Proton Control of Amino Acid Transport and Protein Trafficking

Zhiyi Wu

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

October 2020

Acknowledgements

I would like to thank my supervisor Professor Philip Biggin and Professor Simon Newstead. Without their kind guidance, I wouldn't be able to complete this thesis. I would like to thank Wellcome Trust and Department of Physiology Anatomy & Genetics for the financial support. I would like to thank Dr Joanne Parker and Dr Gurdeep Minhas for their help on the experimental techniques and advices from the experimental side. I would also like to thank Dr Naushad Velgy and Dr Laura Domicevica for their help in familiarising myself with molecular dynamics. I would like to thank Dr Irfan Alibay, Dr Matteo Aldeghi and Dr Joe Bluck for discussion about the principle of free energy calculations. I would like to thank Rocco Meli, Dr Will Glass and Professor Maria Musgaard for discussions about chemistry and quantum chemistry. I would like to thank Dr Anna Duncan, Dr Robin Corey, Tanadet Pipatpolkai for the help with course-grain simulations. I would also like to thank Josh Sauer, Michael Horrell, Dan Quetschlich, Patrick Simcock, Dr Wanlin Song, Dr Afroditi Zaki, Dr Marc Dämgen, Max Epstein, Mina Moniri and Dr Fiona Naughton for their accompany and all the fun times. Finally, I would like to thank my girlfriend Zhiyun Bian for her support during the thesis writing and my parents for their encouragement. I would like to thank again to Professor Philip Biggin and Professor Simon Newstead for their dedication to my project and all the support as well as encouragements.

Abstract

Protons play a key role in various biological process ranging from substrate binding, protein conformational change, oligomerisation processes as well as altering the chemical environment within the protein. Computational techniques such as molecular dynamics or quantum mechanics calculations are excellent tools at tracking the proton as well as examining its effect on the protein.

Secondary active transporters such as the SLC36 superfamily use the proton gradient to drive the transport of an amino acids against its concentration gradient. This thesis shows that the deprotonation of a key glutamate destabilises the inward-occluded state, which leads to the conformational change to the inward-open state. The conformational change clears the steric barrier from the intracellular gate to allow the egress of the substrate.

The KDEL receptor binds ER resident proteins in the Golgi and returns them to the ER. The binding of the ER retention tag (KDEL) occurs in a pH dependent fashion. This thesis shows that the protonation of a key histidine (H12) facilitates the formation of a low-barrier hydrogen bond, which locks the KDEL tag in the receptor, where as the deprotonation of this histidine weakens the hydrogen bond which allows the release of the KDEL tag.

The trafficking of the KDEL receptor between the ER and Golgi has been shown to involve oligomerisation. This thesis shows that the KDEL-bound KDEL receptor and the apo receptor could form dimers with different interfaces. The apo receptor could potentially form a nonamer when another key histidine (H150) is deprotonated.

Contents

List of Figures **v**
List of Tables **vii**
List of Abbreviations **viii**
1 Introduction **1**
1.1 The Proton in Biological Systems 1
1.2 Hydrogen Bonds are Electrostatic by Nature 3
1.2.1 Low-barrier Hydrogen Bond 3
1.2.2 Kinetic Isotope Effects 6
1.3 Proton-coupled Small Molecule Transport 6
1.3.1 Various Transporters Exist in the Human Body 6
1.2.2 The SLC7 Transporter Family 10
1.4 Transport Mechanisms 15
1.4.1 The Elevator Model 19
1.4.2 The Rocker-Switch Model 21
1.4.3 The Toppling Model 23
1.4.4 The Rocking-bundle Model 25
1.5 The Neurotransmitter Sodium Symporters (NSSs) 31
1.5.1 The Substrate Binding Pocket 31
1.5.2 The Allosteric Binding Pocket 35
1.5.3 The Conformational Change 36
1.5.4 Factors Determining Conformational Change 38
1.5.5 The Role of the Termini 40
1.6 Protein Trafficking in the Cell 42
1.6.1 Cellular Distribution 42
1.6.2 ER-Golgi Trafficking 43
1.6.3 The KDEL Receptor 46
1.6.4 The Coat Protein Complex I (COPI) Vesicles 48
1.6.5 The Structure and Function of KDEL Receptor 51
1.6.6 Structural Similarity and Evolution relationship 57
1.7 Aims and Hypothesis 60
2. Methods **62**

2.1 Molecular Dynamics 62
2.1.1 Equations of Motion for Molecular Systems 62
2.1.2 Force Computation 63
2.1.3 Time steps and Constrains 66
2.1.4 Coarse-grained Model 66
2.2 Long-range Forces 67
2.2.1 Reaction Field Method 68
2.2.2 Particle Mesh Ewald (PME) 68
2.3 Isothermal-isobaric Ensemble 69
2.4 Grand Canonical Ensemble and Monte Carlo Sampling 71
2.5 Energy Minimisation and Equilibrium Detection 72
2.6 Ligand Binding Free Energy 72
2.6.1 The Definition of Binding Free Energy 72
2.4.2 Thermodynamic cycle 74
2.7 Charge Correction 81
2.8 Attach, Pull, Release (APR) Scheme 84
2.9 Well-tempered Metadynamics 90
2.10 Quantum Mechanics 91
2.10.1 Solving Schrodinger Equation 91
2.10.2 Wave Function Analysis 94
2.10.2.1 Bond Order Analysis 94
2.10.2.2 Atom-in-molecule (AIM) Analysis 95
2.8.2.3 Electron Localisation Function (ELF) Analysis 97
2.8.2.4 Electrostatic Potential Analysis 98
2.11 QM/MM calculations 99
3. Proton Control of Transitions in an Amino Acid Transporter 101
3.1 Introduction 101
3.2 Material and Methods 104
3.2.1 System Preparation 104
3.3.2 Bilayer Setup 105
3.2.3 All-atom Molecular Dynamics 106
3.2.4 Reparameterisation of Protonated Glutamate 107

3.2.5 Umbrella Sampling Simulations and Potential of Mean Force Calculations 110
3.2.6 Absolute Binding Free Energy Calculations 110
3.2.7 Protonation Free Energy Calculations 112
3.3 Results 115
3.3.1 Protonated Glutamate Stabilised the Transporter in the Inward-occluded State 115
3.3.2 The Inward-open State Permits Substrate Exit 118
3.3.3 PMF Calculations Showed that E115 Plays a Key Role in the Proton-coupling of the Transport Mechanism 123
3.3.4 The Protonation Free Energy Confirmed that Protonation of the E115 Stabilised the Inward-occluded State 124
3.3.5 The Binding Affinity of Alanine is Constant Across Different Conditions 125
3.4 Discussion 126
3.5 Conclusion 128
4. The KDEL Trafficking Receptor Exploits pH to Tune the Strength of an Unusual Short Hydrogen Bond 129
4.1 Introduction 129
4.2 Methods 132
4.2.1 QM/MM Simulations 132
4.2.2 QM Geometry Optimisation 132
4.2.3 QM Interaction Energies 133
4.2.4 QM Proton Transfer Energy 134
4.2.5 MM Protonation Free Energy Calculations. 134
4.2.6 Grand Canonical Monte Carlo (GCMC) Simulations 135
4.3 Results 135
4.3.1 Positioning the Hydrogen in the Protein 135
4.3.2 The Strength of the E127···H—Y158 H-bond Depends on the Protonation State of H12 139
4.3.3 The Short Hydrogen Bond Influences the Affinity of H12 for Protons 144
4.3.4 Protonated Histidine Stabilised More Water in the Intracellular Pocket 147
4.4 Discussion 150

4.5 Conclusion 153

5. Conformational change of the KDEL receptor drives different oligomerization profile 154

5.1 Introduction 154

5.2 Methods 158

5.2.1 Characterisation of the Dimer Interface 158

5.2.2 System Setup and Simulations Details 159

5.2.3 Unbiased Simulations 160

5.2.4 Two-dimensional Potential of Mean Force Calculations 161

5.2.5 Semi-restrained Nonamer Formation 163

5.2.6 Attach Pull Release Calculations 163

5.3 Results 164

5.3.1 Unbiased Simulations Showed Multiple Possible Dimerisation Interfaces 164

5.3.2 Two-dimensional Free Energy Calculations Reveal the Most Populated Dimer Interface 169

5.3.3 Apo Structure Permits a Nonamer Formation 172

5.3.4 Protonation of Histidine Destabilised the Apo Dimer 174

5.4 Discussion 176

5.5 Conclusion 180

6 Conclusion and Future Work 181

7 Published Work 185

Bibliography 186

List of Figures

Figure 1.1: Classification of the hydrogen bonds.

Figure 1.2: Classification of the transporters.

Figure 1.3: Antiporter and uniporter mechanisms.

Figure 1.4: General mechanisms of channels and transporters.

Figure 1.5: Summary of the different transport mechanism.

Figure 1.6: The Elevator Model using GltPh as an example.

Figure 1.7: The Rocker-switch Model using semi-SWEET as an example.

Figure 1.8: Mechanism of the Toppling Model.

Figure 1.9: The topology of CATs, LATs and GkApcT.

Figure 1.10: LeuT structure fold use GkApcT as an example.

Figure 1.11: The transport cycle of rocking-bundle mechanism.

Figure 1.12: The binding pocket in the NSS transporters.

Figure 1.13: Subregion B/C in the binding pocket of the LeuT structural fold.

Figure 1.14: The calnexin/calreticulin cycle of protein quality control.

Figure 1.15: The recruitment of KDELR to the COPI.

Figure 1.16: Ligand binding induced conformational change.

Figure 1.17: Structure comparison of KDELR and SWEET.

Figure 2.1. Using thermodynamic cycle to compute the binding free energy.

Figure 2.2. Using restraint to restraint the relative orientation of ligand to protein.

Figure 2.3: Restraints for the Attach, Pull, Release scheme.

Figure 2.4. Critical points in AIM analysis and ELF analysis.

Figure 3.1. Structure and mechanism of GkApcT.

Figure 3.2. Potential energy surface of the C-C-O-H dihedral of protonated glutamate.

Figure 3.3. Finite size correction for the chloride ion charge annihilation free energy.

