

# **Improving Oral Drug Delivery: Computational Studies of Proton Dependent Oligopeptide Transporters**



**Mohd Firdaus Samsudin**

Department of Biochemistry

University of Oxford

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

Wolfson College

Trinity 2015

Copyright ©2015 Mohd Firdaus Samsudin

The *viva voce* examination was held on Tuesday 10 November 2015 by Dr Philip Biggin (University of Oxford) and Dr Alexander Cameron (University of Warwick). A copy of this thesis has been deposited in the Bodleian Library, Oxford.

*Untuk Mak dan Ayah di kampung...*

## Abstract

Proton dependent oligopeptide transporters (POTs) play a central role in nitrogen homeostasis by coupling the uptake of dipeptides and tripeptides to the proton electrochemical gradient across the plasma membrane. In human, members of this transporter family, PepT1 and PepT2, are critical modulators of drug pharmacokinetics as they facilitate the uptake and retention of numerous orally administered drugs including the  $\beta$ -lactam antibiotics. Rationally designing drugs to target these transporters is therefore an attractive approach to improving bioavailability. To this end, the binding of peptides to a bacterial homolog, PepT<sub>St</sub>, was modelled based on recently determined crystal structures. A range of computational methods to predict the free energy of binding were evaluated and a hybrid approach, where the end-point methods were used to classify peptides into strong and poor binders and a theoretically exact method for refinement, was able to accurately predict ligand affinities. This approach was utilised to investigate the substrate preference of PepT<sub>St</sub> and the results were validated using *in vitro* transport assays. To extend this study to the human peptide transporters, homology models of PepT1 and PepT2 were built using the crystal structures of PepT<sub>So</sub> and mouse and rat extracellular domains (ECDs) as the templates. Essential residues as proposed by various mutational studies align well with the binding cavity, suggesting that the models are structurally sound. Applying the free energy methods to predict the affinities of peptides and drugs to the homology model of PepT1, however, resulted in discrepancies with experimental data, highlighting the importance of a high-resolution crystal structure in binding affinity predictions. Based on the results for PepT<sub>St</sub>, a binding model for peptide prodrugs and  $\beta$ -lactam antibiotics to the human PepT1 was proposed. Overall, this thesis provides a framework for future computational studies using free energy methods to understand drug interactions with pharmaceutically relevant transporters.



## Acknowledgements

I wish to express my formal gratitude to my supervisors: Phil Fowler for his constant support, invaluable constructive feedback, and most importantly for his ceaseless enthusiasm and optimism that drove this project over the four challenging, yet rewarding, years; and to Mark Sansom for his critical advises and interesting ideas that helped shape this research to become what it is.

I am also deeply indebted to all the wonderful members, past and present, of SBCB, especially to my subgroups: PIMP, SysP and SysS, for without their useful comments and suggestions my research would have been stagnant. I also wish to thank certain members whose contributions to my work were very specific: Heidi Koldsø for helpful discussion on membrane transporters, Matteo Aldeghi on free energy methods, and Caroline Dahl for developing interquant.

Experimental collaborations were instrumental in this study; I am therefore immensely grateful to Simon Newstead for his support and confidence in my computational work, Johnny Beale for the structures of the extracellular domains of PepT1 and PepT2, Joanne Parker for validating my computational predictions using transport assays, and Nicolae Solcan and Daniel Grba for the transport assay data of peptides and drugs.

I thank the Terengganu Foundation for their generous funding.

Finally, but really foremost, I want to thank my family in Malaysia for believing in me forever and always; and to my dearest friends Basir, Izzu, Mas, Firdaus and Radzy, whose encouragement stood me in good stead from first to last.

MFS

# List of publications

This thesis is the product of my own work, unless otherwise stated. It is based in part on work described in the following publications.

- Beale JH\*, Parker JL\*, Samsudin MF\*, Barret AL, Senan A, Bird LE, Scott D, Owens RJ, Sansom, MSP, Tucker SJ, Meredith D, Fowler PW and Newstead S. (2015). **Crystal structures of the extracellular domain from PepT1 and PepT2 provide novel insights into mammalian peptide transport.** Structure 23, 1889–1899.  
DOI: <http://dx.doi.org/10.1016/j.str.2015.07.016>
- Samsudin MF, Parker JL, Sansom MSP, Newstead S and Fowler PW. (2015). **Accurate prediction of ligand affinities for a peptide transporter.** Cell Chem. Biol., In press.
- Samsudin MF, Sansom MSP, Newstead S and Fowler PW. **Substrate promiscuity within the proton-dependent oligopeptide transporter family.** Biochemistry, In preparation.

\*joint first author

# Contents

|                                                                                     |             |
|-------------------------------------------------------------------------------------|-------------|
| <b>List of publications</b>                                                         | <b>v</b>    |
| <b>Contents</b>                                                                     | <b>vi</b>   |
| <b>List of Figures</b>                                                              | <b>x</b>    |
| <b>List of Tables</b>                                                               | <b>xiii</b> |
| <b>Nomenclature</b>                                                                 | <b>xiv</b>  |
| <b>1 Introduction</b>                                                               | <b>1</b>    |
| 1.1 Transport across the plasma membrane . . . . .                                  | 1           |
| 1.1.1 Channels and transporters . . . . .                                           | 3           |
| 1.1.2 Primary and secondary active transporters . . . . .                           | 4           |
| 1.1.3 Structural and mechanistic studies of secondary active transporters . . . . . | 6           |
| 1.2 Proton-dependent oligopeptide transporters . . . . .                            | 11          |
| 1.2.1 The SLC15 gene family . . . . .                                               | 11          |
| 1.2.2 Substrate specificity of PepT1 and PepT2 . . . . .                            | 13          |
| 1.2.3 Conserved motifs from bacteria to human . . . . .                             | 17          |
| 1.2.4 Crystal structures of bacterial homologs . . . . .                            | 19          |
| 1.3 Oral drug bioavailability . . . . .                                             | 26          |
| 1.3.1 PepT1 and PepT2 mediated drug uptake . . . . .                                | 29          |
| 1.4 Aim of this thesis . . . . .                                                    | 31          |
| <b>2 Theory and methods</b>                                                         | <b>32</b>   |
| 2.1 Molecular dynamics simulation . . . . .                                         | 32          |
| 2.1.1 Numerical integration for the equations of motion . . . . .                   | 33          |
| 2.1.2 Molecular mechanics force field . . . . .                                     | 34          |

|          |                                                                                       |           |
|----------|---------------------------------------------------------------------------------------|-----------|
| 2.1.3    | Constrained dynamics . . . . .                                                        | 38        |
| 2.1.4    | Periodic boundary conditions . . . . .                                                | 39        |
| 2.1.5    | The particle-mesh Ewald Method . . . . .                                              | 40        |
| 2.1.6    | Temperature and pressure control . . . . .                                            | 41        |
| 2.2      | Free energy of binding . . . . .                                                      | 43        |
| 2.3      | Computational predictions of binding energy . . . . .                                 | 44        |
| 2.3.1    | Scoring Functions . . . . .                                                           | 44        |
| 2.3.2    | End-point free energy methods . . . . .                                               | 46        |
| 2.3.3    | ‘Exact’ free energy methods . . . . .                                                 | 51        |
| 2.4      | Experimental transport assays . . . . .                                               | 56        |
| 2.5      | Homology modelling . . . . .                                                          | 58        |
| 2.6      | Protein-protein docking . . . . .                                                     | 59        |
| <b>3</b> | <b>Binding affinity predictions using free energy methods</b>                         | <b>61</b> |
| 3.1      | Binding affinity and transport . . . . .                                              | 61        |
| 3.2      | Methods . . . . .                                                                     | 64        |
| 3.2.1    | Modelling peptide binding . . . . .                                                   | 64        |
| 3.2.2    | Binding energy prediction . . . . .                                                   | 65        |
| 3.2.3    | Simulation details . . . . .                                                          | 67        |
| 3.2.4    | Statistical Analysis . . . . .                                                        | 68        |
| 3.3      | Results and Discussion . . . . .                                                      | 71        |
| 3.3.1    | Docking scores are inaccurate. . . . .                                                | 71        |
| 3.3.2    | End-point methods accurately predict transport. . . . .                               | 73        |
| 3.3.3    | Thermodynamic integration improves prediction for hydrophobic<br>dipeptides . . . . . | 77        |
| 3.4      | Conclusion . . . . .                                                                  | 81        |
| <b>4</b> | <b>Accurate prediction of peptide binding in a bacterial POT transporter</b>          | <b>83</b> |
| 4.1      | Substrate multispecificity in the POT transporters . . . . .                          | 83        |
| 4.2      | Methods . . . . .                                                                     | 85        |
| 4.2.1    | Thermodynamic integration with a change in net charge . . . . .                       | 85        |
| 4.2.2    | Contact Analysis . . . . .                                                            | 87        |
| 4.2.3    | Competition transport assay . . . . .                                                 | 87        |
| 4.3      | Results and Discussion . . . . .                                                      | 88        |
| 4.3.1    | Neutral dipeptides favoured over charged. . . . .                                     | 88        |
| 4.3.2    | The order of residues does matter. . . . .                                            | 91        |

|          |                                                                             |            |
|----------|-----------------------------------------------------------------------------|------------|
| 4.3.3    | Dipeptide binding sensitive to the N-terminal side chain . . . . .          | 92         |
| 4.3.4    | Lysine containing dipeptides may bind like a tripeptide. . . . .            | 98         |
| 4.3.5    | Tripeptides with bulky residues bind less well. . . . .                     | 103        |
| 4.4      | Conclusion . . . . .                                                        | 105        |
| <b>5</b> | <b>Homology models of human peptide transporters</b>                        | <b>108</b> |
| 5.1      | Previous models of PepT1 and PepT2 . . . . .                                | 108        |
| 5.2      | Methods . . . . .                                                           | 112        |
| 5.2.1    | Templates selection . . . . .                                               | 112        |
| 5.2.2    | Sequence alignment . . . . .                                                | 113        |
| 5.2.3    | Structural evaluation . . . . .                                             | 115        |
| 5.2.4    | Molecular dynamic simulation . . . . .                                      | 115        |
| 5.2.5    | Protein-protein docking . . . . .                                           | 116        |
| 5.3      | Results and Discussion . . . . .                                            | 116        |
| 5.3.1    | Essential residues mapped on homology models . . . . .                      | 124        |
| 5.3.2    | Comparison of PepT1 and PepT2 . . . . .                                     | 136        |
| 5.3.3    | Interactions of ECD with TM region and membrane . . . . .                   | 141        |
| 5.3.4    | PepT1 and PepT2 ECD bind to trypsin . . . . .                               | 145        |
| 5.4      | Conclusion . . . . .                                                        | 147        |
| <b>6</b> | <b>Peptide and drug binding in human PepT1</b>                              | <b>150</b> |
| 6.1      | Current state of drug design for PepT1 . . . . .                            | 150        |
| 6.2      | Methods . . . . .                                                           | 158        |
| 6.2.1    | Modelling peptide and drug binding . . . . .                                | 158        |
| 6.2.2    | MD Simulation . . . . .                                                     | 158        |
| 6.2.3    | Free energy calculation . . . . .                                           | 159        |
| 6.3      | Results and Discussion . . . . .                                            | 160        |
| 6.3.1    | All methods fail to predict transport. . . . .                              | 160        |
| 6.3.2    | General Amber force field for drug parameters. . . . .                      | 163        |
| 6.3.3    | Experimental data from whole cell assays. . . . .                           | 165        |
| 6.3.4    | Crystal structures essential for prediction of binding free energy. . . . . | 166        |
| 6.4      | Insights into drug design from PepT <sub>St</sub> . . . . .                 | 173        |
| 6.4.1    | Prodrug binding via a pseudo-N-terminus . . . . .                           | 176        |
| 6.4.2    | A possible role of His57 in drug transport . . . . .                        | 184        |
| 6.5      | Conclusion . . . . .                                                        | 189        |

---

|                   |                                                         |            |
|-------------------|---------------------------------------------------------|------------|
| <b>7</b>          | <b>Summary and future perspectives</b>                  | <b>191</b> |
| 7.1               | Summary and relevance of research . . . . .             | 191        |
| 7.2               | Future directions . . . . .                             | 193        |
|                   | <b>References</b>                                       | <b>197</b> |
| <b>Appendix A</b> | <b>Repeat simulations for thermodynamic integration</b> | <b>220</b> |
| <b>Appendix B</b> | <b>Sequence alignment for homology modelling</b>        | <b>223</b> |

# List of Figures

|      |                                                                                   |    |
|------|-----------------------------------------------------------------------------------|----|
| 1.1  | A schematic representation of a membrane bilayer . . . . .                        | 2  |
| 1.2  | Examples of membrane channel and transporter . . . . .                            | 5  |
| 1.3  | Primary and secondary active transporters . . . . .                               | 5  |
| 1.4  | Alternating access mechanism . . . . .                                            | 7  |
| 1.5  | Two common structural fold of secondary active transporters . . . . .             | 8  |
| 1.6  | Crystal structures of secondary active transporters in different states . . . . . | 10 |
| 1.7  | Membrane topology and expression of mammalian POTs . . . . .                      | 12 |
| 1.8  | The relationship of PepT1 with other membrane proteins . . . . .                  | 14 |
| 1.9  | Conserved motifs in POT family . . . . .                                          | 18 |
| 1.10 | The crystal structure of PepT <sub>So</sub> . . . . .                             | 19 |
| 1.11 | Substrate binding site of bacterial POTs . . . . .                                | 21 |
| 1.12 | Salt bridges gating the substrate binding site . . . . .                          | 23 |
| 1.13 | A general peptide uptake model by a POT transporter . . . . .                     | 25 |
| 1.14 | Oral drug absorption and metabolism . . . . .                                     | 28 |
| 1.15 | Structural comparison of cefadroxil and triAla . . . . .                          | 30 |
| 2.1  | The components of a molecular mechanics force field . . . . .                     | 36 |
| 2.2  | Two dimensional representation of periodic boundary conditions. . . . .           | 39 |
| 2.3  | A schematic representation of the Ewald summation. . . . .                        | 40 |
| 2.4  | The concept of free energy of binding, $\Delta G$ . . . . .                       | 43 |
| 2.5  | Free energy methods to predict $\Delta G$ . . . . .                               | 44 |
| 2.6  | Thermodynamic cycle for the LIE calculation. . . . .                              | 47 |
| 2.7  | Thermodynamic cycle for the MMPB/GBSA calculation. . . . .                        | 49 |
| 2.8  | Thermodynamic cycle for alchemical free energy method. . . . .                    | 52 |
| 2.9  | The application of soft core potential to avoid singularity. . . . .              | 54 |
| 2.10 | Transport assays used to study POTs . . . . .                                     | 57 |

|      |                                                                                                          |     |
|------|----------------------------------------------------------------------------------------------------------|-----|
| 3.1  | Structures of peptides for binding affinity prediction test with PepT <sub>St</sub> . . . .              | 63  |
| 3.2  | Modelling peptide binding on bacterial peptide transporter, PepT <sub>St</sub> . . . . .                 | 65  |
| 3.3  | Thermodynamic cycle for relative binding free energy calculation on PepT <sub>St</sub> . . .             | 66  |
| 3.4  | Dual topology approach for AlaPhe to AlaAla transformation. . . . .                                      | 67  |
| 3.5  | Schematic illustration of block averaging technique. . . . .                                             | 68  |
| 3.6  | An example of calculating auto-correlation time and equilibration time. . .                              | 70  |
| 3.7  | An example of docking protocol using AutoDock Vina. . . . .                                              | 72  |
| 3.8  | Affinity prediction for PepT <sub>St</sub> by AutoDock . . . . .                                         | 73  |
| 3.9  | Affinity prediction for PepT <sub>St</sub> by end-point methods. . . . .                                 | 75  |
| 3.10 | Variation in the prediction of $\Delta G$ s by end-point methods. . . . .                                | 76  |
| 3.11 | Example of $\frac{\delta U}{\delta \lambda}$ distribution for alchemical transformation MD simulation. . | 78  |
| 3.12 | Affinity prediction refined by thermodynamic integration. . . . .                                        | 80  |
| 3.13 | Performance of affinity prediction methods and their computational time. .                               | 81  |
| 4.1  | Binding orientations of dipeptide and tripeptide to PepT <sub>St</sub> . . . . .                         | 84  |
| 4.2  | TI calculation with a change in total charge. . . . .                                                    | 86  |
| 4.3  | $\Delta G$ of all 400 dipeptides to PepT <sub>St</sub> calculated by the LIE method . . . . .            | 89  |
| 4.4  | Predicted $\Delta G$ of all dipeptides grouped according to chemical properties . .                      | 90  |
| 4.5  | $\Delta G$ for all dipeptides plotted against their sequence-reversed equivalents . .                    | 91  |
| 4.6  | $\Delta \Delta G$ for alchemically transforming AlaAla side chains. . . . .                              | 93  |
| 4.7  | Contact analysis for N- and C-terminal side chain of a dipeptide substrate .                             | 96  |
| 4.8  | Contact analysis for N- and C-terminal side chain based on chemical group                                | 97  |
| 4.9  | IC <sub>50</sub> competition curves for DA, AD and KA, AK . . . . .                                      | 98  |
| 4.10 | Modelling the binding of AlaLys and LysAla based on triAla . . . . .                                     | 100 |
| 4.11 | $\Delta \Delta G$ for alchemically transforming vertical AlaAla into LysAla and AlaLys. .                | 101 |
| 4.12 | $\Delta \Delta G$ for alchemically transforming triAla side chains. . . . .                              | 104 |
| 4.13 | Homology models of PepT <sub>St</sub> bound to FAA and AAF . . . . .                                     | 105 |
| 4.14 | The presence of AAF breaks extracellular gate of PepT <sub>St</sub> . . . . .                            | 106 |
| 5.1  | Previous models built for mammalian peptide transporters. . . . .                                        | 110 |
| 5.2  | Gating residues in PepT <sub>So</sub> , PepT1 and PepT2 . . . . .                                        | 114 |
| 5.3  | Homology models of human PepT1 and PepT2 . . . . .                                                       | 118 |
| 5.4  | Ramachandran plots for PepT1 and PepT2 homology models . . . . .                                         | 119 |
| 5.5  | Secondary structure analysis of the homology models (TM) . . . . .                                       | 121 |
| 5.6  | Secondary structure analysis of the homology models (ECD) . . . . .                                      | 122 |
| 5.7  | RMSD of PepT1 and PepT2 homology models . . . . .                                                        | 123 |



|      |                                                                                               |     |
|------|-----------------------------------------------------------------------------------------------|-----|
| 5.8  | RMSF of PepT1 and PepT2 homology models . . . . .                                             | 124 |
| 5.9  | Density profile of the homology models with respect to membrane . . . . .                     | 125 |
| 5.10 | Key residues for peptide binding in PepT1 . . . . .                                           | 125 |
| 5.11 | Tilted helices in PepT1 homology model . . . . .                                              | 128 |
| 5.12 | Hydrophobic packing between TM helices H7, H9, H10 and H12 . . . . .                          | 129 |
| 5.13 | A water pocket close to His57 in PepT1 homology model . . . . .                               | 132 |
| 5.14 | Proposed Arg282-Asp341 salt bridge in PepT1 homology model . . . . .                          | 133 |
| 5.15 | Intracellular helices in PepT1 homology model . . . . .                                       | 134 |
| 5.16 | Single nucleotide polymorphisms in PepT1 and PepT2 . . . . .                                  | 135 |
| 5.17 | Mapping the results of chimera studies on PepT1 homology models . . . . .                     | 137 |
| 5.18 | Sequence conservation of PepT1 and PepT2 mapped on homology models . . . . .                  | 139 |
| 5.19 | Comparison of PepT1 and PepT2 binding sites . . . . .                                         | 140 |
| 5.20 | Salt bridge interactions between ECD and TM . . . . .                                         | 142 |
| 5.21 | Interaction of ECD with membrane bilayer . . . . .                                            | 144 |
| 5.22 | Possible trypsin binding site based on sequence conservation . . . . .                        | 146 |
| 5.23 | A model of PepT1-trypsin complex . . . . .                                                    | 148 |
| 6.1  | Peptide test set for human PepT1 . . . . .                                                    | 152 |
| 6.2  | Modelling drug binding to PepT1 . . . . .                                                     | 159 |
| 6.3  | Predicted $\Delta G$ of peptide and drug binding to PepT1 . . . . .                           | 162 |
| 6.4  | Analysis of GAFF for simulations of drug compounds . . . . .                                  | 164 |
| 6.5  | Comparison of two PepT1 uptake studies using whole cell assays . . . . .                      | 165 |
| 6.6  | Performance of binding energy predictions for PepT <sub>St</sub> homology models . . . . .    | 169 |
| 6.7  | PepT <sub>St</sub> homology models after a 1 ns simulation. . . . .                           | 170 |
| 6.8  | RMSD of ligands during simulations with PepT <sub>St</sub> homology models . . . . .          | 170 |
| 6.9  | Contacts made by ligands during simulations with PepT <sub>St</sub> homology models . . . . . | 171 |
| 6.10 | Comparison of PepT1 and PepT <sub>St</sub> binding site and peptide transport . . . . .       | 174 |
| 6.11 | A model of peptide prodrug binding to human PepT1 . . . . .                                   | 182 |
| 6.12 | Structural comparison of azidothymidine-dipeptide conjugate and triAla . . . . .              | 184 |
| 6.13 | Examples of cephalosporins $\beta$ -lactam antibiotics . . . . .                              | 185 |
| 6.14 | Transport data of PepT <sub>St</sub> for drug compounds. . . . .                              | 186 |
| 6.15 | A possible role of His57 in drug binding . . . . .                                            | 188 |
| A.1  | $\Delta\Delta G$ for replicate alchemical MD simulations . . . . .                            | 222 |

# List of Tables

|     |                                                                                                                                  |     |
|-----|----------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.1 | Different affinities of peptides to PepT1 and PepT2 . . . . .                                                                    | 16  |
| 1.2 | Available crystal structures of bacterial POTs . . . . .                                                                         | 19  |
| 3.1 | Peptide transport by PepT <sub>St</sub> . . . . .                                                                                | 64  |
| 3.2 | The performance of binding affinity prediction methods for PepT <sub>St</sub> . . . . .                                          | 71  |
| 3.3 | Predicted $\Delta G$ for peptide binding to PepT <sub>St</sub> based on AutoDock . . . . .                                       | 71  |
| 3.4 | Predicted $\Delta G$ for peptide binding to PepT <sub>St</sub> based on end-point methods . . . . .                              | 74  |
| 3.5 | TI results for the alchemical transformation of AlaPhe to AlaAla, PhePhe or AlaTyr. . . . .                                      | 79  |
| 4.1 | Correlation of $\Delta G$ for dipeptides with their sequence-reverse equivalents based on chemical groups. . . . .               | 92  |
| 4.2 | Lennard-Jones (vdw) and Coulomb (el) potentials in the unbound and bound states of peptide ligands. . . . .                      | 92  |
| 4.3 | TI results for the alchemical transformation of AlaAla side chains. . . . .                                                      | 94  |
| 4.4 | IC <sub>50</sub> values based on proton driven competition transport assays. . . . .                                             | 99  |
| 4.5 | TI results for the alchemical transformation of AlaAla in a vertical binding pose . . . . .                                      | 101 |
| 4.6 | TI results for the alchemical transformation of triAla side chains. . . . .                                                      | 104 |
| 5.1 | Bacterial peptide transporters with available crystal structures and their sequence identity with human PepT1 and PepT2. . . . . | 113 |
| 5.2 | Proposed gating residues for PepT1 and PepT2 . . . . .                                                                           | 113 |
| 5.3 | MolProbity analysis for PepT1 and PepT2 homology models . . . . .                                                                | 120 |
| 5.4 | Key residues in human PepT1 from mutational studies . . . . .                                                                    | 126 |
| 5.5 | Percentage amino acid sequence identity for mammalian PepT1 and PepT2 . . . . .                                                  | 138 |
| 5.6 | Salt bridge interaction between ECD and TM domain . . . . .                                                                      | 142 |
| 5.7 | Interaction of ECD with membrane bilayer . . . . .                                                                               | 144 |

---

|     |                                                                             |     |
|-----|-----------------------------------------------------------------------------|-----|
| 6.1 | Peptide transport by human PepT1 . . . . .                                  | 153 |
| 6.2 | Drug transport by human PepT1 . . . . .                                     | 154 |
| 6.3 | The performance of binding affinity prediction methods for human PepT1 .    | 160 |
| 6.4 | Predicted $\Delta G$ for peptide and drug binding to human PepT1 . . . . .  | 161 |
| 6.5 | Templates for PepT <sub>St</sub> homology modelling . . . . .               | 167 |
| 6.6 | Prodrugs designed to target human PepT1 . . . . .                           | 178 |
| A.1 | TI results for replicate alchemical transformation MD simulations . . . . . | 221 |

# Chapter 1

## Introduction

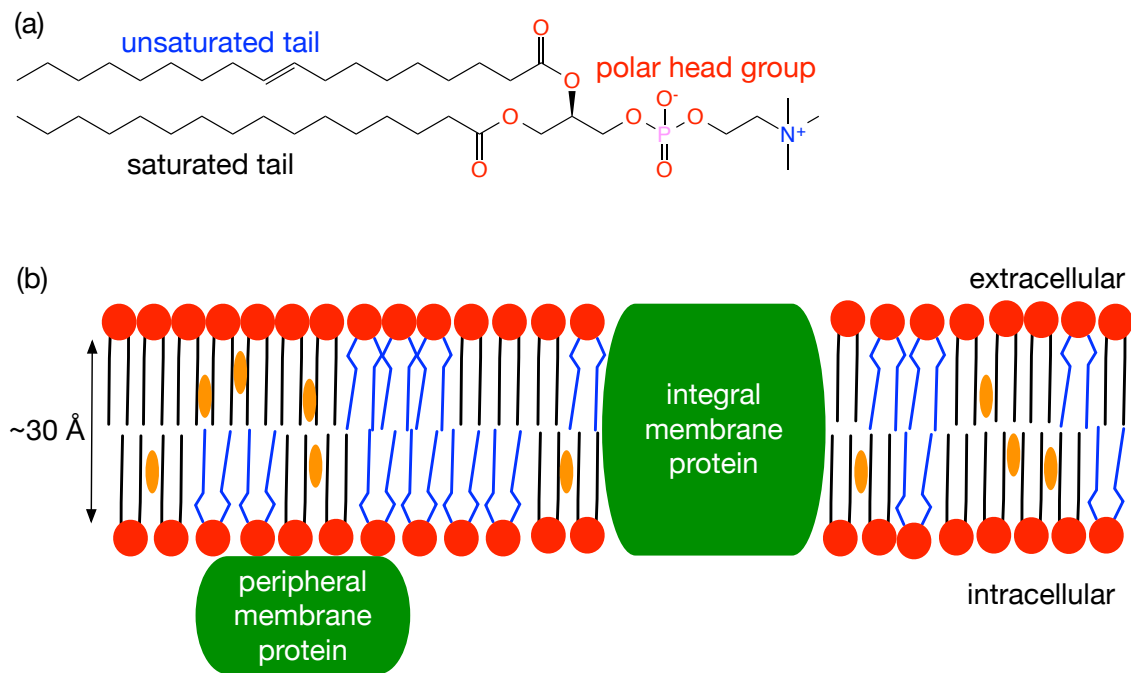
Fundamental to the functioning of the living cell is its ability to control the traffic across its plasma membrane. This thesis uses a *computational microscope* to examine the translocation of peptides and drugs into the cell facilitated by a *membrane transporter*. Studying the behaviour of this protein allows for its future exploitation as a molecular delivery system.

### 1.1 Transport across the plasma membrane

The emergence of biological membranes marks a defining step during the creation of the first living cells in the pre-biotic world (Schrum et al., 2010; Koch and Silver, 2005). Cellular evolution has led to the incorporation of multiple species of lipids, proteins and carbohydrates, essentially increasing the complexity of the membrane, but most importantly making it more versatile—the plasma membrane is not only a structural barrier but also a medium of communication between the intracellular contents and the external world (Luckey, 2008). The major components of the plasma membrane are the amphipathic phospholipids that self-assemble to form a bilayer (Figure 1.1) (Bishop and Bell, 1988). Whilst this is able to concentrate vital compounds within the cell for efficient chemical reactions, it creates a further problem: the influx of nutrients and efflux of waste products are almost completely abolished. The hydrophobic acyl tails of the lipids ensure a tight and impermeable barrier surrounding the cytoplasm. Certain small and/or non-polar species like oxygen, carbon dioxide and water\* can passively diffuse across the plasma membrane, but most polar solutes like sugar and amino acids, as well as ions require a specialised conduit through which they can pass. This is where membrane proteins come into the story. The presence of nu-

---

\*In some tissues, e.g. in the kidney, water is required much faster than passive diffusion can supply, hence the cells express specialised water channels, *aquaporins* (King et al., 2004), to meet this demand.



Figs. 1.1: A schematic of the plasma membrane. (a) The chemical structure of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), an example of a phospholipid. Another class of lipid found in the plasma membrane is sphingolipid, whereby the glycerol core structure is replaced by sphingosine. Other variants of the head group include phosphatidylserine (PS), phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). (b) A cross-sectional representation of a basic membrane bilayer. Membrane proteins (green) are broadly classified into integral and peripheral membrane proteins; the former typically spans the entire width of the bilayer, whilst the latter is attached to one side of the bilayer through electrostatic interactions (e.g. via a pleckstrin homology (PH) domain) or covalent bonds (e.g. glycosylphosphatidylinositol (GPI)-anchored protein). Orange ovals represent cholesterol, which reside in between lipid tails and increase rigidity. Also, the composition of lipids in the plasma membrane varies between different organisms and cell types. For example, a typical membrane of a mammalian cell consists mostly of PC and sphingomyelin (Phillips and Dodge, 1967) with some regions, like the apical side, enriched with PS and cholesterol (Gerl et al., 2012). A typical bacterial membrane, on the other hand, contains only PE, PG and cardiolipin (Goldfine, 1984).

merous integral membrane proteins, broadly classified into *channels* and *transporters* based on their mechanisms, gives the cell absolute control over the entry and exit of compounds across the plasma membrane.

### 1.1.1 Channels and transporters

**Channels** are membrane proteins that form an aqueous pore penetrating the plasma membrane to allow the movement of ions or water molecules (Hille, 2001). Permeation through channels is passive, i.e. it relies only on the electrochemical or osmotic gradient and is not coupled to the consumption of energy. The rate of diffusion across a channel is typically of the order of  $10^6 \text{ s}^{-1}$  and  $10^9 \text{ s}^{-1}$  for ions and water, respectively, comparable to the rate of diffusion in bulk solution. Despite transporting passively, channels are tightly regulated. Firstly, ion channels are usually selective; potassium channels, for example, are capable of distinguishing between  $\text{K}^+$  and  $\text{Na}^+$  such that the ratio of  $\text{K}^+:\text{Na}^+$  traversing the channel is above 1000:1 (Hille, 2001). Furthermore, most aquaporins are only permeable to water molecules and do not allow the diffusion of ions at all (King et al., 2004). This is achieved by a *selectivity filter*, a narrow segment at the beginning of the pore, along which the hydration shell of ions is removed so diffusion can proceed in a single file manner.

Another way by which channels are controlled is via gating; external stimuli can induce opening to allow the flow of ions/water molecules and closing to stop the flux. These stimuli can include ligands, transmembrane pressure, potential, pH and temperature. Two structural components exist to mediate the transduction of the signal from the external stimuli to the opening or closing of the channel: i) a *sensor* that detects the stimuli, for example a ligand binding site and ii) a *gate*, which physically obstructs the flow of permeant species during the closed state. An accurate relay of information from the sensor to the gate is crucial for the precise control of ionic flux. An example of a potassium channel is KcsA from bacteria *Streptomyces lividans*, a homolog of human voltage gated  $\text{K}_v$  channels (Figure 1.2(a)).

**Transporters** are membrane proteins that couple a source of free energy to the movement of solutes, often against their concentration gradient, effectively accumulating compounds, such as nutrients, into the cytoplasm or actively removing undesired wastes into the external milieu. The energy utilised by a membrane transporter comes from ATP (*primary active transporter*) or the electrochemical gradient (*secondary active transporter*). Transporters undergo a cycle of large conformational changes involving substrate binding and unbinding on opposite sites of the membrane, and therefore the rate of uptake is much slower than a flux across a channel, typically between 1 to  $1000 \text{ s}^{-1}$ . Transporters have two

gates, one on the extracellular side and another on the cytoplasmic site, only one of which can open at a time to obviate a continuous flow of substrates that would effectively turn them into channels. Structurally, a transporter protein has a distinct substrate binding site, hence only a discrete number of substrate molecule can be translocated in every cycle, unlike a channel, which allows thousands to millions of ions/water to pass through simultaneously once it is opened. An example of a transporter is the potassium transporter KtrAB found in bacteria, archaea, fungi and plants (Figure 1.2(b)).

As more biophysical and crystallographic data are collected, however, the distinction between transporters and channels is becoming less clear (Ashcroft et al., 2009). This is exemplified by the cystic fibrosis transmembrane conductance regulator (CFTR), which is a chloride channel despite its structural and sequence homology to ATP-binding cassette (ABC) transporters (Gadsby et al., 2006). In addition, the CLC family consists of both chloride channels and proton-chloride exchangers that share similar overall topology and key residues for ion binding (Miller, 2006; Accardi and Picollo, 2010). Furthermore, channel-like ionic leaks have been observed in several neurotransmitter transporters, for example the glutamate transporter and its bacterial homolog (Veruki et al., 2006; Ryan and Mindell, 2007). This evidence suggests that a channel is perhaps just a “broken” transporters that has lost one of its gates through evolution.

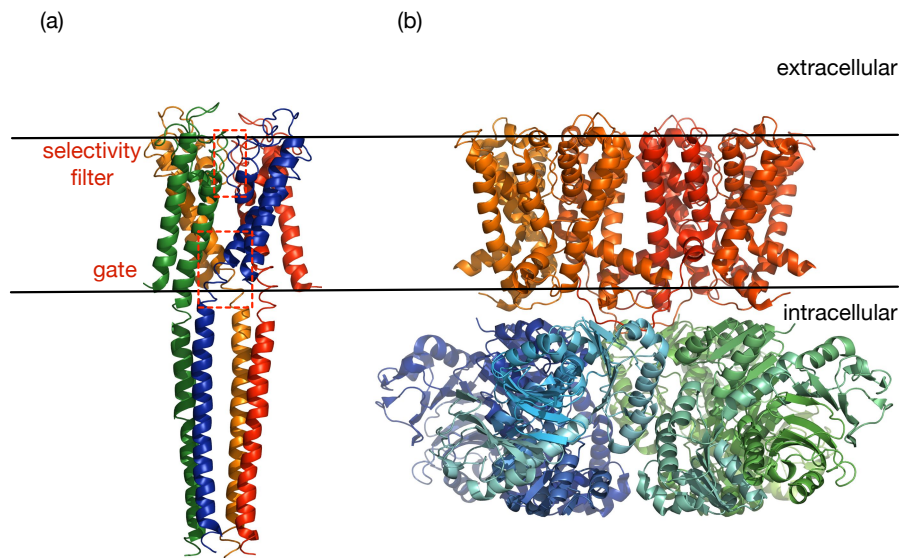
### 1.1.2 Primary and secondary active transporters

**Primary active transporters** utilise a direct source of energy from ATP hydrolysis and are often referred to as *pumps* (Figure 1.3(a)). There are two classes of ATP-powered protein pumps: ATPases and ABC transporters (Lodish et al., 2000); the former responsible to transport ions, while the latter facilitate the uptake of small molecules such as amino acids, sugars, peptides and vitamins.

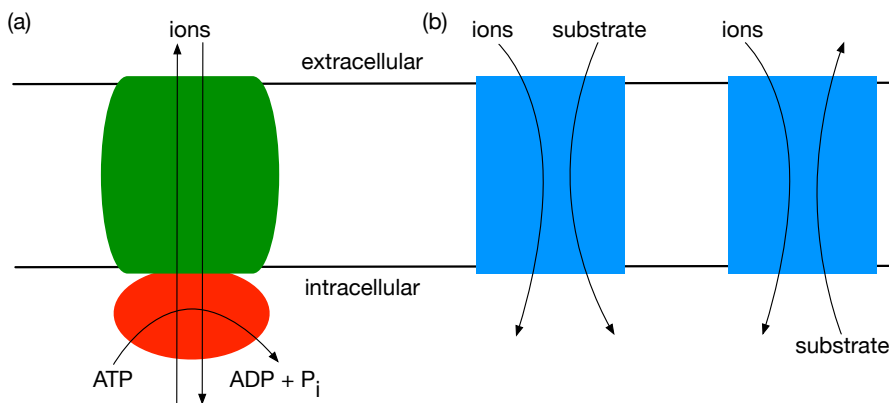
The source of energy employed by **secondary active transporters** is the potential energy stored in a transmembrane ionic gradient (Figure 1.3(b)). The movement of these ions down their concentration gradient instigates a conformational change that simultaneously drives the uphill uptake of substrates. Mechanistically, secondary active transporters can be subdivided into two groups: i) *symporters*, for which the transported substrate and the coupled ion(s) move in the same direction, and ii) *antiporters*, for which substrate and ion(s) are exchanged or move in the opposite direction<sup>†</sup>.

---

<sup>†</sup>Of note, there is another type of transporter, i.e. the *uniporter* that allows the binding and transport of *one* molecule (substrate) at a time down its concentration gradient. A good example is the mammalian glucose transporter GLUT1 (Deng et al., 2014), which when compared to its bacterial homolog, the proton-dependent



Figs. 1.2: Crystal structures of a membrane channel and transporter. (a) KcsA is bacterial potassium channel made of four identical subunits coloured in blue, green, orange and red. The positions of the selectivity filter and the gate are indicated. The structure is taken from Uysal et al. (2011) (PDB ID: 3PJS) and was solved to 3.9 Å resolution. (b) KtrAB is a bacterial potassium transporter made of a homodimeric KtrB membrane transport domain (red and orange) and an octameric KtrA cytoplasmic regulatory domain (blue and green). The structure was solved at 3.5 Å resolution by Vieira-Pires et al. (2013) (PDB ID: 4J7C). Two black lines represents the approximate position of the plasma membrane.



Figs. 1.3: Primary and secondary active transporters. (a) A schematic of a primary active transporters. A nucleotide binding domain (red) is often attached to the transmembrane domain (green) to couple ATP hydrolysis to ionic transport. (b) Two types of secondary active transporter: symport (left) which facilitate the uptake of substrate in the same direction as the ionic gradient, and antiport (right) that transport substrate in the opposite direction compared to the ionic gradient.



The two most common species coupled to substrate transport by these proteins are sodium ions and protons; the former is more prominent than the latter in eukaryotes, whilst both are featured equally in prokaryotes (Saier, 2000). In most tissues of a multi-cellular eukaryote, stringent pH homeostasis maintains a negligible difference between cytoplasmic and external pH (Casey et al., 2010), thus the weak transmembrane proton gradient results in the cell resorting to other ionic gradient, for example that of sodium, as a source of energy. Peptide transporters are examples of the very few eukaryotic membrane transporters that are driven by proton electrochemical gradient. By contrast, bacteria have little control over the extracellular environment within which they grow, and therefore internal and external pH difference can be quite pronounced such that they can harness the transmembrane proton gradient as an energy source (Krulwich et al., 2011).

### 1.1.3 Structural and mechanistic studies of secondary active transporters

Secondary active transporters are essential in various important metabolic processes like nutrient uptake, pH homeostasis and neurotransmitter regulation, and have been implicated in multiple diseases such as depression and cancer. As such, they have been one of the primary foci for structural biology, as understanding their three dimensional architecture is instrumental to elucidating their molecular mechanism.

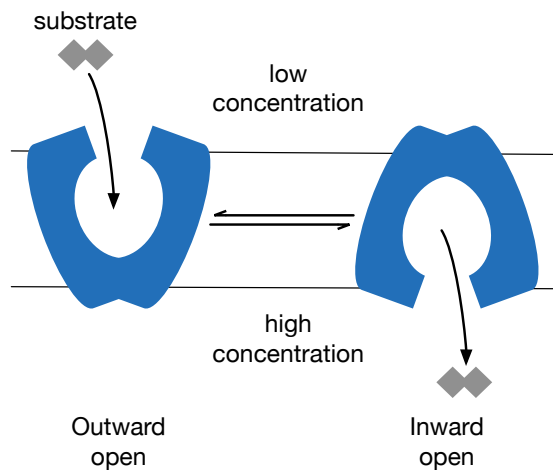
#### Alternating access model

Long before the first X-ray crystal structure of a secondary active transporter was published, a simple mechanistic model was formulated by Jardetzky (1966). The model proposed three simple features that must be met by a transporter (referred to as “pump” in the original paper<sup>‡</sup>) to conduct active transport: i) it must have a binding cavity that fits a substrate molecule, ii) it can adopt two conformational states, one where the binding cavity is exposed to the extracellular side (*outward open*), and another where it is opened to the intracellular side (*inward open*), and iii) the affinity of the binding site for the substrate molecule must differ in these two states. Active transport begins with the substrate binding to the transporter from the side of the membrane where substrate concentration is low. This is followed by a conformational change triggered by an additional process (Jardetzky (1966)

---

xylose transporter, XylE, has an asparagine in place of a key aspartate residue. This suggests that a uniporter is essentially a permanently protonated symporter, hence has lost the ability to conduct active transport.

<sup>‡</sup>At the time this model was proposed, little is known regarding the distinction between the primary active *pumps* and secondary active *transporters*. The sodium-potassium pump was used as an example to illustrate the model, hence the involvement of phosphorylation to drive the conformational change.

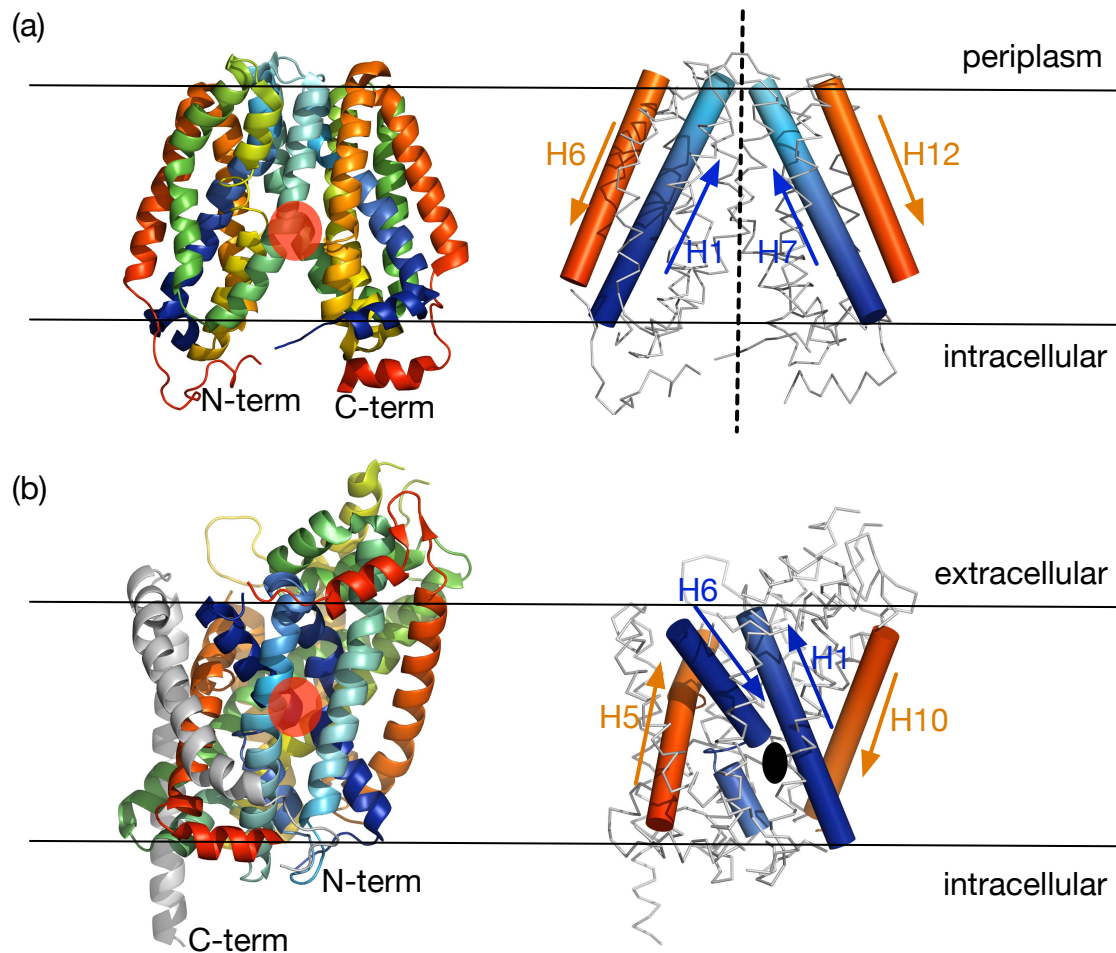


Figs. 1.4: The alternating access mechanism proposed by Jardetzky (1966). The model suggests that a transporter explores two configurational states such that the binding site is exposed to one side of the membrane in one configuration and to the opposite site in the other. Two black lines represent the plasma membrane.

used phosphorylation/ATP hydrolysis as an example), which lowers the binding affinity and results in substrate release to the side of the membrane where its concentration is higher (Figure 1.4). This coupling factor, in the case of secondary active transporter is the ionic gradient, is responsible in driving the uphill uptake of substrate. The model has since been called the “alternating access mechanism” referring to the fact that the binding site can only be accessed from one side of the membrane at any given point.

### Common structural folds

The first X-ray crystal structures of a secondary active transporter were of the proton-dependent lactose permease LacY (Abramson et al., 2003) (Figure 1.5(a)) and the phosphate glycerol-3-phosphate antiporter GlpT (Huang et al., 2003) from *E. Coli*. These two transporters belong to the major facilitator superfamily (MFS) (Pao et al., 1998) and therefore share a similar overall fold. They have twelve transmembrane helices that form two bundles, i.e. the N-terminal bundle consists of H1 to H6 and the C-terminal bundle H7 to H12. These two helical bundles are related by an approximate two-fold pseudo-symmetry axis *perpendicular* to the plane of the membrane (Forrest, 2013). What this means is if the structure is viewed in the plane of the membrane, H1 would appear symmetrical to H7 with respect to a central imaginary axis along the membrane normal. Accordingly, H2 is symmetric to H8, H3 to H9, and so forth. These helical bundles can therefore be superimposed with each other with a relatively low RMSD value, for example, 2.2 Å C $\alpha$  RMSD for LacY. This observation is rather unanticipated given the low sequence homology between the corresponding helices. The crystal structures also revealed a substrate binding site in the middle of the two terminal bundles, which is only accessible to solvent from the cyto-



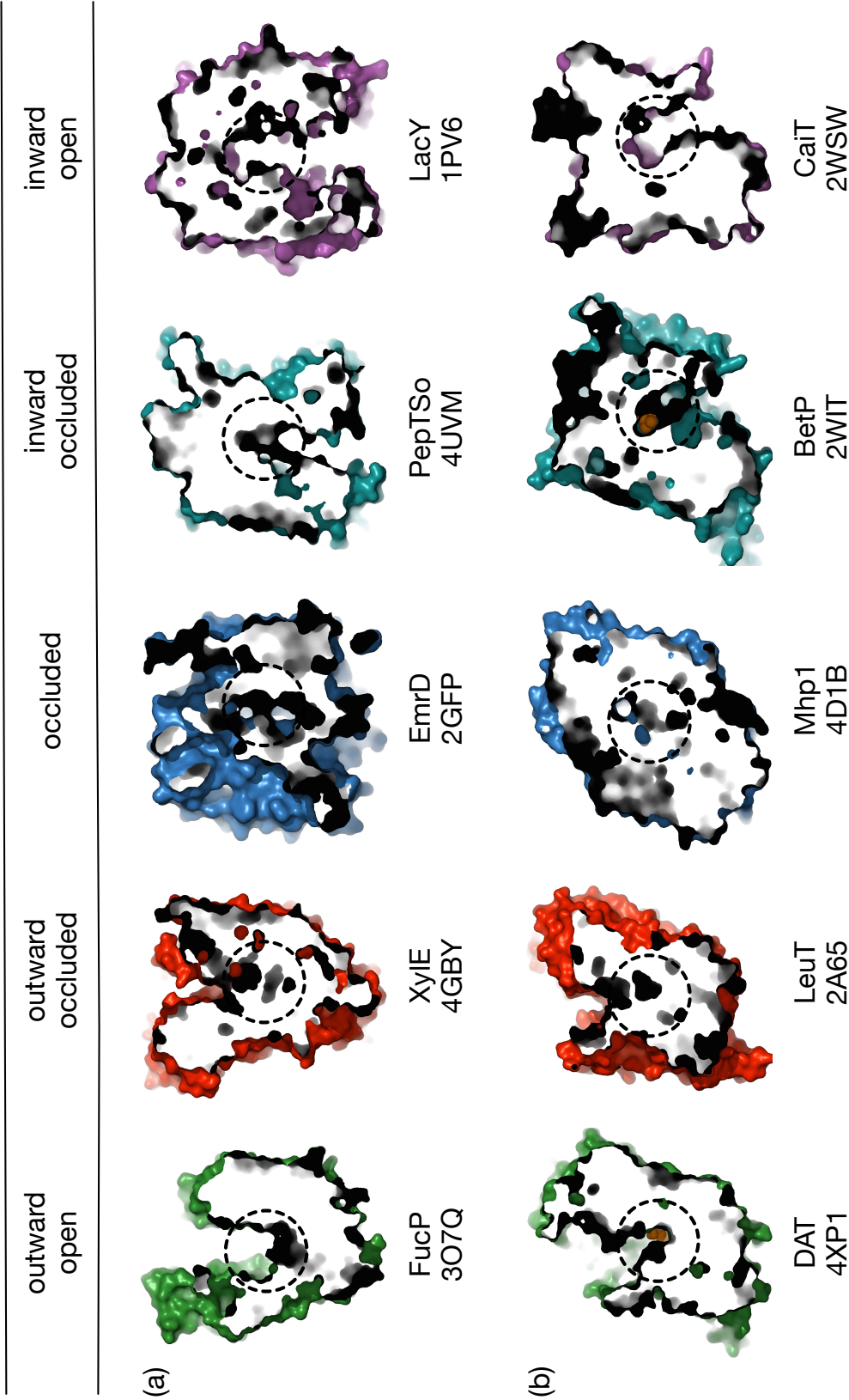
Figs. 1.5: Two common structural folds found in the secondary active transporters. (a) The crystal structure of LacY (PDB ID: 1PV6) representing the MFS fold. On the left, the structure in cartoon representation coloured in rainbow (from blue to red) for each repeat unit. The approximate position of the substrate binding site is indicated by the red circle. On the right, two pairs of helices are highlighted to demonstrate the pseudo-symmetry, with the axis displayed as a dashed line. (b) The crystal structure of LeuT (PDB ID: 2A65) representing the LeuT fold displayed in a similar representation as LacY with the pseudo-symmetrical axis shown by the black ellipsoid. Two black lines represents the approximate position of the plasma membrane.

plasmic side—hence, both structures are termed *inward open*. They were therefore the first structural confirmation of the alternating access model proposed 40 years earlier (Jardetzky, 1966).

Two years later, another crystal structure of a secondary active transporter with a novel structural fold was published. The structure was of a sodium/leucine symporter LeuT (Figure 1.5(b)) from bacteria *Aquifex aeolicus* (Yamashita et al., 2005). LeuT is a homolog of the neurotransmitter/sodium symporters (NSS) (Jack et al., 2000). The structure consists of twelve transmembrane helices, although only ten of which are deemed the core structure. An internal structural repeat can also be observed, where a pseudo-two-fold symmetry axis relates H1 to H5 with H6 to H10. Unlike LacY, however, the axis of symmetry is *parallel* to the membrane. In other words, when viewed along the plane of the membrane, to superimpose the two helical bundles, one has to be turned 180° upside down, in contrast to a 180° sideways rotation for LacY. Substrate and sodium binding sites are found between H1 and H6, where the  $\alpha$ -helices are shortly disrupted. These binding sites are not accessible to solvent from either side of the membrane, but, due to the extracellular side of the transporter being more opened, the structure is categorised as *outward occluded*. Viewed in the context of Jardetzky's alternating access model, the structure introduces a new intermediate configuration through which a transporter must undergo between the outward open and the inward open states, wherein both sides of the binding cavity are blocked (Gouaux, 2009).

Subsequent structural studies elucidated the crystal structures of numerous other secondary active transporters, most of which can be classified into either the MFS or the LeuT fold. These structures expand the alternating access model by identifying new conformational intermediates between the two end states proposed by Jardetzky (1966) (Figure 1.6). Whilst most of the earlier structural data of secondary active transporters were derived from bacteria, the number of crystal structures of eukaryotic transporters has begun to increase in the past few years. Amongst the first to be crystallised were a fungal phosphate transporter PipT (Pedersen et al., 2013), a dopamine transporter from *Drosophila melanogaster* dDAT (Penmatsa et al., 2013), a plant nitrate transporter NRT1.1 (Parker and Newstead, 2014; Sun et al., 2014) and a human glucose transporter GLUT1 (Deng et al., 2014).

Some secondary active transporters have a unique architecture that does not belong to either the MFS or LeuT fold. These include the glutamate transporter homolog Glt<sub>Ph</sub> (Yernool et al., 2004), the sodium/proton antiporter NhaA (Hunte et al., 2005) and the proton-coupled chloride exchangers ClC (Dutzler et al., 2002; Miller, 2006).



Figs. 1.6: Cross sectional surface representations of X-ray crystal structures of various secondary active transporters in different conformational states. (a) Transporters in the MFS fold and (b) transporters in the LeuT fold. Dotted circles represent the approximate position of the binding site. Where available, bound substrates are represented by orange spheres.

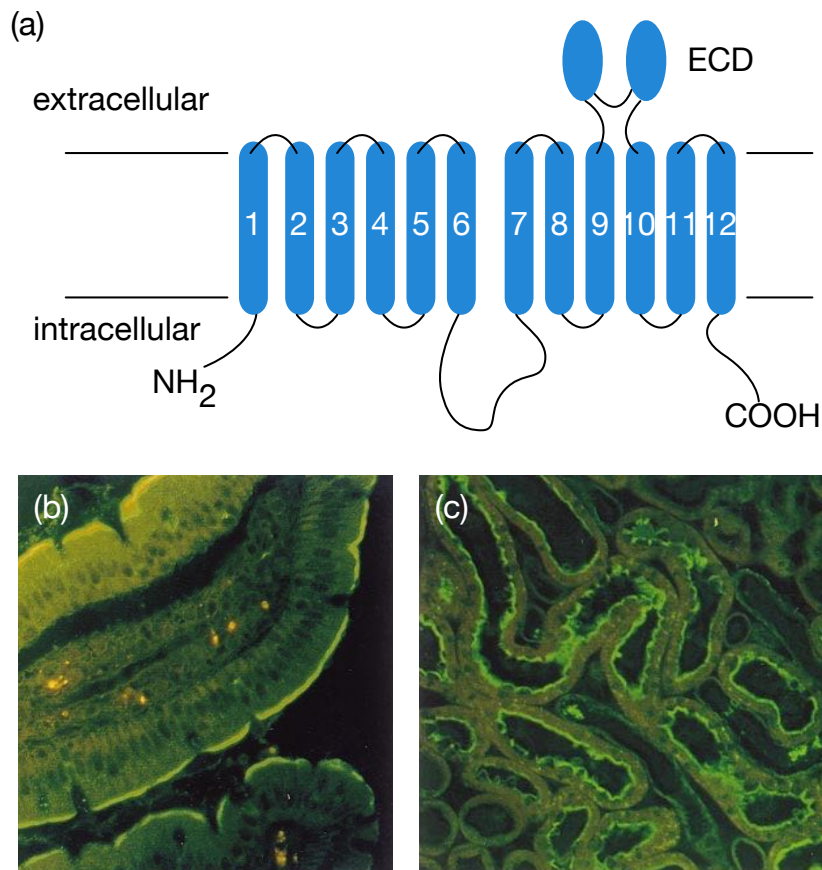
## 1.2 Proton-dependent oligopeptide transporters

Nutrient uptake is a critical component of transmembrane transport for all living organisms. In eukaryotes, specialised cells have evolved to have a membrane studded with transport proteins giving the ability to efficiently absorb nutritional compounds. This is exemplified by the mammalian digestive system, wherein nutrients are broken down into their molecular constituents and taken up into the blood circulation. One key nutrient is protein, which is the building block of tissues and a source of energy. Protein is digested first by salivary enzymes in the oral cavity and then by a mixture of proteases and peptidases in the stomach. It was first thought that all dietary proteins were broken down into amino acids for intestinal absorption, but the time scale of this process is prohibitively long (Frankel, 1916), suggesting that other strategies are also employed. Early studies with extracts from brush border membranes revealed high intracellular peptidase activities, indicating that not all oligopeptides are hydrolysed in the lumen. Di- and tripeptides were found to be accumulated inside the cytoplasm and their transport was separate from that of amino acid (Addison et al., 1975; Himukai et al., 1978; Webb et al., 1992). It was also demonstrated that an increase in intracellular peptide concentration was accompanied by an increase in membrane depolarisation (Boyd and Ward, 1982) and that an inwardly directed proton gradient stimulated peptide uptake (Ganapathy and Leibach, 1983). These data point towards a distinct proton-dependent peptide uptake system—the proton-coupled oligopeptide transporters (POTs).

### 1.2.1 The SLC15 gene family

The dawn of recombinant DNA technology brought forth expression cloning of the first mammalian POT from rabbit small intestine (Fei et al., 1994), followed later by its human homolog (Liang et al., 1995). Homology screening using a probe derived from the cDNA of the intestinal POT isolated another isoform from kidney (Liu et al., 1995). The intestinal isoform was named PepT1 and the renal isoform PepT2. Based on the HUGO Gene Nomenclature Committee (HGNC), these genes encoding PepT1 and PepT2 are designated members of the *SLC15* family (Daniel and Kottra, 2004). So far, this family has four members, PepT1 (*SLC15A1*), PepT2 (*SLC15A2*), PHT1 (*SLC15A4*) and PHT2 (*SLC15A3*). The PHT proteins have very low sequence homology with the rest of the family, and currently very little is known in regards to the function of the proteins; they are therefore beyond the scope of this thesis.

PepT1 and PepT2 share around 50% sequence identity and 70% similarity. Hydropathy analysis (Kyte and Doolittle, 1982) of the primary amino acid sequence for both proteins



Figs. 1.7: Membrane topology and expression of mammalian POTs. (a) The topology of PepT1 and PepT2 as suggested by hydropathy analysis (Fei et al., 1994) and epitope-tagging experiments (Covitz et al., 1998). (b) Immunolocalisation of PepT1 in rat small intestine. (c) Immunolocalisation of PepT2 in rat kidney proximal tubule. Both transporters are found on the apical membrane of brush border cells. Figure (b) and (c) are taken from Shen et al. (1999).

suggested the presence of twelve membrane-spanning domains, typical of MFS transporters (Figure 1.7(a)). Utilising an epitope insertion methodology, Covitz et al. (1998) validated this topology and confirmed the positions of the intra- and extracellular loops as well as the C-terminal. Intriguingly, this topology suggests the existence of a large extracellular domain (ECD) linking H9 and H10 that is unique to mammalian POTs and completely absent in bacterial, fungal and plant homologs. PepT2 is slightly longer with 729 amino acids compared to 708 in PepT1, with the additional residues found on the cytosolic N-terminal. Two and five putative phosphorylation sites are present in PepT1 and PepT2, respectively. PepT1 has seven putative N-glycosylation sites whereas PepT2 has five, and congruently based on Western blot analysis, they exhibit different levels of glycosylation (Boll et al., 1996).

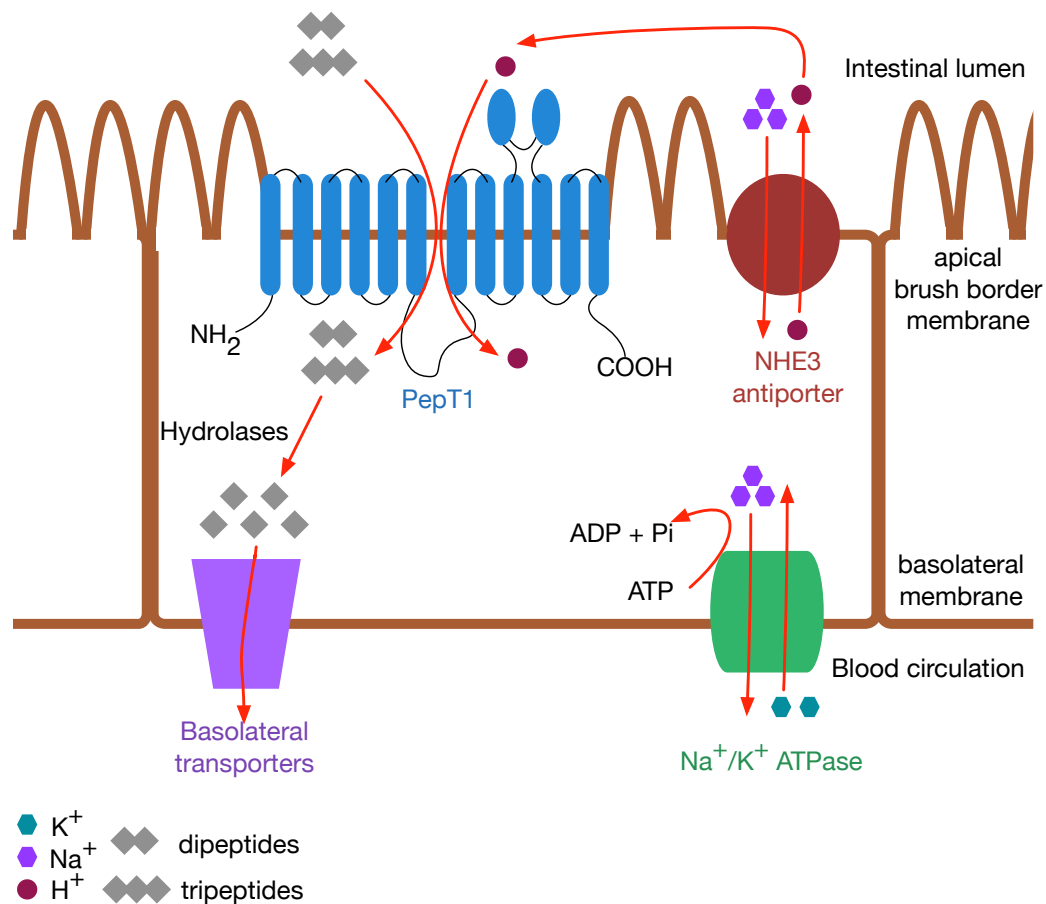
The expression of PepT1 and PepT2, respectively, is not limited to the small intestine and kidney. PepT1 shows a high degree of expression in the liver with a lower level of expression also observed in some tissues in the brain like the cerebellum (Fei et al., 1998). Both transporters are found in the proximal tubule of the kidney but their localisations are different; PepT1 is expressed in a segment closer to the glomerulus whilst PepT2 is found in the proximity of the loop of Henle (Shen et al., 1999). PepT2 is more widely expressed with its mRNA also found in the lung (Groneberg et al., 2001), mammary gland (Groneberg et al., 2002) and cardiomyocytes (Lin and King, 2007). In the brain, PepT2 is expressed in astroglia where it transports the dipeptide CysGly, a precursor for the synthesis of glutathione (Dringen et al., 1998), and in the choroidal epithelial cells where it facilitates the efflux of peptides from cerebrospinal fluid into the blood (Shu et al., 2002). Wherever they are expressed, however, PepT1 and PepT2 are always found on the apical side of the brush border membrane (Figure 1.7 (b) and (c)). A network of coordinated actions is present between the mammalian POTs and other membrane proteins found in the intestinal epithelial cells, such as the sodium-proton exchanger, NHE3, and the sodium-potassium pump, to maintain cellular ion concentrations whilst absorbing nutrients (Figure 1.8).

### 1.2.2 Substrate specificity of PepT1 and PepT2

The substrates of PepT1 and PepT2 are practically all physiological di- and tripeptides; if the twenty naturally occurring amino acids are considered, these amount to a total of 8400 compounds. It is worth acknowledging, however, that not all of these compounds have been experimentally tested for uptake activity with the transporters<sup>§</sup>, and some large tripeptides

<sup>§</sup>Functional studies of PepT1 and PepT2 are usually performed in whole cell assays, e.g. the Caco-2 cell monolayer. Substrate affinities are determined using competition uptake assays with <sup>13</sup>C-Glysar typically used





Figs. 1.8: A schematic representation of epithelial cells in the small intestine, highlighting the interplay between PepT1 and other membrane proteins.  $Na^+/K^+$  ATPase on the basolateral membrane pumps two  $K^+$  ions into and three  $Na^+$  ions out of the cell, creating a  $Na^+$  gradient, i.e. fewer  $Na^+$  ions inside the cell than the outside. To restore the intracellular  $Na^+$  concentration, a  $Na^+/H^+$  exchanger, NHE3, on the apical membrane take up  $Na^+$  ions whilst removing protons, effectively increasing the external proton concentration. PepT1 uses this proton electrochemical gradient to actively transport di- and tripeptide from the intestinal lumen into the epithelial cells, in which they are hydrolysed into amino acids before being moved into the blood circulation by basolateral membrane carriers. The figure is adapted from Daniel and Spanier (2006).

like TrpTrpTrp have been demonstrated to bind with high affinity but not transported, and are therefore competitive inhibitors (Biegel et al., 2005, 2006). Even so, the binding cavity of PepT1 and PepT2 have evolved to be very promiscuous to accommodate such a vast number of different substrates. There are only four transporters involved in the uptake of peptides (the SLC15 family), compared to seven transporters designated for amino acids and 26 for sugars, indicating that most secondary active nutrient transporters are substrate specific and the promiscuity of PepT1 and PepT2 is unique.

Although these transporters are able to translocate a diverse set of compounds, the affinity at which they interact greatly varies according to the size and chemical nature of the substrate (for an in depth review, see Brandsch et al. (2004)). First of all, it was established early on that neither amino acids nor tetra- and pentapeptides bind to the transporters (Adibi and Morse, 1977; Ganapathy et al., 1981). Both the amino and carboxyl termini are indispensable for peptide recognition, with the methylation or acetylation of the former greatly aggravating binding while similar modification to the latter also reduces affinities but to a lesser extent (Addison et al., 1975; Eddy et al., 1995; Brandsch et al., 1999; Meredith et al., 2000). The peptide bond, on the other hand, is not essential. Brandsch et al. (2004) replaced the amide linkage of GlyGly with a ketomethylene group and measured similar  $K_i$  values for PepT1. A similar result was documented for the renal isoform PepT2 (Daniel and Adibi, 1994).

PepT1 and PepT2 have a strong preference for hydrophobic side chains like valine, phenylalanine and tryptophan (Vig et al., 2006; Biegel et al., 2006). Any charged peptide has a significantly lower affinity than its neutral counterpart. For example, both LysLys and GluGlu have inhibition constants,  $K_i$ , in the mM range for PepT1, compared to 80  $\mu$ M for AlaAla (Eddy et al., 1995; Terada et al., 2000b). The binding of anionic and cationic peptides is also pH dependent, as illustrated by the increase in affinity of dipeptides with negatively charged residues at lower pH, suggesting that the transporters favour the zwitterionic form of a charged peptide (Amasheh et al., 1997; Fei et al., 1999). Additionally, measurements of inward current in *X. laevis* oocytes indicated a different stoichiometry between a neutral and an anionic peptide, i.e. 1:1 peptide to proton ratio for the former, but 2:1 for the latter, implying that one additional proton is required to neutralise the side chain (Steel et al., 1997; Chen et al., 1999). Peptides with an N-terminal proline residue showed very weak binding, but positioning the same residue on the C-terminus results in high affinity, suggesting that the binding cavity is not symmetrical (Brandsch et al., 1999).

---

as the reporter substrate. The usual metrics used to compare between different substrates are the inhibition constant  $K_i$  and  $IC_{50}$  values. This will be discussed in depth in §2.4.

Table 1.1: Different affinities of peptides to PepT1 and PepT2

| Peptide   | $K_i$ (mM) |       |
|-----------|------------|-------|
|           | PepT1      | PepT2 |
| AlaAla    | 0.10       | 0.006 |
| AlaAsp    | 0.26       | 0.014 |
| CysGly    | 0.20       | 0.029 |
| LeuPro    | 0.11       | 0.010 |
| PheAla    | 0.11       | 0.016 |
| AlaAlaAla | 0.18       | 0.018 |
| AlaAspAla | 0.72       | 0.079 |
| MetMetMet | 0.10       | 0.002 |

$K_i$  measured in competition transport assays performed in Caco-2 (PepT1) and SPKT (PepT2) cell monolayers using  $^{14}\text{C}$ -Glysar as the reporter substrate (Biegel et al., 2006).

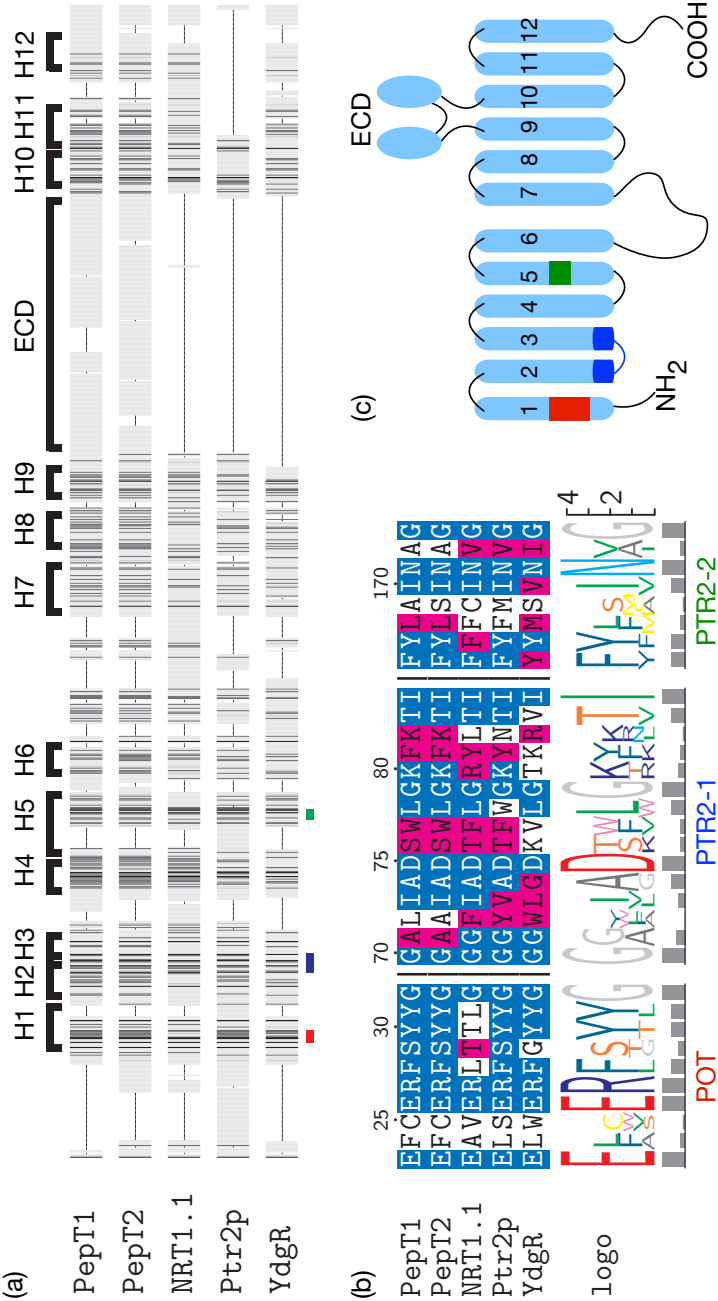
Peptide binding to PepT1 and PepT2 is also stereospecific. L-L-dipeptides or L-L-L-tripeptides display remarkably higher affinity than an equivalent D-amino acid containing peptide (Lister et al., 1995). In a systematic study using ValVal dipeptide and ValValVal tripeptide, Li et al. (1998) demonstrated that isomers having more than two D-Val residues do not interact with PepT1 at all. Dipeptides with only one D-isomer on the N-terminus, however, are accepted, as exemplified by D-Phe-L-Gln and D-Phe-L-Ala (Meredith et al., 2000; Theis et al., 2002).

Whilst the overall substrate preference of PepT1 and PepT2 is virtually identical, binding affinities recorded for the same peptides notably differ. This observation led to the designation of PepT1 as the “high capacity, low affinity” isoform, whereas PepT2 the “high affinity, low capacity” isoform (Smith et al., 2013). These names are given in reference to the striking difference of the kinetics parameters, i.e. the  $K_m$  (affinity) and  $V_{\max}$  (capacity), of these two transporters. For most substrates, affinities to PepT2 are at least ten times higher than to PepT1, as illustrated in Table 1.1. Under comparable experimental settings, the range of  $K_m$  values measured for PepT1 is from 0.2 mM to 10 mM, whereas for PepT2 is from 5  $\mu\text{M}$  to 0.5 mM (Rubio-Aliaga and Daniel, 2002). Nonetheless, little is currently known about the structural origin of these discrepancies, although various chimeric studies have been performed to shed light on this issue (Doring et al., 1996; Döring et al., 2002; Fei et al., 1998). Their results, however, are inconclusive and, at times, contradictory.

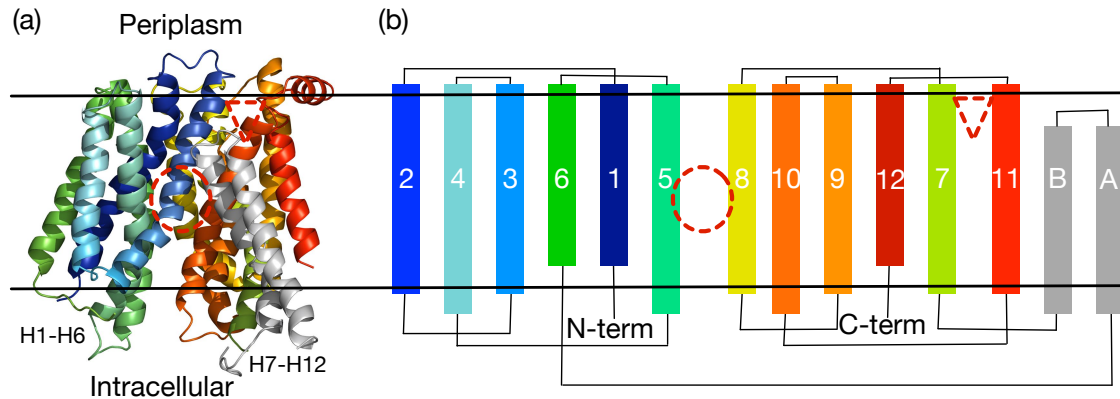
### 1.2.3 Conserved motifs from bacteria to human

In simpler organisms without a fully developed digestive track, a peptide uptake system is also important to acquire nitrogen, which is fundamental for cellular metabolism. Bacteria have three different ways to absorb peptides from its environment: the oligopeptide permease (Opp), the dipeptide permease (Dpp) and the POTs. The first two are ABC transporters; the Opp is able to transport peptides with four to eighteen residues (Detmers et al., 1998), whereas the Dpp can only transport dipeptides. Another proton-dependent symporter catering for tetra- and pentapeptides, the oligopeptide transporter (OPT), is present in yeast, but bears no sequence resemblance to the POT family. To date, only the POT family is ubiquitously found from bacteria to man (Daniel and Spanier, 2006), with it being the exclusive peptide transport system for higher order eukaryotes. In contrast, plants gain nitrogen primarily through the uptake of nitrate from the soil. This is achieved through nitrate transporters that are phylogenetically classed into the broader proton-dependent symporter family called NPF (Nitrate transporter 1/Peptide transporter Family), the substrates of which include nitrate, peptides, amino acids, dicarboxylate and glucosinolate (Léran et al., 2014). In spite of the various ways of acquiring nitrogen across different kingdoms of life, the POT family shows a high degree of sequence similarity.

There are three small stretches of residues, or motifs, that are universally conserved amongst the POT transporters. These are the POT, PTR2-1 and PTR2-2 motifs (Figure 1.9) (Daniel and Spanier, 2006). The POT motif is found in H1, the PTR2-1 motif in the loop linking H2 and H3, while the PTR2-2 motif is located in H5. Functional studies have shown that these conserved domains do affect peptide uptake activity. The crystal structure of a bacterial homolog from *S. thermophilus* PepT<sub>St</sub> revealed that the POT motif forms a significant portion of the binding cavity and point mutations to most residues within this region abolished transport (Solcan et al., 2012). Intriguingly, in a plant nitrate transporter NRT1.1, two aromatic residues (Tyr30 and Tyr31 in human PepT1) are replaced by a threonine and leucine, respectively. Crystallographic data suggest that having smaller side chains at these positions helps nitrate bind and at the same time precludes peptide recognition and transport (Parker and Newstead, 2014; Sun et al., 2014), pointing towards the role of the POT motif in substrate selectivity. Döring et al. (2002) performed a chimeric study on rabbit PepT1 and PepT2 and proposed that residue 60-91, wherein lies the PTR2-1 motif, act as a pH sensor for the protein. Additionally, mutation at Tyr167 of human PepT1, which is found in the PTR2-2 motif, to any other residue eliminated uptake activity (Yeung et al., 1998), whereas alanine scanning mutagenesis of this domain altered the substrate preference of yeast peptide transporter Ptr2p (Hauser et al., 2005).



