classification of Proteins–extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic acids research.* 42. p.304–309.

Frankel, E. M. (1916) A comparative study of the behavior of purified proteins towards proteolytic enzymes. *Journal of Biological Chemistry.* 26 (1). p.31–59.

Fredriksson, R., Nordström, K. J. V., Stephansson, O., Hägglund, M. G. A., and Schiöth, H. B. (2008) The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. *FEBS letters.* 582 (27). p.3811–3816.

Fruton, J. S. and Bergmann, M. (1939) The specificity of pepsin action. *Science.* 87 (127). p.627–641.

Ganapathy, M. E., Huang, W., Wang, H., Ganapathy, V., and Leibach, F. H. (1998) Valacyclovir: a substrate for the intestinal and renal peptide transporters PEPT1 and PEPT2. *Biochemical and biophysical research communications.* 246 (2). p.470–475.

Ganapathy, V. and Leibach, F. H. (1985) Is intestinal peptide transport energized by a proton gradient? *The American journal of physiology.* 249 (1). p.153–60.

Ganapathy, V. and Leibach, F. (1983) Role of pH gradient and membrane potential in dipeptide transport in intestinal and renal brush-border membrane vesicles from the rabbit. Studies with L-carnosine and glycyl-L-proline. *Journal of Biological Chemistry.* 258 (23). p.4189–4192.

Ganapathy, V. and Radhakrishnan, A. N. (1980) Sodium-dependent inhibition of amino acid and dipeptide transport by harmaline in monkey small intestine. *Biochemical pharmacology.* 29 (5). p.713–716.

Ganapathy, V., Mendicino, J., Pashley, D. H., and Leibach, F. H. (1980) Carrier-mediated transport of glycyl-L-proline in renal brush border vesicles. *Biochemical and Biophysical Research Communications.* 97 (3). p.1133–1139.

Ganapthy, V., Mendicino, J., and Leibach, F. H. (1981) Evidence for a dipeptide transport system in renal brush border membranes from rabbit. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 642 (2). p.381–391.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook.* Ed. by J. M. Walker. New York: Humana Press.

Greene, C. M. and McElvaney, N. G. (2009) Proteases and antiproteases in chronic neutrophilic lung disease - relevance to drug discovery. *British journal of pharmacology.* 158 (4). p.1048–58.

Groneberg, D. A., Nickolaus, M., Springer, J., Döring, F., Daniel, H., and Fischer, A. (2001) Localization of the peptide transporter PEPT2 in the lung: implications for pulmonary oligopeptide uptake. *The American journal of pathology.* 158 (2). p.707–714.

Groneberg, D. A., Eynott, P. R., Döring, F., Dinh, Q. T., Oates, T., Barnes, P. J., Chung, K. F., Daniel, H., and Fischer, A. (2002) Distribution and function of the peptide transporter PEPT2 in normal and cystic fibrosis human lung. *Thorax.* 57 (1). p.55–60.

Grosse-Kunstleve, R. W. and Adams, P. D. (2003) Substructure search procedures for macromolecular structures. *Acta crystallographica. Section D, Biological crystallography.* 59 (11). p.1966–1973.

Guettou, F., Quistgaard, E. M., Trésaugues, L., Moberg, P., Jegerschöld, C., Zhu, L., Jong, A. J. O., Nordlund, P., and Löw, C. (2013) Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO reports.* 14 (9). p.804–810.

Guinier, A. (1939) La diffraction des rayons X aux tres petits angles : application al'etude de phenomenes ultramicroscopiques. *Annals of Physics.* 3 (12). p.161–237.

Hagting, A., Kunji, E. R., Leenhouts, K. J., Poolman, B., and Konings, W. N. (1994) The di- and tripeptide transport protein of Lactococcus lactis. A new type of bacterial peptide transporter. *The Journal of biological chemistry.* 269 (15). p.11391–11399.

Hamilton, J. A. and Benson, M. D. (2001) Transthyretin: a review from a structural perspective. *Cellular and molecular life sciences.* 58 (10). p.1491–1521.

Harding, S. E. (1995) On the hydrodynamic analysis of macromolecular conformation. *Biophysical Chemistry.* 55 (1). p.69–93.

Hediger, M. A., Romero, M. F., Peng, J.-B., Rolfs, A., Takanaga, H., and Bruford, E. A. (2004) The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *European journal of physiology.* 447 (5). p.465–468.