Figure 3.4. Unbiased simulation of GkApcT in a lipid bilayer.

Figure 3.5. Work done as a function of time during the steered MD.

Figure 3.6. The rebinding of alanine to the binding site.

Figure 3.7. The transition from inward-occluded state to the inward-open state.

Figure 4.1. Structures of the KDEL receptor.

Figure. 4.2. The electrostatic potentials on the 0.001 a.u. contour of the molecular electron density.

Figure. 4.3. Energy of proton transfer from the Y158 to E127.

Figure 4.4. Influence of the short H-bond on H12 affinity.

Figure 4.5. The water network in the cytoplasmic water pocket.

Figure 4.6. Purposed scheme of KDEL-mediated ER-Golgi trafficking.

Figure 5.1. A model of the trafficking of LMAN1 between ER and Golgi.

Figure 5.2. Two dihedrals characterising the dimer interface.

Figure 5.3. Equilibration stages of the two-dimensional potential of mean force calculations and the nonamer formation.

Figure 5.4. Unbiased oligomerisation simulations.

Figure 5.5. Unbiased dimerisation simulations.

Figure 5.6. Two-dimensional occupancy and free energy landscape.

Figure 5.7. Nonamer formation.

Figure 5.8. Protonation state of the H150 influences the dimerisation process.

Figure 5.9. A schematics of the shuttling of the KDEL receptor between ER and Golgi.

List of Tables

Table 1.1 Mammalian Amino Acid Transport System

Table 3.1. Summary of free energy to protonate the E115 sidechain.

Table 3.2. Absolute free energies for alanine.

Table 4.1. Free energy and geometry of the hydrogen bonds.

Table 4.2. Properties of the hydrogen bond.

Table 4.3. ΔΔG of (in kcal/mol) of deprotonation* of H12 in the KDEL-bound and protonation** of H12 in apo state under different conditions.

List of Abbreviations

AMBER Assisted Model Building with Energy Refinement

APC Amino acid-Polyamine-organoCation

Arf1 ADP Ribosylation Factor 1

ATP Adenosine TriPhosphate

BiP immunoglobulin heavy chain-Binding Protein

CAT Cationic Amino acid Transporter

CamKII Ca^{2+}/calmodulin-dependent protein Kinase II

CDC42 Cell Division Cycle 42

COPII Coat Protein complex II

CV Collective Variable

DAT DopAmine Transporter

EDEM ER Degradation-Enhancing α-Mannosidase-like protein

EMC ER Membrane protein Complex

ER Endoplasmic Reticulum

ERAD ER Associated Degradation

ERGIC ER-Golgi Intermediate Compartment

ERS ER Retention Signal

GAT GABA Transporter

GABA γ-AminoButyric Acid

GAP GTPase-Activating Protein

GDP Guanosine Di-Phosphate

GEF Guanine nucleotide Exchange Factor

GPCR G Protein-Coupled Receptor

GROMACS GROningen MAchine for Chemical Simulations

GTP Guanosine TriPhosphate

HAT Heteromeric Amino Acid Transporter

KDEL Lys-Asp-Glu-Leu peptide

KDELR KDEL receptor

LAT L-type Amino acid Transporter

L-DOPA L-3,4-DihydrOxyPhenylAlanine

LeuT Leucine Transporter

LINCS LINear Constraint Solver

NET NorEpinephrine Transporter

NSS Neurotransmitter Sodium Symporter

PDI Protein Disulfide Isomerase

PHB ProHiBitin-1

PIP_2 PhosphatidylInositol 4,5-biPhosphate

PME Particle Mesh Ewald

PMF Potential of Mean Force

POPC 1-Palmitoyl-2-Oleoyl-*sn*-glycero-3-PhosphoCholine

POPE 1-Palmitoyl-2-Oleoyl-*sn*-glycero-3-PhosphoEthanolamine

POPG 1-Palmitoyl-2-Oleoyl-sn-glycero-3-Phospho-(1'-rac-Glycerol)

QM Quantum Mechanics

RMSD Root Mean Square Deviation

SERT SERotonin Transporter

SLC Solute Carriers

SRP Signal Recognition Particle

SWEET Sugars Will Eventually be Exported Transporter

TM TransMembrane helices

TOG Transporter-Opsin-G protein-coupled receptor

UPR Unfolded Protein Response

WHAM Weighted Histogram Analysis Method

ZPE Zero-Point Energy

1 Introduction

1.1 The Proton in Biological Systems

Despite its small size and mass, the proton is one of the most significant particles in biology. It can act as a molecular glue that holds proteins together or allow signals to propagate through cells. It can also play a vital role in determining the affinity of a small molecule for a protein receptor, typically via the protonation state of key amino acids. The chemical property of a proton allows enzymes to facilitate reactions and allows the conversion of sugar or light into energy and its subsequent storage. The energy can then be used to “power” various biological processes.

In physiological conditions, molecules are brought together mainly via three types of interactions: the ionic bond, van der Waals interactions and the hydrogen bond. Among these interactions, hydrogen bonds are found to play a key role in physiological processes as the strength of the hydrogen bonds could be modulated by many factors such as the environmental pH. The hydrogen bond is a type of polar interaction where the proton, which is an electropositive atom, is shared between two electronegative atoms. In most cases, the hydrogen bond is directional in that the proton forms a covalent bond with the more electronegative atom (hydrogen bond donor) and electrostatically interacts with the less electronegative atom (hydrogen bond acceptor). The strength of the hydrogen bond is tuneable. In enzymes, the closing up of the catalytic site desolvates the hydrogen bond, leading to a stronger sharing of the proton between the donor and the acceptor and thus, a stronger hydrogen bond (1) and in the case of ketosteroid isomerase, a 10,000-fold increase in acidity (2). In contrast, when a hydrogen bond is

highly solvated, water will compete for the potential interaction sites leading to a weakening of the hydrogen bond in question (~0.15 kcal/mol (3)).

Proteins can exploit pH (the proton concentration) to alter the strength of a hydrogen bond and in turn affect functional properties. For example, the hydrogen bond may only exist in a certain pH range and thus, provides attractive interaction between the two protonatable residues. The protein GkApcT is stabilised by the hydrogen bond forming between protonated E115 and deprotonated D237 at pH 4.0 (4). The state of a protonated E115 and deprotonated D237 only exist within a narrow range of pH, where in the lower pH, however, both residues become deprotonated, resulting in stacked negative charges and thus, exerting repulsive forces between the two residues (5). In higher pH range, on the other hand, the interaction between protonated E115 and D237 cannot offer sufficient stabilisation to the protein resulting in the protein having an optimal working pH.

Protons mediate the breakage of polysaccharide bonds as part of the lysosome-mediated pathogen elimination process (6). Lysosome is an organelle, where there is significant pH difference between the lumen and the cytoplasm. One role of the lysozyme is to eliminate external pathogens as part of the immune response. To ensure the effective elimination of bacteria, lysosomes digest the cell wall of bacteria by cutting open oligosaccharide chains. The hydrolysis reaction is catalysed by forcing the sugar into the catalytic site, where aspartate from the lysosome forms a covalent bond with the C1 carbon of a sugar molecule. The breakage of the polysaccharide bond is then stabilised as glutamate donates a proton to the next sugar down the oligosaccharide chain. A water is then split to restore the proton to glutamate and break the sugar free from aspartate (7).

In an environment where there is a highly-ordered hydrogen-bonding network, protons can traverse the space efficiently. In proteins, the arrangement of the polar and ionic residues could create an ordered hydrogen-bonding network, where Grotthuss shuttling could happen. During Grotthuss shuttling (8), by moving a proton from hydrogen bond donor to the hydrogen bond acceptor in each hydrogen bonding pair, the overall effect would be shuttling a proton across a long distance. It has been shown that proton transferring by Grotthuss shuttling (happening at μs time scale (9)) happens more than 40 times faster in a protein than in water. This exploitation of specific hydrogen-bond donors and acceptors for fast proton transfer is called a proton-wire and is found to play a key role in proton pumps such as cytochrome c oxidase (10).

1.2 Hydrogen Bonds are Electrostatic by Nature

Hydrogen bond is formed between a pair of atoms that have high affinity for electrons, where a hydrogen atom sits in the middle of this pair of atoms. One atom of the pair, which can be fluorine, nitrogen or oxygen atoms, bond covalently to the hydrogen atom and is called hydrogen bond donor. Due to the difference in electron affinity between the hydrogen and the bonded atom, the electrons are shared unequally and the hydrogen atom takes up a slightly positive charge. The other atom in this pair, which could also be fluorine, nitrogen or oxygen has a lone pair of electrons, which gives it a slightly negative charge and is called hydrogen bond acceptor. Thus, the interactions between the hydrogen bond acceptor and the hydrogen bond donor are mainly electrostatic interactions between the opposite charges.

1.2.1 Low-barrier Hydrogen Bond

Though hydrogen bonds are mainly electrostatic by nature, strong hydrogen bond could still display significant covalent characteristics. The ordinary hydrogen bond is

asymmetric, where the hydrogen atom is bonded to the hydrogen bond donor heavy atom and interact electrostatically with the hydrogen bond acceptor. The hydrogen atom is confined to the donor heavy atom and the zero-point vibration energy (ZPE) only allows the proton to vibrate around the donor heavy atom (Figure 1.1A). This is the most common type of hydrogen bonds and weakly couple the acceptor to the donor. In certain cases, where the donor heavy atom and the acceptor are in the same chemical environment, a symmetrical hydrogen bond can be observed. For symmetrical hydrogen bonds, the proton is equally likely to bond to the donor heavy atom or the acceptor. However, since the donor heavy atom and acceptor are far apart, the energy barrier for the proton to go from the donor to the acceptor is higher than the ZPE and this proton transfer cannot happen spontaneously (Figure 1.1B). Reduction in the distance between the donor heavy atom and the acceptor would lower the energy barrier of this transfer, which gives the low-barrier hydrogen bond (LBHB). In the case of LBHB, the energy barrier for proton transfer is lower than the ZPE, which allows the proton to moving freely between the donor and the acceptor (Figure 1.1C). In this case, though the donor still interacts with the acceptor mainly via electrostatic interactions, some covalent characteristics can be observed, which allows LBHB to be stronger than ordinary hydrogen bond. In the most extreme case, where the donor and acceptor are close to each other, the hydrogen is covalently bonded to both of the donor and acceptor and is named as single-well hydrogen bond (Figure 1.1D). This single-well hydrogen bond, due to its strong covalent characteristics, has the strength comparable to covalent bond.