Figs. 1.9: Conserved motifs in the POT transporter family. (a) An amino acid sequence alignment of human PepT1 (Uniprot ID: P46059), human PepT2 (Q16348), a plant nitrate transporter from *A. thaliana* NRT1.1 (Q05085), a yeast peptide transporter Ptr2p (E7Q6E4) and an *E. coli* peptide transporter YdgR (P77304), represented as fingerprints with conserved sequence coloured in dark grey. The three conserved motifs, POT, PTR2-1 and PTR2-2 are highlighted in red, blue and green respectively. (b) Sequence alignment of the conserved motifs; identical residues shown in blue and similar residues in red. Consensus sequence shown at the bottom in WebLogo representation ([weblogo.berkeley.edu/logo.cgi](http://weblogo.berkeley.edu/logo.cgi)) (Crooks et al., 2004) and the grey bars indicate the degree of conservation. (c) The approximate positions of the three conserved motifs mapped onto the topology of PepT1.



Figs. 1.10: The crystal structure of PepT<sub>So</sub>. (a) The structure is depicted in cartoon representation and coloured in rainbow from blue (N-terminal) to red (C-terminal), and additional helices HA and HB illustrated in grey. The position of the peptide binding site and extracellular cavity are marked with a dashed circle and inverted triangle, respectively. (b) A topology diagram of PepT<sub>So</sub> displayed in the same colour as (a). The approximate position of the membrane bilayer is highlighted by the two black bars.

### 1.2.4 Crystal structures of bacterial homologs

Table 1.2: Available crystal structures of bacterial POTs

| Name                | Conformation    | Resolution (Å) | PDB ID                       |
|---------------------|-----------------|----------------|------------------------------|
| PepT <sub>So</sub>  | Inward occluded | 3.0-3.6        | 2XUT, 4UVM                   |
| PepT <sub>St</sub>  | Inward open     | 2.4-3.3        | 4APS, 4D2B, 4D2C, 4D2D       |
| GkPOT               | Inward open     | 1.9-2.4        | 4IKV, 4IKW, 4IKX, 4IKY, 4IKZ |
| PepT <sub>So2</sub> | Inward open     | 3.2-3.9        | 4LEP, 4TPJ, 4TPH, 4TPK       |
| YbgH                | Inward open     | 3.4            | 4Q65                         |
| YePepT              | Inward open     | 3.0            | 4W6V                         |

Given the high sequence homology within the POT family, as exemplified by the three signature motifs, it is reasonable to use POT transporters from simpler organisms as model systems to study human PepT1 and PepT2. To this end, X-ray crystallography has been employed to elucidate the three dimensional structure of bacterial POT transporters, from which the molecular details of how they operate can be inferred. So far, the structures of five peptide transporters from four different species of bacteria have been solved (Table 1.2).

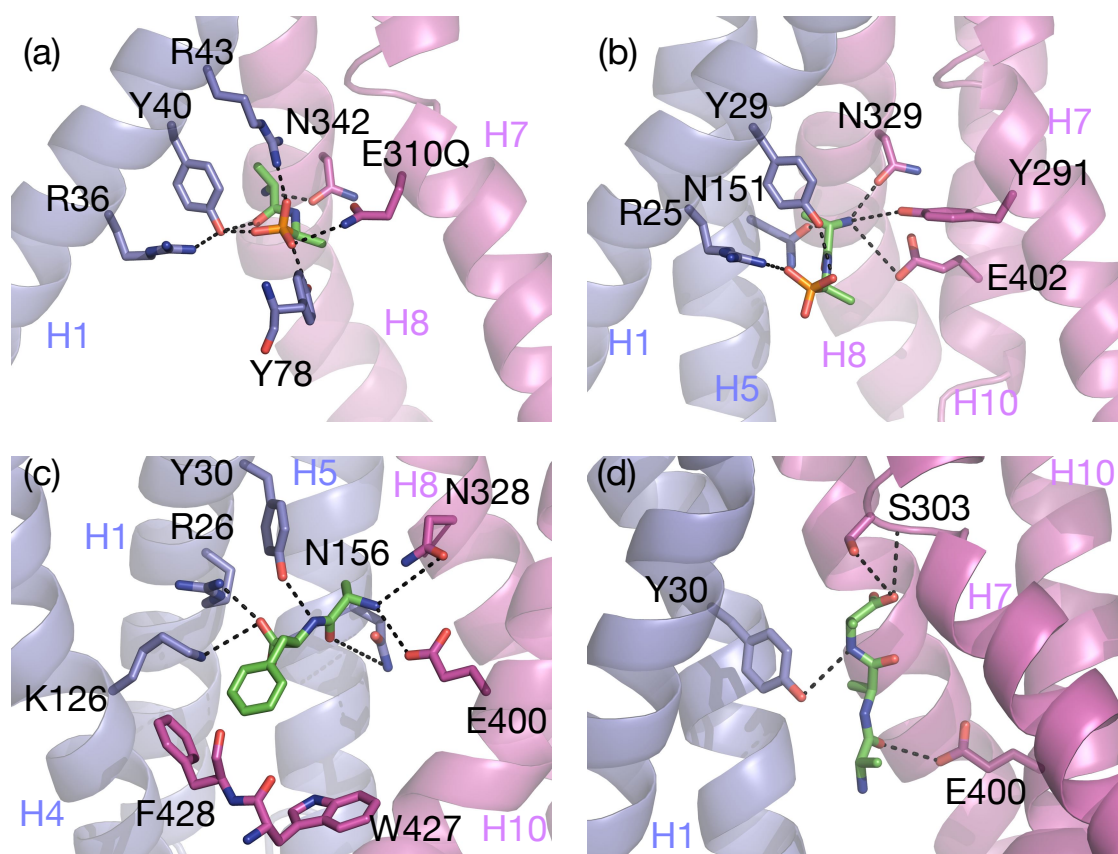
The first of these structures was PepT<sub>So</sub> from *Shewanella oneidensis*, which was solved to 3.6 Å resolution by Newstead et al. (2011). This paper provided a first look at the molecular architecture of the POT family and confirmed that they belong to the MFS. As il-

illustrated in Figure 1.10, PepT<sub>So</sub> is made of a twelve-transmembrane-domain core, within which lie the pseudo-symmetric repeats similar to the prototypical MFS transporter LacY (Figure 1.5(a)). Intriguingly, the structure showed two additional transmembrane helices inserted between the N- and C-terminal bundles, called the HA and HB. They are shorter than the other helices and form a hairpin on the periphery of the core domains. The central cavity is occupied by electron density the size of a dipeptide, although due to the low resolution, the identity of this compound could not be confirmed. This cavity was designated the peptide binding site owing to its similarity to where a sugar analog binds to LacY. Most of the key residues are also found lining the cavity, lending further support to this claim. The structure was described as “inward occluded” as the binding site is not accessible from both sides of the membrane but is more open on the cytoplasmic end. An additional cavity exists between H7 and H11 and is exposed to the extracellular side, hence called the extracellular cavity. A new crystal structure of the same transporter was recently solved to a higher resolution of 3.0 Å (Fowler et al., 2015). Following PepT<sub>So</sub>, Solcan et al. (2012) solved the crystal structure of PepT<sub>St</sub> from *Streptococcus thermophilus* to a resolution of 3.3 Å in an inward open structure, providing a new conformational state for the POT family.

### Substrate binding site

What was missing from the crystal structures of PepT<sub>So</sub> and PepT<sub>St</sub> is a detailed illustration of how a substrate molecule interacts with the transporter in the binding cavity. Succeeding crystal structures of two other bacterial POTs and peptide bound PepT<sub>St</sub> tried to address this question. The first depictions of ligand binding to a POT transporter came from a complex of the dipeptide analog Alafosfalin and a POT from the bacterium *Geobacillus kaustophilus* (GkPOT) (Figure 1.11(a)) (Doki et al., 2013). The crystals diffracted to a high resolution of 1.9 Å. The structure revealed several major interactions made between Alafosfalin with residues from the POT motif in H1. These were assumed to stabilise the phosphate terminal of the molecule, strengthening the argument that this invariably conserved region is instrumental in substrate recognition. Doki et al. (2013) also proposed that the essential residue Glu310 on H7 is crucial for proton coupling, as E310Q binds to the phosphate group, suggesting that a protonated glutamate residue at this position stabilises the carboxyl terminus of a peptide.

About the same time as this structure was published, another Alafosfalin bound POT structure was elucidated by Guettou et al. (2013) (Figure 1.11(b)), albeit to a lower resolution (3.2 Å). The structure is of another homolog from *S. oneidensis*, PepT<sub>So2</sub>, and is also in an inward open state. Compared to the GkPOT structure, the position of Alafosfalin appears



Figs. 1.11: Substrate binding site of bacterial POTs, highlighting key residues interacting with the ligand and helices on which they are found. (a) The binding of a peptide analog Alafosfalin to GkPOT (PDB ID: 4IKZ) (Doki et al., 2013). (b) The binding of Alafosfalin to PepT<sub>So2</sub> (4LEP) (Guettou et al., 2013). (c) The binding of a dipeptide AlaPhe to PepT<sub>St</sub> (4D2C) (Lyons et al., 2014). (d) The binding of a tripeptide AlaAlaAla to PepT<sub>St</sub> (4D2D) (Lyons et al., 2014). Residues from the N-terminal bundle are shown in purple and the C-terminal bundle in magenta, while ligands are shown in green.

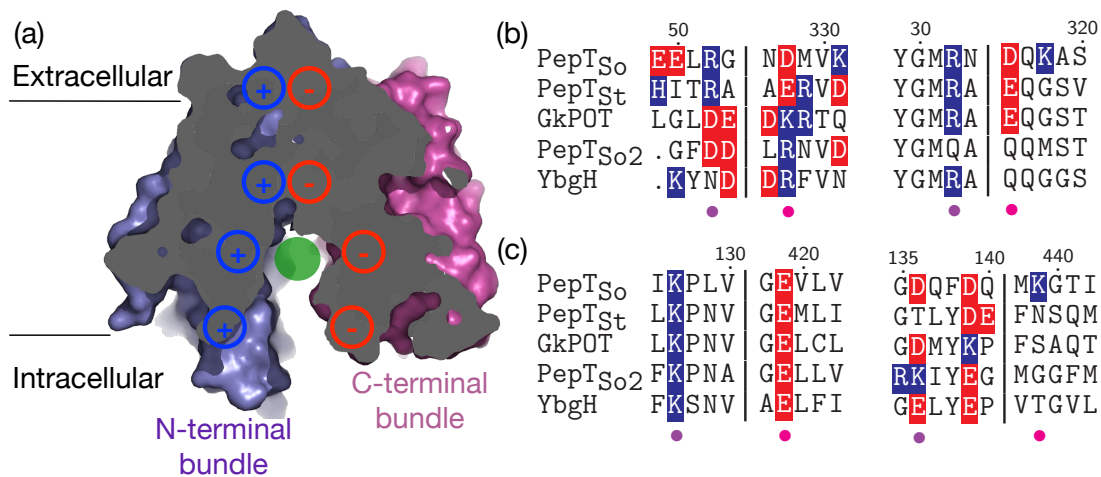


to be shifted towards the cytoplasmic side by about 4.5 Å, as a result of which, different interactions are observed especially with the C-terminal phosphate. Despite these differences, the general orientation of Alafosfalin in the binding cavity remains largely similar with the N-terminus projecting towards the C-terminal bundle of the transporter and the phosphate group towards the N-terminal bundle. One interpretation of the discrepancies is that both transporters accommodate substrate compound in distinct ways, perhaps due to subtle structural differences in the binding cavity. Later on, the crystal structures of PepT<sub>S02</sub> in complex with three peptides, AlaAlaAla, AlaTyrAla and AlaTyr, were determined to a resolution of 3.2 Å, 3.9 Å and 3.2 Å, respectively (Guettou et al., 2014). The study reported a similar binding orientation for all three peptides, which overlaps with the previously identified binding of Alafosfalin. The study defined three binding pocket into which each peptide side chain is projected and suggest that these pockets have subtly different preference for amino acid residues.

Lyons et al. (2014) solved two structures of peptide bound PepT<sub>St</sub> to a high resolution of 2.5 Å. The peptides were AlaPhe (representing a dipeptide) and AlaAlaAla (representing a tripeptide). The crystal structures revealed two strikingly different binding poses: AlaPhe was positioned laterally in the binding site whereas AlaAlaAla adopted a vertical orientation (Figure 1.11(c) and (d)). The binding of AlaPhe is in good agreement with the PepT<sub>S02</sub> peptide complexes, with both the N- and C-termini coordinated in a similar fashion. Also apparent from the structure is that AlaAlaAla makes significantly fewer interactions with the transporter compared to AlaPhe, which is perhaps why the latter has a higher affinity. Nevertheless, the residues to which these peptides interact are largely similar, suggesting that despite the different orientations, the same binding pockets are employed to accommodate the ligands.

Two conclusions can be made with regards to peptide binding to the bacterial POTs: i) the same peptide can adopt different binding poses in different transporters, as exemplified by Alafosfalin in GkPOT and PepT<sub>S02</sub> and AlaAlaAla in PepT<sub>S02</sub> and PepT<sub>St</sub>, and ii) the same transporter can interact with different peptides in different ways, as demonstrated by PepT<sub>St</sub> crystal complex with AlaPhe and AlaAlaAla. Whilst most key residues are preserved, subtle structural differences do exist in the binding site of these transporters, which could explain why a particular substrate binds differently. That dipeptides and tripeptides are oriented in two distinct modes in PepT<sub>St</sub> indicates the versatility of the binding site in accommodating ligands of different structures, which is likely to be the fundamental aspect in achieving substrate promiscuity amongst the POT family.

### Intracellular and extracellular gates



Figs. 1.12: Salt bridges gating the substrate binding site. (a) A schematic depiction of gating salt bridges found on the intracellular and extracellular side of the substrate binding site (green). The N-terminal bundle is depicted in purple and C-terminal bundle magenta. (b) Sequence alignment of the five bacterial POTs with known crystal structures, highlighting two pairs of potential extracellular gating residues. (c) Two pairs of potential residues involved in intracellular gating. Residues are coloured based on their electrostatic charges, i.e. blue (positive) and red (negative). Gating residues highlighted by purple and magenta bullets corresponding to the N- and C-terminal bundles, respectively.

Controlling the access to the central substrate binding site are a series of transient constrictions formed by several transmembrane helices from both the N- and C-terminal bundles—the intracellular and extracellular gates—that are stabilised by specific key salt bridges (Figure 1.12). This flexible gating mechanism is not limited to the POT family, but is widespread amongst the MFS transporters, as exemplified by LacY (Stelzl et al., 2014). Four pairs of such salt bridges exist in PepT<sub>So</sub>, of which two are on the extracellular side (Arg52-Asp328 and Arg32-Glu316) and another two on the intracellular side (Lys127-Glu419 and Asp136-Lys439) (Newstead et al., 2011). Comparable gating residues are found in PepT<sub>St</sub>, except for the lower extracellular gate, where the aspartate residue is replaced by a threonine (N-terminal) and the lysine residue is replaced by an asparagine (C-terminal).

The functional roles of each of these gating residues have been extensively investigated (Solcan, 2013). The upper extracellular gate appears to be not crucial for both transporters as point mutation on either residue only slightly decreased transport, suggesting a potential supportive rather than essential role. The lower extracellular gate, on the other hand, is important for proton-coupled uptake in PepT<sub>St</sub> and the glutamate residue at this position is

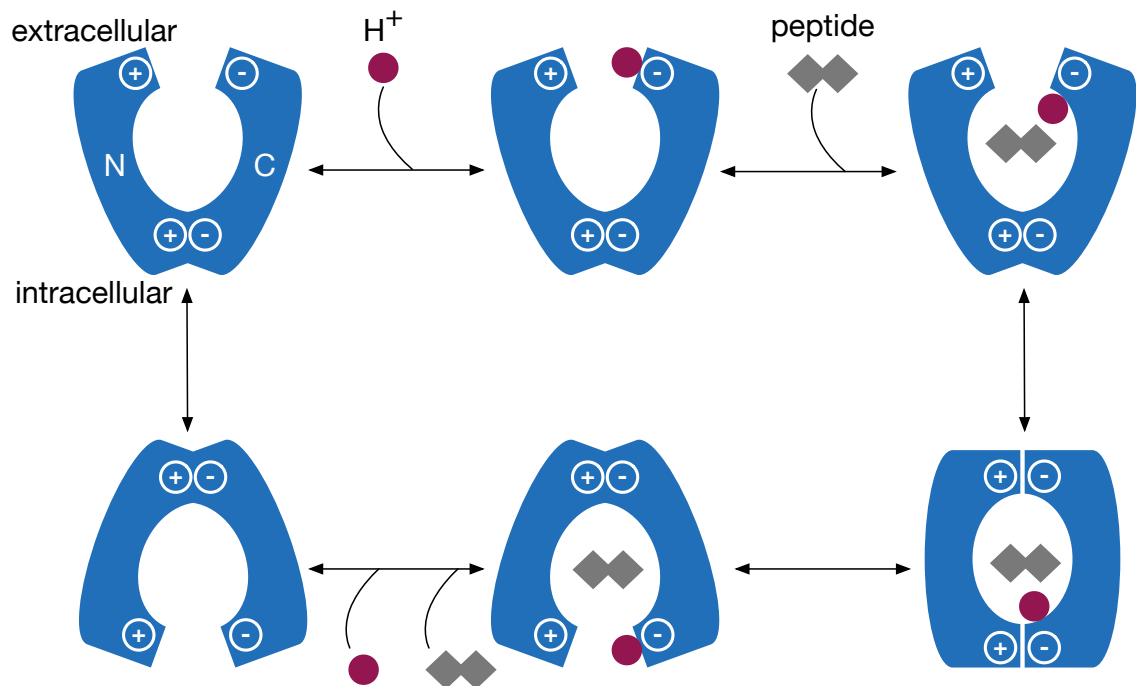
proposed to be the protonation site. Interestingly, the R32A mutant of PepT<sub>So</sub> (lower extracellular gate residue) is capable of performing normal transport with no reduction in activity at all. A possible explanation to this observation is that the adjacent His61 in its protonated form could substitute for the arginine residue and act as the gating residue for the N-terminal bundle. For both transporters, the universally conserved intracellular gate is absolutely vital for peptide uptake as mutation at either position severely reduced or abolished transport.

A similar study has also been performed on the lower extracellular gate of GkPOT by Doki et al. (2013). Their results suggested that the Glu310 is a potential protonation site as point mutation inactivated proton driven peptide uptake, congruent with the crystal structure wherein E310Q binds to the phosphate group of Alafosfalin. The charges of the upper extracellular gate of GkPOT are swapped compared to PepT<sub>So</sub> and PepT<sub>St</sub>, suggesting that the exact charge at this position is not important, as long as oppositely charged residues are present. A similar observation is observed on another POT, PepT<sub>So2</sub>, whereby an acidic residue is positioned on H1 and a basic residue on H7. Whilst PepT<sub>So</sub>, PepT<sub>St</sub> and GkPOT have at least three salt bridges around the binding site, PepT<sub>So2</sub> and YbgH have only two and one, respectively. In PepT<sub>So2</sub>, apart from the two gates, a methionine residue on H11, M426, was proposed to form an additional intracellular gate by hydrophobic interactions with a leucine and phenylalanine from the N-terminal bundle, based on comparison to the occluded structure of PepT<sub>So</sub>. For YbgH, our sequence alignment shows no apparent extracellular salt bridge; It is possible that a similar hydrophobic gating mechanism suggested in PepT<sub>So2</sub> is also present in this transporter.

By measuring inter-helical distances and the intra- and extracellular openings of all of the MFS transporters with known crystal structures, Fowler et al. (2015) showed that the first two helices in each pseudo-symmetric repeat unit contribute to the formation of gates. This means H1 and H2 pack against H7 and H8 to form the extracellular gate, while H4 and H5 pack against H10 and H11 to form the intracellular gate. Taken together with the sequence alignment shown in Figure 1.12(b) and (c), it is unlikely that key salt bridges found in PepT<sub>So</sub> are conserved for all transporters within the MFS superfamily, or even within the POT family itself (as shown by the lack of corresponding residues in YbgH), but this topology based gating model could be the underlying foundation of the alternating access mechanism.

### **Proposed peptide transport mechanism**

Based on the currently available structural, biophysical and biochemical data, a general model of how peptides are transported by peptide transporters across the membrane can be



Figs. 1.13: A general model for proton dependent peptide uptake by a POT transporter. Intra- and extracellular salt bridge gates are depicted by the positive and negative charges.

inferred (Figure 1.13). In the outward open conformation, the extracellular gate is open while the intracellular gate remains closed; this configuration exposes the extracellular gating residue to the bulk solution for protonation. The extracellular pH may therefore play an instrumental role in initiating substrate transport as a lower pH (higher proton concentration) will increase the likelihood of proton binding. Indeed, increasing extracellular pH from 6 to 8 reduced the activity of PepT1 by half (Boll et al., 1994) and that of PepT2 by 5-fold (Boll et al., 1996).

A peptide, via its C-terminal carboxylate, interacts with the protonated residue, as suggested by the binding of Alafosfalin phosphate group to Gln310 in GkPOT (Doki et al., 2013). The side chains of the peptide occupies distinct binding pockets as has been observed in PepT<sub>St</sub> (Lyons et al., 2014) and PepT<sub>So2</sub> (Guettou et al., 2014). The extracellular gate subsequently closes and an occluded state is formed, with the proton transferred to another negatively charged residue within the binding cavity. A series of glutamate/aspartate residues on the C-terminal bundle, like in PepT<sub>St</sub>, allows for proton transfer to the intracellular gate, which effectively disrupts the salt bridge, inducing an inward open state. In this conformation, the peptide is released into the cytoplasm. The proton concentration in the intracellular environment is low, therefore the negatively charge residue deprotonates and

the intracellular gate is restored to form an empty occluded transporter, which cycles back to the outward open state.

Some aspects of this model are not yet fully understood due to various pieces of missing information. For example, there is currently no crystal structure of a POT transporter in an outward open conformation. As such, the initial protonation site and peptide binding orientation remain unknown. The mechanism of how an empty inward facing transporter returns to the outward open state is also not known. It is possible that the transporter stochastically interconverts between these two conformations and proton binding stabilises the outward open state and favours subsequent substrate binding and transport. Another critical piece of information that is often neglected in studies of POT transporters is the number of protons translocated during each cycle. Only very recently has such data become available from pH dependent uptake assays of PepT<sub>St</sub> (Parker et al., 2014). The study suggested two distinct proton coupling mechanisms, one for dipeptides and another for tripeptides, in harmony with crystal structures showing that they interact with the transporter differently (Lyons et al., 2014). That each tripeptide is coupled to three protons and each dipeptide to four or five protons suggest that most titratable side chains in the binding cavity are protonated during the transport cycle, corroborating the current model that protons are relayed from one negatively charge residue to another to reach the intracellular gate.

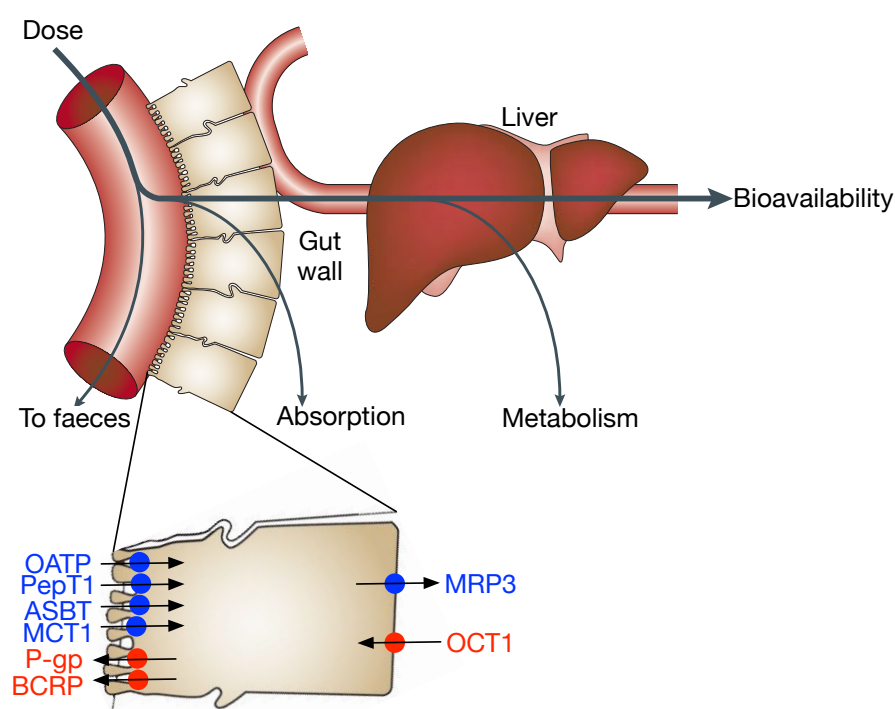
### 1.3 Oral drug bioavailability

Other than nutritional compounds like peptides, orally administered drugs are also taken up in the gut. When a drug is taken orally, it faces a great number of obstacles before it can reach the target. The fraction of unaltered drug compounds that successfully reaches the blood circulation compared to its administered dosage—the bioavailability—is the primary aspect of pharmacokinetics (Figure 1.14). The four main factors affecting the bioavailability of a drug are its absorption, distribution, metabolism and excretion or, for short, its ADME properties. Ensuring high oral bioavailability amongst drug candidates is an ongoing battle for the pharmaceutical and biotechnology industries (van de Waterbeemd and Gifford, 2003; Thomas et al., 2006). In fact, many compounds fail in pre-clinical and clinical stages due to poor bioavailability, one of the leading causes of which is poor absorption in the gastrointestinal tract (van De Waterbeemd et al., 2001). The gut wall presents a formidable barrier between the digestive tract and the systemic circulation, preventing the entry of undesired exogenous substances.

There are several ways by which a drug molecule can traverse the intestinal barrier. Most

drugs can pass through the space between epithelial cells, often referred to as paracellular permeation. Some drugs, however, have to adopt a transcellular approach, whereby they cross the apical and then the basolateral epithelial membranes to reach the blood vessels. Historically, it was thought that passive transmembrane diffusion is the principal mechanism of transcellular permeation, owing to the lipophilic nature of most drug compounds. However, in recent years, growing bodies of evidence indicate the co-existence of carrier mediated drug transport across the plasma membrane (Dobson and Kell, 2008; Sugano et al., 2010). Indeed, compelling clinical studies point towards the potential roles of nutrient transporters, not only in drug absorption and distribution, but also in therapeutic efficacy and potential adverse reactions (Shitara and Sugiyama, 2002; Cusatis et al., 2006). The US Food and Drug Administration (FDA) have published a guideline for the pharmaceutical industries that contains recommendation for pre-clinical *in vivo* and *in vitro* assessments of transporter mediated drug-drug interactions (U.S. Department of Health and Human Services Food and Drug Administration et al., 2012). Clearly, carrier mediated cellular uptake of drugs is gaining credibility and attention in drug development.

A number of transporters, from both the ABC and SLC superfamilies, have been experimentally shown to influence drug disposition (reviewed by Dobson and Kell (2008) and Giacomini et al. (2010)) (Figure 1.14). Chief amongst these is P-gp that actively exports hydrophobic drugs from the luminal membrane of the small intestine and the blood brain barrier (Zhou, 2008). Like P-gp, the breast cancer resistance protein (BCRP) is another ABC transporter that pumps out xenobiotics from the gastrointestinal tract and other tissues like kidney and the brain endothelium (Vlaming et al., 2009). In contrast, several transporters are able to take up these drug compounds from the lumen into the cell. These include the organic anion transporting polypeptide (OATP) transporters from the *SLCO* gene family that absorb drugs such as the anti-cancer agent methotrexate and the HMG-CoA reductase inhibitor statins (Hagenbuch and Gui, 2008). Other transporters that are implicated in drug uptake across the apical brush border membrane are the monocarboxylic acid transporter 1 (MCT1) from the *SLC16* family (Halestrap, 2013) and the apical sodium/bile acid co-transporter (ASBT) from the *SLC10* family (Claro Da Silva et al., 2013). The basolateral side of the epithelia, on the other hand, is populated by a group of carrier proteins that are capable of regulating drug retention in the blood stream. For example, the organic cation transporter 1 (OCT1) from the *SLC22* family assists in the removal of drugs from the systemic circulation back into the lumen (Koepsell, 2013). Its substrates include the anti-diabetic metformin and the chemotherapeutic agent paclitaxel. The ATP-dependent efflux pump, multidrug resistance protein 3 (MRP3), acts in the opposite way, i.e. it facilitates



Figs. 1.14: A schematic diagram of oral drug absorption and metabolism. Orally administered drug compounds are absorbed through the gut wall into the systemic circulation and subsequently metabolised in the liver. Bioavailability represents the fraction of the dosage that becomes available to interact with the target. Enlarged image depicts an intestinal epithelial cell highlighting several membrane embedded carrier proteins relevant in drug uptake. The figure is adapted from van de Waterbeemd and Gifford (2003) and Giacomini et al. (2010)

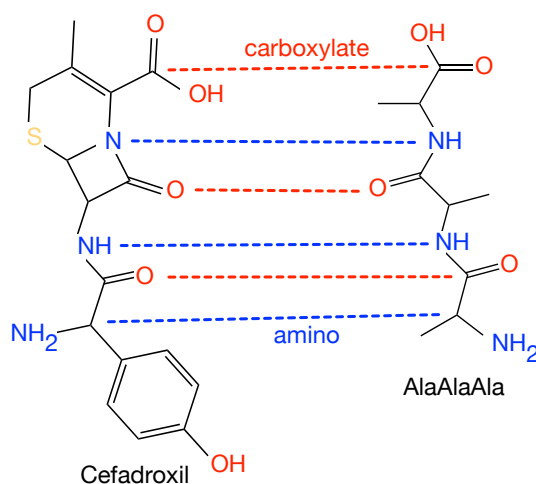
drug disposition (e.g. emtricitabine used in HIV therapy) into the blood vessels (Sodani and Patel, 2012).

### 1.3.1 PepT1 and PepT2 mediated drug uptake

Given the substrate polyspecificity prevalent within the POT family, it is not surprising that PepT1 can also play a role in drug absorption along the lining of the gut. The key to being recognised by this transporter is having certain structural features similar to either di- or tripeptides. The best known examples of PepT1 drug substrates are the  $\beta$ -lactam antibiotics. As early as in the 1990's, evidence of cephalosporin  $\beta$ -lactam transport by PepT1 surfaced via a series of experiments measuring proton dependent uptake of cephadrine, cefadroxil and cephalexin in whole cell assays over-expressing PepT1 (Kato et al., 1989; Thwaites et al., 1994; Fei et al., 1994; Boll et al., 1994). This observation potentially stems from the peptide mimetic properties of the cephalosporins, i.e. the carboxylate and amino ends, and a peptide bond linking the lactam ring and the side chain (Figure 1.15). Subsequent studies provided a better understanding of the preference PepT1 has for a range of these antibiotics. For example, Wenzel et al. (1996) showed that zwitterionic  $\beta$ -lactams (cefadroxil, amoxicillin) are the best substrates, while their anionic counterparts (benzylpenicillin, cefixime) are less preferred. Bretschneider et al. (1999) assayed 23 different  $\beta$ -lactams and demonstrated a good correlation between affinity (as measured by inhibition of glysar transport) and bioavailability (as measured by transepithelial flux). Similarly, its renal isoform, PepT2, has also been established as a drug carrier. As has been shown with di- and tripeptides, PepT2 possesses a significantly higher affinity for the  $\beta$ -lactam antibiotics compared to PepT1, albeit with the overall rank ordering preserved (Luckner and Brandsch, 2005).

In addition to the  $\beta$ -lactam antibiotics, PepT1 and PepT2 also transport the tumor suppressors bestatin (Inui et al., 1992), floxuridine (Tsume et al., 2008) and gemcitabine (Song et al., 2005), the imaging substance  $\delta$ -aminolevulinic acid (Döring et al., 1998), the antiviral agents valacyclovir (Ganapathy et al., 1998) and valganciclovir (Sugawara et al., 2000) and the anti-hypotensive drug midodrine (Tsuda et al., 2006). Some of these compounds are actually prodrugs, which are created by attaching an amino acid or a dipeptide to the active core drug, essentially to mimic a dipeptide or a tripeptide, respectively, to gain affinity to the POT transporters. *In vivo*, the active form of the drug is released by an enzymatic hydrolysis using esterases that cleaves off the amino acid or dipeptide. This approach is particularly useful to enhance the bioavailability of drugs that are not naturally a substrate of PepT1 or PepT2. For example, attaching a phenylalanine residue to L-dopa, which is widely used to





Figs. 1.15: Structural comparison of the  $\beta$ -lactam antibiotic cefadroxil and the tripeptide triAla, highlighting similarities including the carboxylate and amino groups.

alleviate the symptoms of Parkinson's disease, enhanced its intestinal absorption by almost 40-fold (Tamai et al., 1998). As PepT1 in the small intestine controls the entry of these substances into the body and PepT2 in the kidney regulates their retention in the systemic circulation, understanding the molecular basis of interactions between these drugs and the POT transporters is of paramount importance for drug design to improve oral drug delivery.

## 1.4 Aim of this thesis

Physiologically, PepT1 and PepT2 are not essential. In fact, multiple studies with knock-out mice reported no pathological phenotypes despite the reduced peptide uptake (Rubio-Aliaga et al., 2003; Nässl and Rubio-Aliaga, 2011). The motivation in studying these transporters comes from their ability and potential to improve oral bioavailability. The wealth of data, for example the availability of crystal structures from multiple bacterial homologs and the vast number of transport assays, make PepT1 and PepT2 suitable candidates for more systematic and rational design of drugs in the future.

The principal aim of this thesis is to predict substrate transport by elucidating the molecular mechanism of interactions between the POT transporters and their substrates, i.e. peptides and, most importantly, drugs. The main approach is to employ molecular dynamic simulations and a range of free energy calculation methods, alongside verification from experimental data in the form of transport assays, to characterise the binding affinities of various ligands to the transporters and provide structural justifications of their substrate preference. To achieve this aim, this thesis is divided into the five following chapters:

- **Chapter 2** Recapitulate the theory behind molecular dynamics simulation and free energy calculations.
- **Chapter 3** Establish which methods can be used to accurately predict the binding affinities of ligands using a model POT transporter from bacteria.
- **Chapter 4** Perform an in-depth analysis of the substrate preference of the model bacterial POT transporter using these verified methods.
- **Chapter 5** Build a homology model of human PepT1 and PepT2 using a bacterial POT as the template.
- **Chapter 6** Apply the prediction methods to PepT1 homology model to understand drug interactions with the transporter.

# Chapter 2

## Theory and methods

In this thesis, computational studies of peptide transporters are conducted primarily from the perspective of protein-ligand interactions using two principal methods: i) atomistic molecular dynamics simulations and ii) free energy calculations. Molecular dynamics are used to investigate how the binding site of the transporter adapts to various ligands, which is the key to understanding the structural origin of the promiscuity prevalent in this protein family. The calculation of binding affinities of peptides and drugs to the transporter are performed using methods encompassing a vast range of accuracy and computational requirements. Other methods utilised in this research include homology modelling to build a model structure for human PepT1 and protein-protein docking to propose how trypsin binds to its extracellular domain. This chapter will recapitulate the theories behind these computational techniques and describe how they are used in the context of the research.

### 2.1 Molecular dynamics simulation

Molecular dynamics (MD) simulations are capable of providing a detailed view of how atoms interact with each other and, on a larger scale, how biomolecules behave in their natural environment. Thus MD is sometimes referred to as *computational microscope* (Lee et al., 2009). Classical MD computes the motion of a set of interacting particles based on Newton's second law of motion, which states that force equals the rate of change of momentum or  $\mathbf{F}_i = m_i \mathbf{a}_i$ . In a simple system, the force acting on each particle can be

determined by solving the second order differential equation:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \quad (2.1)$$

However, the force acting on each particle changes not only when its position is altered, but also when other surrounding particles with which it interacts move, resulting in intricately coupled motions within the system. This phenomenon gives rise to the *many-body problem*, whereby the complex dependency of particles in a system makes it impossible to solve the equation of motion analytically for all the particles. This issue is especially aggravated in a large system, for example in a biomolecular simulation, wherein lies tens of thousands of atoms. To circumvent this problem, the equation of motion is solved numerically using a *finite difference* method.

### 2.1.1 Numerical integration for the equations of motion

The finite difference method relies on approximating the positions,  $\mathbf{r}$ , and dynamic properties such as velocities,  $\mathbf{v}$ , and acceleration,  $\mathbf{a}$ , as Taylor series expansion:

$$\begin{aligned} \mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \frac{\delta t}{1!} \mathbf{r}'(t) + \frac{\delta t^2}{2!} \mathbf{r}''(t) \dots \\ \mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{\delta t^2}{2} \mathbf{a}(t) \dots \end{aligned} \quad (2.2)$$

The equivalent equation for  $\mathbf{r}$  at time  $(t - \delta t)$  is:

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{\delta t^2}{2} \mathbf{a}(t) \dots \quad (2.3)$$

Adding and subtracting, respectively, Equation 2.2 and Equation 2.3 will give us expressions for  $\mathbf{r}(t + \delta t)$  and  $\mathbf{v}(t)$ :

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t) \quad (2.4)$$

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t} \quad (2.5)$$

This algorithm, however, suffers from numerical inaccuracies as the calculations involve differences between large similar numbers, that are  $2\mathbf{r}(t) - \mathbf{r}(t - \delta t)$  in Equation 2.4 and  $\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)$  in Equation 2.5. The most widely implemented approach in GROMACS is the *leap-frog* algorithm (Hockney, 1970), which does not have this disadvantage. The expressions for positions and velocities are given by:

$$\mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t - \frac{1}{2}\delta t) + \frac{\delta t}{m}\mathbf{F}(t) \quad (2.6)$$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t + \frac{1}{2}\delta t) \quad (2.7)$$

The algorithm utilises two dynamic properties at different time points: the position  $\mathbf{r}$  at time  $t$ , and the velocity  $\mathbf{v}$  at time  $t - \frac{1}{2}\delta t$ . Using forces  $\mathbf{F}$  at time  $t$ , from which the acceleration  $\mathbf{a}$  can be determined, it first calculates the velocity at time  $t + \frac{1}{2}\delta t$  (Equation 2.6). Then this information is used to derive the position at time  $t + \delta t$  (Equation 2.7). As such, the velocities and positions are always deduced in a staggered interval of  $\frac{1}{2}\delta t$ , thus “leap-frogging” over each other. One disadvantage of this approach is that the positions and velocities are not synchronised, therefore it is not possible to determine the kinetic energy corresponding to the defined coordinates, from which potential energy is calculated.

### 2.1.2 Molecular mechanics force field

The force  $\mathbf{F}_i$  acting on a particle  $i$  can be determined by differentiating the potential energy,  $U$ :

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}_i)}{\partial \mathbf{r}_i} \quad (2.8)$$

This potential energy is described by a physical model—the force field—that defines each particle in the system being simulated. Atomistic force fields describe each atom as a single particle\*. Molecular mechanics force fields use classical mechanics to describe the potential energy function of a set of particles based on their position in space,  $\mathbf{r}$ . This potential energy is generally modelled as a sum of the intra- and intermolecular forces, respectively forming the bonded and non-bonded energy components. The former is divided into three constituents: i) bond length, ii) bond angle, and iii) bond dihedral, whilst the latter

---

\*A variation of this model is called the *united-atom* approach, which groups together hydrogens (only the ones that are not capable of forming hydrogen bonds) with heavy atoms, for example in a methyl or methylene group, into a single bead to reduce the total number of particles in the system. An example of a united-atom biomolecular force field is GROMOS (Oostenbrink et al., 2004).

is made of two parts: i) electrostatic, and ii) van der Waals (Equation 2.9 and Figure 2.1). The bonded energy terms describe the interactions between atoms that are linked by at most three covalent bonds and complementarily, the non bonded energy terms model interactions between atoms on different molecules or on the same molecule but more than three covalent bonds apart from each other<sup>†</sup>.

$$U(\mathbf{r}) = U_{\text{bonds}}(\mathbf{r}) + U_{\text{angle}}(\mathbf{r}) + U_{\text{torsion}}(\mathbf{r}) + U_{\text{electrostatics}}(\mathbf{r}) + U_{\text{vdW}}(\mathbf{r}) \quad (2.9)$$

In this thesis, the Amber biomolecular force field will be used, specifically the ff99SB version with improved torsion potentials from Lindorff-Larsen et al. (2010). The functional form of the force field is given by Equation 2.10.

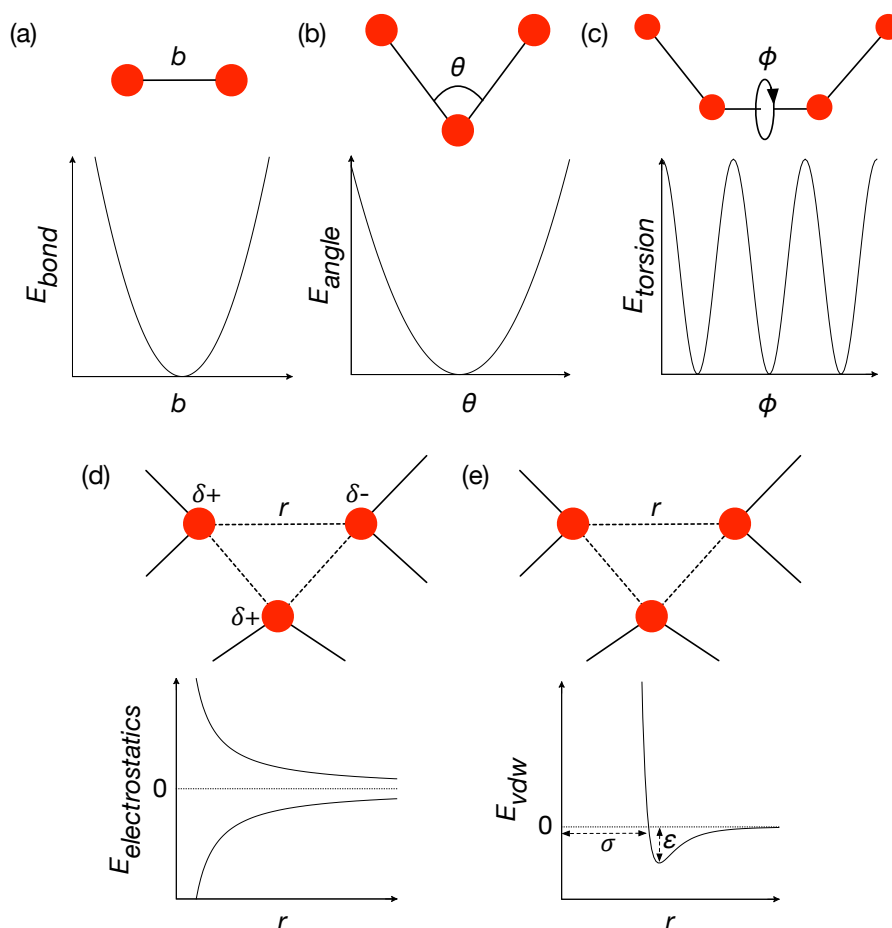
$$U(\mathbf{r}) = \overbrace{\sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}^{\text{Bonded}} + \underbrace{\sum_{i < j} \epsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i < j} \frac{q_i q_j}{\epsilon_0 r_{ij}}}_{\text{Non-bonded}} \quad (2.10)$$

### Bonded interactions

The bond length term, the first term in Equation 2.10, describes how much a bond,  $b$ , can stretch with respect to a reference value,  $b_0$ , corresponding to the length adopted when all other terms in the force field are set to zero (Figure 2.1(a)). For simplicity, this is modelled using a simple harmonic potential with a force constant of  $K_b$  that calculates the energetic penalty sustained by the bond if it deviates from the reference length. The second term in Equation 2.10 denotes the bond angle energy, which measures the energy associated with the bending of two adjacent bonds as described by the angle between them,  $\theta$  (Figure 2.1(b)). Again, a harmonic potential is employed to characterise this energy, whereby a force constant  $K_\theta$  and a reference value  $\theta_0$  are included in the calculation. Compared to other components of the force field, a substantial amount of energy is needed to distort bond lengths and angles, and as such, structural variations during a simulation are usually contributed from the torsional and non-bonded terms.

The bond dihedral or torsional energy, outlined by the third term in Equation 2.10, takes

<sup>†</sup>In some force fields, the intermolecular interactions between an atom and its third covalently bonded neighbours ( $i$  and  $i + 3$ ) are reduced or scaled to avoid strong repulsion, which could deform the molecule.



Figs. 2.1: A schematic representation of the components of a molecular mechanics force field and their energy functions: (a) bond length, (b) bond angle, (c) bond torsion or dihedral, (d) electrostatics interactions and (e) van der Waals interaction. The functional form of the energy function is given by Equation 2.10. Figure adapted from Levitt (2001).

into account three adjacent bonds and computes the ability of the middle bond to rotate with respect to the other two flanking bonds (Figure 2.1(c)). The distribution of rotational conformations can be modelled using a cosine function where  $V_n$  represents the height of the energetic barrier for a rotation to happen,  $\gamma$  denotes the position of energy minimum and  $n$  depicts the number of minima available as the bond rotates through  $360^\circ$  (also known as the multiplicity). For example, if a simple system of ethane is considered, there are nine ( $3^2$ ) identical torsions of H-C-C-H that reach an energetic minimum in a staggered conformation and an energetic maximum in an eclipsed conformation. If the position of the first energy-minimum staggered conformation, or  $\gamma$  is set to zero, and the bond is rotated through  $360^\circ$ , there will be a total of three minima ( $n=3$ ) at  $0^\circ$ ,  $120^\circ$  and  $240^\circ$ . Correspondingly, the energy-maximum eclipsed structures will be observed three times at  $60^\circ$ ,  $180^\circ$  and  $300^\circ$ . Improper dihedrals are introduced to cater for aromatic molecules with an electron-rich  $\pi$  system. In such molecules, some peripheral atoms, for example the oxygen atom on the side chain of tyrosine, must be kept in the same plane as the aromatic ring for an optimum configuration. To this end, “improper” cosine functions are added; they are similar to the previously described dihedral angles, except that the four atoms under consideration are not linearly connected.

### Non-bonded interactions

The last two terms in Equation 2.10 refer to the non-bonded potential energy which sums the interactions between all atom pairs in the system ( $i$  and  $j$ ) based on their inter-atomic distances,  $r_{ij}$ . The first of these terms represents the van der Waals potential, which is modelled by a Lennard-Jones 12-6 function with  $\sigma$  denoting the distance for which the potential energy is zero and  $\varepsilon$  depicting the depth of the energy well (Figure 2.1(e)). The van der Waals interactions are made of two components, i.e. attractive and repulsive forces. The former is caused by the dispersive forces that arise due to the fluctuations in the electron clouds; the formation of an instantaneous dipole on one molecule induces dipoles on adjacent molecules. This component is calculated by the inverse sixth power function of distance ( $r^{-6}$ ). The repulsive force, on the other hand, originates from the Pauli exclusion principle, which states that no two particles in a system can have the same set of quantum numbers, thus prohibiting atoms from occupying the same region of space. The contribution of this force is computed using the inverse twelfth power function of distance ( $r^{-12}$ )<sup>‡</sup>. The

<sup>‡</sup>The inverse twelfth power function describing the repulsive part of the Lennard-Jones potential is not directly related to any theoretical arguments. In fact, quantum mechanic calculations suggest an exponential function, as used in Buckingham potential. However, the Lennard-Jones potential is more widely used for



Lennard-Jones potential is short-ranged and rapidly decreases with increasing inter-atomic distance. In MD simulations, therefore, its contribution is often calculated only within a specified radius, the *cutoff distance*, beyond which the potential is assumed to have converged to zero.

The second component of the non-bonded interactions is the electrostatic potential, which is described using Coulomb's law with partial atomic charges of  $q_i$  or  $q_j$  and the permittivity of free space,  $\epsilon_0$  ( $8.854 \times 10^{-12} \text{ F m}^{-1}$ ) (Figure 2.1(d)). In a molecule, some atoms are more electronegative than others and thus attract more electrons, creating an unequal distribution of charge (permanent dipole) that are represented by partial atomic charges. In the Amber force field, these partial charges are obtained via a semi-empirical approach called the restrained electrostatic potential (RESP), which fits point charges onto atoms to reproduce the quantum mechanically determined electrostatic potential at the surface of a molecule (Bayly et al., 1993). Unlike the van der Waals potential, the electrostatic potential is reciprocal to the first power function of distance ( $r^{-1}$ ), as a result of which the energy decays much more slowly with respect to increasing distance. A similar cutoff approach used for van der Waals potential, therefore, cannot be applied here. Instead, the calculation of electrostatic forces can be accelerated using the Ewald sum technique, which shall be briefly described in §2.1.5. Even so, calculating the electrostatic interactions remains the most computationally demanding step of any MD simulation.

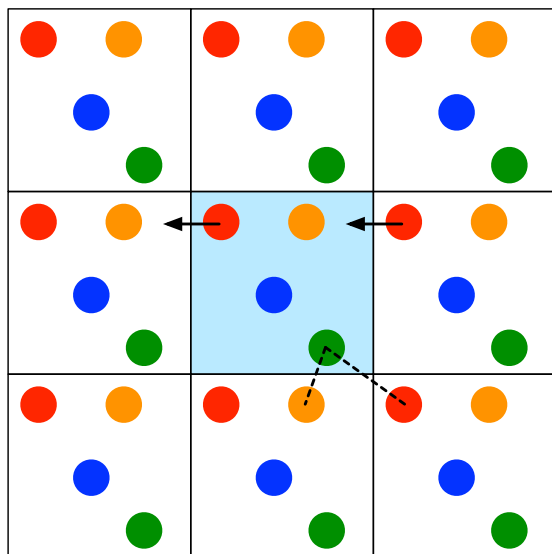
### 2.1.3 Constrained dynamics

The size of the time step,  $\delta t$ , is limited by the fastest motion in the system. For a biomolecular simulation, this is represented by the vibration of hydrogens, H, attached to heavy atoms, X. By keeping the X-H bond length constant, the value of  $\delta t$  can be increased, therefore the integrator can advance more quickly. This is typically performed using constrained dynamics algorithms<sup>§</sup>. The default constraints imposed in GROMACS is the LINCS algorithm (Hess et al., 1997). Briefly, this algorithm resets the intramolecular bond lengths involving hydrogen atoms to its original value after each integration step. Numerical integration is first used to update the position of hydrogens without any constraints. Then, the algorithm calculates the amount of force required to correct any deviations and applies it along the axis of the bonds, such that the bond length will return to the starting value. This approach

---

ease of computation as  $r^{-12}$  can be rapidly calculated by squaring the  $r^{-6}$  term.

<sup>§</sup>Another way to accelerate the calculation is by using the multiple time stepping approach, where a smaller  $\delta t$  is used to account for faster motions like the vibration of hydrogen, and a larger  $\delta t$  is used for other calculations (Humphreys et al., 1994).

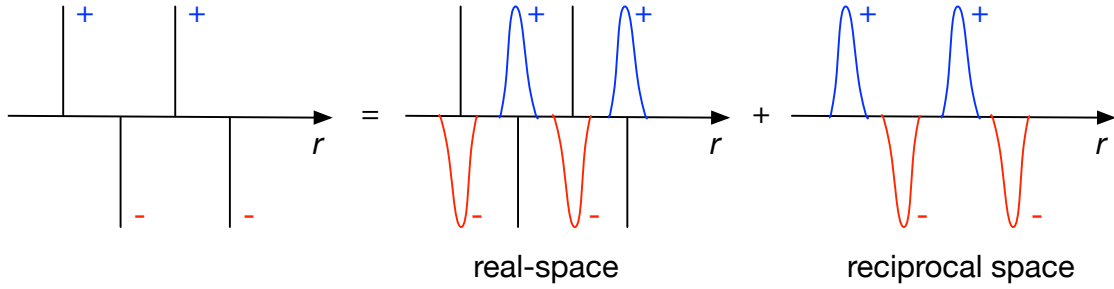


Figs. 2.2: A two dimensional schematic representation of the periodic boundary conditions. The central unit cell is colour in light blue for clarity. Four particles exist in the simulation box; the red particle exits the box from the left side and re-enters from the right side, as illustrated by the arrow. Based on the minimum image convention, the intermolecular forces acting on the green particles from the red and orange ones are indicated by the dashed lines.

allows for a  $\delta t$  of 2 fs, as is used in this thesis, compared to 1 fs without the constraints.

#### 2.1.4 Periodic boundary conditions

To ensure bulk behaviours are not lost in a simulation of a system with a limited number of particles, the simulation box is replicated in all directions following the periodic boundary conditions. This technique produces a periodic array of infinite copies of the simulation box. For example, if the system has a cubic simulation box, in two dimensions, this box will be surrounded by eight identical boxes on all sides (Figure 2.2), whereas in three dimensions, 26 boxes. A particle leaving the simulation box from one side will “reappear” from the opposite side with the same velocity, hence the total number of particles in the central box stays constant. Despite the infinitely repeated boxes, the *minimum image convention* is applied to the calculation of short-ranged intermolecular interactions—each particle “sees” only one image of every other particle in the system, that is the closest image either from the same box or from an adjacent periodic box. Also, the periodic boundary conditions are not capable of reproducing structural characteristics on a scale larger than the unit cell, for example long-wavelength membrane undulations, as the repetitious nature of the periodic boxes suppresses such fluctuations. GROMACS implements various different box shapes that can adhere to the periodic boundary conditions like cube, rhombic dodecahedron and truncated octahedron. The former represents the most widely used unit cell, whilst the latter two are particularly beneficial in the simulation of a spherical system, as the shape is closer to a sphere, therefore less water and ions are required for solvation. In this thesis, as a



Figs. 2.3: A schematic representation of the Ewald summation used to compute electrostatic potential. A Gaussian charge distribution is placed on top of each initial set of charges (calculated in real space), which is counter-balanced by an additional charge distribution of opposite sign (calculated in reciprocal space).

system with a planar membrane bilayer is simulated, a cubic box will be used for all the simulations.

### 2.1.5 The particle-mesh Ewald Method

As discussed in §2.1.2, unlike the van der Waals potential, the electrostatic potential converges to zero significantly less quickly as it varies inversely with  $r$ , therefore the interactions at a long distance is far from negligible. As such, introducing a cutoff radii scheme as applied to the van der Waals potential will generate spuriously truncated forces. The long-ranged nature of electrostatic potential means all charged atoms in a system can interact not only with each other, but also with their images in periodic cells because the forces extend well beyond half the size of simulation box. The number of pairwise forces that need to be computed, therefore, scales with the number of particles,  $N$ , by a factor of  $O(N^2)$ , inevitably making explicit direct calculation prohibitively expensive.

The total electrostatic energy contributed by charge-charge interactions in the central box and between the central box and its periodic images is given by:

$$U_{el} = \frac{1}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i \cdot q_j}{4\pi\epsilon_0 \mathbf{r}_{ij,\mathbf{n}}} \quad (2.11)$$

where  $\mathbf{n} = (n_x, n_y, n_z)$  is the periodic box index vector while the asterisk denotes that whenever  $i = j$ , the interactions is calculated only when  $i$  and  $j$  are periodic image of each other. The distance indicated by  $\mathbf{r}_{ij}$  is the real distance between the particles instead of the minimum image distance. The summation in Equation 2.11, however, does not converge. The

basis of Ewald summation (Ewald, 1921) is reformulating the summation into two series, each of which converges significantly more rapidly:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r} \quad (2.12)$$

where  $f(r)$  represents the introduction of new functions that converts the original sum into two quickly-converging terms. In practice, the Ewald summation surrounds each atomic charges with a neutralising Gaussian charge distribution of equal magnitude but of opposite sign, represented by the first term in Equation 2.12. To correct for this addition, a second term is added by virtue of a cancelling charge distribution, which is calculated in reciprocal space using Fourier transforms (Figure 2.3). Also, a third term is added to subtract the interaction of each Gaussian with itself. Despite the Ewald sum as a whole converging as  $O(N^{\frac{3}{2}})$ , the reciprocal term still scales as  $O(N^2)$  and therefore the method is still computationally expensive to implement. To improve upon this method, the particle-mesh Ewald (PME) method (Darden et al., 1993; Essmann et al., 1995) interpolates each atomic charge onto a three-dimensional grid, which allows for a Fast Fourier Transform (FFT) technique to be applied to compute the summation in reciprocal space. This algorithm scales as  $O(N \log N)$ , considerably reducing the computational cost of calculating electrostatic potential. Therefore, the PME method shall be used to compute electrostatic forces throughout this thesis.

### 2.1.6 Temperature and pressure control

So far all the simulations discussed have a constant number of particles at a condition of constant volume and energy, which is referred to as the NVE or *microcanonical* ensemble. Most biological phenomena, however, occur under constant temperature and pressure (the NPT or *isothermal-isobaric* ensemble). Another common alternative ensemble is the NVT or *canonical* ensemble, for which the volume the simulation box is kept constant instead of pressure. Methods that maintain a constant temperature are called a *thermostat* while methods that maintain a constant pressure are called a *barostat*.

#### Thermostat

GROMACS implements three coupling schemes, from the weaker Berendsen thermostat (Berendsen et al., 1984) to the more rigorous velocity rescaling (Bussi et al., 2007) and Nose-Hoover (Nosé, 1984; Hoover, 1985) schemes. The Berendsen algorithm couples the simulation unit cell to a large water bath, to which thermal energy is dissipated and from

which energy is supplied, respectively, following an increase and decrease in temperature. This scheme scales the velocities of all particles such that the rate of change of temperature is proportional to the difference in measured temperatures of the system and the water bath (Equation 2.13). The coupling parameter,  $\tau_T$ , determines how fast this scheme works and thus how tightly coupled the bath and the system are. This thermostat suppresses the fluctuation of the kinetic energy and therefore leads to an energetically incorrect sampling.

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_{bath} - T(t)) \quad (2.13)$$

A similar thermostat that generates a rigorous canonical sampling is the velocity rescaling scheme. This thermostat is essentially a Berendsen thermostat with an additional stochastic term introduced to ensure a correct kinetic energy distribution:

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}} \quad (2.14)$$

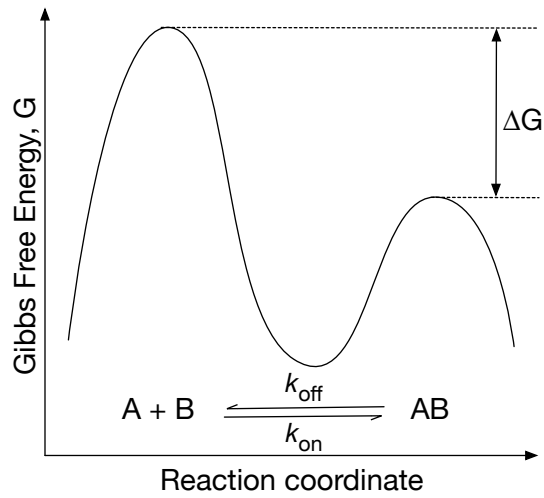
where  $K$  is the kinetic energy,  $N_f$  is the number of degrees of freedom and  $dW$  a Wiener process (a continuous-time stochastic process). No additional parameter is required except for a random seed. Apart from producing a correct canonical ensemble, this thermostat retains the advantage of a Berendsen thermostat, i.e. first order decay of temperature deviations with no oscillations. This thermostat therefore shall be used for most of the simulations performed in this thesis. To prevent differences in temperature arising between the different components of the simulated system (the hot solvent, cold solute phenomenon), which is prevalent in velocity-based rescaling methods, group temperature coupling is applied, whereby all protein, membrane and solvent atoms are coupled to independent water baths.

### Barostat

Similarly, the pressure of the simulation unit cell can also be coupled to a “pressure bath” to maintain a constant value. The rate of change of pressure is given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_{bath} - P(t)) \quad (2.15)$$

where  $\tau_P$  is the coupling constant,  $P_{bath}$  and  $P(t)$  are the pressure of the bath and the pressure of the simulation box at time  $t$ , respectively. In this thesis, the Parrinello-Rahman barostat (Parrinello and Rahman, 1981) shall be used as it is more rigorous and capable of sampling from true NPT ensemble. Simulations with a membrane bilayer are coupled *semi-isotropically*, which means the  $x$  and  $y$  dimensions (the membrane plane) are rescaled



Figs. 2.4: The fundamental concept of free energy of binding,  $\Delta G$ . Two units, A and B, bind to form a complex AB with lower free energy.  $\Delta G$  represents the free energy difference between the bound and unbound state and corresponds to the strength of binding.  $k_{\text{on}}$  and  $k_{\text{off}}$  are the rate of association and dissociation, respectively.

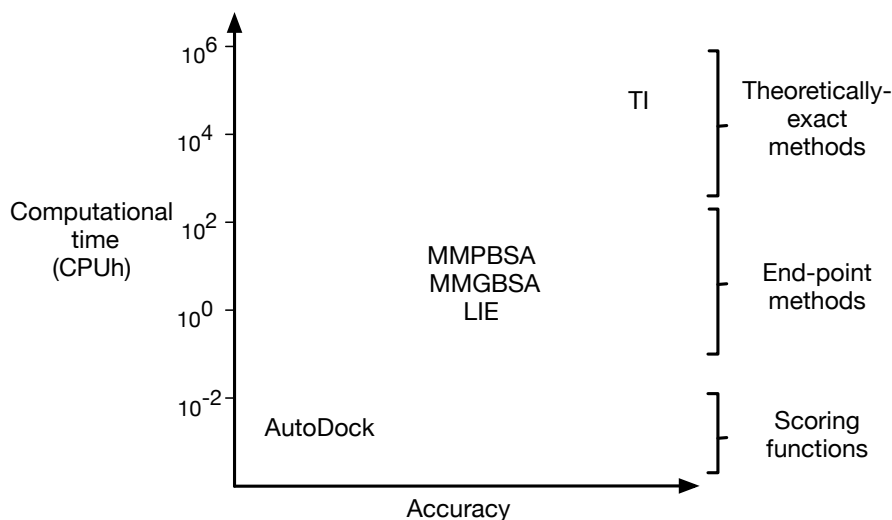
together, independent of the  $z$  dimension (the membrane normal). This approach allows for a proper lateral pressure profile in the bilayer.

## 2.2 Free energy of binding

Gibbs free energy,  $G$ , describes the amount of useful work that can be extracted from a thermodynamically closed system at a constant temperature and pressure (NPT). The Gibbs free energy of binding,  $\Delta G$ , represents the strength of binding of two bodies, in which case if the free energy of the complex is lower than its constituents, i.e.  $\Delta G < 0$ , the binding is favourable (Figure 2.4). Measuring  $\Delta G$  is essential to study the physical nature of interactions between two molecules. At chemical equilibrium, the concentration of both unbound (A and B) and bound (AB) states remains constant and the position of the equilibrium is quantified by the equilibrium constant,  $K$  as described by Equation 2.16.  $\Delta G$  is related to  $K$  by Equation 2.17, where  $R$  is the gas constant ( $8.314 \text{ J K}^{-1} \text{ mol}^{-1}$ ) and  $T$  is temperature (K). A favourable binding designated by a negative  $\Delta G$  corresponds to a large value of  $K$ , which means the concentration of AB is higher than the product concentrations of its subunits, and as such A and B prefer to bind to each other than to be separated.

$$K = \frac{[AB]}{[A] \cdot [B]} \quad (2.16)$$

$$\Delta G = -RT \ln K \quad (2.17)$$



Figs. 2.5: Available computational methods to calculate the free energy of binding ( $\Delta G$ ) arranged according to their accuracy and estimated computational expenses.

## 2.3 Computational predictions of binding energy

Computer modelling has an enormous potential for the study of biomolecular interactions. By using the three-dimensional structure of a target protein, either from crystallography or homology modelling, with its ligand of interest, one can obtain not only the binding free energy but also detailed insights into the molecular interactions. In this section, the available methods to calculate the free energy of binding between a protein and a small-molecule ligand will be reviewed (Figure 2.5), focusing on the accuracy of each method, their sources of errors and the amount of computational resources required to perform each calculation.

### 2.3.1 Scoring Functions

Molecular docking programs were developed to predict the way two biomolecules bind to each other and subsequently estimate binding affinities for the proposed binding orientations (see reviews by Bello et al. (2013) and Yuriev and Ramsland (2013)). The method is generally divided into two components: a search algorithm and a scoring function. The former suggests several optimum binding poses, usually based on spatial complementarity, and the latter ranks these poses based on their affinity scores. Scoring functions are fast and highly-simplified mathematical models. For the purpose of this thesis, the  $\Delta G$  predictions will be performed using the AutoDock Vina docking program (Trott and Olson, 2010), which

has an in-house scoring function that is largely inspired by X-score (Wang et al., 2002). This function sums the intermolecular and intramolecular contributions of all pair of atoms within three covalent bonds from each other according to a general function:

$$\Delta G = \sum_{i < j} f_{t_i t_j}(r_{ij}) \quad (2.18)$$

where  $t_x$  is the assigned atom type for  $x$  following the X-Score atom-typing,  $f_{t_i t_j}$  denotes the interaction functions and  $r_{ij}$  represents the interatomic distance. The interaction functions,  $f_{t_i t_j}$  consist of equations describing steric interactions, the hydrophobic effect, hydrogen bonds and deformation effects. Each of these equations is multiplied by an empirical constant, identical for all atom pairs, derived by regression analyses using a training set from PDBbind (Wang et al., 2004, 2005). Most of these terms are highly simplified, for example, the hydrophobic contribution equals 1 when  $r_{ij} < 0.5 \text{ \AA}$  and 0 when  $r_{ij} > 1.5 \text{ \AA}$ , and linearly interpolated for when  $0.5 \text{ \AA} \leq r_{ij} \leq 1.5 \text{ \AA}$ .

Due to the simplistic nature of these calculations, estimating  $\Delta G$  using scoring functions is extremely fast. On a quad core machine with Intel<sup>®</sup> Xeon<sup>®</sup> processors, docking a small peptide onto the bacterial POT transporter, PepT<sub>St</sub>, and estimating  $\Delta G$  for all proposed binding orientations take only around 30 seconds. This method therefore is the usual choice to perform a high-throughput virtual screening involving a large library of ligands. For example, Schlessinger et al. (2011) computationally screened more than 6,000 small molecules in the Kyoto Encyclopedia of Genes and Genomes (KEGG) Drug Database (Ogata et al., 1999) using the DOCK docking procedure (Ewing et al., 2001) and found chemically novel substrates for the norepinephrine transporter (NET), which could potentially be developed into antidepressants and psychostimulants. Chen and Shoichet (2009) used the same docking protocol to screen all commercially available fragment libraries containing more than 300,000 molecules and discover novel  $\mu\text{M}$ -range inhibitors against a  $\beta$ -lactamase. These indicate that scoring functions are adequately reliable at isolating potential binders of a protein from a large database of compounds with a wide range of structural and chemical properties. Here, the affinity scores are used only to filter out low-scoring non-binders and therefore the individual  $\Delta G$  value *per se* is not important.

However, comparing these predicted  $\Delta G$  values to experimental data has always been problematic. A comparative assessment of all major scoring functions showed that the maximum correlation coefficient between their binding scores and experimental data was only around 0.5 with the best scoring functions, X-score and ChemScore, correctly ranking protein-ligand complexes 50% of the time (Cheng et al., 2009). The typical root mean



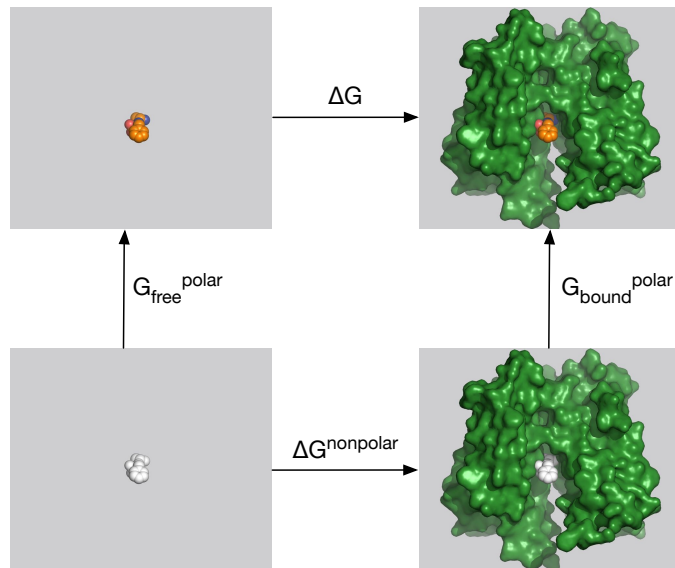
square error (RMSE) achieved by scoring functions is usually around 2 kcal/mol (Wang et al., 2002). One source of errors in scoring functions is the empirical constants derived from a training set of a limited number of protein-ligand complexes. Using the same weighting factors for proteins from a different family or with ligands of different chemical properties often yields questionable results (Ross et al., 2013). Scoring functions also neglect any solvation effect, which can be central to ligand binding as water molecules can be a part of the interaction interface. Indeed, including crystallographic waters during docking protocols does significantly improve accuracy (Thilagavathi and Mancera, 2010; Santos et al., 2010). To reduce computational cost, most docking programs by default consider the protein as a rigid body and the ligand as flexible, following the traditional lock-and-key binding model (Fischer, 1894). However, some methods (Totrov and Abagyan, 1997; Zavodszky and Kuhn, 2005) allows users to specify residues in the binding site, for which the side chains are made flexible. Still, these protocols omit the motion of the protein backbone, which can be a crucial component during ligand recognition, especially within a promiscuous binding site that undergoes large conformational changes to accommodate various ligands. To include this flexibility, MD simulations have to be employed.

### 2.3.2 End-point free energy methods

As the name suggests, the end-point methods require simulations of the end points of a binding event, i.e. the unbound states (free protein and ligand) and the bound state (protein-ligand complex). These methods provide a compromise between speed and accuracy as they are faster than the theoretically-rigorous free energy calculations that require simulations of intermediate states, and yet are more accurate than scoring functions as they take into account protein-ligand dynamics. An ensemble of configurations is generated by MD simulations and a thermodynamic average is calculated from them. Three examples of end-point methods will be discussed, largely following reviews from Guitierrez-de Teran and Aqvist (2012), Srinivasan et al. (1998) and Miller III et al. (2012).

#### Linear Interaction Energy (LIE)

LIE considers ligand binding as a partition problem: a ligand in solvent,  $l$  (free state), is transferred into a water-solvated binding site of a protein (bound state) and the total change in free energy associated with the ligand is calculated (Aqvist et al., 1994). The fundamental equation is derived from a thermodynamic cycle (Figure 2.6), whereby the transfer process involves i) stripping from the ligand any intermolecular electrostatic interactions ii)



Figs. 2.6: The thermodynamic cycle used to calculate  $\Delta G$  using the LIE method. Protein (PepT<sub>St</sub>) is shown in green surface representation and ligand (AF) in orange sphere representation. Ligand without intermolecular electrostatic interactions is depicted in white.

ligand binding where only non-polar interactions are present and iii) restoring the electrostatic components of the ligand. The free energy of binding,  $\Delta G$ , can be estimated simply as the sum of polar and non-polar energy changes between the two states of the ligand (Equation 2.19). This method takes advantage of the way an MD simulation is performed, whereby the force field similarly separate the polar and non-polar components of the intermolecular interactions into Coulomb (el) and Lennard-Jones (vdw) potentials, respectively (Equation 2.9).

$$\begin{aligned}\Delta G(l) &= \Delta G^{\text{nonpolar}} + \Delta G^{\text{polar}} \\ &= \alpha \cdot (U_{\text{bound}}^{\text{vdw}} - U_{\text{unbound}}^{\text{vdw}}) + \beta \cdot (U_{\text{bound}}^{\text{el}} - U_{\text{unbound}}^{\text{el}})\end{aligned}\quad (2.19)$$

To convert these potential energies,  $\Delta U$  into free energies,  $\Delta G$ , each  $\Delta U$  term is factored by an empirical constant. The original value of  $\beta$  was set to 0.5 based on the linear response theory for electrostatic forces (Warshel and Russell, 1984). As this theory was first derived for solvated ions, ligand molecules which display more chemical variety, have shown a slight divergence. For dipolar compounds, the  $\beta$  value varies according to the number of hydroxyl groups present in the compound, due to their ability to interact with the hydrogen bond network of the solvent (Equation 2.20) (Carlson and Jorgensen, 1995; Aqvist and

Hansson, 1996; Hansson et al., 1998). Further refinement by Almlöf et al. (2007) considered other chemical groups like amides, amines, and carboxylic acids, and compute the  $\beta$  values for compounds containing a mixture of these chemical groups.

$$\beta = \begin{cases} 0.5 & \text{if } q \neq 0 \\ 0.43 & \text{if } q = 0, n\text{OH} = 0 \\ 0.37 & \text{if } q = 0, n\text{OH} = 1 \\ 0.33 & \text{if } q = 0, n\text{OH} \geq 2 \end{cases} \quad (2.20)$$

The non-polar scaling constant,  $\alpha$ , was first proposed to be 0.18, as derived empirically using a training set of 18 molecules (Aqvist et al., 1994). This has since been successfully reproduced for various protein-ligand systems, e.g. neutral ligands of cytochrome P450cam (Almlöf et al., 2004), charged potassium channel blockers (Osterberg and Aqvist, 2005; Luzhkov and Aqvist, 2001) and large peptide-mimetics inhibitors of aspartate protease plas-mepsins (Bjelic et al., 2007). A study to further refine the value of  $\alpha$  suggests a correlation to the hydrophobicity of the binding site (Wang et al., 1999). More recently, an “adapted” LIE method was introduced by Linder et al. (2013), where both  $\alpha$  and  $\beta$  coefficients are fitted based on system-dependent polarity descriptors in order to increase the accuracy of predictions. Additionally, an offset parameter,  $\gamma$ , is often added to Equation 2.19 to fit the predicted  $\Delta G$  values to experimental data, although without  $\gamma$ , the method can still predict relative binding affinities or ranking of ligands.

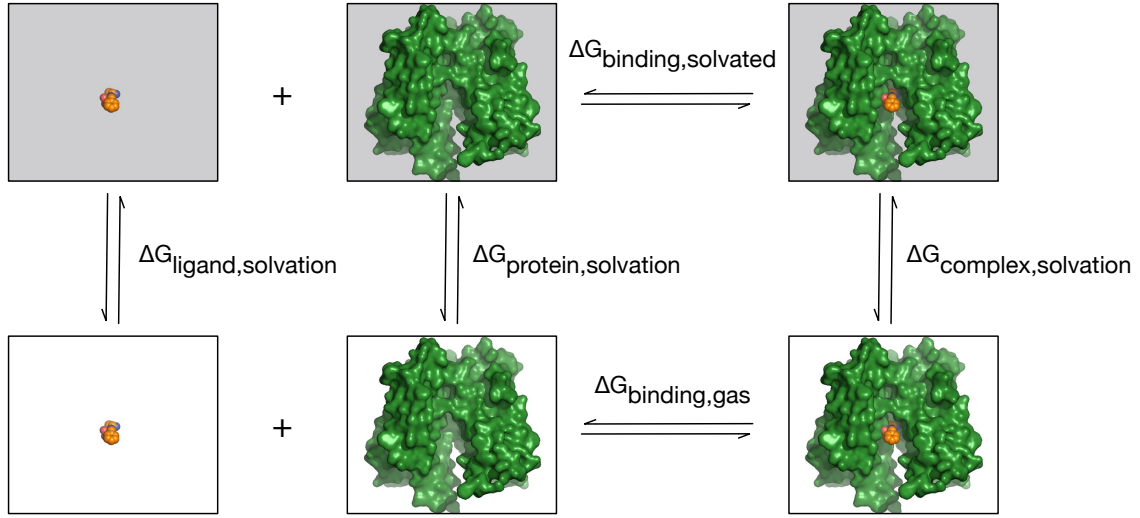
### Molecular Mechanics Poisson-Boltzmann/Generalized Born Surface Area (MMPB/GBSA)

Both the MMPBSA and MMGBSA methods are developed using the same fundamental principle: the implicit solvation of protein, ligand and their complex (Figure 2.7) (Srinivasan et al., 1998; Gouda et al., 2003). These methods calculate  $\Delta G$  by subtracting the free energy of the solvated unbound states from the free energy of the solvated bound state. The central equation is:

$$\Delta G_{\text{binding,solvated}} = G_{\text{complex,solvated}} - (G_{\text{protein,solvated}} + G_{\text{ligand,solvated}}) \quad (2.21)$$

$$\Delta G_{\text{solvated}} = E_{\text{MM}} + \Delta G_{\text{solvation,polar}} + \Delta G_{\text{solvation,nonpolar}} - TS_{\text{solute}} \quad (2.22)$$

where  $E_{\text{MM}}$  denotes the internal energies from the molecular mechanical force field,  $\Delta G_{\text{solvation}}$  represents the free energy of solvation, which is further decomposed into polar and non-



Figs. 2.7: Thermodynamic cycle for free energy of binding prediction using the MMPB/GBSA method. Solvated systems are illustrated in grey boxes, while systems in the gas phase are shown in white boxes. Figure adapted from Miller III et al. (2012).

polar components, and  $S_{\text{solute}}$  indicates the entropic contribution from the solute. The MMPBSA and MMGBSA methods use the Poisson-Boltzmann equation and generalised Born model, respectively, to estimate the polar component of  $\Delta G_{\text{solvation}}$ . The non-polar component is calculated based on the solvent accessible surface area (SASA). Finally, analysis of the normal modes is usually employed to estimate the solute entropy. This calculation, however, is associated with very large standard errors, often an order of magnitude larger than that of other energetic terms (Kar et al., 2011), resulting in unreliable  $\Delta G$  values. Additionally, the improvement in the agreement with experimental data when entropy is included has been shown to be small (Oehme et al., 2012). It is therefore common for relative free energy calculations between related system to assume similar solute entropies, thereby obviating the need to explicitly calculate them.