Hellier, M. D., Thirumalai, C., and Holdsworth, C. D. (1973) The effect of amino acids and dipeptides on sodium and water absorption in man. *Gut.* 14 (1). p.41–5.

Helliwell, J. (2005) *Macromolecular Crystallography with Synchrotron Radiation.* 1st ed. Cambridge: Cambridge University Press. p.181–210.

Hennebry, S. C. (2009) Evolutionary changes to transthyretin: structure and function of a transthyretin-like ancestral protein. *The FEBS journal.* 276 (19). p.5367–5379.

Hennebry, S. C., Law, R. H. P., Richardson, S. J., Buckle, A. M., and Whisstock, J. C. (2006) The crystal structure of the transthyretin-like protein from Salmonella dublin, a prokaryote 5-hydroxyisourate hydrolase. *Journal of molecular biology.* 359 (5). p.1389–1399.

Higgins, C. F., Hiles, I. D., Whalley, K., and Jamieson, D. J. (1985) Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. *The EMBO journal.* 4 (4). p.1033–1039.

Higgins, C. F., Hiles, I. D., Salmond, G. P., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., and Bell, A. W. (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature.* 323 (6087). p.448–450.

Hiles, I. D., Gallagher, M. P., Jamieson, D. J., and Higgins, C. F. (1987) Molecular characterization of the oligopeptide permease of Salmonella typhimurium. *Journal of molecular biology.* 195 (1). p.125–142.

Himukai, M, Suzuki, Y, and Hoshi, T (1978) Differences in characteristics between glycine and glycylglycine transport in guinea pig small intestine. *The Japanese journal of physiology.* 28 (4). p.499–510.

Hinder, R. A. and Kelly, K. A. (1977) Canine gastric emptying of solids and liquids. *The American journal of physiology.* 233 (4). p.335–340.

Hogarth, B. G. and Higgins, C. F. (1983) Genetic organization of the oligopeptide permease (opp) locus of Salmonella typhimurium and Escherichia coli. *Journal of bacteriology.* 153 (3). p.1548–1551.

Holm, L. and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic acids research.* 38. p.545–549.

Holton, J. M. (2009) A beginner's guide to radiation damage. *Journal of synchrotron radiation.* 16 (2). p.133–142.

Hu, Y., Shen, H., Keep, R. F., and Smith, D. E. (2007) Peptide transporter 2 (PEPT2) expression in brain protects against 5-aminolevulinic acid neurotoxicity. *Journal of neurochemistry.* 103 (5). p.2058–2065.

Hu, Y., Smith, D. E., Ma, K., Jappar, D., Thomas, W., and Hillgren, K. M. (2008) Targeted disruption of peptide transporter PepT1 gene in mice significantly reduces dipeptide absorption in intestine. *Molecular pharmaceutics.* 5 (6). p.1122–1130.

Huang, Y., Lemieux, M. J., Song, J., Auer, M., and Wang, D.-N. (2003) Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science.* 301 (5633). p.616–20.

Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., Tsutakawa, S. E., Jenney, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature methods.* 6 (8). p.606–612.

Hyde, S. C., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D. R., Hubbard, R. E., and Higgins, C. F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature.* 346 (6282). p.362–365.

Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation.* 16 (6). p.872–879.

Inui, K., Okano, T., Takano, M., Saito, H., and Hori, R. (1984) Carrier-mediated transport of cephalexin via the dipeptide transport system in rat renal brush-border membrane vesicles. *Biochimica et biophysica acta.* 769 (2). p.449–454.

Inui, K., Okano, T., Maegawa, H., Kato, M., Takano, M., and Hori, R. (1988) H+ coupled transport of p.o. cephalosporins via dipeptide carriers in rabbit intestinal brush-border membranes: difference of transport characteristics between cefixime and cephradine. *The Journal of pharmacology and experimental therapeutics.* 247 (1). p.235–241.

Ireton, G. C. and Stoddard, B. L. (2004) Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta crystallographica. Section D, Biological crystallography.* 60 (3). p.601–605.

Jardetzky, O. (1966) Simple Allosteric Model for Membrane Pumps. *Nature.* 211 (5052). p.969–970.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography.* 26 (6). p.795–800.

Kabsch, W. (2010) XDS. *Acta crystallographica. Section D, Biological crystallography.* 66 (Pt 2). p.125–132.

Kantardjieff, K. A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science.* 12 (9). p.1865–1871.

Karplus, P. A. and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science.* 336 (6084). p.1030–1033.

Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., Nocek, B., Cuff, M., Osipiuk, J., Tan, K., Fan, Y., Bigelow, L., Maltseva, N., Wu, R., Borovilos, M., Duggan, E., Zhou, M., Binkowski, T. A., Zhang, R. G., and Joachimiak, A. (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nature methods.* 5 (10). p.853–854.

Kohn, P. G., Smyth, D. H., and Wright, E. M. (1968) Effects of amino acids, dipeptides and disaccharides on the electric potential across rat small intestine. *The Journal of physiology.* 196 (3). p.723–746.

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. *Journal of Applied Crystallography.* 36 (5). p.1277–1282.

Kong, W., McConalogue, K., Khitin, L. M., Hollenberg, M. D., Payan, D. G., Bohm, S. K., and Bunnett, N. W. (1997) Luminal trypsin may regulate enterocytes through proteinase-activated receptor 2. *Proceedings of the National Academy of Sciences.* 94 (16). p.8884–8889.

Koshikawa, N., Hasegawa, S., Nagashima, Y., Mitsuhashi, K., Tsubota, Y., Miyata, S., Miyagi, Y., Yasumitsu, H., and Miyazaki, K. (1998) Expression of trypsin by epithelial cells of various tissues, leukocytes, and neurons in human and mouse. *The American journal of pathology.* 153 (3). p.937–944.

Kowalczuk, S., Bröer, A., Tietze, N., Vanslambrouck, J. M., Rasko, J. E. J., and Bröer, S. (2008) A protein complex in the brush-border membrane explains a Hartnup disorder allele. *FASEB Journal.* 22 (8). p.2880–2887.

Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography.* 60 (12). p.2256–2268.

Krissinel, E. and Henrick, K. (2005) *Computational Life Sciences.* Ed. by M. R. Berthold, R. C. Glen, K. Diederichs, O. Kohlbacher, and I. Fischer. Vol. 3695. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. p.163–174.

Krogh, A, Larsson, B, von Heijne, G, and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology.* 305 (3). p.567–80.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics.* 23 (21). p.2947–2948.

Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science.* p.90–125.

Law, C. J., Maloney, P. C., and Wang, D. N. (2008) Ins and Outs of Major Facilitator Superfamily Antiporters. *Annual review of microbiology.* 62. p.289–305.

Law, C. J., Yang, Q., Soudant, C., Maloney, P. C., and Wang, D. N. (2007) Kinetic evidence is consistent with the rocker-switch mechanism of membrane transport by GlpT. *Biochemistry.* 46 (43). p.12190–12197.

Law, C. J., Almqvist, J., Bernstein, A., Goetz, R. M., Huang, Y., Soudant, C., Laaksonen, A., Hovmöller, S., and Wang, D. N. (2008) Salt-bridge Dynamics Control Substrate-induced Conformational Change in the Membrane Transporter GlpT. *Journal of Molecular Biology.* 378 (4). p.828–839.

Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins.* 61 (4). p.704–721.

Li, L., Wang, D., Zhang, X., and Song, X. (2013) Expression of PEPT2 mRNA in the lung of rat with bleomycin-induced pulmonary fibrosis. *The Chinese-German Journal of Clinical Oncology.* 12 (10). p.468–472.

Liang, R., Fei, Y. J., Prasad, P. D., Ramamoorthy, S., Han, H., Yang-Feng, T. L., Hediger, M. A., Ganapathy, V., and Leibach, F. H. (1995) Human intestinal H+/peptide cotransporter. Cloning, functional expression, and chromosomal localization. *The Journal of biological chemistry.* 270 (12). p.6456–6463.

Liu, W., Liang, R., Ramamoorthy, S., Fei, Y., Ganapathy, M., Hediger, M., Ganapathy, V., and Leibach, F. (1995) Molecular cloning of PEPT 2, a new member of the H+/peptide cotransporter family, from human kidney. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 1235 (2). p.461–466.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by C-alpha geometry: phi,psi and C-beta deviation. *Proteins.* 50 (3). p.437–450.

Lunin, V. Y. (1988) Use of the information on electron density distribution in macromolecules. *Acta Crystallographica Section A Foundations of Crystallography.* 44 (2). p.144–150.

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 reports.* 15 (8). p.886–893.

Maiden, M. C., Davis, E. O., Baldwin, S. A., Moore, D. C., and Henderson, P. J. (1987) Mammalian and bacterial sugar transport proteins are homologous. *Nature.* 325 (6105). p.641–643.