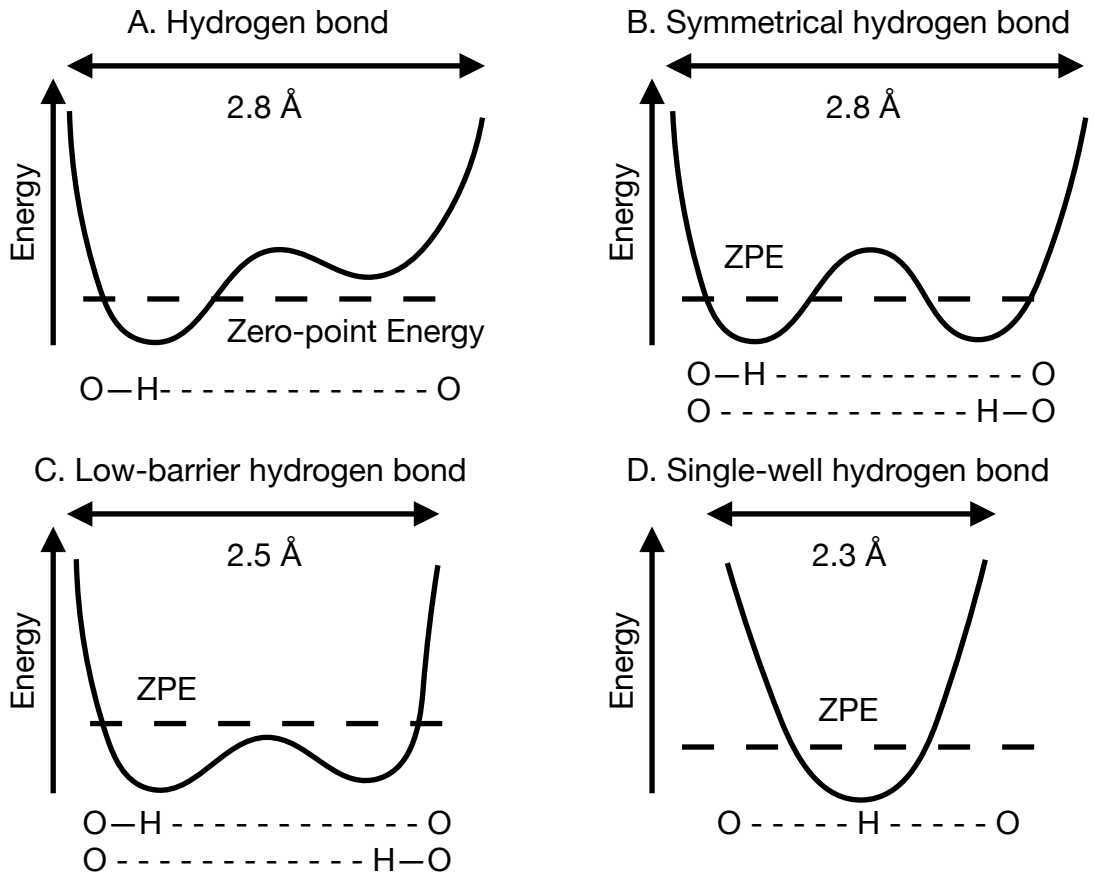


Figure 1.1*: Classification of the hydrogen bonds. A. Ordinary hydrogen bond, where the hydrogen bonded to one oxygen. B. Symmetrical hydrogen bond, where hydrogen has equal probability being close to both oxygens, but the energy barrier is higher than the zero-point energy (ZPE). C. Low-barrier hydrogen bond, where the distance between the two oxygens are shortened such that the energy barrier is lower than the ZPE and the proton can hop between the two oxygens. D. Single-well hydrogen bond, where the distance between the two oxygens are shortened such that the proton is bonded to both oxygens.*

1.2.2 Kinetic Isotope Effects

LBHB is purposed to have a significant effect on enzymatic reactions (11). Experimental characterisation of the LBHB could be done by utilising the kinetic isotope effects (KIE). As is shown in the previous section, the key factor distinguishing the LBHB from the ordinary hydrogen bond is whether the ZPE of the proton is higher than the energy barrier of the proton transfer. The zero-point vibration energy is the lowest possible energy of the system and is the ground state energy, which depends on the reduced mass of the molecule. Thus, the heavier the atom, the lower the frequency of the vibration and the smaller the ZPE. The lighter atoms, on the other hand, will have a higher frequency of vibration and a larger ZPE. By increasing the mass of the hydrogen, the ZPE of the proton could be lowered such that it is unable to cross the energy barrier. Thus, in KIE experiments, when the protein with LBHB is immersed in the deuterated water solution. The hydrogen forming the LBHB, due to its lower mass, could cross the energy barrier and leaver the protein with its higher ZPE and is replaced with the deuterium in the water. However, due to its lager mass, the deuterium would not be able to leave the protein, resulting in deuterium enrichment, which could be observed in the NMR experiments.

1.3 Proton-coupled Small Molecule Transport

1.3.1 Various Transporters Exist in the Human Body

In addition to the generation of ATP, the proton gradient can also be used to power other biological processes, such as the movement of substances against their electrochemical gradient. Many biological processes such as nitric oxide synthesis require the influx of arginine into the cell. However, the cells are enclosed by a lipid membrane and due to the hydrophobic nature of the lipids, soluble substances, such as amino acids, cannot readily diffuse across the membrane. Thus, transporters exist to open up a passage in the lipid

membrane for these soluble materials to cross the hydrophobic barrier. The transporters can be divided into roughly two families: passive and active transporters. The passive transporters, which are also sometimes called facilitated diffusers or just facilitators, allow the substance to diffuse down its electrochemical gradient. Active transporters on the other hand enable the transport of substances against their electrochemical gradient (Figure 1.2).

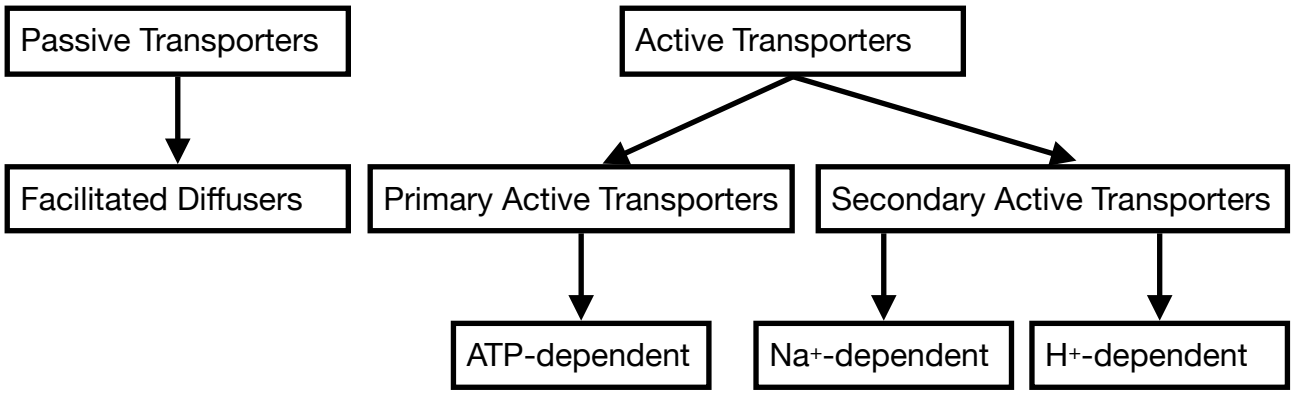


***Figure 1.2**: Classification of the transporters. The transporters could be classified into two major categories: the passive transporter which is also called the facilitated diffusers and the active transporters, which are further classified into primary active transporters that depend on ATP and secondary active transporters, which could be coupled to either sodium concentration or proton concentration.*

According to the law of thermodynamics, substances can spontaneously diffuse down their concentration gradient. To transport substances against their electrochemical gradients, another energy source has to be provided. Depending on the energy source, the active transporters can be further divided into two categories. The first is the primary active transporters, which couple the transport of substance to the hydrolysis of adenosine triphosphate (ATP). The other category is termed secondary active transporters, which couple the uphill movement of one substance against its electrochemical gradients to the downhill movement of another substance down its electrochemical gradients. The substance moving down its electrochemical gradients is often sodium ions. The sodium ion concentration gradient is maintained by the sodium-potassium pump. Another example could be the proton concentration gradient, which could be generated by P-type ATPase.

It is estimated that 10% of all human genes encode for transporter proteins (12). The amino acid-polyamine-organocation (APC) superfamily is a family of secondary active transporters that shuttle cellular metabolites across the membrane (13). Recently, the structure of a proton-coupled amino acid transporter from bacteria has been solved (4). This transporter is classified as belonging to the APC superfamily and since it was found in the bacteria *Geobacillus Kaustophilus*, it is termed GkApcT. The transporters from the APC superfamily are sometimes classified via the solute carriers (SLC) nomenclature (14). A transporter is said to belong to a certain family if it has 20% sequence identity to the other members of the family. The GkApcT was found to share over 40% amino acid sequence identity with the eukaryotic SLC7 family of transporters (4).

1.2.2 The SLC7 Transporter Family

The SLC7 cationic amino acid transporter/glycoprotein-associated amino-acid transporter family can be subdivided into the cationic amino acid transporters (CATs) and the L-type amino acid transporters (LATs). The CATs are thought to function as a monomer. LATs on the other hand, have to combine with the heavy subunits (glycoproteins) from the SLC3 family to form functional transporters. Since the transporter is composed of two different (light and heavy) subunits, these complexes are also called heteromeric amino acid transporters (HATs).