The Poisson equation describes the three dimensional electrostatic potential distribution as:

$$\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{\rho(\mathbf{r})}{\epsilon_0 \epsilon} \quad (2.23)$$

where  $\rho(\mathbf{r})$  represents the local charge density of the solute,  $\epsilon_0$  is the permittivity of free space and  $\epsilon$  is the dielectric constant. The distribution of the ions is accounted for by a

Boltzmann distribution as follows:

$$f_B = 4\pi \sum_i^N q_i c_i e^{\frac{-q_i \psi_i}{k_B T}} \quad (2.24)$$

where  $c_i$  is the bulk concentration of ionic species,  $k_B$  is the Boltzmann constant and  $T$  the temperature. Solving the Poisson Boltzmann equation is computationally costly, therefore an approximation is developed in the form of the Generalised Born model, which follows the functional form:

$$G_s = \frac{1}{8\pi} \cdot \left( \frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \cdot \sum_{i,j}^N \frac{q_i q_j}{f_{GB}}$$

where

$$f_{GB} = \sqrt{r_{ij}^2 + a_{ij}^2} e^{-D}$$

$$D = \left( \frac{r_{ij}}{2a_{ij}} \right), a_{ij} = \sqrt{a_i a_j} \quad (2.25)$$

where  $q_i$  is the atomic charge of particle  $i$ ,  $r_{ij}$  is its distance from particle  $j$  and  $a_i$  is its effective Born radius, which characterises the degree of burial of an atom. The accuracy of the model lies in the accurate estimation of the effective Born radius (Onufriev et al., 2002). When  $\Delta G$  is computed using the MMPBSA and the MMGBSA methods, determining the electrostatic component of the implicit solvation is the slowest step during the calculation. For a system with the POT transporter, PepT<sub>St</sub>, the MMGBSA method, on average, takes around 5 minutes, whilst the MMPBSA method around 30 minutes. In comparison, the LIE method converts the electrostatic potential directly into electrostatic force by a simple multiplication step, therefore each calculation can be performed under a minute.

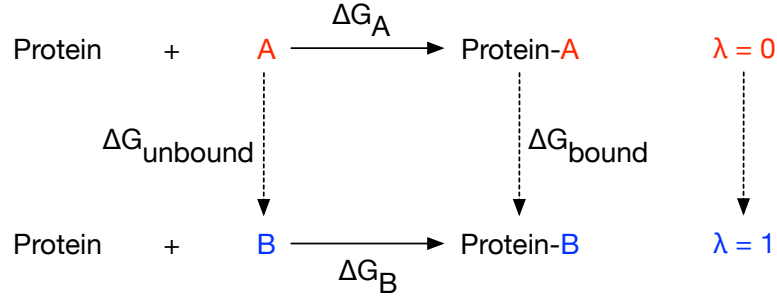
Equations 2.21 and 2.22 reflect the energies of the system at a single time point, but in practice, just like the LIE method, the MMPB/GBSA method uses the averages from an ensemble of configurations generated by MD simulations. To generate these ensembles, there are two different approaches: i) the single trajectory protocol (STP) and ii) the multiple trajectory protocol (MTP). STP requires only one simulation of the protein-ligand complex, from which the ensembles of free protein and ligand are extracted. This approach assumes that the protein and ligand sample a largely similar conformational space in the unbound state compared to the bound state. It is not only less computationally expensive, but be-

cause the configurations of the bound and unbound ensembles are identical, the internal energy terms cancel out, reducing the overall noise and improving convergence of the free energy of binding (Hou et al., 2011). For the MTP approach, three separate simulations are performed: protein-only, ligand-only and protein-ligand complex. The MTP is appropriate if there are large conformational changes involved during ligand binding, although this approach requires significantly longer simulations for the three components to converge.

In general, the end-point methods are more reliable than scoring functions, but the use of empirical constants and implicit solvation means quantitative comparison to experimental data is still difficult as the magnitude of  $\Delta G$  predicted could be way off from expected values. Kuhn et al. (2005) demonstrated that MMPBSA is useful to distinguish between good and weak binders of markedly different  $IC_{50}$  values, but rarely capable of reproducing small free energy differences. A good overall correlation to experimental data can be achieved when the ligand test set comprises a group of compounds of diverse binding characteristics (Brown and Muchmore, 2009). Intriguingly, MMGBSA performed better at ranking ligands compared to the more computationally expensive MMPBSA (Hou et al., 2011; Oehme et al., 2012). Additionally, the  $\Delta G$  values predicted by end-point methods are often not reproducible, such that different repeats of the same system can vary by as much as 10 kcal/mol (Wright et al., 2013). Therefore, for the most accurate and precise affinity predictions, theoretically rigorous ‘exact’ methods have to be employed.

### 2.3.3 ‘Exact’ free energy methods

The ‘exact’ free energy methods are derived from rigorous theoretical underpinnings without any empirical parameters. They are also frequently described as the “alchemical” methods as the chemical paths through which the calculation is performed are unphysical. As depicted in Figure 2.8, four free energies are considered: the binding energies of A and B to the protein (horizontal arrows), and the energies required to convert A to B in the unbound and the bound states (vertical arrows). As the free energy is a state function, the sum around the close thermodynamic cycle is zero. Therefore, instead of directly calculating the individual binding energies of A and B, which is computationally inefficient, the alchemical methods compute the difference by measuring the energies needed for the alchemical transformations (Equation 2.26). This approach can be applied to obtain a *relative* binding free energy ( $\Delta\Delta G$ ), in which case A and B are two different ligands, or an *absolute* binding free energy ( $\Delta G$ ), for which the ligand is transformed into a dummy molecule. The alchemical conversion is coupled to a collective variable,  $\lambda$ , which varies from 0 (initial state) to



Figs. 2.8: A thermodynamic cycle illustrating the use of alchemical transformation (dashed lines) to indirectly calculate the difference in binding free energy (solid lines) of A and B to a protein (see Equation 2.26).

1 (end state), that is used to scale the non-bonded interactions of the ligand to the protein. Individual simulations are performed for over a range of  $\lambda$  values and the free energy profile between successive  $\lambda$ s should overlap to ensure sufficient sampling. In this thesis, the ‘exact’ free energy method used is thermodynamic integration (TI).

$$\begin{aligned}\Delta G_A + \Delta G_{\text{bound}} - \Delta G_B - \Delta G_{\text{unbound}} &= 0 \\ \Delta G_B - \Delta G_A &= \Delta G_{\text{bound}} - \Delta G_{\text{unbound}}\end{aligned}\tag{2.26}$$

### Thermodynamic Integration (TI)

To understand the theoretical basis of this method, the two free energies must be first considered: Gibbs,  $G$ , and Helmholtz,  $A$ :

$$G \equiv H - TS; \quad A \equiv U - TS \tag{2.27}$$

where  $H$  is enthalpy,  $U$  is the internal energy,  $T$  is the temperature and  $S$  is the entropy. The former applies to the NPT ensemble, whereas the latter applies to the NVT ensemble. To understand the difference between these two free energies, the partition function,  $Q$ , which represents the sum of all the energies of the microstates existing in a system will be considered:

$$Q_{NPT} = \int \int e^{-\beta(U+pV)} dU dV; \quad Q_{NVT} = \int e^{-\beta U} dU \quad (2.28)$$

where  $\beta = \frac{1}{k_B T}$ . All of the simulations performed in this thesis are conducted in the NPT ensemble to match the experimental conditions, where the pressure is constant. As the volume is allowed to change during the simulation, the partition function,  $Q$ , contains the integration over the volume and the Boltzmann factor has a  $pV$  term, making the derivation of equation of thermodynamic integration more convoluted. This equation shall therefore be derived in the NVT ensemble to simplify the math, and later a corresponding expression for the NPT ensemble will be quoted by analogy.

First, the Helmholtz free energy,  $A$ , can be written in terms of the partition function,  $Q$ :

$$A = -\frac{1}{\beta} \ln Q \quad (2.29)$$

Assuming  $A$  is a continuous function of the coupling parameter,  $\lambda$ , where  $0 \leq \lambda \leq 1$ , we can write:

$$\Delta A = \int_0^1 \frac{\partial A}{\partial \lambda} d\lambda \quad (2.30)$$

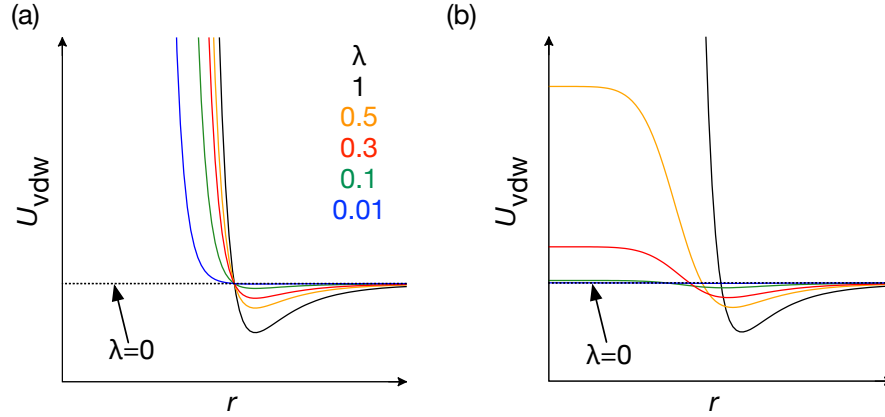
Substituting for  $A$  using Equation 2.29 gives:

$$\begin{aligned} \Delta A &= -\frac{1}{\beta} \int_0^1 \frac{\partial \ln Q}{\partial \lambda} d\lambda \\ &= -\frac{1}{\beta} \int_0^1 \frac{1}{Q} \cdot \frac{\partial Q}{\partial \lambda} d\lambda \end{aligned} \quad (2.31)$$

Finally, substituting the partition function from Equation 2.28 into Equation 2.31 gives:

$$\begin{aligned} \Delta A &= -\frac{1}{\beta} \int_0^1 \frac{-\beta}{e^{-\beta U}} \cdot \frac{e^{-\beta U} \partial U}{\partial \lambda} d\lambda \\ &= \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \end{aligned} \quad (2.32)$$





Figs. 2.9: The application of soft-core potential to avoid singularity. (a) Using the Lennard-Jones 12-6 function, when  $\lambda \rightarrow 0$ , the van der Waals potential is slowly reduced but the singularity remains until  $\lambda = 0$ , resulting in an abrupt change in energy. (b) The soft core van der Waals potential removes the singularity for a smooth transition.

By analogy, for the NPT ensemble, the Gibbs free energy can be written as:

$$\Delta G = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (2.33)$$

assuming that the pressure and volume of the system is independent of  $\lambda$ . In practice, this means that the ensemble average of the derivative of the potential energy of the system,  $U$ , with respect to  $\lambda$ , for several values of  $\lambda$  must be determined. These  $\frac{\partial U}{\partial \lambda}$  values are then plotted against  $\lambda$  and the area under the curve, or the integral, is the value of  $\Delta G$ . In this thesis, this integration is performed numerically using the trapezoid rule.

The alchemical MD simulation proceeds by linearly scaling the potential energy function of the molecule being transformed via  $\lambda$  to the potential energy of the initial structure, A, and the final structure, B:

$$U = (1 - \lambda)U_A + \lambda U_B \quad (2.34)$$

A problem arises with the van der Waals forces as the  $\lambda$  value approaches zero (or one in the case of  $(1-\lambda)$ ). The Lennard-Jones 12-6 function (fourth term in Equation 2.10) introduces a singularity at  $r = 0$ , brought on by the  $r^{-6}$  and  $r^{-12}$  terms. When this function is multiplied by a value that is close to zero, the volume encapsulated by this singularity becomes increasingly small (Figure 2.9(a)), but the singularity remains until  $\lambda = 0$ . Therefore, there is an extremely abrupt change in  $\frac{\partial U}{\partial \lambda}$  from  $\lambda$  close to zero to  $\lambda = 0$ , such that

an accurate numerical integration around this region becomes impossible. This also means that  $A$  is no longer a continuous function of  $\lambda$ . One way to solve this problem is to use a soft-core van der Waals potential (Beutler et al., 1994):

$$U_{\text{vdw}} = 4\lambda\epsilon \left( \frac{\sigma_{ij}^{12}}{[\alpha(1-\lambda)\sigma_{ij}^6 + r_{ij}^6]^2} - \frac{\sigma_{ij}^6}{\alpha(1-\lambda)\sigma_{ij}^6 + r_{ij}^6} \right) \quad (2.35)$$

where  $\alpha$  is a positive constant that denotes the “softness” of the potential. This potential function ensures that the van der Waals potential approaches a finite value as the inter-atomic distance  $r$  reaches zero, as a result of which, the transition of  $\frac{\partial U}{\partial \lambda}$  from  $\lambda$  close to zero to  $\lambda = 0$  becomes smoother (Figure 2.9(b)). The soft-core potential shall therefore be applied in the alchemical simulations in this thesis to avoid end-point integration errors.

Another important practical aspect when performing alchemical simulation is to avoid turning off electrostatic charges and van der Waals potentials at the same time. If two oppositely charged atoms, A and B, are considered, scaling back the van der Waals interactions will remove the repulsive force between them. The electrostatic potentials, however, scale much slower, thus A and B are still attracted to each other due to their residual partial charges. As the Coulomb potential is inversely proportional to  $r$ , when  $r = 0$  A and B will be trapped on top of each other with an infinite energy, which in practice, causes the simulation to crash. To avoid this problem, each alchemical transformation in this thesis is divided into three steps: i) removing the atomic partial charges, ii) removing the van der Waals potentials of the disappearing atoms whilst adding those of the emerging atoms, and iii) adding back the atomic partial charges. As such, predicting  $\Delta G$  using the TI method is significantly more computationally demanding than both the scoring functions and the end-point methods.

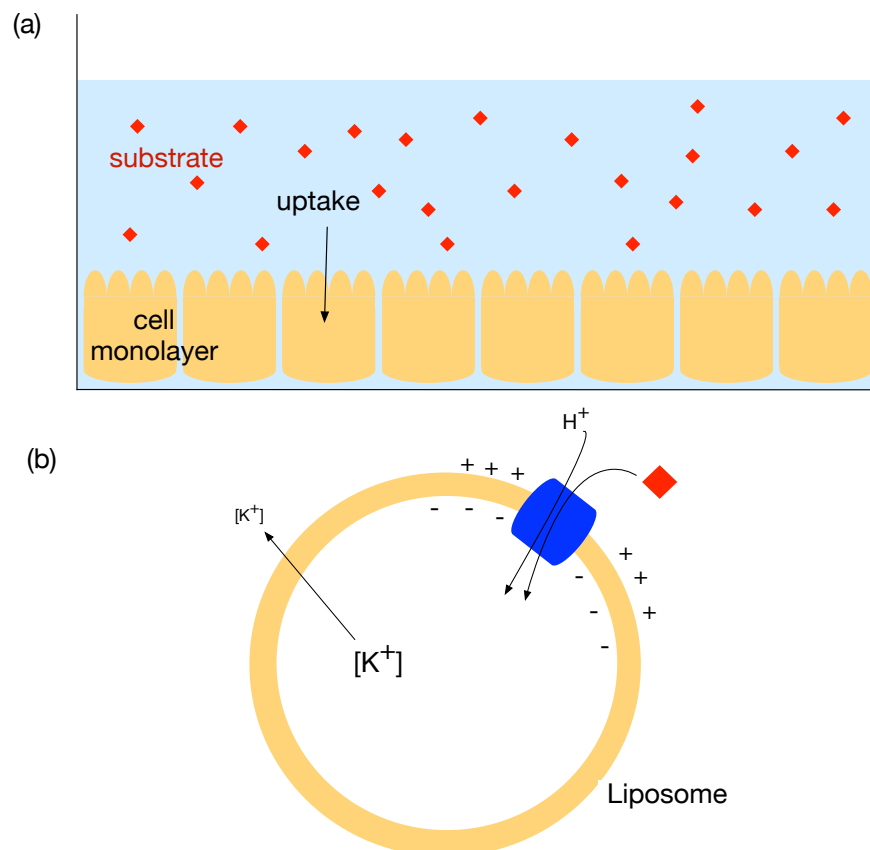
Due to their rigorous theoretical framework, the ‘exact’ free energy methods produce the most reliable results. For example, Rizzo et al. (2000) used these methods to study the binding of the HIV drug sustiva to HIV-1 reverse transcriptase and managed to compute  $\Delta G$  within 1-2 kcal/mol from experimental data. These alchemical free energy methods have also been used to validate a homology model of a G-protein-coupled receptor by calculating the relative  $\Delta G$  of several known inhibitors (Hénin et al., 2006). To gain mechanistic insights into the  $F_1F_0$ -ATPase, Yang et al. (2003) conducted several relative  $\Delta G$  calculations and identified the tight binding site of ATP. Nevertheless, as these methods require a prodigious amount of computational resources, the number of ligands studied in each paper is usually small and the calculations are mostly limited to relative binding free energies.

## 2.4 Experimental transport assays

In this thesis, binding free energies computed using methods discussed in §2.3 will be compared to experimentally determined affinities. To date, transport assays in whole cells (*in vivo*) or proteoliposomes (*in vitro*) remain the standard approach for the functional studies of peptide transporters, including to measure substrate affinities.

**Whole cell uptake assays** are performed by over-expressing the protein of interest in a heterologous system, e.g. the human colon adenocarcinoma (Caco-2) or Madin-Darby canine kidney (MDCK) epithelial cell lines for human membrane transporters (Volpe, 2011). Studies involving human PepT1 (Eddy et al., 1995; Knütter et al., 2004; Biegel et al., 2005; Vig et al., 2006) are usually conducted in a stably-transfected Caco-2 cell monolayer uptake assay (Figure 2.10a). A similar technique, albeit using *E. coli* cells as the expression system, has been used to study bacterial POTs such as PepT<sub>So</sub> (Newstead et al., 2011), YjdL (Jensen et al., 2012; Ernst et al., 2009), YdgR (Weitz et al., 2007; Malle et al., 2011; Bippes et al., 2013) and YhiP (Harder et al., 2008). One downside of this uptake assay is the limited choice of reporter substrates that can be used. As naturally occurring peptides are hydrolysed inside the cells, which renders uptake measurement impossible, non-physiological peptide analogues are used instead e.g. glycylsarcosine (GlySar) and  $\beta$ -Ala-Lys-AMCA. Borrowing the technique from ion channels, some studies (Knu et al., 2009; Ilka et al., 2008) have resorted to using two electrode voltage clamp electrophysiology, exploiting the electrogenic nature of these transporters. For this approach, the protein is over-expressed in *Xenopus laevis* oocytes and the proton influx that occurs concurrently with substrate uptake is measured. This method eliminates the need for a reporter substrate as the current measured is a direct indication of transport. It can also differentiate between a true substrate and a competitive inhibitor. For both types of *in vivo* assays, however, it is imperative to rule out the effects of other endogenous transporters, which might cause false positive results. Also, these methods vary from lab to lab due to different cell origins and culture conditions and between repeats due to variable protein expression levels. Therefore, comparison of results are often difficult.

**Proteoliposome transport assays** are a more robust system (Figure 2.10b) (Newman and Wilson, 1980). This assay uses an artificially imposed electrochemical gradient created by the addition of valinomycin, a potassium-specific ionophore that causes efflux of K<sup>+</sup>. The resulting positive-out/negative-in environment across the membrane of the liposome activates the transporters (Lolkema et al., 1991). Passive transport, i.e. non-electrogenic uptake driven by the concentration gradient of the substrate, can also be studied by us-



Figs. 2.10: Two types of transport assays used in the study of POT family transporters. (a) Whole cell uptake assay involves growing a layer of epithelial cells inside a well and incubate them with a medium containing substrates (red diamond). Substrates are labelled (radioactively or by a fluorescence tag) to quantitate the amount of uptake. Figure adapted from Volpe (2011) (b) Transporter proteins (blue square) are purified and reconstituted into a liposome containing a high concentration of  $K^+$ . Introduction of valinomycin induces efflux of  $K^+$  and creates an electrochemical gradient across the membrane that drive the uptake of substrates. These substrates are also labelled for measurement.

ing a high concentration of substrate inside the liposome that triggers the transporters to exchange them with external substrates (counterflow assay). Like the traditional *in vivo* uptake assay, substrates are often radio-labelled. However, as they are not hydrolysed after transport, physiological peptides like AlaAla can be used. Recent studies (Lee et al., 2013; Parker et al., 2014) employed pH sensitive dye like pyranine to directly measure the accumulation of protons inside the liposome. Proteoliposome-based assays are not amenable to electrophysiology as the small size of the system makes accurate measurement of current impractical. Still, this *in vitro* assay is popular amongst bacterial transporters (Hagting et al., 1997; Fann et al., 2003; Dang et al., 2010; Doki et al., 2013) as it allows for a better control of experimental conditions, e.g. substrate and ion concentrations, which results in more meaningful comparisons between different studies. Also, there is no influence of other transporters on substrate uptake. However, this approach has yet to be applied to human POTs due to the difficulties to purify the protein, and therefore current studies on PepT1 and PepT2 are limited to cell-based assays.

A competition assay is the most common approach to compare the affinities of various substrates. By measuring the residual uptake of a reporter substrate, e.g.  $^3\text{H}$ -AlaAla or  $^{14}\text{C}$ -GlySar, the transporter's affinity for a competing substrate of interest can be inferred. The typical metric for a competition transport assay is  $\text{IC}_{50}$ , which is the concentration of a competing substrate required to reduce the uptake of a reporter substrate by half. A low  $\text{IC}_{50}$  value corresponds to a strong competition and thus high substrate affinity and, by extension, more negative binding energy. In this thesis,  $\text{IC}_{50}$  values are used as the experimental data to which the computational predictions of binding free energy are compared.

## 2.5 Homology modelling

Since the first high resolution crystal structure of a protein was solved in 1958 (Kendrew et al., 1958), the number of unique proteins with structures determined by X-ray crystallography has increased exponentially to  $\sim 110,000$  (Berman et al., 2000). This number, however, is small compared to  $\sim 54$  millions currently annotated protein sequences (Bairoch and Apweiler, 2000). As protein structure is much more conserved than sequence during evolution, homology modelling is frequently employed to predict unknown protein structures. For an in depth review on homology modelling, see a book by Orry and Abagyan (2012).

First, a template structure is selected based on sequence similarity. If a suitable template is not known, the sequence of the protein of interest can be used as a query for searching a database using the basic local alignment search tool (BLAST). The higher the sequence ho-

mology, the higher the confidence in the modelled structure, with a 30% sequence identity generally accepted as a lower threshold to generate a useful model (Eswar et al., 2003). The sequence of this template is then aligned to that of the protein to be modelled. Automated sequence algorithms like Probcons (Do et al., 2005) are frequently utilised at this step, although manual curation of the alignment is often necessary, especially when the homology between the two sequences is low. A multiple sequence alignment consisting of the target and template proteins from various organisms is recommended as it can highlight the evolutionary relationship between the proteins, and therefore increases the probability for key residues to be correctly aligned.

Based on the crystal structure of the template and the sequence alignment, the model is built by mapping the position of the target residue to the coordinates of the corresponding amino acids on the template. The backbone of the residue is modelled first, while the side chain is later added using rotamer libraries. Additionally, spatial restraints, often obtained from biophysical studies like double electron-electron resonance (DEER) spectroscopy, can be included during this step to increase the validity of the model. The regions in the protein without any template information are modelled *de novo*, typically as loops, and represent the least reliable regions in the model. Finally, the homology model is evaluated, for example, using the discrete optimised protein energy (DOPE) score (Shen et al., 2006). The DOPE scores assess the atomic distances in a model and compare them to the ones observed in available crystal structures. Other metrics regularly used to evaluate a homology model are the backbone  $\phi$  and  $\psi$  angles using the Ramachandran plot (Ramachandran et al., 1963) and the Molprobity scores (Chen et al., 2010). In this thesis, the Modeller program (Sali and Blundell, 1994) will be used to generate a homology model of human peptide transporters.

## 2.6 Protein-protein docking

Protein-protein interactions are ubiquitous in biology (Jones and Thornton, 1996). Predicting the structure of a protein-protein complex, however, is one of the most challenging tasks in computational modelling. When the three-dimensional structures of two independently crystallised (or modelled) proteins are docked to each other, the number of possible complexes is so enormous that distinguishing the native structures from the false positives is almost impossible. Experimental inputs, for example the residues involved in the interactions, are crucial to delineate the binding interface, effectively reducing the number of degrees of freedom during docking. The models of a protein complex are usually evaluated and ranked using scoring functions to find the best configuration.

In this thesis, the ClusPro program developed by Comeau et al. (2004) will be used to perform protein-protein docking. Cluspro takes the coordinates of two proteins and uses the DOT rigid-body docking algorithm (Mandell et al., 2001) to dock them together based on shape complementarity. The resultant models are filtered based on the electrostatic and desolvation free energy, both of which are calculated empirically using simplified energy functions (Camacho et al., 2000; Zhang et al., 1997). As many as 2000 candidate complexes are retained for clustering, where one protein is “fixed” (designated as the “receptor”) and the complexes are grouped based on the positions of the other protein (“ligand”). The complex with the ligands having the most neighbours is then selected as the representative structure of the cluster. Finally, these are subject to a short energy minimisation step using CHARMM (Brooks et al., 1983) to remove clashes between side chains. ClusPro is a fully automated docking program that can be accessed at <http://cluspro.bu.edu/login.php>.

## Chapter 3

# Binding affinity predictions using free energy methods

A range of binding affinity prediction methods were used to predict the preference of a bacterial peptide transporter, PepT<sub>St</sub>, to a group of peptides. By comparing the results to an *in vitro* competition transport assay, it was shown that end-point methods, such as molecular mechanics Poisson Boltzmann surface area (MMPBSA), are capable of distinguishing between good and poor substrates. More computationally demanding methodologies like thermodynamic integration (TI) can subsequently be used to improve the results within the subset of good substrates. Scoring functions, albeit being the fastest approach, failed to replicate experimental results and essentially produce near random predictions. By plotting correlation to experimental data against computational expense, it appears that the most pragmatic way to predict binding affinities is through a hybrid approach—end-point methods to broadly classify the ligands and theoretically exact methods for refining the results.

### 3.1 Binding affinity and transport

In Chapter 2, the importance of measuring the free energy of binding,  $\Delta G$ , in the study of biomolecular interactions was discussed. In recent years, predicting binding affinity using computational methods has become a popular option to study molecules that are difficult to synthesise and also to enhance understanding on specific intermolecular interactions that cannot be directly observed (Schames et al., 2004; Fowler et al., 2007; Wang et al., 2006a;

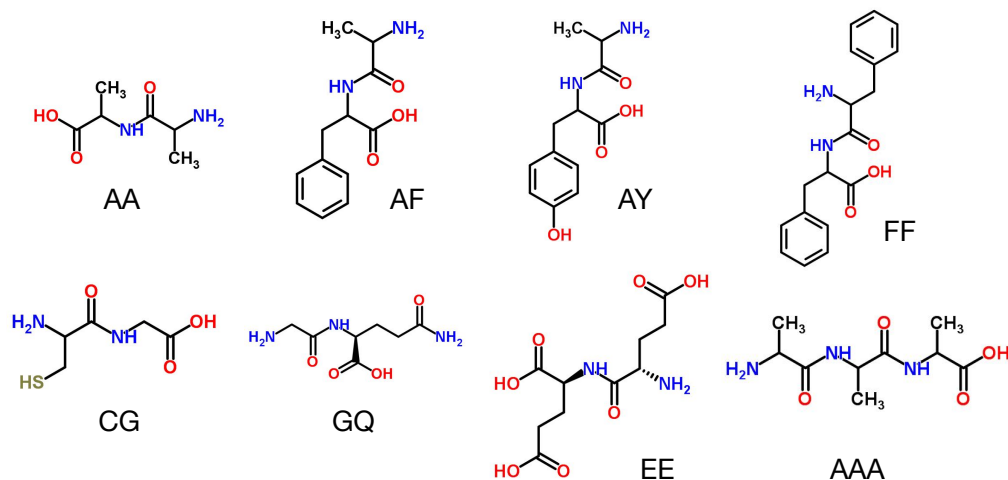


Ge et al., 2012; Wright et al., 2013; Spiliotopoulos et al., 2012). For membrane transporters, however, estimating binding affinity *in silico* is currently not widespread for both practical and theoretical reasons. It is currently challenging to experimentally determine the  $\Delta G$  of binding of a small molecule ligand to a membrane protein as a lipid bilayer is required for the protein to function properly. As such, data to which computationally predicted  $\Delta G$  values can be quantitatively compared are scarce.

Ligand affinities to a membrane transporter are typically measured using competition transport assays, whereby the ability of a certain ligand of interest to compete the transport of a reporter substrate is measured. The most common metric used to compare between different ligands is  $IC_{50}$ , which is the concentration required to reduce the transport of the reporter substrate by half. Ligands with a high affinity will therefore have a small  $IC_{50}$  value as they need a small amount to outcompete the reporter substrate. It is worth noting, however, that  $IC_{50}$  is not a measure of how well a ligand is transported by the membrane transporter. In human peptide transporters, for example, large hydrophobic tripeptides like TrpTrpTrp displayed low  $IC_{50}$  values but are not transported, and are therefore competitive inhibitors (Biegel et al., 2005, 2006).

During a complete transport cycle, a transporter protein must undergo significant conformational changes to move a substrate from one side of the membrane to the other. The relationship between  $IC_{50}$  value and  $\Delta G$  of binding to one structure of the transporter during this cycle, if any, is poorly understood. As initial substrate binding occurs in the outward facing conformation, the competition between a ligand of interest (for which the  $IC_{50}$  value is measured) with the reporter substrate is most likely to happen at this stage. It is therefore probable that the  $IC_{50}$  value is somewhat related to  $\Delta G$  of binding in the outward facing conformation; whether such a relationship persists in other conformational states (occluded and inward facing) is still up for debate.

The apparent affinity of substrates also depends on the rate of conformational change as this will determine the distribution of transporters that are accessible to the ligands for binding to occur (Eraly, 2008). It is so far unclear how the strength of binding of a ligand correlates to the rate of conformational change of a transporter. The crystal structure of a plant nitrate transporter, NRT1.1, suggests that the overall turnover rate is dependent on the gating efficiency (Parker and Newstead, 2014). A recent study of the glutamate transporter homolog, Glt<sub>Ph</sub>, proposed an inverse relationship between substrate affinity and transport domain motion (Akyuz et al., 2014). For human PepT1, it has been shown that certain compounds like the angiotensin-converting enzyme (ACE) inhibitors can act as competitive inhibitors for the transporter (Brandsch, 2009), indicating that ligand binding is not a



Figs. 3.1: The structures of peptides that were used to assess binding affinity prediction methods.

guarantee for a conformational change leading to transport.

Additionally, as the molecular architecture of the binding site changes during a conformational change, it is also likely that the binding affinity of a ligand differs at every step. The glucose transporter GLUT1, for example, has approximately 10 times higher apparent affinity for D-glucose in the outward facing state compared to the inward facing state (Brahm, 1983). To complicate things even further, the transport of substrate is coupled to the movement of one or more ions in either a parallel (symport) or an anti-parallel (antiport) direction, which means that more than one association and dissociation events should be considered. Therefore, deriving an “experimental”  $\Delta G$  from an  $IC_{50}$  value is at best approximate. It is more sensible to use experimentally measured  $IC_{50}$  as a proxy for binding affinity and compare them to computational predictions only in a qualitative manner.

Here, the question of which computational method can best predict the binding affinity of peptide to the POT transporters is addressed, using a bacterial homolog  $PepT_{St}$  as a model system. Based on the availability of experimental data, eight peptides are selected for the test set (Figure 3.1), which represent a wide range of transport properties as depicted by the  $IC_{50}$  values from competition transport assays (Solcan et al., 2012) (Table 3.1). Five different computational methods for calculating  $\Delta G$  are chosen, which will be validated using this test set. These methods vary in the amount of computational resources required, from the fast structure based scoring function in AutoDock (Trott and Olson, 2010), to the intermediate end-point methods, which are linear interaction energy (LIE), molecular mechanics generalised Born surface area (MMGBSA), and molecular mechanics Poisson Boltzmann

Table 3.1: Peptide transport by PepT<sub>St</sub>

| Peptide   | IC <sub>50</sub> ( $\mu$ M) |
|-----------|-----------------------------|
| PhePhe    | 5.5                         |
| AlaPhe    | 21.5                        |
| AlaAla    | 36.0                        |
| GlyGln    | 37.1                        |
| AlaTyr    | 69.9                        |
| CycGly    | 342.0                       |
| AlaAlaAla | 404.3                       |
| GluGlu    | 597.6                       |

IC<sub>50</sub> values were measured by competition transport assays in re-constituted proteoliposome using <sup>3</sup>H-Ala-Ala as the reporter substrate (Solcan et al., 2012).

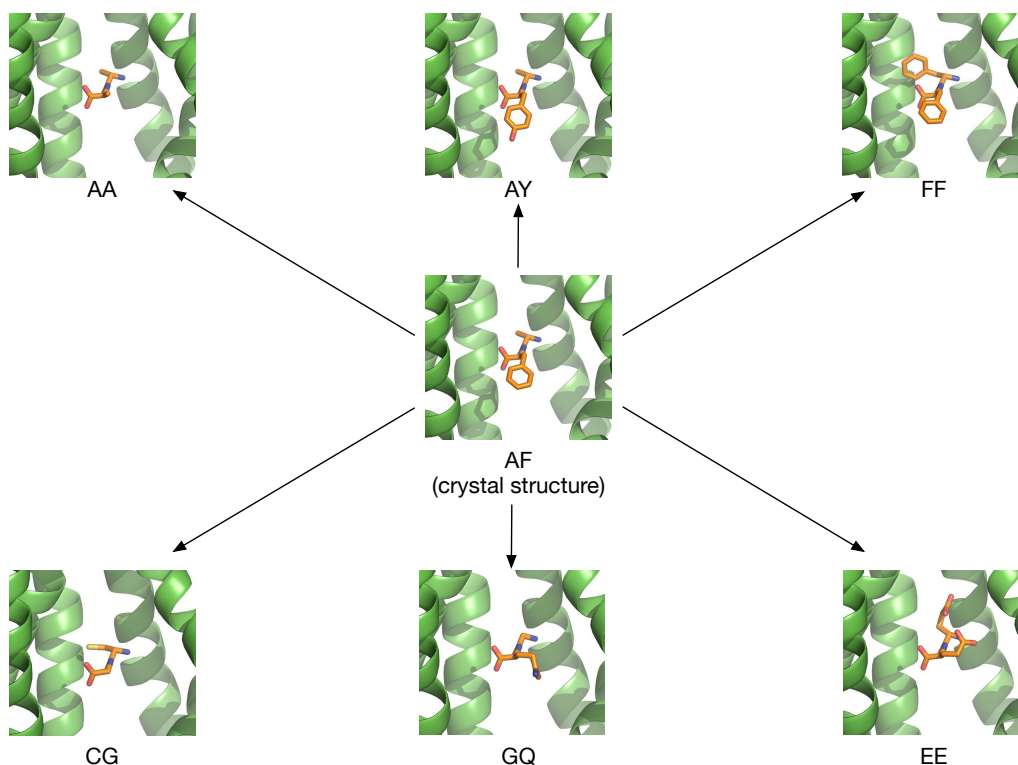
surface area (MMPBSA), to the rigorous exact method, thermodynamic integration (TI).

## 3.2 Methods

For completeness, the methods discussed in Chapter 2 are outlined briefly here, focussing on the parameters that were used for the MD simulations and free energy calculations, and the statistical methods to analyse the results.

### 3.2.1 Modelling peptide binding

Using the crystal structure of PepT<sub>St</sub> bound to AlaPhe (PDB ID: 4D2C (Lyons et al., 2014)), the binding of six dipeptides, i.e. PhePhe, AlaAla, GlyGln, AlaTyr, CysGly and GluGlu, were modelled using Modeller version 9v9 (Sali and Blundell, 1994) (Figure 3.2). By visual inspection, we ensured that the backbone of these peptide substrates made the same contacts with residues in the binding site as the backbone of AlaPhe. As a result, only the side chain of the peptide substrates was altered, whilst the structures of the rest of the protein and substrate were kept identical to the original crystal structure.

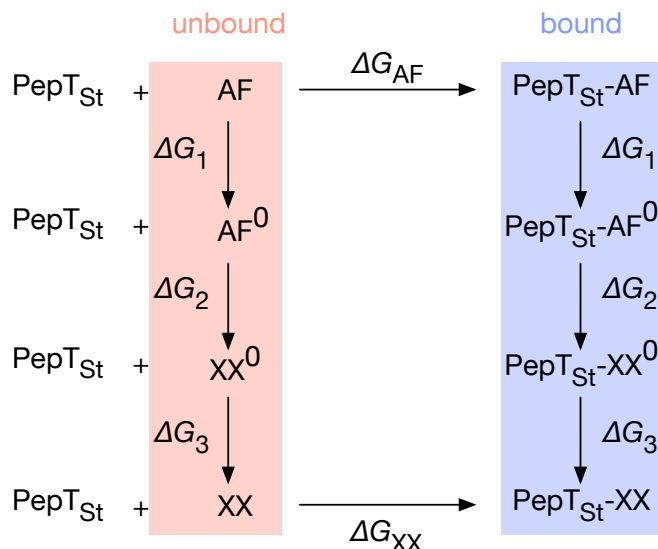


Figs. 3.2: Modelling dipeptides based on the crystal complex of PepT<sub>St</sub>-AlaPhe. For clarity, only the binding site of the protein is shown with H2, H11, HA and HB removed.

### 3.2.2 Binding energy prediction

**Scoring Functions** AutoDock Vina (Trott and Olson, 2010) was used to dock each peptide to the binding site. Each docking procedure produced 9 binding poses, which were then compared to the crystal structures (for AlaPhe and triAla) and models (other peptides). The score for the most similar binding pose (based on all-atom RMSD) is taken as the affinity value. No MD simulations were involved in this protocol.

**End-point methods** Two MD simulations were performed for each peptide: i) peptide bound to PepT<sub>St</sub> and ii) the free peptide in solution. The LIE method uses the trajectories from both simulations, whilst for MMGBSA and MMPBSA, only the former is used following the single trajectory protocol (Hou et al., 2011), whereby the trajectories of unbound PepT<sub>St</sub> and free peptide were extracted from it. A GROMACS tool, *g\_lie*, was used to calculate  $\Delta G$  values along the simulation trajectory based on the LIE method, while for MMGBSA and MMPBSA, a python program integrated inside the AmberTools package, *MMPBSA.py*, was used (Miller III et al., 2012).

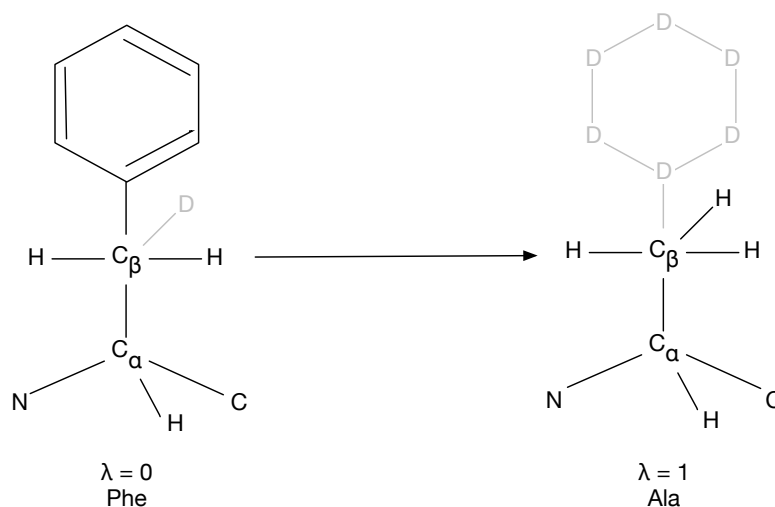


Figs. 3.3: A thermodynamic cycle to calculate the relative binding free energy of dipeptide XX with respect to AlaPhe. The transformation of AlaPhe to XX is divided into three steps:  $\Delta G_1$  for removing the electrostatic charge,  $\Delta G_2$  for altering the van der Waals interactions, and  $\Delta G_3$  for adding the electrostatic charge.

**Thermodynamic integration** The TI method was applied to only three hydrophobic dipeptides, AlaAla, PhePhe and AlaTyr, to calculate the relative free energy of binding,  $\Delta\Delta G$ , with respect to AlaPhe, based on the thermodynamic cycle shown in Figure 3.3. As discussed in Chapter 2, free energy is a state function, therefore to calculate  $\Delta\Delta G$  of AlaPhe and XX, the difference in energies required to transform AlaPhe to XX in the bound and unbound state was measured.

Alchemical transformation MD simulations were performed using the PepT<sub>St</sub>-AlaPhe crystal structure for the starting coordinates and the dipeptide ligand was morphed to either AlaTyr, PhePhe or AlaAla. The dual topology approach was employed, whereby the vanishing and growing atoms were represented separately. For example, during the transformation of AlaPhe to AlaAla, both the phenyl ring and the methyl group were attached to the second amino acid residue at the same time. However, to minimise the number of alchemical transformations, branching was done at the  $C_\beta$  atom rather than the  $C_\alpha$  atom. The phenyl ring was morphed to a ring of dummy atoms while another dummy atom was transformed into a hydrogen to form a complete methyl group (Figure 3.4). All alchemical transformations involved only the non-bonded interactions while bonded interactions were kept the same throughout the simulations (Boresch and Karplus, 1999; Boresch, 2002).

Each transformation was further divided into three steps: i) removing the electrostatic

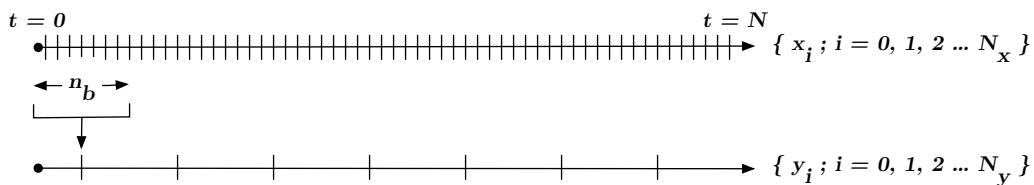


Figs. 3.4: Dual topology approach for the transformation of AlaPhe to AlaAla. For clarity only the C $\alpha$  and side chain of the second residue was shown. Dummy atoms are represented as “D” and coloured in silver.

charge on the disappearing chemical groups, ii) removing the van de Waals envelope of the disappearing groups while growing that of the emerging chemical groups, and iii) recharging the emerging groups (Figure 3.3). The transformation was done by coupling the non-bonded potential terms to a scaling parameter  $\lambda$ , whereby at  $\lambda = 0$ , the non-bonded terms of AlaPhe was used whilst at  $\lambda = 1$ , the non-bonded terms of XX was used. A soft core potential (Beutler et al., 1994) was used in step (ii) to avoid singularities and instabilities. For each step, eleven independent 5 ns MD simulations with varying  $\lambda$  values were performed i.e. ( $\lambda = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ ). The  $\frac{\partial U}{\partial \lambda}$  values were computed every 1 ps and these were used to estimate equilibrium and correlation times. After discarding data from the equilibration period, de-correlated  $\frac{\partial U}{\partial \lambda}$  values were integrated using the trapezoid rule to obtain a  $\Delta G$  value along the 5 ns trajectory. In total, for each  $\Delta \Delta G$ , 66 simulations (330 ns in total) was used for this calculation.

### 3.2.3 Simulation details

PepT<sub>St</sub> and all peptides were parametrized using the standard Amber99SB force field (Lindorff-Larsen et al., 2010). Peptide-bound PepT<sub>St</sub> was embedded in a pre-equilibrated 381-molecule POPC model membrane from Dickson et al. (2012) inside a simulation box of dimensions 10 nm  $\times$  13 nm  $\times$  15 nm using the GROMACS g\_membed protocol (Wolf et al., 2010). The simulation box was hydrated by TIP3P water (Jorgensen et al., 1983) and neutralised by



Figs. 3.5: Schematic illustration of the block averaging method used to coarse-grain MD output data to gain independent samples.

0.15 M  $\text{Na}^+$  and  $\text{Cl}^-$  ions. The steepest-descent protocol was used to minimise the system before being subject to a 1 ns equilibration simulation, during which heavy atoms were positionally restrained. This restraint was removed for a production MD simulation, the trajectory of which was used in the free energy calculations. Peptide-only simulations underwent the same protocol, albeit without the membrane and in a smaller simulation box (size 3 nm  $\times$  3 nm  $\times$  3 nm). The length of simulations performed for the end-point methods was 1 ns and for the TI method, 5 ns for each  $\lambda$ . The temperature was kept at 310 K using the velocity-rescaling thermostat with a time constant of 0.1 ps (Bussi et al., 2007). The pressure was kept at 1 atm by a semi-isotropic coupling to the Parrinello-Rahman barostat with a time constant of 1.0 ps and compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  (Parrinello, 1981). The lengths of all bonds with hydrogen atoms were constrained using the LINCS algorithm (Hess et al., 1997), so an integration time step of 2 fs could be used. Long-range electrostatic interactions were calculated using particle mesh Ewald (PME) method (Essmann et al., 1995) using a real space cutoff of 12 Å, and the van der Waals forces were cutoff at 12 Å. The energy data were written out every 1 ps.

### 3.2.4 Statistical Analysis

**Convergence analysis and error estimation** For binding affinity prediction methods that uses MD simulations as their input, it is essential to calculate two timescales: i) the equilibration time,  $t_{\text{eq}}$ , which is the time required for the simulation to reach equilibrium and ii) the autocorrelation time,  $\tau$ , which is the time between effectively independent samples. The former is used to discard non-equilibrated data while the latter is important to calculate statistical errors. Both of these values were estimated using the method of reverse cumulative averaging (Yang et al., 2004) that uses block averaging (Flyvbjerg and Petersen, 1989) and statistical inefficiency (Friedberg and Cameron, 1970).

A series of measurements from an MD simulation is highly correlated. The block averaging technique is used to “coarse-grain” these data until consecutive points become independent of one another, which is a prerequisite for accurate subsequent analyses. A set of  $N$  original MD output values of  $x_0, x_1, x_2 \dots, x_N$ ,  $\{x_i\}$  with a variance of  $\sigma_x^2$  is divided into sequential and non-overlapping bins of equal width  $n_b$ . The mean of  $x_n$  inside each bin is calculated to form a second dataset  $\{y_i\}$  with a variance of  $\sigma_y^2$  (Figure 3.5). This procedure is reiterated with increasing  $n_b$  values until  $\{y_i\}$  becomes uncorrelated. Statistical inefficiency,  $\phi$  is given by equation 3.1 and correlation time,  $\tau$  is the limiting value of  $\phi$  as  $n_b \rightarrow \infty$ .

$$\phi = n_b \cdot \frac{\sigma_y^2}{\sigma_x^2} \quad (3.1)$$

This block averaging technique is performed from the last frame,  $t = N$ , to the first frame,  $t = 0$ , effectively generating a reverse coarse-grained data set. Multiple block averaging steps are conducted and, for each step, an increasing portion of initial data are excluded. The fundamental idea behind this method is that when non-equilibrated data are included during the calculation,  $\phi$  will not have an asymptote. This is caused by the lower frequency motions often found in non-equilibrated system that are correlated over a longer timescale. By determining the smallest amount of initial data that has to be excluded for a converging value of  $\phi$ , the value of  $t_{eq}$  can be estimated. Once the data before  $t_{eq}$  is eliminated, the standard error can be calculated using equation 3.2. See Figure 3.6 for an example of how these statistical analyses were used in this thesis.

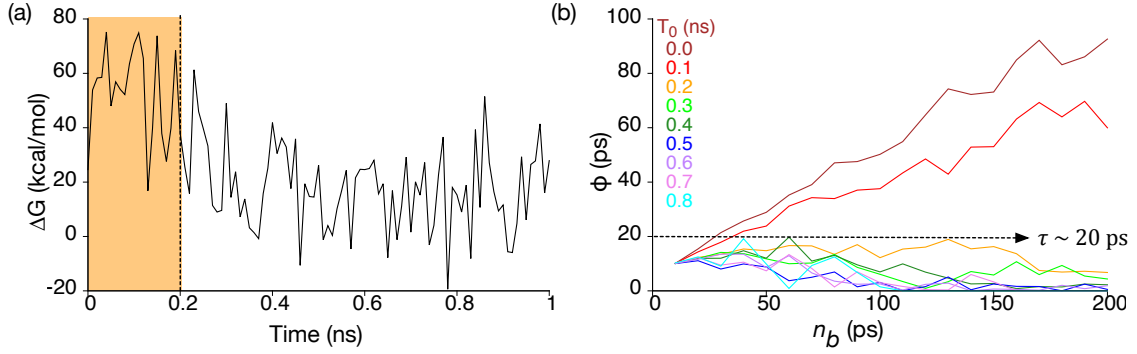
$$\sigma = \sqrt{\sigma_x^2 \cdot \frac{\tau}{N}} \quad (3.2)$$

**Performance analysis for prediction methods** The performance of each binding affinity prediction method was assessed by two statistical metrics: i) Spearman’s rank correlation coefficient (Lehmann and D’Abrera, 1998),  $\rho$  and ii) Pearson product-moment correlation coefficient (Pearson, 1895),  $r$ .

$\rho$  is a measure of correlation between the *rank* of two variables and it does not directly take into account the actual value of the raw data. In this case, for example, the  $IC_{50}$  values,  $X_i$  and predicted  $\Delta G$ s,  $Y_i$  were ranked from the best substrate to the poorest substrate and the two rankings were compared using this equation:

$$\rho = 1 - \frac{6 \cdot \sum (x_i - y_i)^2}{n(n^2 - 1)} \quad (3.3)$$





Figs. 3.6: An example of using the reverse cumulative averaging method to calculate  $t_{eq}$  and  $\tau$ . (a)  $\Delta G$  data calculated using the LIE method from a 1 ns simulation of PepT<sub>St</sub> bound to AlaPhe. The non-equilibrated region is shown in orange. (b) Nine statistical inefficiency analyses, each excluding either 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 or 0.8 ns of the initial simulation data. An asymptote is reached at 20 ps (corresponds to  $\tau$ ) when  $T_0 = 0.2$  ps ( $t_{eq}$ ).

whereby  $x_i$  is the ranking for  $X_i$ ,  $y_i$  is the ranking for  $Y_i$  and  $n$  is the size of the dataset.

Whilst  $\rho$  provides the simplest qualitative way of comparing between experimental and simulation data,  $r$  is a more stringent statistical test, whereby correlation between the *covariance* of the two datasets is evaluated, essentially computing a linear correlation. The calculation for  $r$  is described below:

$$r = \frac{\Sigma(X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \cdot \Sigma(Y_i - \bar{Y})^2}} \quad (3.4)$$

whereby  $\bar{X}$  is the mean value of  $X_i$  and  $\bar{Y}$  is the mean value of  $Y_i$ . As previously discussed, due to the intricate nature of the mechanism of a membrane transporter, it is not known exactly how  $\Delta G$  varies with  $IC_{50}$ . For the sake of comparison between the different binding affinity prediction methods, here it shall be assumed that  $\Delta G$  varies linearly with the log of  $IC_{50}$  following the relationship described by Equation 2.13.

For both analyses, a value of 1 represents perfect positive correlation, 0 represents no correlation (random prediction) and -1 represents perfect negative correlation.

### 3.3 Results and Discussion

The summary of how all the prediction methods perform is shown in Table 3.2. All end-point methods and the TI method show good agreements with experimental data, while AutoDock gives inferior performance as shown by the low correlation coefficients.

Table 3.2: The performance of binding affinity prediction methods for PepT<sub>St</sub>

| Correlation measure | AutoDock | LIE  | MMGBSA | MMPBSA | TI   |
|---------------------|----------|------|--------|--------|------|
| $\rho$              | 0.43     | 0.74 | 0.67   | 0.71   | 1.00 |
| $r$                 | 0.54     | 0.66 | 0.78   | 0.69   | 0.99 |

$\rho$  represents Spearman's rank correlation coefficient calculated using Equation 3.3, whilst  $r$  represents Pearson correlation coefficient calculated using Equation 3.4..

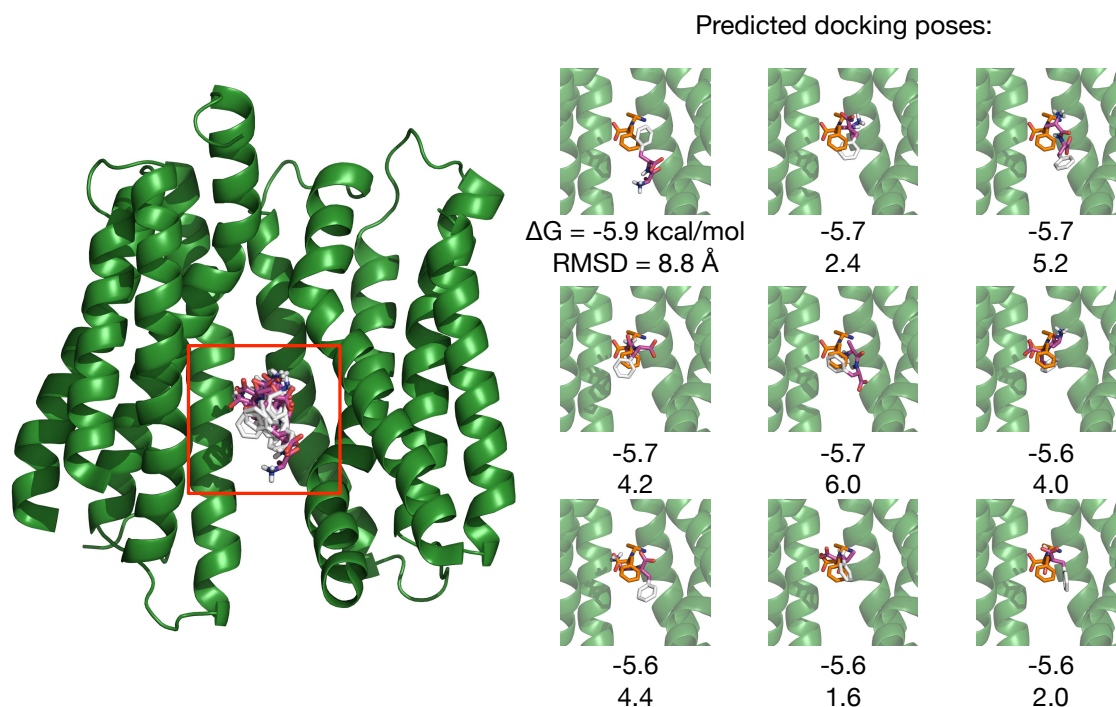
#### 3.3.1 Docking scores are inaccurate.

AutoDock is able to reproduce the binding orientation observed in the crystal structure to an RMSD of less than 2 Å (for all atoms), but the score for the correct pose is not always the highest amongst all suggested poses (Figure 3.7). Also, when the scores for all eight peptides in the test set are compared to the experimental IC<sub>50</sub> values, it is obvious that AutoDock fails to rank the ligands correctly (Figure 3.8). This is exemplified by the poor substrate, GluGlu, which is predicted to have a similar affinity to a good substrate, AlaPhe ( $\Delta G = -5.4$  and  $-5.7$  kcal/mol, respectively).

Table 3.3: Predicted  $\Delta G$  for peptide binding to PepT<sub>St</sub> based on AutoDock

| Peptide   | $\Delta G$ (kcal/mol) |
|-----------|-----------------------|
| PhePhe    | -6.5                  |
| AlaPhe    | -5.6                  |
| AlaAla    | -4.4                  |
| AlaTyr    | -5.6                  |
| GlyGln    | -5.0                  |
| CysGly    | -4.0                  |
| AlaAlaAla | -5.2                  |
| GluGlu    | -5.4                  |

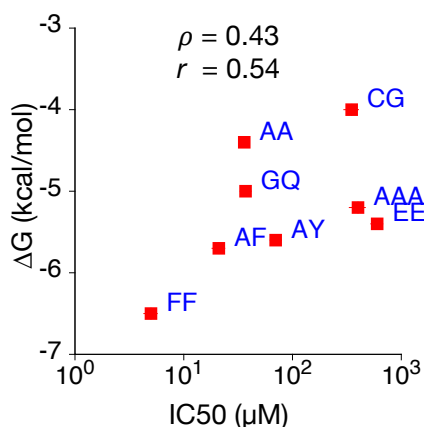
There are at least two reasons why AutoDock does not perform well in this test: i) it uses a simplified scoring function and ii) it takes no account of dynamics in the calculation.



Figs. 3.7: An example of a docking protocol using AutoDock Vina for AlaPhe. The search space is illustrated by the red box and all nine binding poses predicted by the program are shown. These are ranked based on  $\Delta G$  estimated by a scoring function. The proposed AlaPhe binding orientations (pink) are compared to the crystal structure (orange) and the all-atom RMSD is shown. For clarity, H2, H11, HA and HB are removed.

AutoDock uses a scoring function that is inspired by the X-SCORE (Trott and Olson, 2010; Wang et al., 2002). It was developed using a training set of 190 protein-ligand complexes of known binding affinities by a regression analysis, which tend to pick up only common interaction characteristics from the training set and therefore is likely to miss others that are not included in the study. This is apparent when they used the scoring function against a test set of 30 protein-ligands complexes, for which the correlation to experiment was only 0.36, comparable to the results obtained in this thesis.

A comparative study of various scoring functions by Cheng et al. (2009) showed that X-score yielded reasonable results with certain types of protein-ligand complexes, for example trypsin ( $\rho=0.8$ ) and thrombin ( $\rho=0.6$ ) complexes, for which there exists a notable correlation between binding constants and the size of ligands. In other cases where such correlation is not obvious, X-score performed poorly, for example for HIV protease complexes ( $\rho=0.3$ ). This suggests that X-score relies too heavily on physical properties of the ligand for its prediction. Our peptide test set also displays low correlation between the size



Figs. 3.8: Binding affinities of peptides to PepT<sub>St</sub>, as predicted by AutoDock. X-error bars indicate standard errors from three repeat measurements.

of ligands and IC<sub>50</sub> values: AlaTyr and GluGlu have similar molecular weights but very different IC<sub>50</sub> values. This might contribute towards the poor performance of the scoring function.

AutoDock also uses only one snapshot of the protein-ligand complex and as such ignores any rearrangements that occur during binding. This method will only be accurate if the protein is rigid and the binding of a ligand follows the “lock-and-key” model (Fischer, 1894). A promiscuous protein such as PepT<sub>St</sub>, however, undergoes considerable conformational changes to accommodate different ligands. As discussed earlier, this method also failed with HIV protease, in which ligand binding also involves large conformational changes (Cheng et al., 2009). It can therefore be concluded that AutoDock is inadequate to accurately predict peptide binding affinities for PepT<sub>St</sub> due to the simplified scoring function and lack of dynamics information.

### 3.3.2 End-point methods accurately predict transport.

A short 1 ns simulation of peptide-bound PepT<sub>St</sub> and peptide in solution was performed for each peptide using the parameters discussed in the previous section. Statistical analysis suggests that most of the simulations take a very short time to equilibrate ( $\sim 300$  ps), except the simulations of PepT<sub>St</sub>-AlaAla and PepT<sub>St</sub>-GluGlu, which reaches equilibrium only after 800 and 600 ps, respectively, leaving only 200 and 400 ps for data analysis. Also, all 8 simulations have  $\tau \leq 20$  ps.

Unlike AutoDock, all the end-point methods manage to agree reasonably well with experimental data with encouraging  $\rho$  of around 0.7 (Figure 3.9). The predicted  $\Delta G$  values are also distributed within a larger range ( $\sim 60$  kcal/mol), as a result of which the good and poor substrates are better distinguished. All methods consistently produce better rankings

Table 3.4: Predicted  $\Delta G$  for peptide binding to PepT<sub>St</sub> based on end-point methods

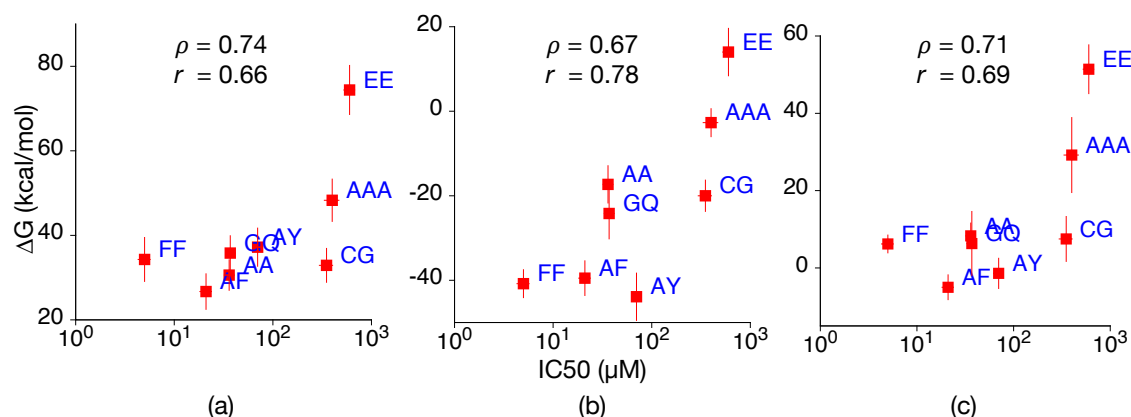
| Peptide   | $\Delta G$ (kcal/mol) |        |        |
|-----------|-----------------------|--------|--------|
|           | LIE                   | MMGBSA | MMPBSA |
| PhePhe    | 34.3                  | -40.8  | 6.2    |
| AlaPhe    | 26.7                  | -39.5  | -5.0   |
| AlaAla    | 30.6                  | -17.3  | 8.3    |
| AlaTyr    | 37.2                  | -43.9  | -1.4   |
| GlyGln    | 35.8                  | -24.2  | 6.3    |
| CysGly    | 32.9                  | -20.0  | 7.5    |
| AlaAlaAla | 48.3                  | -2.7   | 29.2   |
| GluGlu    | 74.4                  | 14.0   | 51.4   |

As these methods rely on empirical constants and implicit solvation, the magnitude of  $\Delta G$ s calculated may seem dubiously large (or small) and therefore should not be considered as the *true* binding energy.

than AutoDock and are able to separate GluGlu and triAla (the two worst substrates) from a group of hydrophobic dipeptides, AlaPhe, AlaTyr and PhePhe, (the best substrates). These methods also suggest a stronger linear correlation between  $\Delta G$  and the log of IC<sub>50</sub> with  $r$  value around 0.7.

Interestingly, all three end-point methods overestimated the binding affinity of CysGly. From the transport assay, CysGly can be classed as a poor substrate with a comparable IC<sub>50</sub> value to triAla. The predictions, however, rank this dipeptide amongst the good substrates such as AlaAla. When CysGly is removed from the calculation of  $\rho$ , correlation is improved, for example for LIE, from 0.74 to 0.89. One possible explanation for this poor estimation is that CysGly might bind in a different binding pose compared to other dipeptides, rendering the model and simulation to be inaccurate. While this hypothesis is not explored further for the CysGly peptide, it will be shown in Chapter 4, that some dipeptides may potentially bind in a tripeptide binding mode, owing to a large side chain.

One weakness of the end-point methods is the possible poor reproducibility of the  $\Delta G$  data (Wright et al., 2013). To evaluate how wide the distribution of  $\Delta G$ s predicted by these methods and if these variations mask the difference between peptides, the simulations of PepT<sub>St</sub> with AlaPhe and GluGlu were repeated for 50 times varying only the initially assigned atomic velocities. These two peptides were chosen to represent the best and worst substrates to investigate if the separation between the good and poor substrates can be reproduced. Assuming that all of these simulations have the same  $\tau$  as the previous simulations,



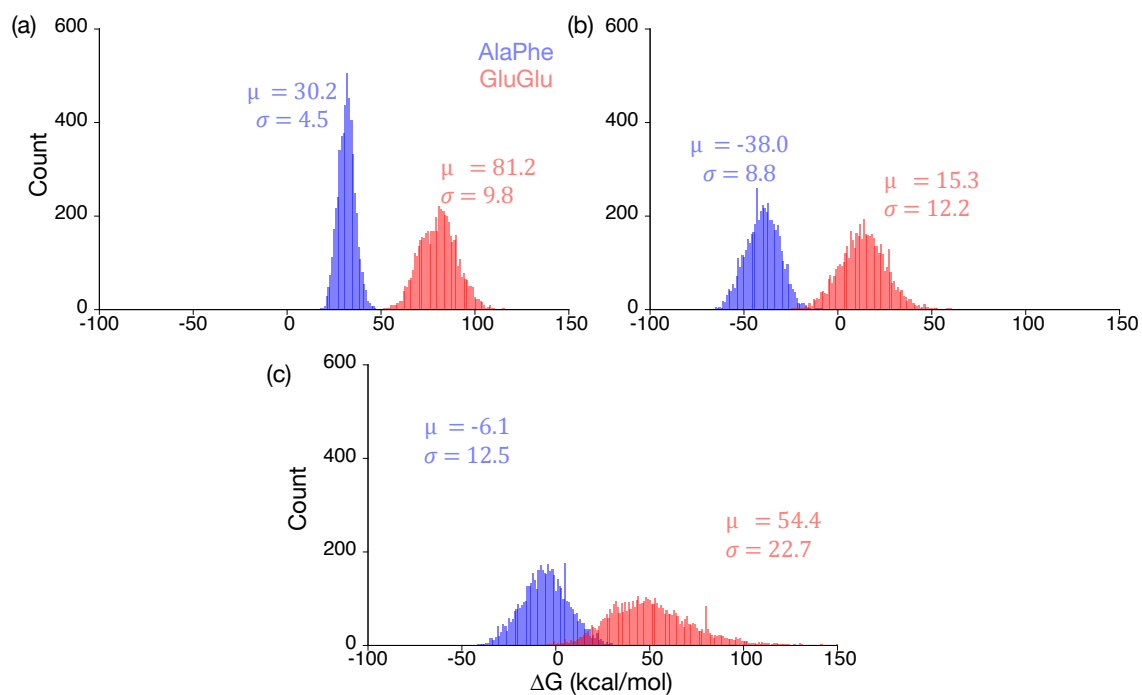
Figs. 3.9: Binding affinities of peptides to bacterial peptide transporter PepT<sub>St</sub>, as predicted by end-point methods; (a) LIE, (b) MMGBSA and (c) MMPBSA. Y-error bars indicate the standard error from de-correlated  $\Delta G$  data during the 1 ns MD simulations whereas X-error bars represent standard error from three repeat measurements.

all de-correlated  $\Delta G$  values generated along the 50 trajectories after 200 ps of equilibration period are plotted (Figure 3.10)\*.

The graphs show that there are indeed large variations produced by all three end-point methods for the replica simulations. LIE has the smallest standard deviation ( $\sigma$ ) of 4.5 and 9.8 kcal/mol for AlaPhe and GluGlu, respectively. MMGBSA and MMPBSA, on the other hand, can vary by as much as 10-20 kcal/mol, comparable to the results of Wright et al. (2013). For all methods, the variation for the GluGlu dipeptide is consistently larger than AlaPhe, perhaps due to the larger side chains that contribute to more flexibility in the binding site, and therefore more conformations sampled during the simulations. Despite these variations, however, the predicted means ( $\mu$ ) for both peptides are always separated far apart. For example, when LIE was used, the difference of the mean  $\Delta G$  values was around 50 kcal/mol. As such, no overlap exists between the  $\Delta G$  distributions of AlaPhe and GluGlu from the LIE calculations, and very marginal overlaps for MMGBSA and MMPBSA. These results suggest that the probability of predicting similar  $\Delta G$  values for ligands with different binding affinities is low even if the variability of the predictions is large. The end-point methods can therefore reproducibly distinguish between good and poor substrates of PepT<sub>St</sub>.

These results point towards the potential use of these end-point methods as a fast fil-

\*As these 50 simulations were not subject to the same statistical analysis to determine their  $t_{eq}$  and  $\tau$  values individually, there is a risk of including non-equilibrated and/or correlated data. However, the main point of these replica simulations is to gain a rough distribution of  $\Delta G$ , not the exact mean value for each simulation. Also, the long combined simulation time (50 ns) will hopefully diminish any errors resulting from the inclusion of non-equilibrated data.



Figs. 3.10: Distribution of  $\Delta G$  values as predicted by (a) LIE, (b) MMGBSA and (c) MMPBSA for PepT<sub>St</sub>-AlaPhe (blue) and PepT<sub>St</sub>-GluGlu (red). A 1 ns simulation for each dipeptide was repeated 50 times with different initial atomic velocities.  $\Delta G$  values computed along the trajectories are divided into 1 kcal/mol bins.  $\mu$  represents the mean and  $\sigma$  represents the standard deviation of the energy distribution, each quoted in kcal/mol.

tering step to categorise dipeptides and tripeptides into good and bad substrates of PepT<sub>St</sub>. As these techniques are relatively fast and require only moderate amounts of computational resources, one can also use them in a high-throughput manner to understand the general preference of peptide transporters. Ito et al. (2013) designed an empirical affinity prediction model for a yeast peptide transporter, Ptr2p, based on physicochemical properties of dipeptides and their transport data from a fluorescence-based uptake assay. Even though this prediction model is considerably faster than the end-point methods, it has been shown to be limited to dipeptides and dipeptide-mimetic compounds. Their results underline the weakness of empirically-derived prediction models, which only perform well for compounds with similar molecular and chemical properties. A prediction based on the end-point methods, on the other hand, would not be constrained in this way as it takes into account the structure of both the transporter and the ligand, and therefore should work for a wide range of compounds.

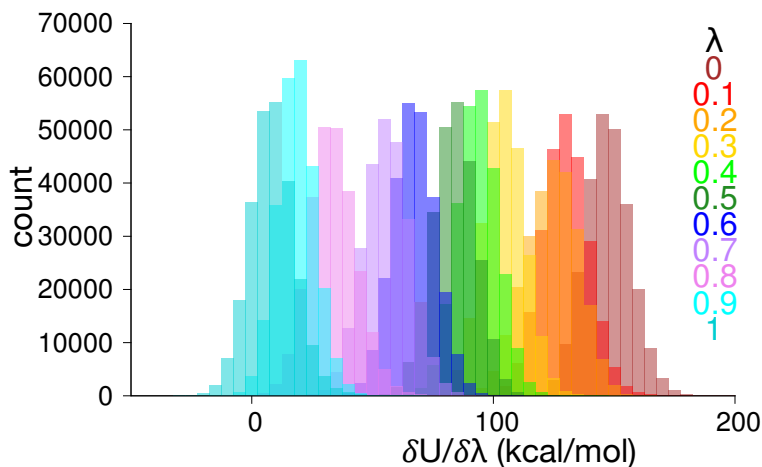
### 3.3.3 Thermodynamic integration improves prediction for hydrophobic dipeptides

Upon closer inspection on the results of the end-point methods (Figure 3.9), it was found that the predictive ability deteriorates when only substrates with similar IC<sub>50</sub> values are considered. This is true for a group of hydrophobic dipeptides such as AlaAla, AlaPhe, AlaTyr and PhePhe, which all have IC<sub>50</sub> values less than 100  $\mu$ M. For this subset of peptides, the predicted vs. experimental  $\rho$  correlation value drops to 0.00 (random ranking) and the  $r$  correlation to -0.27 (negative linear correlation) for the MMPBSA method. The decrease in accuracy suggests that while the end-point methods are adequate at distinguishing good and weak substrates, they cannot separate good substrates from one another.

As dipeptides share a similar structural scaffold, they are excellent candidates for alchemical transformation MD simulation for the more rigorous TI calculations. To investigate whether TI would improve affinity prediction for this subset of peptides, the relative free energy of binding,  $\Delta\Delta G$ , of each dipeptide with respect to AF was calculated and this was subsequently added to the original  $\Delta G$  value obtained from the MMPBSA method<sup>†</sup>. Simulations of length 5 ns were performed for each  $\lambda$  window. They were first analysed using the same statistical methods used for the short simulations of end-point methods to

<sup>†</sup>Technically, there is no need to add the  $\Delta\Delta G$  value from TI calculation to the  $\Delta G$  calculated by the end-point methods as it will not affect the value of correlation coefficients of the TI results to the experimental data. This step was performed, nevertheless, to make graphical comparison to the end-point methods easier as is shown in Figure 3.12.





Figs. 3.11: Distribution of  $\frac{\delta U}{\delta \lambda}$  for the alchemical transformation of AlaPhe to PhePhe. Data is taken from step 1 (discharging) of the PepT<sub>St</sub>-bound system. The overlapping energy distribution between neighbouring  $\lambda$  windows suggests that adequate sampling has been achieved to calculate a reliable  $\Delta G$  value.

determined the  $t_{eq}$  and  $\tau$  values. All peptide-only simulations (unbound state) reached equilibrium immediately after the production simulations began, and the correlation times,  $\tau$ , are very short ( $\sim 10$  ps), which is characteristic of a simple system. For the PepT<sub>St</sub>-bound systems, step 1 and step 3, which involves decoupling and re-coupling of the electrostatic forces take a longer time to equilibrate than step 2 (disabling/enabling Lennard-Jones terms). For example, in the AlaPhe to AlaTyr transformation, the former steps have a  $t_{eq}$  of around 1 ns, while the latter equilibrates almost instantly. Additionally, all of these simulations have relatively large correlation times ( $\sim 100$ -200 ps), which is reflected in the larger statistical errors (see Table 3.5 for full results).

To get a reliable  $\Delta G$  value for each step of the transformation, adequate sampling within each window must be performed. To assess this, the distribution of energy for each value of  $\lambda$  within each step is plotted (an example is shown in Figure 3.11). In all 3 alchemical simulations, it was found that the  $\frac{\delta U}{\delta \lambda}$  values overlap between neighbouring  $\lambda$  windows through all steps, which indicates that substantial regions of the free energy landscape between the initial and final structures have been explored. This was achieved by maximising the similarity between states by decoupling atoms from  $C_\beta$  instead of  $C_\alpha$  of the amino acid residue. Although converting the ring structure of phenylalanine into hydrogen (alanine) involves relatively large changes, by using the dual topology approach, the need to break or close the ring alchemically is avoided. Instead, the ring was removed entirely and a soft core potential has been shown to accurately deal with this transformation (Shirts et al., 2003). To test the reproducibility of the TI method, several alchemical transformation MD simulations were repeated, both in forward and reverse directions. All of the  $\Delta \Delta G$  values from these replicate simulations were in agreement with the original results (Appendix A).