Malhotra, A. (2009) Tagging for protein expression. *Methods in enzymology.* 463. p.239–58.

Marasco, W. A., Phan, S. H., Krutzsch, H., Showell, H. J., Feltner, D. E., Nairn, R., Becker, E. L., and Ward, P. A. (1984) Purification and identification of formyl-methionyl-leucyl-phenylalanine as the major peptide neutrophil chemotactic factor produced by Escherichia coli. *The Journal of biological chemistry.* 259 (9). p.5430–5439.

Matthews, B. (1968) Solvent content of protein crystals. *Journal of Molecular Biology.* 33 (2). p.491–497.

Matthews, D. M., Burston, D., and Addison, J. M. (1972) Transport of glycylsarcosine by hamster small intestine in vitro. *Clinical science.* 42 (5). p.29.

McCoy, A. J. (2002) New applications of maximum likelihood and Bayesian statistics in macromolecular crystallography. *Current opinion in structural biology.* 12 (5). p.670–673.

McCoy, A. J., Storoni, L. C., and Read, R. J. (2004) Simple algorithm for a maximum-likelihood SAD function. *Acta crystallographica. Section D, Biological crystallography.* 60 (7). p.1220–1228.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of applied crystallography.* 40 (4). p.658–674.

McPherson, A. (1976) The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. *Methods of biochemical analysis.* 23. p.249–345.

McPherson, A. (1990) Current approaches to macromolecular crystallization. *European Journal of Biochemistry.* 189 (1). p.1–23.

Meisburger, S. P., Warkentin, M., Chen, H., Hopkins, J. B., Gillilan, R. E., Pollack, L., and Thorne, R. E. (2013) Breaking the radiation damage limit with Cryo-SAXS. *Biophysical journal.* 104 (1). p.227–36.

Meissner, B., Boll, M., Daniel, H., and Baumeister, R. (2004) Deletion of the intestinal peptide transporter affects insulin and TOR signaling in Caenorhabditis elegans. *The Journal of biological chemistry.* 279 (35). p.36739–36745.

Meredith, D. and Price, R. A. (2006) Molecular modeling of PepT1–towards a structure. *The Journal of membrane biology.* 213 (2). p.79–88.

Merlin, D., Steel, A., Gewirtz, A. T., Si-Tahar, M., Hediger, M. A., and Madara, J. L. (1998) hPepT1-mediated epithelial transport of bacteria-derived chemotactic peptides enhances neutrophil-epithelial interactions. *The Journal of clinical investigation.* 102 (11). p.2011–2018.

Merlin, D., Si-Tahar, M., Sitaraman, S. V., Eastburn, K., Williams, I., Liu, X., Hediger, M. A., and Madara, J. L. (2001) Colonic epithelial hPepT1 expression occurs in inflammatory bowel disease: transport of bacterial peptides influences expression of MHC class 1 molecules. *Gastroenterology.* 120 (7). p.1666–1679.

Miller, R., Gallo, S. M., Khalak, H. G., and Weeks, C. M. (1994) SnB: crystal structure determination via shake-and-bake. *Journal of Applied Crystallography.* 27 (4). p.613–621.

Miyazaki, K., Ohtani, K., Umeniwa, K., and Arita, T. (1982) Intestinal absorption of several beta-lactam antibiotics. III. Competitive inhibition behavior among zwitterionic beta-lactam antibiotics in the rat intestinal absorption. *Journal of pharmacobiodynamics.* 5 (8). p.555–563.

Monaco, H. L., Rizzi, M., and Coda, A. (1995) Structure of a complex of two plasma proteins: transthyretin and retinol-binding protein. *Science.* 268 (5213). p.1039–41.

Mueller, M., Wang, M., and Schulze-Briese, C. (2012) Optimal fine ϕ-slicing for single-photon-counting pixel detectors. *Acta crystallographica. Section D, Biological crystallography.* 68 (1). p.42–56.

Murray, J. W., Garman, E. F., and Ravelli, R. B. G. (2004) X-ray absorption by macromolecular crystals: the effects of wavelength and crystal composition on absorbed dose. *Journal of Applied Crystallography.* 37 (4). p.513–522.

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography.* 53 (3). p.240–255.

Naider, F. and Becker, J. M. (1975) Multiplicity of oligopeptide transport systems in Escherichia coli. *Journal of bacteriology.* 122 (3). p.1208–1215.