Compared with the LATs, the transporter GkApcT is more similar to the CATs in terms of amino acid sequence identity (4). The CATs are passive transporters and mediate the sodium independent transport of positively charged amino acids. There are four types of CATs in human, CAT-1, CAT-2A/B and CAT-3. The CAT-2 family has two different splice variants, namely the CAT-2A and CAT-2B. The CAT transporters show the characteristics of trans-stimulation. The trans-stimulation is defined as the phenomenon where the transport rate is enhanced by the presence of substrate at the trans side of the transporter (15). The trans-stimulation could be considered as the combination of the uniport and antiport. For uniporter type of transporters, after transporting the substrate from the cis-side of the membrane to the trans-side, the transporter could efficiently reorientate itself back to accept substrate from the cis-side. The energy barrier for this conformational change is low enough for this process to happen spontaneously. The antiporter, on the other side, cannot reorientate itself back after shuttling one substrate from the cis-side to the trans-side, due to the prohibitively large energy barrier (Figure 1.3). However, upon the binding of the substrate at the trans-side, the energy barrier is reduced significantly, which allows the antiporter to carry the substrate back to the cis-side and complete the transport cycle. The CATs have a low enough energy barrier for this

conformational change to happen and, thus, can be termed as a uniporter. However, this energy barrier could be lowered further by the binding of substrate at the trans-side and, thus, leads to a faster completion of transport cycle and higher transport rate.

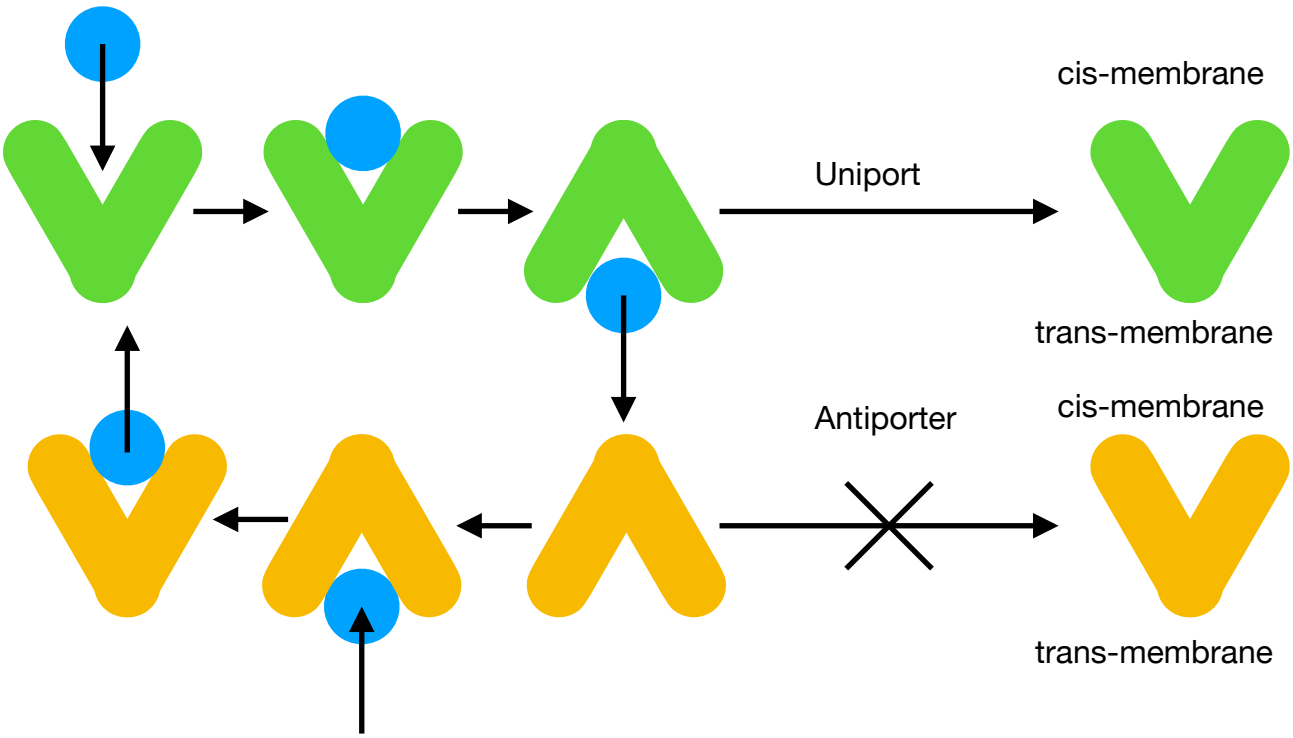


***Figure 1.3**: Antiporter and uniporter mechanisms. After transporting the ligand from the cis side of the membrane to the trans side of the membrane, the uniporter could spontaneously return to face the cis side of the membrane. The antiporter, on the other hand, would require the another ligand from the trans side of the membrane for it to reorientate back to the cis side of the membrane.*

Furthermore, The CAT transporters are shown to have a high affinity for arginine (K_m in the range 38–450 μM, which is similar to the concentration of arginine as is expected in the blood plasma). The K_m value for lysine was found to be at a similar level. CAT-1 was found to have the highest affinity for arginine and the strongest level of trans-stimulation. The CAT-2A family on the other hand, have the lowest affinity for arginine and very little trans-stimulation. There is also different preference in terms of substrate specificity for different CATs, where the CAT-1 can transport histidine, while CAT-3 cannot (16). Though very weak, the CATs have been shown to able to transport neutral and negatively charged amino acids.

The HATs transporters are obligatory exchangers and are found to operate on a wide range of amino acids (17) (Table 1.1, adopted from (18)). The system L and asc transporter can transport neutral amino acids. The arpAT transporter can transport aromatic amino acids. The x_c^- transporter system can transport negatively charged amino acids. The y^+L and $b^{0,+}$ transporters can transport both cationic and neutral amino acids. The y^+L transporter couples the movement of amino acids to the sodium gradient. The HATs were found to localise in renal and small intestinal epithelia and mediates the re-absorption of positively charged amino acids from the lumen to the extracellular space.

Table 1.1 Mammalian Amino Acid Transport System

System	Na+-dependent	Typical substrates
B	Yes	Threonine, alanine
ASC	Yes	Alanine, glycine
IMINO	Yes	Proline, pipecolic acid
β	Yes, with Cl	β-Alaninc
X_c^-	Yes, with Cl	Glutamate, aspartate
A	Yes	MeAIB
y^+	No	Arginine, lysine
$b^{0,+}$	No	Lysine, leucine
L	No	BCH, phenylalanine, leucine

Though in terms of the sequence identity, the GkApcT is more closely associated to the CATs instead of the LATs, in the crystal structure, the GkApcT was solved in complex with a single trans-membrane helix YneM, a protein found in *Escherichia coli*. The pattern of transporters requiring a transmembrane helix to function resembles the HATs and is not observed to be the case for CATs. The heavy subunit of the HATs, which linked to the LATs via a disulphide bond, are type II membrane N-glycoproteins. The heavy subunit is composed of a single transmembrane helix and a glycoprotein in the C-terminus (19). This heavy subunit is thought to be critical for the transport of the transporter (light subunit) to the plasma membrane (20). Aside from marking the transporter for plasma membrane, the heavy subunit also participates in cell proliferation (21), cell fusion (22) and β-integrin signaling (23). The cystines, which form the disulfide bond between the heavy subunit and the light subunit, are located between the transmembrane helix and the C-terminal glycoprotein.

Among the various kind of HATs, the heteromeric transporter of 4F2hc/y^+LAT1, exchanges the cationic amino acid with the neutral amino acid (24). Thus, together with the CATs, HATs are able to maintain the cationic amino acid and neutral amino acid balance in the cell (25). The LAT1 transporter was also found to play a role in assisting foetal development through supplying thyroid hormones. Furthermore, the wide range of substrates which could be transported by LAT1 allows it to be used for drug delivery (e.g. helping L-DOPA (L-3,4-dihydroxyphenylalanine) to cross the blood brain barrier) (26).

1.4 Transport Mechanisms

Despite the enormous amount of cellular and biochemical work being done on transporters, there is still much unknown about the detail of how transporters work at the atomic level. Though it is recognised that different transporters work differently, it has also

been suggested that there is also considerable similarity between some transporters, which allows classification according to their proposed mechanism. Compared with channels, which open up a continuous pathway from the extracellular side to the intracellular side of the membrane, transporters work by providing alternating access to the extracellular side or the intracellular side and thus adopting the generalised transport process referred to as the alternating-access mechanism (Figure 1.4). Within this broad envelope of the alternating-access mechanism, four specific mechanisms exists; i) the elevator mechanism, ii) the rocking-switch mechanism iii) the toppling mechanism and iv) the rocking-bundle model (also known as the thin/thick gates model) (Figure 1.5) (27).