Table 3.5: TI results for the alchemical transformations of (a) AlaPhe to AlaAla, (b) AlaPhe to PhePhe and (c) AlaPhe to AlaTyr. All values are quoted in kcal/mol with a precision of 0.1 kcal/mol. The calculated errors represent errors of integration of  $\frac{\partial U}{\partial \lambda}$ .

| (a)   |                           |                             |                   |
|-------|---------------------------|-----------------------------|-------------------|
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta \Delta G$ |
| 1     | $146.2 \pm 0.3$           | $144.1 \pm 0.1$             | $2.1 \pm 0.3$     |
| 2     | $-2.9 \pm 0.1$            | $-4.2 \pm 0.0$              | $1.4 \pm 0.1$     |
| 3     | $-147.5 \pm 0.1$          | $-144.0 \pm 0.1$            | $-3.4 \pm 0.1$    |
| Total |                           |                             | $0.1 \pm 0.4$     |

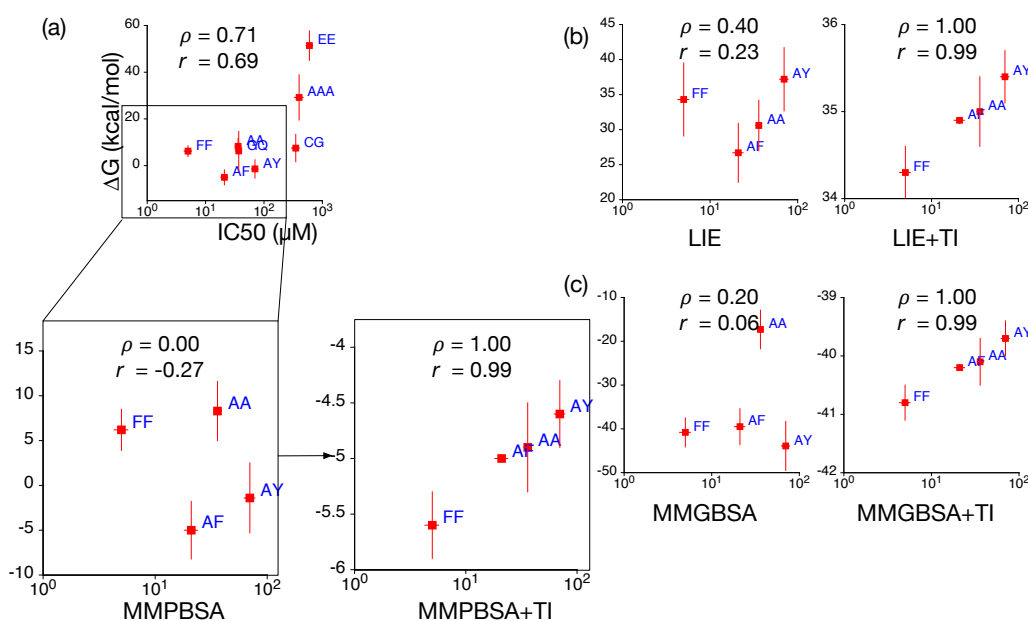
  

| (b)   |                           |                             |                   |
|-------|---------------------------|-----------------------------|-------------------|
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta \Delta G$ |
| 1     | $82.9 \pm 0.1$            | $68.0 \pm 0.0$              | $14.9 \pm 0.1$    |
| 2     | $3.0 \pm 0.2$             | $4.7 \pm 0.0$               | $-1.7 \pm 0.2$    |
| 3     | $-90.6 \pm 0.2$           | $-76.9 \pm 0.1$             | $-13.7 \pm 0.2$   |
| Total |                           |                             | $-0.6 \pm 0.3$    |

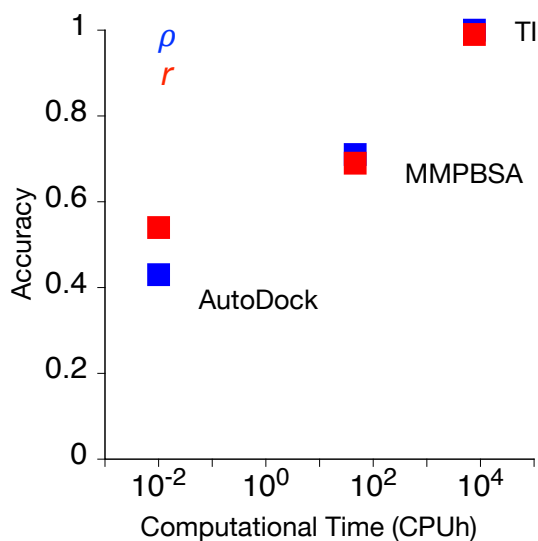
  

| (c)   |                           |                             |                   |
|-------|---------------------------|-----------------------------|-------------------|
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta \Delta G$ |
| 1     | $-83.9 \pm 0.3$           | $-91.3 \pm 0.0$             | $7.4 \pm 0.3$     |
| 2     | $-1.0 \pm 0.0$            | $-0.6 \pm 0.0$              | $-0.5 \pm 0.0$    |
| 3     | $63.3 \pm 0.1$            | $69.7 \pm 0.0$              | $-6.5 \pm 0.1$    |
| Total |                           |                             | $0.4 \pm 0.3$     |

As illustrated in Figure 3.12, TI calculation indeed improves the results and able to replicate the experimental ranking ( $\rho = 1.00$ ). Also evident from these results is the ability of this method to discern small structural differences between ligands, for example between a hydrogen and a hydroxyl group during the calculation of  $\Delta\Delta G$  of AlaTyr. It is worth noting, however, the immense amount of computational resources that was employed for these calculations compared to other methods. Each of the alchemical transformations requires 330 ns of simulation time as compared to 2 ns for any of the end-point methods. The reason for such a marked difference is that the end-point methods only need simulations of the “end-states” i.e. the bound and unbound conformations, whilst for TI, the intermediate unphysical states must also be explicitly simulated. In principle, apart from using TI to measure the relative binding energy of a peptide with respect to another peptide, one can also calculate the *absolute* binding free energy of each peptide by decoupling the electrostatic and van der Waals of the whole molecule, effectively turning the compound into a group of dummy atoms. As this task involves a large transformation, much longer simulations are needed for a fully converged result, and therefore this approach is not pursued in this thesis.



Figs. 3.12: Binding affinity predictions refined by rigorous free energy calculations. (a) Enlarged image highlights the inability of the end-point method (MMPBSA) to discriminate substrates of similar  $IC_{50}$  values.  $\Delta\Delta G$  values calculated using the TI method is added to correct the MMPBSA results. Y-error bars for the MMPBSA+TI graph indicate the error of integration whereas X-error bars represent standard errors from three repeat measurements. (b) and (c) show similar assessments with the LIE and MMGBSA methods, respectively.



Figs. 3.13: The performance of all binding affinity prediction methods and the computational time spent on each of them. The degree of accuracy here is represented by the correlation coefficients of predicted  $\Delta G$ s (or in the case of TI,  $\Delta\Delta G$ s) with experimental  $IC_{50}$ s. The measured computational hour was based on a quad core processor. The three end-point methods have similar  $\rho$  and  $r$  values and require the same amount of computational resources, therefore only the data for MMPBSA are shown on the graph.

### 3.4 Conclusion

A range of methods to predict the binding affinity of peptides to a bacterial peptide transporter, PepT<sub>St</sub>, were assessed and the results were compared to experimental data in the form of  $IC_{50}$  values from *in vitro* competition uptake assay. It is no surprise that the more computational input fed into the methods, the more accurate the results become. A scoring function, which uses only the static structures of the protein and ligand, produces glaringly inferior results. When some dynamics during binding is taken into account, just by considering the bound and unbound states in the end-point methods, an encouraging increase of accuracy for the predictions was observed. These approaches are able to separate the well- and poorly transported substrates. The theoretically rigorous TI method, for which simulations of intermediate states are required, further refines the results by discerning peptides with small structural differences and similar transport behaviours. To quantify and compare the exact amount of resources required for these methods, computational usage in CPUh was estimated based on the GROMACS performance for MD simulation of AlaPhe bound PepT<sub>St</sub> using an Intel® quad core Xeon® processor (Figure 3.13). Affinity prediction by AutoDock does not require any MD simulation and as each docking protocol takes around 30 s, this is plotted as  $10^{-2}$  CPUh in the graph.

It is evident from these results that simulation-based binding affinity prediction methods require a dramatically larger amount of resources than a scoring function. The CPUh needed to perform MMPBSA is almost 4 order of magnitude more than for AutoDock, whilst TI is 6 order of magnitude more expensive. With increasing investment of computational time,

however, the level of accuracy also increases. Interestingly, the increase in correlation values (notably for  $\rho$ ) is more significant from AutoDock to MMPBSA than from MMPBSA to TI, which indicates that the end-point methods are an excellent compromise between good performance and low cost. Nevertheless, the TI method is indeed very effective at accurately measuring energetic differences involving small changes in structure, which is not perceptible by any of the end-point methods. It is therefore more pragmatic to adopt a hybrid approach, whereby the end-point methods are used as a “first-pass” prediction, perhaps to rule out poor substrates, and then utilise the more rigorous TI for further refinement of specific predictions.

Collectively, these results support the hypothesis that computationally-determined binding affinities are useful to rank ligands of the proton-coupled oligopeptide symporters. These validated approaches shall be used in the next chapters to (i) study bacterial peptide transporter, PepT<sub>St</sub> in greater detail (Chapter 4) and (ii) test it with the medically relevant human transporter, PepT1 (Chapter 6).

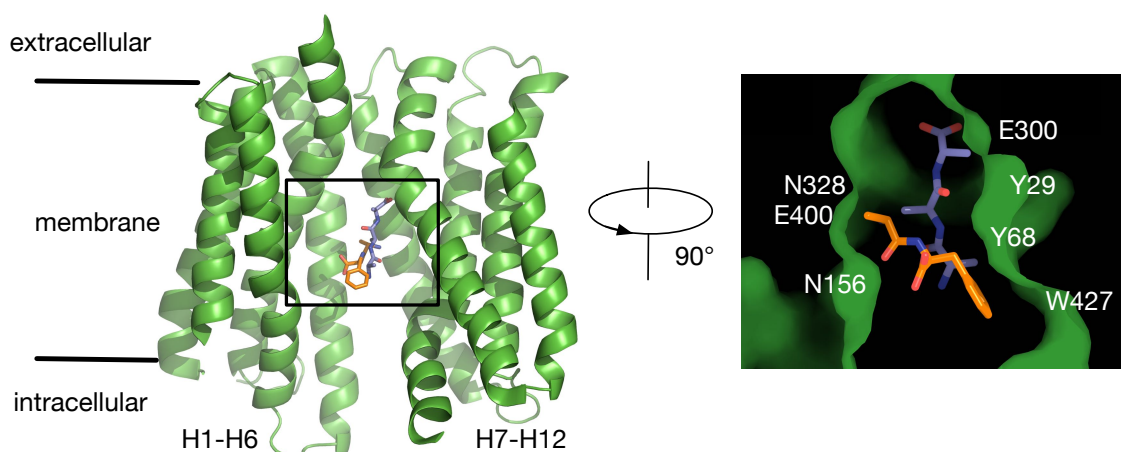
## Chapter 4

# Accurate prediction of peptide binding in a bacterial POT transporter

The binding of all 400 possible dipeptides to PepT<sub>St</sub> was modelled. The linear interactions energy (LIE) method was used to classify these dipeptides into good and poor substrates. The most preferred substrates are neutral peptides, whilst acidic peptides bind less well. Interestingly, the order of amino acid residues does influence affinity, and as such the contribution of each side chain towards binding selectivity was investigated using the thermodynamic integration (TI) method. The results suggest that the N-terminal side chain is more sensitive to perturbation than the C-terminal side chain. For positively charged dipeptides, however, the LIE and TI results show discrepancies to experimental data, suggesting they may bind differently. Indeed, modelling the binding of AlaLys and LysAla based on the triAla complex structure restored agreement, which indicates that basic dipeptides, perhaps due to their extended side chains, may adopt a tripeptide-like pose. For tripeptides, a bulky residue at the C-terminus results in unfavourable binding, consistent with experiments showing no transport for large tripeptides and drugs.

### 4.1 Substrate multispecificity in the POT transporters

Central to the functioning of the POT transporters is their ability to recognise and accommodate a wide variety of ligands, which is achieved by having a promiscuous binding site. Despite the promiscuity, the POT transporters retain some specificity, such that there are



Figs. 4.1: Crystal structures reveal two distinct binding orientations exist for PepT<sub>St</sub>, one for the dipeptide AlaPhe (orange) and another for the tripeptide tri-Ala (purple). The position of key residues are shown on the right. For clarity, H2, H11, HA and HB are removed. Figure adapted from Lyons et al. (2014).

certain structural features to which a compound must adhere in order to become a substrate (see §1.2.2). Due to its enormous potential in rational drug design for improved bioavailability, great efforts have been put forth in recent years towards building a substrate binding template for the human PepT1. For example, a simplified two-dimensional substrate model was proposed by Bailey et al. (2000) suggesting that a high affinity substrate should have its amino and carboxylic acid groups separated by about 6 Å, which matches the length of the backbone of a dipeptide. This was followed by more detailed three dimensional pharmacophore models (Biegel et al., 2005; Vig et al., 2006; Foley et al., 2010; Pedretti et al., 2008).

A growing body of evidence, however, indicates that the POT transporters do not operate using a single mechanism and therefore a universal binding model might not be sufficient. The peptide bound crystal complexes of a bacterial homolog, PepT<sub>St</sub>, clearly elucidated two distinct binding orientations (Figure 4.1): the dipeptide AlaPhe binds “horizontally” with respect to the plane of the membrane whilst the tripeptide triAla binds “vertically” (Lyons et al., 2014) (discussed in §1.2.4). Despite occupying distinct poses, AlaPhe and triAla essentially interact with similar key residues of the transporter, although the latter adopts a less tightly coordinated position that could explain its lower affinity. Further corroborating the notion of multiple transport mechanisms in the POT family, Parker et al. (2014) recently suggested that PepT<sub>St</sub> employed two different proton:peptide coupling stoichiometries—the transport of a dipeptide is coupled to four or five protons whereas for a tripeptide, only three

protons are used. These recent studies hint at two distinct transport strategies utilised by this transporter, one for dipeptides and another for tripeptides.

Even though this evidence explains in part how the POT transporters achieve their promiscuity, it raises several important questions, e.g. how do the numerous other di- and tripeptides and, more importantly, drugs interact with the transporter, and how do their affinities compare to each other? Experimental data from transport assays for PepT<sub>St</sub> are currently limited to the eight peptides shown in Table 3.1. Having determined using these test set of peptides which free energy methods are most accurate, these methods shall now be used to make some predictions, which will be tested experimentally. One of the end-point methods, linear interaction energy (LIE), is chosen to broadly classify all dipeptides into good and poor binders. Following these results, the theoretically rigorous method, thermodynamic integration (TI), is utilised to understand the contributions of each side chain of a peptide substrate to its overall binding selectivity.

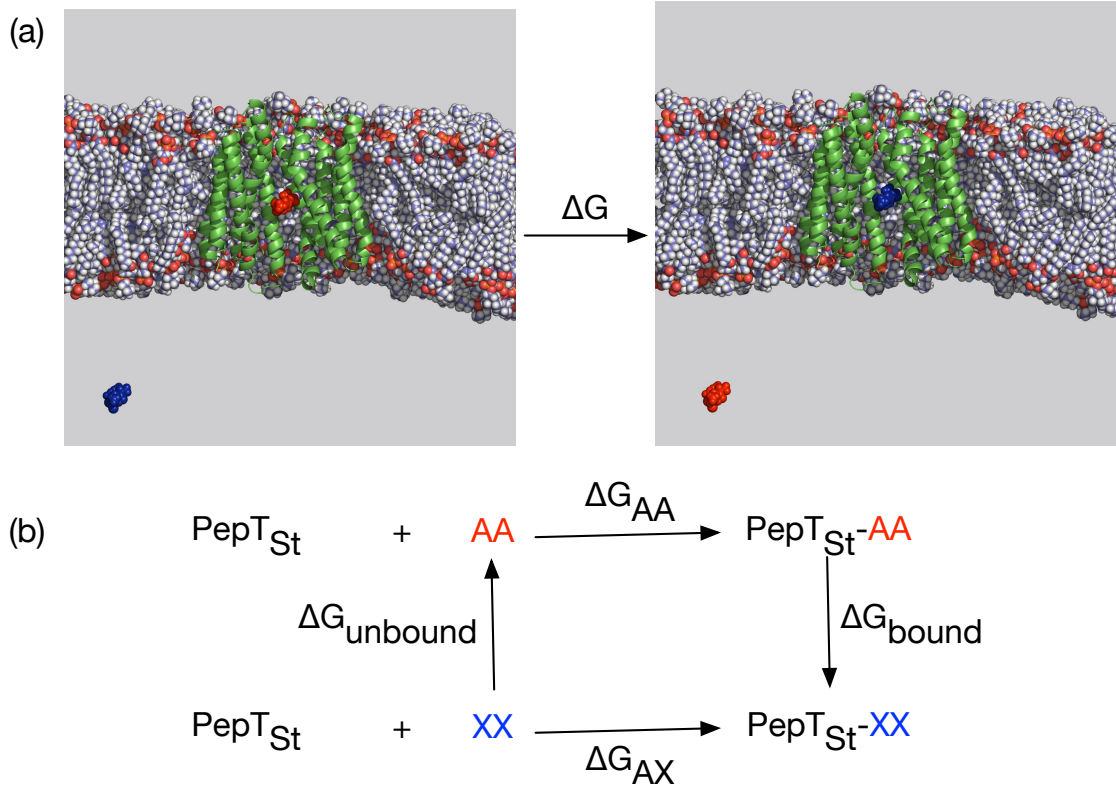
## 4.2 Methods

First, the binding of all 400 dipeptides to PepT<sub>St</sub> was modelled as outlined in §3.2.1. These models were used in the prediction of the free energy of binding,  $\Delta G$ , using the LIE method, which was selected over MMGBSA and MMPBSA for two reasons: i) it produces the most accurate ranking compared to the experimental data (see Figure 3.9) and ii) it is integrated in GROMACS via the `g_lie` tool. The parameters for MD simulations and free energy calculations used in Chapter 3 were repeated here, unless otherwise stated.

### 4.2.1 Thermodynamic integration with a change in net charge

The TI method was applied to investigate the sensitivity of each side chain of a dipeptide substrate to changes. PepT<sub>St</sub>-AlaAla model from the previous chapter was used for the starting coordinates. Alchemical simulations were performed to convert the alanyl side chain at either the N- or the C-terminus to phenylalanine, aspartate, glutamate or lysine. The TI method was used to calculate relative free energy of binding,  $\Delta\Delta G$ , of the resultant dipeptides with respect to AlaAla. The same approach was applied to study the side chain of a tripeptide substrate, for which the crystal structure of PepT<sub>St</sub>-triAla was used as the starting coordinates and the N-terminal side chain (first residue) or the C-terminal side chain (third residue) was morphed into either phenylalanine or tyrosine.





Figs. 4.2: TI calculation with a change in the total charge of the substrate. (a) While in the bound state, AlaAla (red) is converted to another dipeptide XX (blue), the reverse transformation is simultaneously performed on XX in the bulk to ensure that the total charge of the system does not change throughout the calculations. PepT<sub>St</sub> is in green cartoon representation and lipids are in spheres with phosphate atoms shown in red. (b) A thermodynamic cycle for these transformations, simplified from Figure 3.3.

$$\Delta G_{AA} + \Delta G_{\text{bound}} - \Delta G_{XX} + \Delta G_{\text{unbound}} = 0$$

$$\Delta \Delta G = \Delta G_{XX} - \Delta G_{AA} = \Delta G_{\text{bound}} + \Delta G_{\text{unbound}} \quad (4.1)$$

For transformations to the side chain of aspartate, glutamate and lysine, there is a change in the total charge of the peptide substrate. The PME method requires that the net charge of the system remains neutral throughout the entire simulation. To this end, a similar technique as described in Rashid et al. (2013) was adopted. Both bound and unbound alchemical simulations were performed in the same simulation box (Figure 4.2). As the PepT<sub>St</sub>-bound AlaAla was converted to another dipeptide, XX, a reverse transformation was simultane-

ously conducted in the bulk solution (XX converted to AlaAla). The former transformation resulted in a gain of charge and the latter a loss of charge, essentially cancelling each other out to ensure that the net charge of the system remained unchanged throughout the simulation. To get the value of  $\Delta\Delta G$ , instead of calculating the difference in  $\Delta G$  of the bound and unbound simulations as before, the sum of both terms is calculated (Equation 4.1), which in practice equals to the  $\Delta G$  of the transformation of the whole system (Figure 4.2(a)). Following the same protocol as in Chapter 3, this transformation was also divided into three steps, each of which consisted of eleven simulations of different  $\lambda$  values. This approach therefore reduced the total number of simulations by half (33 instead of 66).

For predictions with the alternative binding mode of lysine containing dipeptides, the crystal structure of PepT<sub>St</sub>-triAla was used as the starting coordinates. An alanine residue was removed from triAla either from the N- or the C-terminus. Using the same “one-box” approach, either the N- or the C-terminal side chain of the resulting AlaAla ligand was transformed to lysine. All simulations were subject to the reverse cumulative averaging analysis (Yang et al., 2004), as described in §3.2.4.

### 4.2.2 Contact Analysis

To understand the structural basis of the different contributions from each side chain to the binding selectivity of a dipeptide, a contact analysis was performed using a VMD plugin, *interquant*. This program quantifies the amount of interactions between the physical centre of the analysis, or the “origin”, with a specified group of atoms surrounding it, or the “target”, based on a stipulated distance cutoff. For this analysis, the side chain of interest of the peptide substrate was set as the origin, whilst all residues of PepT<sub>St</sub> were set as the target. A cut-off radius of 3.5 Å was implemented. The contact analyses of 20 simulations of X-Ala was averaged to elucidate the residues in the transporter to which the N-terminal side chain interacts, and for the C-terminal side chain, the simulations of Ala-X were used.

### 4.2.3 Competition transport assay

To validate the computational predictions, Joanne Parker performed proton-driven competition transport assays for two pairs of dipeptides i.e. AlaAsp, AspAla and AlaLys, LysAla, following methods described in Solcan et al. (2012). Proteoliposomes with 5 µg protein in internal buffer (20 mM Potassium Phosphate, 100 mM Potassium Acetate, 2 mM Magnesium Sulphate, pH 6.5) were diluted into external buffer (120 mM Sodium Phosphate, 2 mM Magnesium Sulphate, pH 6.5) containing 25 µM <sup>3</sup>H labelled AlaAla (reporter substrate) in

the presence of increasing concentration of peptides (substrate of interest). Reaction was started by the addition of 10  $\mu\text{M}$  valinomycin and stopped after 4 minutes through dilution into 0.1 M Lithium Chloride and collected on nitrocellulose filters prior to scintillation counting. The radioactive  $^3\text{H}$  signal was converted to molar concentrations of peptides using standard curves for each substrate.

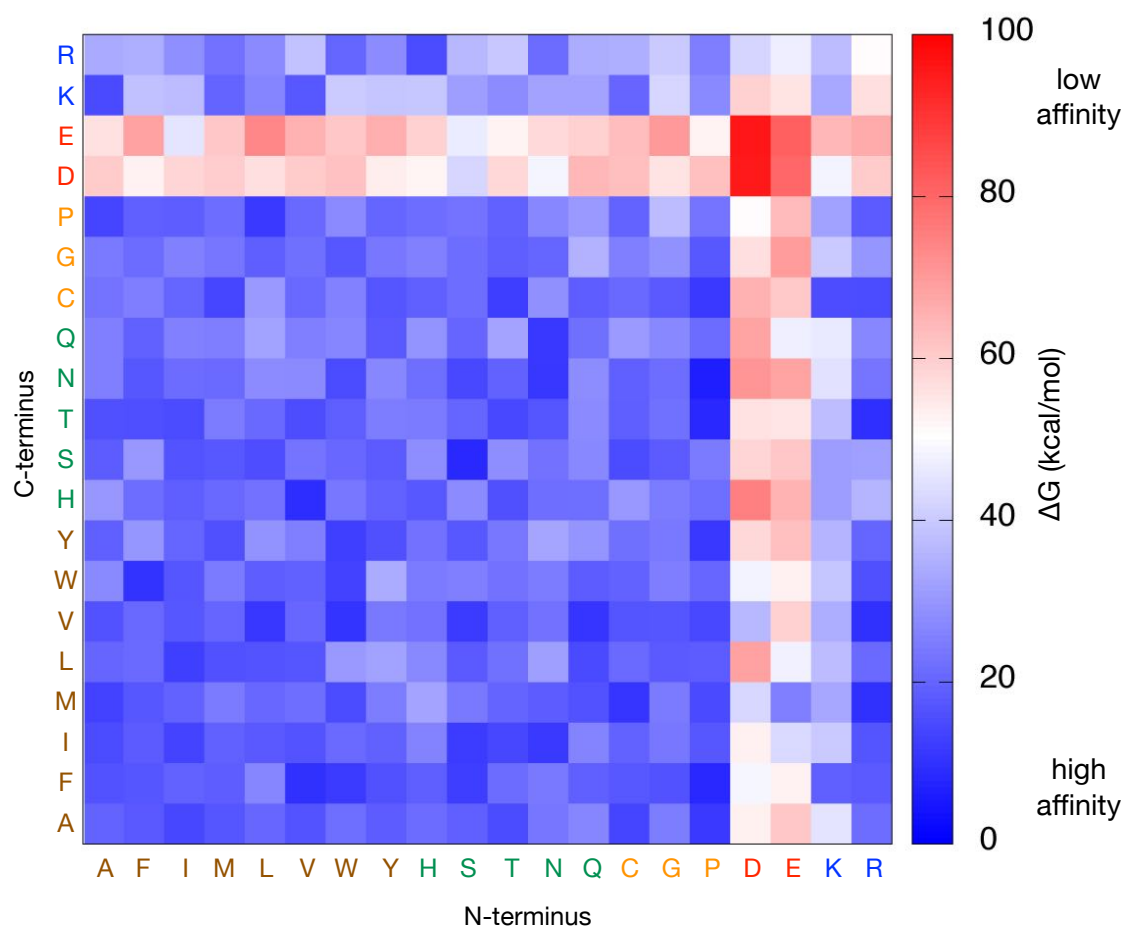
## 4.3 Results and Discussion

First, the general preference of PepT<sub>St</sub> was investigated by classifying all naturally occurring dipeptides into high- and low-affinity substrates. Following the prediction that the sequence of residues affects binding affinity, the contributions of residues at specific position towards the overall selectivity were calculated and the results validated by experimental transport assays. A new model for positively charged dipeptides was proposed to justify the discrepancies of these results with experiments. Finally, a similar study on tripeptides revealed that bulky residues might aggravate binding.

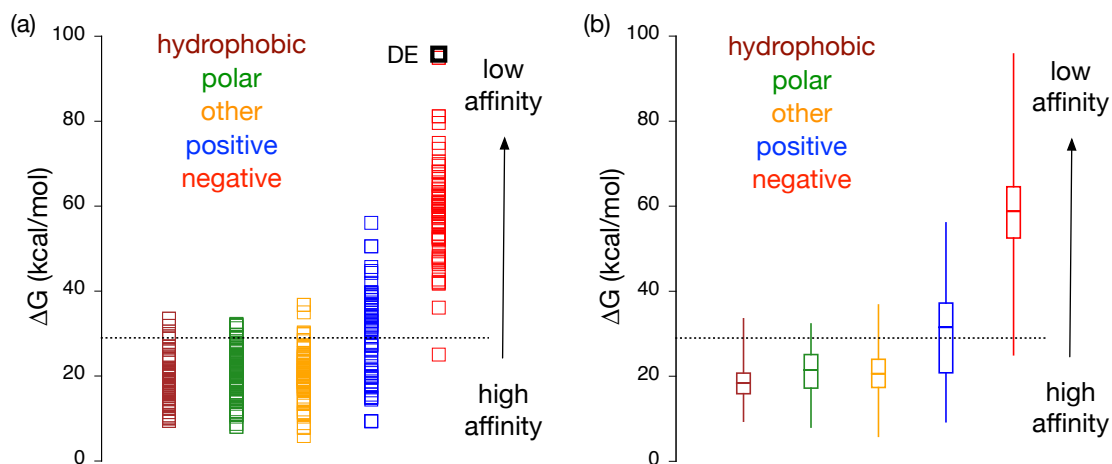
### 4.3.1 Neutral dipeptides favoured over charged.

Based on the 20 physiological amino acids, the binding of all 400 possible dipeptides to PepT<sub>St</sub> was modelled using the co-crystal structure of PepT<sub>St</sub>-AlaPhe. For each peptide, a short 1 ns simulation was performed using the same parameters used in the Chapter 3. The LIE method was then employed to estimate  $\Delta G$  (Figure 4.3). For this calculation, it was assumed that all simulations became fully equilibrated after 200 ps and de-correlated after 10 ps, which was true for most simulations of the peptide test set. Due to the empirical nature of the LIE method, the magnitude of  $\Delta G$ s predicted are all positive and, in some cases, may seem dubiously large. Therefore, they should not be considered as the *true* binding energy of the ligands to the transporter, and are used only for qualitative comparison between different ligands.

To understand the preference of the transporter, these peptides were grouped according to the chemical nature of the amino acid residues and their  $\Delta G$ s replotted according to these groups (Figure 4.4). The results suggest that neutral dipeptides made of hydrophobic, polar and other (cysteine, glycine, proline) amino acids are the preferred substrates for PepT<sub>St</sub> as they show the lowest  $\Delta G$  values. Almost all of them have  $\Delta G$ s below the overall average. Comparing between these uncharged peptides, the hydrophobic ones appear to have the smallest interquartile range which fall well below 20 kcal/mol, suggesting that they are the



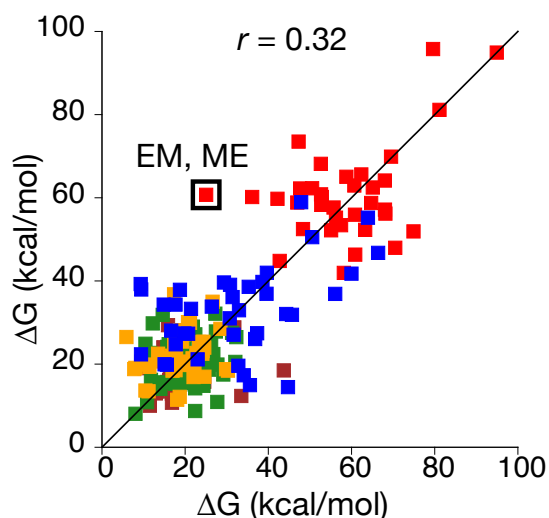
Figs. 4.3: The binding free energy,  $\Delta G$ , of all 400 physiological dipeptides to PepT<sub>St</sub> as calculated by an end-point method, LIE. High affinity binding (low  $\Delta G$ ) is shown in blue whilst low affinity binding (high  $\Delta G$ ) in red. Amino acid residues at the N- and C-terminal are coloured based on their chemistries; hydrophobic in brown, polar in green, negatively-charged in red, positively-charged in blue and others in orange.



Figs. 4.4: Predicted  $\Delta G$  of all 400 dipeptides plotted based on their chemical properties. (a) A scatter plot of all 400  $\Delta G$  values from Figure 4.3. (b) A box-and-whisker plot to highlight the position of the median and the distribution. Dotted line represent the overall mean.

best substrates. Polar and “other” dipeptides have a slightly more spread-out distribution of  $\Delta G$ s, although they are still within the high-affinity binding region. Both positive- and negatively charged dipeptides, on the other hand, are predicted to have highly varied binding affinities, with both distributions spanning more than twice that of the neutral peptides. Acidic dipeptides bind least well with almost all  $\Delta G$  values lying above the overall average. In contrast, almost half of the basic dipeptides appear to bind as tightly as their neutral counterparts, whilst the other half are poor binders with affinities similar to a negatively charged peptide.

Collectively, these results suggest that PepT<sub>St</sub> has similar selectivity profile compared to other POT transporters. The mammalian PepT1 and PepT2 transporters also have higher affinity for neutral peptides, especially hydrophobic, and transport charged peptides less well (Gebauer and Hartrodt, 2003; Vig et al., 2006; Biegel et al., 2006). Ito et al. (2013) performed a similar “high-throughput” assessment with a peptide transporter from *Saccharomyces cerevisiae*, Ptr2p, using a fluorescence based uptake assay on more than 300 dipeptides. Classifying these dipeptides into high- and low-affinity groups, they found that peptides with a hydrophobic residue like tryptophan are present in high frequency in the former, whilst the latter are mostly populated by negatively charged dipeptides. The results in this thesis also echo the study of DtpT, a peptide transporter from *Lactococcus lactis*, by Fang et al. (2000) that found the highest affinity substrates are peptides with at least one hydrophobic residue.



Figs. 4.5: Predicted binding free energy,  $\Delta G$  of all 400 dipeptides plotted against their sequence-reversed equivalents, e.g. GluMet and MetGlu. These peptides are again coloured according to their chemical properties as in Figure 4.4. The line drawn represents perfect linear correlation.

The predictions with basic dipeptides, however, are in conflict with the results of an *in vitro* transport assay. Solcan et al. (2012) showed that a positively charged dipeptide, LysLys has a very high  $IC_{50}$  value ( $\sim 2000 \mu M$ ), even when compared to other poor substrates like GluGlu ( $IC_{50}=600 \mu M$ ). Similarly, based on whole cell uptake assays, PepT1 shows very poor affinity for the same peptide as indicated by the very high  $K_i$  value (Eddy et al., 1995; Terada et al., 2000b). One possible reason for this discrepancy is that the binding model currently applied for all dipeptides might not be suitable for these basic dipeptides, which shall be discussed later in the chapter.

### 4.3.2 The order of residues does matter.

In an effort to build a substrate binding model for PepT1, Foley et al. (2010) suggested that the binding pocket is contiguous and therefore substrates are predominantly recognised by their peptide backbone rather than their side chains. This conclusion was derived based on the observation that AlaTrp and TrpAla have similar affinities to the transporter. As such, they suggested the sequence of amino acids making up a dipeptide will not affect its affinity. To check if the current results with PepT<sub>St</sub> agree with this model, the  $\Delta G$  values for the 400 dipeptides are replotted in pairs, each comprised of  $\Delta G$ s for two dipeptides composed of the same residues but in the opposite sequence, i.e. Ala-X vs. X-Ala. If order did not matter, then all the points should fall on a single line indicating that these  $\Delta G$  pairs are highly correlated.

Interestingly, there is a low correlation ( $r = 0.32$ ) between the  $\Delta G$  values of dipeptides made of the same residues, as illustrated by the widespread scatter of the points in Fig-

Table 4.1: Correlation of  $\Delta G$  for dipeptides with their sequence-reverse equivalents based on chemical groups.

| Chemistry   | $r$  |
|-------------|------|
| Hydrophobic | 0.24 |
| Polar       | 0.21 |
| Other       | 0.23 |
| Positive    | 0.45 |
| Negative    | 0.46 |

ure 4.5. Some of these peptide pairs have significantly different  $\Delta G$ s, e.g. GluMet and Met-Glu. This observation is true regardless of the chemical nature of the residues (Table 4.1). Peptide pairs with a large difference in  $\Delta G$  values, however, have very similar Lennard-Jones and Coulomb potentials profiles in their unbound state (Table 4.2), which suggests that these  $\Delta G$  differences indeed arise from the protein-bound state. The overall binding affinity of a dipeptide to PepT<sub>St</sub> is therefore influenced by the order of amino acids. It is also possible that the amino acid side chain at the N- and C-termini contributes differently towards the selectivity of binding and play distinct roles during peptide recognition.

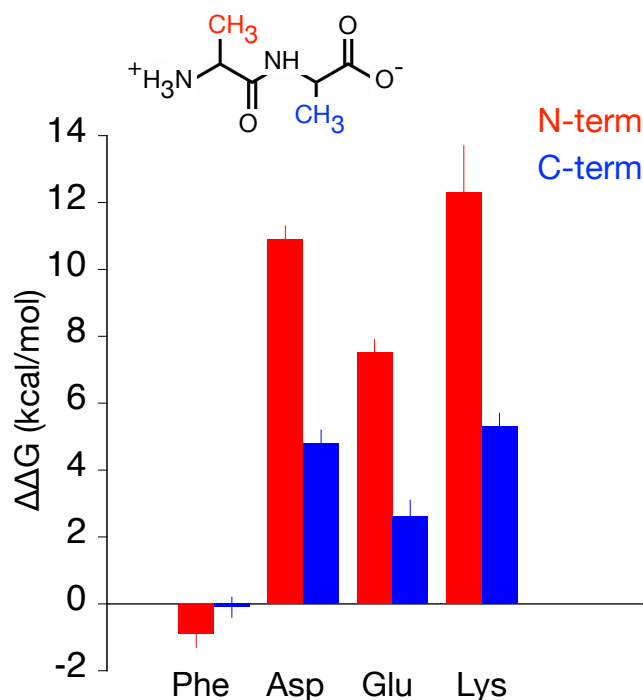
Table 4.2: Lennard-Jones (vdw) and Coulomb (el) potentials in the unbound and bound states of peptide ligands.

| Peptide | $U_{\text{unbound}}$ |        | $U_{\text{bound}}$ |        | $\Delta G$ |
|---------|----------------------|--------|--------------------|--------|------------|
|         | vdw                  | el     | vdw                | el     |            |
| EM      | 1.0                  | -206.1 | -16.7              | -147.2 | 26.3       |
| ME      | 0.9                  | -209.6 | -15.9              | -82.2  | 60.7       |
| KA      | -0.4                 | -165.9 | -12.9              | -72.0  | 60.7       |
| AK      | -0.6                 | -164.1 | -14.3              | -130.2 | 14.5       |
| WL      | -9.2                 | -142.2 | -32.7              | -30.7  | 43.7       |
| LW      | -12.1                | -135.9 | -21.4              | -89.0  | 18.5       |

Data extracted using the Gromacs `g_energy` tool and are shown in kcal/mol. The LIE method uses these data to calculate  $\Delta G$  based on Equation 2.15.

### 4.3.3 Dipeptide binding sensitive to the N-terminal side chain

To elucidate the contributions of each side chain of the peptide substrate to the binding selectivity, a series of alchemical transformation MD simulations and TI calculations were



Figs. 4.6: Relative free energy of binding,  $\Delta\Delta G$  of alchemically transforming the N- or C-terminus of the AlaAla dipeptide into either phenylalanine, aspartate, glutamate or lysine. Error bars indicate errors of integration of  $\frac{\partial U}{\partial \lambda}$ .

performed, whereby the N- or C-terminal side chain of AlaAla was transmuted into phenylalanine, aspartate, glutamate or lysine. A different simulation setup (described in details in the Method section) was used for alchemically converting alanine to a charged residue to preserve the total charge of the system. Each simulation was first subject to a reverse cumulative averaging statistical analysis (Yang et al., 2004) to determine its equilibration and correlation times, and from that, its statistical error. All results are shown in Figure 4.6 and Table 4.3.

A small negative  $\Delta\Delta G$  value was obtained when either one of the alanine side chains was substituted with a phenylalanine, which indicates a stronger binding for the hydrophobic residue. A transformation into a charged residue, on the other hand, resulted in a positive  $\Delta\Delta G$  and therefore a weaker binding, in agreement with the LIE results discussed earlier. What is more interesting, however, is that when the N-terminus was altered, the magnitude of  $\Delta\Delta G$  was always larger than the identical change at the C-terminus. The  $\Delta\Delta G$  value for an N-terminus transformation is at least twice as large as the equivalent transformation of the C-terminus. These results suggest that the N-terminal residue is more sensitive to changes and hence is crucial in determining the binding selectivity of a dipeptide to  $\text{PepT}_{\text{St}}$ .

To investigate the structural reasoning behind these results, the key residues in the bind-



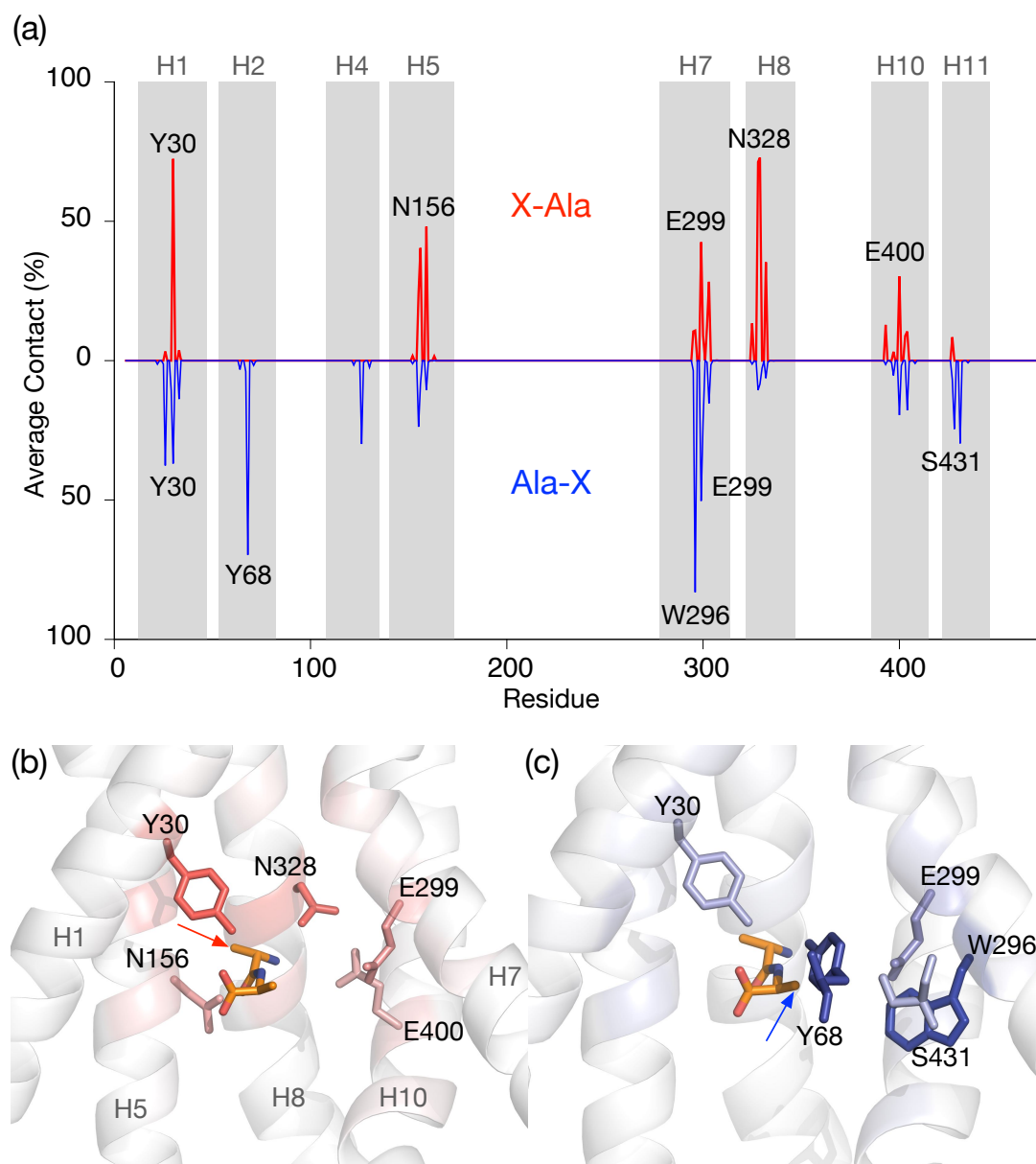
Table 4.3: TI results for the alchemical transformations of the N- or C-terminal side chain of AlaAla into (a) phenylalanine, (b) aspartate, (c) glutamate and (d) lysine. All values are quoted in kcal/mol. The calculated errors represent errors of integration of  $\frac{\partial U}{\partial \lambda}$ .

| (a)   |                  |                  |
|-------|------------------|------------------|
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $13.2 \pm 0.2$   | $-2.1 \pm 0.1$   |
| 2     | $-1.4 \pm 0.2$   | $-2.1 \pm 0.1$   |
| 3     | $-12.8 \pm 0.3$  | $4.0 \pm 0.3$    |
| Total | $-0.9 \pm 0.4$   | $-0.1 \pm 0.3$   |
| (b)   |                  |                  |
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-296.4 \pm 0.3$ | $-217.3 \pm 0.4$ |
| 2     | $-0.4 \pm 0.2$   | $-1.4 \pm 0.1$   |
| 3     | $307.6 \pm 0.3$  | $300.0 \pm 0.1$  |
| Total | $10.9 \pm 0.4$   | $4.8 \pm 0.5$    |
| (c)   |                  |                  |
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-302.9 \pm 0.3$ | $-290.5 \pm 0.4$ |
| 2     | $-1.2 \pm 0.2$   | $-2.4 \pm 0.1$   |
| 3     | $311.6 \pm 0.2$  | $295.5 \pm 0.3$  |
| Total | $7.5 \pm 0.4$    | $2.6 \pm 0.5$    |
| (d)   |                  |                  |
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-168.0 \pm 1.2$ | $-169.4 \pm 0.3$ |
| 2     | $-3.8 \pm 0.3$   | $0.1 \pm 0.2$    |
| 3     | $184.1 \pm 0.6$  | $174.6 \pm 0.2$  |
| Total | $12.3 \pm 1.4$   | $5.3 \pm 0.4$    |

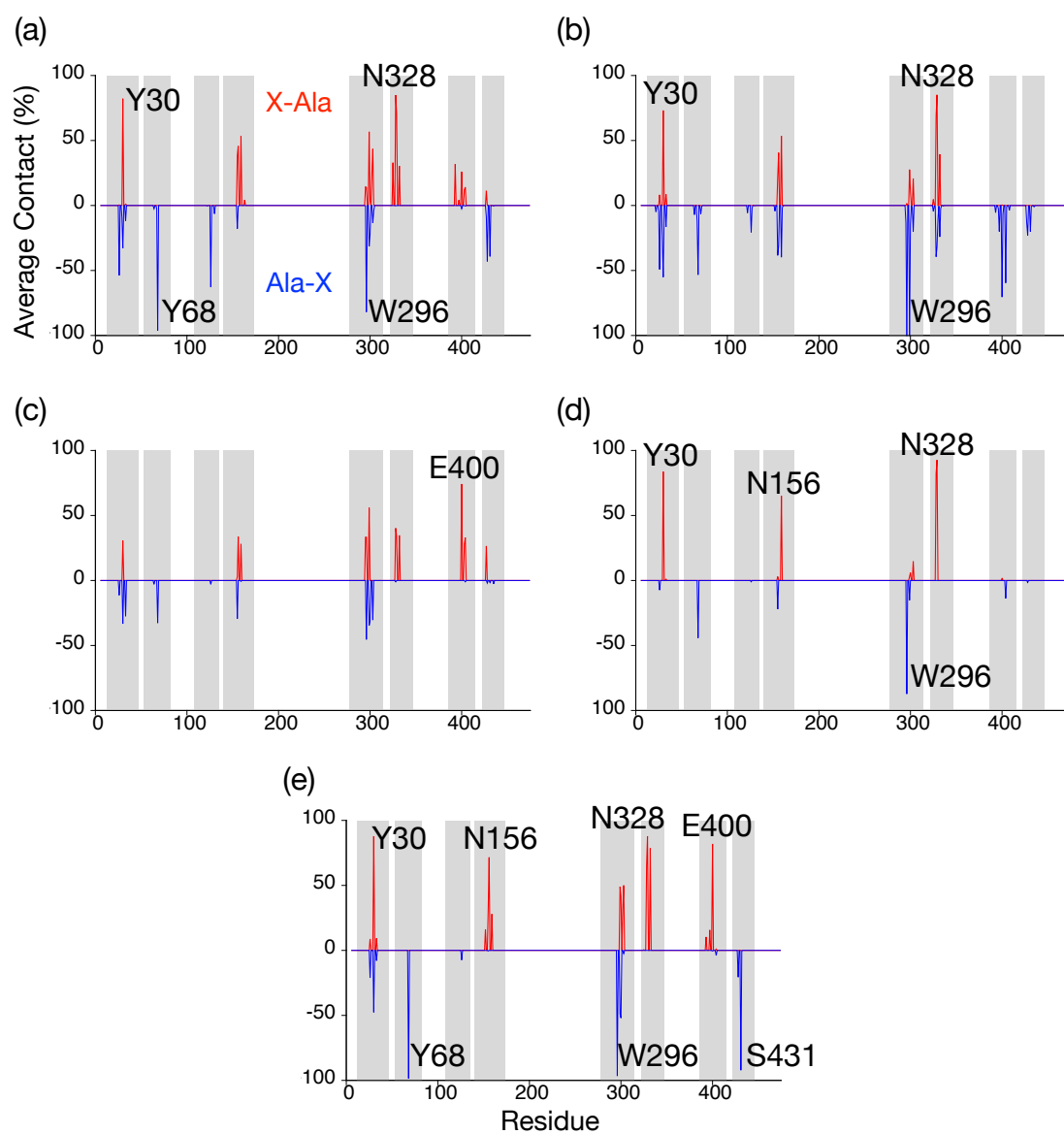
ing site of PepT<sub>St</sub> that interact with the N- and C-terminal side chain of the dipeptide substrate were identified. To do this for all twenty amino acid side chains, the 1 ns simulations performed for the LIE calculations were used. The results illustrated in Figure 4.7 and Figure 4.8 were averaged over the 1 ns simulation time and over the number of simulations involved. It is apparent from this contact analysis that the N-terminal side chain interacts with a group of polar residues consisting of Tyr30, Asn156 and Asn328. The C-terminal side chain, on the other hand, mainly occupies a hydrophobic pocket as per suggested by the crystal structure of AlaPhe complex (Lyons et al., 2014) and form hydrophobic interactions primarily with Tyr68 and Trp296.

Overall, the analysis implies that the side chain at the N-terminus potentially binds more strongly to the binding site via multiple electrostatic interactions, and by doing so, it becomes more sensitive to perturbation. In contrast, the side chain at the C-terminus is accommodated inside an electro-neutral cavity that perhaps can adopt different chemical groups without any large energetic penalty. Previous studies on mammalian PepT1 have documented the importance of the N-terminus for peptide recognition. Borner et al. (1998) and Meredith et al. (2000) suggested that the amino group plays an instrumental role in aligning the rest of the compound in the binding pocket. For the yeast peptide transporter, Ptr2p, Ito et al. (2013) categorised a library of dipeptides into high- and low-affinity groups based on their  $K_i$  and found that the N-terminal residue displayed a higher propensity towards determining to which group a substrate belongs. The *in silico* study presented in this thesis further corroborates this idea by showing that the N-terminal side chain is the major contributor to the overall binding selectivity of dipeptides to PepT<sub>St</sub>.

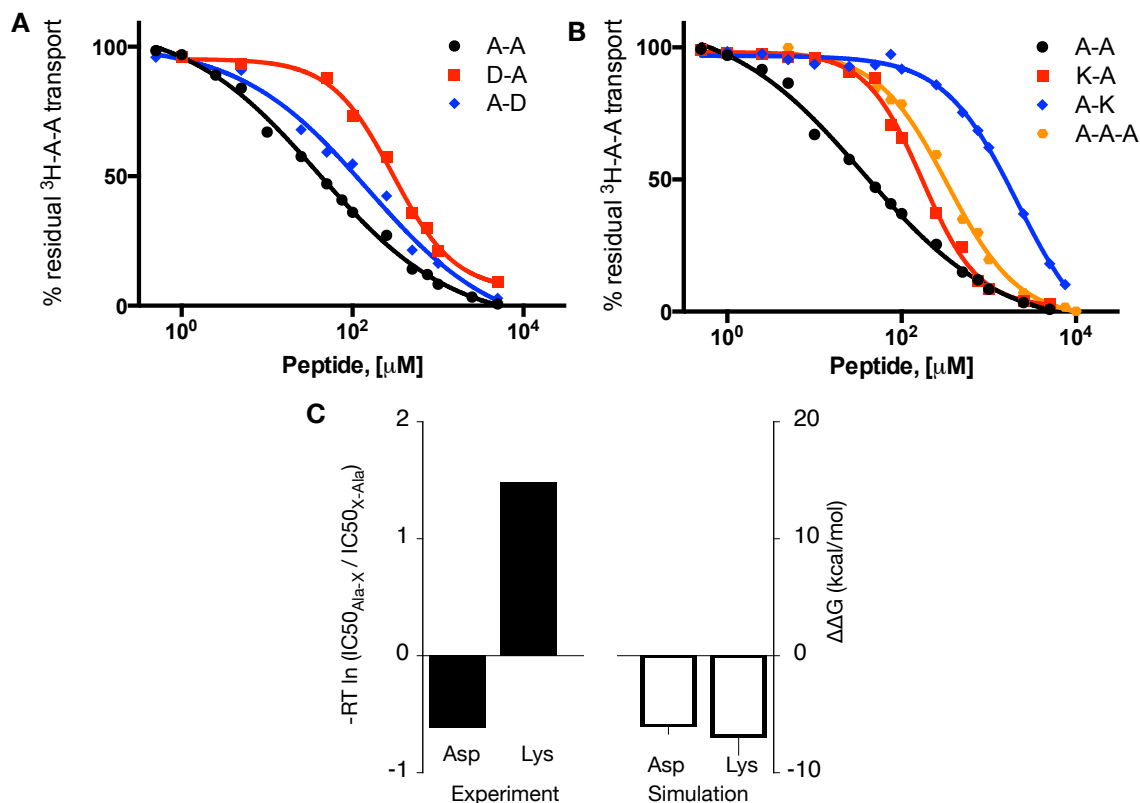
To validate these predictions, *in vitro* proton-driven competition transport assays were performed by Joanne Parker. Two pairs of dipeptides made of the same residues but in opposite sequence—AspAla, AlaAsp and LysAla, AlaLys—were chosen to represent the different charges. Similar assays for two neutral peptides, AlaAla and triAla, were also conducted for comparison. For all experiments, the residual uptake of a reporter substrate, radio-labelled  $^3\text{H}$ -AlaAla, was measured (Figure 4.9 (a) and (b)) and the  $\text{IC}_{50}$  value determined (Table 4.4). To make comparison with the computational predictions easier, the ratio of  $\text{IC}_{50}$  values between AspAla and AlaAsp as well as between LysAla and AlaLys, was determined and the apparent “ $\Delta\Delta G$ ” values calculated (Figure 4.9(c)). In agreement with these predictions, having an aspartate residue at the N-terminus is more detrimental to transport compared to having the same residue at the C-terminal, as indicated by the negative apparent “ $\Delta\Delta G$ ” value. For lysine containing dipeptides, however, the opposite trend was obtained. Positioning a lysine residue at the C-terminus (AlaLys) appears to be extremely aggravating



Figs. 4.7: Contact analysis of the N- and C-terminal side chain of a dipeptide ligand on PepT<sub>St</sub>. (a) Analysis of interactions made by the side chain of dipeptides Ala-X (red) and X-Ala (blue), whereby X is one of the 20 amino acids. This is averaged over 20 simulations of 1 ns length (one for each amino acid). Residues that made significant contacts are mapped to the binding side in (b) and (c), while red and blue arrows point towards the N- and C-terminal side chain, respectively. The cut-off for contacts is set to 3.5 Å.



Figs. 4.8: Contact analysis of the N- and C-terminal side chain based on chemical group: (a) hydrophobic, (b) polar, (c) other, (d) negative and (e) positive.



Figs. 4.9: Proton driven competition uptake assays. (A)  $\text{IC}_{50}$  competition curves for AlaAsp and AspAla showing residual uptake of  $^3\text{H}$ -AlaAla in a proteoliposome-based transport assay. (B) Similar  $\text{IC}_{50}$  competition curves for AlaLys, LysAla and triAla. (C) Apparent “ $\Delta\Delta G$ ” values derived based on the ratio of  $\text{IC}_{50}$  of Ala-X and X-Ala, compared to the corresponding calculated  $\Delta\Delta G$  values from simulation.

for its transport, as exemplified by the very high  $\text{IC}_{50}$  value, resulting in a positive “ $\Delta\Delta G$ ” when compared to LysAla, which is in conflict with the computational predictions.

#### 4.3.4 Lysine containing dipeptides may bind like a tripeptide.

So far, the computational predictions for dipeptides with positively charged residues do not agree well with experimental data. The LIE method overestimated the binding affinity of most dipeptides containing lysine and arginine and ranked some of them similar to neutral dipeptides. Additionally, the TI calculations for LysAla and AlaLys gave the opposite trend to *in vitro* transport assays. These discrepancies could be as a result of many different factors such as inadequate sampling and force field errors. One other possible interpretation of these results is that positively charged dipeptides do not bind to  $\text{PepT}_{\text{St}}$  in the same “horizontal”

Table 4.4: IC<sub>50</sub> values based on proton driven competition transport assays.

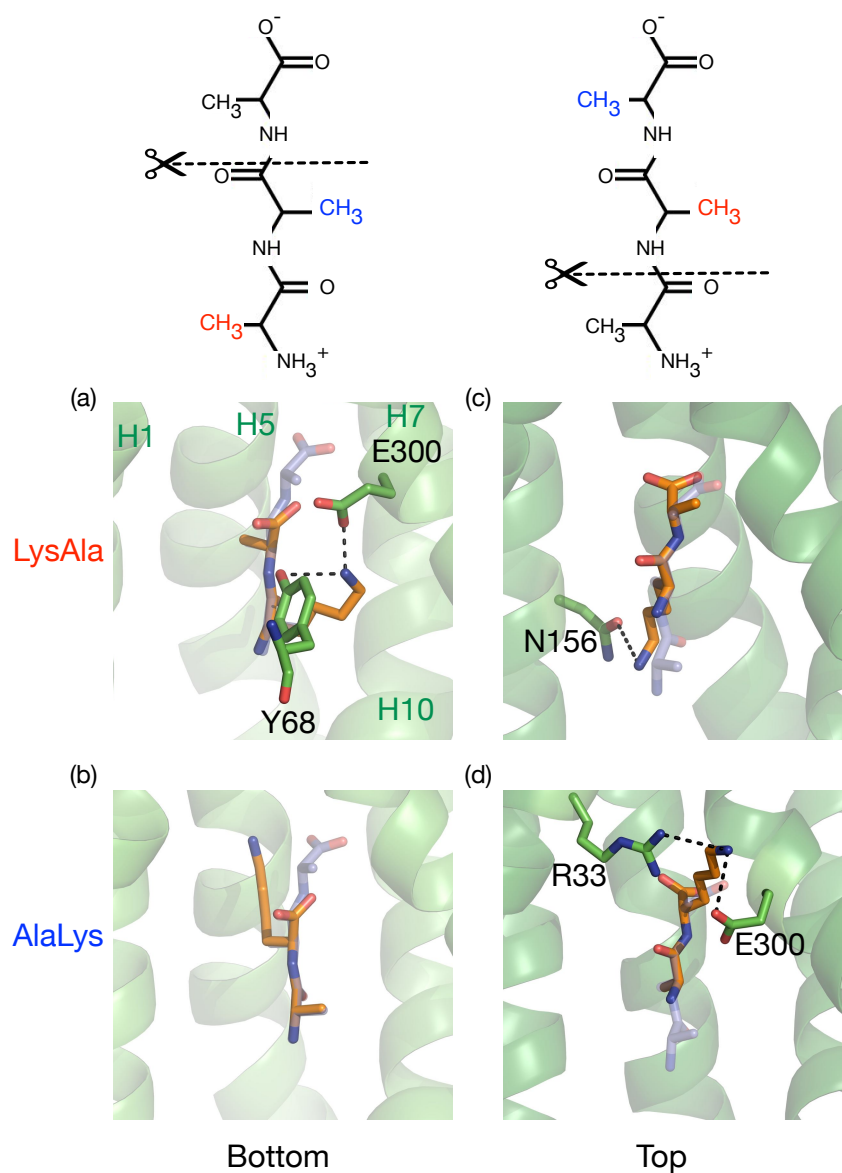
| Peptide   | IC <sub>50</sub> ( $\mu$ M) |
|-----------|-----------------------------|
| AlaAla    | 36                          |
| AlaAlaAla | 400                         |
| AspAla    | 300                         |
| AlaAsp    | 100                         |
| LysAla    | 150                         |
| AlaLys    | 2100                        |

pose as other dipeptides, which explains why using the AlaPhe complex crystal structure as the modelling template produced inaccurate predictions. We hypothesise that for a basic dipeptide, binding occurs in a “vertical” orientation instead, like the tripeptide triAla.

Alternative binding models for AlaLys and LysAla were built using the complex crystal structure of PepT<sub>St</sub>-triAla (Lyons et al., 2014). As triAla has three residues while AlaLys and LysAla two, there are two possible ways to perform this modelling as depicted in Figure 4.10: i) the C-terminal residue was removed and modelling was conducted using the first and second residues as the template, or ii) the N-terminal residue was removed while the second and third residues became the template. Using the former approach, the dipeptide model would be positioned closer to the cytoplasmic side, and henceforth is called the bottom models. On the other hand, the resulting dipeptide for the latter approach would be placed nearer to the extracellular side, and so is called the top models.

As before, thermodynamic integration was performed to alchemically convert either the N- or C-terminal side chain of AlaAla in this vertical binding modes into lysine, and subsequently calculated the  $\Delta\Delta G$  between LysAla and AlaLys (Table 4.5). Intriguingly, the results for both the bottom and top models show qualitative agreement with the experimental transport assays. AlaLys has a larger  $\Delta G$  value and therefore bind less tightly than LysAla, as indicated by the positive  $\Delta\Delta G$  values (Figure 4.11). Further inspection of the alternative binding modes suggests that the stronger binding of LysAla stems from the formation of favourable salt bridges between the  $\epsilon$ -amino group of the peptide substrate and residue Tyr68 and Glu300 in the bottom model or Asn156 in the top model (Figure 4.10(a) and (c)). The binding of AlaLys, on the other hand, is unfavourable due to the lysine side chain protruding into a cavity that is in close proximity to Arg33, and thereby resulting in electrostatic repulsion (Figure 4.10(d)).

While from a modelling perspective, two possible vertical binding sites exist for a dipeptide, i.e. the bottom and top models, it is acknowledged that in reality this might not be the



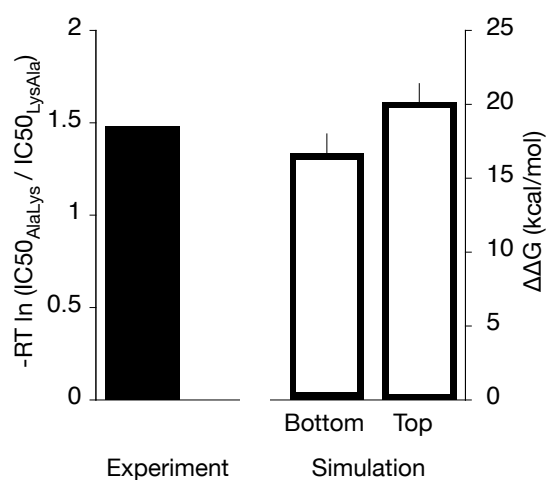
Figs. 4.10: Two ways to model the binding of a dipeptide based on the triAla binding pose. (a,b) Bottom models, whereby the C-terminal residue is removed and the first and second residues are used as template and (c,d) Top models, whereby the N-terminal residue is removed and the second and third residues are used as template. The figures overlay the new models of LysAla and AlaLys (orange) on top of triAla from the crystal complex (purple). Also shown are residues in PepT<sub>S</sub> binding site (green) that are within 4 Å of the ε-amino group of the lysine side chain and can therefore result in electrostatic interactions or repulsions. No residues were found in this vicinity for the AlaLys Bottom model.

Table 4.5: TI results for the alchemical transformations of the N- or C-terminal side chain of AlaAla in a vertical binding pose into lysine based on (a) the bottom model and (b) the top model. All values are quoted in kcal/mol with a precision of 0.1 kcal/mol. The calculated errors represent errors of integration of  $\frac{\partial U}{\partial \lambda}$ .

| (a)   |                  |                  |
|-------|------------------|------------------|
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-190.6 \pm 0.7$ | $-152.1 \pm 0.6$ |
| 2     | $-4.8 \pm 0.4$   | $3.6 \pm 0.5$    |
| 3     | $188.4 \pm 0.6$  | $158.5 \pm 0.6$  |
| Total | $-8.3 \pm 1.0$   | $8.4 \pm 0.9$    |

| (b)   |                  |                  |
|-------|------------------|------------------|
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-190.7 \pm 0.7$ | $-154.7 \pm 0.7$ |
| 2     | $-4.0 \pm 0.4$   | $-1.6 \pm 0.1$   |
| 3     | $179.9 \pm 0.6$  | $163.1 \pm 0.6$  |
| Total | $-14.1 \pm 1.0$  | $6.0 \pm 0.9$    |



Figs. 4.11: The TI method was used to calculate  $\Delta\Delta G$  of alchemically transforming the N- or C-terminal side chain of AlaAla (in both bottom and top models) into lysine. These are compared to the apparent “ $\Delta\Delta G$ ” values derived based on the ratio of  $IC_{50}$  of AlaLys and LysAla from the transport assays.



case. In the crystal structure of PepT<sub>St</sub>-triAla complex, the binding of triAla is poorly coordinated with residues in the binding site, indicating a weak protein-ligand interactions (Lyons et al., 2014). Similarly in the simulation, a dipeptide bound in this orientation was able to move up and down, suggesting that there are no distinct “bottom” and “top” binding sites, instead just one large vertical cavity. Therefore, one would expect to obtain a similar  $\Delta\Delta G$  value for both models. That this is not the case in the simulations suggests that the TI calculations have not converged, which is not surprising given the likely timescale for the ligand to explore the vertical pocket. Nevertheless, as they both show the same trend as experiment, these current results suggest that AlaLys and LysAla may bind to PepT<sub>St</sub> in a vertical fashion like the tripeptide triAla.

A ligand-based substrate template of PepT1 proposed by Bailey et al. (2000) suggested that the best peptide ligands should have their amino and carboxyl ends separated by about 6 Å. As this approximately matches the length of a dipeptide backbone, it is implied that dipeptides have better affinities than tripeptides. Some basic dipeptides like LysLys, however, have been shown to have worse transport characteristics than neutral tripeptides (Eddy et al., 1995; Terada et al., 2000b), for which this substrate template failed to account. This evidence also points towards the importance of the overall chemistry of a peptide substrate in determining its affinity, instead of just the number of residues it possesses. Trying to shed some light into the poor transport of basic dipeptides, Brandsch et al. (1999) modified the lysine residue in LysPro and LysAla into norleucine (Nle) by removing the  $\epsilon$ -amino group, and formed NlePro and NleAla, respectively. Interestingly, this step tremendously increased affinities, which suggests that the transporter disfavours the positively charged  $\epsilon$ -amino group, which they speculated due to electrostatic repulsion in the binding site.

Here, we propose that the poor affinities of these ligands may be caused by the way they interact with the transporter. Unlike neutral and negatively charged dipeptides, a dipeptide with a lysine residue may adopt a similar binding pose to triAla. Positioning these dipeptides laterally like AlaPhe would cause unfavourable steric clashes due to the large side chain. The length of a fully extended lysine side chain is around 5 Å. If a dipeptide has one lysine residue, its total length would be approximately 10 Å, resembling the length of a tripeptide backbone. As such, the only way it can fit into the binding cavity is perhaps by aligning itself vertically like a tripeptide. As this pose is less optimally coordinated with the residues in the binding site (Lyons et al., 2014) and potentially coupled to fewer protons (Parker et al., 2014), its binding affinity is reduced. Another basic residue, arginine, has a side chain that is longer than lysine and is also capped by a positive charge. Thus we conjecture that arginine containing dipeptides may also bind in this pose, which would explain their similarly poor

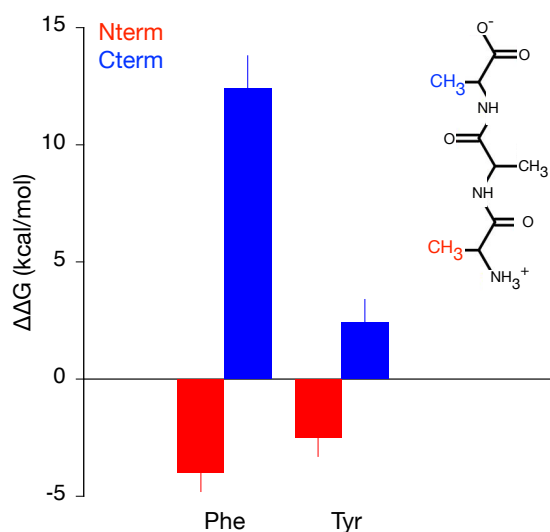
uptake. The current results complement the study conducted by Brandsch et al. (1999) as the shorter side chain of Nle might have restored binding into the horizontal orientation and thereby increased affinities. Collectively, this means the overall size of a peptide, not just the length of its backbone, determine the way it interacts with a POT transporter and therefore its affinity and transport characteristics. Nevertheless, a crystal structure of PepT<sub>St</sub> bound to either lysine or arginine is still required to verify this model.

#### 4.3.5 Tripeptides with bulky residues bind less well.

Tripeptides are weaker substrates of PepT<sub>St</sub> when compared to dipeptides, as has been demonstrated in proteoliposome transport assays, whereby the tripeptide triAla inhibited <sup>3</sup>H-AlaAla uptake ten times more poorly compared to the dipeptide AlaAla (Solcan et al., 2012). Crystallographic data rationalised this observation by showing that the tripeptide triAla bound in a less-tightly coordinated pose compared to the dipeptide AlaPhe (Lyons et al., 2014). Further illuminating this issue, a recent biochemical study using pH-dependent uptake assays suggested that the uptake of a tripeptide is coupled to fewer protons than that of a dipeptide (Parker et al., 2014). In this section, this topic is explored by addressing a similar question as before i.e. how do other tripeptides interact with PepT<sub>St</sub>, and how are their affinities different from triAla?

Based on the twenty amino acids, there are 8000 possible tripeptides that could be the substrate of PepT<sub>St</sub>. However, as this is a very large number, modelling the binding and running the simulation of all tripeptides, as have been done with dipeptides, would be very time consuming. Instead, only tripeptides with bulky residues are studied for two reasons: i) most POT transporters, including PepT<sub>St</sub>, favour bulky and hydrophobic side chains in a dipeptide substrate (Eddy et al., 1995; Biegel et al., 2006; Ito et al., 2013; Solcan et al., 2012), but this has not been studied for a tripeptide substrate, and ii) tripeptides with bulky residues are structurally similar to the  $\beta$ -lactam antibiotics, and therefore studying their binding is of pharmaceutical interest. Using similar techniques that have been applied to dipeptides, alchemical simulations and TI calculations of converting either the N- or the C-terminal residue of triAla into either phenylalanine or tyrosine were performed. The results are shown in Figure 4.12 and Table 4.12.

The transformation of the N-terminal side chain of triAla into phenylalanine and tyrosine resulted in negative  $\Delta\Delta G$  values, which correspond to stronger binding. Upon closer examinations, this bulky side chains were inserted into the hydrophobic pocket, much like the C-terminal side chain of a dipeptide. As such, they form favourable hydrophobic inter-



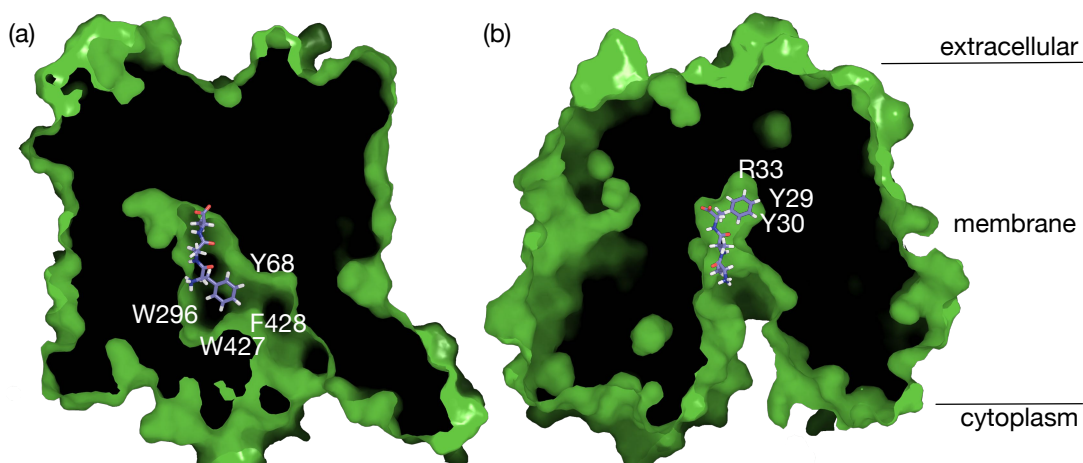
Figs. 4.12: Relative binding free energy,  $\Delta\Delta G$  of alchemically transforming the N- or C-terminus of the tripeptide triAla into either phenylalanine or tyrosine. Y-error bars indicate the error of integration

Table 4.6: TI results for the alchemical transformations of the N- or C-terminal side chain of triAla into (a) phenylalanine and (b) tyrosine. All values are quoted in kcal/mol with a precision of 0.1 kcal/mol. The calculated errors represent errors of integration of  $\frac{\partial U}{\partial \lambda}$ .

| (a)   |                  |                  |
|-------|------------------|------------------|
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-187.8 \pm 0.6$ | $-136.7 \pm 0.8$ |
| 2     | $-0.1 \pm 0.4$   | $5.5 \pm 0.6$    |
| 3     | $183.8 \pm 0.3$  | $143.5 \pm 1.0$  |
| Total | $-4.0 \pm 0.8$   | $12.4 \pm 1.4$   |

| (b)   |                  |                  |
|-------|------------------|------------------|
|       | N-terminus       | C-terminus       |
| Step  | $\Delta\Delta G$ | $\Delta\Delta G$ |
| 1     | $-216.4 \pm 0.6$ | $-136.1 \pm 0.8$ |
| 2     | $0.7 \pm 0.2$    | $-3.8 \pm 0.3$   |
| 3     | $213.2 \pm 0.5$  | $142.3 \pm 0.6$  |
| Total | $-2.5 \pm 0.8$   | $2.4 \pm 1.0$    |

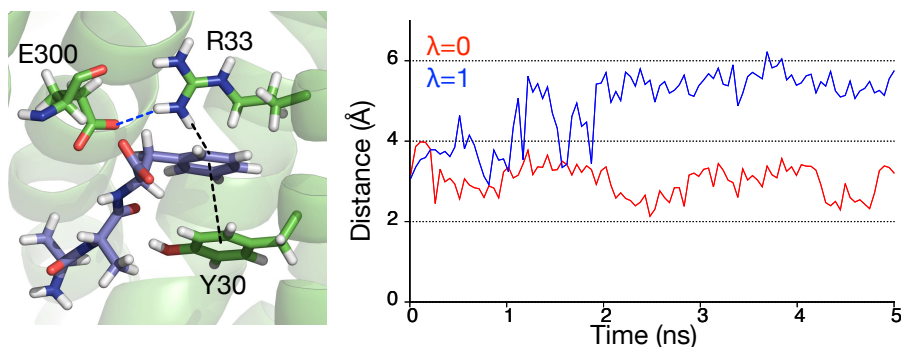


Figs. 4.13: Homology models of PepT<sub>St</sub> bound to (a) PheAlaAla and (b) AlaAlaPhe, in surface representation. Tripeptide substrates are depicted in purple stick representation and residues within the vicinity of the phenylalanine side chain are highlighted.

actions with key residues in the binding site, e.g. Trp296 (Figure 4.13(a)), which explains their better binding affinities than triAla. Transmuting the C-terminal side chain, on the other hand is unfavourable, as indicated by the positive  $\Delta\Delta G$  values for both phenylalanine and tyrosine simulations. The bulky side chain protrudes into a cavity surrounded by Tyr29, Tyr30 and Arg33 (Figure 4.13(b)). It is not clear from the model alone, why this orientation is not favourable. Upon looking at the alchemical simulation of triAla to AlaAlaPhe, it was found that the phenyl ring forms a  $\pi$ - $\pi$  interaction with Tyr30 and a cation- $\pi$  interaction with Arg33 (Figure 4.14). By doing so, however, the phenyl ring disrupts an existing salt bridge between Arg33 and Glu300, which forms the crucial extracellular gate of PepT<sub>St</sub> (Solcan et al., 2012). This loss of salt bridge interaction is perhaps energetically unfavourable, hence the positive  $\Delta\Delta G$ . If the alternating access mechanism is considered, a tripeptide with an electron-rich  $\pi$  system (phenylalanine, tyrosine and tryptophan) at its C-terminus could prevent the formation of the extracellular gate and therefore deter the completion of a transport cycle, which may explain why they are not transported at all.

## 4.4 Conclusion

In this chapter, the binding affinity prediction methods validated in Chapter 3 were used to understand the molecular basis of promiscuity of peptide binding by the bacterial POT transporter, PepT<sub>St</sub>. The crystal complexes solved by Lyons et al. (2014) were used as



Figs. 4.14: The phenyl ring on AlaAlaPhe (purple stick representation), as seen from the extracellular side, interacts with the binding site by forming  $\pi$ - $\pi$  interaction with Tyr30 and cation- $\pi$  interaction with Arg33, depicted as the black dashed lines. The latter breaks the extracellular gate formed by Arg33 and Glu300 (blue dashed line). The graph shows distance between Arg33 and Glu300 during simulations at  $\lambda=0$ , whereby the ligand is triAla, and  $\lambda=1$ , whereby the ligand is AlaAlaPhe.

the starting structures to model the binding of other peptides. Whilst both PepT<sub>St</sub>-AlaPhe and PepT<sub>St</sub>-triAla co-crystal structures were refined to a maximum resolution of 2.5 Å, the quality of electron density maps around their binding site were notably different. In PepT<sub>St</sub>-triAla complex, the electron density map of the ligand was poorly defined due the low occupancy of this peptide in the binding site. As a result, the positions of the N- and C-termini could not be resolved unambiguously. The C-terminus was modelled to face the extracellular side following functional studies of peptide recognition in PepT1 where the C-terminus was suggested to interact with the extracellular gating residues (Meredith et al., 2000). The electron density of AlaPhe ligand, on the other hand, was clearer due to the higher occupancy of this peptide in the binding site as it is a high affinity ligand. The interactions made by this peptide substrate with residues in the binding site were therefore better resolved, allowing us to distinguish the positions of the termini. As all of our models of peptide binding to PepT<sub>St</sub> in this thesis were built using these two structures, our results are currently limited to the interpretation of electron density map by the authors of the paper.

Despite the ambiguous orientation of triAla in the binding site, it is clear that at least two binding modes exist in this transporter. Our computational study adds another layer of complexity by proposing that not all dipeptides bind in the same way, and instead peptide binding is potentially dealt on a case-by-case basis depending on the overall size of the ligands. That a larger dipeptide may interact with the transporter in a similar manner to a tripeptide demonstrates a flexible way of employing the same binding cavity to cater for a wide range of substrates. Human PepT1 shares around 80% sequence identity with PepT<sub>St</sub> within the peptide binding site, suggesting that it is likely PepT1 has a similar substrate

recognition mechanism. The implications of these findings with PepT<sub>St</sub> here on the rational design of drugs targeting PepT1 shall be discussed. However, as there is currently no crystal structure available for human PepT1, a homology model will be first built in the next chapter.

## Chapter 5

# Homology models of human peptide transporters

Full length homology models of the human peptide transporters, PepT1 and PepT2, were built using the crystal structure of PepT<sub>So</sub> as a template for the transmembrane (TM) domain, with extracellular domains (ECD) from mouse PepT1 and rat PepT2. Molecular dynamics (MD) simulations demonstrated a low overall RMSD progression and a good secondary structure preservation, indicating that these are good structural models for the transporters. Essential residues for peptide transport as suggested by various mutational studies align well around the hydrophilic binding cavity. Additionally, a water pocket exists close to His57 and may play a role in the opening of the extracellular gate. By comparing PepT1 and PepT2 models, possible structural explanations for their different kinetics were suggested. The potential interactions between the ECD and TM domains as well as the membrane bilayer were explored to elucidate any potential functional significance. Protein-protein docking was employed to model the interaction of the ECD with intestinal protease trypsin and a working model of how these membrane transporters function within a multicellular organism was proposed.