Nakanishi, T., Tamai, I., Sai, Y., Sasaki, T., and Tsuji, A. (1997) Carrier-mediated Transport of Oligopeptides in the Human Fibrosarcoma Cell Line HT1080. *Cancer Research.* 57 (18). p.4118–4122.

Nakashima, E., Tsuji, A., Mizuo, H., and Yamana, T. (1984) Kinetics and mechanism of in vitro uptake of amino-beta-lactam antibiotics by rat small intestine and relation to the intact-peptide transport system. *Biochemical pharmacology.* 33 (21). p.3345–3352.

Nam, H. J., Han, S. K., Bowie, J. U., and Kim, S. (2013) Rampant exchange of the structure and function of extramembrane domains between membrane and water soluble proteins. *PLoS computational biology.* 9 (3). p.e1002997.

Nässl, A.-M., Rubio-Aliaga, I., Sailer, M., and Daniel, H. (2011) The intestinal peptide transporter PEPT1 is involved in food intake regulation in mice fed a high-protein diet. *PloS one.* 6 (10). p.e26407.

Navaza, J. and Vernoslova, E. (1995) On the fast translation functions for molecular replacement. *Acta Crystallographica Section A Foundations of Crystallography.* 51 (4). p.445–449.

Nettleship, J. E., Assenberg, R., Diprose, J. M., Rahman-Huq, N., and Owens, R. J. (2010) Recent advances in the production of proteins in insect and mammalian cells for structural biology. *Journal of structural biology.* 172 (1). p.55–65.

Newey, H and Smyth, D. H. (1962) Cellular mechanisms in intestinal transfer of amino acids. *The Journal of physiology.* 164. p.527–51.

Newstead, S. (2014) Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. *Biochimica et biophysica acta.*

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., Baldwin, S. A., and Iwata, S. (2011) Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *The EMBO journal.* 30 (2). p.417–426.

Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols.* 2 (9). p.2212–2221.

Nikaido, H. (2011) Structure and mechanism of RND-type multidrug efflux pumps. *Advances in enzymology and related areas of molecular biology.* 77. p.1–60.

Ocheltree, S. M., Shen, H., Hu, Y., Keep, R. F., and Smith, D. E. (2005) Role and relevance of peptide transporter 2 (PEPT2) in the kidney and choroid plexus: in vivo studies with glycylsarcosine in wild-type and PEPT2 knockout mice. *The Journal of pharmacology and experimental therapeutics.* 315 (1). p.240–247.

Ogihara, H., Saito, H., Shin, B. C., Terado, T., Takenoshita, S., Nagamachi, Y., Inui, K., and Takata, K. (1996) Immuno-localization of H+/peptide cotransporter in rat digestive tract. *Biochemical and biophysical research communications.* 220 (3). p.848–852.

Okamura, M., Terada, T., Katsura, T., Saito, H., and Inui, K. I. (2003) Inhibitory Effect of Zinc on PEPT1-Mediated Transport of Glycylsarcosine and β-Lactam Antibiotics in Human Intestinal Cell Line Caco-2. *Pharmaceutical Research.* 20 (9). p.1389–1393.

Okano, T., Inui, K., Maegawa, H., Takano, M., and Hori, R. (1986) H+ coupled uphill transport of aminocephalosporins via the dipeptide transport system in rabbit intestinal brush-border membranes. *The Journal of biological chemistry.* 261 (30). p.14130–14134.

Olsson, M. H. M., Sø ndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *Journal of Chemical Theory and Computation.* 7 (2). p.525–537.

Ortega, A., Amorós, D., and García de la Torre, J (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophysical journal.* 101 (4). p.892–898.

Paithankar, K. S., Owen, R. L., and Garman, E. F. (2009) Absorbed dose calculations for macromolecular crystals: improvements to RADDOSE. *Journal of synchrotron radiation.* 16 (2). p.152–162.

Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., and Salemme, F. R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *Journal of biomolecular screening.* 6 (6). p.429–440.

Pao, S. S., Paulsen, I. T., and Saier, M. H. (1998) Major facilitator superfamily. *Microbiology and molecular biology reviews.* 62 (1). p.1–34.

Parker, J. L. and Newstead, S. (2014) Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature.* 507 (7490). p.68–72.

Paulsen, I. T. and Skurray, R. A. (1994) The POT family of transport proteins. *Trends in Biochemical Sciences.* 19 (10). p.404.