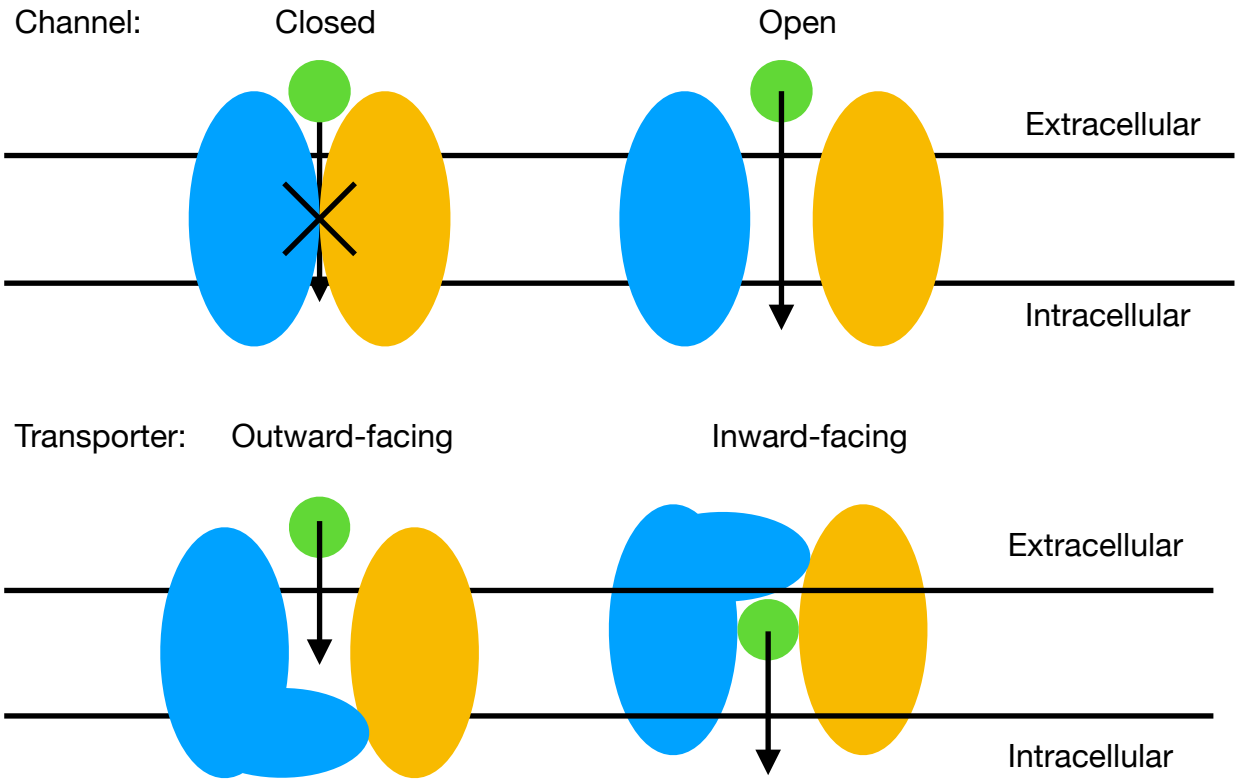


***Figure 1.4**: General mechanisms of channels and transporters. Channels can adopt open and closed conformations. In the closed state, the channel is occluded from both extracellular and intracellular side. In the open state, both extracellular and intracellular gates are open, which allow ions (green) to pass through. The transporters, on the other hand, have access to the extracellular and intracellular side alternatively. In the outward-facing state, the extracellular gate opens up to allow substrate (green) to enter. In the inward-facing state, the intracellular gate opens up to allow substrate (green) to exit.*

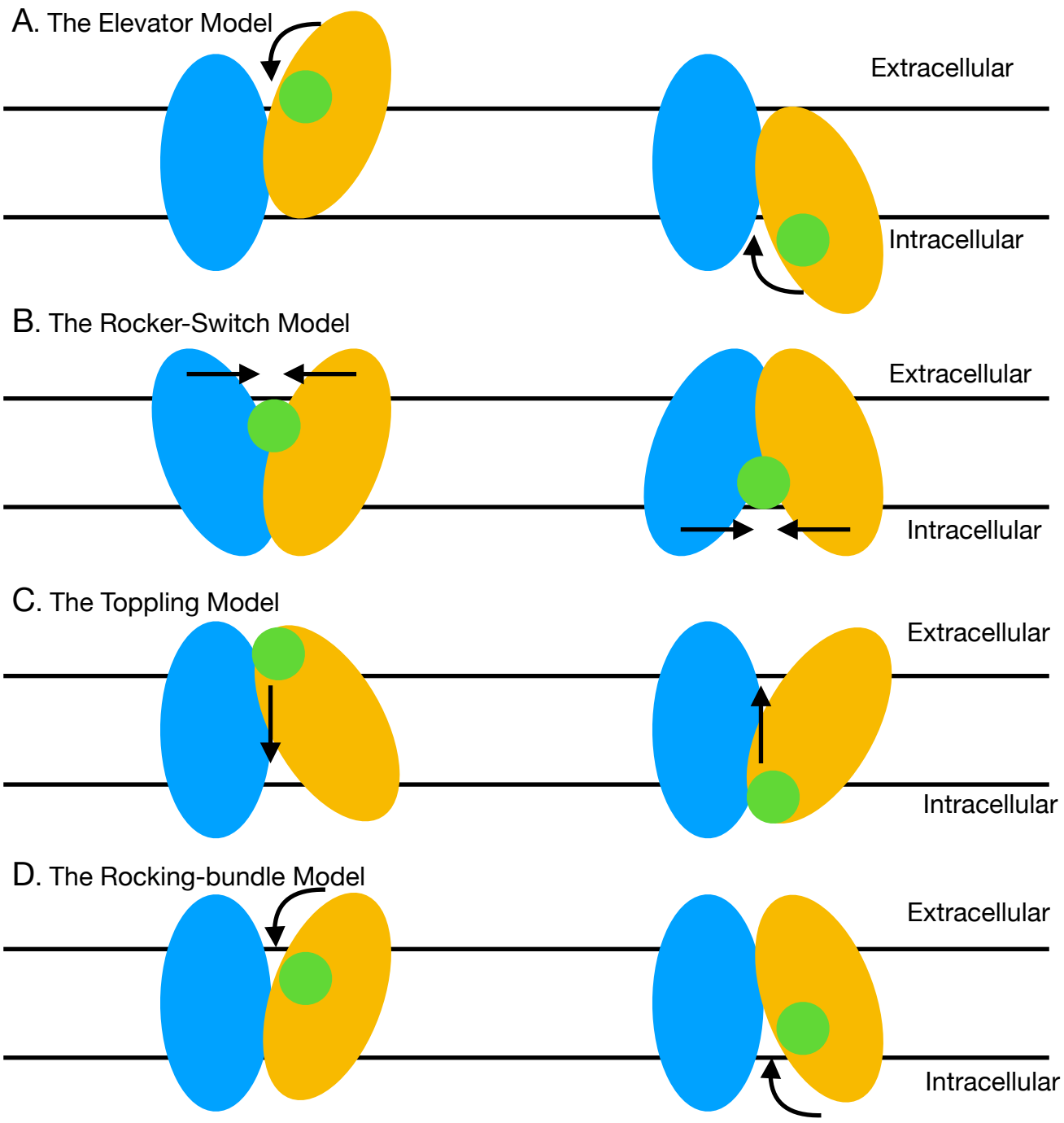


***Figure 1.5**: Summary of the different transport mechanism. A. The elevator model is characterised by the vertical movement of the transport domain (orange) with respect to the scaffold domain (blue), which shuttles the substrate (green) across the membrane (black lines). B. In the rocker-switch model, there is no scaffold domains and the substrate is exposed to different sides of the membrane through the concurrent horizontal motion of all the helices. C. In the toppling mode, the tip of the transport domain slide against the scaffold domain to transport the substrate across the membrane. D. In the rocking-bundle model, the tip of the transport domain moves horizontally towards or away from the scaffold domain to deliver the substrate.*

1.4.1 The Elevator Model

In the elevator model, the substrate is confined to a pocket which slides against the scaffold to deliver the substrate across the membrane. One example of a transporter that employs the elevator model is the sodium–aspartate symporter from *Pyrococcus horikoshii* (Glt_{Ph}). The structure of Glt_{Ph} can be classified into two domains - the scaffold domain and the transport domain. The substrate binds exclusively to the transport domain, which undergoes rotational and translational movement and moves the binding pocket roughly 18 Å towards the intracellular side (Figure 1.5A). The scaffold domain, on the other hand, stays immobile and is thought to participate in the dimerisation or trimerisation process. Glt_{Ph} consists of 8 TMs and two helical hairpins, HP1 and HP2. The structure of Glt_{Ph} consists of two inverted repeats, which are TM1-3 and TM4-6 or HP1-TM7 and HP2-TM8. TM1-2 and TM4-5 forms the scaffold domain, whereas the rest forms the transport domain (Figure 1.6). TM7 and TM8 hosts the substrate binding site, which contains unwinded region to accommodate different substrates.

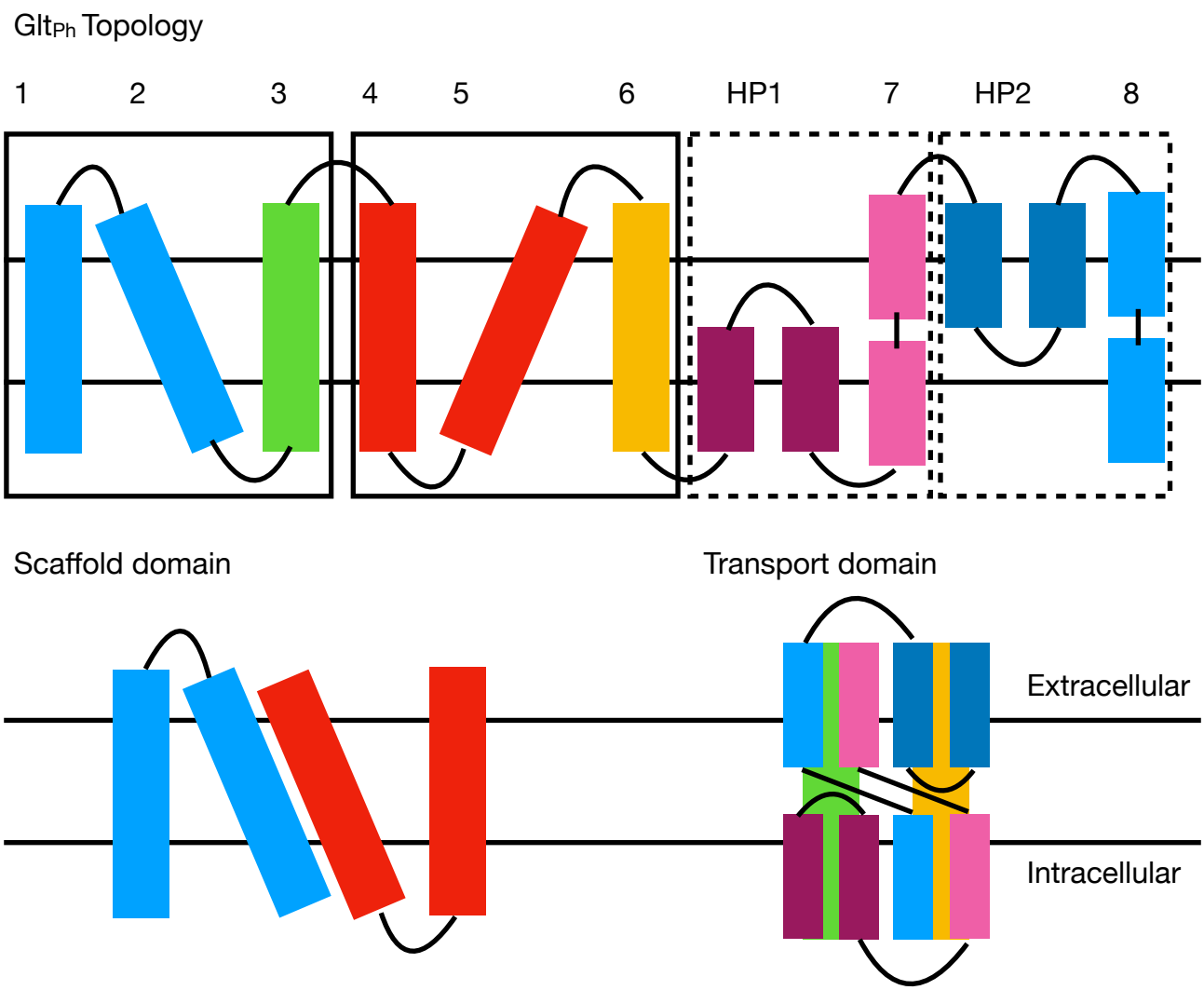


***Figure 1.6**: The Elevator Model using Glt_{Ph} as an example. The topology of Glt_{Ph} is consisted of two inverted repeats, which are TM1-3 and TM4-6 (solid square) and HP1, TM7 and HP2, TM8 (dashed square). The TM1-2 and TM4-5 form the scaffold domain, while the rest of structure forms the transport domain. Figure adapted from (35).*