### 5.1 Previous models of PepT1 and PepT2

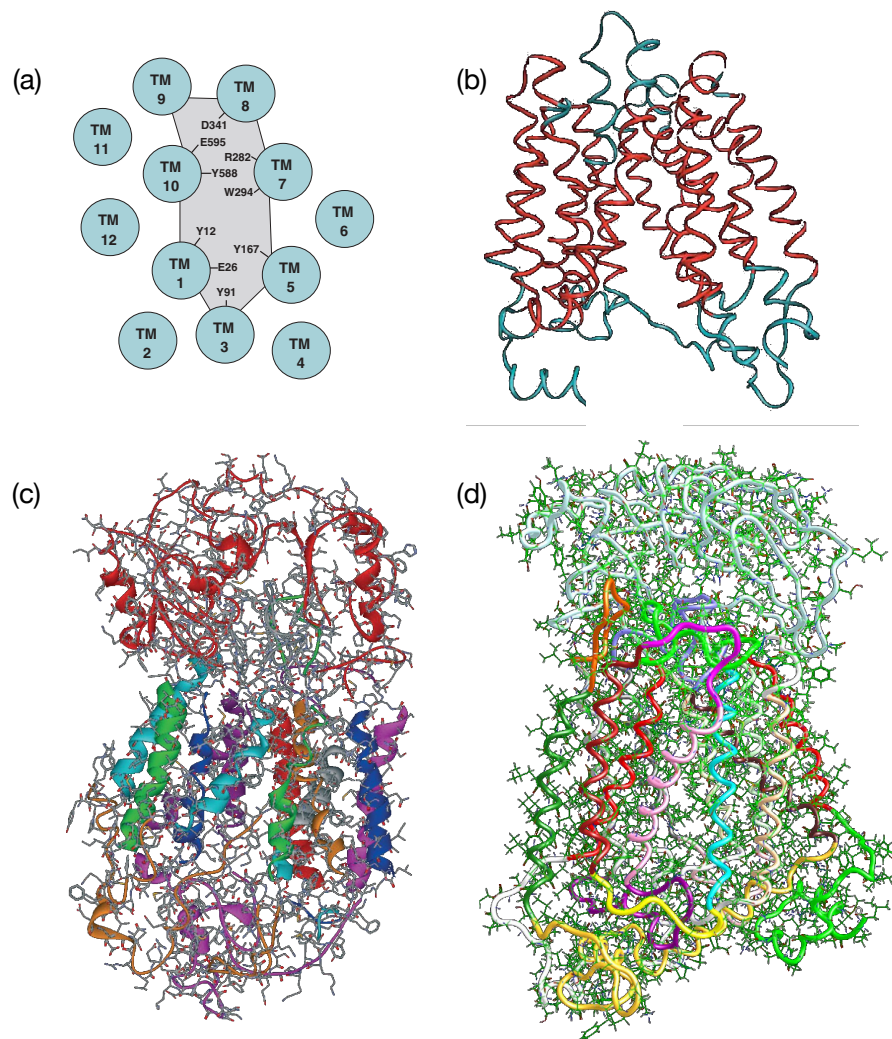
X-ray crystallography is the current gold standard to determine the three-dimensional structure of a protein in atomistic detail. For mammalian integral membrane proteins, however,

obtaining a high resolution structure remains challenging due to several bottlenecks such as poor protein expression, inefficient extraction, low purification yield and difficulty in getting well-ordered crystals (Moraes et al., 2014). For the POT transporters, in particular, several crystal structures have been determined for the bacterial homologs (see §1.2.4); nevertheless, the structures of mammalian PepT1 and PepT2 remain elusive.

Due to the interest in designing drugs targeting human PepT1 for better bioavailability, computer modelling has been employed to predict the structure of this transporter. The first model of PepT1 by Bolger et al. (1998) and Yeung et al. (1998) focused on the position of the TM helices and the residues lining the binding cavity (Figure 5.1(a)). To model the packing of the helices, they first described each helix using a helical wheel representation, where there are seven distinct “faces” through which a helix can interact with its adjacent neighbour. Hence, there are 49 ways ( $7^2$ ) for two helices to be positioned with respect to one another. They then calculated which orientation had the lowest interacting energy and iterated this approach in a pair wise manner for all the helices. Amphipathic helices were arranged to form a central “channel” region, from which key residues were subsequently determined and tested in whole cell uptake assays. One disadvantage of this method is the implicit assumption that all adjacent helices in the sequence are also adjacent in the three dimensional structure. As PepT1 has long intra- and extracellular loops connecting the TM helices, there is no such restriction in the way the helices pack against each other. Despite this limitation, this approach successfully identified and validated several essential residues in the binding site, which became a reference for later modelling studies.

Structural determination of the lactose permease, LacY (Abramson et al., 2003) and the glycerol-3-phosphate transporter, GlpT (Huang et al., 2003) from *Escherichia coli* provided the first insights into the three dimensional structure of the MFS transporter superfamily, of which PepT1 is a member. Using these crystal structures, Meredith and Price (2006) built a homology model of truncated rat PepT1. They took an indirect approach in order to remove extra residues in PepT1, primarily the inter-helical loops, that have no corresponding residues in LacY. They first aligned the LacY sequence to another member of the SLC15 transporter family, PHT1, and removed unmatched regions to form a truncated version of PHT1. This version was subsequently aligned to the sequence of rat PepT1, and again, regions of PepT1 that are not represented in PHT1 were cut out. The modified sequence of PepT1 was finally threaded onto the merged structure of LacY and GlpT to produce a model (Figure 5.1b). Their homology model indeed predicted that many sequentially adjacent helices do not interact with each other. Additionally, they observed H2 and H4 form the binding site despite the previous model suggesting them to lie on the periphery of





Figs. 5.1: Structural models of mammalian peptide transporters built in previous studies. (a) The arrangement of twelve TM helices of PepT1 proposed by Bolger et al. (1998) and Yeung et al. (1998). Figure adapted from a review written by Lee et al. (1999). (b) A homology model of truncated rabbit PepT1 based on the merged structures of LacY and GlpT (Meredith and Price, 2006). (c) A homology model of full-length human PepT1 built using fragmental modelling (Pedretti et al., 2008). (d) Similar approach applied to produce a homology model of human PepT2 (Pedretti et al., 2011).

the protein. Having said that, due to the low sequence identity between LacY and GlpT with PepT1 (12% and 13% respectively), the accuracy of this model is still highly questionable. Also, they omitted important loop regions during the truncation step, especially the ECD, due to the absence of any suitable structural template.

The first full length homology model of PepT1 was generated by Pedretti et al. (2008) using a “fragmental” approach. The primary sequence of PepT1 was first divided into 25 segments: twelve TM domains, six extracellular loops, five intracellular loops and two terminal segments. These segments were then modelled separately using parts of various proteins including a cytochrome for H2, a transcription factor for H7 and a sucrose phosphatase for the ECD\*. These small fragments are assembled together using the backbone of LacY as the template to form the final model (Figure 5.1c). Apparent from this hybrid model is the much shorter transmembrane helices compared to the previous homology model by Meredith and Price (2006), which is a result of using soluble proteins to model the TM region. They predicted the ECD fully covers the extracellular side of the TM domain. Pedretti et al. (2008) docked 50 known PepT1 ligands to the central binding site to construct a structure-based pharmacophore. However, most of the key interacting residues derived from their pharmacophore map were different from previous mutational studies of PepT1, suggesting that this is an erroneous model and the fragmental approach might not be suitable for PepT1. They later built the first and, to date, only homology model of human PepT2 (Figure 5.1d) by applying the same methodology (Pedretti et al., 2011).

The current lack of an accurate structural model to study human peptide transporters calls for a new attempt at homology modelling, especially with the recent determination of the crystal structures of bacterial homologs. In this chapter, homology models of both human PepT1 and PepT2 will be built in conjunction with the X-ray structural characterisation of the ECDs from mouse PepT1 and rat PepT2 by Johnny Beale. The structure of a bacterial POT transporter from *Shewanella oneidensis*, PepT<sub>So</sub>, will be used as a template for the TM domain. The reliability of these models shall be discussed based on various functionally essential residues from the literature and highlight other potentially critical sites for future analysis. Finally, the interaction of the ECD with trypsin will be modelled to suggest how these transporters operate in a multicellular environment.

---

\*The authors of the paper justified this somewhat eccentric approach as “...avoiding the construction of models that lose their structural peculiarities in being forced to conform with the structure of a unique global template.”

## 5.2 Methods

First the selection of a template structure to model the TM domain will be described and then a sequence alignment will be built. The structure-validation programs and molecular dynamics (MD) simulation protocols conducted to assess the models will be outlined. Finally, the protein-protein docking procedure performed to model the interaction of the ECD with trypsin is discussed.

### 5.2.1 Templates selection

The primary sequences of *Homo sapiens* PepT1 (Uniprot ID: P46059) and PepT2 (Uniprot ID: Q16348) were retrieved from the Uniprot protein databases. Five possible candidates consisting of bacterial peptide transporters with available X-ray crystal structures were considered for the template of the TM domain (Table 5.1). Their sequences were aligned with PepT1 and PepT2 and percentage sequence identity calculated using Jalview (Waterhouse et al., 2009). In general, these bacterial POT transporters have moderate overall sequence identity of around 20–30% to their human homologs. Most of the functionally important residues, however, are conserved, therefore these transporters are better than GlpT and LacY that were used in previous work (Meredith and Price, 2006). As PepT<sub>S0</sub> shows the highest percentage identity amongst these candidates, it was selected as the template. The newer crystal structure of PepT<sub>S0</sub> from Fowler et al. (2015) was used instead of the original (Newstead et al., 2011) due to its higher resolution (3.0 Å compared to 3.6 Å, respectively). The coordinates for this structure was fetched from the PDB under the ID code 4UVM. As human peptide transporters are predicted to have twelve TM helices (Fei et al., 1994; Liu et al., 1995), the additional helices, HA and HB, in PepT<sub>S0</sub> (residue 226 to 285) were removed from the structure.

The crystal structures of the ECD of *Mus musculus* PepT1 (Uniprot ID: Q9JIP7) and *Rattus norvegicus* PepT2 (Uniprot ID: Q63424) have been solved to a resolution of 2.1 Å and 2.0 Å, respectively (Beale et al., 2015). The percentage sequence identity of mouse and human PepT1 in the ECD region is 67%, whereas for rat and human PepT2 the score is 70%. As such, these crystal structures, for the first time, provide a good structural template to model an accurate full-length human peptide transporters.

Table 5.1: Bacterial peptide transporters with available crystal structures and their sequence identity with human PepT1 and PepT2.

| Name                | Sequence Identity (%) |       | Uniprot ID |
|---------------------|-----------------------|-------|------------|
|                     | PepT1                 | PepT2 |            |
| PepT <sub>So</sub>  | 31.3                  | 31.4  | Q8EKT7     |
| PepT <sub>St</sub>  | 23.4                  | 25.9  | Q5M4H8     |
| GkPOT               | 24.1                  | 24.7  | Q5KYD1     |
| PepT <sub>So2</sub> | 22.9                  | 22.0  | Q8EHE6     |
| YbgH                | 22.8                  | 21.6  | P75742     |

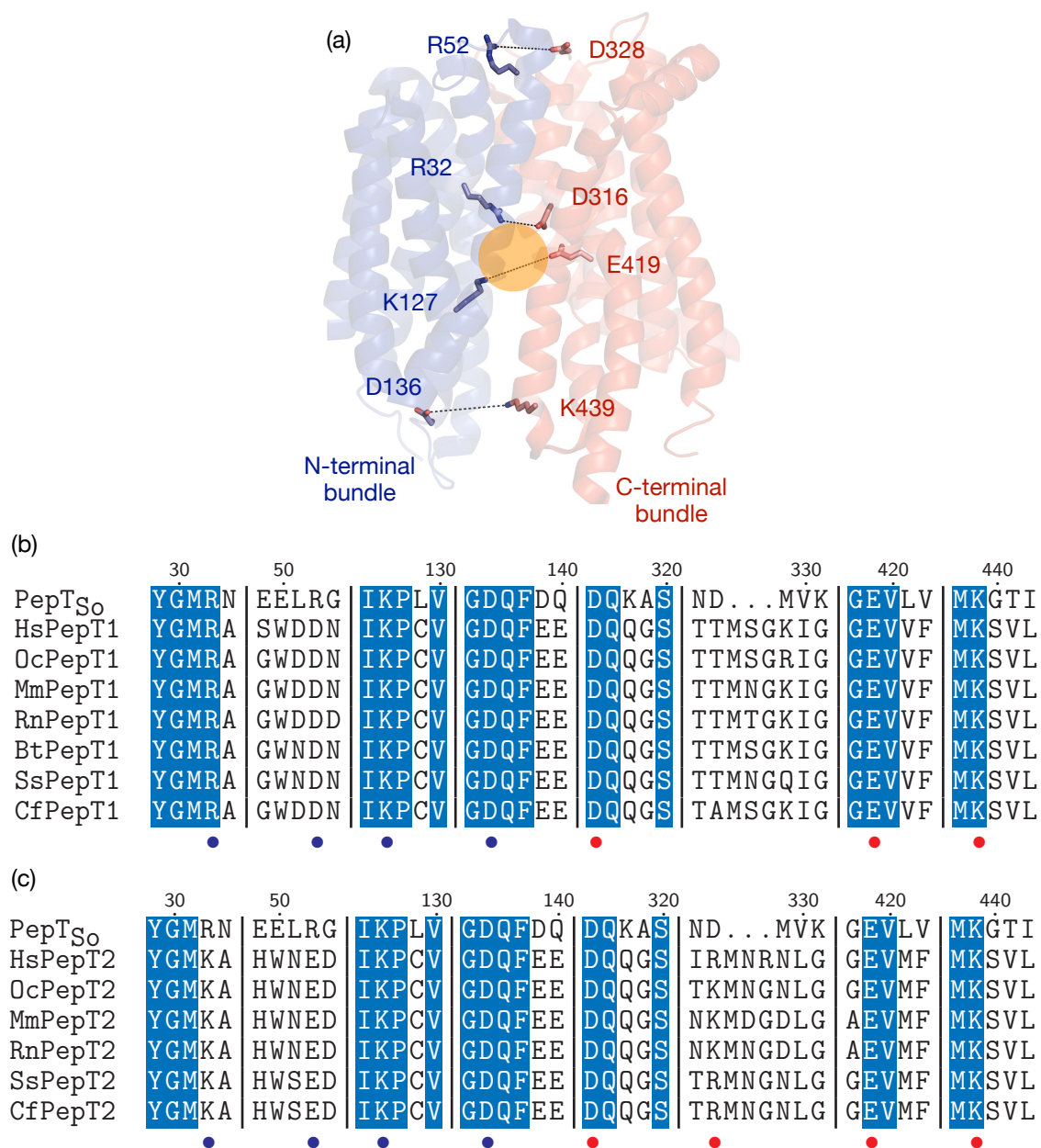
Table 5.2: Proposed gating residues for PepT1 and PepT2

| PepT1  |        | PepT2  |        |
|--------|--------|--------|--------|
| N-term | C-term | N-term | C-term |
| Asp49  | -      | Glu79  | Arg329 |
| Arg34  | Asp298 | Lys64  | Asp317 |
| Lys140 | Glu595 | Lys161 | Glu622 |
| Asp149 | Lys615 | Asp170 | Lys642 |

## 5.2.2 Sequence alignment

As described in §1.2.4, the alternating access mechanism prevalent amongst the MFS transporters is controlled by a series of salt bridges linking the N- and C-terminal bundles of helices, which form gates that occlude the central binding cavity from either side of the membrane (Law et al., 2008; Stelzl et al., 2014; Fowler et al., 2015). The sequence alignment used in this homology modelling was built by aligning the gating residues of PepT<sub>So</sub> with those of PepT1 and PepT2. In PepT<sub>So</sub>, four pairs of residues have been suggested to be involved in gating (Newstead et al., 2011) (Figure 5.2(a)). As gating residues are supposed to be conserved across different species, multiple sequence alignments were generated comprising several mammalian sequences such as human, rabbit (*Oryctolagus cuniculus*), mouse, rat, cow (*Bos taurus*), pig (*Sus scrofa*) and dog (*Canis familiaris*) to identify possible candidates. By subsequently including PepT<sub>So</sub> in these alignments (Figure 5.2(b) and (c)), all comparable salt bridges are found except for the outer periplasmic/extracellular gate, where no corresponding positively-charged residue was found in PepT1 for Asp328 in PepT<sub>So</sub> (the implication of which shall be discussed in §5.3.3). These gating residues are listed in Table 5.2.

Primary sequences from the crystal structures of truncated PepT<sub>So</sub>, mouse PepT1 ECD



Figs. 5.2: Identifying gating residues in PepT1 and PepT2. (a) Residues forming the periplasmic and cytoplasmic gates in PepT<sub>S0</sub>; N-terminal gating residues are coloured in blue, whilst C-terminal in red. The position of the binding site is indicated by the orange circle. (b) Corresponding gating residues were elucidated for PepT1 by aligning amino acid sequences from mammalian transporters with PepT<sub>S0</sub>. Proposed gating residues are highlighted by the blue and red bullets representing the N- and C-terminal gates, respectively. Conserved residues are highlighted in light blue. (c) A similar analysis was conducted to determine the salt bridge pairs for PepT2. Uniprot IDs: PepT<sub>S0</sub> (Q8EKT7), HsPepT1 (P46059), OcPepT1 (P36836), MmPepT1 (Q9JIP7), RnPepT1 (P51574), BtPepT1 (A5D7E5), SsPepT1 (F1RP40), CfPepT1 (Q8WMX5), HsPepT2 (Q16348), OcPepT2 (P46029), MmPepT2 (Q9ES07), RnPepT2 (Q63424), SsPepT2 (A5A4L7) and CfPepT2 (E2QWX1).

and rat PepT2 ECD were extracted using the WHAT IF web server (<http://swift.cmbi.ru.nl/servers/html/soupir.html>) (Vriend, 1990). These sequences were aligned to the amino acid sequence of human PepT1 and PepT2, using the gating residues as points of reference. The Probcons algorithm (Do et al., 2005) was used to generate an initial sequence alignment and this was manually edited in Jalview (Waterhouse et al., 2009). The final alignments used for homology modelling are shown in Appendix B.

Using secondary structure prediction algorithms and point mutagenesis, Xu et al. (2010) predicted that the long intracellular loop linking the N- and C-terminal bundles (between H6 and H7) is made of two amphipathic  $\alpha$ -helices. As such, these helices were modelled following the positions suggested by this study. Modeller-9v9 (Sali and Blundell, 1994) was employed to generate 100 homology models each for PepT1 and PepT2 and they were ranked based on discrete optimised protein energy (DOPE) score (Shen et al., 2006). Models with the lowest DOPE scores were chosen for further assessments.

### 5.2.3 Structural evaluation

We chose 10 homology models, each for PepT1 and PepT2, to be subject to stereochemical and geometrical assessments. Ramachandran analysis (Ramachandran et al., 1963) was performed on the Rampage web server (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell and Davis, 2003), followed by contacts, conformation and geometry analyses on the MolProbity website (<http://molprobity.biochem.duke.edu>) (Chen et al., 2010). PepT1 and PepT2 models with the highest overall MolProbity score were chosen and the results of these structural analyses are shown in Figure 5.4 and Table 5.3. Further structural assessment was done by simulating the homology models inside a membrane bilayer.

### 5.2.4 Molecular dynamic simulation

The N-terminal loop (residue 1-14 for PepT1 and residue 1-42 for PepT2) and C-terminal loop (residue 688-708 in PepT1 and residue 713-729 in PepT2) were cleaved as they do not have any secondary structure. Using the g\_membed protocol, each homology model was inserted into a pre-equilibrated model membrane bilayer consisting of 381 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules (Dickson et al., 2012) inside a simulation box of dimension 10 nm  $\times$  13 nm  $\times$  15 nm. The box was solvated with TIP3P water (Jorgensen et al., 1983) and 0.15 M Na<sup>+</sup> and Cl<sup>-</sup> ions. The system was relaxed through a steepest-descent energy minimisation protocol. An equilibration simulation was performed for 1 ns, whereby heavy atoms of the homology models were positionally-restrained. Re-

moving these restraints, three repeats of production simulations were run, each for 50 ns. The temperature was kept at 310 K by applying the velocity-rescaling thermostat with a time constant of 0.1 ps (Bussi et al., 2007). To maintain pressure at 1 atm, a semi-isotropic coupling was employed using a Parrinello-Rahman barostat with a time constant of 1.0 ps and compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  (Parrinello, 1981). A LINCS algorithm was used to constraint the length of all bonds (Hess et al., 1997) and a time step of 2 ps was used to integrate the equation of motion. Long-range electrostatic interactions were calculated by the particle mesh Ewald (PME) method (Essmann et al., 1995) with a real-space cut off of 12 Å. Simulations were conducted in Gromacs 4.5.4 (Hess et al., 2008) using the Amber99SB biomolecular force field (Lindorff-Larsen et al., 2010). Structural analysis was conducted using Gromacs tools `do_dssp`, `g_rms` and `g_rmsf` and the Visual Molecular Dynamics (VMD) plug-in `interquant` (Humphrey and Dalke, 1996).

### 5.2.5 Protein-protein docking

Using two residues as points of reference—Glu549 and Glu572 in PepT1, and Asp576 and Glu599 in PepT2—we modelled the binding of PepT1 and trypsin by docking the crystal structure of human trypsin (PDB ID: 1TRN) (Gaboriaud et al., 1996) to the homology model on ClusPro 2.0 web server (<http://cluspro.bu.edu/login.php>) (Comeau et al., 2004). The procedure produced 120 models, from which the best representative model was selected based on several guidelines: i) the trypsin active site, or the catalytic triad (His57, Asp102 and Ser195), must be accessible to solvent to allow access for incoming polypeptides, ii) the position of trypsin along the z-axis must not be too low such that it will interrupt the membrane bilayer, and iii) there must be a large buried surface area between trypsin and the ECD for a tight binding.

A simulation of PepT1-trypsin complex was set up using the same protocols described in §5.2.4. To remove steric clashes at the interface of ECD and trypsin, a 1 ns equilibration MD simulation was performed, whereby C $\alpha$  atoms of both proteins were positionally-restrained but side chain atoms were allowed to move.

## 5.3 Results and Discussion

The homology models of human PepT1 and PepT2 are shown in Figure 5.3 with their relative position in a model membrane bilayer highlighted. These two models are structurally similar to each other (C $\alpha$  RMSD of 1.4 Å), as expected due to the same structural template

for the TM region and the ECD from both proteins are alike. The TM domain matches the thickness of a POPC bilayer and most extracellular loops are exposed to the bulk solution with no direct contact with the membrane. One of the two intracellular helices between the N- and C terminal bundles was predicted to interact with the inner leaflet of the bilayer due to its amphipathicity. Sitting upright above the TM domain, the ECD does not form any apparent interaction with the outer leaflet of the bilayer.

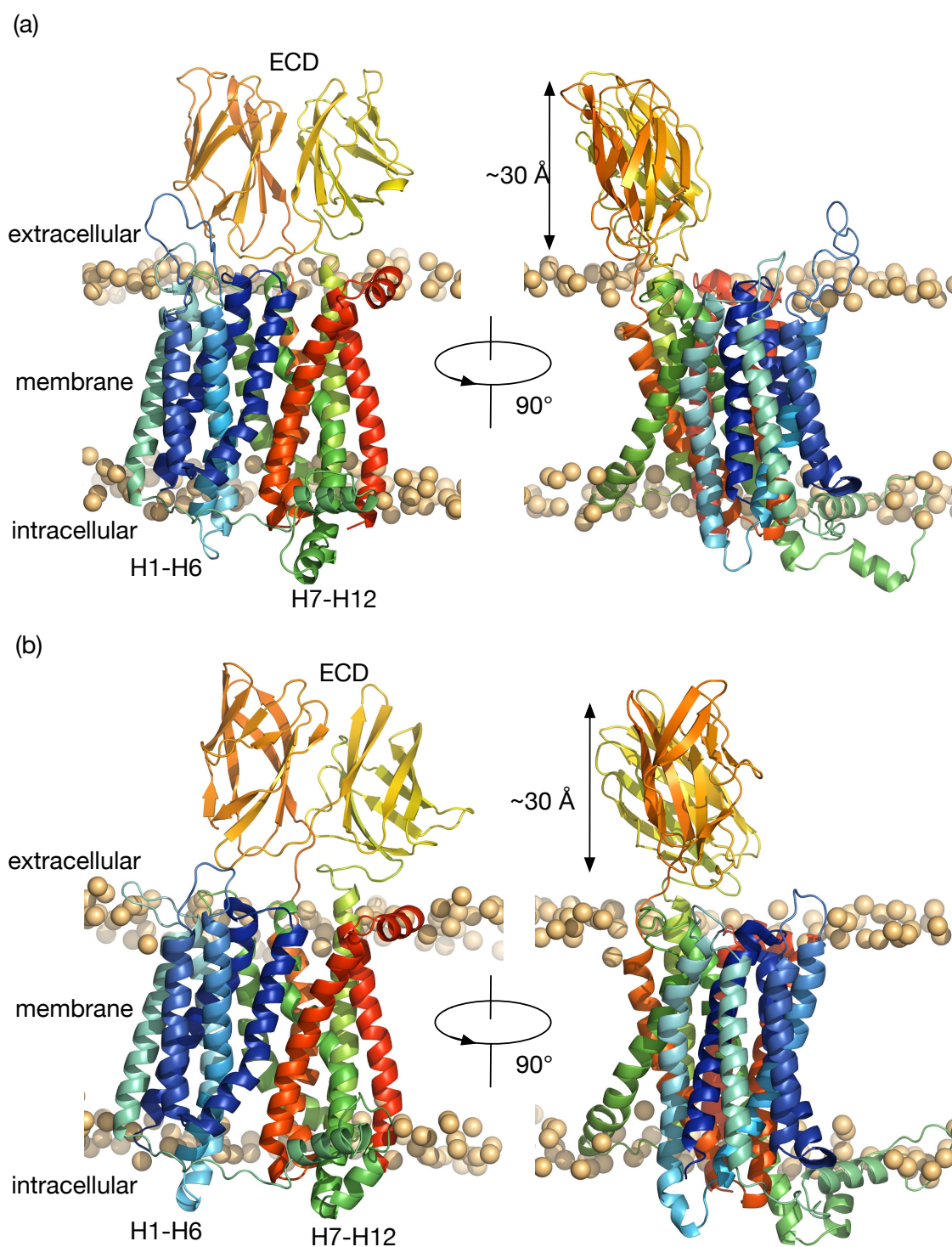
Our structural evaluation using Rampage (Lovell and Davis, 2003) and Molprobit (Chen et al., 2010) web servers (Figure 5.4 and Table 5.3, respectively) suggest that in general, these are poor protein structures as indicated by the low percentile of MolProbit overall scores<sup>†</sup> (11<sup>th</sup> and 15<sup>th</sup> for PepT1 and PepT2, respectively). This is not surprising given they are homology models instead of high resolution X-ray crystal structures. Considering the poor homology between PepT<sub>So</sub> and PepT1 or PepT2, however, these models indeed performed quite well. Most of the Ramachandran outliers were found in unstructured interhelical loop regions (Figure 5.4) where there were no structural templates and therefore Modeller cannot model them accurately. These are mainly the extracellular loops, which are significantly longer in the mammalian sequences, and the large intracellular loop between H6 and H7. Additionally, both PepT1 and PepT2 models have no clashing residues as indicated by the zero Clashscore<sup>‡</sup>.

Three repeats of 50 ns MD simulations were performed with the homology models of PepT1 and PepT2 inserted into a POPC model membrane. Running similar structural evaluations after these MD simulations show a moderate improvement whereby the number of Ramachandran outliers are reduced to 6 and 3, respectively, for PepT1 and PepT2 models, while the Molprobit scores increase to the 96<sup>th</sup> and 92<sup>nd</sup> percentiles. The ability of these homology models to retain their secondary structure during the simulation was also assessed. The analysis was first performed on the TM region by the Gromacs tool `do_dssp`. This program classifies each residue into  $\alpha$ -helical or other for the first and last 5 ns of the simulation (Figure 5.5). All  $\alpha$ -helices demonstrated a dynamic progression during the simulation, where unfolding and refolding occurred repeatedly especially around the ends of helices, as exemplified by H12. Looking at the total percentage of  $\alpha$ -helical residues, however, there is no significant changes between the beginning and the end of the simulations, suggesting the TM region of the homology models is structurally robust when embedded into the membrane. A similar analysis was then applied to the ECD region (Figure 5.6) and

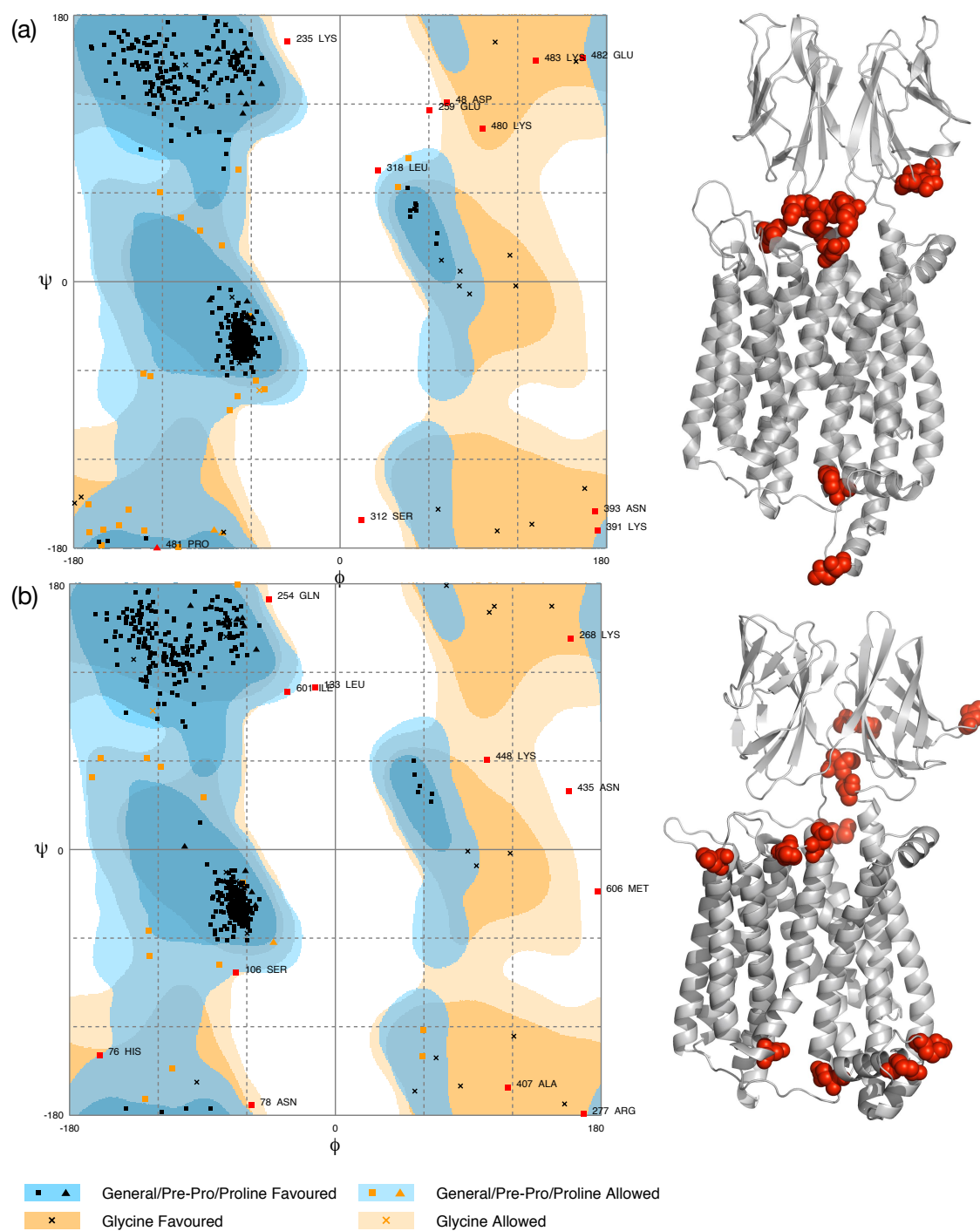
<sup>†</sup>MolProbit scores are based on atom clashes, arrangement of rotamers and Ramachandran evaluations. These are ranked based on crystal structures selected in 2006, whereby 100<sup>th</sup> quartile represents the best structures whilst 0<sup>th</sup> the worst.

<sup>‡</sup>Number of serious steric overlaps ( $>0.4$  Å) per 1000 atoms





Figs. 5.3: Homology models of (a) PepT1 and (b) PepT2 from human. The models are coloured by rainbow from red (N-terminal) to blue (C-terminal) with the ECD in yellow and orange. Relative position of the membrane is indicated by the spheres representing the phosphate head group of POPC lipid molecules.



Figs. 5.4: Ramachandran plots for PepT1 and PepT2 homology models generated in Rampage web server (Lovell and Davis, 2003). The positions of outliers are highlighted in red spheres.

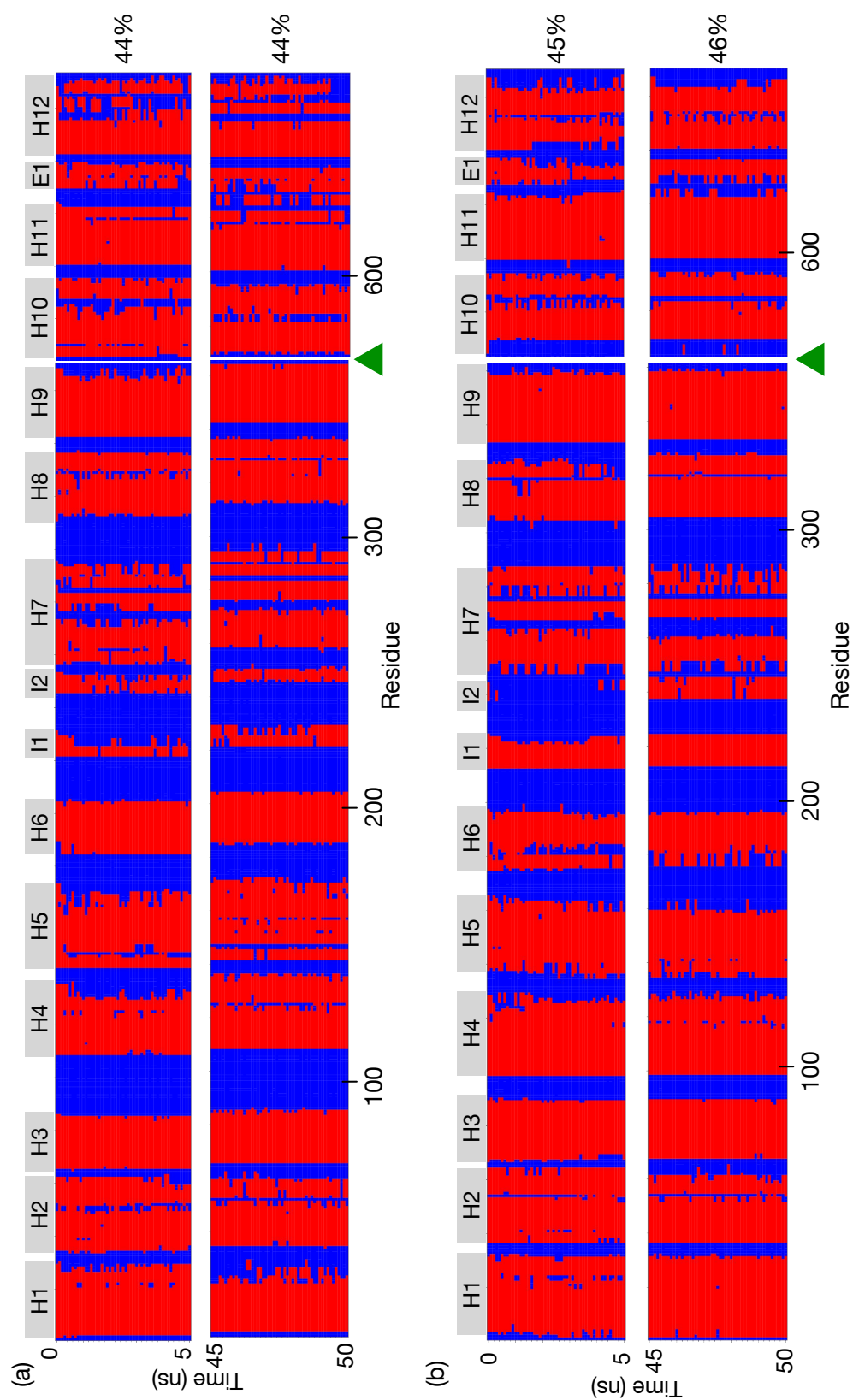
Table 5.3: MolProbity analysis for PepT1 and PepT2 homology models

|                               |          | PepT1            | PepT2            |
|-------------------------------|----------|------------------|------------------|
| Ramachandran analysis         | Favoured | 635 (94.6%)      | 640 (95.8 %)     |
|                               | Allowed  | 25 (3.7%)        | 16 (2.4%)        |
|                               | Outlier  | 11 (1.6% )       | 12 (1.8%)        |
| Clashscore                    |          | 0                | 0                |
| Poor rotamers                 |          | 28 (4.1%)        | 26 (3.9%)        |
| C $\beta$ deviations          |          | 15 (2.3%)        | 9 (1.3%)         |
| Backbone bond deviations      |          | 0                | 0                |
| Backbone angle deviations     |          | 12 (1.8%)        | 10 (1.5%)        |
| MolProbity score (Percentile) |          | 11 <sup>th</sup> | 15 <sup>th</sup> |

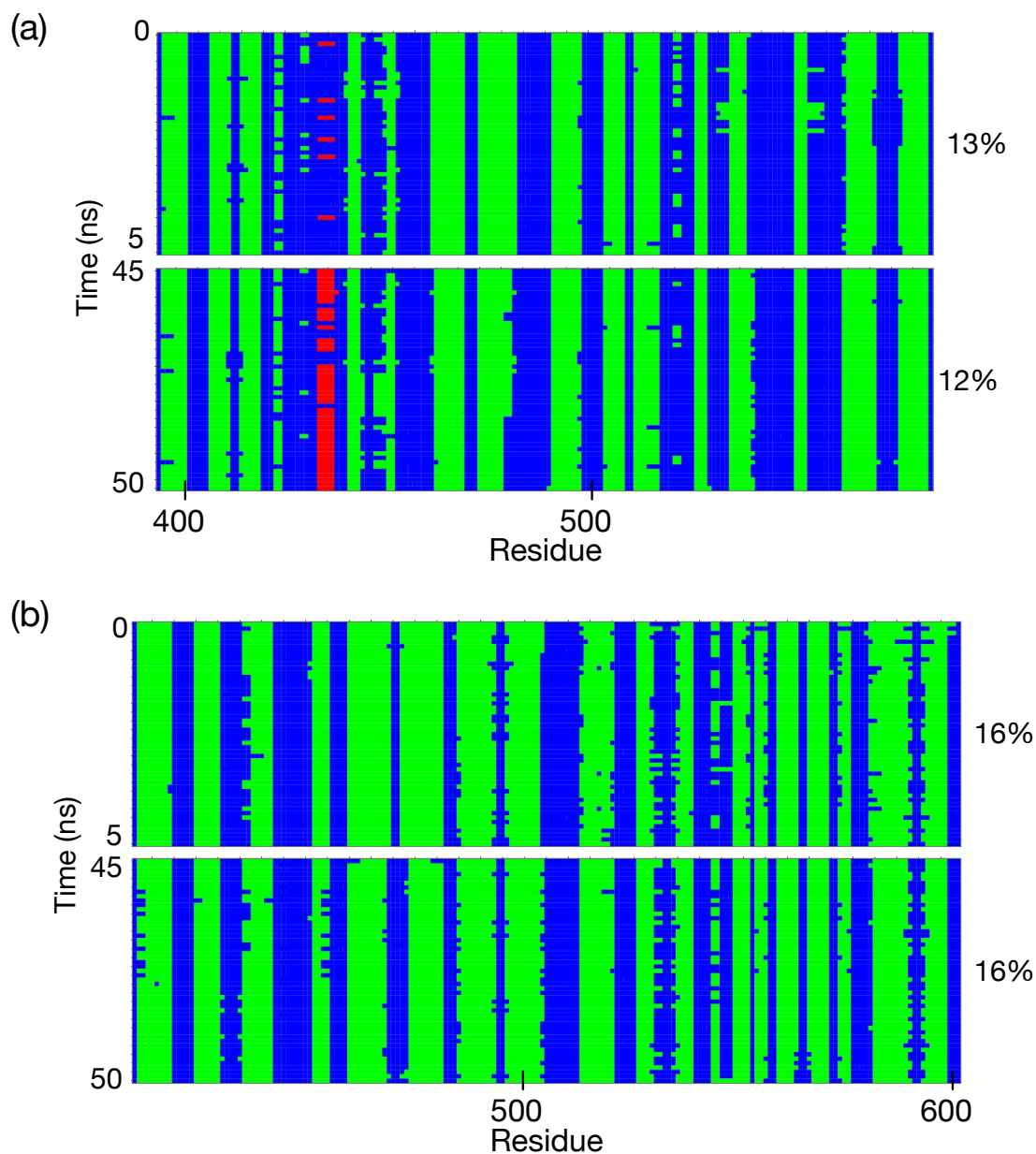
a strong preservation of the  $\beta$ -sheet structures was observed throughout the entire simulation with very minimal changes, especially in PepT2. For PepT1, the short  $\alpha$  helix present in the mouse crystal structure (residue 433-435) unfolded when the simulation began, but re-folded towards the end. Overall, both the TM and ECD regions retain their main secondary structural features without any major unfolding.

Another way to assess the structural quality of the homology models is by measuring the root mean square displacement (RMSD), which indicates the deviation the models undergo along the timeline of the simulation, with respect to a reference structure. The analysis was performed on the C $\alpha$  atoms of the protein and the original structure of the homology models was used as the reference. The results were compared to a simulation of PepT<sub>S0</sub> under the same condition (Figure 5.8). Both homology models have a higher RMSD compared to PepT<sub>S0</sub>, which is expected as the latter is a high resolution crystal structure and therefore is more structurally sturdy. For homology models that are moderately large ( $\sim 700$  residues), the RMSD values of PepT1 and PepT2 are encouragingly low. Interestingly, the overall RMSD for PepT2 is around 1 Å lower than that of PepT1, but if only the TM region is considered, both models show comparable values. This suggests that while both TM domains of PepT1 and PepT2 behave similarly, the ECD of PepT1 is more mobile during the simulation than that of PepT2.

To investigate further, the root mean square fluctuations (RMSF) (Figure 5.8), which indicate the total deviations of each residue throughout the entire simulations with respect

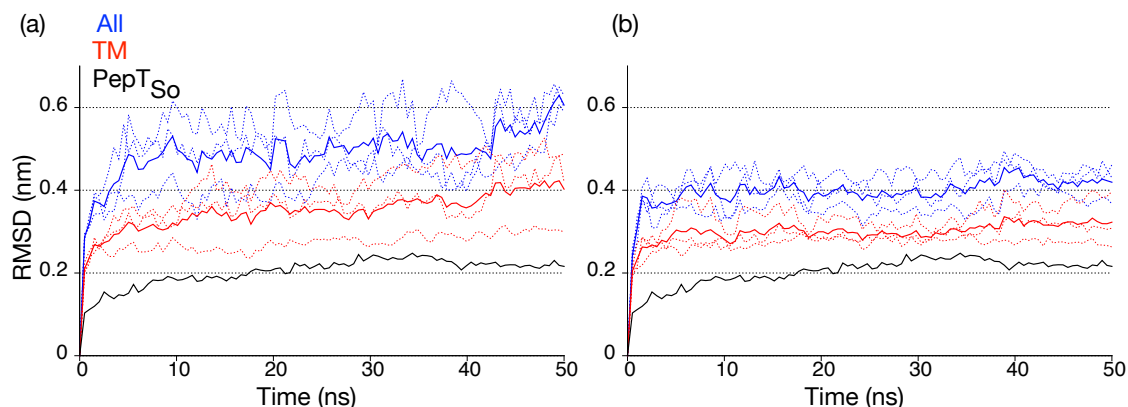


Figs. 5.5: Secondary structure analysis of the TM region of (a) PepT1 and (b) PepT2 homology models during one of the three simulations in a model membrane. The top bar of each panel represents the first 5 ns of the simulation and the bottom bar the last 5 ns. An  $\alpha$  helical structure is coloured in red and any other structure in blue. Predicted position of each TM helix is shown in grey rectangle whilst the position of the ECD is marked with a green triangle. Percentage residues forming  $\alpha$  helical structure (out of all residues) for the 5 ns time frame is shown on the right.



Figs. 5.6: Secondary structure analysis of the ECD from (a) PepT1 and (b) PepT2 homology models during one of the three simulations with a model membrane. The top bar of each panel represents the first 5 ns of the simulation and the bottom bar the last 5 ns. A  $\beta$ -sheet structure is coloured in green,  $\alpha$ -helical in red and any other structure in blue. Percentage residues forming  $\beta$ -sheet structure (out of all residues) for the 5 ns time frame is shown on the right.

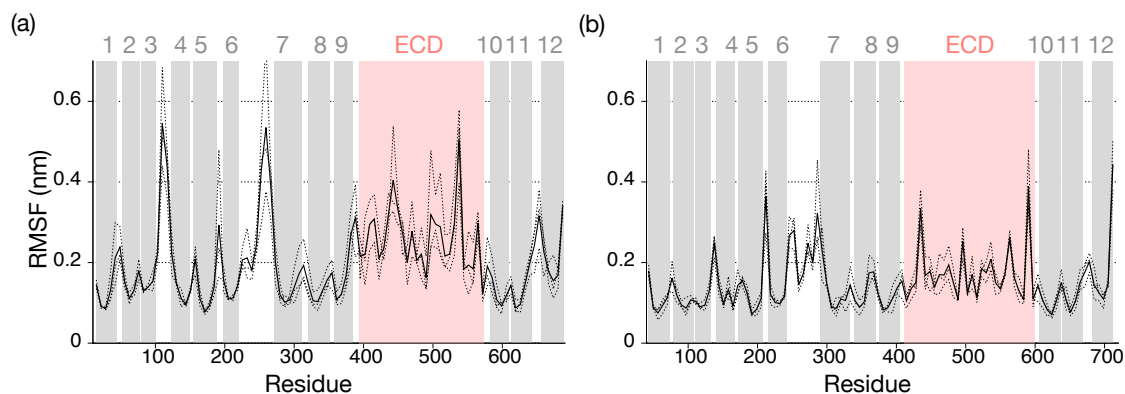




Figs. 5.7: Root mean square displacement (RMSD) of the  $C\alpha$  atoms of (a) PepT1 and (b) PepT2 homology models during 50 ns simulations in a POPC model membrane. The RMSD progression of all  $C\alpha$  atoms is shown in blue, whilst  $C\alpha$  within the TM region is in red. Three dotted lines represent three repeats of the simulations and to ease visualisation, an average value is calculated and shown as the solid line. RMSD progression of all  $C\alpha$  atoms of PepT<sub>So</sub> is shown in solid black line as a reference.

to the starting structure, were measured. Indeed, the ECD region (shown in pink) of PepT1 displays higher RMSF values than PepT2, further corroborating earlier suggestion that the PepT1 ECD moved more during the simulation. However, this movement is likely to be insignificant as the difference between PepT1 and PepT2 is only around 1 Å, which suggests that this is caused by small local changes within the ECD rather than an orchestrated movement of the whole domain. Also of high RMSF are two loops: the extracellular loop linking H3 and H4 and the intracellular loop between H6 and H7. In PepT1 these loops are longer and therefore more flexible as indicated by the higher RMSF compared to PepT2. The RMSF analysis also displays low fluctuations for residues within the TM helices (shown in grey) i.e. less than 2 Å, consistent with the secondary structure analysis discussed earlier that demonstrated strong conservation of  $\alpha$ -helical structures.

The ECD was modelled into PepT1 and PepT2 such that it is positioned above the TM domain. Due to the long linker between H9 and H10 and the ECD, for which there are currently no structural information, the ECD is free to move during the simulation. To elucidate the structural conformations sampled by this large domain, the density profiles of the protein relative to the membrane were plotted (Figure 5.9). As postulated earlier, there is no large concerted movement of the ECD with respect to the TM region, but some dynamics within the ECD itself can be observed. The density profile also suggests no significant interaction between the ECD and the membrane. This supports the original models that

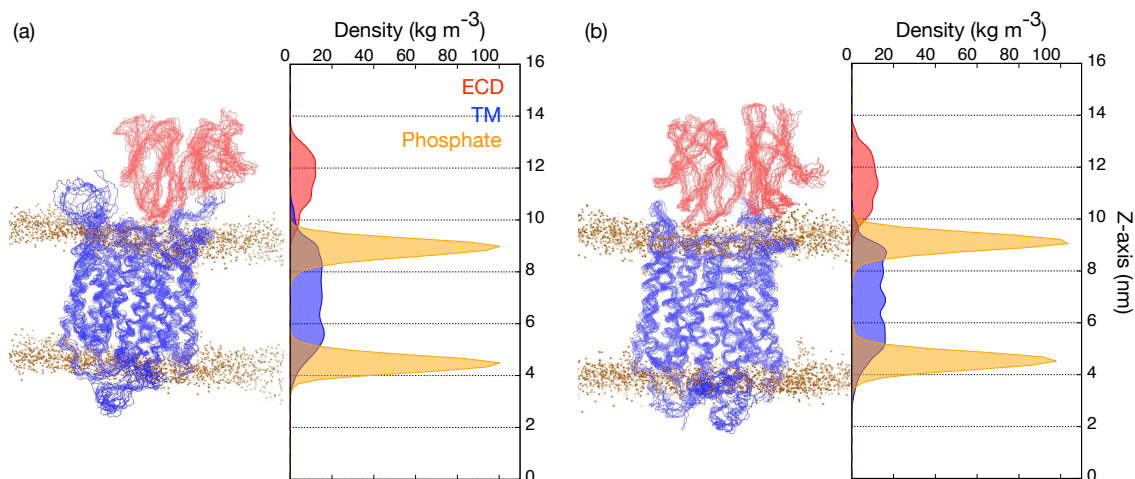


Figs. 5.8: Root mean square fluctuation (RMSF) of the C $\alpha$  atoms of (a) PepT1 and (b) PepT2 homology models during 50 ns simulations in a POPC model membrane. The positions of all TM helices are indicated in grey whilst the ECD in pink. Three dotted lines represents three repeats of the simulations and to ease visualisation, an average value is calculated and shown as the solid line.

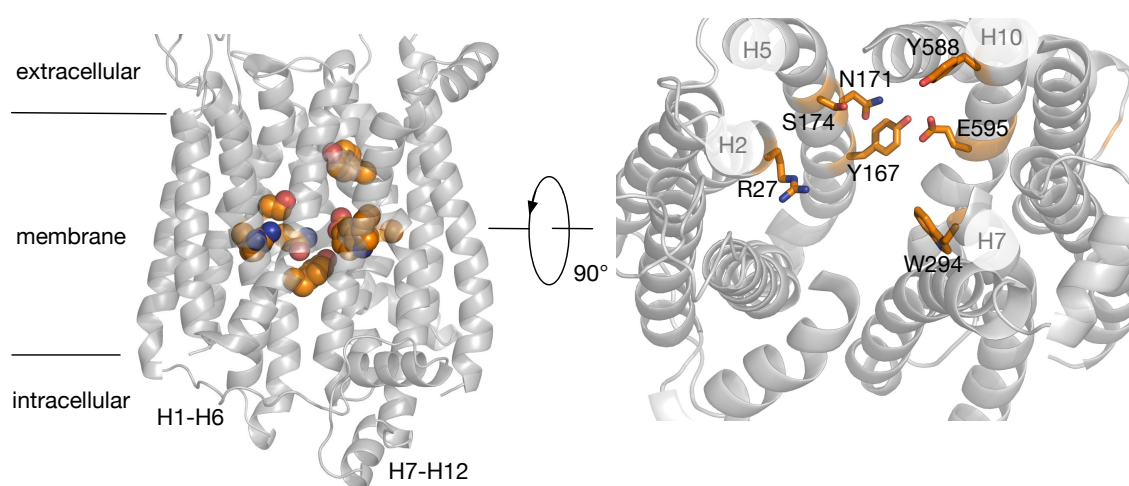
the ECD sits upright above the TM region, possibly through specific interactions with other extracellular loops (explored in §5.3.3). The implications of this behaviour of the ECD shall be discussed later in the chapter.

### 5.3.1 Essential residues mapped on homology models

Given the interest to design drugs with improved oral bioavailability, these transporters, particularly PepT1, have been the subject of numerous functional studies by virtue of applying point mutations and creating chimeric proteins. Due to the difficulty in purifying mammalian transporters for reconstitution into an *in vitro* system, transport assays were generally performed in transiently transfected cell lines, for example the MDCK cells (Irvine et al., 1999), or in *X. laevis* oocytes. The uptake of a reporter substrate, usually  $^{14}\text{C}$ -Glysar, by a mutated or chimeric transporter was then compared against the wild type. The results of these functional studies will be discussed in the context of the homology models.



Figs. 5.9: Density profile of the TM (blue) and ECD (red) for (a) PepT1 and (b) PepT2 with respect to the membrane bilayer, as indicated by the phosphate atoms (orange). Structural evolution of the models during the 50 ns simulations were shown as traces produced every 1 ns. Membrane density is represented by the phosphate head group.



Figs. 5.10: Key residues for peptide binding in PepT1 as proposed by various mutational studies (Table 5.4) mapped onto the homology model.



Table 5.4: Key residues in human PepT1 from mutational studies

| Residue | Position | Suggested Function    | Reference                                                          |
|---------|----------|-----------------------|--------------------------------------------------------------------|
| His27   | H1       | Substrate binding     | Terada et al. (2004)                                               |
| His57   | H2       | Proton coupling       | Terada et al. (1996); Fei et al. (1997); Uchiyama et al. (2003)    |
| Tyr58   | H2       | Proton coupling       | Chen et al. (2000); Pieri et al. (2009)                            |
| Tyr91   | H3       | Proton coupling       | Pieri et al. (2009)                                                |
| Tyr167  | H5       | Substrate binding     | Yeung et al. (1998); Bolger et al. (1998); Kulkarni et al. (2003a) |
| Asn171  | H5       | Substrate binding     | Kulkarni et al. (2003a)                                            |
| Ser174  | H5       | Substrate binding     | Kulkarni et al. (2003a)                                            |
| Glu267  | IH2      | Conformational change | Xu et al. (2010)                                                   |
| Arg282  | H7       | Proton coupling       | Meredith (2004); Pieri et al. (2008)                               |
| Phe293  | H7       | Helix packing         | Kulkarni et al. (2003b)                                            |
| Trp294  | H7       | Substrate binding     | Bolger et al. (1998)                                               |
| Leu296  | H7       | Helix packing         | Kulkarni et al. (2003b)                                            |
| Phe297  | H7       | Helix packing         | Kulkarni et al. (2003b)                                            |
| Gly594  | H10      | Substrate binding     | Xu et al. (2009)                                                   |
| Glu595  | H10      | Substrate binding     | Xu et al. (2009); Bolger et al. (1998)                             |
| Tyr588  | H10      | Substrate binding     | Xu et al. (2009)                                                   |

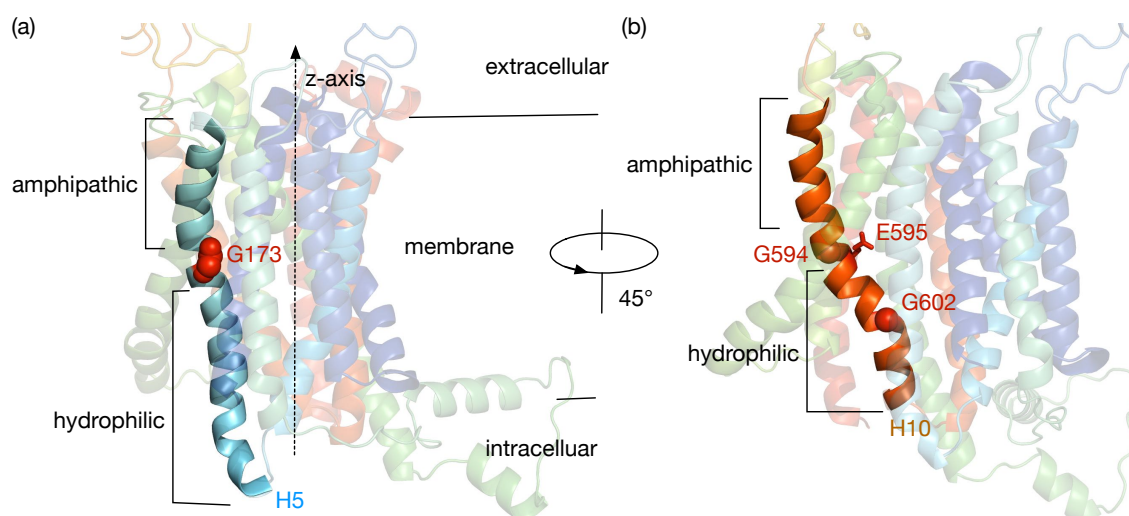
**Substrate translocation pathway** Considerable effort has been placed on characterising the peptide binding site of PepT1, as knowing the residues involved in binding will pave the way to understanding the structural requirements for a compound to become a substrate. The homology model built by Bolger et al. (1998) provided the first guide for mutational studies of residues along the translocation pathway. This model predicted that five TM helices, i.e. H1, H3, H5, H7 and H10 collectively form the putative “channel”, within which substrate binding occurs.

Most of these helices have been subject to cysteine scanning mutagenesis, where each residue along the helix is individually mutated to a cysteine and the effects on transport are measured in whole cell assays<sup>§</sup>. Kulkarni et al. (2003a) performed this analysis on H5 and found a significant decrease in transport for Y167C, N171C and S174C mutations. When these three residues were mapped on the homology model, they were indeed positioned around the binding site and their side chains pointing towards the hydrophilic cavity (Figure 5.10). Using substituted cysteine accessibility method (SCAM)<sup>¶</sup>, they proposed the helix is slightly tilted from the vertical axis, with the extracellular half facing both the hydrophobic membrane and the hydrophilic binding site (amphipathic), but the cytoplasmic half only exposed to the translocation pathway. Our homology model is in good agreement with their prediction, as illustrated in Figure 5.11(a), where there is a helix tilt facilitated by a glycine residue, Gly173.

A similar analysis was performed on H7 by Kulkarni et al. (2003b) where they found three detrimental mutations, i.e. F293C, L296C and F297C, that resulted in negligible uptake. Due to the hydrophobicity of these residues, the study suggested that they are not involved in substrate translocation, and instead are crucial for hydrophobic packing between helices. The homology model supports this prediction by showing that H7 is packed against H9, H10 and H12 via an extensive hydrophobic network involving these three residues and Phe372, Leu591, Phe598, Ile665 and Phe669 (Figure 5.12). These interactions hold together the C-terminal bundle. As H7 is proposed to be an essential part of the C-terminal gate in MFS transporters (Fowler et al., 2015), mutating any of these hydrophobic residues will disrupt the relative movement of the bundle during the transport cycle and therefore

<sup>§</sup>Due to the sensitivity of membrane proteins, a mutation at any position in a helix is likely, to some degree, to give a negative impact. As such, a residue is considered of interest only if it severely reduced or completely abolished protein function.

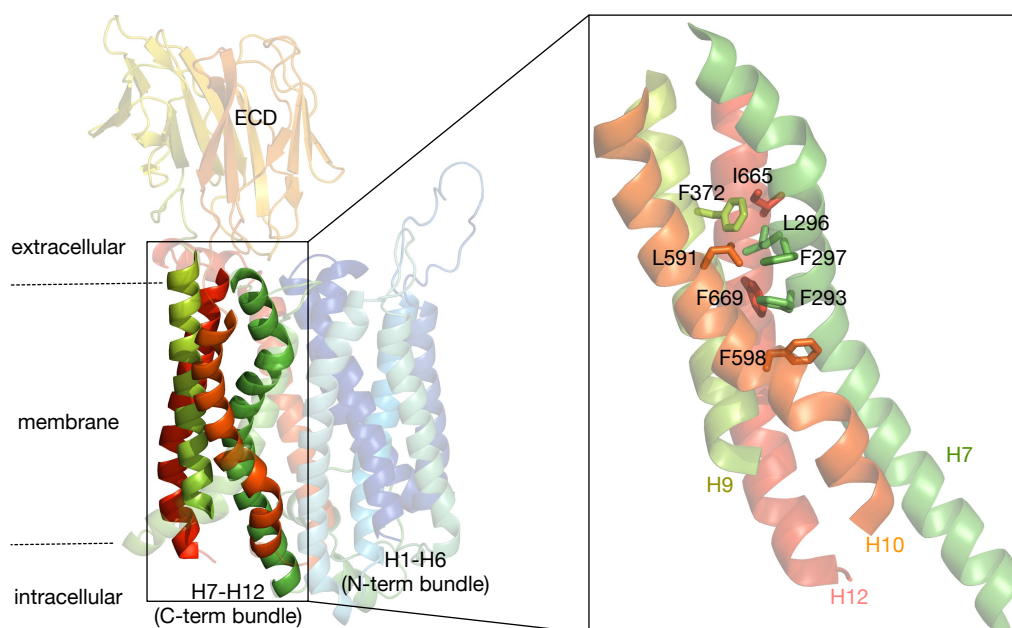
<sup>¶</sup>The SCAM technique (Karlin and Akabas, 1998) employs bulky methane thiosulfonate (MTS) reagent to determine whether a residue is solvent accessible or not. Upon a cysteine mutation, if a reaction with the MTS reagent reduces transport, the residue is assumed to be solvent accessible, as the bulky MTS will covalently attach to the cysteine and block substrate translocation pathway. This analysis is often performed in conjunction with cysteine scanning mutagenesis to map the topology of the TM domains, as has been done with LacY (Sahin-Tóth and Kaback, 1993) and GLUT1 (Hruz and Mueckler, 2000).



Figs. 5.11: Tilted helices in PepT1 homology model as predicted by experiment. SCAM studies by Kulkarni et al. (2003a) and Xu et al. (2009) demonstrated helical tilts in (a) H5 (cyan) and (b) H10 (orange), respectively. Glycine residues causing these kinks are displayed in red spheres. The position of Glu595, which is important for peptide binding and gating, is highlighted in a red stick representation.

eliminate transport. Kulkarni et al. (2003b), however, underestimated the length of H7 by almost twenty residues, therefore it is difficult to interpret the results of their SCAM analysis in the context of the homology model. Additionally, Bolger et al. (1998) highlighted the importance of another residue on H7, that is Trp294, for which a mutation to alanine completely abolished transport. The model is consistent with this result as Trp294 is positioned at the centre of the binding site (Figure 5.10). The equivalent residue in PepT<sub>St</sub>, Trp296, forms a hydrophobic pocket that accommodates the AlaPhe dipeptide in the crystal structure (Lyons et al., 2014). It is unclear, however, why a mutation to cysteine at this position does not affect binding (Kulkarni et al., 2003b).

Another helix important for gating is H10, for which a cysteine scanning mutagenesis study has been conducted by Xu et al. (2009). Three residues on this helix were found to be sensitive to mutation: Glu595, Gly596 and Tyr588, all of which, in the homology model, project into the binding cavity (Figure 5.10). Mutation of Glu595 to a neutral or basic residue eliminated transport, which points towards the importance of a negative charge at the position. From the sequence alignment earlier, it was proposed that Glu595 forms one of the intracellular gates for PepT1 (Figure 5.2(b) and Table 5.2). Glu595 interacts with Lys140 in the outward open state of the transporter, hence the need for an acidic residue at this position. The crystal structure of PepT<sub>St</sub> bound to AlaPhe shows that the equivalent residue, Glu400,



Figs. 5.12: Extensive hydrophobic interaction network involving the TM H7 of PepT1, as demonstrated by Kulkarni et al. (2003b). The residues of interest, Phe293, Leu296 and Phe297, form the core of the C-terminal bundle of the transporter.

is involved in coordinating the amino terminus of the dipeptide, which would also explain the importance of the negative charge. Xu et al. (2009) also predicted a tilt for H10, whereby much like H5, the exoplasmic half is amphipathic, while the cytoplasmic half is hydrophilic. Again, our homology model confirms this structural feature and illuminates that the kink is caused by two glycine residue, Gly594 and Gly602 (Figure 5.11(b)). Mutation of the former totally depleted peptide uptake, suggesting that the tilt at this particular position may be relevant to the positioning of the Glu595 residue in the binding cavity.

Links et al. (2007) performed cysteine scanning mutagenesis on H3 as it was also suggested to be a part of the substrate translocation pathway based on the homology model by Bolger et al. (1998). They showed that all but three mutations resulted in a significant loss of transport, and proposed that this helix packs against H5 via multiple  $\pi$ - $\pi$  stacking interactions of aromatic residues. In our homology model of PepT1, however, H3 was found to be one of the scaffold helices, which interacts primarily with the membrane and H1, H4 and H6 instead of forming the binding site. This arrangement is consistent with other MFS transporters with known crystal structures (Abramson et al., 2003; Solcan et al., 2012; Doki et al., 2013; Parker and Newstead, 2014). That H3 forms the core of the N-terminal bundle and thus facilitates its motion during the alternating access mechanism might instead

explain why any mutation to cysteine on this helix is damaging to transport.

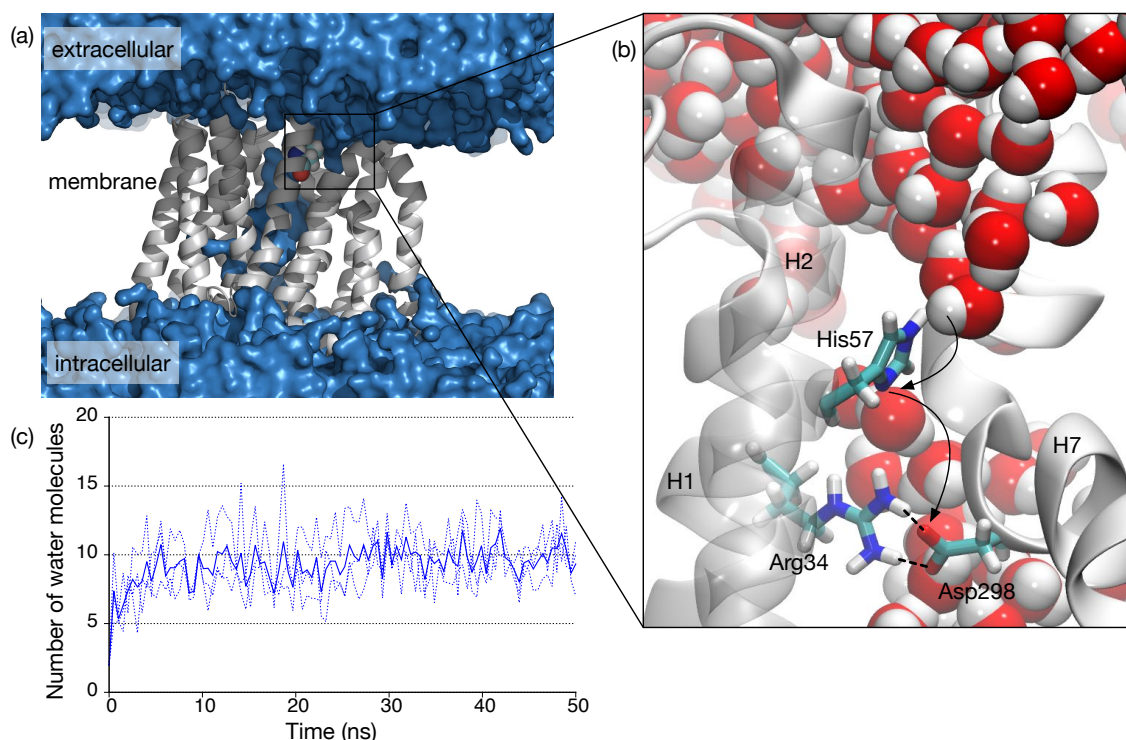
Bolger et al. (1998) also predicted H1 to be a part of the putative binding site, which agrees with our homology model, but surprisingly there are very few studies performed on this TM helix. The original positioning of this helix from residue 7 to 25 by Fei et al. (1998) has been controversial as a later study showed no protein expression when an epitope tag was inserted at residue 39, which was supposed to be in the extracellular loop according to this topology (Covitz et al., 1998). Bolger et al. (1998) performed point mutagenesis on residue Tyr12, which would be in the middle of the helix based on the topology, and found very little effect on transport. Our homology model predicts that H1 starts later in the amino acid sequence and spans from residue 16 to 45. As such, residue 39 is mid helix, which would explain the lack of expression when it was epitope-tagged, whilst residue 12 lies on the N-terminal loop, and hence might not be essential for peptide uptake. For PepT2, Terada et al. (2004) have looked at the functional consequence of a genetic variation in H1 involving residue Arg57 (Arg27 in PepT1) and discovered completely diminished transport activity with variant R57H. According to our model, Arg27 in PepT1 is oriented towards the centre of the hydrophilic cavity (Figure 5.10), and so is potentially involved in substrate binding. The arginine residue is also a part of a consensus motif in the POT family, the ExxERFxYY motif, found to make up a significant proportion of peptide binding site in the bacterial homologs (Newstead et al., 2011; Solcan et al., 2012; Doki et al., 2013; Guettou et al., 2013). As this motif is highly conserved from bacteria to man, it is highly likely that these residues perform a similar function.

**Protonation site** The uptake of substrate by PepT1 and PepT2 is coupled to the proton electrochemical gradient, therefore understanding where proton(s) binds on the protein is of paramount importance. Site directed mutagenesis and pH specific uptake assays are often used to identify protonation sites. Various evidence points towards the importance of histidine residues in transporters that utilise proton(s) as their driving force, such as the sodium proton exchanger (Ganapathy et al., 1987) and the folate transporter (Said and Mohammadkhani, 1993). There are nine histidine residues in PepT1, three of which are conserved in mammals: His57, His121 and His260. Mutation of His57 and His121 to glutamine eliminated the uptake of the antibiotic ceftibuten and  $^{14}\text{C}$ -Glysar in *X. laevis* (Terada et al., 1996). Fei et al. (1997) repeated this experiment in *X. laevis* and HeLa cells but reported conflicting results, whereby only His57 is obligatory for the functioning of PepT1, and mutation of the other two conserved histidines produced the same results as wild type. They also showed that His87 in PepT2, which is the equivalent of His57 in PepT1, is also essential for transport. Later, Uchiyama et al. (2003) provided more solid evidence by per-

forming uptake assays at different pHs: At pH 6, mutations of His57 to lysine or arginine stopped transport altogether, while at pH 7.4 the identical mutations did not alter transport behaviours. At low pH, basic residues, unlike histidine, cannot bind and release protons due to the higher  $pK_a$  of their side chains. Increasing the local pH closer to their  $pK_a$  will allow them to work as a histidine, and thus restored the transport properties. Their study therefore strongly suggests that PepT1 requires a proton donor/acceptor at this position to function properly.

In the PepT1 homology model, His57 is positioned adjacent to an extracellular cavity formed between H2, H7, H9, H11 and H12 (Figure 5.13(a)), as characterised by Newstead et al. (2011) in the crystal structure of PepT<sub>So</sub>. Intriguingly, upon examining the MD simulations of PepT1 in a POPC bilayer, this extracellular cavity was found to be fully hydrated throughout the entire simulation, essentially forming a water pocket in the vicinity of His57. As shown in Figure 5.13(c), it took around 5 ns for water to enter the cavity and the average number of water molecules occupying the pocket was around ten. Closer inspection suggests that this pocket was highly dynamic—water molecules entered and left continuously and no single molecule showed significantly long occupancy. Also nearby His57 is the proposed extracellular gate made of Arg34 and Asp298 (Figure 5.13(b)). As the model was built based on an inward open structure, this gate remained closed during the entire simulation. Taken together, it is likely that a water molecule from the extracellular cavity donates a proton to His57, which in turn protonates Asp298 to break the extracellular gate, and subsequently initiate a conformational change from inward open to outward open. A similar proton relay system was also observed in the crystal structure of a eukaryotic proton coupled phosphate transporter, PipT (Pedersen et al., 2013), in which a tunnel lined by a series of negatively charged residues exists to allow proton release into the cytoplasm.

Other residues that have been implicated in proton coupling are Tyr56, Tyr91 and Arg282. Chen et al. (2000) demonstrated that the transport of GlyLeu was severely reduced and lost completely for Y56F and Y56A mutants, respectively. As this residue lies adjacent to the primary protonation site, His57, they suggested that the phenolic side chain of Tyr56 is crucial to stabilise the positive charge on the protonated histidine. The other tyrosine residue, Tyr91, was suggested to play a similar role (Pieri et al., 2009). Mutation at this position to phenylalanine only reduced peptide uptake slightly and it is still coupled to proton gradient. Y91Q and Y91R mutants, on the other hand, showed a significantly lower transport that is no longer proton-dependent, i.e. same transport profile at different pHs, suggesting that the hydrophobicity of Tyr91 was required to somehow maintain a suitable environment for protonation to occur at His57. This speculation stems from the original PepT1 model

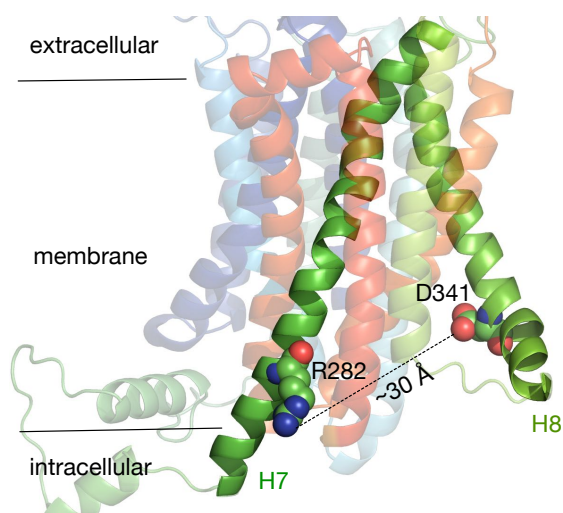


Figs. 5.13: The proposed protonation site of PepT1, His57, is positioned close to a hydrophilic extracellular cavity. (a) Structure of PepT1 homology model after 50 ns simulation in a model membrane, highlighting the water in surface representation. His57 is displayed in sphere representation. (b) Enlarged image of the hydrated extracellular cavity. Also in close proximity to His57 is the extracellular gate (Arg34-Asp298). Arrows indicate the suggested proton transfer from water to His57 to Asp298, that triggers the opening of the gate. (c) The number of water occupying the extracellular cavity during the 50 ns simulation. Data for three separate runs are shown as dotted lines, and for clarity, the average is shown as the solid line.

by Bolger et al. (1998), which proposed that Tyr91 is in close proximity to His57. In our homology model, however, Tyr91 is a part of H3, which is a scaffold helix, and it is not only far away from the histidine residue, but also not solvent accessible. As such, it is likely that this residue does not play a direct role in protonation. In contrast, the hydrophobicity of Tyr91 is perhaps a vital component for helix-helix packing within the N-terminal bundle, which explains the reduction in transport. As peptide uptake was so low for the glutamine and arginine mutants, the difference between experiments conducted at different pHs might be less discernible, leading to the misinterpretation that transport was independent of proton gradient.

A site directed mutation of Arg282 to a glutamate residue in rabbit PepT1 uncoupled peptide transport from the proton gradient and peptides could not be accumulated above



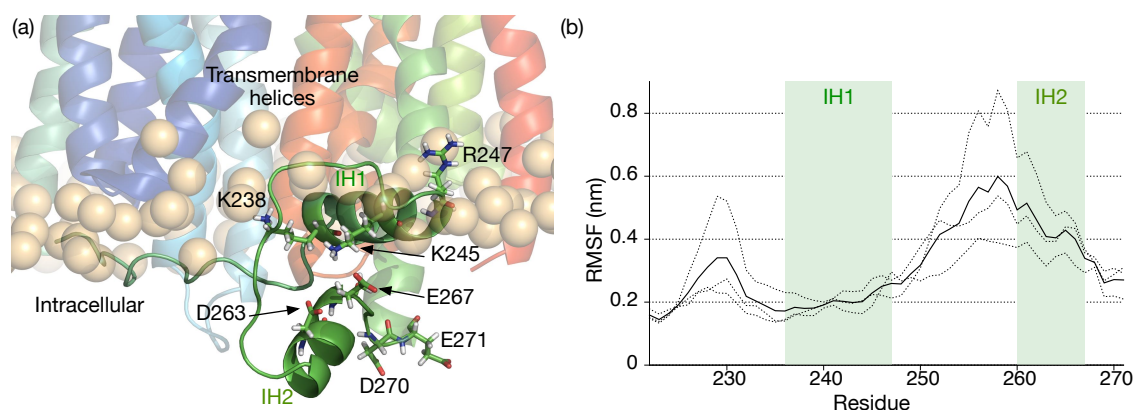


Figs. 5.14: The proposed charge pair interaction between Arg282 on H7 and Asp341 on H8 based on Kulkarni et al. (2007). The model predicted that these residues are found on the opposite end of the transporter and they are  $\sim 30$  Å apart from each other.

equilibrium, suggesting its possible role in proton coupling (Meredith, 2004). Later, Kulkarni et al. (2007) proposed a charge-pair interaction between this residue and Asp341 in human PepT1. Mutation of either residues reduced Glysar uptake, but a double mutant, R282D/D341R, restored wild type activity. Based on both studies, they speculated that this salt bridge exists in the outward open state and that translocated proton interacts with the salt bridge, breaks it and induces conformational change to inward open. In our homology model, however, these residues are  $30$  Å apart from each other, which makes forming a salt bridge interaction impossible (Figure 5.14). H7 and H8, on which they resided, are close to each other on the extracellular half, but due to a major tilt on both helices, the cytoplasmic half are positioned on the opposite sides of the transporter. Both residues are located on the cytoplasmic end of these helices. It is worth considering that this model is in the inward facing conformation, in which according to Kulkarni et al. (2007) the Arg282-Asp341 interaction is temporarily broken. It is possible that through a large conformational change into the outward open conformation, the cytoplasmic ends of these helices are brought together to form the suggested salt bridge. Nevertheless, this remains speculative until an outward open crystal structure of a POT transporter is solved.