Payne, J. and Gilvarg, C. (1971) Peptide transport. *Advances in enzymology and related areas of molecular biology.* 35. p.187.

Pedersen, B. P., Kumar, H., Waight, A. B., Risenmay, A. J., Roe-Zurz, Z., Chau, B. H., Schlessinger, A., Bonomi, M., Harries, W., Sali, A., Johri, A. K., and Stroud, R. M. (2013) Crystal structure of a eukaryotic phosphate transporter. *Nature.* 496 (7446). p.533–536.

Poirier, A., Belli, S., Funk, C., Otteneder, M. B., Portmann, R., Heinig, K., Prinssen, E., Lazic, S. E., Rayner, C. R., Hoffmann, G., Singer, T., Smith, D. E., and Schuler, F. (2012) Role of the intestinal peptide transporter PEPT1 in oseltamivir absorption: in vitro and in vivo studies. *Drug metabolism and disposition.* 40 (8). p.1556–1565.

Powell, H. R. (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta crystallographica. Section D, Biological crystallography.* 55 (10). p.1690–1695.

Powell, H. R., Johnson, O., and Leslie, A. G. W. (2013) Autoindexing diffraction images with iMosflm. *Acta crystallographica. Section D, Biological crystallography.* 69 (7). p.1195–1203.

Quarcoo, D., Fischer, T. C., Heppt, W., Lauenstein, H. D., Pilzner, C., Welte, T., and Groneberg, D. A. (2009) Expression, localisation and functional implications of the transporter protein PEPT2 in the upper respiratory tract. *Respiration; international review of thoracic diseases.* 77 (4). p.440–446.

Rambo, R. P. and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers.* 95 (8). p.559–571.

Rambo, R. P. and Tainer, J. A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 496 (7446). p.477–481.

Read, R. J. (2001) Pushing the boundaries of molecular replacement with maximum likelihood. *Acta crystallographica. Section D, Biological crystallography.* 57 (10). p.1373–1382.

Read, R. J. and McCoy, A. J. (2011) Using SAD data in Phaser. *Acta crystallographica. Section D, Biological crystallography.* 67 (4). p.338–344.

Receveur-Brechot, V. and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science.* 13 (1). p.55–75.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I., and Saier, M. H. (2012) The major facilitator superfamily (MFS) revisited. *The FEBS journal.* 279 (11). p.2022–2035.

Ren, Q. and Paulsen, I. T. (2007) Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *Journal of molecular microbiology and biotechnology.* 12 (3). p.165–179.

Rossmann, M. G. and Blow, D. M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica.* 15 (1). p.24–31.

Rossmann, M. G. (2001) Molecular replacement - historical background. *Acta Crystallographica Section D Biological Crystallography.* 57 (10). p.1360–1366.

Rubino, A., Field, M., and Shwachman, H. (1971) Intestinal Transport of Amino Acid Residues of Dipeptides. 1. Influx of the glycine residue of glycyl-L-proline across mucosal borber. *J. Biol. Chem.* 246 (11). p.3542–3548.

Saito, H., Terada, T., Okuda, M., Sasaki, S., and Inui, K. I. (1996) Molecular cloning and tissue distribution of rat peptide transporter PEPT2. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 1280 (2). p.173–177.

Sakata, K., Yamashita, T., Maeda, M., Moriyama, Y., Shimada, S., and Tohyama, M. (2001) Cloning of a lymphatic peptide/histidine transporter. *Biochemical Journal.* 356 (1). p.53.

Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology.* 234 (3). p.779–815.

Sauter, N. K., Grosse-Kunstleve, R. W., and Adams, P. D. (2004) Robust indexing for automatic data collection. *Journal of applied crystallography.* 37 (3). p.399–409.

Scapin, G. (2013) Molecular replacement then and now. *Acta crystallographica. Section D, Biological crystallography.* 69 (11). p.2266–2275.

Schneider, T. R. and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta crystallographica. Section D, Biological crystallography.* 58 (10). p.1772–1779.

Schomaker, V. and Trueblood, K. N. (1968) On the rigid-body motion of molecules in crystals. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry.* 24 (1). p.63–76.

Schrodinger, L. (2010) *The PyMOL Molecular Graphics System.*

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal.* 78 (3). p.1606–1619.

Sheff, M. F. and Smyth, D. H. (1955) An apparatus for the study of in vivo intestinal absorption in the rat. *The Journal of physiology.* 128 (3). p.67.