1.4.2 The Rocker-Switch Model

Another type of transport mechanism is called the rocker-switch mechanism, where two symmetrical bundles moves around the central binding pocket to translocate the substrate. On example would be the Sugars Will Eventually be Exported Transporters (SWEETs) (28). The SWEET is a dimer consisted of two three transmembrane segments. In the middle of TM1 of both segments exists the PQ-loop motif (29), which divides TM1 into TM1a and TM1b (30). The two segments are joined by the packing of TM1a against TM2 (31). During the conformational change from outward-facing to inward-facing, the TM1a moves with the symmetry pair in the other segment, while TM1b moves with its own segment (31). This conformational change switch the opening from one end of the membrane to the other end of the membrane (32) (Figure 1.7).

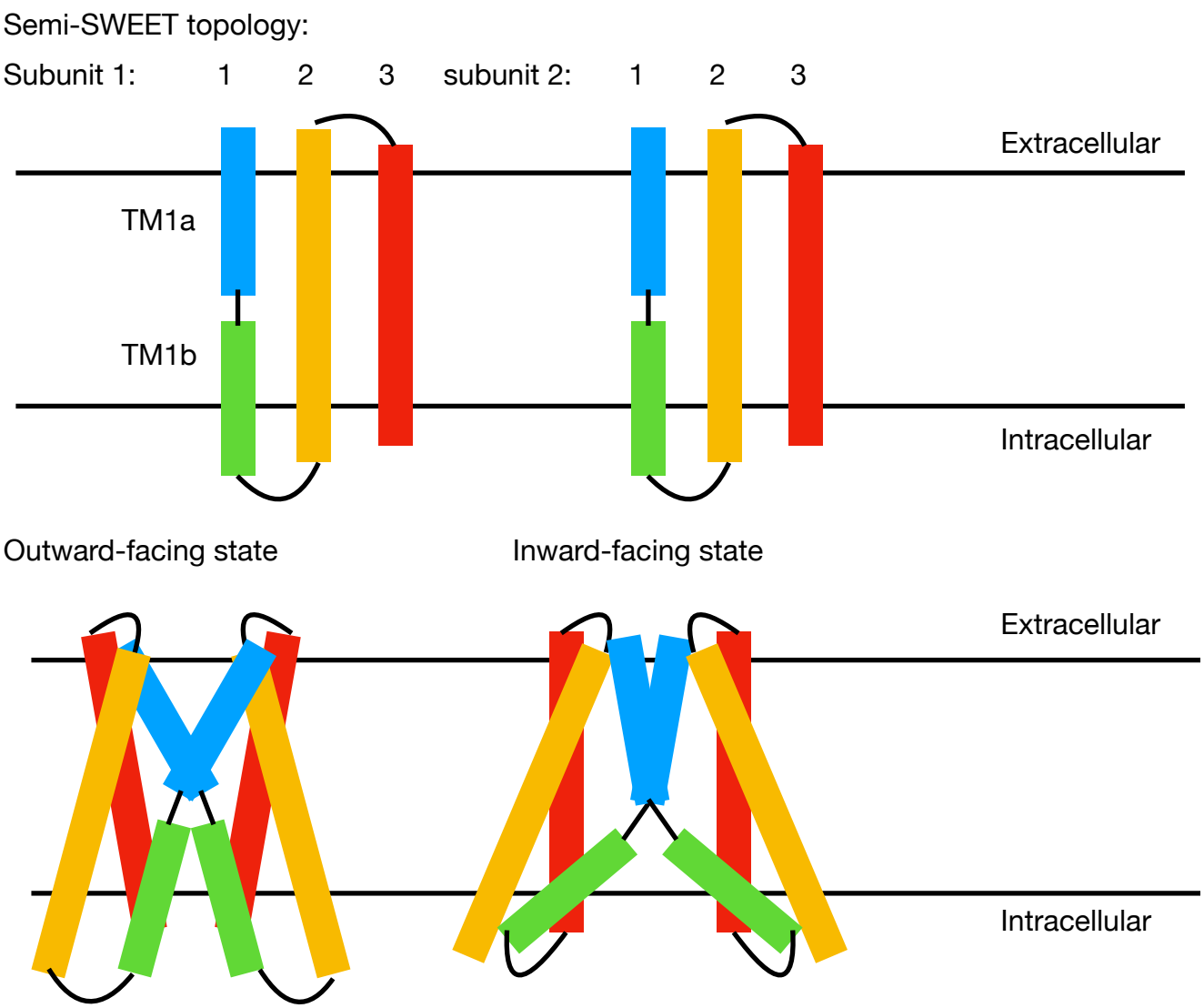


***Figure 1.7**: The Rocker-switch Model using semi-SWEET as an example. The semi-SWEET transporter is a dimer of two 3 TM subunits, where the TM1 is unwound in the middle. The conformational change from outward-facing to inward-facing is characterised by the mirrored motion between two subunits. Figure adapted from (13).*

1.4.3 The Toppling Model

The key feature of the toppling model is the toppling motion of the tip of the transporter which allows the substrate to shuttle across the membrane. A sample transporter would be the ECF transporters (33). The ECF transporters are consisted of the ECF module and the S-component. The ECF module modulates the binding of S-component in an ATP-dependent fashion, where the apo S-component can only be released in the presence of the ATP. The dissociated S-component then orientates towards the extracellular side which allows the substrate to bind. The ECF modules without bound ATP would then bind the S-component and stabilise the S-component such that the binding pocket is facing the intracellular side, which leads to the release of the substrate. The binding of the ATP to ECF modules then releases the S-component for another cycle of transport (Figure 1.8).

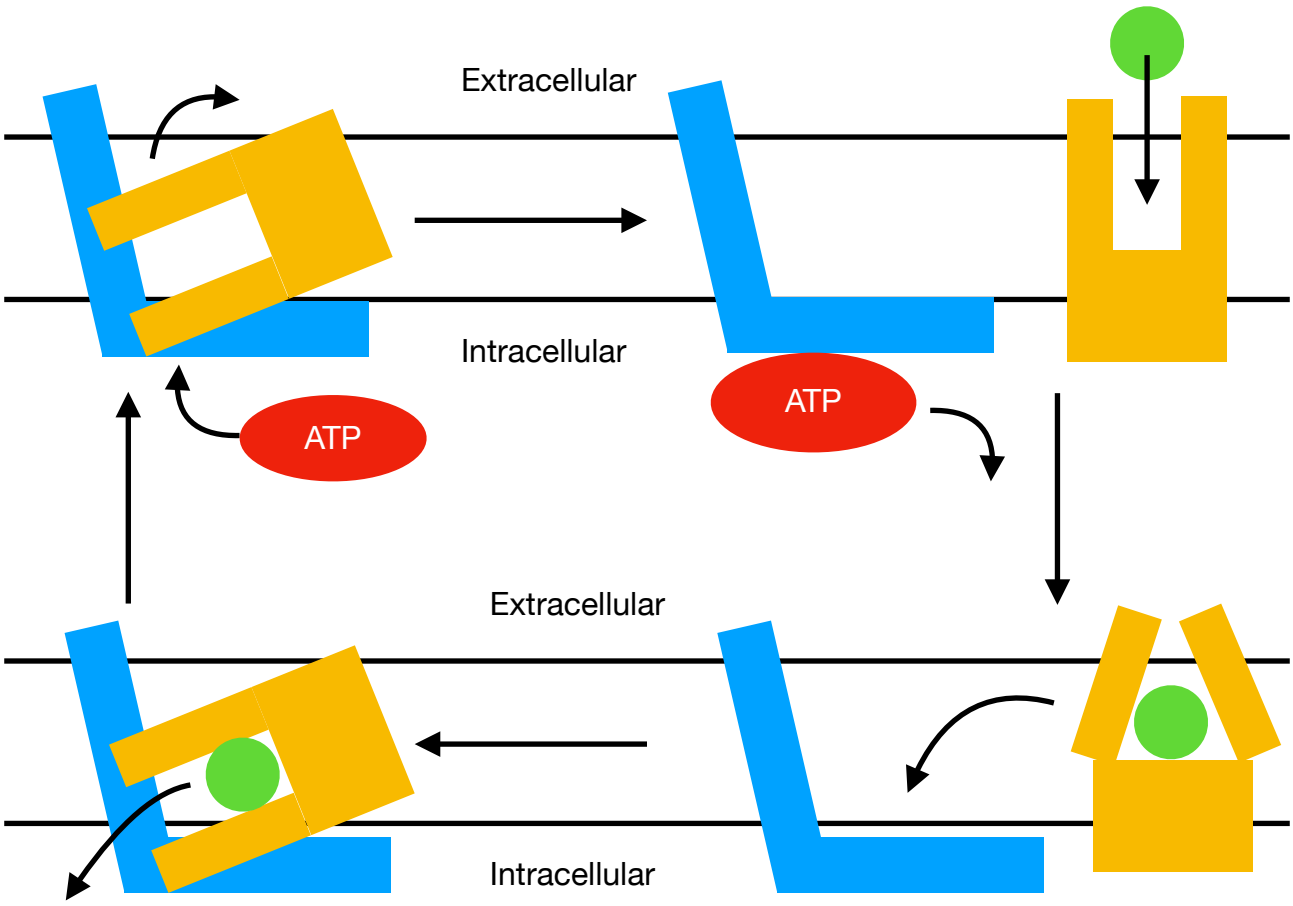


***Figure 1.8**: Mechanism of the Toppling Model. The binding of ATP (red) releases the S-component (orange) from the ECF module (blue), which allows the substrate (green) to enter the S-component. The substrate bound S-component proceeds to bind the ECF module, which leads to the release of the substrate. Figure adapted from (23).*