**Intracellular helices** In the building of these homology models, the largest region without any structural template is the intracellular loop linking H6 and H7. Using secondary structural predictions, Xu et al. (2010) predicted two  $\alpha$ -helices in this region. Interestingly, the distribution of charged residues in this loop is asymmetric—the N-terminal half is rich in basic residues while all five acidic residues are found on the C-terminal half. Mutation of three basic residues, which according to the model all reside along the first helix, IH1, reduced Glysar uptake. This positively charged helix is positioned laterally and in the MD

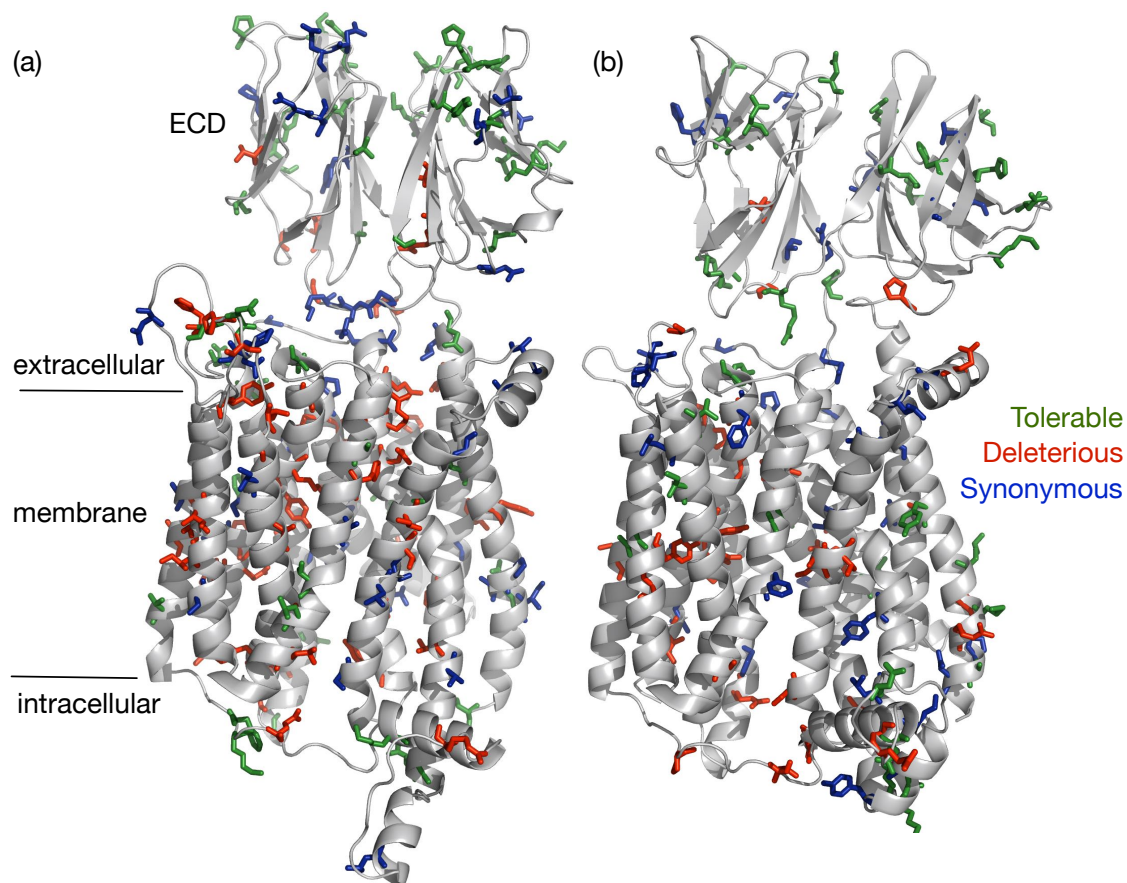




Figs. 5.15: Intracellular loop between H6 and H7 is made of two helices. (a) The position of residues that was subject to mutational studies by Xu et al. (2010). Phosphate head group are represented as orange spheres. (b) RMSF of all residues on this loop, highlighting the positions of IH1 and IH2 in green. Data from three separate runs are shown as dotted lines, whilst the solid line represents the average.

simulations, it interacted with the inner leaflet of the bilayer (Figure 5.15(a)). Quantitative lipidomics indicate that the inner leaflet of the apical epithelial membrane, wherein PepT1 is primarily expressed, is enriched with negatively charged phosphatidylserine (PS) (Gerl et al., 2012). Physiologically, it is possible that IH1 anchors PepT1 to the membrane via electrostatic interactions with the PS lipids and mutation of the basic residues therefore destabilises the protein and reduces transport efficiency. A similar lateral helix is also observed in the crystal structure of a plant homolog of the POT family, the nitrate transporter NRT1.1 (Parker and Newstead, 2014; Sun et al., 2014). That helix also contains potential stabilising anchors consisting of arginine and tryptophan residues, adding weight to our hypothesis.

Mutation of four negatively charged residues on the second intracellular helix, IH2, also impaired substrate uptake, especially Glu267 for which a mutation to alanine or lysine decreased transport by 80% compared to wild type (Xu et al., 2010). This helix is positioned in bulk solution and so is more mobile in simulations compared to IH1, as seen from the higher RMSF value (Figure 5.15(b)). In the NRT1.1 structure, no structural information could be resolved for this region, further suggesting that it is highly flexible. It is difficult to interpret how this mobile intracellular helix affect transport at all, but Xu et al. (2010) suggested IH1 and IH2 interact with each other via electrostatic interactions, drawing the N- and C-terminal bundles of the protein together to form an outward open state. Again, this assumption requires a structure in the outward facing state for validation.



Figs. 5.16: Single nucleotide polymorphisms (SNPs) of the *SLC15A1* and *SLC15A2* genes mapped onto the homology model of (a) PepT1 and (b) PepT2, respectively. Tolerable, deleterious and synonymous SNPs are coloured in green, red and blue, respectively.

**Genetic polymorphism** The 1000 genomes project catalogs human DNA sequence variations for genetic studies to understand inherited diseases, evolution and even human history (The 1000 Genomes Project Consortium, 2010). From their public database, single nucleotide polymorphism (SNP) data were retrieved for genes encoding PepT1 and PepT2, i.e. *SLC15A1* and *SLC15A2*, respectively. *SLC15A1* has 127 missense and 57 synonymous variants in the coding region, whilst for *SLC15A2*, there are 222 missense and 92 synonymous variants. The number of non-synonymous SNPs does not translate directly to the number of amino acid variants in a protein as some variations at DNA level can occur at the same amino acid residue. Indeed, at protein level, there are more variations in PepT1, despite *SLC15A1* having less SNPs. These variants were plotted onto the homology model and colour them according to their predicted effects<sup>||</sup>(Figure 5.16). As expected, most of the

<sup>||</sup>Genetic variations are categorised into tolerable and deleterious based on two algorithms: Sorting Intoler-

deleterious SNPs are clustered in the TM regions, whilst most tolerable SNPs are located in the loop regions, including the ECD. The TM domains are important for the protein function, therefore variations in this region are potentially damaging. The loop regions, on the other hand, are more tolerant to changes as they are not directly involved in substrate uptake, suggesting that they might play some role in determining different transport capacity between individuals.

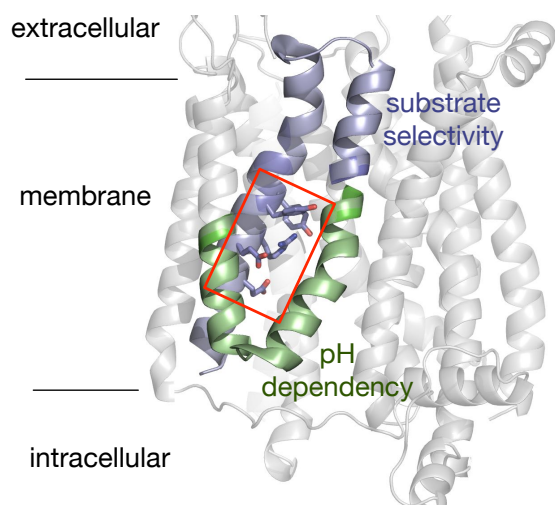
### 5.3.2 Comparison of PepT1 and PepT2

The two primary isoforms of human peptide transporters share around 50% overall sequence identity and 60% identity within the TM region. In spite of their similarity in the primary amino acid sequence, they show fundamentally different characteristics in peptide uptake—PepT1 is a low-affinity, high-capacity transporter, while in contrast PepT2 is a high-affinity, low-capacity transporter (Ganapathy et al., 1995; Smith et al., 2013). Additionally, PepT2 is more sensitive to extracellular pH as its activity decreased 5-fold from pH 6.0-8.0, whereas the same pH changes only reduced PepT1 uptake by a half (Boll et al., 1994, 1996). Nevertheless, there are currently no concrete structural explanations to account for these differences. In the past, a series of chimera studies were conducted to address this question, whereby varying domains of PepT1 and PepT2 were exchanged to delineate which part of the transporter makes them different from one another. Here, the results of these studies will be discussed in the context of our homology models, and other possible regions of conserved sequence differences will be suggested for future studies.

**Chimera studies** Similar to cysteine scanning mutagenesis and SCAM analysis, chimeric transporters were constructed to elucidate the putative substrate binding site, but more specifically which part of the binding site contributes to the different apparent substrate affinities between PepT1 and PepT2. Most chimera studies suggest the importance of the N-terminal of the transporters towards their phenotypic differences. Doring et al. (1996) showed that a transporter with H1-H9 from PepT2 with ECD and H10-H12 from PepT1 functions like PepT2. Further corroborating their results, Fei et al. (1998) narrowed down the responsible region to H7-H9. Their HR2 construct (H1-H9 and ECD from PepT1 with H10-12 from PepT2) behaved like PepT1, whilst HR3 construct (H1-H6 from PepT1 with H7-H12 and ECD from PepT2) showed the characteristics of PepT2. However, Terada et al. (2000a) demonstrated contradicting results. They showed that if a chimera has the N-terminal bundle (H1-H6) from PepT1, it functioned like PepT1, and likewise for PepT2,

---

ance from Tolerance (SIFT) (Shen et al., 2006) and Polymorphism Phenotyping (Polyphen) (Adzhubei et al., 2010).



Figs. 5.17: The proposed regions responsible for the difference in substrate selectivity (purple) and pH dependency (green) of PepT1 and PepT2 based on a chimera study by Döring et al. (2002). The position of ExxERFxYY motif is highlighted in the red box.

indicating the role of the first six helices instead. Later, Döring et al. (2002) identified a small region (residue 1-59) that contributes significantly towards different substrate affinities found in the mammalian peptide transporters. Only three of these four studies looked at the difference in pH dependency (Döring et al., 1996; Terada et al., 2000a; Döring et al., 2002). They unanimously agreed that the N-terminal region is associated with adapting to the extracellular pH changes, with one study by Döring et al. (2002) specifically isolating residues 60-91.

These results should be interpreted with caution, not only due to the conflicting data, but also because mixing two different proteins, no matter how similar, is likely to introduce confounding variables to the experiment. The first chimeric transporter by Döring et al. (1996) used residues 1-401 from PepT2 and residues 402-707 in PepT1. When plotted on the homology models, however, there are almost ten residues missing in the ECD region due to the fact that PepT1 ECD starts at position 395, raising serious doubts regarding the reliability of their results. Also, these four studies used a range of different compounds as reporter substrates, and therefore comparing their results is difficult. Having said that, the position of residues, which have been proposed by Döring et al. (2002), to play a role in determining affinities and pH dependency were mapped (Figure 5.17). Residues 1-59 spans H1 with part of H2 and includes the consensus ExxERFxYY motif. This motif contains two tyrosine residues implicated in determining substrate specificities in PepT<sub>St</sub> (Solcan et al., 2012), lending further support to their hypothesis. The position of the proposed pH sensor, however, is a bit dubious. Residues 60-91 lies on the cytoplasmic side of H2 and H3, with no exposure to the extracellular side at all. This stretch of protein also does not contain any histidine residue, which is usually associated with pH sensing, suggesting that these results

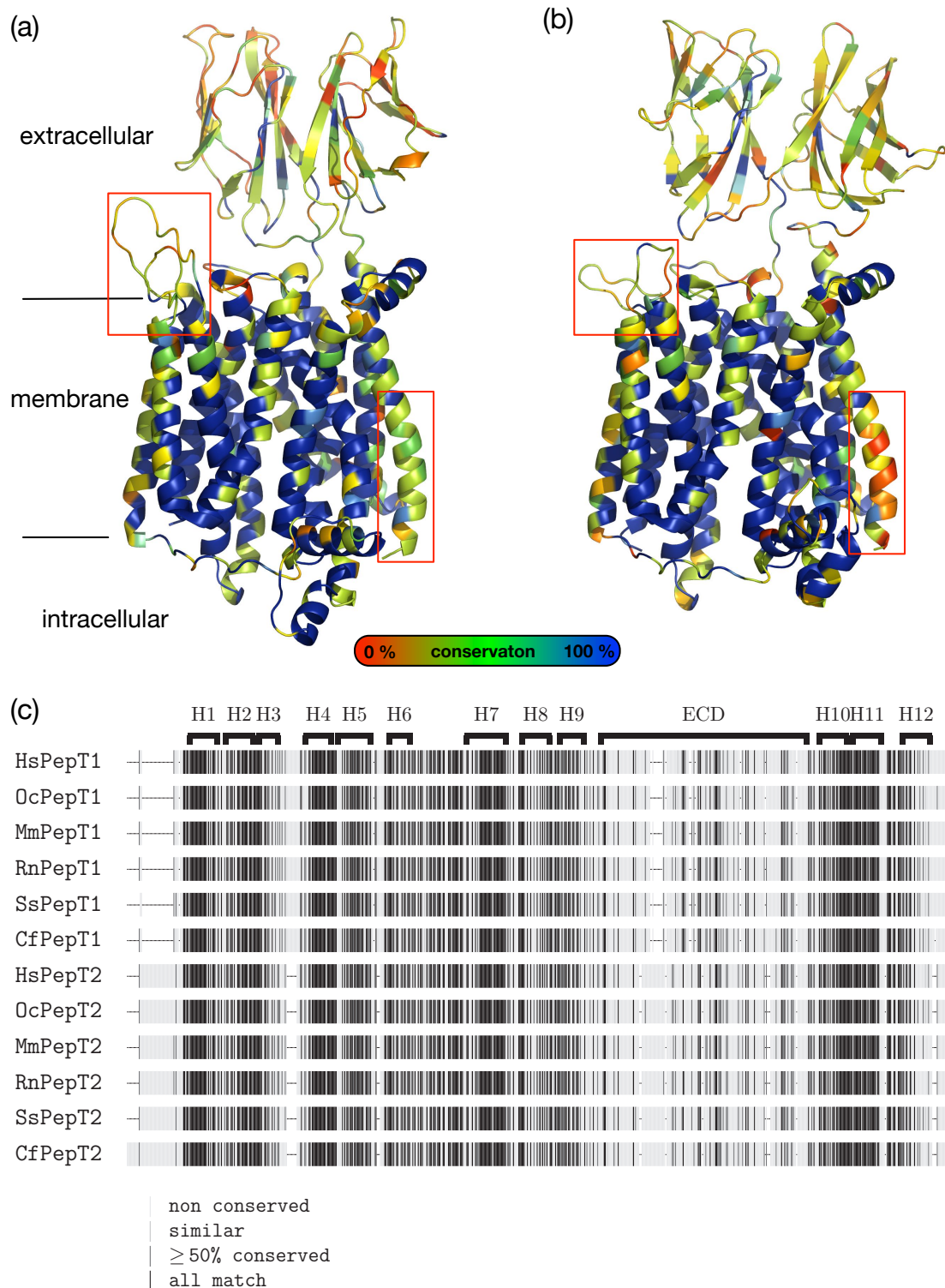
Table 5.5: Percentage amino acid sequence identity for mammalian PepT1 and PepT2

| Region | Sequence Identity |
|--------|-------------------|
| H1     | 91.2              |
| H2     | 88.1              |
| H3     | 77.7              |
| H4     | 92.5              |
| H5     | 83.9              |
| H6     | 85.2              |
| H7     | 88.9              |
| H8     | 85.5              |
| H9     | 83.2              |
| ECD    | 14.7              |
| H10    | 92.1              |
| H11    | 84.8              |
| H12    | 55.5              |

demand further verification.

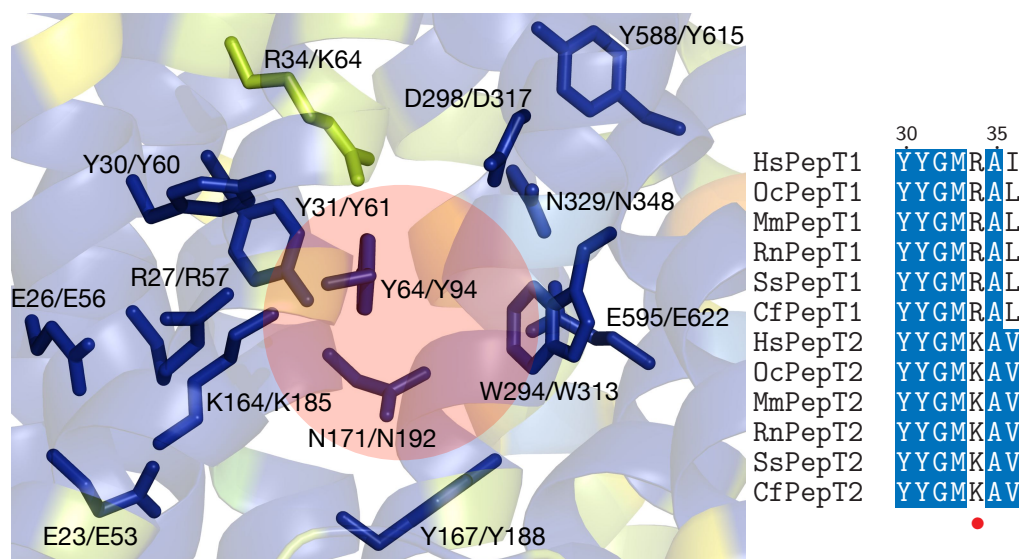
**Sequence analysis** The lack of any meaningful results from these chimera studies signifies a big gap in our current understanding of the different phenotypic behaviours of the two peptide transporters. In light of this, an analysis based on a multiple sequence alignment of PepT1 and PepT2 were performed to find conserved similarities and differences between them, and subsequently map these onto the homology models (Figure 5.18). The sequences of PepT1 and PepT2 from six mammals were used: human, rabbit, mouse, rat, pig and dog. The percentage of sequence identity for each TM helix and the ECD between these twelve sequences is shown in Table 5.5.

As expected sequence conservation is highest in the TM domains, with the helices forming the binding site showing the best homology. When key residues in substrate binding as suggested by various mutational studies of PepT1, PepT2 and bacterial POTs were mapped, all but one residue are totally conserved (Figure 5.19). In PepT1, the N-terminal extracellular gate is formed by residue Arg34, but this residue is replaced by a lysine in PepT2. Whilst both residues are positively charged, it is still very intriguing to see that they are conserved differently in both transporters. Further analysis was conducted by adding 20 other sequences from vertebrates and it was found that this pattern prevailed. A residue at this position is proposed to be involved in controlling the extracellular gate; it could be that the arginine in PepT1 functions more efficiently than the lysine in PepT2, thereby increas-



Figs. 5.18: Sequence conservation between mammalian PepT1 and PepT2 mapped on their homology models ((a) and (b)) and coloured in rainbow, where red represents no conservation in either transporter, green represents conserved differently in each transporter and blue represents conserved similarly in both. Red boxes highlight regions of conserved sequence difference. (c) Sequence alignment used in this analysis, highlighting the regions conserved in both transporters in dark grey.





Figs. 5.19: Comparing the sequence conservation in the binding site of PepT1 and PepT2. Key residues in peptide binding are shown in stick representation and coloured using the same colour scheme as Figure 5.18. The approximate position of the binding site is indicated by the red circle. Multiple sequence alignment around the Arg34/Lys64 (marked by the red bullet) is shown on the right. Conserved residues are coloured in blue.

ing the rate of peptide uptake and making it the high-capacity transporter. Similarly in the nitrate transporter NRT1.1, phosphorylation alters the efficiency of the formation and breakage of a salt bridge gate and this significantly affects the kinetics of nitrate uptake (Parker and Newstead, 2014).

While the first eleven helices are highly conserved between the two isoforms, the most distal helix, H12, shows a degree of sequence divergence especially on the cytoplasmic half. According to the homology model, H12 is a scaffold helix and the cytoplasmic end interacts with H7 and H9, with one side exposed to the membrane. The cytoplasmic ends of H7 and H9 in PepT1 and PepT2 share a high sequence similarity, so it is difficult to draw a meaningful conclusion as to why sequence of H12 diverges. Additionally, chimera studies discussed previously demonstrated that this helix does not affect the characteristics of the transporter (Fei et al., 1998).

Apart from the ECD, two other extracellular loops display sequence divergence. The loop linking H3 and H4 in PepT1 has a considerably conserved 9-residue insert, making it significantly longer than the same loop in PepT2. This region is negatively charged as it contains up to four aspartate residues and the current model predicts it to be unstructured as there is currently no structural information available (the same loop in PepT<sub>So</sub> is comparably shorter). A neighbouring loop between H5 and H6 in PepT1 also has a conserved sequence

insertion that is absent in PepT2, albeit only made of three residues. These two inserts in PepT1 are relatively short and most probably do not have any secondary structure, so it is hard to imagine that they would alter the behaviour of the transporter. However as they are positioned extracellularly, they could still be a target to study the difference in pH dependency between the two transporters, more so as both inserts contain histidine residues.

Another region of conserved difference is the intracellular N-terminus, which was not modelled in this thesis due to the lack of structural template. For PepT1, H1 starts at residue 15, whereas the beginning of H1 in PepT2 is at position 45, leaving PepT2 with 30 extra residues in its N-terminal region. Interestingly, this long N-terminus is highly conserved amongst mammalian PepT2 sequences. A secondary structural prediction using JPred (Cole et al., 2008) predicted two  $\beta$ -strands in this region, suggesting that it might not be just a long unstructured loop. While the function of this domain is not yet known, it is unlikely to have a direct influence on peptide transport due to its distant position from the substrate binding site. Another MFS transporter with a similar terminal domain outside of the TM core is the drug efflux protein from *E. Coli*, YajR, which has a 65-residue C-terminal ferredoxin-like domain, the function of which has yet to be elucidated (Jiang et al., 2013).

### 5.3.3 Interactions of ECD with TM region and membrane

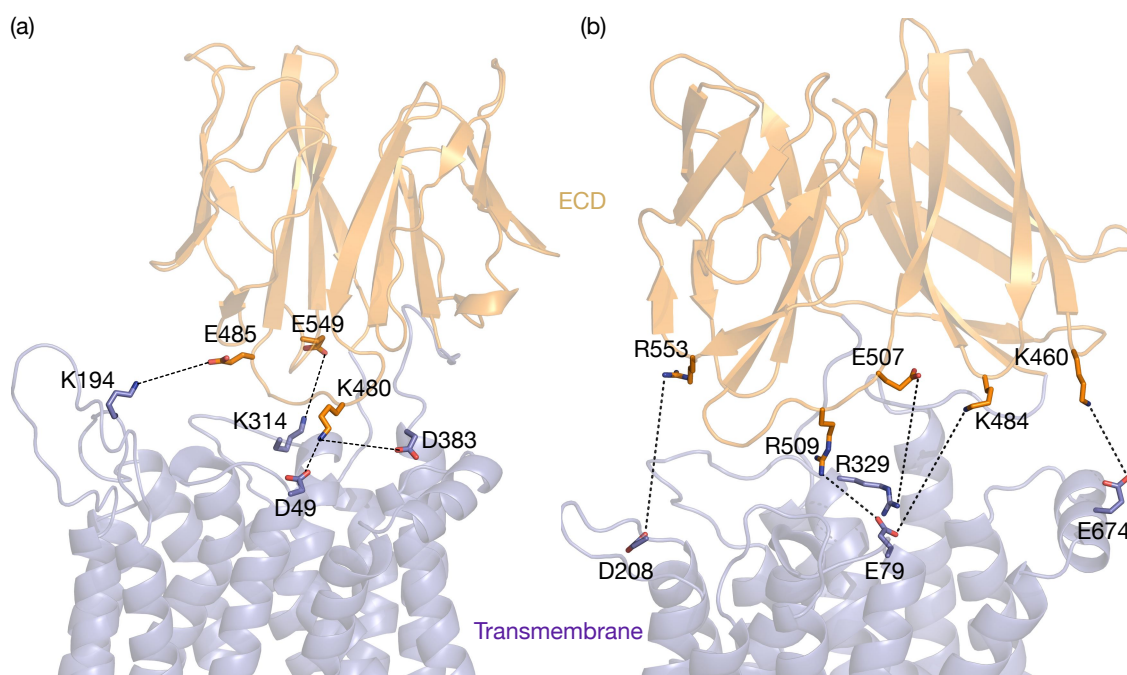
One of the key strengths of our homology models is the accurate structural representation of the large extracellular loop between H9 and H10, i.e. the ECD. The crystal structures of the ECDs from mouse PepT1 and rat PepT2 reveal two compact  $\beta$ -sandwich immunoglobulin-like folds linked by a short connecting loop (Beale et al., 2015). Despite sharing only 22% sequence identity, the two ECDs closely resemble each other with a C $\alpha$  RMSD of 1.8 Å. These structures, for the first time, allow for accurate homology modelling of the full-length transporters. The MD simulations suggest the ECD resides on one side of the transporter and adopts a vertical orientation relative to the plane of the membrane (Figure 5.9). Here a more thorough analysis of how the ECD interacts with the rest of the protein and the surrounding membrane bilayer were performed.

Using VMD (Humphrey and Dalke, 1996), the 50 ns simulations were analysed to identify salt bridge interactions made by residues on the ECD with the rest of the protein. How long each interaction lasted during each simulation was determined and the average over three replicates was calculated. Interactions with a significant percentage duration of contact are listed in decending order in Table 5.6. The positions of these salt bridges are mapped onto the models in Figure 5.20.



Table 5.6: Salt bridge interaction between ECD and TM domain

| PepT1       |           | PepT2       |           |
|-------------|-----------|-------------|-----------|
| Salt bridge | % contact | Salt bridge | % contact |
| K480-D383   | 32.2      | R509-E79    | 35.6      |
| E485-K194   | 15.2      | R553-D208   | 27.3      |
| E549-K314   | 14.8      | K484-E79    | 18.2      |
| K480-D49    | 1.7       | K460-E674   | 16.7      |
|             |           | E507-R329   | 2.5       |
| Total       | 63.9      | Total       | 100.3     |



Figs. 5.20: Salt bridge interactions made by residues in the ECD (orange) of (a) PepT1 and (b) PepT2 with the rest of the protein (purple) as listed in Table 5.6. This is based on three repeats of 50 ns MD simulations.

In general, the ECD interacted mostly with residues in other extracellular loops or residues from the end of a TM helix. These interactions, however, were mostly transient with only three and four salt bridges persisting for longer than 5 ns (or 10% of the simulation) for PepT1 and PepT2, respectively. One possible reason for this is the potential low accuracy of the homology models around the extracellular loops as they lack any structural template. Due to their poor coordination with surrounding residues and their exposure to the bulk solution, these loops were highly mobile in simulation, which aggravated any potential interaction with the ECD. Comparing PepT1 and PepT2, it was found that the ECD in PepT2 made more contacts with the rest of the protein, and these contacts, in total, lasted longer. These salt bridge interactions restricted the movement of the ECD during the simulations, contributing to the lower overall RMSF value of PepT2 (Figure 5.8).

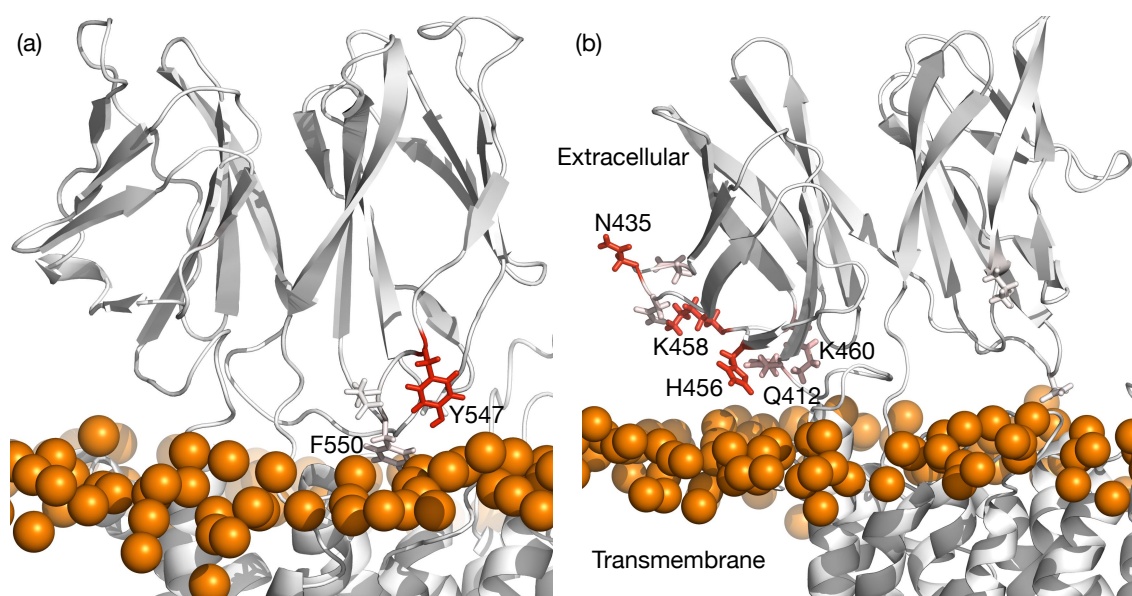
Of potential interest are the contacts made between the ECD and gating residues in the TM domain. In PepT2, the most persistent salt bridge was between residue Arg509 in the ECD and Glu79 in H2. The latter, interestingly, is proposed to be the N-terminal extracellular gate residue based on a sequence alignment to PepT<sub>So</sub> (Figure 5.2 and Table 5.2). A similar interaction is also observed in PepT1, that is Lys480 in the ECD forms a salt bridge with Asp49 in H2, albeit with a much shorter duration. As both Lys480 and Arg509 are highly conserved in mammalian PepT1 and PepT2, respectively, it is possible that they control the opening and closing of the extracellular gate and, by extension, the transport cycle. Additionally, the sequence alignment also revealed that Asp49 in PepT1 lacks any interacting partner on the C-terminal side of the TM domain as the closest positively charged residue is Arg314, positioned five residues away due to an insertion of a highly conserved sequence of TTMxG, which is absent in PepT2 or PepT<sub>So</sub>. This lends further support that the ECD could be essential for peptide transport, at least in PepT1. Whether the same applies to PepT2 is still debatable, as the N-terminal gate, Glu79, already has a partner on the C-terminal side from H2 (Arg329), making another gate from the ECD seem redundant.

To characterise the interaction between the ECD and the membrane, the interquant VMD plug-in (Humphrey and Dalke, 1996) was used with lipid residues on the upper leaflet set as “origin” and all residues on the ECD set as “target”. A cut-off of 4 Å was applied. Again, the length of each contact was measured and it was averaged over three replicates of 50 ns simulations. The list of residues that interact with the membrane is shown in Table 5.7 and mapped to the homology models in Figure 5.21.

Just like the interactions with the rest of the protein, the ECD contacted the lipid bilayer in a transitory manner. In total, PepT1 ECD made more interactions with the membrane, mostly mediated by residue Tyr547. Upon further inspection of the multiple sequence align-

Table 5.7: Interaction of ECD with membrane bilayer

| PepT1   |           | PepT2   |           |
|---------|-----------|---------|-----------|
| Residue | % contact | Residue | % contact |
| Y547    | 28.6      | H456    | 3.4       |
| F550    | 2.8       | N435    | 3.3       |
| S523    | 1.5       | K458    | 3.1       |
| E549    | 0.8       | Q412    | 1.0       |
|         |           | K460    | 1.0       |
|         |           | N433    | 0.5       |
|         |           | N436    | 0.4       |
|         |           | N512    | 0.2       |
|         |           | L574    | 0.1       |
| Total   | 33.7      | Total   | 13.0      |



Figs. 5.21: Interactions made between residues in the ECD of (a) PepT1 and (b) PepT2 with the inner leaflet of the membrane bilayer. Residues are coloured based on the duration of contacts made with the membrane, from red (longest interaction) to white (no interaction) (Table 5.7), based on three repeats of 50 ns MD simulations. The position of the phosphate head group of the membrane is highlighted as orange spheres.

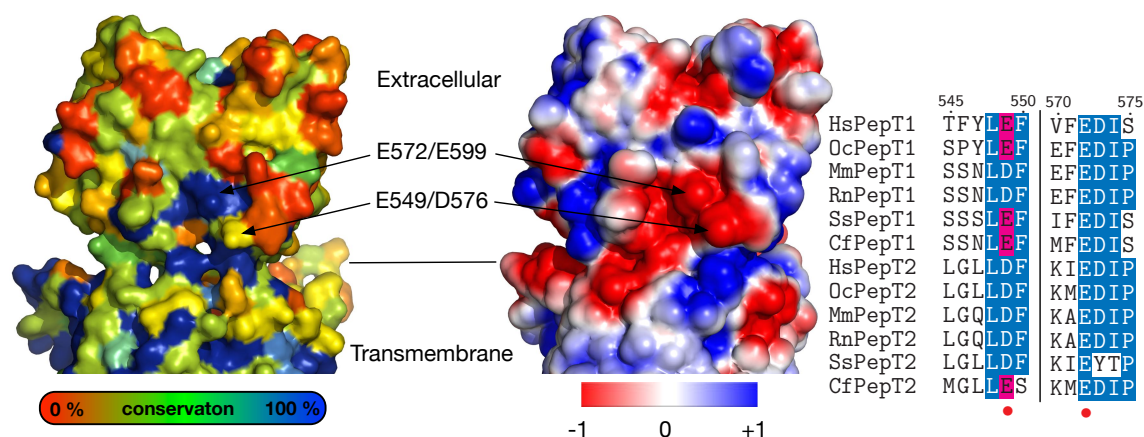
ment, all mammalian PepT1 sequences have a polar residue at this position, suggesting it might play a role in anchoring the ECD to the membrane via a hydrogen bond to a phosphate head group. Another residue of interest is Phe550 that potentially forms hydrophobic interactions with lipid tails and is also totally conserved. Significantly less contacts was observed between the ECD of PepT2 and the membrane. This is not surprising as the ECD made more interactions with the rest of PepT2 as discussed earlier, and therefore was tilted more toward the TM region of the protein. Some residues showed more prominent contacts with the membrane than others, like His456, Lys458 and Asn435, although these were all very short-termed (around 1-2 ns each) and thereby potentially non-specific.

### 5.3.4 PepT1 and PepT2 ECD bind to trypsin

Using two-electrode voltage clamp technique, our collaborators measured the transport characteristics of PepT1 and PepT2 and demonstrated that the ECD is dispensable for peptide and drug transport (Beale et al., 2015). Deleting the ECD from PepT2 or swapping the region from PepT1 to PepT2 did not change the apparent affinity for the uptake of Glysar or the antibiotic cefaclor. Deleting the ECD from PepT1 or swapping the region from PepT2 to PepT1, on the other hand, resulted in a poor level of expression, suggesting it plays a structural role and is required for the correct folding of the protein. As such, whether the ECD affects the transport of PepT1 remains to be determined.

Given the immunoglobulin-like structure of the ECD and the intestinal/renal lumen environment in which these transporters are expressed, it is possible that the ECD interacts with other extracellular protein(s) like enzymes and hormones. Our collaborators performed binding assays using surface plasmon resonance (SPR) and microscale thermophoresis (MST) and found that this domain binds to the intestinal protease trypsin with a  $K_D$  of around 8  $\mu$ M and 6  $\mu$ M for PepT1 and PepT2, respectively. Further experiments with a higher salt concentration abolished this interaction, suggesting that the ECD-trypsin binding is mediated by an electrostatic interface. Using protein-protein docking, this binding was then modelled and subsequently the physiological relevance of the interaction to the function of mammalian PepT1 and PepT2 was suggested.

Mapping sequence conservation for six mammalian homologs on the mouse PepT1 and rat PepT2 ECD crystal structures, a patch of highly conserved negatively charged surface was found. Interestingly, if the full-length transporter is considered, this conserved surface does not lie directly above the TM domain, but on the opposite side facing the bulk solution. It does make sense, however, that trypsin should not bind directly above the TM region



Figs. 5.22: Identifying a possible binding site for the interaction of PepT1 and PepT2 ECD with trypsin, based on sequence conservation. Human PepT1 homology model is shown in surface representation, whereby on the left it is coloured based on sequence homology, and on the right based on electrostatics. Multiple sequence alignment used in this analysis is shown highlighting the two conserved negatively charged residues by the red bullets.

as this will obstruct the peptide translocation pathway and therefore hinder peptide binding and transport. Two key residues were identified: Asp550 and Glu573 in mouse PepT1 and Asp576 and Glu599 in rat PepT2. Mutations of these residues to alanine severely reduced the strength of interaction with trypsin, lending further support that this is a plausible binding site. The equivalent residues in human PepT1 and PepT2 are Glu549, Glu572 and Asp576, Glu599, respectively, as illustrated in Figure 5.22.

We then performed protein-protein docking on the ClusPro web server (Comeau et al., 2004) using these two negatively charged residues as a guide, to build a model of full-length PepT1-trypsin complex. This model was simulated inside a POPC bilayer to determine its relative position with respect to the plasma membrane. As illustrated in Figure 5.23(a), trypsin sits between the ECD of PepT1 and the membrane with its active site oriented towards the bulk solution. Based on this snapshot, a working model for a multi-protein complex of human peptide transporters with trypsin enzyme emerges. Long-chained polypeptides from acid hydrolysis and pepsin digestion in the stomach enter the small intestine, in which they bind to and get cleaved by trypsin to form shorter peptides. Trypsin is a specific enzyme that recognises positively charged residues and cleaves the downstream peptide bond, essentially producing peptides with lysine or arginine on its C-terminus. Di- and tripeptides containing these basic residues are poor substrates of peptide transporters possibly due to the less optimal binding mode as has been suggested in §4.3.4. Therefore, tethering trypsin using the ECD confers a huge advantage to the transporter as it can in-

crease the local concentration of these low affinity peptides, and thereby improving uptake. Negatively charged patches exist adjacent to the trypsin catalytic site and above the peptide translocation pathway to retain these basic peptides until they bind to the transporter and get transported into the cytoplasm (Figure 5.23(b)). Nevertheless, more robust structural and biophysical studies involving the full length transporter are required to verify this model.

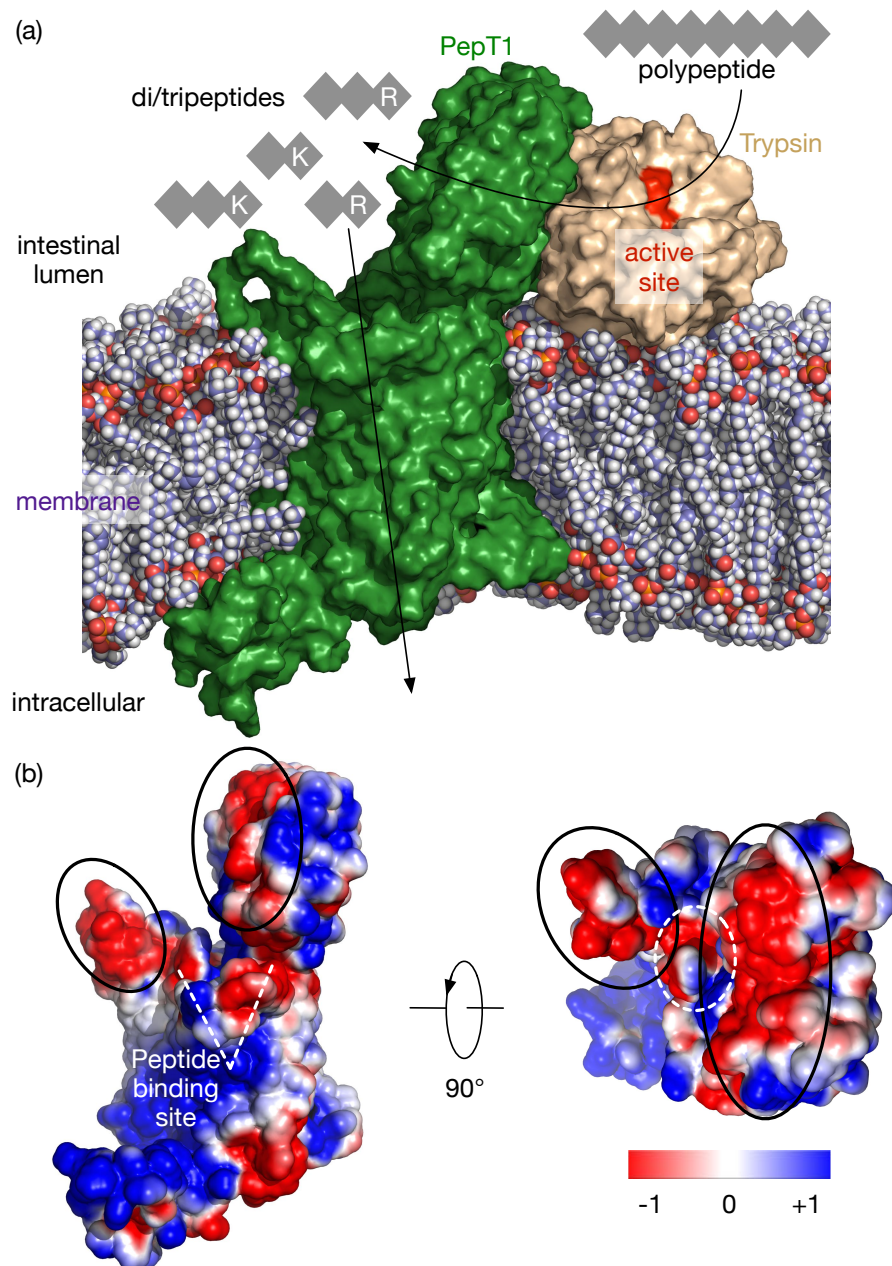
## 5.4 Conclusion

Here, full length homology models for the human peptide transporters, PepT1 and PepT2, were built using as templates the crystal structures of a bacterial POT transporter, PepT<sub>So</sub>, and the extracellular domain of mouse PepT1 or rat PepT2. To date, these models are the most complete and accurate three dimensional representation of these transporters owing to the use of a template with a high sequence identity for the TM domain and, more importantly, the crystal structures of the mammalian ECDs that were absent in previous models. The only large regions without any structural templates are the N- and C-termini, although arguably these could be unstructured loops, and the intracellular loop between H6 and H7 that has been proposed to contain two  $\alpha$  helices (Xu et al., 2010). The homology models were then thoroughly vetted using stereochemical and geometrical evaluations, as well as molecular dynamic simulations, which all indicated that they are structurally robust.

One key implication of the current modelling study is the ability to accurately map residues that, through various independent studies, are proposed to be essential for peptide transport. Knowing their positions with respect to the binding site provides possible explanations for their functional roles. Most mutational studies so far used the crude homology model by Bolger et al. (1998) as a guide. This model, however, assumed the binding site is made of H1, H3, H5, H7 and H10, which according to an increasing number of crystallographic structures of POT transporters, and MFS transporters in general, is incorrect. Additionally, the starting position of H1 was also wrongly predicted. As a result, there are several regions of interest, based on our homology models, that are still left unexplored. One good example is the consensus ExxERFxYY motif on H1 that forms the majority of the substrate binding site. Numerous studies on bacterial POTs have shown its importance in peptide recognition and selectivity but to date, no studies have been performed on these residues in PepT1 or PepT2. Our new homology models, therefore, will not only allow for better interpretations of available mutational data but most importantly, can become a reference to design future experiments.

PepT1 and PepT2 belong to the same transporter family, but functional studies have





Figs. 5.23: A model of the putative PepT1-trypsin complex was built by docking human trypsin crystal structure to PepT1 homology model using ClusPro (Comeau et al., 2004). (a) The relative position of the complex inside a membrane bilayer, highlighting the position of trypsin active site in red and the proposed mechanism of cooperativity between both proteins. (b) Negatively charged patches exist directly above the putative peptide binding site of PepT1, suggested to be involved in maintaining a high local concentration of positively charged peptides for more efficient transport.

shown that they have distinct transport kinetics (Boll et al., 1994, 1996). Efforts to pinpoint the structural origins of these differences have so far been inconclusive. This is especially challenging for the current study given the homology models for both proteins are very similar. Nonetheless, there are a few interesting regions in the binding site and around the extracellular loops that were highlighted, for which different residues are conserved between the two isoforms. It is possible that both transporters share a very similar overall architecture and that their distinct characteristics arise from small local differences. The nitrate transporter, NRT1.1, which is a plant homolog from the POT family, switches from low-affinity to high-affinity transport mode just by phosphorylation of a tyrosine residue that causes a localised disruption to the N-terminal bundle (Parker and Newstead, 2014)—a good example of how a small structural alteration can reverberate across the whole transporter and significantly influence its fundamental behaviour.

*In vitro* binding assays revealed that the ECDs of PepT1 and PepT2 show physiologically relevant affinity for the intestinal protease trypsin. The interaction of trypsin with the intestinal mucosa has been documented as far back as 1968, when it was suggested to be adsorbed to the cell surface membrane (Goldberg et al., 1968). Our homology models allow for the building of a PepT1-trypsin complex, subsequently providing a first look at how trypsin is potentially tethered to the membrane bilayer via the ECD. It was proposed that by capturing trypsin locally, PepT1 increases its efficiency of transport for di- and tripeptides with basic residues. This is, to date, a unique example of an MFS transporter evolving to include an additional functional domain outside of its canonical 12 TM helices in order to assist uptake activity.

Ultimately, the interest to study human peptide transporters comes from their ability to transport pharmacologically relevant compounds, and hence enhance oral drug delivery. Fitting with the overarching theme of this thesis, the PepT1 homology model shall now be used, along with the previously validated free energy methods, to study drug binding.



## Chapter 6

# Peptide and drug binding in human PepT1

The binding of seven peptides and 14 drugs to a homology model of human PepT1 was modelled based on the crystal structures of peptide bound complexes of the bacterial POT, PepT<sub>St</sub>. A scoring function and three end-point free energy methods were evaluated to estimate the binding affinities of these compounds, but poor correlation to *in vivo* competition transport assays was observed. Our results clearly indicate the need for a high resolution crystal structure to conduct a reliable protein-ligand simulation and accurately estimate the free energy of binding. Results from previous studies on the bacterial homolog, PepT<sub>St</sub>, was used as a basis to suggest how PepT1 recognises drug compounds. The successful prodrug approach may stem from the attached amino acid forming a pseudo N-terminus that anchors the drug to the transporter whilst a large hydrophobic pocket in the binding cavity acts as a refuge for active drug moieties. The presence of His57 as a part of the extracellular gate may facilitate the uptake of peptidomimetics by allowing a more efficient transition of conformational states.

### 6.1 Current state of drug design for PepT1

In the past decades, the field of computational chemistry has contributed significantly to the early stage of drug development by means of ligand-based modelling techniques, such as pharmacophore mapping and three-dimensional quantitative structure-activity relationship

(3D-QSAR) analyses (Caporuscio and Tafi, 2011). These approaches rely on identifying the physico-chemical similarities between known ligands to develop a common template, usually by aligning them all together in their energy minimised state. The correlation between molecular descriptors on the template and the biological activities of the ligands is then quantified (Acharya et al., 2011). This substrate template can be used in virtual screening to discover novel drug candidates. Potential compounds are scored based on how well they fit this template, and top-ranking hits are subsequently synthesised for lab testing. Apart from applying these techniques to drug targets like receptors or enzymes, they have also been used to study drug transporters.

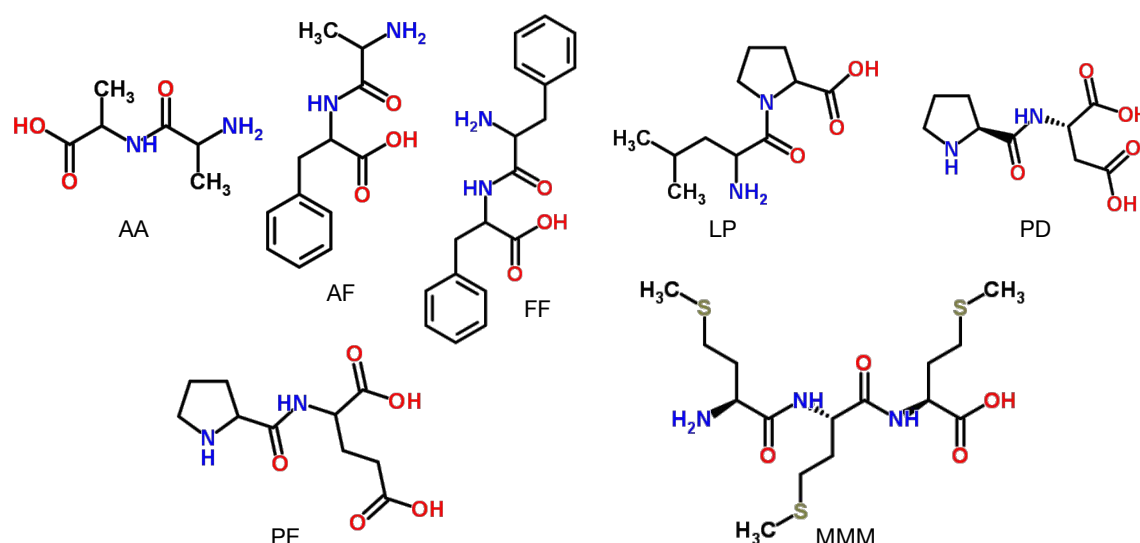
As discussed in Chapter 1, PepT1 is an important target for designing drugs that are orally active. Bailey et al. (2000) compiled the results of various experimental studies consisting of 100 compounds and proposed the first simple substrate template for PepT1 that they claimed could determine whether a substrate will be transported with high, medium or low affinity. In contrast, Ekins et al. (2005) employed just three well known PepT1 substrates, glysar, bestatin and enalapril\* to generate a common-feature pharmacophore map using the HIPHOP program (Clement and Mehl, 2000). Despite the small sample, their model successfully identified two novel inhibitors for the transporter: the anti-diabetic repaglinide and the HMG-CoA reductase inhibitor fluvastatin, which were later experimentally verified.

Using 3D-QSAR methodologies such as comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA), Gebauer and Hartrodt (2003) used a training set of 79 dipeptide-like substrates to identify six recognition elements favouring binding to PepT1. Cross-validation with a test set of 19 compounds demonstrated a correlation coefficient of  $\sim 0.7$  with experimentally measured transport data. The model was later improved by adding tripeptides and  $\beta$ -lactam antibiotics into the training set and a higher predictive power was obtained ( $r^2 = 0.8$ ) (Biegel et al., 2005). A more sophisticated pharmacophore map that can distinguish between substrate binding and transport was built by Vig et al. (2006) by applying a two-pronged approach—competition assays as an input for binding affinities and membrane depolarisation assays as an input for efficiency of transport. The study identified key structural features for a compound to be transported by PepT1.

While these methodologies are good at clarifying the molecular descriptors that substrates tend to share, they provide very little information on the binding site of the protein

---

\*A more recent study by Brandsch (2009) showed that despite showing moderate competition with Glysar uptake in Caco-2 cells, enalapril only induced 5% of the maximal currents elicited by a typical substrate in a two-electrode voltage clamp experiment using *X. laevis*, suggesting that it is either a very poor substrate or a competitive inhibitor.



Figs. 6.1: The structures of peptides used as a test set to assess binding affinity prediction methods with human PepT1 homology model.

itself and therefore offer little or no structural justifications for the results. Recently, the determination of X-ray crystal structures of pharmaceutically important membrane proteins has shifted the paradigm from these indirect ligand based approaches towards direct structure based approaches. One example relevant for this thesis is, of course, computing the free energy of binding (reviewed in detail in Chapter 2). With the ever increasing computational power, the implementation of free energy calculations in structure based drug design will continue to expand to include more proteins and larger compound libraries (Shirts et al., 2010).

The structure of human PepT1 itself is not yet available. Nevertheless, the growing number of crystal structures of bacterial POTs provide great model systems to understand the molecular basis of ligand binding in the POT family. In Chapter 4, the interaction of numerous peptides with PepT<sub>St</sub> was characterised and the specificity profile of this transporter was determined using a range of free energy methods. More importantly, this approach provided interesting insights, which could not have been gained by ligand based studies; for example, that the low affinity of basic dipeptides may stem from a less optimal binding orientation. Having built a reasonable homology model of PepT1 in the previous chapter, the same technique shall now be extended to human PepT1.

In the first instance, a scoring function (via AutoDock) and three types of end-point free energy methods (linear interaction energy (LIE), molecular mechanics Generalised Born surface area (MMGBSA) and molecular mechanics Poisson Boltzmann surface area

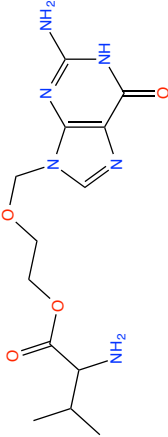
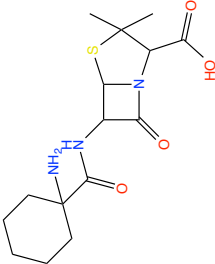
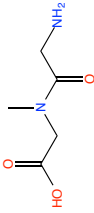
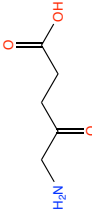
Table 6.1: Peptide transport by human PepT1

| Peptide   | IC <sub>50</sub> (mM) | Reference                   |
|-----------|-----------------------|-----------------------------|
| AlaPhe    | 0.07                  | Vig et al. (2006)           |
| AlaAla    | 0.08                  | Vig et al. (2006)           |
| MetMetMet | 0.10                  | Biegel et al. (2005)        |
| PhePhe    | 0.11                  | Gebauer and Hartrodt (2003) |
| LeuPro    | 0.18                  | Biegel et al. (2005)        |
| ProAsp    | 9.80                  | Biegel et al. (2005)        |
| ProGlu    | 20.00                 | Biegel et al. (2005)        |

IC<sub>50</sub> values were measured by competition transport assays in Caco-2 cells monolayer using <sup>14</sup>C-Glysar as the reporter substrate.

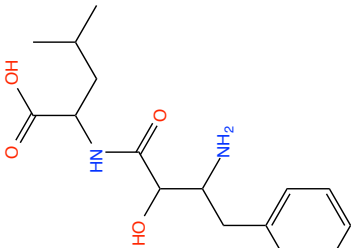
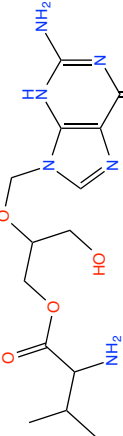
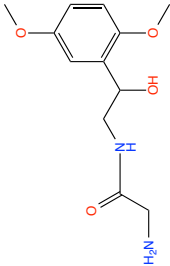
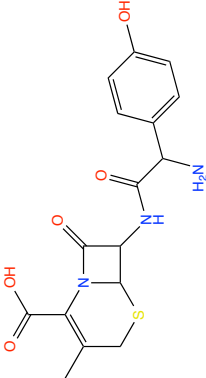
(MMPBSA)) will be used to estimate the free energy of binding,  $\Delta G$ , for a test set of ligands. The vast amount of transport data available in the literature for PepT1 allows us to include not only di- and tripeptides, but also some drug compounds, in the test set. Seven peptides featured in either Gebauer and Hartrodt (2003), Biegel et al. (2005) or Vig et al. (2006) were selected (Figure 6.1 and Table 6.1). Six  $\beta$ -lactam antibiotics studied in Bretschneider et al. (1999) are also chosen along with eight other drugs and prodrugs (Table 6.2). Using transport data from more than one source could introduce errors. However, there is currently no single study that has measured the uptake of a comprehensive list of peptides and drugs by human PepT1. Nonetheless, to maintain consistency between these different studies, we made sure that they all used the same assays and similar conditions, i.e. competition transport assays in Caco-2 cells monolayer using <sup>14</sup>C-Glysar as a reporter substrate. This test set was chosen to cover a large range of peptide/drug selectivity, i.e. IC<sub>50</sub> from 0.07 to 40 mM. Afterwards, the suitability of using these affinity prediction methods on our homology model will be discussed.

Table 6.2: Drug transport by human PepT1

| Drug                | Abbreviation | Structure                                                                            | IC <sub>50</sub> (mM) | Reference                    |
|---------------------|--------------|--------------------------------------------------------------------------------------|-----------------------|------------------------------|
| Valacyclovir        | VCV          |    | 0.5                   | Ganapathy et al. (1998)      |
| Cyclacillin         | CYC          |    | 0.5                   | Bretschnneider et al. (1999) |
| Glycylsarcosine     | GSC          |   | 1.1                   | Vig et al. (2006)            |
| Aminolevulinic acid | AMV          |  | 1.5                   | Irie et al. (2001)           |

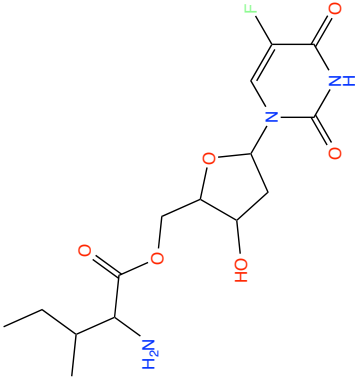
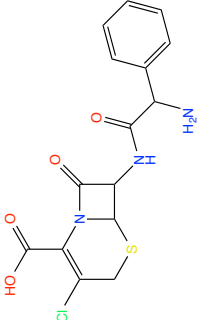
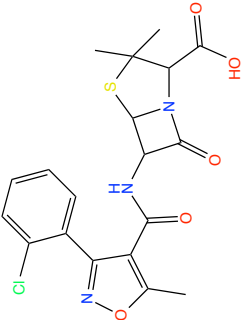
*Continued on next page*

Table 6.2 (Continued)

| Drug           | Abbreviation | Structure                                                                            | IC <sub>50</sub> (mM) | Reference                   |
|----------------|--------------|--------------------------------------------------------------------------------------|-----------------------|-----------------------------|
| Bestatin       | BES          |    | 1.5                   | Terada et al. (2000b)       |
| Valganciclovir | VGC          |    | 1.7                   | Sugawara et al. (2000)      |
| Midodrine      | MID          |   | 2.8                   | Tsuda et al. (2006)         |
| Cefadroxil     | CFD          |  | 7.2                   | Bretschneider et al. (1999) |

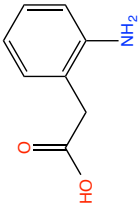
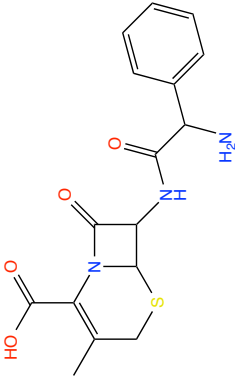
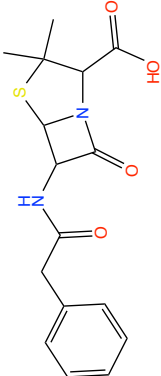
Continued on next page

Table 6.2 (Continued)

| Drug        | Abbreviation | Structure                                                                           | IC <sub>50</sub> (mM) | Reference                    |
|-------------|--------------|-------------------------------------------------------------------------------------|-----------------------|------------------------------|
| Floxuridine | FLO          |   | 7.3                   | Tsume et al. (2008)          |
| Cefaclor    | CFR          |   | 11                    | Bretschnneider et al. (1999) |
| Oxacillin   | OXA          |  | 12                    | Bretschnneider et al. (1999) |

Continued on next page

Table 6.2 (Continued)

| Drug                   | Abbreviation | Structure                                                                           | IC <sub>50</sub> (mM) | Reference                    |
|------------------------|--------------|-------------------------------------------------------------------------------------|-----------------------|------------------------------|
| Aminophenylacetic acid | APA          |  | 14                    | Borner et al. (1998)         |
| Cefalexin              | CFL          |   | 15                    | Bretschnneider et al. (1999) |
| Benzy1 Penicillin      | BPC          |   | 40                    | Bretschnneider et al. (1999) |



## 6.2 Methods

Briefly, most of the protocols for modelling, simulation and free energy calculation follow closely the ones explained in §3.2. Emphasis will be given on how drug compounds were dealt with as this has not been described in previous chapters.

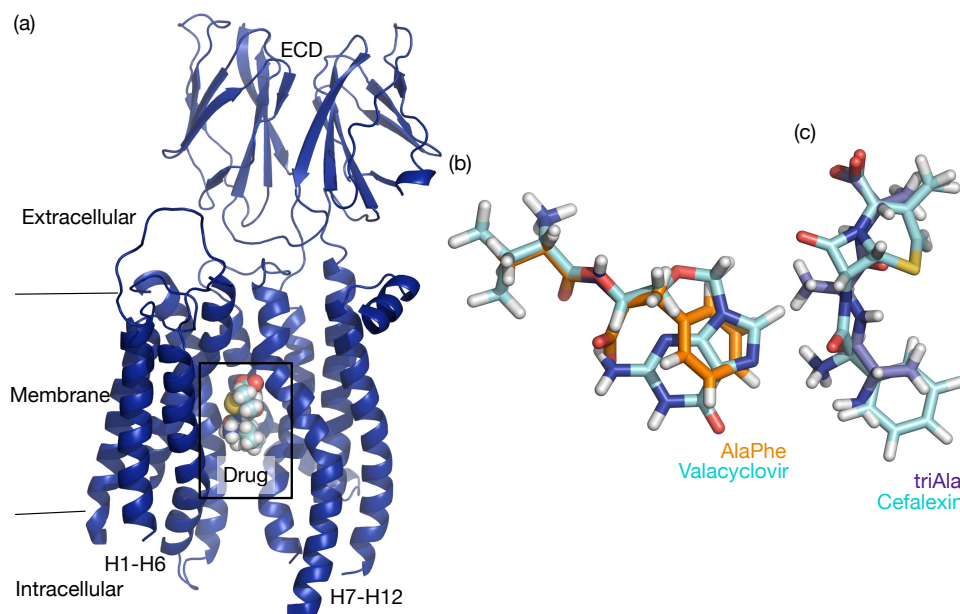
### 6.2.1 Modelling peptide and drug binding

As human PepT1 shares 80% sequence identity with PepT<sub>St</sub> around the binding cavity, it is assumed that ligand interactions would also be preserved. The homology model of PepT1 was aligned to the PepT<sub>St</sub>-AlaPhe (PDB ID: 4D2C) and PepT<sub>St</sub>-triAla (4D2D) complex structures (Lyons et al., 2014) in PyMOL to build models of PepT1-AlaPhe and PepT1-triAla, respectively. The coordinates of AlaPhe and triAla were then used as a template to model other peptides. Only the side chains of the peptide substrates were altered whilst the structure of the rest of the homology model remained the same.

To model the binding of a drug compound to the homology model of PepT1, the structure was determined by inspection if it resembles AlaPhe or triAla. Small drugs, i.e. AMV, APA and GSC, as well as prodrugs, i.e. VCV, VGC, MID and FLO are structurally analogous to AlaPhe. Each of these compounds was aligned to AlaPhe in the “horizontal” binding pose using a flexible ligand alignment program, LigAlign (Heifets and Lilien, 2010) (Figure 6.2(b)). On the other hand, the  $\beta$ -lactam antibiotics (CFD, CFL, CFR, CYC, OXA and BPC) and BES are more similar to triAla, and therefore were aligned to the vertical binding pose (Figure 6.2(c)). These alignments were copied to the coordinates of the homology model to create a series of PepT1-drug complexes.

### 6.2.2 MD Simulation

The PepT1 homology model and peptides were parametrised using the standard Amber99SB force field (Lindorff-Larsen et al., 2010). Drug molecules, on the other hand, were parametrised using the generalised Amber force field (GAFF) (Wang et al., 2004). These parameters were generated in the Antechamber program (Wang et al., 2006b) whereby the AM1-BCC charge model was utilised (Jakalian et al., 2002). Each ligand-bound transporter protein was inserted into a pre-equilibrated 381-molecule POPC bilayer (Dickson et al., 2012) in a cubic box of dimension 10 nm  $\times$  13 nm  $\times$  15 nm using the g\_membed protocol (Wolf et al., 2010). TIP3P water (Jorgensen et al., 1983) and 0.15 M NaCl ions were added to the simulation box. The system was energy-minimised by a steepest-descent protocol and further



Figs. 6.2: Modelling the binding of drug compounds to PepT1 homology model. (a) The binding site of human PepT1 with H2 and H11 removed for clarity. (b) Small drugs and prodrugs are modelled based on the binding of AlaPhe in the horizontal mode, while (c)  $\beta$ -lactams are modelled based on the binding of triAla in the vertical mode.

equilibrated in a 1 ns simulation, with the heavy atoms of the protein and ligand positionally restrained. Upon removing this restraint, a 1 ns production simulation was performed for each ligand as an input for the end-point methods. All simulations were performed in Gromacs 4.5.4 (Hess et al., 2008) with settings as outlined in §5.2.4.

### 6.2.3 Free energy calculation

As described in §3.2.2, binding free energy prediction was performed for each ligand in the test set using either AutoDock Vina (Trott and Olson, 2010), the `g_lie` tool in Gromacs 4.5.4 (Hess et al., 2008) or the MMPBSA.py program in Amber (Miller III et al., 2012).

## 6.3 Results and Discussion

The graphs of predicted  $\Delta G$  vs experimental  $IC_{50}$  are shown in Figure 6.3. The predicted  $\Delta G$  values for all ligands in the test set are listed in Table 6.4. The performance of each binding affinity prediction method is assessed using the Spearman's correlation coefficient,  $\rho$  (Lehmann and D'Abrera, 1998), and the Pearson correlation coefficient,  $r$  (Pearson, 1895) (explained in §3.2.4). This is summarised in Table 6.3.

Table 6.3: The performance of binding affinity prediction methods for human PepT1

| Correlation measure | AutoDock | LIE  | MMGBSA | MMPBSA |
|---------------------|----------|------|--------|--------|
| $\rho$              | -0.08    | 0.07 | 0.07   | 0.08   |
| $r$                 | -0.14    | 0.00 | 0.10   | 0.00   |

$\rho$  represents Spearman's rank correlation coefficient calculated using Equation 3.3, whilst  $r$  represents Pearson correlation coefficient calculated using Equation 3.4..

### 6.3.1 All methods fail to predict transport.

Similar to the results for PepT<sub>St</sub> in Chapter 3, AutoDock failed to accurately rank the ligands with both Spearman's and Pearson correlation coefficients close to zero, indicating no predictive power (Figure 6.3(a)). As seen previously, the predicted  $\Delta G$  values span a relatively small range from 0 to -10 kcal/mol and most ligands are predicted to have a binding affinity of between -2 to -8 kcal/mol. A poorly transported  $\beta$ -lactam antibiotic, BPC, is estimated to have a very similar  $\Delta G$  to a well transported dipeptide, AF (-5.9 kcal/mol and -5.7 kcal/mol, respectively). These results strengthen our argument in §3.3.1 that scoring functions, due to the use of an over simplified mathematical model and lack of consideration for protein-ligand binding dynamics, are insufficient to distinguish good and poor substrates of peptide transporters.

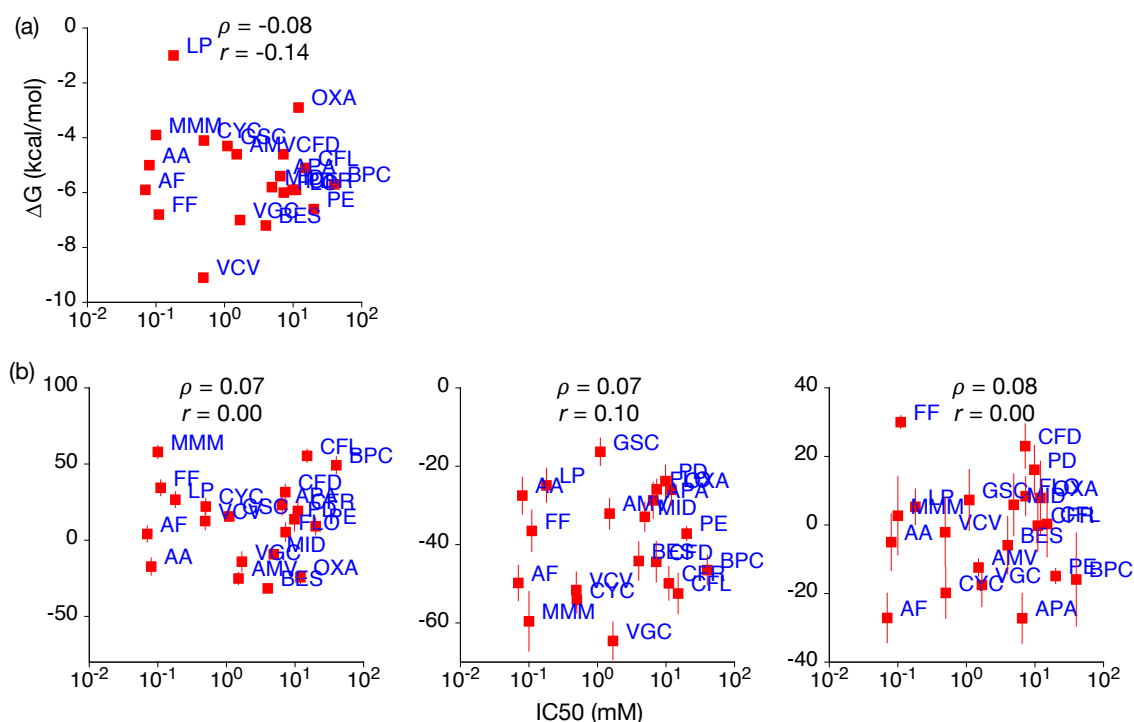
For binding affinity predictions using the end-point free energy methods, a 1 ns simulation was included in an attempt to account for conformational changes that happen when a compound binds to PepT1. Previously, in §3.3.2, it was demonstrated that for most of the ligand bound PepT<sub>St</sub> simulations,  $\Delta G$  data became de-correlated after 10 ps and reached equilibrium after 200 ps. It shall be assumed that these characteristics apply to all the ligand bound PepT1 simulations and so the same auto-correlation time,  $\tau$ , and equilibrium time,  $t_{eq}$ , will be used for all of the simulations in this section.

To our surprise, the performance of all three end-point methods was comparable to AutoDock. In contrast to the results for PepT<sub>St</sub>, LIE, MMGBSA and MMPBSA were un-

Table 6.4: Predicted  $\Delta G$  for peptide and drug binding to human PepT1

| Compound | $\Delta G$ (kcal/mol) |       |        |        |
|----------|-----------------------|-------|--------|--------|
|          | AutoDock              | LIE   | MMGBSA | MMPBSA |
| AF       | -5.9                  | 4.1   | -49.8  | -27.1  |
| AA       | -5.0                  | -17.3 | -27.5  | -5.0   |
| MMM      | -3.9                  | 57.9  | -59.6  | 2.7    |
| FF       | -6.8                  | 34.4  | -36.5  | 30.0   |
| LP       | -1.0                  | 26.6  | -24.9  | 5.3    |
| PD       | -5.9                  | 13.8  | -23.8  | 16.1   |
| PE       | -6.6                  | 9.2   | -37.2  | -14.9  |
| VCV      | -9.1                  | 12.6  | -51.6  | -2.1   |
| CYC      | -4.1                  | 22.1  | -54.1  | -19.8  |
| GSC      | -4.3                  | 15.6  | -16.3  | 7.3    |
| AMV      | -4.6                  | -25.1 | -32.1  | -12.4  |
| VGC      | -7.0                  | -14.1 | -64.6  | -17.5  |
| BES      | -7.2                  | -31.6 | -44.2  | -5.9   |
| MID      | -5.8                  | -9.2  | -32.9  | 5.9    |
| APA      | -5.4                  | 23.1  | -28.7  | -27.2  |
| CFD      | -4.6                  | 31.6  | -44.4  | 23     |
| FLO      | -6.0                  | 5.3   | -25.9  | 8.4    |
| CFR      | -5.9                  | 19.2  | -49.9  | -0.2   |
| OXA      | -2.9                  | -24.3 | -26    | 7.9    |
| CFL      | -5.1                  | 55.4  | -52.5  | 0.3    |
| BPC      | -5.7                  | 49.2  | 46.5   | -15.9  |

The end-point methods use empirical constants and implicit solvation, therefore the magnitude of  $\Delta G$ s calculated may seem dubiously large (or small). These should not be considered as the *true* binding energy.



Figs. 6.3: Binding affinities of peptides and drugs to human PepT1 as predicted by (a) AutoDock and (b) end-point free energy methods (from left: LIE, MMGBSA and MMPBSA) compared to experimental  $IC_{50}$  values. Error bars in (b) represents standard error from de-correlated  $\Delta G$  data for a 1 ns MD simulation.

able to reproduce the experimental ranking of ligands as indicated by the Spearman's and Pearson correlation coefficients all having values around zero, suggesting that the prediction is random. Inspecting the range of predicted  $\Delta G$  values, it was found that a similarly large span for each method as observed previously. For example, LIE predicts a distribution of  $\Delta G$  from -50 to 50 kcal/mol, i.e. a 100 kcal/mol range. In practice, this should allow us to distinguish between strong and weak binders. However, due to the inaccurate prediction for each ligand, the methods could not differentiate for example, the low affinity BPC antibiotics from high affinity peptides like AA, AF and FF. Overall, the performance of end-point free energy methods for the PepT1 model is very poor, especially considering the computational requirements are at least three orders of magnitude higher than that of AutoDock. The additional protein-ligand dynamics evidently failed to improve predictions by any appreciable degree as the results obtained are on par with using AutoDock and a static structure. As such, it seems sensible not to pursue with the theoretically rigorous method, thermodynamic integration (TI). Instead, possible reasons for this loss of predictive ability of the end-point methods shall be explored.

### 6.3.2 General Amber force field for drug parameters.

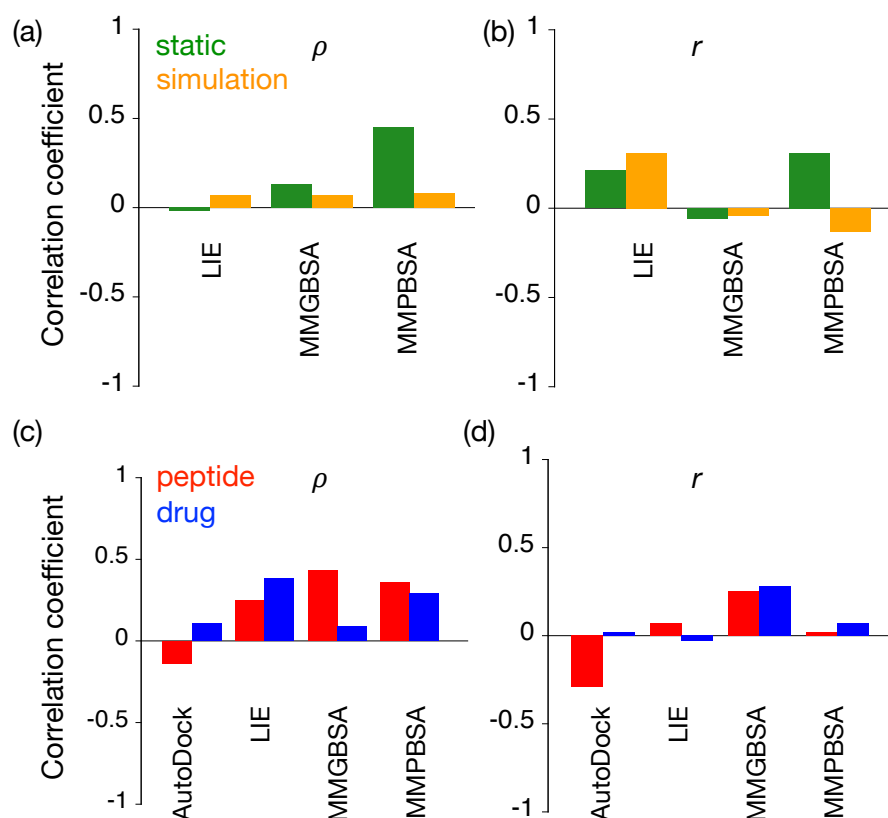
In this study, all drug compounds are parametrised using the general Amber force field (GAFF)<sup>†</sup> (Wang et al., 2004). This force field has a limited number of atom types and yet covers a vast range of chemistries from organic to pharmaceutical molecules. In practice, some of the atom types will be too generic to describe the chemical environment of a specific group, for example, the  $\beta$ -lactam ring. The peptide ligands, on the other hand, are parametrised using the well established Amber protein force field that was developed to specifically fit the chemical properties of amino acids, and therefore is more accurate (Lindorff-Larsen et al., 2010). We speculate that the poor predictive ability of the end-point free energy methods stems from the incorrect parametrisation of the drug molecules causing inaccurate sampling during MD simulations.