Shen, H., Smith, D. E., Yang, T., Huang, Y. G., Schnermann, J. B., and Brosius, F. C. (1999) Localization of PEPT1 and PEPT2 proton-coupled oligopeptide transporter mRNA and protein in rat kidney. *The American journal of physiology.* 276 (5). p.658–665.

Shi, Y. (2013) Common folds and transport mechanisms of secondary active transporters. *Annual review of biophysics.* 42. p.51–72.

Shu, C. (2002) Role of PEPT2 in Peptide/Mimetic Trafficking at the Blood-Cerebrospinal Fluid Barrier: Studies in Rat Choroid Plexus Epithelial Cells in Primary Culture. *Journal of Pharmacology and Experimental Therapeutics.* 301 (3). p.820–829.

Sigrist-Nelson, K. (1975) Dipeptide transport in isolated intestinal brush border membrane. *Biochimica et biophysica acta.* 394 (2). p.220–226.

Silbernagl, S., Ganapathy, V., and Leibach, F. H. (1987) H+ gradient-driven dipeptide reabsorption in proximal tubule of rat kidney. Studies in vivo and in vitro. *The American journal of physiology.* 253 (3). p.448–457.

Sliz, P., Harrison, S. C., and Rosenbaum, G. (2003) How does Radiation Damage in Protein Crystals Depend on X-Ray Dose? *Structure.* 11 (1). p.13–19.

Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta crystallographica. Section D, Biological crystallography.* 68 (4). p.368–380.

Smith, D. E., Pavlova, A., Berger, U. V., Hediger, M. A., Yang, T., Huang, Y. G., and Schnermann, J. B. (1998) Tubular localization and tissue distribution of peptide transporters in rat kidney. *Pharmaceutical research.* 15 (8). p.1244–1249.

Smith, D. E., Clémençon, B., and Hediger, M. A. (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Molecular aspects of medicine.* 34 (2). p.323–336.

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Analytical biochemistry.* 150 (1). p.76–85.

Solcan, N. (2013) “Biochemical and biophysical studies of the prokaryotic proton dependent oligopeptide transporters”. phD. Oxford. p.111–173.

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 oligopeptide transporters. *The EMBO journal.* 31 (16). p.3411–3421.

Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *Journal of Chemical Theory and Computation.* 7 (7). p.2284–2295.

Steiner, H. Y., Naider, F, and Becker, J. M. (1995) The PTR family: a new group of peptide transporters. *Molecular microbiology.* 16 (5). p.825–834.

Stockley, R. A. (2014) Alpha1-antitrypsin review. *Clinics in chest medicine.* 35 (1). p.39–50.

Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification.* 41 (1). p.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 (7490). p.73–77.

Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y., and Yan, N. (2012) Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature.* 490 (7420). p.361–366.

Svergun, D. I. and Koch, M. H. J. (2003) Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics.* 66 (10). p.1735–1782.

Swaan, P. W., Bensman, T., Bahadduri, P. M., Hall, M. W., Sarkar, A., Bao, S., Khantwal, C. M., Ekins, S., and Knoell, D. L. (2008) Bacterial peptide recognition and immune activation facilitated by human peptide transporter PEPT2. *American journal of respiratory cell and molecular biology.* 39 (5). p.536–542.

Tarlow, M. J., Hadorn, B, Arthurton, M. W., and Lloyd, J. K. (1970) Intestinal enterokinase deficiency. A newly-recognized disorder of protein digestion. *Archives of disease in childhood.* 45 (243). p.651–655.

Terada, T., Saito, H., Sawada, K., Hashimoto, Y., and Inui, K. (2000a) N-terminal halves of rat H+/peptide transporters are responsible for their substrate recognition. *Pharmaceutical research.* 17 (1). p.15–20.

Terada, T., Sawada, K., Irie, M., Saito, H., Hashimoto, Y., and Inui, K. (2000b) Structural requirements for determining the substrate affinity of peptide transporters PEPT1 and PEPT2. *Pflügers Archive: European journal of physiology.* 440 (5). p.679–684.

Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta crystallographica. Section D, Biological crystallography.* 56 (8). p.965–972.

Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *Journal of synchrotron radiation.* 11 (Pt 1). p.49–52.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography.* 64 (1). p.61–69.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica. Section D, Biological crystallography.* 65 (6). p.582–601.

Teuscher, N. S., Keep, R. F., and Smith, D. E. (2001) PEPT2-mediated uptake of neuropeptides in rat choroid plexus. *Pharmaceutical research.* 18 (6). p.807–813.