1.4.4 The Rocking-bundle Model

The last type of mechanism is called the rocking-bundle mechanism. A large number of transporters employ this transport mechanisms exemplified by the Leucine transporter (LeuT). These transporters are thus sometimes also referred to as the “LeuT fold” transporters and the GkApcT is one example of this structural fold. Topologically, all the transporters in the ApcT superfamily shared a similar topology. The GkApcT is composed of 12 transmembrane helices (TM), while the CATs are composed of of 14 transmembrane helices (34) and the LATs have 12 transmembrane helices (35). The first 12 transmembrane helices in the GkApcT, LATs and CATs are all quite similar and the last 2 helices in the CATs are thought to be a duplication of the last 2 helices in the first 12 transmembrane helices (36) (Figure 1.9). The bacterial homolog of LATs was found to lack the key cystine which forms disulphide bridge between the light and heavy subunits (37). However, given the reducing environment in the bacteria, it is questionable as to whether the disulfide bridge will have any importance.

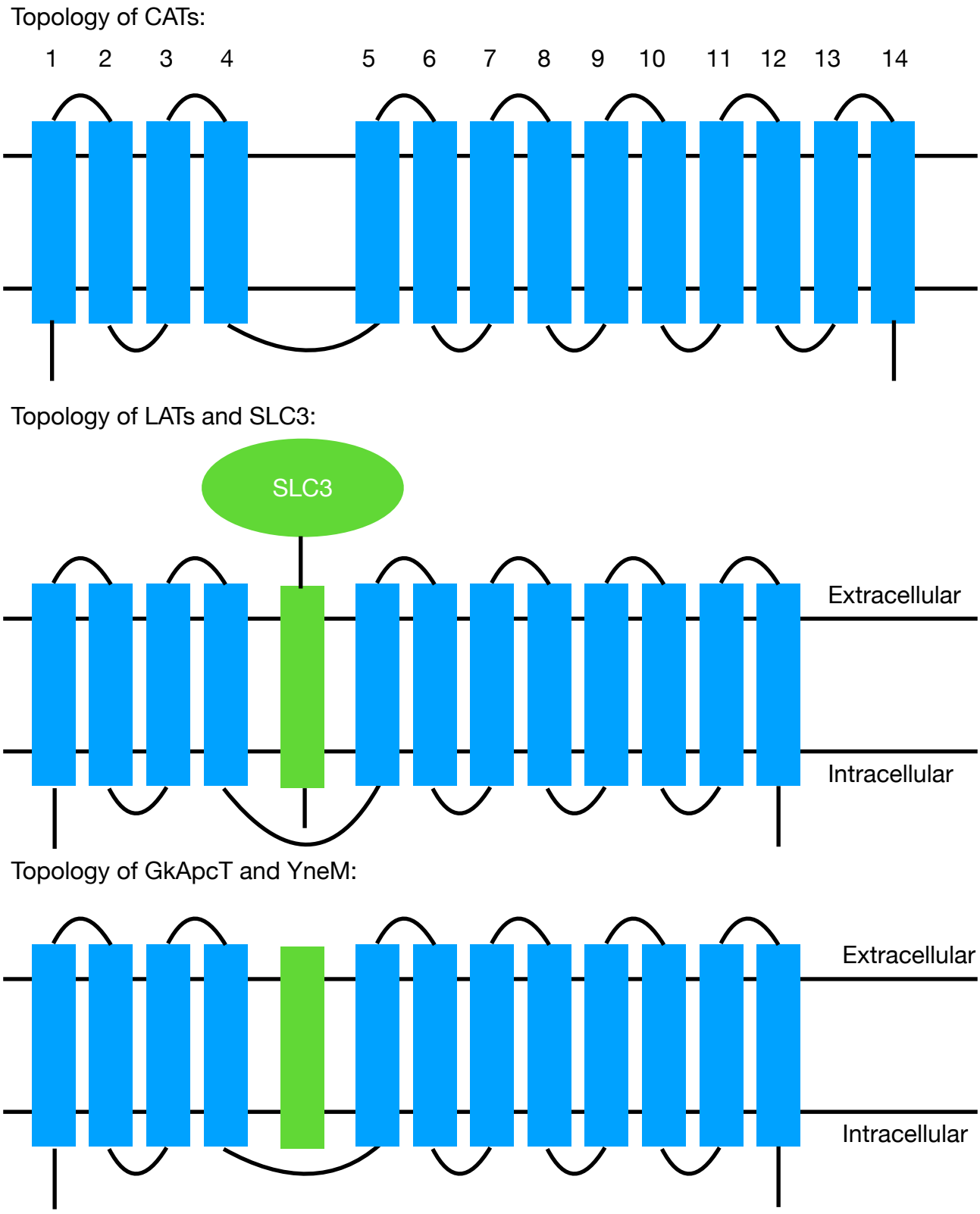


***Figure 1.9**: The topology of CATs, LATs and GkApcT. The CATs have 14 TMs, whereas LATs and GkApcT contain 12 TMs. The last 2 TMs in the CATs are a replica of TM11-12. The TM of glycoprotein SLC3 and YneM both binds the interface of TM4-5 in LATs and GkApcT.*

Several homolog in terms of sequence similarity of the GkApcT has been solved, including bacterial APC transporters AdiC (arginine/agmatine antiporter) (38-41), ApcT (H+-coupled amino acid transporter) (42) and GadC (glutamate/GABA antiporter) (43). These transporters all share the LeuT fold (44). Considering that among all transporters that share this structure fold, very little sequence identity is observed (<10%) (45) and given the abundance in the literature of the neurotransmitter sodium symporter (NSS) (27), we will use NSSs as examples for discussing the LeuT fold. The LeuT fold consists of an inverted repeat of five helices. The first five transmembrane helices are structurally similar to the next five transmembrane helices and is symmetric with respect to the plane of the membrane. These 10 helices form the LeuT fold and are the first 10 helices of the GkApcT, CATs and LATs (46) (Figure 1.10).

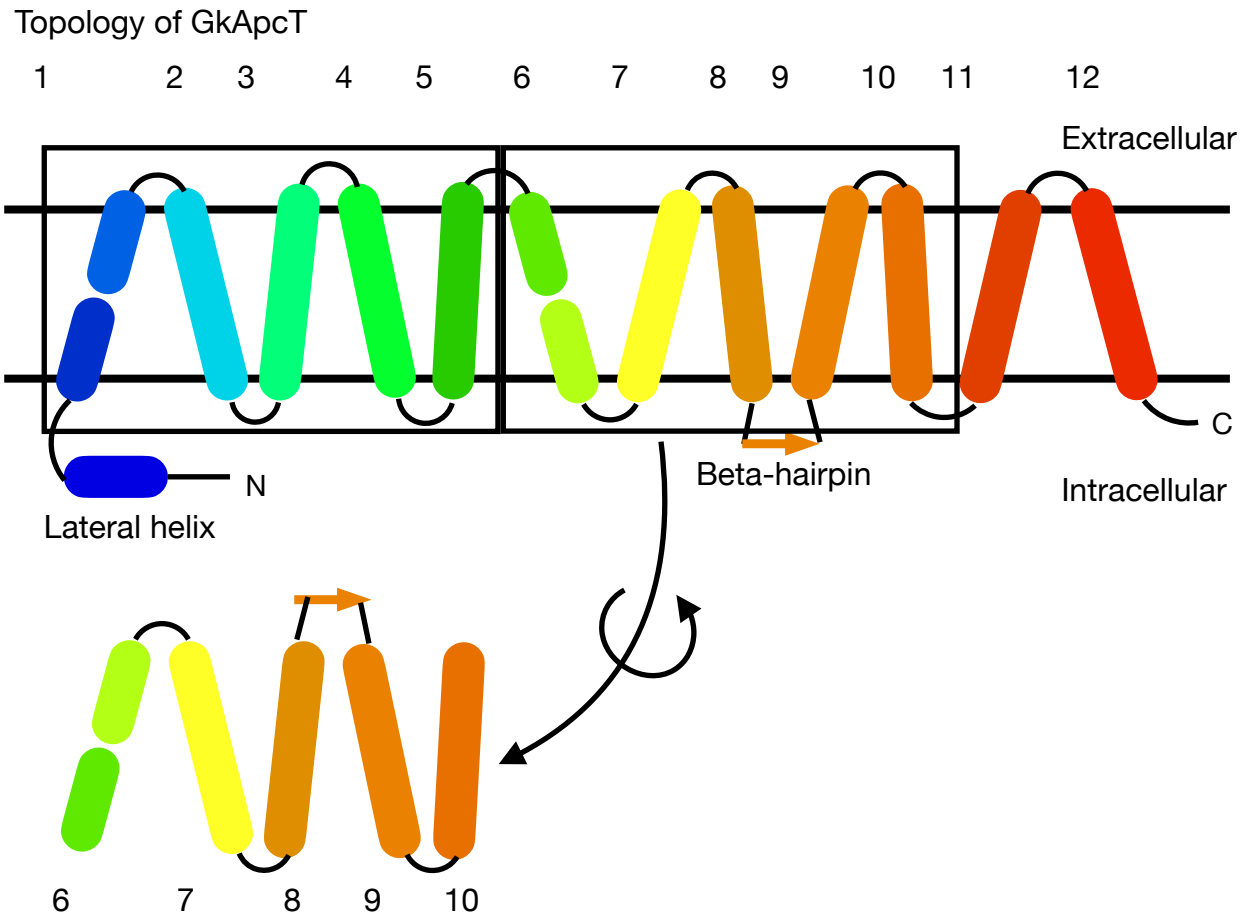


***Figure 1.10**: LeuT structure fold use GkApcT as an example. The canonical LeuT fold is composed of an inverted repeat of the five helices. In the GkApcT, the first 10 TMs follows the LeuT structural fold, where TM6-10 (second black square) is the inverted repeat of the TM1-5 (first black square). Figure adapted from (2).*

The transport cycle of a transporter sharing the LeuT fold could be generalised to these stages. In the resting state, the extracellular gate of the transporter is open with the intracellular gate being firmly closed, which is termed the outward-open state. The substrate as well as proton or any co-transported ion comes in through the extracellular gate and sits in the binding pocket. The extracellular gate then closes resulting in an occluded state. The intracellular gate opens up to allow the substrate and proton or any co-transported ion to be release to the intracellular solution, which is given the name of inward-open state. The transporter then performs a conformational change to restore the transporter to the outward-open state (Figure 1.11). The occluded state could also be further subdivided into inward-occluded state, which is when the extracellular gate is closed by interwinding helices and the intracellular gate closed by several residues and an outward-occluded state, which is the opposite of the inward-occluded state.