A previous study by Kuhn et al. (2005) found a slight degradation in predictive ability of MMPBSA when using ensembles of snapshots generated by MD simulations compared to using a single structure. The study employed GAFF and MAB force fields to parametrise 16 ligands of the p38 MAP kinase and suggested the ligand parameters were to blame. To test this, all three end-point methods was performed on the static structure of PepT1-ligand complex built based on homology modelling (Figure 6.4(a) and (b)). For LIE and MMG-BSA, the analysis indicates no significant different between using either the static structure or snapshots from MD simulations, with all  $\rho$  values remain close to zero. Interestingly a somewhat similar degradation in predictions for the MMPBSA method was observed, i.e. calculations using the static structures are better than using MD ensembles. However, whether this observation is significant is questionable since the MMPBSA performance for single structures is still considerably lower than what was previously observed for PepT<sub>St</sub>.

To investigate further, the ligands in the test set were separated into peptides and drugs and the values of the  $\rho$  and  $r$  correlation coefficients for these subsets were recalculated. We wanted to see if one subset performed better than the other, which would indicate an underlying force field issue. Unfortunately, as illustrated in Figure 6.4(b), similarly discouraging results were found for both the peptide and drug subsets. For the LIE method,  $\Delta G$  predictions for the drug test set are actually slightly better than for the peptide test set.

---

<sup>†</sup> Another force field typically employed to parametrise small molecules is CHARMM general force field or CGenFF (Vanommeslaeghe et al., 2010). Compared to GAFF, it boasts a larger library of atom types to cater for more specific chemical environments. However, at the time this study was performed, CGenFF did not have any parameters for the  $\beta$ -lactam ring, making it impossible to parametrise the  $\beta$ -lactam antibiotics in our test set, therefore we proceeded with GAFF. Alternatively, one can perform individual parametrisation for each ligand to get the best representation, but this would be very computationally demanding and time consuming.

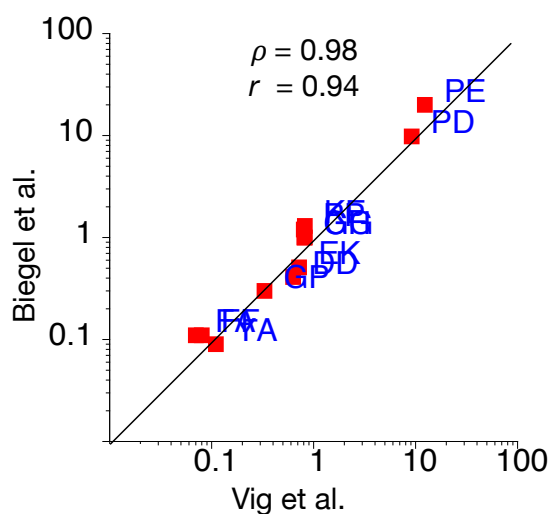


Figs. 6.4: Analysis of the reliability of GAFF in simulation for free energy calculations. (a) The performance of binding affinity prediction methods for a static structure (orange) and for an ensemble from a 1 ns MD simulation (green). The metric used for comparison is the Spearman's correlation coefficient,  $\rho$ . (b) A similar comparison for the Pearson's correlation coefficient,  $r$ . (c) The performance of the methods for peptides (red) and drugs (blue) compared based on their  $\rho$  values. (d) A similar plot for the  $r$  values.

Additionally, comparing the correlation coefficients for the peptide test set to the predictions involving  $\text{PepT}_{\text{St}}$  (Table 3.2 and Figure 3.9), it is clear that the results for the  $\text{PepT1}$  homology model are notably inferior, despite having a smaller data set (the peptide test set for  $\text{PepT}_{\text{St}}$  contains eight compounds while the  $\text{PepT1}$  test set has only seven). Moreover, Wright et al. (2013) demonstrated encouraging results when MMPBSA was performed using trajectories from MD simulations using GAFF for nine inhibitors of the HIV-1 protease. Collectively, we conclude that the poor performance of end-point methods in predicting the binding affinities of peptides and drugs to  $\text{PepT1}$  cannot be attributed to the force field used to parametrise the drug molecules.

### 6.3.3 Experimental data from whole cell assays.

As previously mentioned, the experimental data to which the  $\Delta G$  predictions are compared were collated from several different *in vivo* studies. Whilst all of these studies were performed in Caco-2 cell monolayers and used the same reporter substrate ( $^{14}\text{C}$ -Glysar), it is possible that different laboratories would measure different values for the same compound, leading to errors. In whole cell uptake assays, peptide and drug transport can vary due to numerous reasons, such as cell source, culture media, monolayer age, buffer composition and other experimental conditions (Volpe, 2011). Unlike the *in vitro* proteoliposome uptake assays, it is also more difficult to precisely quantify the expression level of the transporter protein in the cell-based assay, introducing a further variable to the measurement. It is likely, therefore, that using  $\text{IC}_{50}$  values from different studies together to rank the ligands might elicit dubious results, and ultimately reduce the correlation to the computed  $\Delta G$ s.



Figs. 6.5: Comparing the  $\text{IC}_{50}$  values for twelve dipeptides obtained from two different studies (Vig et al. (2006) and Biegel et al. (2005)) using the Caco-2 cell monolayer and  $^{14}\text{C}$ -Glysar. The unit on both axes is mM and the line drawn represents perfect linear correlation.



To approximate the degree of variability between these experimental studies, the results of Biegel et al. (2005) and Vig et al. (2006), which contain the most extensive data for di- and tripeptide uptake by human PepT1 in Caco-2 cells, were compared. Between the two studies, there are twelve identical dipeptides but no identical tripeptides. The  $IC_{50}$  values for these twelve dipeptides from both studies were plotted, and if this *in vivo* assay is reproducible, then the points should all fall on a straight line. Indeed, there is an almost perfect correlation between the two sets of transport data with  $\rho = 0.98$  and  $r = 0.94$  (Figure 6.5). It is worth noting that the exact  $IC_{50}$  values for each peptide from the two studies do differ, for example, for ProGlu, Vig et al. (2006) obtained 12.3 mM whereas Biegel et al. (2005) reported 20 mM. Despite this discrepancy, the rank-ordering of all the twelve peptides is very similar. This analysis strongly indicates that, whilst the inter-laboratory variations cause differences in individual measurement, the overall trend of peptide transport remains the same, suggesting that, at least for peptides, using uptake data from more than one study will not significantly affect comparison to the calculated  $\Delta G$  values.

Repeating the analysis for drug uptake, however, proved more difficult as often each study only consisted of one drug compound. Only the uptake of the  $\beta$ -lactam antibiotics has been repeated in several studies. Nevertheless, in most cases the experiments were performed using different protocols, making comparison between them rather problematic. For example, different cell lines were used: Terada et al. (1997) measured the inhibition of  $^{14}C$ -Glysar uptake by nine antibiotics in LLC-PK<sub>1</sub> cells stably transfected with rat PepT1; Han et al. (1999), on the other hand, used CHO cells over expressing human PepT1 and  $^3H$ -Glysar to study the transport of cefaclor and cephalexin. Despite this, the analysis of the peptide uptake data alone suggests it is unlikely that the usage of  $IC_{50}$  values from different *in vivo* studies to be the reason for the low correlations observed in the predictions with PepT1.

#### 6.3.4 Crystal structures essential for prediction of binding free energy.

Another possible reason for the poor performance of the prediction methods is the lack of a high resolution crystal structure, as was used in Chapter 3 and Chapter 4 for PepT<sub>St</sub>. We described in Chapter 5 how a homology model of human PepT1 was built based on the structure of PepT<sub>So</sub>, since the crystal structure is currently unavailable. While various tests like geometrical analyses and MD simulation suggested that our model is overall structurally reasonable, it is likely that the precise orientations of residues in the binding site, especially the amino acid side chains interacting with the ligand molecule, are inaccurate. Additionally,

Table 6.5: Templates for PepT<sub>St</sub> homology modelling

| Name               | Sequence identity (%) | C $\alpha$ RMSD ( $\text{\AA}$ ) | PDB ID        |
|--------------------|-----------------------|----------------------------------|---------------|
| PepT <sub>St</sub> | 100                   | 0.0                              | 4D2C and 4D2D |
| GkPOT              | 52.9                  | 1.2                              | 4IKV          |
| LacY               | 25.2                  | 4.8                              | 1PV6          |

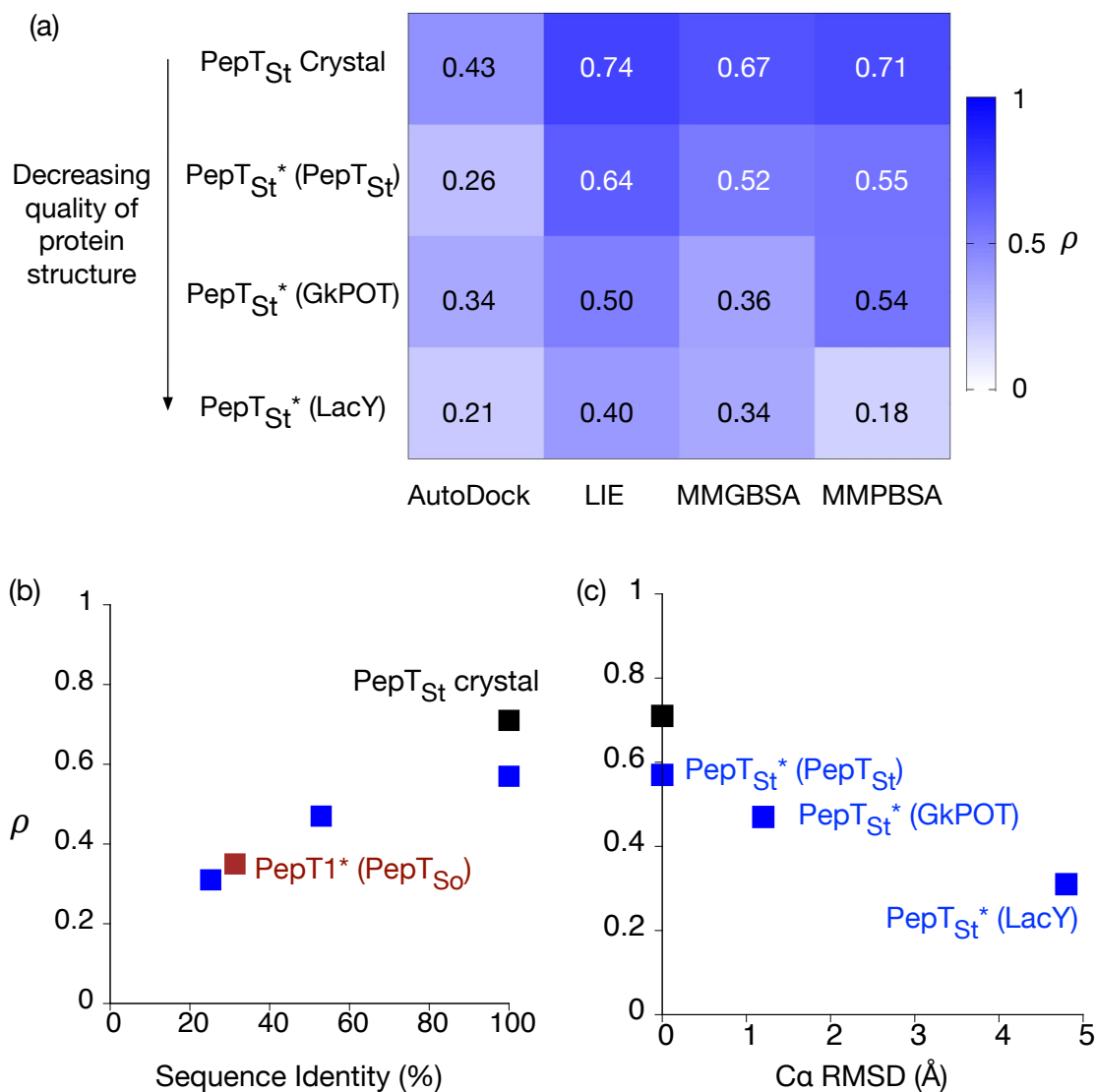
the current study is based on the assumption that di- and tripeptides occupy the binding site of PepT1 in a similar way to PepT<sub>St</sub>, as elucidated by Lyons et al. (2014), and that drugs bind to the transporter either like a dipeptide or a tripeptide depending on their size and chemistry. Whilst these are sensible assumptions given that both PepT1 and PepT<sub>St</sub> share most key residues in the binding site, it is plausible that subtle structural deficiencies in the homology model could result in erroneous binding affinity predictions. This raises several important questions: How sensitive are *in silico* affinity predictions to the structural integrity of the protein model? Can MD simulations correct any structural errors in a homology model and ameliorate predictions? Or will MD simulations of an incorrect homology model induce spurious sampling that ultimately aggravates the problem?

In order to address these questions, we have to go back to the bacterial peptide transporter, PepT<sub>St</sub> for a few reasons: i) there is a crystal structure, to which the quality of a homology model can be compared, ii) the binding orientations of peptides are known, thus eliminating another variable, and iii) the high-quality transport assay data from proteoliposome-based experiments. Three homology models of PepT<sub>St</sub> were built based on different templates of varying sequence identity and RMSD (Table 6.5). The first model was built using the PepT<sub>St</sub> crystal structure itself as a template. This should be the best homology model possible for the protein. The second model uses the crystal structure of GkPOT, which shares a high sequence identity (>50%) with PepT<sub>St</sub>, as a template. The crystal structure of GkPOT is also very similar to that of PepT<sub>St</sub> (C $\alpha$  RMSD = 1.2 $\text{\AA}$ ), representing an intermediate-quality model. The third homology model was built using LacY as a template. Despite having a similar MFS canonical fold, LacY shares a low sequence identity ( $\sim$ 25%) and structural similarity (C $\alpha$  RMSD = 4.8 $\text{\AA}$ ) with PepT<sub>St</sub>. As such, it represents a low-quality model. All chosen templates are in the inward facing conformation. For consistency, the same eight peptides as in Chapter 3 were used as the test set (Table 3.1) and their binding to the homology models is modelled based on the protocol described in §6.2.1. For each protein-peptide complex, the binding affinity was predicted using AutoDock, LIE, MMGBSA and MMPBSA as described in §3.2.2. For each calculation with the end-point methods, a simulation of length 1 ns was performed following the parameters outlined in

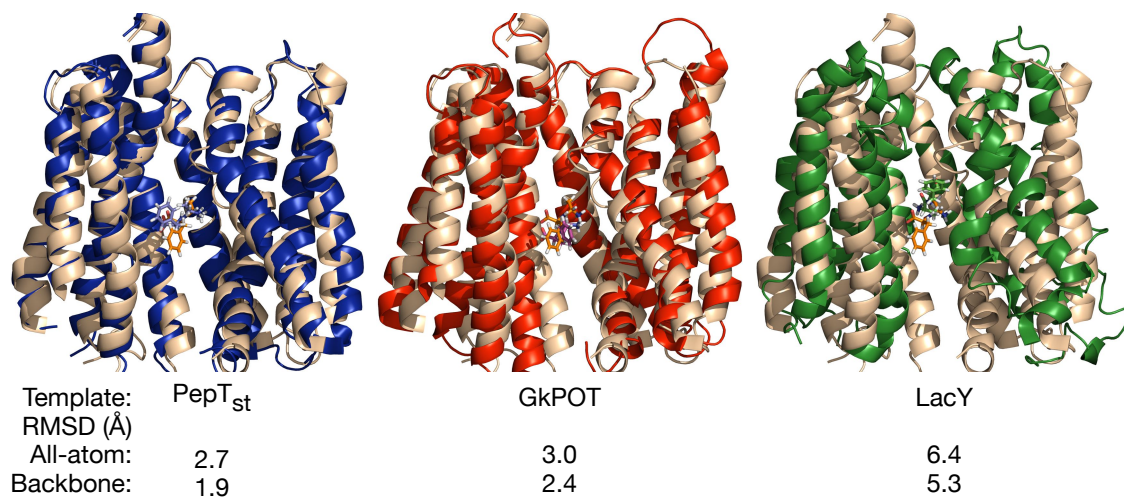
§3.2.3. As before, predicted  $\Delta G$  values were compared to experimental  $IC_{50}$  values from Solcan et al. (2012) and the Spearman's correlation coefficient was calculated as the measure of accuracy of the predictions.

It seems clear from Figure 6.6(a) that the predictive ability of the end-point methods is proportional to the quality of the protein model used for the calculations. Compared to the  $PepT_{St}$  crystal structure, all the homology models show lower  $\rho$  values for all the three methods. It is interesting that even when the homology model is built using the  $PepT_{St}$  crystal structure as a template, there is a decrease in the correlation coefficient values. This reduction of  $\rho$  is larger for the model based on GkPOT and larger still when LacY is employed as the structural template. When the average  $\rho$  values were plotted for the three end-point methods vs. the sequence identity of the template with  $PepT_{St}$  (Figure 6.6(b)), it becomes apparent that the higher the sequence identity, the better the  $\Delta G$  predictions. This trend is in agreement with the predictions made by the  $PepT_1$  homology model:  $PepT_1$  and its template ( $PepT_{So}$ ) have a percentage sequence identity of  $\sim 30\%$ , which is slightly higher than the  $PepT_{St}$ -LacY model but lower than the  $PepT_{St}$ -GkPOT model, and congruently the  $\rho$  values for the peptide test set falls between these two models. The performance of the end-point methods, however, reaches a plateau of  $\rho \sim 0.6$  for the best homology model, i.e. a percentage sequence identity of 100% when  $PepT_{St}$  is used as the template. A reciprocal trend is observed when the RMSD between the crystal structure of the template and that of  $PepT_{St}$  was considered (Figure 6.6(c)); the more divergent the two structures the worse the predictive ability. The results obtained using the AutoDock Vina scoring function, on the other hand, demonstrated a consistently low  $\rho$  values across all homology models, as was previously seen for  $PepT_{St}$  crystal structure (§3.3.1), and is not affected by the use of different templates during homology modelling.

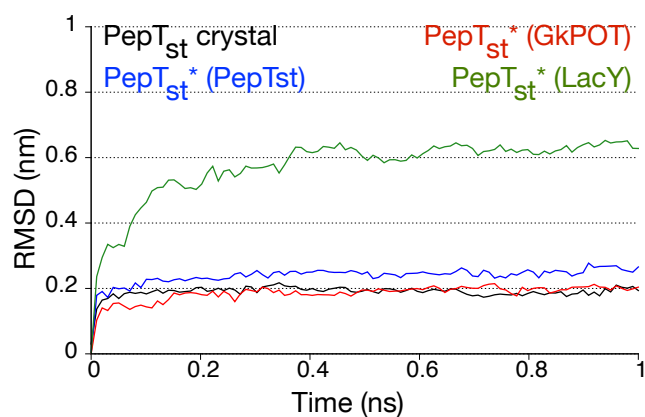
To understand the structural origin of the errors in end-point free energy predictions using these homology models, the conformational space explored by the protein and ligand during the 1 ns MD simulations were examined. Comparing the overall structures after a 1 ns simulation (Figure 6.7), it is apparent that the homology models sample different conformations from the crystal structure. For the best homology model, whereby  $PepT_{St}$  itself is used as the template, there is a minor difference in structure: the observed all-atom RMSD is 2.7 Å, whilst backbone RMSD is 1.9 Å, suggesting that while the protein backbone remain largely similar, the amino acid side chains could be the primary contributor towards the structural difference. A model based on GkPOT reports a larger RMSD as the template has a slightly different fold, and a model based on LacY displays the most structural difference with  $RMSD \sim 6$  Å, as expected due to the largely dissimilar structure.



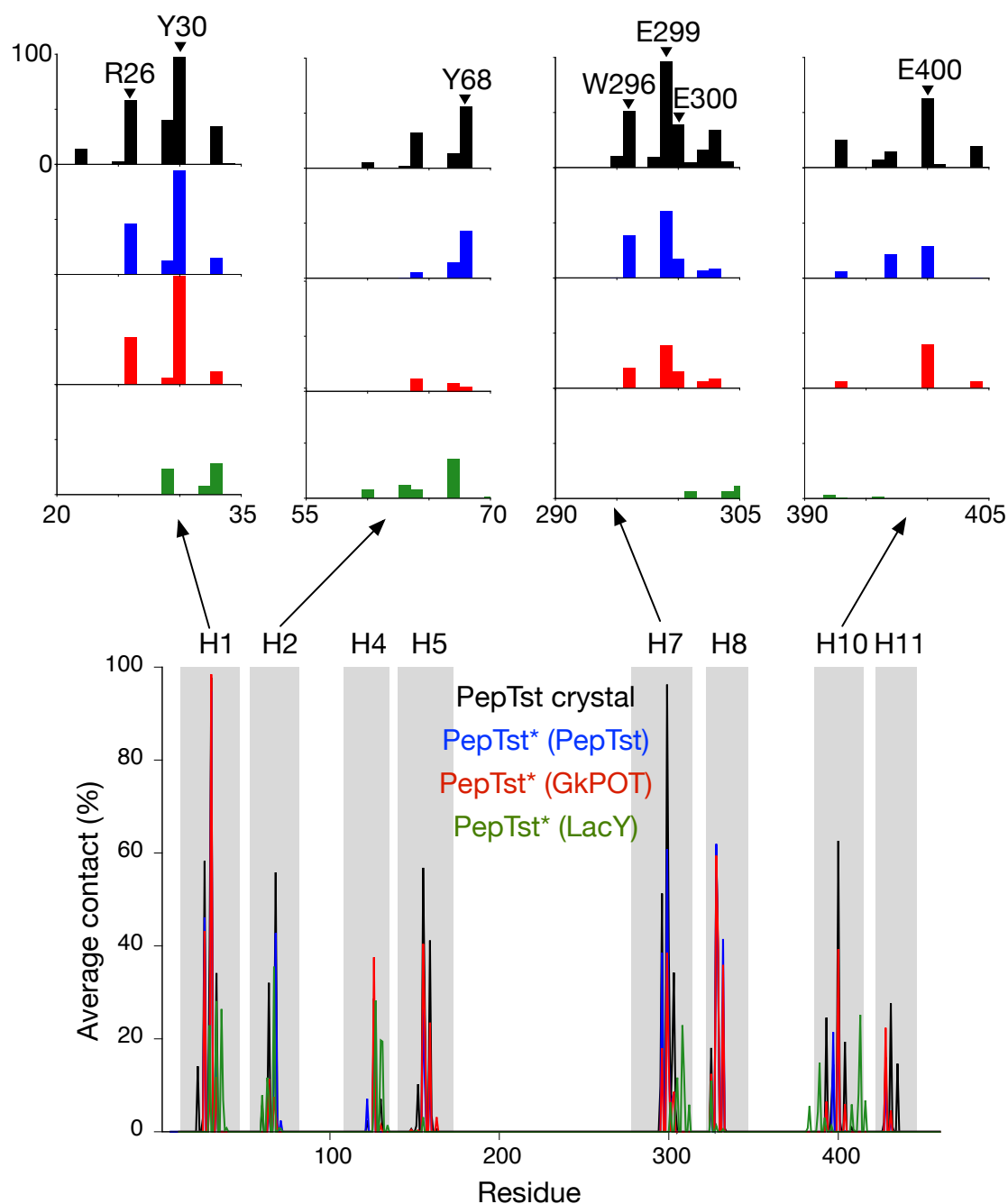
Figs. 6.6: Comparison of the performance of binding energy predictions using a scoring function (AutoDock) and end-point methods (LIE, MMPBSA and MMGBSA) for homology models of PepT<sub>St</sub>. (a) The number in each box represents Spearman's correlation coefficient,  $\rho$ , and is coloured from white (0.0) to blue (1.0). The predictive ability of the end-point methods is dependent on (b) the sequence identity between the template and PepT<sub>St</sub> and (c) the C $\alpha$  RMSD of the template and the crystal structure of PepT<sub>St</sub> (see Table 6.5). The plotted value of  $\rho$  is averaged over the three end-point methods, and for PepT1, only the  $\rho$  values from peptide test set are considered. Homology models are indicated by an asterisk with its template structure is given in parentheses.



Figs. 6.7: The structure of the PepT<sub>st</sub> homology models (blue, red and green) bound to AlaPhe after a 1 ns simulation aligned to the complex crystal structure from Lyons et al. (2014) (wheat) after a similar simulation protocol. The peptide ligand, AlaPhe, is shown in stick representation, while H2, H11, HA and HB are removed for clarity.



Figs. 6.8: The progression of average RMSD of eight peptide ligands during the 1 ns simulation performed for the calculation of  $\Delta G$  by end-point free energy methods. The coordinate of ligands at time = 0 ns was used as the reference structure.



Figs. 6.9: Contacts made by peptide ligands with residues in the binding site of PepT<sub>St</sub> during the 1 ns simulations performed using the crystal structure and three homology models, which was used for the calculation of  $\Delta G$  by end-point methods. Residues with significant interactions with the peptide ligands are highlighted. The cut-off for contact analysis was set to 3.5Å and the percentage contacts were averaged over eight peptides in the test set.

Based on this analysis, it was also found that the position of the peptide ligand, AlaPhe, shifted during the simulation such that it occupied a different binding orientation compared to the simulation using the crystal structure as the starting coordinates. To inspect the stability of the peptide within the homology model, the average RMSD for all eight peptides throughout the 1 ns trajectory was plotted (Figure 6.8). For the PepT<sub>St</sub> homology models based on the PepT<sub>St</sub> and GkPOT crystal structures, the peptide conformation is reasonably stable with RMSD values of around 2 Å, comparable to the simulations of the PepT<sub>St</sub> crystal structure. With the LacY based homology model, however, there is a major shift away from the starting configuration (RMSD~6 Å), suggesting weak binding. This is not unexpected as the LacY crystal structure is more open on the cytoplasmic side compared to either PepT<sub>St</sub> or GkPOT, and therefore it is possibly more difficult for peptide substrate to form strong interactions with residues in the binding site.

Protein-ligand interactions in the binding site form the basis of the end-point free energy calculations to determine  $\Delta G$ . As such, a thorough contact analysis was performed using the VMD plug-in interquant (Humphrey and Dalke, 1996) for all eight peptides and subsequently the average values for the three homology models and the crystal structure were plotted (Figure 6.9). The analysis for the simulations using the crystal complex identified key residues as suggested by functional studies in Solcan et al. (2012) and Lyons et al. (2014). These include Tyr30 in H1, Tyr68 in H2, Trp296 in H7 and Glu400 in H10. An interesting trend emerges when the results for these residues for the homology models were inspected. The PepT<sub>St</sub> based homology model showed a small but noticeable decrease in contacts compared to the crystal structure, while the GkPOT based homology model showed an even more conspicuous reduction. For example, Tyr68 in H2 made around 60% contacts during the simulation involving the crystal structure, but only 40% for PepT<sub>St</sub> based model and less than 10% for GkPOT based model. The same trend also prevailed for residues in H7 and H10. It seems that for the best and intermediate quality homology models, the similarity of protein-ligand contacts with the crystal structure is proportional to the structural integrity of the homology models. A similar analysis for the lowest quality homology model based on LacY, on the other hand, returned no significant contacts for any residues in the binding site. The poor coordinations of ligand with protein residues in this model explains the unstable binding as depicted by the RMSD analysis discussed earlier. Taken together, these results suggest that the lack of a good protein structure is the most likely reason for the poor correlation between the predicted  $\Delta G$  and experimental IC<sub>50</sub> values for peptide and drug binding to human PepT1.

The use of homology models in free energy calculation is not without precedent. Gen-

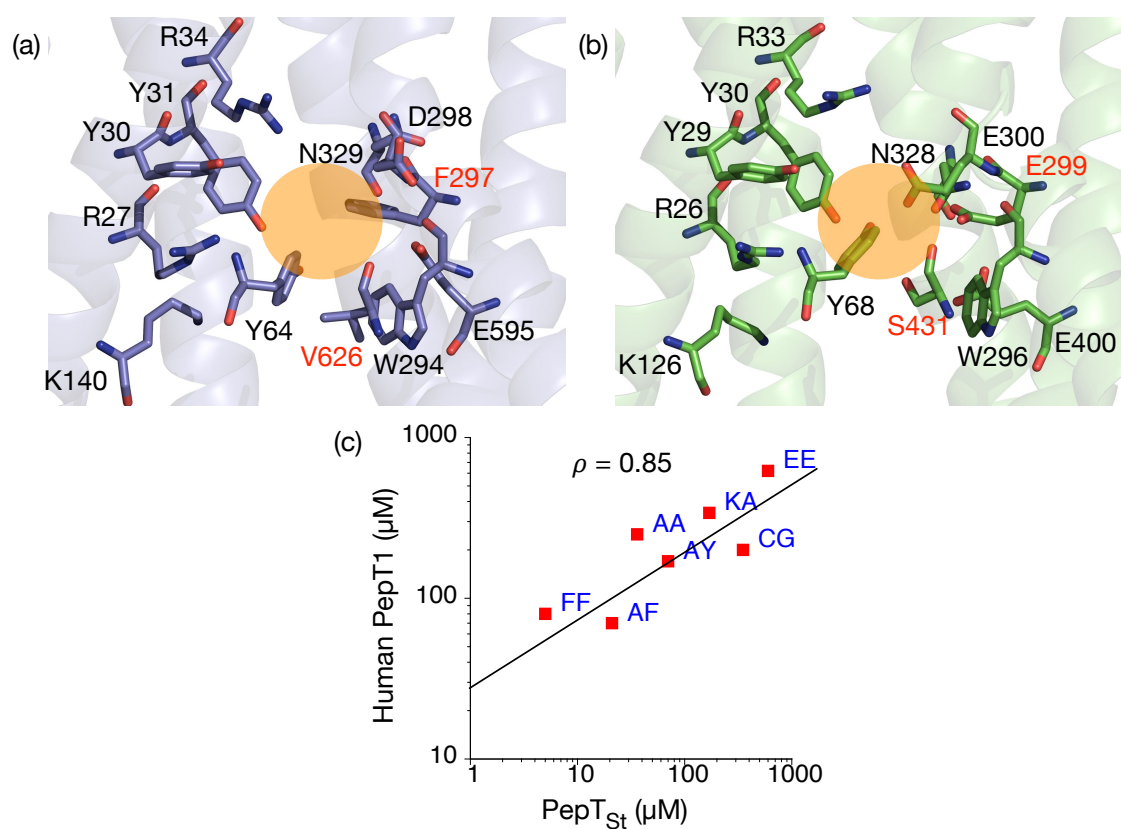
heden (2012) generated several homology models for the factor Xa protein. They calculated relative binding affinities of different ligands to both proteins using TI and subsequently compared the results to similar calculations using published crystal structures. A large difference (8 kJ/mol) is obtained when a three-residue loop region is incorrectly modelled, demonstrating that a small and localised structural error can indeed affect free energy calculations. However, some studies utilising homology models have proven successful. For example, Osterberg and Aqvist (2005) used a homology model of the human voltage gated hERG potassium channel based on the structure of a bacterial KvAP channel to predict the affinity of six ligands using AutoDock scoring function and the LIE method. While AutoDock failed to discriminate between good and weak binders of the ion channel, LIE produced a good correlation to experimental data. Additionally, Huang et al. (2012) built a homology model of the M5 muscarinic acetylcholine receptor and employed the MMPBSA method to determine the binding energy difference between two pharmaceutically-relevant ligands. Encouragingly, the results obtained is within 0.3 kcal/mol to experimentally derived value. Therefore, the use of a homology model in the calculation of binding free energy is probably system-dependent. It is tempting to generate such models in the absence of a crystal structure, especially for medically relevant membrane proteins which are often difficult to crystallise, but they should be used with a degree of caution. Experimental validation, for example by virtue of binding assays, is the only way to evaluate whether the model is fit to study protein-ligand interactions.

## 6.4 Insights into drug design from PepT<sub>St</sub>

Following our discovery that the accuracy of the free energy methods is critically dependent on the quality of the structure, we decided to take a step backward and investigate the bacterial POT transporter, PepT<sub>St</sub>. The POT transporter family traverses from bacteria to man with essential residues in peptide transport conserved across different species (Daniel and Spanier, 2006). In Chapter 4, using a range of affinity prediction methods, several key characteristics of binding selectivity were elucidated and how dipeptides and some tripeptides interact with the transporter was suggested. To understand how much of these characteristics are likely to apply to PepT<sub>1</sub>, the binding sites of the two homologs and their transport preference will be first compared.

As illustrated in Figure 6.10(a) and (b), the molecular architecture of the binding site is almost identical, as most of the residues that have been shown to be vital for transport via various functional studies are preserved in both transporters. Nevertheless, two differences





Figs. 6.10: The binding site of (a) the homology model of PepT1 and (b) PepT<sub>St</sub>, highlighting residues that are important in peptide binding. Different residues are labelled in red and the approximate position of substrate binding site is marked by the orange circle. (c) The IC<sub>50</sub> of eight dipeptides from transport assays of human PepT1 and PepT<sub>St</sub>.

are found: i) Phe297/Glu299 on H7 and ii) Val626/Ser431 on H11.

Positioned between two highly conserved negatively charged residues (Glu300 and Glu400) on the C-terminal bundle of PepT<sub>St</sub> is another acidic residue, Glu299. It has been proposed that the cluster of three glutamates may form a proton relay network (Solcan, 2013); protonated Glu300 transfers its proton to Glu299 and forms the extracellular gate with Arg33 to induce the outward facing to inward facing conformational change, followed by Glu400 accepting this proton from Glu299 to break the intracellular gate and release the proton to the cytoplasmic space, completing the transition. This hypothesis, however, remains speculative as mutation of the Glu299 residue to alanine, glutamine or even aspartate, did not express and thus no transport data is available. Also, crystallographic data shows that the backbone carbonyl of Glu299 interacts with the C-terminus of triAla (Lyons et al., 2014). Interestingly, at the equivalent position in PepT1, a hydrophobic residue, Phe297 is found instead. Further inspection of the multiple sequence alignment of mammalian PepT1 and PepT2 indicates that this residue is totally conserved in PepT1 but is replaced by another hydrophobic residue, leucine, in PepT2. It is worth noting that in other bacterial homologs, for example PepT<sub>So</sub>, a hydrophobic residue is also found at this position, suggesting that the role Glu299 plays in PepT<sub>St</sub> is likely to be specific to this transporter alone. Mutation of Phe297 to cysteine in human PepT1 abolished Glysar uptake activity (Kulkarni et al., 2003b), further corroborating the importance of this residue in peptide transport.

Another residue not conserved between the binding sites of human PepT1 and PepT<sub>St</sub> is Val626 and Ser431, respectively. In the crystal structure of PepT<sub>St</sub>-AlaPhe complex, the polar residue, Ser431, is found within the hydrophobic pocket that accommodates the phenylalanine side chain of the peptide substrate. Mutation of this residue to alanine severely reduced proton-driven AlaAla uptake, indicating that it is central to peptide binding (Lyons et al., 2014). In mammalian PepT1 and PepT2, a hydrophobic residue, either valine or isoleucine, is conserved at this position. However, no mutational studies has been performed on this particular residue for human PepT1. It is likely that the presence of a more hydrophobic side chain in this region would augment the hydrophobicity of the pocket, making PepT1 prefer hydrophobic ligands more than PepT<sub>St</sub>.

Finally, the transport data for human PepT1 and PepT<sub>St</sub> are compared. The IC<sub>50</sub> values for PepT1 are derived from whole cell competition uptake assays performed by Biegel et al. (2005) or Vig et al. (2006), whereas for PepT<sub>St</sub>, data from proteoliposome transport assays are used, all of which are conducted by either Solcan et al. (2012) or Joanne Parker (see Chapter 4). To date, data for tripeptide transport remains scarce for both transporter homologs, therefore comparison is only made for dipeptides. As these experimental assays

are fundamentally different in their setups, i.e. *in vivo* vs. *in vitro*, a direct quantitative comparison is not possible. Instead, we are more interested in the substrate rankings for both transporters to understand if there exist any discrepancies between their peptide selectivities. Again, the Spearman's rank correlation coefficient (Equation 3.5) (Lehmann and D'Abrera, 1998) is used as a statistical metric. There is a moderately high correlation between the two data sets ( $\rho = 0.85$ ), suggesting at least for dipeptides, the substrate preference of human PepT1 and PepT<sub>St</sub> are very similar (Figure 6.10(c)). This is not entirely surprising given the high amino acid sequence similarity in the binding site, as discussed earlier. The two differences found (Phe297/Glu299 and Val626/Ser431) evidently do not cause a major divergence in substrate selectivity. Therefore, it is reasonable to assume that the key characteristics for peptide binding in PepT<sub>St</sub> will transfer to human PepT1.

Next, the findings for the bacterial homolog (Chapter 4) will be interpreted in terms of their implications towards drug design targeting PepT1. There are two types of drug compounds that are known substrates of PepT1: i) peptide prodrugs like the antiviral valacyclovir and ii) peptidomimetics like the  $\beta$ -lactam antibiotics. We will propose how these drugs interact with the transporter and suggest essential features for molecular recognition.

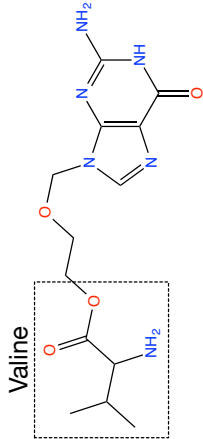
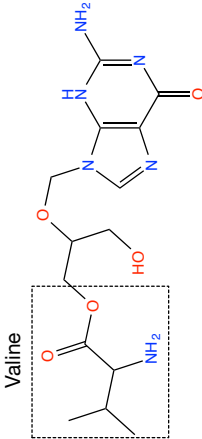
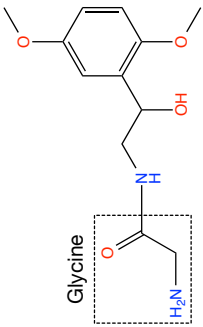
#### 6.4.1 Prodrug binding via a pseudo-N-terminus

Prodrugs are inactive and chemically altered versions of a pharmacological agent that undergo an enzymatic reaction *in vivo* to become an active drug compound (Rautio et al., 2008). Prodrugs are developed to improve various properties of an initial lead compound, for example enhancing their chemical solubility in an aqueous environment (physicochemical) or increasing their lipophilicity for better permeability into the cells (pharmacokinetic). A drug compound is covalently attached to a "promoiety" group through a bio-reversible link that will be enzymatically cleaved upon entrance into a cell. The choice of link depends upon the chemical groups on the drug compound itself that are amenable to chemical reaction; for example a carboxyl group on a drug can be attached to a promoiety via an ester, amide and carbonate link. Ester prodrugs are the most common owing to their straightforward synthesis and the ubiquitous *in vivo* esterases that assure high recovery of the active compound (Liederer and Borchardt, 2005).

Designing prodrugs to improve oral bioavailability has been performed for decades, although earlier efforts depended on a trial and error approach. Often, a polar group on a drug compound is attached to a more hydrophobic moiety to mask polarity and therefore improve membrane permeability. In recent years, however, it has been realised that carrier-

mediated cellular uptake plays an instrumental role in drug absorption (Dobson and Kell, 2008). One example is, of course, through the peptide transporter PepT1 in the endothelial cells of the small intestine. Valacyclovir, a prodrug of an antiviral acyclovir, was created by attaching a valine amino acid residue to a carboxyl group of the drug compound via an esterification reaction. This increases the bioavailability three- to five-fold (Wang et al., 1996; Beutner et al., 1995), although during its conception in 1995, little was known about the mechanism that led to the improved absorption. A few years later, a study by Ganapathy et al. (1998) provided the first evidence that this prodrug is a substrate of PepT1 and PepT2, whereas its parent drug, acyclovir, is not transported. Their results explained the enhanced bioavailability and triggered developments of numerous other prodrugs to target PepT1.

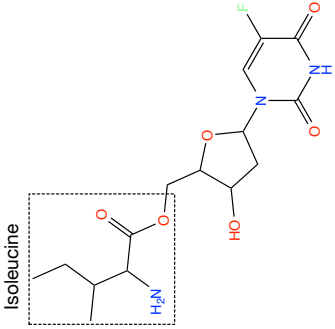
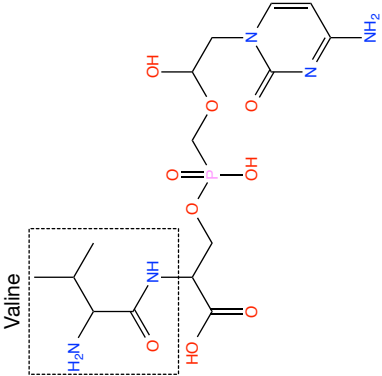
Table 6.6: Prodrugs designed to target human PepT<sub>1</sub>

| Name           | Structure                                                                            | Bioavailability <sup>a</sup> (%) |         | IC <sub>50</sub> (mM) | Reference               |
|----------------|--------------------------------------------------------------------------------------|----------------------------------|---------|-----------------------|-------------------------|
|                |                                                                                      | Drug                             | Prodrug |                       |                         |
| Valacyclovir   |   | 10                               | 55      | 0.45                  | Ganapathy et al. (1998) |
| Valganciclovir |   | 5                                | 60      | 1.68                  | Sugawara et al. (2000)  |
| Midodrine      |  | 50                               | 93      | 2.8                   | Tsuda et al. (2006)     |

<sup>a</sup>Retrieved from <http://www.medscape.com>, IV; Administered intravenously, NA; Not available (Prodrugs in pre-clinical stage)

*Continued on next page*

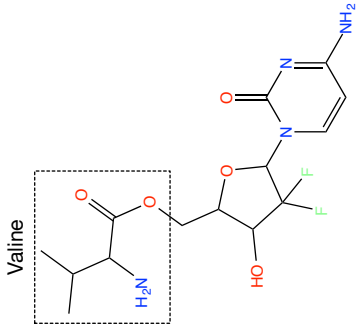
Table 6.6 (Continued)

| Name        | Structure                                                                            | Bioavailability <sup>a</sup> (%) |         | IC <sub>50</sub> (mM) | Reference             |
|-------------|--------------------------------------------------------------------------------------|----------------------------------|---------|-----------------------|-----------------------|
|             |                                                                                      | Drug                             | Prodrug |                       |                       |
| Floxuridine |   | IV                               | NA      | 0.7                   | Tsume et al. (2008)   |
| Cidofovir   |  | IV                               | NA      | 2.7                   | McKenna et al. (2005) |

<sup>a</sup>Retrieved from <http://www.medscape.com>, IV; Administered intravenously, NA; Not available (Prodrugs in pre-clinical stage)

Continued on next page

Table 6.6 (Continued)

| Name        | Structure                                                                           | Bioavailability <sup>a</sup> (%) |         | IC <sub>50</sub> (mM) | Reference          |
|-------------|-------------------------------------------------------------------------------------|----------------------------------|---------|-----------------------|--------------------|
|             |                                                                                     | Drug                             | Prodrug |                       |                    |
| Gemcitabine |  | IV                               | NA      | 0.14                  | Song et al. (2005) |

<sup>a</sup>Retrieved from <http://www.medscape.com>, IV; Administered intravenously, NA; Not available (Prodrugs in pre-clinical stage)

Table 6.6 lists six such prodrugs, three of which have been approved by the United States Food and Drug Administration (FDA), while four others are currently in pre-clinical studies. Valacyclovir, valganciclovir and cidofovir are used to treat various kinds of viral infections, midodrine is a vasopressor agent, while floxuridine and gemcitabine are utilised in chemotherapy for lung and colorectal cancer, respectively. Even though the function of the active drugs are distinct from each other, one common feature persists amongst all of these prodrugs—an amino acid residue is attached to the end of the molecule. With the exception of midodrine, the residue are linked to the drug compound through a hydroxyl, carboxyl or phosphoryl group via an ester linkage. For midodrine, an amide bond is used instead. Interestingly, the attached amino acid is always hydrophobic, i.e. valine for valacyclovir, valganciclovir, cidofovir and gemcitabine<sup>‡</sup>, glycine for midodrine and isoleucine for floxuridine<sup>§</sup>.

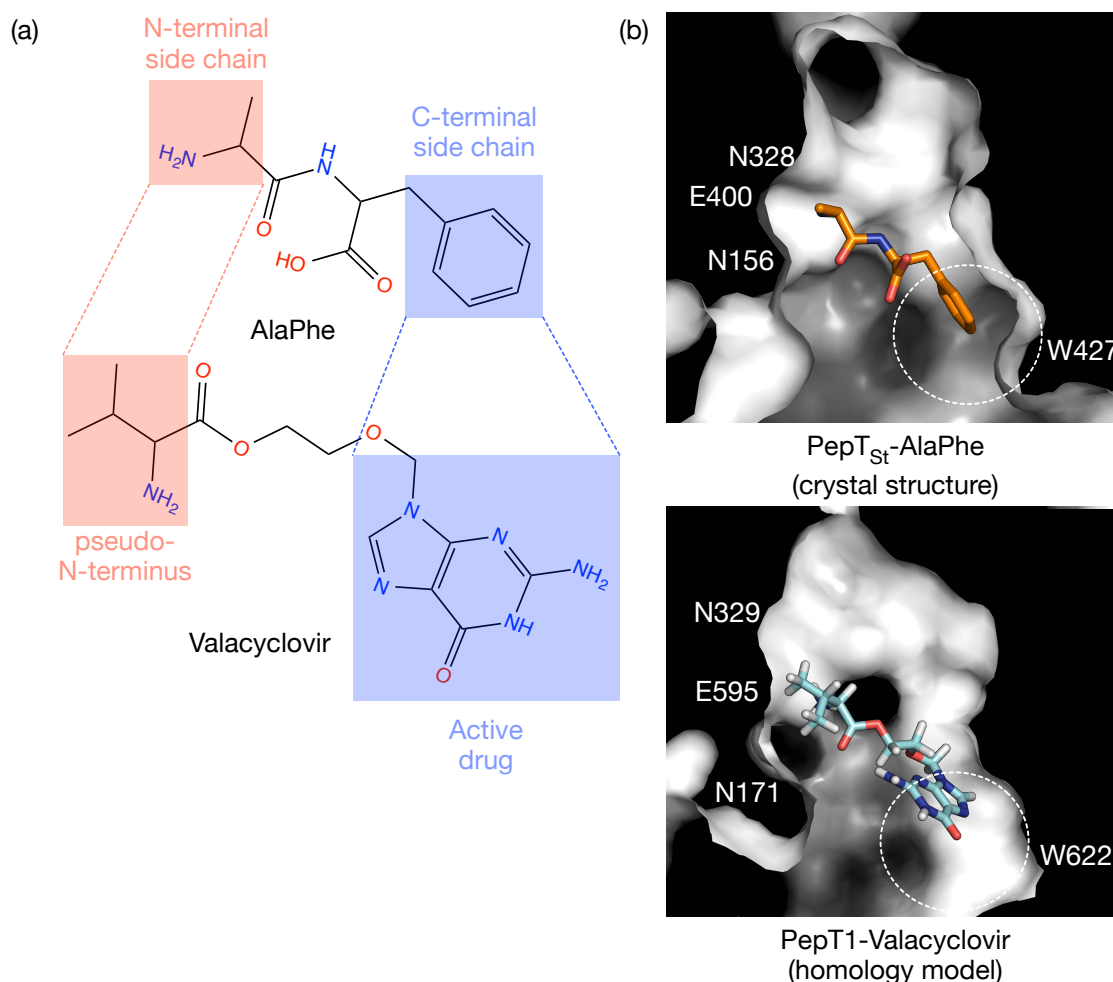
Structurally, these prodrugs resemble a dipeptide, with the additional amino acid mimicking the N-terminal residue and the active drug matching the position of the C-terminal side chain (Figure 6.11(a)). In Chapter 4, it was shown that in PepT<sub>St</sub> the binding of most dipeptides is sensitive to the N-terminus as changing the N-terminal side chain results in a larger perturbation in the overall binding free energy (Figure 4.6). Further examination revealed a strong interaction of the N-terminal side chain with a group of polar residues in the binding site of the protein (Figure 4.7). We propose peptide prodrugs bind to PepT1 in a similar way. The attached amino acid acts as a pseudo-N-terminus that strongly interacts with residues in the cavity, i.e. Asn329, Glu595 and Asn171 (Figure 6.11(b)). This hypothesis is further corroborated by previous findings by Borner et al. (1998) and Meredith et al. (2000) who demonstrated the central role of the N-terminus in peptide recognition in human PepT1, and suggested that the amino group is responsible in aligning the rest of the substrate in the binding site. It makes sense that all of the attached amino acid in the peptide prodrugs are hydrophobic as dipeptides with hydrophobic residues are the most preferred by this family of transporter, as was shown in Figure 4.4 and 4.6 for PepT<sub>St</sub> and previous studies on human PepT1 (Gebauer and Hartrodt, 2003; Vig et al., 2006; Biegel et al., 2005). Having a hydrophobic residue at this position of the prodrug will therefore guarantee a high affinity binding to the transporter.

Our free energy calculations also indicated that binding is less sensitive to the C-terminus

<sup>‡</sup>Isoleucyl-prodrug of gemcitabine also demonstrated an increase in uptake, but lower than the valyl counterpart (six- and eleven-fold, respectively) (Song et al., 2005).

<sup>§</sup>During the development of floxuridine (Tsume et al., 2008), other hydrophobic amino acid residues like valine, phenylalanine, leucine and glycine have been tested as well, but isoleucine gave the lowest IC<sub>50</sub> value i.e. highest affinity.



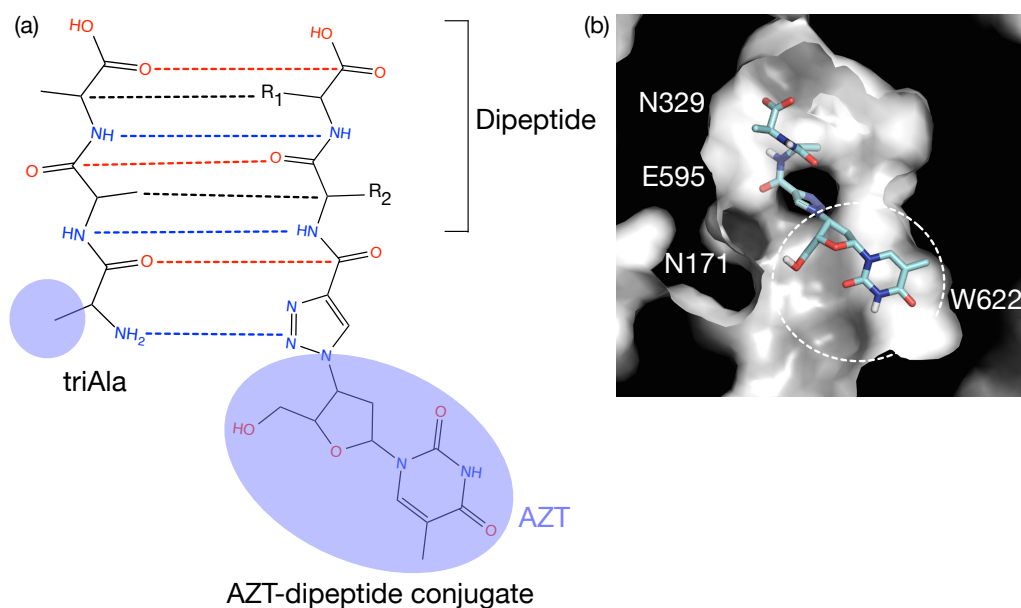


Figs. 6.11: A model of peptide prodrug binding to human PepT1. (a) Structural comparison of an antiviral prodrug, valacyclovir, with a dipeptide, AlaPhe reveals that the attached valine residue on the prodrug acts as a pseudo-N-terminus, whereas the active drug mimics the position of the C-terminal side chain of the dipeptide. (b) Surface representation of the binding site of PepT<sub>St</sub>-AlaPhe (crystal structure) and PepT1-valacyclovir (homology model) highlighting the position of phenylalanine side chain and acyclovir inside a hydrophobic cavity (dotted circles).

of a dipeptide. The structural reasoning behind this observation is the C-terminal side chain is accommodated inside an electro-neutral hydrophobic cavity that can adapt to various chemical group without a large energetic penalty (Figure 4.7). Our homology model of PepT1 predicted a similar hydrophobic pocket made of residues such as Tyr64, Trp294, Trp622, Leu623 and Val626. Tyrosine and tryptophan are both aromatic and polar and therefore the pocket, in principle, could adopt various different chemical groups. Assuming the peptide prodrugs bind to PepT1 like a dipeptide, then the active drug moiety will be positioned inside this large hydrophobic cavity. As this pocket is neutral in charge, it should allow a variety of different compounds inside it without a large change in the overall affinity. This would explain how PepT1 can accommodate a wide range of active drug agent of diverse chemistries. Taken together, this analysis suggests that the peptide prodrug approach has been successful owing to two reasons: i) the addition of amino acid gives the drug compound an N-terminal anchor to attain high affinity to the transporter and ii) the adaptable hydrophobic cavity in PepT1 can accommodate a wide range of drug moieties.

A recent study by Zhang et al. (2012) presented a new approach in the prodrug design for PepT1—instead of adding an amino acid residue, a dipeptide is attached to the pharmaceutically active agent. The drug of interest, azidothymidine (AZT), is an inhibitor of the reverse transcriptase enzyme and is used for antiretroviral therapy against HIV. A library of 61 different dipeptides is attached to AZT via an azide-alkyne cycloaddition reaction (Rostovtsev et al., 2002). Intriguingly, out of these dipeptides, the best affinity, measured by competition assays in Hela cells, is achieved by the addition of SerGlu. Conversely, the predictions for PepT<sub>St</sub> in Chapter 4 suggests that acidic dipeptides are very poor binders (Figure 4.4) and the same trend is also observed in human PepT1 (Biegel et al., 2005) and yeast Ptr2p (Ito et al., 2013). One possible explanation is that the addition of AZT changes the chemistry of the glutamate side chain, for example, by increasing its pK<sub>a</sub> resulting in a protonation, masking the negative charge.

Structural comparison of the AZT-dipeptide conjugate reveals that, unlike other prodrugs discussed earlier, this compound is more similar to a tripeptide (Figure 6.12(a)). The AZT moiety resembles the position of the first amino acid side chain, whilst the attached dipeptide acts as the second and third residues. The crystal complex of PepT<sub>St</sub>-triAla demonstrated that despite the different binding orientation compared to AlaPhe ('vertical' vs. 'horizontal'), the residues contacted in the binding cavity are essentially very similar (Lyons et al., 2014). The first and second side chains of triAla occupy the same binding pockets as the C-terminal and N-terminal side chains, respectively, of AlaPhe. By performing a structural alignment of the AZT-dipeptide prodrug with triAla, it was found that in our

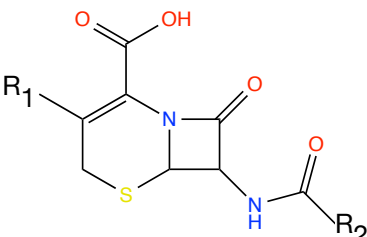


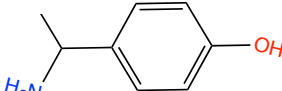
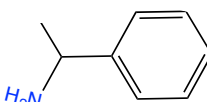
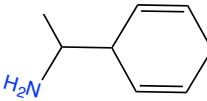
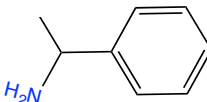
Figs. 6.12: The structure of azydothymidine (AZT)-dipeptide conjugate developed by Zhang et al. (2012) to target PepT1. (a) The prodrug is analogous to triAla where the AZT drug moiety mimics the position of the first side chain. (b) Proposed binding pose of AZT prodrug according to the vertical binding orientation of triAla. The hydrophobic pocket is highlighted the dotted circle.

PepT1 model, the AZT group is projected into the large hydrophobic cavity (Figure 6.12(b)), which is also occupied by drug moieties from other prodrugs. Hence, we propose that in spite of this different strategy, the molecular mechanism behind the success of these two approaches is fundamentally the same—that the large hydrophobic pocket provides a refuge for drug molecules to be transported by PepT1.

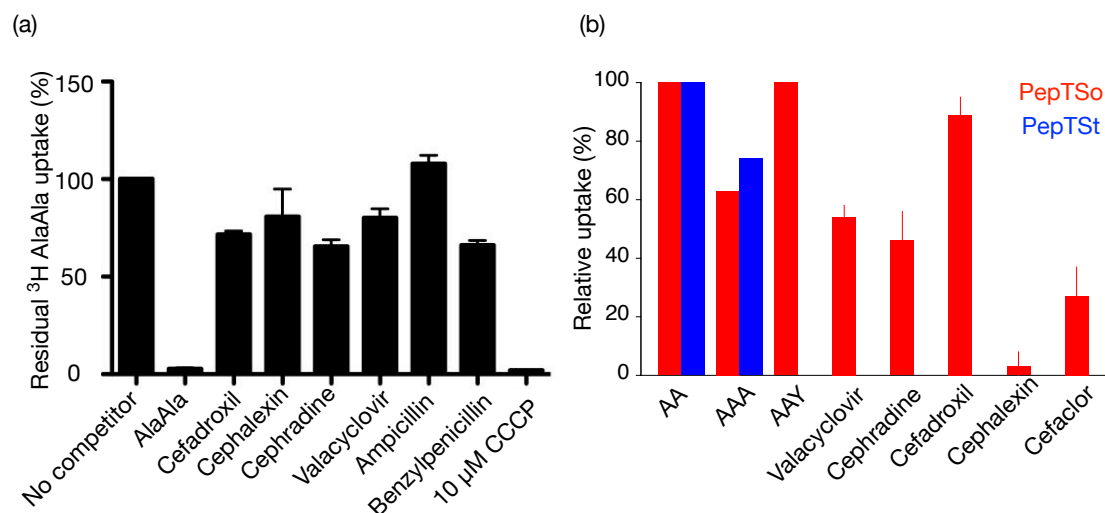
#### 6.4.2 A possible role of His57 in drug transport

Another category of drug that is a substrate of the human peptide transporters is the peptidomimetics, i.e. compounds that are structurally similar to peptides. The largest family of drugs that falls in this category is the  $\beta$ -lactam antibiotics (Biegel et al., 2005; Luckner and Brandsch, 2005). These drugs are prescribed to combat bacterial infections by inhibiting the biosynthesis of their cell wall. The core structure of a  $\beta$ -lactam antibiotics resembles the backbone of a tripeptide by having features like a peptide bond, a carboxylic acid group and an amino group (Figure 1.15). The two classes of  $\beta$ -lactam antibiotics, which demonstrate affinity to human PepT1, are the cephalosporins and penicillins, with the former generally displaying a lower  $K_i$  and higher oral bioavailability (Bretschneider et al., 1999). Four examples of cephalosporins with high intestinal absorption are shown in Figure 6.13, high-



| Name       | R <sub>1</sub>    | R <sub>2</sub>                                                                      | Bioavailability (%) |
|------------|-------------------|-------------------------------------------------------------------------------------|---------------------|
| Cefadroxil | CH <sub>3</sub> - |  | 100                 |
| Cephalexin | CH <sub>3</sub> - |  | 95                  |
| Cephradine | CH <sub>3</sub> - |  | 100                 |
| Cefaclor   | Cl-               |  | 100                 |

Figs. 6.13: Examples of the  $\beta$ -lactam antibiotics from cephalosporins family with high bioavailability. The core structure of cephalosporins is shown at the top. Percentage oral bioavailability is taken from Bretschneider et al. (1999).



Figs. 6.14: Experimental transport data of PepT<sub>St</sub> for several drug compounds from proteoliposome-based uptake assays. (a) Percentage residual uptake of <sup>3</sup>H-AlaAla from proton-driven competition transport assays (Solcan, 2013). (b) A direct measurement of peptide and drug uptake using pH based transport assays for PepT<sub>St</sub> (blue) and PepT<sub>So</sub> (red). Uptake was measured relative to AlaAla (AA). No transport was observed for AlaAlaTyr (AAY), valacyclovir, cephradine, cefadroxil, cephalixin and cefaclor by PepT<sub>St</sub> (Grba et. al, unpublished data).

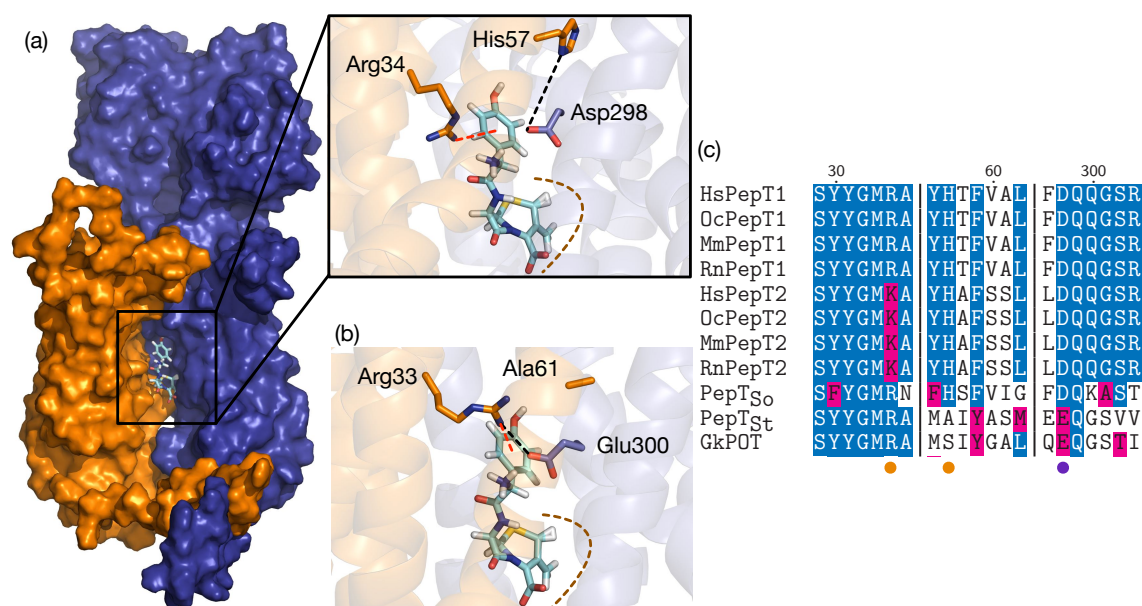
lighting the structures of their side chains. Several ligand-based studies have been performed to identify structural determinants for binding to human PepT1 by utilising the vast amount experimental data available in the literature; for example, Biegel et al. (2005) highlighted the importance of the size of the side chain attached to the lactam ring to determine affinities and that certain chemical groups like aminothiazole in ceftibuten can enhance binding.

Intriguingly, transport data from proteoliposome based uptake assays indicate that the  $\beta$ -lactam antibiotics (and drugs, in general) are much less favoured by the bacterial homolog, PepT<sub>St</sub> (Figure 6.14). *In vitro* competition assays using <sup>3</sup>H-AlaAla as a reporter substrate demonstrated marginal inhibition by drugs like cefadroxil and cephalixin while others such as ampicillin did not affect peptide uptake at all (Solcan, 2013), implying that these drugs have very poor binding affinities to this transporter. A further study using pH based assays that directly measured drug uptake into proteoliposomes corroborated this findings by showing no transport at all for all tested drugs and also for bulky tripeptides (Grba et. al, unpublished data). Their results suggest that drug binding to PepT<sub>St</sub> occurs mostly at low affinity (if at all), and once bound, the compounds are not transported across the membrane, essentially acting as a weak competitive inhibitor. These observations, however, are not applicable to all bacterial peptide transporters, as exemplified by PepT<sub>So</sub>. Grba et. al showed a

significant uptake for all tested drug compounds as well as bulky tripeptides in their PepT<sub>So</sub> transport assays (Figure 6.14(b)). This raises further questions of what makes PepT<sub>St</sub> different and how PepT1 and PepT<sub>So</sub> can transport bulky tripeptides, and most importantly, drug compounds?

One of the principle differences between PepT1 and PepT<sub>So</sub> with PepT<sub>St</sub> is a histidine residue on H2. His57 in human PepT1 (or His61 in PepT<sub>So</sub>) has been implicated in proton coupling as the effects of mutations at this position on peptide transport is pH dependent (Uchiyama et al., 2003; Newstead et al., 2011). This histidine residue is highly conserved in the mammalian peptide transporters PepT1 and PepT2 but is only present in select bacterial homologs, such as PepT<sub>So</sub>. At this position in PepT<sub>St</sub> and GkPOT, the histidine residue is replaced by an alanine and a serine, respectively (Figure 6.15(c)). In Chapter 5, it was proposed that His57 forms a proton tunnel due to the adjacency to an extracellular water pocket; the residue is suggested to transfer a proton from a water molecule to the extracellular gating residue, Asp298, initiating a conformational change from inward facing (gate close) to outward facing (gate open) (Figure 5.13). Consequently, it is possible that PepT1 and PepT<sub>So</sub> have a more efficient proton coupling mechanism compared to other POT transporters without a histidine at this position like PepT<sub>St</sub>. The transport of tripeptides, and by extension drugs, is coupled to fewer protons in PepT<sub>St</sub> (Parker et al., 2014). The lack of an efficient proton transfer strategy, therefore, could potentially exacerbate or even eliminate drug uptake, which would explain why PepT<sub>St</sub> does not transport drugs but PepT1 and PepT<sub>So</sub> do.

Another possible interpretation of this observation is that His57 itself can be a gating residue. This brings us back to the simulations of PepT<sub>St</sub> with tripeptides containing aromatic side chains, i.e. phenylalanine and tyrosine, in §4.3.5. Positioning either one of these residues on the C-terminus is detrimental to binding as depicted by the positive values of  $\Delta\Delta G$  (Table 4.4 and Figure 4.11). Upon closer inspection, it was found that a gating residue, Arg33, interacts with the benzene/phenyl ring via a cation- $\pi$  interaction, which potentially disrupts the formation of the extracellular gate, therefore preventing the transport cycle. Arg33 has been shown to be instrumental in peptide transport as the R33A mutation abolished proton-coupled uptake activity (Solcan et al., 2012). Similar results are obtained for GkPOT when the equivalent residue, Arg43, is mutated (Doki et al., 2013). These arginine residues are proposed to form the extracellular gate for these bacterial transporters by interacting with Glu300 (PepT<sub>St</sub>) and Glu310 (GkPOT). Interaction with a peptide substrate, much like these mutations, will deprive the glutamate residues on the C-terminal bundle of a salt bridge pair on the N-terminal bundle, essentially interrupting the conformational



Figs. 6.15: A possible role of His57 as an extracellular gate in drug binding. (a) Proposed binding orientation of a  $\beta$ -lactam antibiotics, cefadroxil (cyan) based on the vertical pose of triAla. Arg34 can form a cation- $\pi$  interaction (red dashed line) with the phenyl group of the drug and His57 potentially substitutes this residue as a gating residue and interacts with Asp298 (black). The approximate position of the hydrophobic pocket is indicated by a brown dashed line. The N-terminal bundle is coloured in orange and the C-terminal bundle in purple. H2 and H11 are removed for clarity. (b) A similar binding pose of cefadroxil to PepT<sub>St</sub> with Arg33-Glu300 as the extracellular gate and Ala61 in place of His57. (c) Sequence alignment of mammalian PepT1 and PepT2 with three bacterial homologs, highlighting the gate residues in orange (N-terminal) and purple (C-terminal).

change. As illustrated in Figure 6.13, all of the  $\beta$ -lactam antibiotics tested in the uptake assay of PepT<sub>St</sub> (Figure 6.14) have an aromatic ring on the R<sub>2</sub> position<sup>¶</sup>. This electron-rich  $\pi$  system could form a similar cation- $\pi$  interaction with Arg33 (Figure 6.15(b)), as has been observed in the simulations of PepT<sub>St</sub> with AlaAlaPhe and AlaAlaTyr, which would explain the lack of transport.

In PepT<sub>So</sub>, however, Arg32 appears to be less critical to transport as mutation to alanine did not hinder peptide uptake (Solcan, 2013). While the equivalent residue in human PepT1 (Arg33) has not been subject to any functional studies, we expect a similar result. The presence of His57 in PepT1 (or His61 in PepT<sub>So</sub>) on H2 in the N-terminal bundle increases the versatility of the extracellular gate as it can also be the salt bridge partner to Asp298 (Asp316) in the C-terminal bundle. Hence, the arginine residues could be employed to stabilised the binding of the  $\beta$ -lactam antibiotic (Figure 6.15(a)), which would explain their high affinities and relative uptake activities in PepT1 and PepT<sub>So</sub>. For this hypothesis to be true, the  $\beta$ -lactam antibiotic must be positioned vertically in the triAla binding orientation (Lyons et al., 2014) with the lactam ring projected into the hydrophobic pocket and the R<sub>2</sub> group oriented towards the extracellular side. The large size of the lactam ring means there is a limit as to how big the R<sub>1</sub> group can be if the hydrophobic cavity is to accommodate it, which justifies the results from Biegel et al. (2005) who discovered that the size of this side chain affects binding affinities. This would also explain why cephalosporins with extended aromatic R<sub>1</sub> side chains like cefsulodin and cefodizime have poor oral bioavailability (Bretschneider et al., 1999).

## 6.5 Conclusion

The promising results for binding affinity predictions in PepT<sub>St</sub> using a range of free energy methods (Chapter 3 and Chapter 4) provides a blueprint to study the pharmaceutically important human PepT1. Applying the same methodologies to the homology model of PepT1 built in Chapter 5, however, yields in disappointing results. All end-point methods fail to distinguish good substrates from poor ones and the correlation to experimental data is in the same range as the AutoDock Vina scoring function. Several limitations in the current study that was not previously present in the prediction with PepT<sub>St</sub> were acknowledged. These include the use of a generic force field to parametrise the drug compounds and the less ac-

---

<sup>¶</sup>While most cephalosporins and penicillins have an electron-rich ring system on their side chain, some compounds like ceftibuten and cyclacillin have a side chain made of a fully saturated ring, and yet still display high affinities towards human PepT1 (Bretschneider et al., 1999), suggesting that other chemical groups and residues are also involved in drug binding to PepT1.



curate experimental data from whole cell uptake assays, both of which, however, are shown to not significantly affect the results.

Further analyses point towards the importance of a high resolution crystal structure for binding free energy predictions or even for protein-ligand simulations in general. Homology modelling, whilst good at predicting the overall fold of a protein given an adequate sequence similarity, is not sufficient to accurately model the coordinates of the amino acid side chains, which is the key component of ligand interactions in this protein family. Even at the upper limit, when a model is built using the crystal structure of itself as a template, the slight structural differences at the beginning of an MD simulation can potentially propagate during the simulation and induce erroneous sampling of the ligand molecule and ultimately incorrect  $\Delta G$  calculations. Considering the difficulty in crystallising mammalian membrane proteins, the current lack of structural data will inevitably pose a significant bottleneck in affinity predictions for pharmacologically relevant transporters.

The last resort is to establish a good model system—in the case of the human PepT1, using available crystal structures of the bacterial homologs—from which meaningful deductions on the interactions of the protein with pharmaceutical agents can be made. Based on comparison to PepT<sub>St</sub>, we propose that peptide prodrugs bind to human PepT1 like a dipeptide, i.e. in a lateral pose with the attached amino acid providing an N-terminal anchor to the binding site whilst the active drug moiety itself is accommodated inside a hydrophobic pocket. We also underline the possible role of His57 as an additional gate residue to drive the more efficient transport of drug compounds, especially the  $\beta$ -lactam antibiotics, across the plasma membrane. These hypotheses rely on the assumption that peptide and drug binding in PepT1 occurs in a largely similar way to PepT<sub>St</sub>. Future studies in the form of biophysical and biochemical assays are required to verify these key chemical groups and residues, and ultimately a high resolution crystal structure of the transporter bound to a peptide or a drug is a crucial step forward in the field of drug design for PepT1.

# Chapter 7

## Summary and future perspectives

In 2011 when this research project started, the first crystal structure of a proton-dependent oligopeptide transporter, PepT<sub>So</sub>, had just been published by my experimental collaborator, Simon Newstead (Newstead et al., 2011). Not long afterwards, they solved a structure of a bacterial homolog, PepT<sub>St</sub>, in a different conformational state (Solcan et al., 2012), and collected diffraction data from crystals of the extracellular domain (ECD) of mouse PepT1 (Beale et al., 2015). These structural data opened a myriad of possibilities for studying computationally the POT transporters with the ultimate goal of utilising them as a molecular delivery system for orally administered drug compounds. At the beginning, this thesis branched out in two main directions. The first of these was to build a full-length model of the human peptide transporter PepT1 in collaboration with Johnny Beale, a DPhil student in Simon Newstead's group, with an aim of understanding the role of the ECD in peptide transport. The second direction was to computationally predict the binding affinities, and by extension, the transport of substrates, i.e. peptides and more importantly, drugs, using a range of free energy methods. This part was made possible by the recent publication of co-crystallised structures of PepT<sub>St</sub>-AlaPhe and PepT<sub>St</sub>-AlaAlaAla, which provided the first insights into how ligands interact with this transporter (Lyons et al., 2014). In this final chapter, we shall first summarise each of the preceding results chapters briefly, discuss the relevance of the research, and then suggest potential future directions.

### 7.1 Summary and relevance of research

In Chapter 3, a range of binding affinity prediction methods were assessed using a set of experimental data from competition transport assays of PepT<sub>St</sub>. This chapter laid the ground work by determining which methods were most reliable at reproducing the ranking of lig-

ands: these were then employed in subsequent chapters. The promiscuity of the POT transporter family was studied in depth in Chapter 4 using the model bacterial transporter, PepT<sub>St</sub>, by modelling the binding and predicting the affinities of all 400 possible dipeptides. Our predictions in this chapter were validated by proteoliposome competition transport assays performed by Simon Newstead's group. The building of the homology models of human PepT1 and PepT2 was described in Chapter 5. To date, they represent the most complete three dimensional structural characterisation of these pharmaceutically relevant transporters. Using the PepT1 homology model, we then modelled the binding of peptides and drugs in Chapter 6, and attempted to compute their binding affinities. This chapter also extended our findings on the model transporter, PepT<sub>St</sub>, in Chapter 4 to the rational design of drugs targeting PepT1.

This research project has advanced our understanding of peptide transporters primarily in two ways. Methodologically, it was shown that free energy methods can indeed be used to accurately predict the binding affinities of ligands to a peptide transporter. The end-point methods like the linear interaction energy (LIE) method are a good compromise between accuracy and computational expenses as they are able to broadly classify ligands into strong and weak binders. The more rigorous method, thermodynamic integration, is useful for refinement of results or to answer specific questions, for example which part of the ligand molecule is sensitive to perturbation. The use of free energy methods to study protein-ligand interactions *per se* is not novel, although their application to membrane transporters is, to the best of our knowledge. Membrane transporters undergo complex conformational changes, which raises the question of whether the binding affinity of a ligand to one particular conformation is representative of the whole transport cycle. Also, in the absence of binding assays, which are typically difficult to set up for membrane proteins, ligand affinity data are limited to indirect measurements, for example the IC<sub>50</sub> values from competition transport assays, which might not be correlated to binding energy. In this thesis, we showed that, at least for the bacterial peptide transporter PepT<sub>St</sub>, *in silico* binding energy predictions are not only able to reproduce ligand ranking based on available IC<sub>50</sub> values, but also able to expand our current understanding on the different ways ligands can interact with the transporter. It must be emphasised that the accurate computation of binding affinities is contingent on the availability of a crystal structure; applying the same methods to the human PepT1 homology model returned discouraging results. This highlights the limitation of the current homology modelling algorithms and the importance of protein structural determination by X-ray crystallography.

Another key advancement contributed by this thesis lies in the biology of the peptide

transporter itself. From the complex crystal structures solved by Lyons et al. (2014), we know that the promiscuity of this transporter family arises, in large part, from their ability to accommodate ligands in various orientations using the same binding cavity. In this work, it was demonstrated that a dipeptide with a large side chain, like lysine, may be oriented vertically like a tripeptide, further corroborating the idea that this transporter utilises its binding cavity in a flexible fashion. Also, we delineated the origin of binding selectivity for dipeptides to their N-terminal side chains through a series of alchemical transformations. The binding of the C-terminal side chain was found to be less sensitive to structural changes, suggesting the energetic landscape of the binding site is asymmetrical. These results have important implications for rational drug design targeting human PepT1. The current approach of making prodrugs by attaching an amino acid residue to an active drug moiety can be justified as the amino acid residue mimics the N-terminus for the drug compound and therefore the drug gains affinity for the transporter. As such, the choice of amino acid to be added to the drug compounds is of utmost importance to obtain the highest affinity. Whilst the binding of tripeptides were not studied in as much detail as that of dipeptides in this thesis, it was shown that certain bulky residues like phenylalanine and tyrosine can potentially reduce affinities. These results will be relevant in the design of prodrugs mimicking a tripeptide, in which case a dipeptide is attached to the drug moiety, as performed by Zhang et al. (2012). Nevertheless, as all of these results were obtained from the model peptide transporter, PepT<sub>St</sub>, it is essential to validate them with human PepT1 once the structure becomes available.