Thamotharan, M., Bawani, S. Z., Zhou, X., and Adibi, S. A. (1999) Hormonal regulation of oligopeptide transporter Pept-1 in a human intestinal cell line. *American Journal of Physiology - Cell Physiology.* 276 (4). p.821–826.

Urban, E. and Weser, E. (1980) Intestinal adaptation to bowel resection. *Advances in internal medicine.* 26. p.265–91.

Usón, I. and Sheldrick, G. M. (1999) Advances in direct methods for protein crystallography. *Current Opinion in Structural Biology.* 9 (5). p.643–648.

Vagin, A. and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography.* 30 (6). p.1022–1025.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 59 (4). p.687–696.

Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure.* 14 (11). p.1617–1622.

Wang, B. C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods in enzymology.* 115. p.90–112.

Wang, H., Fei, Y. J., Ganapathy, V., and Leibach, F. H. (1998) Electrophysiological characteristics of the proton-coupled peptide transporter PEPT2 cloned from rat brain. *American Journal of Physiology - Cell Physiology.* 275 (4). p.967–975.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25 (9). p.1189–1191.

Webb, K. E., Matthews, J. C., and DiRienzo, D. B. (1992) Peptide absorption: a review of current concepts and future perspectives. *Journal of animal science.* 70 (10). p.3248–3257.

Wenzel, U., Kuntz, S., and Daniel, H. (2001) Flavonoids with Epidermal Growth Factor-Receptor Tyrosine Kinase Inhibitory Activity Stimulate PEPT1-Mediated Cefixime Uptake into Human Intestinal Epithelial Cells. *Journal of Pharmacology and Experimental Therapeutics.* 299 (1). p.351–357.

Wernimont, A. and Edwards, A. (2009) In situ proteolysis to generate crystals for structure determination: an update. *PloS one.* 4 (4). Ed. by H. Song. p.5094.

Whitman, W. (1973) *Walt Whitman's Camden Conversations.* Ed. by W Teller. New Jersey: Rutgers University Press.

Wiggans, D. S. and Johnston, J. M. (1959) The absorption of peptides. *Biochimica et Biophysica Acta.* 32. p.69–73.

Wilson, T. H. and Wiseman, G. (1954) The use of sacs of everted small intestine for the study of the transference of substances from the mucosal to the serosal surface. *The Journal of physiology.* 123 (1). p.116–125.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography.* 67 (4). p.235–242.

Winter, G. (2009) Xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography.* 43 (1). p.186–190.

Wojtczak, A, Cody, V, Luft, J. R., and Pangborn, W (1996) Structures of human transthyretin complexed with thyroxine at 2.0 A resolution and 3',5'-dinitro-N-acetyl-L-thyronine at 2.2 A resolution. *Acta crystallographica. Section D, Biological crystallography.* 52 (4). p.758–765.

Xu, L., Haworth, I. S., Kulkarni, A. A., Bolger, M. B., and Davies, D. L. (2009) Mutagenesis and cysteine scanning of transmembrane domain 10 of the human dipeptide transporter. *Pharmaceutical research.* 26 (10). p.2358–2366.

Xu, L., Li, Y., Haworth, I. S., and Davies, D. L. (2010) Functional role of the intracellular loop linking transmembrane domains 6 and 7 of the human dipeptide transporter hPEPT1. *The Journal of membrane biology.* 238 (1). p.43–49.

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochimica et biophysica acta.* 1804 (4). p.996–1010.

Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology Journal.* 7 (5). p.620–634.

Zhang, K. Y. J. and Main, P. (1990) Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Crystallographica Section A Foundations of Crystallography.* 46 (1). p.41–46.

Zhang, Z., Sauter, N. K., van den Bedem, H., Snell, G., and Deacon, A. M. (2006) Automated diffraction image analysis and spot searching for high-throughput crystal screening. *Journal of Applied Crystallography.* 39 (1). p.112–119.

Ziegler, T. R., Fernández-Estívariz, C., Gu, L. H., Bazargan, N., Umeakunne, K., Wallace, T. M., Diaz, E. E., Rosado, K. E., Pascal, R. R., Galloway, J. R., Wilcox, J. N., and Leader, L. M. (2002) Distribution of the H+/peptide transporter PepT1 in human intestine: up-regulated expression in the colonic mucosa of patients with short-bowel syndrome. *The American journal of clinical nutrition.* 75 (5). p.922–930.