## 7.2 Future directions

Our computational study of peptide transporter in this thesis is by no means complete. There are still many aspects that are not covered directly in this thesis and can potentially be addressed in future work. First and foremost is the protonation of the residues in the binding site. There are six protonatable residues in the binding site of PepT<sub>St</sub> and five in human PepT1. Parker et al. (2014) recently showed that the transport of tripeptide by PepT<sub>St</sub> is coupled to three protons whereas the transport of dipeptide is coupled to four or five protons. Mutagenesis studies of PepT<sub>St</sub> have shown that removing any of these protonatable residues abolishes proton-coupled transport (Solcan et al., 2012; Solcan, 2013), suggesting they are all protonated at some point during the transport cycle and at least three and four/five residues are deprotonated to release tripeptide and dipeptide, respectively, from the binding site. Due to the resolution of the available X-ray crystal structures, however, the

protonation state of these residues in the crystallised inward open structure of PepT<sub>St</sub> is not known. For simplicity, we used a structure where all of these residues are deprotonated in our modelling and simulations. It would be interesting to repeat the free energy calculation with a structure where some or all of the protonatable residues are protonated to see how this affects the affinity of ligand to the transporter. As outlined in §1.2.4, protonation of the gating residues is proposed to play a crucial role in the conformational changes of the transporter and it is possible that this alters the strength of binding of the ligand molecule to facilitate its release into the cytoplasm. We showed in Chapter 3 that the thermodynamic integration method is sufficiently accurate to discern small structural changes, therefore one way to investigate this is to perform alchemical transformation of the protonatable residues in the binding site, for example Glu400 to GluH400, and calculate the relative free energy of binding of a ligand. Understanding the effect of protonation on ligand binding would advance our knowledge on the coupling of proton transport to substrate translocation by the POT family transporters.

Another aspect that has not been addressed in this thesis is the role of water in ligand binding. Water is crucial in coordinating a ligand in the binding site via a complex network of hydrogen bonds with the protein (Lu et al., 2007). The accurate placement of water molecules in the binding site during an MD simulation is therefore essential for a reliable prediction of binding affinities (Luccarelli et al., 2010) and, by extension, successful computational design of drugs (de Beer et al., 2010). Based on our simulations, the binding cavity of PepT<sub>St</sub> are fully solvated in the presence of ligand due to the existence of several polar residues in the binding site such as Tyr68 and Asn328. The polar nature of the backbone of a peptide substrate suggests that some water molecules might serve as stabilising bridges in the interface of the transporter and peptide substrate. We also showed in §4.3.1 that dipeptides with polar side chains are amongst the best substrates of PepT<sub>St</sub>, further hinting at the potential role of water in peptide binding. Due to the moderate resolution of the crystal complexes of PepT<sub>St</sub>-AlaPhe and PepT<sub>St</sub>-AlaAlaAla, however, no water molecules were resolved in the binding cavity. As such, there is currently no experimental data on structurally conserved water molecules for this transporter. To approach this problem computationally, we could map the water density in the binding site from our MD simulation of the transporter in the ligand-bound and ligand-unbound states. This would help identify any water molecules that are not displaced during peptide binding and thus are potentially key to the formation of protein-ligand complex. We could then compare these results to existing computational methods that predict the position of water in the binding site, for example the WaterDock algorithm developed by Ross et al. (2012). These predictions could be validated

when higher resolution crystal structures become available in the future.

It was shown for the sugar transporter GLUT1 that apparent binding affinity of glucose differs in the outward facing and the inward facing states (Brahm, 1983). Given the structural similarity of GLUT1 and the POT family transporters, it is highly likely that the same observation also applies to PepT<sub>St</sub>. Hence, it would be interesting to computationally predict the binding affinities of ligands to this transporter in other conformational states than the inward open state that has been performed in this thesis. Of course, we are currently hampered by the lack of crystal structure of PepT<sub>St</sub> in the outward open configuration. The repeat swapped modelling technique introduced by Forrest (2013) has been successful in generating an alternative conformation of a transporter from available crystal structures. However, this approach was not pursued in this thesis as we have shown that binding affinity predictions using homology models invariably produces inferior results for this transporter family. Also, it is not known whether the binding pose of a peptide ligand to PepT<sub>St</sub> in the outward open conformation is the same as the inward open conformation. Nevertheless, when such a structure becomes available, extending our binding energy prediction to the outward facing state would elucidate the changes in affinities during the transport cycle and help us understand the mechanism of initial binding and eventual release of substrates.

While this research project focused primarily on protein-ligand binding, there are various other aspects regarding the biology of peptide transporters that can be studied computationally. One example is investigating large-scale conformational changes. Various advanced computational techniques such as non-equilibrium steered MD and enhanced sampling protocols like metadynamics have been employed to simulate large structural transitions of membrane transporters (Moradi and Tajkhorshid, 2015). Long-timescale unbiased atomistic MD simulations are also capable of generating an alternative conformation, for example the partially outward facing state of PepT<sub>So</sub> (Fowler et al., 2015). Apart from illuminating the molecular mechanism of transport cycle, such studies also provide inputs for cross linking studies to capture the transporters in a particular conformation for crystallisation. To date, most studies have focused on simulating the transitional change of an *apo* transporter; it would be intriguing to perform such simulation in a ligand-bound state to understand the different binding poses explored by the substrate during a transport cycle. These results could then feed into the binding affinity prediction of ligands in different conformational states of transporters.

Additionally, specific protein-lipid interactions can also be studied using MD simulations. Several integral membrane proteins have been shown to bind specifically to certain species of lipids such as phosphatidyl inositol phosphate and cholesterol (Laganowsky et al.,

2014; Fantini and Barrantes, 2013). For example, the crystal structure of the *Drosophila melanogaster* dopamine transporter includes a cholesterol molecule inserted between its transmembrane helices and it is speculated this stabilises the outward open conformation (Penmatsa et al., 2013). Little is currently known about the interaction of lipids with peptide transporters. *In silico*, this could be examined by performing simulations of the transporter in a complex lipid bilayer with a composition mimicking the plasma membrane. The density of each lipid species throughout the simulation could then be mapped around the protein to pinpoint any selective binding. Certain lipid species like phosphatidyl serine are enriched in the apical epithelial membrane where PepT1 is found (Gerl et al., 2012), and thus could potentially play a role in its function.

Finally, the ultimate goal of this thesis is to predict binding and transport of drug compounds by the human peptide transporter PepT1 to improve oral bioavailability. Alas, we showed in §6.3.1 and §6.3.4 that homology modelling does not produce a sufficiently reliable structure for an accurate prediction of binding affinities. The determination of the structure of mammalian peptide transporters by X-ray crystallography would therefore significantly advance this field. The availability of a high-resolution crystal structure would enable the binding of drug molecules to the human transporter to be modelled. A high-throughput screening procedure could then be set up using the end-point methods to filter out strong binders from a compound library, and theoretically rigorous free energy methods could be subsequently applied to optimise drug design. With the ever-growing speed and availability of modern computational resources, we expect this protocol to be applied at the front line of drug discovery in the near future, not only for PepT1, but also for other medically relevant membrane transporters.

# References

- Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., and Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* (80-. ). *301*, 610–5.
- Accardi, A. and Picollo, A. (2010). CLC channels and transporters: Proteins with borderline personalities. *Biochim. Biophys. Acta* *1798*, 1457–1464.
- Acharya, C., Coop, A., Polli, J.E., and MacKerell Jr., A.D. (2011). Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. *Curr. Comput. Aided Drug Des.* *7*, 10–22.
- Addison, J., Burston, D., Dalrymple, J., Matthews, D., Payne, J., Sleisenger, M., and Wilkinson, S. (1975). A common mechanism for transport of di- and tri-peptides by hamster jejunum in vitro. *Clin. Sci. Mol. Med.* *49*, 313–22.
- Adibi, S.A. and Morse, E.L. (1977). The number of glycine residues which limits intact absorption of glycine oligopeptides in human jejunum. *J. Clin. Inv.* *60*, 1008–1016.
- Adzhubei, I., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* *7*, 248–9.
- Akyuz, N., Georgieva, E.R., Zhou, Z., Stolzenberg, S., Cuendet, M.a., Khelashvili, G., Altman, R.B., Terry, D.S., Freed, J.H., Weinstein, H., et al. (2014). Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature* *518*, 68–73.
- Almlof, M., Brandsdal, B.O., and Aqvist, J. (2004). Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J. Comp. Chem.* *25*, 1242–54.
- Almlof, M., Carlsson, J., and Aqvist, J. (2007). Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* *3*, 2162–2175.
- Amasheh, S., Wenzel, U., Boll, M., Dorn, D., Weber, W., Clauss, W., and Daniel, H. (1997). Transport of Charged Dipeptides by the Intestinal H<sup>+</sup>/Peptide Symporter PepT1 Expressed in *Xenopus laevis* Oocytes. *J. Membr. Biol.* *155*, 247–256.
- Aqvist, J. and Hansson, T. (1996). On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* *100*, 9512–9521.
- Aqvist, J., Medina, C., and Samuelsson, J. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* *73*, 385–91.



- Ashcroft, F., Gadsby, D., and Miller, C. (2009). The blurred boundary between channels and transporters. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 364, 145–147.
- Bailey, P.D., Boyd, C.A., and Bronk, J.R. (2000). How to Make Drugs Orally Active: A Substrate Template for Peptide Transporter PepT1. *Angew. Chem. Int. Ed* 39, 505–508.
- Bairoch, A. and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28, 45–48.
- Bayly, C.I., Cieplak, P., Cornell, W.D., and Kollman, P.A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 97, 10269–10280.
- Beale, J., Parker, J.L., Samsudin, M.F., Barrett, A.L., Senan, A., Bird, L.E., Scott, D., Owens, R.J., Sansom, M.S., Tucker, S.J., et al. (2015). Crystal structures of the extracellular domain from PepT1 and PepT2 provide novel insights into mammalian peptide transport. *Structure* 23, 1889–1899.
- Bello, M., Martínez-Archundia, M., and Correa-Basurto, J. (2013). Automated docking for novel drug discovery. *Expert Opin. Drug Discov.* 8, 821–34.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- Beutler, T., Mark, A., and van Schaik, R. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* 222, 529–539.
- Beutner, K.R., Friedman, D.J., Forszpaniak, C., Andersen, P.L., and Wood, M.J. (1995). Valaciclovir compared with acyclovir for improved therapy for herpes zoster in immunocompetent adults. *Antimicrob. Agents Chemother* 39, 1546–1553.
- Biegel, A., Gebauer, S., Hartrodt, B., Brandsch, M., Neubert, K., and Thondorf, I. (2005). Three-dimensional quantitative structure-activity relationship analyses of beta-lactam antibiotics and tripeptides as substrates of the mammalian H<sup>+</sup>/peptide cotransporter PEPT1. *J. Med. Chem.* 48, 4410–4419.
- Biegel, A., Knütter, I., Hartrodt, B., Gebauer, S., Theis, S., Luckner, P., Kottra, G., Rastetter, M., Zebisch, K., Thondorf, I., et al. (2006). The renal type H<sup>+</sup>/peptide symporter PEPT2: structure-affinity relationships. *Amino Acids* 31, 137–56.
- Bippes, C., Ge, L., and Meury, M. (2013). Peptide transporter DtpA has two alternate conformations, one of which is promoted by inhibitor binding. *Proc. Natl. Acad. Sci. U.S.A.* 3978–3986.
- Bishop, W.R. and Bell, R.M. (1988). Assembly of phospholipids into cellular membranes: biosynthesis, transmembrane movement and intracellular translocation. *Annu. Rev. Cell. Biol.* 4, 579–610.

- Bjelic, S., Nervall, M., Gutiérrez-de Terán, H., Ersmark, K., Hallberg, A., and Aqvist, J. (2007). Computational inhibitor design against malaria plasmepsins. *Cell. Mol. Life Sci.* 64, 2285–305.
- 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., et al. (1998). Structure, function, and molecular modeling approaches to the study of the intestinal dipeptide transporter PepT1. *J. Pharm. Sci.* 87, 1286–91.
- 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. *Proc. Natl. Acad. Sci. U.S.A.* 93, 284–9.
- Boll, M., Markovich, D., Weber, W.M., Korte, H., Daniel, H., and Murer, H. (1994). Expression cloning of a cDNA from rabbit small intestine related to proton-coupled transport of peptides, beta-lactam antibiotics and ACE-inhibitors. *Pflügers Arch.* 429, 146–9.
- Boresch, S. (2002). The Role of Bonded Energy Terms in Free Energy Simulations - Insights from Analytical Results. *Mol. Sim.* 28, 13–37.
- Boresch, S. and Karplus, M. (1999). The Role of Bonded Terms in Free Energy Simulations: 1. Theoretical Analysis. *J. Phys. Chem. A* 103, 103–118.
- Borner, V., Fei, Y.J., Hartrodt, B., Ganapathy, V., Leibach, F.H., Neubert, K., and Brandsch, M. (1998). Transport of amino acid aryl amides by the intestinal H<sup>+</sup>/peptide cotransport system, PEPT1. *Eur. J. Biochem.* 255, 698–702.
- Boyd, C.A. and Ward, M.R. (1982). A micro-electrode study of oligopeptide absorption by the small intestinal epithelium of *Necturus maculosus*. *J. Physiol.* 324, 411–428.
- Brahm, J. (1983). Kinetics of glucose transport in human erythrocytes. *J. Physiol.* 339, 339–354.
- Brandsch, M. (2009). Transport of drugs by proton-coupled peptide transporters: pearls and pitfalls. *Expert Opin Drug Metab Toxicol.* 5, 887–905.
- Brandsch, M., Knutter, I., and Leibach, F.H. (2004). The intestinal H<sup>+</sup>/peptide symporter PEPT1: structure–affinity relationships. *Eur. J. Pharm. Sci.* 21, 53–60.
- Brandsch, M., Knutter, I., Thünecke, F., Hartrodt, B., Born, I., Borner, V., Hirche, F., Fischer, G., and Neubert, K. (1999). Decisive structural determinants for the interaction of proline derivatives with the intestinal H<sup>+</sup>/peptide symporter. *Eur. J. Biochem.* 508, 502–508.
- Bretschneider, B., Brandsch, M., and Neubert, R. (1999). Intestinal Transport of Beta-Lactam Antibiotics: Analysis of the Affinity at the Proton-Peptide Symporter (PepT1), the Uptake into Caco-2 Cell Monolayers and the Transepithelial Flux. *Pharm. Res.* 16, 55–61.

- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4, 187–217.
- Brown, S.P. and Muchmore, S.W. (2009). Large-scale application of high-throughput molecular mechanics with Poisson-Boltzmann surface area for routine physics-based scoring of protein-ligand complexes. *J. Med. Chem.* 52, 3159–3165.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101.
- Camacho, C.J., Kimura, S.R., Delisi, C., and Vajda, S. (2000). Kinetics of Desolvation-Mediated Protein-Protein Binding. *Biophys. J.* 78, 1094–1105.
- Caporuscio, F. and Tafi, A. (2011). Pharmacophore modelling: a forty year old approach and its modern synergies. *Curr. Med. Chem* 18, 2543–53.
- Carlson, H.A. and Jorgensen, W.L. (1995). An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* 10667–10673.
- Casey, J.R., Grinstein, S., and Orlowski, J. (2010). Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* 11, 50–61.
- 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. Cryst. D* 66, 12–21.
- Chen, X.Z., Steel, A., and Hediger, M.A. (2000). Functional roles of histidine and tyrosine residues in the H(+)-peptide transporter PepT1. *Biochem. Biophys. Res. Commun.* 272, 726–730.
- Chen, X.Z., Zhu, T., Smith, D.E., and Hediger, M.A. (1999). Stoichiometry and kinetics of the high-affinity H+-coupled peptide transporter PepT2. *J. Biol. Chem.* 274, 2773–2779.
- Chen, Y. and Shoichet, B.K. (2009). Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* 5, 358–64.
- Cheng, T., Li, X., Li, Y., Liu, Z., and Wang, R. (2009). Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* 49, 1079–93.
- Claro Da Silva, T., Polli, J.E., and Swaan, P.W. (2013). The solute carrier family 10 (SLC10): Beyond bile acid transport. *Mol. Asp. Med.* 34, 252–269.
- Clement, O.O. and Mehl, A.T. (2000). HipHop: Pharmacophore based on multiple common-feature alignments. In O.F. Guner, ed., *Pharmacophore Perception, Dev. Use Drug Des.*, 69–84 (San Diego, CA: TBS The Book Service Ltd).
- Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 36, 197–201.
- Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004). ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* 32, 96–99.

- Covitz, K.M.Y., Amidon, G.L., and Sadée, W. (1998). Membrane topology of the human dipeptide transporter, hPEPT1, determined by epitope insertions. *Biochemistry* 37, 15214–15221.
- Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190.
- Cusatis, G., Gregorc, V., Li, J., Spreafico, A., Ingersoll, R.G., Verweij, J., Ludovini, V., Villa, E., Hidalgo, M., Sparreboom, A., et al. (2006). Pharmacogenetics of ABCG2 and adverse reactions to gefitinib. *J. Natl. Cancer Inst.* 98, 1739–1742.
- Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J., and Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature* 467, 734–738.
- Daniel, H. and Adibi, S.A. (1994). Functional separation of dipeptide transport and hydrolysis in kidney brush border membrane vesicles. *FASEB J.* 8, 753–759.
- Daniel, H. and Kottra, G. (2004). The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflügers Arch.* 447, 610–8.
- Daniel, H. and Spanier, B. (2006). From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology* 21, 93–102.
- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092.
- de Beer, S.B.A., Vermeulen, N.P.E., and Oostenbrink, C. (2010). The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* 10, 55–66.
- Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M., and Yan, N. (2014). Crystal structure of the human glucose transporter GLUT1. *Nature* 510, 121–125.
- Detmers, F.J.M., Kunji, E.R.S., Lanfermeijer, F.C., Poolman, B., and Konings, W.N. (1998). Kinetics and specificity of peptide uptake by the oligopeptide transport system of *Lactococcus lactis*. *Biochemistry* 37, 16671–16679.
- Dickson, C.J., Rosso, L., Betz, R.M., Walker, R.C., and Gould, I.R. (2012). GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* 8, 9617.
- Do, C.B., Mahabhashyam, M.S.P., Brudno, M., and Batzoglou, S. (2005). ProbCons : Probabilistic consistency-based multiple sequence alignment. *Genome Res.* 330–340.
- Dobson, P.D. and Kell, D.B. (2008). Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat. Rev. Drug Discov.* 7, 205–220.
- Doki, S., Kato, H.E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N., Tsukazaki, T., Sugita, Y., Kandori, H., et al. (2013). Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc Natl Acad Sci U S A* 110, 11343–8.

- 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. *J. Physiol.* *497*.3, 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. *J. Membr. Biol.* *186*, 55–62.
- Döring, F., Walter, J., Will, J., Föcking, M., Boll, M., Amasheh, S., Clauss, W., and Daniel, H. (1998). Delta-aminolevulinic acid transport by intestinal and renal peptide transporters and its physiological and clinical implications. *J. Clin. Inv.* *101*, 2761–2767.
- 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. *J. Neurochem.* *71*, 388–393.
- Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T., and MacKinnon, R. (2002). X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* *415*, 287–294.
- Eddy, E.P., Wood, C., Miller, J., Wilson, G., and Hidalgo, I.J. (1995). A comparison of the affinities of dipeptides and antibiotics for the di- / tripeptide transporter in Caco-2 cells. *Int. J. Pharm* *115*, 79–86.
- Ekins, S., Johnston, J.S., Bahadduri, P., D’Souza, V.M., Ray, A., Chang, C., and Swaan, P.W. (2005). In vitro and pharmacophore-based discovery of novel hPEPT1 inhibitors. *Pharm. Res.* *22*, 512–7.
- Eraly, S.A. (2008). Implications of the alternating access model for organic anion transporter kinetics. *J. Membr. Biol.* *226*, 35–42.
- Ernst, H.A., Pham, A., Hald, H., Kastrup, J.S., Rahman, M., and Mirza, O. (2009). Ligand binding analyses of the putative peptide transporter YjdL from *E. coli* display a significant selectivity towards dipeptides. *Biochem. Biophys. Res. Com.* *389*, 112–116.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* *103*, 8577.
- Eswar, N., John, B., Mirkovic, N., Fiser, A., Ilyin, V.a., Pieper, U., Stuart, A.C., Marti-Renom, M.A., Madhusudhan, M.S., Yerkovich, B., et al. (2003). Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res.* *31*, 3375–3380.
- Ewald, P.P. (1921). Die Berechnung optischer und elektrostatische Gitterpotentiale. *Ann. Phys.* *64*, 253–287.
- Ewing, T.J.A., Makino, S., Skillman, A.G., and Kuntz, I.D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* *15*, 411–428.
- Fang, G., Konings, W., and Poolman, B. (2000). Kinetics and Substrate Specificity of Membrane-Reconstituted Peptide Transporter DtpT of *Lactococcus lactis*. *J. Bacteriol.* *182*.

- Fann, M.C., Busch, A., and Maloney, P.C. (2003). Functional Characterization of Cysteine Residues in GlpT, the Glycerol 3-Phosphate Transporter of *Escherichia coli*. Functional Characterization of Cysteine Residues in GlpT, the Glycerol 3-Phosphate Transporter of *Escherichia coli*. *J. Bacteriol.* 185, 3863–3870.
- Fantini, J. and Barrantes, F.J. (2013). How cholesterol interacts with membrane proteins: An exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front. Physiol.* 4 FEB, 1–9.
- Fei, Y.J., Kanai, Y., Nussberger, S., and Ganapathy, V. (1994). Expression cloning of a mammalian proton-coupled oligopeptide transporter. *Nature* 368, 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. *Biochem. Biophys. Res. Commun.* 246, 39–44.
- Fei, Y.J., Liu, W., Prasad, P.D.D., and Kekuda, R. (1997). Identification of the Histidyl Residue Obligatory for the Catalytic Activity of the Human H<sup>+</sup>/Peptide Cotransporters PEPT1 and PEPT2. *Biochemistry* 2960, 452–460.
- Fei, Y.J., Nara, E., Liu, J.C., Boyd, C.A.R., Ganapathy, V., and Leibach, F.H. (1999). Preferential recognition of zwitterionic dipeptides as transportable substrates by the high-affinity peptide transporter PEPT2. *Biochim. Biophys. Acta.* 1418, 344–351.
- Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme [Influence of Configuration on the Action of Enzyme]. *Ber. Dt. Chem. Ges.* 27, 2985 – 2293.
- Flyvbjerg, H. and Petersen, H.G. (1989). Error estimates on averages of correlated data. *J. Chem. Phys.* 91, 461.
- Foley, D., Rajamanickam, J., Bailey, P.D., and Meredith, D. (2010). Bioavailability through PepT1: the role of computer modelling in intelligent drug design. *Curr Comp-Aided Drug Des* 6, 68–78.
- Forrest, L.R. (2013). (Pseudo-) Symmetrical Transport. *Science* (80-. ). 399, 399–402.
- Fowler, P., Geroult, S., Jha, S., Waksman, G., and Coveney, P. (2007). Rapid, accurate, and precise calculation of relative binding affinities for the SH2 domain using a computational grid. *J. Chem. Theory Comput.* 3, 1193–1202.
- Fowler, P., Orwick-Rydmark, M., Radestock, S., Solcan, N., Dijkman, P., Lyons, J., Kwok, J., Caffrey, M., Watts, A., Forrest, L., et al. (2015). Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Structure* 23, 290–301.
- Frankel, E.M. (1916). A Comparative Study of the Behavior of Purified Proteins Towards Proteolytic Enzymes. *J. Biol. Chem.* 26, 31–59.
- Friedberg, R. and Cameron, J.E. (1970). Test of the Monte Carlo Method: Fast Simulation of a Small Ising Lattice. *J. Chem. Phys.* 52, 6049–6058.

- Gaboriaud, C., Serre, L., Guy-Crotte, O., Forest, E., and Fontecilla-Camps, J.C. (1996). Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151. *J. Mol. Biol.* 259, 995–1010.
- Gadsby, D.C., Vergani, P., and Csanády, L. (2006). The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* 440, 477–483.
- Ganapathy, M.E., Brandsch, M., Prasad, P.D., Ganapathy, V., and Leibach, F.H. (1995). Differential recognition of beta-lactam antibiotics by intestinal and renal peptide transporters, PEPT 1 and PEPT 2. *J. Biol. Chem.* 270, 25672–25677.
- 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. *Biochem. Biophys. Res. Commun. Comm* 246, 470–5.
- Ganapathy, V., Balkovetz, D.F., Ganapathy, M.E., Mahesh, V.B., Devoe, L.D., and Leibach, F.H. (1987). Evidence for histidyl and carboxy groups at the active site of the human placental Na<sup>+</sup>-H<sup>+</sup> exchanger. *Biochem. J.* 245, 473–477.
- Ganapathy, V. and Leibach, F.H. (1983). Role of pH gradient and membrane potential in dipeptide transport in intestinal and renal brush-border membrane vesicles from the rabbit. *J. Biol. Chem.* 258, 14189–14192.
- Ganapathy, V., Mendicino, J.F., and Leibach, F.H. (1981). Transport of glycyl-L-proline into intestinal and renal brush border vesicles from rabbit. *J. Biol. Chem.* 256, 118–124.
- Ge, Y., Wu, J., Xia, Y., Yang, M., Xiao, J., and Yu, J. (2012). Molecular Dynamics Simulation of the Complex PBP-2x with Drug Cefuroxime to Explore the Drug Resistance Mechanism of *Streptococcus suis* R61. *PLoS One* 7, e35941.
- Gebauer, S. and Hartrodt, B. (2003). Three-dimensional quantitative structure-activity relationship analyses of peptide substrates of the mammalian H<sup>+</sup>/peptide cotransporter PEPT1. *J. Med. Chem.* 46, 5725–5734.
- Genheden, S. (2012). Are homology models sufficiently good for free-energy simulations? *J. Chem. Inf. Model.* 52, 3013–21.
- Gerl, M.J., Sampaio, J.L., Urban, S., Kalvodova, L., Verbavatz, J., Binnington, B., Lindemann, D., Lingwood, C.a., Shevchenko, A., Schroeder, C., et al. (2012). Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. *J. Cell Biol.* 196, 213–221.
- Giacomini, K.M., Huang, S.M., Tweedie, D.J., Benet, L.Z., Brouwer, K.L.R., Chu, X., Dahlin, A., Evers, R., Fischer, V., Hillgren, K.M., et al. (2010). Membrane transporters in drug development. *Nat. Rev. Drug Discov.* 9, 215–36.
- Goldberg, D.M., Campbell, R., and Roy, A.D. (1968). Binding of trypsin and chymotrypsin by human intestinal mucosa. *Biochim. Biophys. Acta* 167, 613–615.
- Goldfine, H. (1984). Bacterial membranes and lipid packing theory. *J. Lipid Res.* 25, 1501–1507.

- Gouaux, E. (2009). Review. The molecular logic of sodium-coupled neurotransmitter transporters. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 364, 149–154.
- Gouda, H., Kuntz, I.D., Case, D.A., and Kollman, P.A. (2003). Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* 68, 16–34.
- Groneberg, D.a., Döring, F., Theis, S., Nickolaus, M., Fischer, A., and Daniel, H. (2002). Peptide transport in the mammary gland: expression and distribution of PEPT2 mRNA and protein. *Am. J. Physiol. Endocrinol. Metab.* 282, E1172–E1179.
- 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. *Am. J. Pathol.* 158, 707–714.
- Guetton, F., Quistgaard, E.M., Raba, M., Moberg, P., Löw, C., and Nordlund, P. (2014). Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat. Struct. Mol. Biol.* 21, 728–731.
- Guetton, 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 Rep.* 14, 804–10.
- Guitierrez-de Teran, H. and Aqvist, J. (2012). Linear interaction energy: method and applications in drug design. In *Methods Mol Biol.*, volume 819, 305–23 (Springer New York).
- Hagenbuch, B. and Gui, C. (2008). Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. *Xenobiotica* 38, 778–801.
- Hagting, A., Knol, J., Hasemeier, B., Streutker, M.R., Fang, G., Poolman, B., and Konings, W.N. (1997). Amplified expression, purification and functional reconstitution of the dipeptide and tripeptide transport protein of *Lactococcus lactis*. *Euro. J. Biochem.* 247, 581–7.
- Halestrap, A.P. (2013). The SLC16 gene family – Structure, role and regulation in health and disease. *Mol. Asp. Med.* 34, 337–349.
- Han, H.K., Rhie, J.K., Oh, D.M., Saito, G., Hsu, C.P., Stewart, B.H., and Amidon, G.L. (1999). CHO/hPEPT1 cells overexpressing the human peptide transporter (hPEPT1) as an alternative in vitro model for peptidomimetic drugs. *J. Pharm. Sci.* 88, 347–350.
- Hansson, T., Marelius, J., and Aqvist, J. (1998). Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des.* 12, 27–35.
- Harder, D., Stolz, J., Casagrande, F., Obrdlik, P., Weitz, D., Fotiadis, D., and Daniel, H. (2008). DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of *Escherichia coli*. *FEBS J.* 275, 3290–3298.
- Hauser, M., Kauffman, S., Naider, F., and Becker, J.M. (2005). Substrate preference is altered by mutations in the fifth transmembrane domain of Ptr2p, the di/tri-peptide transporter of *Saccharomyces cerevisiae*. *Mol. Memb. Biol.* 22, 215–227.



- Heifets, A. and Lilien, R.H. (2010). LigAlign: Flexible ligand-based active site alignment and analysis. *J. Mol. Graph. Model.* *29*, 93–101.
- Hénin, J., Maigret, B., Tarek, M., Escrieut, C., Fourmy, D., and Chipot, C. (2006). Probing a model of a GPCR/ligand complex in an explicit membrane environment: the human cholecystokinin-1 receptor. *Biophys. J.* *90*, 1232–1240.
- Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* *18*, 1463–1472.
- Hess, B., Kutzner, C., and Spoel, D.V.D. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* *4*, 435–447.
- Hille, B. (2001). *Ion Channels of Excitable Membranes* (Sunderland MA. USA: Sinauer Associates Inc.).
- Himukai, M., Suzuki, Y., and Hoshi, T. (1978). Differences in characteristics between glycine and glycyglycine transport in guinea pig small intestine. *Jpn. J. Physiol.* *28*, 499–51.
- Hockney, R.W. (1970). The Potential Calculation and Some Applications. *Methods Comp. Phys.* *9*, 136–211.
- Hoover, W.G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* *31*, 1695–1697.
- Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* *51*, 69–82.
- Hruz, P.W. and Mueckler, M.M. (2000). Cysteine-scanning mutagenesis of transmembrane segment 11 of the GLUT1 facilitative glucose transporter. *Biochemistry* *39*, 9367–9372.
- Huang, X., Zheng, G., and Zhan, C. (2012). Microscopic binding of M5 muscarinic acetylcholine receptor with antagonists by homology modeling, molecular docking, and molecular dynamics simulation. *J. Phys. Chem.* *116*, 532–41.
- Huang, Y., Lemieux, M.J., Song, J., Auer, M., and Wang, D. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* (80-. ). *301*, 616–20.
- Humphrey, W. and Dalke, A. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* *15*, 33–38.
- Humphreys, D.D., Friesner, R.A., and Berne, B.J. (1994). A Multiple-Time-Step Molecular Dynamics Algorithm for Macromolecules. *J. Phys. Chem.* *98*, 6885–6892.
- Hunte, C., Screpanti, E., Venturi, M., Rimon, A., Padan, E., and Michel, H. (2005). Structure of a Na<sup>+</sup>/H<sup>+</sup> antiporter and insights into mechanism of action and regulation by pH. *Nature* *435*, 1197–1202.

- Ilka, K., Wollesky, C., Kottra, G., Hahn, M., Fischer, W., Zebisch, K., Neubert, R., Daniel, H., and Bandsch, M. (2008). Transport of Angiotensin-Converting Enzyme Inhibitors by H<sup>+</sup> / Peptide Transporters Revisited. *J. Pharmacol. Exp. Ther.* 327, 432–441.
- Inui, K., Tomita, Y., Katsura, T., Okano, T., Takano, M., and Hori, R. (1992). H<sup>+</sup> coupled active transport of bestatin via the dipeptide transport system in rabbit intestinal brush-border membranes. *J. Pharmacol. Exp. Ther.* 260, 482–6.
- Irie, M., Terada, T., Sawada, K., Saito, H., and Inui, K. (2001). Recognition and transport characteristics of nonpeptidic compounds by basolateral peptide transporter in Caco-2 cells. *J. Pharmacol. Exp. Ther.* 298, 711–717.
- Irvine, J.D., Takahashi, L., Lockhart, K., Cheong, J., Tolan, J.W., Selick, H.E., and Grove, J.R. (1999). MDCK (Madin-Darby canine kidney) cells: A tool for membrane permeability screening. *J. Pharm. Sci.* 88, 28–33.
- Ito, K., Hikida, A., Kawai, S., Lan, V.T.T., Motoyama, T., Kitagawa, S., Yoshikawa, Y., Kato, R., and Kawarasaki, Y. (2013). Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library. *Nat. Commun.* 4, 2502.
- Jack, D.L., Paulsen, I.T., and Saier, J. (2000). The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* 146, 1797–1814.
- Jakalian, A., Jack, D.B., and Bayly, C.I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comp. Chem.* 23, 1623–41.
- Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* 211, 969–970.
- Jensen, J.M., Ismat, F., Szakonyi, G., Rahman, M., and Mirza, O. (2012). Probing the Putative Active Site of YjdL: An Unusual Proton-Coupled Oligopeptide Transporter from *E. coli*. *PLoS One* 7.
- Jiang, D., Zhao, Y., and Wang, X. (2013). Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. *Proc. Natl. Acad. Sci. U.S.A.* 110, 14664–14669.
- Jones, S. and Thornton, J.M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13–20.
- Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926.
- Kar, P., Lipowsky, R., and Knecht, V. (2011). Importance of polar solvation for cross-reactivity of antibody and its variants with steroids. *J. Phys. Chem. B* 115, 7661–7669.
- Karlin, A. and Akabas, M.H. (1998). Substituted-cysteine accessibility method. *Methods Enzym.* 293, 123–145.

- Kato, M., Maegawa, H., Okano, T., Inui, K., and Hori, R. (1989). Effect of various chemical modifiers on H<sup>+</sup> coupled transport of cephadrine via dipeptide carriers in rabbit intestinal brush-border membranes: role of histidine residues. *J. Pharmacol. Exp. Ther* 251, 745–749.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181, 662–666.
- King, L.S., Kozono, D., and Agre, P. (2004). From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell. Biol.* 5, 687–698.
- Knu, I., Kottra, G., Fischer, W., Daniel, H., and Brandsch, M. (2009). High-Affinity Interaction of Sartans with H<sup>+</sup> / Peptide Transporters. *Drug Metab. Dispos.* 37, 143–149.
- Knütter, I., Hartrodt, B., Theis, S., Foltz, M., Rastetter, M., Daniel, H., Neubert, K., and Brandsch, M. (2004). Analysis of the transport properties of side chain modified dipeptides at the mammalian peptide transporter PEPT1. *Eur. J. Pharm. Sci.* 21, 61–67.
- Koch, A.L. and Silver, S. (2005). The first cell. *Adv. Microb. Biol.* 50, 227–59.
- Koepsell, H. (2013). The SLC22 family with transporters of organic cations, anions and zwitterions. *Mol. Asp. Med.* 34, 413–435.
- Krulwich, T.A., Sachs, G., and Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* 9, 330–343.
- Kuhn, B., Gerber, P., Schulz-Gasch, T., and Stahl, M. (2005). Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* 48, 4040–8.
- Kulkarni, A., Davies, D.L., Links, J.S., Patel, L.N., Lee, V.H.L., and Haworth, I.S. (2007). A charge pair interaction between Arg282 in transmembrane segment 7 and Asp341 in transmembrane segment 8 of hPepT1. *Pharm. Res.* 24, 66–72.
- Kulkarni, A., Haworth, I.S., and Lee, V.H.L. (2003a). Transmembrane segment 5 of the dipeptide transporter hPepT1 forms a part of the substrate translocation pathway. *Biochem. Biophys. Res. Commun.* 306, 177–185.
- Kulkarni, A., Haworth, I.S., Uchiyama, T., and Lee, V.H.L. (2003b). Analysis of transmembrane segment 7 of the dipeptide transporter hPepT1 by cysteine-scanning mutagenesis. *J. Biol. Chem.* 278, 51833–40.
- Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.
- Laganowsky, A., Reading, E., Allison, T.M., Ulmschneider, M.B., Degiacomi, M.T., Baldwin, A.J., and Robinson, C.V. (2014). Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* 510, 172–175.
- Law, C.J., Maloney, P.C., and Wang, D.N. (2008). Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* 62, 289–305.

- Lee, C., Kang, H.J., von Ballmoos, C., Newstead, S., Uzdavinyis, P., Dotson, D.L., Iwata, S., Beckstein, O., Cameron, A.D., and Drew, D. (2013). A two-domain elevator mechanism for sodium/proton antiport. *Nature* .
- Lee, E.H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery through the computational microscope. *Structure* 17, 1295–306.
- Lee, V.H., Chu, C., Mahlin, E.D., Basu, S.K., Ann, D.K., Bolger, M.B., Haworth, I.S., Yeung, A.K., Wu, S.K., Hamm-Alvarez, S., et al. (1999). Biopharmaceutics of transmucosal peptide and protein drug administration: role of transport mechanisms with a focus on the involvement of PepT1. *J. Control. Release* 62, 129–40.
- Lehmann, E.L. and D'Abrera, H.J.M. (1998). *Nonparametrics: Statistical Methods Based on Ranks, Revised* (Englewood Cliffs, NJ: Prentice-Hall).
- Léran, S., Varala, K., Boyer, J.C., Chiurazzi, M., Crawford, N., Daniel-Vedele, F., David, L., Dickstein, R., Fernandez, E., Forde, B., et al. (2014). A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci.* 19, 5–9.
- Levitt, M. (2001). The birth of computational structural biology. *Nat. Struct. Mol. Biol.* 8, 392–393.
- Li, J., Tamura, K., Lee, C.P., Smith, P.L., Borchardt, R.T., and Hidalgo, I.J. (1998). Structure-affinity relationships of Val-Val and Val-Val-Val stereoisomers with the apical oligopeptide transporter in human intestinal Caco-2 cells. *J. Drug Target.* 5, 317–327.
- Liang, R., Fei, Y.J., Prasad, P.D.D., Ramamoorthy, S., Han, H., Yang-Feng, T.L.L., Hediger, M.A., Ganapathy, V., and Leibach, F.H.H. (1995). Human intestinal H<sup>+</sup>/peptide cotransporter. *J. Biol. Chem.* 270, 6456.
- Liederer, B.M. and Borchardt, R.T. (2005). Enzymes Involved in the Bioconversion of Ester-Based Prodrugs. *J. Pharm. Sci.* 95, 1177–1195.
- Lin, H. and King, N. (2007). Demonstration of functional dipeptide transport with expression of PEPT2 in guinea pig cardiomyocytes. *Pflugers Arch.* 453, 915–922.
- Linder, M., Ranganathan, A., and Brinck, T. (2013). “Adapted Linear Interaction Energy”: A Structure-Based LIE Parametrization for Fast Prediction of Protein–Ligand Affinities. *J. Chem. Theory Comput.* 9, 1230–1239.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O., and Shaw, D.E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950–8.
- Links, J.S., Kulkarni, A., Davies, D.L., Lee, V.H.L., and Haworth, I.S. (2007). Cysteine scanning of transmembrane domain three of the human dipeptide transporter: implications for substrate transport. *J. Drug Target.* 15, 218–25.
- Lister, N., Sykes, a.P., Bailey, P.D., Boyd, C.A., and Bronk, J.R. (1995). Dipeptide transport and hydrolysis in isolated loops of rat small intestine: effects of stereospecificity. *J. Physiol.* 484, 173–82.

- Liu, W., Liang, R., Ramamoorthy, S., and Fei, Y.J. (1995). Molecular cloning of PEPT 2, a new member of the H<sup>+</sup>/peptide cotransporter family, from human kidney. *Biochim. Biophys. Acta* 1235, 461–466.
- Lodish, H., Berk, A., and SL, Z. (2000). Active Transport by ATP-Powered Pumps. In *Mol. Cell Biol.*, chapter Section 15 (New York: W.H.Freeman), 4th edition.
- Lolkema, J.S., Carrasco, N., and Kaback, H.R. (1991). Kinetic analysis of lactose exchange in proteoliposomes reconstituted with purified lac permease. *Biochemistry* 30, 1284–1290.
- Lovell, S. and Davis, I. (2003). Structure Validation by C-alpha Geometry : phi, psi and C-beta Deviation. *Proteins* 450, 437–450.
- Lu, Y., Wang, R., Yang, C.Y., and Wang, S. (2007). Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* 47, 668–675.
- Luccarelli, J., Michel, J., Tirado-Rives, J., and Jorgensen, W.L. (2010). Effects of Water Placement on Predictions of Binding Affinities for p38 $\alpha$  MAP Kinase Inhibitors. *J. Chem. Theory Comput.* 6, 3850–3856.
- Luckey, M. (2008). *Membrane Structural Biology with Biochemical and Biophysical Foundations* (Cambridge University Press).
- Luckner, P. and Brandsch, M. (2005). Interaction of 31 beta-lactam antibiotics with the H<sup>+</sup>/peptide symporter PEPT2: analysis of affinity constants and comparison with PEPT1. *FEBS J.* 59, 17–24.
- Luzhkov, V.B. and Aqvist, J. (2001). Mechanisms of tetraethylammonium ion block in the KcsA potassium channel. *FEBS Lett.* 495, 191–6.
- Lyons, J.A., Parker, J.L., Solcan, N., Brinth, A., Li, D., Shah, S.T., Caffrey, M., and Newstead, S. (2014). Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO Rep.* 1–8.
- Malle, E., Zhou, H., Neuhold, J., Spitzenberger, B., Klepsch, F., Pollak, T., Bergner, O., Ecker, G.F., and Stolt-Bergner, P.C. (2011). Random mutagenesis of the prokaryotic peptide transporter YdgR identifies potential periplasmic gating residues. *J. Biol. Chem.* 286, 23121–23131.
- Mandell, J.G., Roberts, V.a., Pique, M.E., Kotlovyy, V., Mitchell, J.C., Nelson, E., Tsigelny, I., and Ten Eyck, L.F. (2001). Protein docking using continuum electrostatics and geometric fit. *Protein Eng.* 14, 105–113.
- McKenna, C.E., Kashemirov, B.A., Eriksson, U., Amidon, G.L., Kish, P.E., Mitchell, S., Kim, J.S., and Hilfinger, J.M. (2005). Cidofovir peptide conjugates as prodrugs. *J. Organomet. Chem.* 690, 2673–2678.
- Meredith, D. (2004). Site-directed Mutation of Arginine 282 to Glutamate Uncouples the Movement of Peptides and Protons by the Rabbit Proton-peptide Cotransporter PepT1. *J. Biol. Chem.* 279, 15795–15798.

- Meredith, D. and Price, R.A. (2006). Molecular modeling of PepT1—towards a structure. *J. Membr. Biol.* 213, 79–88.
- Meredith, D., Temple, C.S., Guha, N., Sword, C.J., Boyd, C.A., Collier, I.D., Morgan, K.M., and Bailey, P.D. (2000). Modified amino acids and peptides as substrates for the intestinal peptide transporter PepT1. *Eur. J. Biochem.* 267, 3723–3728.
- Miller, C. (2006). ClC chloride channels viewed through a transporter lens. *Nature* 440, 484–489.
- Miller III, B., Jr, T.M., Swails, J.M., Homeyer, N., Gohlke, H., and Roitberg, A.E. (2012). MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* 8, 3314–3321.
- Moradi, M. and Tajkhorshid, E. (2015). Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems. *J. Chem. Theory Comput.* 10, 2866–2880.
- Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S., and Stewart, P.D.S. (2014). Membrane protein structure determination - the next generation. *Biochim. Biophys. Acta* 1838, 78–87.
- Nässl, A. and Rubio-Aliaga, I. (2011). Amino acid absorption and homeostasis in mice lacking the intestinal peptide transporter PEPT1. *Am. J. Physiol. Gastrointest. Liver Physiol.* 301, 128–137.
- Newman, M.J. and Wilson, T.H. (1980). Solubilisation and Reconstitution of the Lactose Transport System. *Bio. Chem.* .
- Newstead, S., Drew, D., Cameron, A.D., Postis, V.L.G., Xia, X., Fowler, P.W., Ingram, J.C., Carpenter, E.P., Sansom, M.S.P., McPherson, M.J., et al. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* 30, 417–426.
- Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52, 255–268.
- Oehme, D.P., Brownlee, R.T.C., and Wilson, D.J.D. (2012). Effect of atomic charge, solvation, entropy, and ligand protonation state on MM-PB(GB)SA binding energies of HIV protease. *J. Comput. Chem.* 33, 2566–80.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34.
- Onufriev, A., Case, D.A., and Bashford, D. (2002). Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comp. Chem.* 23, 1297–1304.
- Oostenbrink, C., Villa, A., Mark, A.E., and van Gunsteren, W.F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25, 1656–76.

- Orry, A.J.W. and Abagyan, R. (2012). *Methods in Molecular Biology: Homology Modelling*, volume 1.
- Osterberg, F. and Aqvist, J. (2005). Exploring blocker binding to a homology model of the open hERG K<sup>+</sup> channel using docking and molecular dynamics methods. *FEBS Lett.* 579, 2939–44.
- Pao, S., Paulsen, I., and Saier, M. (1998). Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* 62.
- Parker, J.L., Mindell, J.A., and Newstead, S. (2014). Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *Elife* 1–13.
- Parker, J.L. and Newstead, S. (2014). Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature* 507, 68–72.
- Parrinello, M. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182.
- Parrinello, M. and Rahman, A. (1981). Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* 52, 7182–7190.
- Pearson, K. (1895). Note on Regression and Inheritance in the Case of Two Parents. *Proc. R. Soc. London* 58, 240–242.
- 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., et al. (2013). Crystal structure of a eukaryotic phosphate transporter. *Nature* 496, 533–536.
- Pedretti, A., Luca, L.D., Marconi, C., Negrisoli, G., Aldini, G., and Vistoli, G. (2008). Modeling of the Intestinal Peptide Transporter hPepT1 and Analysis of Its Transport Capacities by Docking and Pharmacophore Mapping. *ChemMedChem* 3, 1913–1921.
- Pedretti, A., Luca, L.D., Marconi, C., Regazzoni, L., Aldini, G., and Vistoli, G. (2011). Fragmental modeling of hPepT2 and analysis of its binding features by docking studies and pharmacophore mapping. *Bioorg. Med. Chem.* 19, 4544–4551.
- Penmatsa, A., Wang, K.H., and Gouaux, E. (2013). X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* 503, 85–90.
- Phillips, G.B. and Dodge, J.T. (1967). Composition of phospholipids and of phospholipid fatty acids in human red cells. *J. Lipid Res.* 8, 676–681.
- Pieri, M., Gan, C., Bailey, P., and Meredith, D. (2009). The transmembrane tyrosines Y56, Y91 and Y167 play important roles in determining the affinity and transport rate of the rabbit proton-coupled peptide transporter PepT1. *Int J Biochem Cell Biol.* 41, 2204–13.
- Pieri, M., Hall, D., Price, R., Bailey, P.D., and Meredith, D. (2008). Site-directed mutagenesis of Arginine282 suggests how protons and peptides are co-transported by rabbit PepT1. *Int. J. Biochem. Cell Biol.* 40, 721–730.
- Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95–99.

- Rashid, M.H., Heinzelmann, G., Huq, R., Tajhya, R.B., Chang, S.C., Chhabra, S., Pennington, M.W., Beeton, C., Norton, R.S., and Kuyucak, S. (2013). A potent and selective peptide blocker of the Kv1.3 channel: Prediction from free-energy simulations and experimental confirmation. *PLoS One* 8.
- Rautio, J., Kumpulainen, H., Heimbach, T., Oliyai, R., Oh, D., Järvinen, T., and Savolainen, J. (2008). Prodrugs: design and clinical applications. *Nat. Rev. Drug Discov.* 7, 255–70.
- Rizzo, R.C., Wang, D.P., Tirado-rives, J., and Jorgensen, W.L. (2000). Validation of a Model for the Complex of HIV-1 Reverse Transcriptase with Sustiva through Computation of Resistance Profiles. *J. Am. Chem. Soc.* 12898–12900.
- Ross, G., Morris, G., and Biggin, P. (2013). One size does not fit all: The limits of structure-based models in drug discovery. *J. Chem. Theory Comput.* 9, 4266–4274.
- Ross, G.a., Morris, G.M., and Biggin, P.C. (2012). Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* 7, 1–13.
- Rostovtsev, V.V., Green, L.G., Fokin, V.V., and Sharpless, K.B. (2002). A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed* 41, 2596–2599.
- Rubio-Aliaga, I. and Daniel, H. (2002). Mammalian peptide transporters as targets for drug delivery. *Trends Pharmacol. Sci.* 23, 434–440.
- Rubio-Aliaga, I., Frey, I., Boll, M., Groneberg, D.A., Eichinger, H.M., Balling, R., and Daniel, H. (2003). Targeted disruption of the peptide transporter Pept2 gene in mice defines its physiological role in the kidney. *Mol. Cell Biol.* 23, 3247–3252.
- Ryan, R.M. and Mindell, J.A. (2007). The uncoupled chloride conductance of a bacterial glutamate transporter homolog. *Nat. Struct. Mol. Biol.* 14, 365–371.
- Sahin-Tóth, M. and Kaback, H.R. (1993). Cysteine scanning mutagenesis of putative transmembrane helices IX and X in the lactose permease of *Escherichia coli*. *Protein Sci.* 2, 1024–1033.
- Said, H.M. and Mohammadkhani, R. (1993). Folate transport in intestinal brush border membrane: involvement of essential histidine residue(s). *Biochem. J.* 290, 237–240.
- Saier, M.H. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* 64, 354–411.
- Sali, A. and Blundell, T. (1994). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Santos, R., Hritz, J., and Oostenbrink, C. (2010). Role of Water in Molecular Docking Simulations of Cytochrome P450 2D6. *J. Chem. Inf. Model.* 50, 146–154.
- Schames, J.R., Henchman, R.H., Siegel, J.S., Sottriffer, C.A., Ni, H., and McCammon, J.A. (2004). Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* 47, 1879–81.



- Schlessinger, A., Geier, E., Fan, H., Irwin, J.J., Shoichet, B.K., Giacomini, K.M., and Sali, A. (2011). Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 15810–15815.
- Schrum, J.P., Zhu, T.F., and Szostak, J.W. (2010). The Origins of Cellular Life. *Cold Spring Harb. Perspect. Biol.* *2*, a002212–a002212.
- 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. *Am. J. Physiol.* *276*, F658–F665.
- Shen, M., Devos, D., Melo, F., and Sali, A. (2006). A composite score for predicting errors in protein structure models. *Protein Sci.* *15*, 1653–1666.
- Shirts, M.R., Mobley, D.L., and Brown, S.P. (2010). Free-energy calculations in structure-based drug design. In *Drug Des. Struct. Ligand-Based Approaches*, 61–86 (Cambridge University Press).
- Shirts, M.R., Pitner, J.W., Swope, W.C., and Pande, V.S. (2003). Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.* *119*, 5740.
- Shitara, Y. and Sugiyama, Y. (2002). Inhibition on the transporter-mediated hepatic uptake as a mechanism of drug-drug interaction. *J. Pharmacol. Exp. Ther* *301*, 610–616.
- Shu, C., Shen, H., Teuscher, N.S., Lorenzi, P.J., Keep, R.F., and Smith, D.E. (2002). Role of PEPT2 in peptide/mimetic trafficking at the blood-cerebrospinal fluid barrier: studies in rat choroid plexus epithelial cells in primary culture. *J. Pharmacol. Exp. Ther* *301*, 820–829.
- Smith, D.E., Cl  men  on, B., and Hediger, M.A. (2013). Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol. Asp. Med.* *34*, 323–36.
- Sodani, K. and Patel, A. (2012). Multidrug resistance associated proteins in multidrug resistance. *Chin. J. Cancer.* *31*, 58–73.
- Solcan, N. (2013). Biochemical and biophysical studies of the prokaryotic proton dependent oligopeptide transporters. Ph.D. thesis, University of Oxford.
- 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. *EMBO J.* 1–11.
- Song, X., Lorenzi, P.L., Landowski, C.P., Vig, B.S., Hilfinger, J.M., and Amidon, G.L. (2005). Amino acid ester prodrugs of the anticancer agent gemcitabine: Synthesis, bio-conversion, metabolic bioevasion, and hPEPT1-mediated transport. *Mol. Pharm.* *2*, 157–167.
- Spiliotopoulos, D., Spitaleri, A., and Musco, G. (2012). Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study. *PLoS One* *7*, e46902.

- Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A., and Case, D.A. (1998). Continuum Solvent Studies of the Stability of DNA , RNA , and Phosphoramidate - DNA Helices. *J. Am. Chem. Soc.* *120*, 9401–9409.
- Steel, A., Nussberger, S., Romero, M.F., Boron, W.F., Boyd, C.A., and Hediger, M.a. (1997). Stoichiometry and pH dependence of the rabbit proton-dependent oligopeptide transporter PepT1. *J. Physiol.* *498* ( Pt 3, 563–569.
- Stelzl, L.S., Fowler, P.W., Sansom, M.S.P., and Beckstein, O. (2014). Flexible gates generate occluded intermediates in the transport cycle of LacY. *J. Mol. Biol.* *426*, 735–751.
- Sugano, K., Kansy, M., Artursson, P., Avdeef, A., Bendels, S., Di, L., Ecker, G.F., Faller, B., Fischer, H., Gerebtzoff, G., et al. (2010). Coexistence of passive and carrier-mediated processes in drug transport. *Nat. Rev. Drug Discov.* *9*, 597–614.
- Sugawara, M., Huang, W., Fei, Y.J., Leibach, F.H., Ganapathy, V., and Ganapathy, M.E. (2000). Transport of valganciclovir, a ganciclovir prodrug, via peptide transporters PEPT1 and PEPT2. *J. Pharm. Sci.* *89*, 781–9.
- Sun, J., Bankston, J.R., Payandeh, J., Hinds, T.R., Zagotta, W.N., and Zheng, N. (2014). Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* *507*, 73–77.
- Tamai, I.K., Nakanishi, T.A., Nakahara, H.A., Sai, Y.O., Ganapathy, V.A., Leibach, F.R.H., and Tsuji, A.K. (1998). Improvement of L -Dopa Absorption by Dipeptidyl Derivation , Utilizing Peptide Transporter PepT1. *J. Pharm. Sci.* *87*, 1542–1546.
- Terada, T., Irie, M., Okuda, M., and Inui, K. (2004). Genetic variant Arg57His in human H<sup>+</sup>/peptide cotransporter 2 causes a complete loss of transport function. *Biochem. Biophys. Res. Commun.* *316*, 416–20.
- Terada, T., Saito, H., Mukai, M., and Inui, K.I. (1996). Identification of the histidine residues involved in substrate recognition by a rat H<sup>+</sup>/peptide cotransporter, PEPT1. *FEBS Lett.* *394*, 196–200.
- Terada, T., Saito, H., Mukai, M., and Inui, K.i. (1997). Recognition of beta-lactam antibiotics by rat peptide transporters, PepT1 and PepT2, in LLC-PK1 cells. *Am. J. Physiol.* *273*, F706–F711.
- Terada, T., Saito, H., Sawada, K., Hashimoto, Y., and Inui, K. (2000a). N-terminal halves of rat H<sup>+</sup>/peptide transporters are responsible for their substrate recognition. *Pharm. Res.* *17*, 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 Arch.* *440*, 679–84.
- The 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. *Nature* *467*, 1061–73.
- Theis, S., Hartrodt, B., Kottra, G., Neubert, K., and Daniel, H. (2002). Defining minimal structural features in substrates of the H(+)/peptide cotransporter PEPT2 using novel amino acid and dipeptide derivatives. *Mol. Pharm.* *61*, 214–221.

- Thilagavathi, R. and Mancera, R.L. (2010). Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Comput. Sci.* *50*, 415–421.
- Thomas, V., Bhattachar, S., and Hitchingham, L. (2006). The road map to oral bioavailability : an industrial perspective. *Expert Opin Drug Metab Toxicol.* *2*, 591–608.
- Thwaites, D.T., Hirst, B.H., and Simmons, N.L. (1994). Substrate specificity of the di/tripeptide transporter in human intestinal epithelia (Caco-2): identification of substrates that undergo H(+)-coupled absorption. *Br. J. Pharmacol.* *113*, 1050–1056.
- Totrov, M. and Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* *29*, 215–220.
- Trott, O. and Olson, A.J. (2010). Software News and Update AutoDock Vina : Improving the Speed and Accuracy of Docking with a New Scoring Function , Efficient Optimization , and Multithreading. *J. Comput. Chem.* *31*, 455–461.
- Tsuda, M., Terada, T., Irie, M., Katsura, T., Niida, A., Tomita, K., Fujii, N., and Inui, K.i. (2006). Transport characteristics of a novel peptide transporter 1 substrate, antihypertensive drug midodrine, and its amino acid derivatives. *J. Pharmacol. Exp. Ther.* *318*, 455–460.
- Tsume, Y., Vig, B.S., Sun, J., Landowski, C.P., Hilfinger, J.M., Ramachandran, C., and Amidon, G.L. (2008). Enhanced Absorption and Growth Inhibition with Amino Acid Monoester Prodrugs of Floxuridine by Targeting hPEPT1 Transporters. *Molecules* *13*, 1441–1454.
- Uchiyama, T., Kulkarni, A.A., Davies, D.L., and Lee, V.H.L. (2003). Biophysical evidence for His57 as a proton-binding site in the mammalian intestinal transporter hPepT1. *Pharm. Res.* *20*, 1911–6.
- U.S. Department of Health and Human Services Food and Drug Administration, (CDER), C.f.D.E., and Research (2012). Guidance for industry. Drug interaction studies study design, data analysis, implications for dosing, and labeling recommendations. *Clin. Pharm.* *79*.
- Uysal, S., Cuello, L.G., Cortes, D.M., Koide, S., Kossiakoff, A.A., and Perozo, E. (2011). Mechanism of activation gating in the full-length KcsA K<sup>+</sup> channel. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 11896–11899.
- van de Waterbeemd, H. and Gifford, E. (2003). ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* *2*, 192–204.
- van De Waterbeemd, H., Smith, D.A., Beaumont, K., and Walker, D.K. (2001). Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* *44*, 1313–33.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., and Darian, E. (2010). CHARMM General Force Field : A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* *31*, 671–690.

- Veruki, M.L., Morkve, S.H., and Hartveit, E. (2006). Activation of a presynaptic glutamate transporter regulates synaptic transmission through electrical signaling. *Nat. Neurosci.* 9, 1388–1396.
- Vieira-Pires, R.S., Szollosi, A., and Morais-Cabral, J.H. (2013). The structure of the KtrAB potassium transporter. *Nature* 496, 323–8.
- Vig, B.S., Stouch, T.R., Timoszyk, J.K., Quan, Y., Wall, D., Smith, R.L., and Faria, T.N. (2006). Human PEPT1 pharmacophore distinguishes between dipeptide transport and binding. *J. Med. Chem.* 49, 3636–44.
- Vlaming, M.L.H., Lagas, J.S., and Schinkel, A.H. (2009). Physiological and pharmacological roles of ABCG2 (BCRP): Recent findings in Abcg2 knockout mice. *Adv. Drug Deliv. Rev.* 61, 14–25.
- Volpe, D.A. (2011). Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. *Futur. Med. Chem.* 3, 2063–77.
- Vriend, G. (1990). WHAT IF : A molecular modeling and drug design program. *J. Mol. Graph.* 8, 52–56.
- Wang, J., Deng, Y., and Roux, B. (2006a). Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* 91, 2798–814.
- Wang, J., Wang, W., Kollman, P.A., and Case, D.A. (2006b). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–60.
- Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P., and Case, D. (2004). Development and testing of a general amber force field. *J. Comp. Chem.* 25, 1157–74.
- Wang, L.H., Schultz, M., Weller, S., Smiley, M.L., and Blum, M.R. (1996). Pharmacokinetics and safety of multiple-dose valaciclovir in geriatric volunteers with and without concomitant diuretic therapy. *Antimicrob. Agents Chemother* 40, 80–85.
- Wang, R., Fang, X., Lu, Y., Yang, C.Y., and Wang, S. (2005). The PDBbind database: Methodologies and updates. *J. Med. Chem.* 48, 4111–4119.
- Wang, R., Lai, L., and Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* 16, 11–26.
- Wang, W., Wang, J., and Kollman, P.A. (1999). What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins* 34, 395–402.
- Warshel, A. and Russell, S.T. (1984). Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* 17, 283–422.
- 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, 1189–91.

- Webb, K.E., Matthews, J.C., Dirienzos, D.B., and DiRienzo, D.B. (1992). Peptide Absorption: A Review of Current Concepts and Future Perspectives. *J. Anim. Sci.* 70, 3248–3257.
- Weitz, D., Harder, D., Casagrande, F., Fotiadis, D., Obrdlik, P., Kelety, B., and Daniel, H. (2007). Functional and structural characterization of a prokaryotic peptide transporter with features similar to mammalian PEPT1. *J. Biol. Chem.* 282, 2832–9.
- Wenzel, U., Gebert, I., Weintraut, H., Weber, W.M., Clauss, W., and Daniel, H. (1996). Transport characteristics of differently charged cephalosporin antibiotics in oocytes expressing the cloned intestinal peptide transporter PepT1 and in human intestinal Caco-2 cells. *J. Pharmacol. Exp. Ther* 277, 831–839.
- Wolf, M.G., Hoeffling, M., Aponte-santamaría, C., Grubmüller, H., and Groenhof, G. (2010). *g\_membed* : Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *J. Comp. Chem.* 31, 2169–2174.
- Wright, D.W., Hall, B.A., Kenway, O.A., Jha, S., and Coveney, P.V. (2013). Computing Clinically Relevant Binding Free Energies of HIV1 Protease Inhibitors. *J. Chem. Theory Comput.* 10, 1228–1241.
- 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. *Pharm. Res.* 26, 2358–66.
- 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. *J. Membr. Biol.* 238, 43–9.
- Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., and Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na<sup>+</sup>/Cl<sup>-</sup>-dependent neurotransmitter transporters. *Nature* 437, 215–223.
- Yang, W., Bitetti-Putzer, R., and Karplus, M. (2004). Free energy simulations: use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.* 120, 2618–28.
- Yang, W., Gao, Y.Q., Cui, Q., Ma, J., and Karplus, M. (2003). The missing link between thermodynamics and structure in F1-ATPase. *Proc. Natl. Acad. Sci. U.S.A.* 100, 874–879.
- Yernool, D., Boudker, O., Jin, Y., and Gouaux, E. (2004). Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* 431, 811–818.
- Yeung, A.K., Basu, S.K., Wu, S.K., Chu, C., Okamoto, C.T., Hamm-Alvarez, S.F., von Grafenstein, H., Shen, W.C., Kim, K.J., Bolger, M.B., et al. (1998). Molecular identification of a role for tyrosine 167 in the function of the human intestinal proton- coupled dipeptide transporter (hPepT1). *Biochem. Biophys. Res. Com.* 250, 103–107.
- Yuriev, E. and Ramsland, P.A. (2013). Latest developments in molecular docking: 2010-2011 in review. *J. Mol. Recognit.* 26, 215–239.

- Zavodszky, M.I. and Kuhn, L.A. (2005). Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci.* *14*, 1104–1114.
- Zhang, C., Vasmatzis, G., Cornette, J.L., and DeLisi, C. (1997). Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* *267*, 707–726.
- Zhang, L., Zhang, L., Luo, T., Zhou, J., Sun, L., and Xu, Y. (2012). Synthesis and evaluation of a dipeptide-drug conjugate library as substrates for PEPT1. *ACS Comb. Sci.* *14*, 108–114.
- Zhou, S.F. (2008). Structure, function and regulation of P-glycoprotein and its clinical relevance in drug disposition. *Xenobiotica* *38*, 802–832.

## **Appendix A**

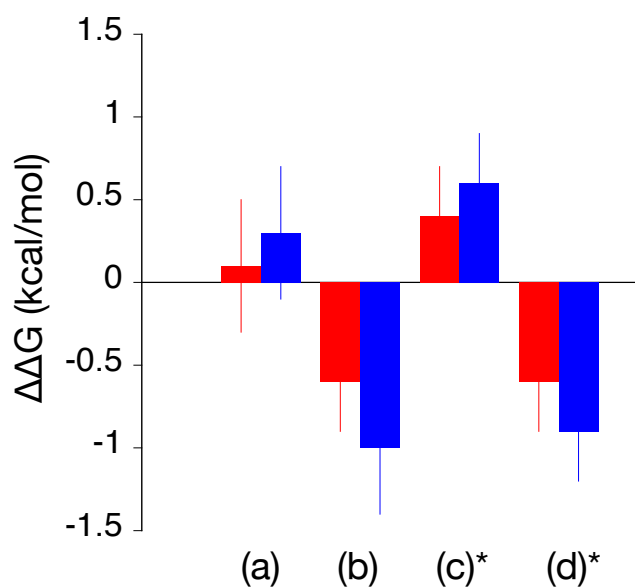
# **Repeat simulations for thermodynamic integration**

Several alchemical transformation MD simulations in Chapter 3 were repeated to demonstrate the reproducibility of the TI results. These include the transformation of AlaPhe to AlaAla and AlaPhe to PhePhe. Reversed transformations were conducted for two cases: AlaTyr to AlaPhe and PhePhe to AlaPhe.

Table A.1: TI results for replicate alchemical transformations of (a) AlaPhe to AlaAla, (b) AlaPhe to PhePhe, (c) AlaTyr to AlaPhe, and (d) PhePhe to AlaPhe. All values are quoted in kcal/mol with a precision of 0.1 kcal/mol, and the error values represent the error of integration.

| (a)   |                           |                             |                  |
|-------|---------------------------|-----------------------------|------------------|
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta\Delta G$ |
| 1     | $147.1 \pm 0.3$           | $144.5 \pm 0.1$             | $2.6 \pm 0.3$    |
| 2     | $-2.3 \pm 0.1$            | $-4.0 \pm 0.1$              | $1.7 \pm 0.1$    |
| 3     | $-140.8 \pm 0.3$          | $-144.8 \pm 0.1$            | $-4.0 \pm 0.1$   |
| Total |                           |                             | $0.3 \pm 0.4$    |
| (b)   |                           |                             |                  |
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta\Delta G$ |
| 1     | $81.5 \pm 0.3$            | $67.1 \pm 0.1$              | $14.4 \pm 0.3$   |
| 2     | $3.5 \pm 0.2$             | $5.1 \pm 0.1$               | $-1.6 \pm 0.2$   |
| 3     | $-89.8 \pm 0.2$           | $-76.0 \pm 0.1$             | $-13.8 \pm 0.1$  |
| Total |                           |                             | $-1.0 \pm 0.4$   |
| (c)   |                           |                             |                  |
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta\Delta G$ |
| 1     | $84.1 \pm 0.2$            | $91.0 \pm 0.2$              | $-6.9 \pm 0.2$   |
| 2     | $0.9 \pm 0.0$             | $0.5 \pm 0.1$               | $0.4 \pm 0.1$    |
| 3     | $-62.3 \pm 0.1$           | $-68.2 \pm 0.1$             | $5.9 \pm 0.1$    |
| Total |                           |                             | $-0.6 \pm 0.3$   |
| (d)   |                           |                             |                  |
| Step  | $\Delta G_{\text{bound}}$ | $\Delta G_{\text{unbound}}$ | $\Delta\Delta G$ |
| 1     | $-84.2 \pm 0.3$           | $-68.9 \pm 0.1$             | $-15.2 \pm 0.2$  |
| 2     | $-2.5 \pm 0.2$            | $-3.5 \pm 0.1$              | $1.0 \pm 0.2$    |
| 3     | $86.5 \pm 0.2$            | $71.4 \pm 0.1$              | $15.1 \pm 0.1$   |
| Total |                           |                             | $0.9 \pm 0.3$    |





Figs. A.1: TI results for repeat alchemical MD simulations (blue) compared to that for the original simulations (red) as shown in Chapter 3 (Table 3.5). (a) Transformation of AlaPhe to AlaAla, (b) AlaPhe to PhePhe, (c) reversed transformation of AlaTyr to AlaPhe, and (d) reverse transformation of PhePhe to AlaPhe. Results for reverse transformations are multiplied by -1 (marked by asterisk) so the magnitude of  $\Delta\Delta G$  can be compared to the original simulations. Error bars indicate the error of integration.

## Appendix B

### Sequence alignment for homology modelling

The sequence alignment used to build the homology model of human PepT1 and PepT2 in Chapter 5. Gating residues on the N-terminal and C-terminal bundles of helices are indicated in blue and red, respectively. Residues lining the binding site are highlighted by the green circles. The secondary structure of the human PepT1 homology model is drawn on the top of the alignment. For mouse PepT1 and rat PepT2, only the sequence of the extracellular domain (ECD) is shown. Conserved residues are highlighted in light blue and similar residues in pink. Uniprot IDs: PepT<sub>So</sub> (Q8EKT7), HsPepT1 (P46059), HsPepT2 (Q16348), MmPepT1 (Q9JIP7) and RnPepT1 (Q63424).

|                    |                                                   |             |    |
|--------------------|---------------------------------------------------|-------------|----|
|                    |                                                   | H1          |    |
|                    |                                                   | <i>oooo</i> |    |
| PepT <sub>So</sub> | .....PKWPRQIPYII                                  |             | 18 |
| HsPepT1            | MGMSK.....SHSFFGYPLSIFFIV                         |             | 20 |
| HsPepT2            | MNPFQKNESKETLFSPVSIEEVPPRPPSPKKPSPTICGSNYPLSIAFIV |             | 50 |
| MmPepT1            | .....                                             |             | 0  |
| RnPepT2            | .....                                             |             | 0  |

|                    | H1                                                      | H2 |    |
|--------------------|---------------------------------------------------------|----|----|
| PepT <sub>So</sub> | ASEACERFSFYGMRNILLTPFLMTALLLSIPEELRGAVAKDVFHSFVIGVY     |    | 68 |
| HsPepT1            | VNEFCERFSYYGMRAILLILYFTNF . . . . ISWDDN.LSTAIYHTFVALCY |    | 64 |
| HsPepT2            | VNEFCERFSYYGMKAVLLILYFLYF . . . . LHWNED.TSTSIYHAFSSLCY |    | 94 |
| MmPepT1            | . . . . .                                               |    | 0  |
| RnPepT2            | . . . . .                                               |    | 0  |

|                    | H2                  | H3                                  |                        |
|--------------------|---------------------|-------------------------------------|------------------------|
|                    | <i>oooooooooooo</i> | <i>oooooooooooooooooooooooooooo</i> |                        |
| PepT <sub>So</sub> | FFPLGGWIADRFFGKYN   | TILWLSLIYCVGHAF                     | LAI FE..... 106        |
| HsPepT1            | LTPILGALIADSWLGKFK  | TIVSLSIVYTIGQAV                     | TSVSSINDLTDH NHDGT 114 |
| HsPepT2            | FTPILGAALIADSWLGKFK | TI IYLSLVYVLGHV                     | IKSLGALPI LGG..... 138 |
| MmPepT1            | .....               | .....                               | ..... 0                |
| RnPepT2            | .....               | .....                               | ..... 0                |

|                    | H5                                    | H6            |     |
|--------------------|---------------------------------------|---------------|-----|
|                    |                                       |               |     |
| PepT <sub>So</sub> | MFYFTINFGSFFASLSMPLLL...KNFGAAV       | AFGIPGVLMFVAT | 192 |
| HsPepT1            | IFYLAINAGSLLSTIITPMLRVQ.QCGIHSKQACYPL | AFGVPAALMAVAL | 213 |
| HsPepT2            | VFYLSINAGSLISTFITPMLRGDVQC...FGEDCYAL | AFGVPGLLMVIAL | 232 |
| MmPepT1            | .....                                 | .....         | 0   |
| RnPepT2            | .....                                 | .....         | 0   |

[illegible]

| ECD                |                                                    |     |
|--------------------|----------------------------------------------------|-----|
| PepT <sub>So</sub> | .....                                              | 399 |
| HsPepT1            | NAFMT.FDVNKLTRINISSPG.....SP.VTAVTDDFKQGQRHTLL     | 463 |
| HsPepT2            | KSFQKTPHYSKL...HLKTKSQDFHFLKYHNLSLYTEHSVQEKNWYSLV  | 490 |
| MmPepT1            | DTFMT.FDIDKLTSINISSSG.....SPGVTTVAHDFEQGHRHTLL     | 71  |
| RnPepT2            | SSFQNTTHYSKL...HLEAKSQDLHFHLKYNSLSVHNDHSVEEKNCYQLL | 84  |

| ECD                |                                                     |     |
|--------------------|-----------------------------------------------------|-----|
| PepT <sub>So</sub> | .....                                               | 399 |
| HsPepT1            | VWA..PNHY.QVVKDGLNQKPEKGENCIRFVNTFNELITITMSGKVYANI  | 510 |
| HsPepT2            | IREDGNSSISSMMVKD.TESRTTNGMTTVRFVNTLHKDVNISLSTDTSNLV | 539 |
| MmPepT1            | VWN..PSQY.RVVKDGLNQKPDKGENCIRFVNTLNEMVTIKMSGKVYENV  | 118 |
| RnPepT2            | IHQDGESISSMLVKD.TGIKPANGMAAIRFITNLHKDLNISLDTDAPLSV  | 133 |

| ECD                |                                                   |     |
|--------------------|---------------------------------------------------|-----|
| PepT <sub>So</sub> | .....                                             | 399 |
| HsPepT1            | S.SYNASTYQFFPSGIKGFTISSTEIPPQCQPN....FNTFYLEFGSAY | 554 |
| HsPepT2            | GEDYGVSAVRTVQRGEYPA.....VHCRTEDKNFSLNLGLLDFGAAY   | 581 |
| MmPepT1            | T.SHNASCYQFFPSGEKQYTINTTAVAPTCLTD....FKSSNLDFGSAY | 162 |
| RnPepT2            | GKDYGVSAYRTVLRGKYPA.....VHCETEDKVFSLDLGLDFGTTY    | 175 |

| ECD                |                                                     |     | H10   |  |     |
|--------------------|-----------------------------------------------------|-----|-------|--|-----|
| PepT <sub>So</sub> | .....                                               |     | ..... |  | 428 |
| HsPepT1            | TYIVQRKNDSCPEVKVFEDISANTVNMALQIPQYFLLTCGEVVF SVTGLE | 604 | ..... |  |     |
| HsPepT2            | LFVITNNTNQGLQAWKIEDIPANKMSIAWQLPQYALVTAGEVMF SVTGLE | 631 | ..... |  |     |
| MmPepT1            | TYVIRRASDGCLVKEFEDIPPNT.....                        | 186 | ..... |  |     |
| RnPepT2            | LFVITNITSQGLQAWKAEDI.....                           | 195 | ..... |  |     |

| H10                |                                                     |     | H11   |  |     |
|--------------------|-----------------------------------------------------|-----|-------|--|-----|
| PepT <sub>So</sub> | .....                                               |     | ..... |  | 478 |
| HsPepT1            | FSYSQAPSKAMKGTIMSFWTLSVTVGNLWVLLANVSVKSPTVTEQIVQTGM | 642 | ..... |  |     |
| HsPepT2            | FSYSQAPSSMKSVLQAAWLLTIAVGNIIIVLVVAQ.....FSGL        | 669 | ..... |  |     |
| MmPepT1            | .....                                               | 186 | ..... |  |     |
| RnPepT2            | .....                                               | 195 | ..... |  |     |

|                    |                                                                                         |                                             |     |
|--------------------|-----------------------------------------------------------------------------------------|---------------------------------------------|-----|
|                    |                                                                                         | H12                                         |     |
|                    |                                                                                         | <u>oooooooooooooooooooooooooooooooooooo</u> |     |
| PepT <sub>So</sub> | .SVTAFQMFF <b>F</b> AGFAILAAIV <b>F</b> ALY <b>A</b> RS <b>Y</b> QMQ.....               |                                             | 509 |
| HsPepT1            | SKQW.AEYIL <b>F</b> AALLLVVCVI <b>F</b> AIM <b>A</b> RF <b>Y</b> TYINPAEIEAQFDEDEKKNRL  |                                             | 691 |
| HsPepT2            | .VQW.AEFIL <b>F</b> SCLLLVICLI <b>F</b> SIM <b>C</b> YYY <b>V</b> VPVKTEDMRGP...ADKHIPH |                                             | 714 |
| MmPepT1            | .....                                                                                   |                                             | 186 |
| RnPepT2            | .....                                                                                   |                                             | 195 |
|                    |                                                                                         |                                             |     |
| PepT <sub>So</sub> | .....DHY.....                                                                           | 512                                         |     |
| HsPepT1            | EKSNPYFMSGANSQKQM                                                                       | 708                                         |     |
| HsPepT2            | IQGN.MIKLE.TKKTKL                                                                       | 729                                         |     |
| MmPepT1            | .....                                                                                   | 186                                         |     |
| RnPepT2            | .....                                                                                   | 195                                         |     |