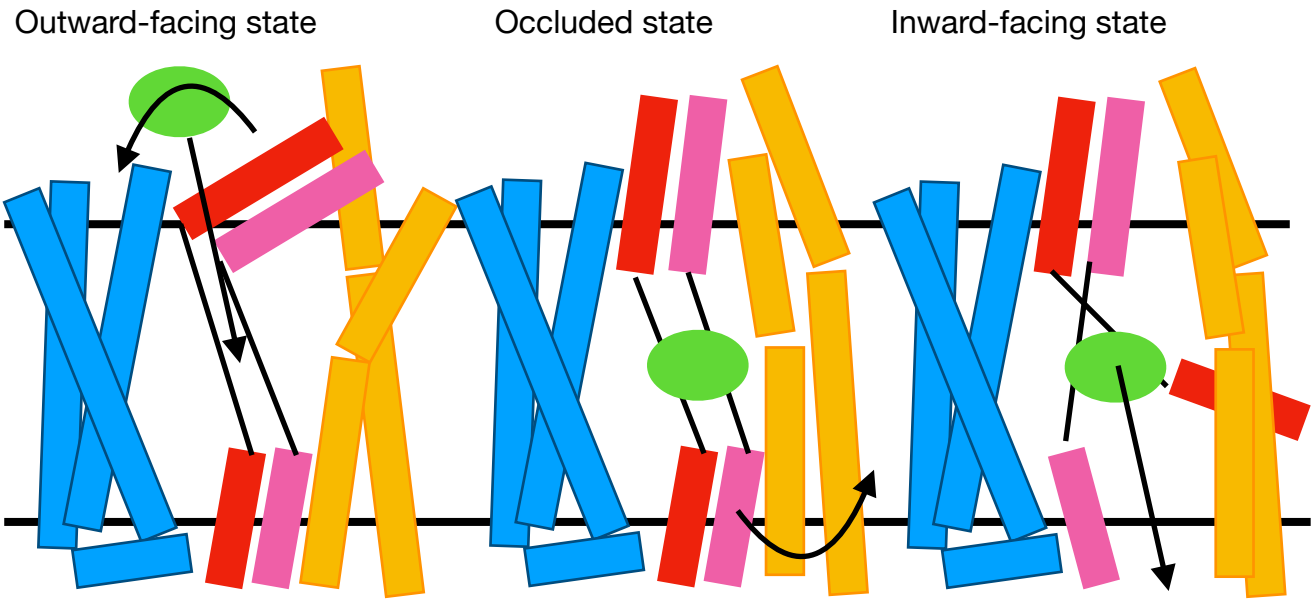


***Figure 1.11**: The transport cycle of rocking-bundle mechanism. During the transport cycle, the substrate (green) comes into the unwound region in the middle of TM1 (red) and TM6 (pink). TM1b,6a move toward the scaffold domain (blue) to close the extracellular gate (from the outward-facing state to the occluded state). The subsequent movement of TM1a,6b away from the scaffold domain opens the intracellular gate (from the occluded state to the Inward-facing state), which allows the substrate to exit. Figure adapted from (36).*

In the rocking-bundle mechanism, the opening and closing of the extracellular or intracellular gate are manifested mainly as the movement of TM1 and TM6. For the LeuT fold, the first 10 TMs could be conceptually divided into two parts, the scaffolding part, which is the TMs other than the TM1 and TM6 and is thought to stay still during the transport cycle. The unwound region in the middle of TM1 and TM6 divides these two TMs into two helices (a and b part) that can be independently moved and allow the binding of substrate. The opening and closing of the extracellular gate is considered as the tilting of TM1b and TM6a (80), which opens up to allow the extracellular solution to access the extracellular vestibule. Similarly the opening and closing of the intracellular gate is gated by TM1a and TM6b (Figure 1.11).

1.5 The Neurotransmitter Sodium Symporters (NSSs)

1.5.1 The Substrate Binding Pocket

The same structural fold suggests the same substrate recognition mechanism is shared between all the LeuT fold transporters. The binding pocket could be subdivided into three regions (Figure 1.12). The subregion A lines the bottom of the binding pocket. Since GkApcT is in inward-facing state, the bottom of the binding pocket is open. The subregion B is the main pocket hosting the side chain of the substrate and is lined by the scaffolding helices (TM3, 8 and 10). The subregion C is the unwound region in TM1 and TM6. The subregion C forms specific recognition site for the carboxyl terminal of amino acid substrate.

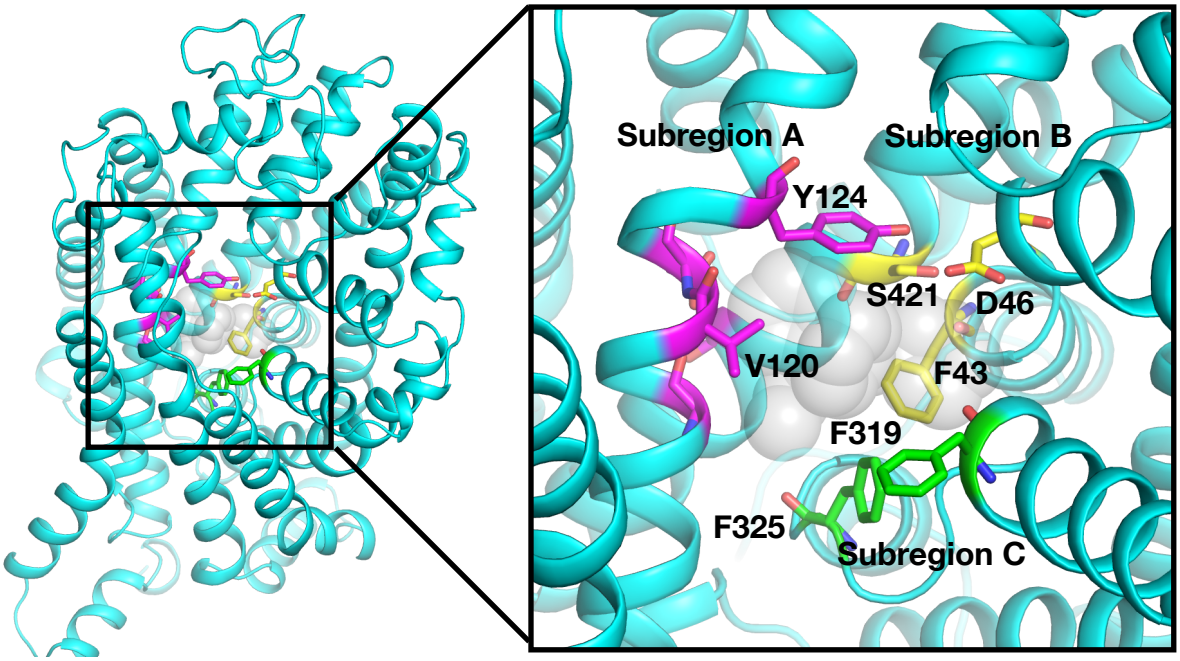


***Figure 1.12**: The binding pocket in the NSS transporters. The binding pocket of Drosophila melanogaster dopamine transporter (dDAT) could be divided into three subregion (A. View from extracellular side). The substrate is shown as transparent grey spheres. The subregion A (green) are the hydrophobic residues that form the intracellular gate. The subregion B (magenta) has the most variety and accommodates the side chain of the substrate. The subregion C (yellow) encompasses the carboxyl moiety of the substrate.*

Subregion B has the most diversity amongst all three subregions. The residues lining the subregion B were found to be crucial in terms of changing the specificity for different substrates. In the wild-type GkApcT, the binding pocket could only host alanine and the functional data has also shown that the transporter could only transport alanine. However, in the human CATs, one of the residues lining the binding pocket is a serine instead of the methionine as is observed in the GkApcT. Thus, mutating the methionine to serine in the GkApcT allows the transporter to transport arginine, which is the endogenous substrate of CATs (4) (Figure 1.13A). In LeuT, hydrophobic residues such as W104 and I359 forms the hydrophobic pocket for hosting the hydrophobic side chain of leucine and phenylalanine. In the case of phenylalanine, F259 forms pi-pi stacking with the aromatic ring of phenylalanine (47). The subregion B is relatively polar in dopamine transporter (DAT) (D121 and S426), which is thought to accommodate the catechol ring of dopamine. These polar residues are changed into A173 and L443 in serotonin transporter (SERT), which reflects the different substrate specificity (48) (Figure 1.13B). When the residues lining the binding pocket of the norepinephrine transporter (NET) are changed to the corresponding residues from SERT, the affinity for the SERT inhibitor S-citalopram is greatly enhanced. NET inhibitors such as nisoxetine and atomoxetine, are, on the other hand, less effective in this mutant (49). Similar effects has also been purposed in DAT and NET (50).

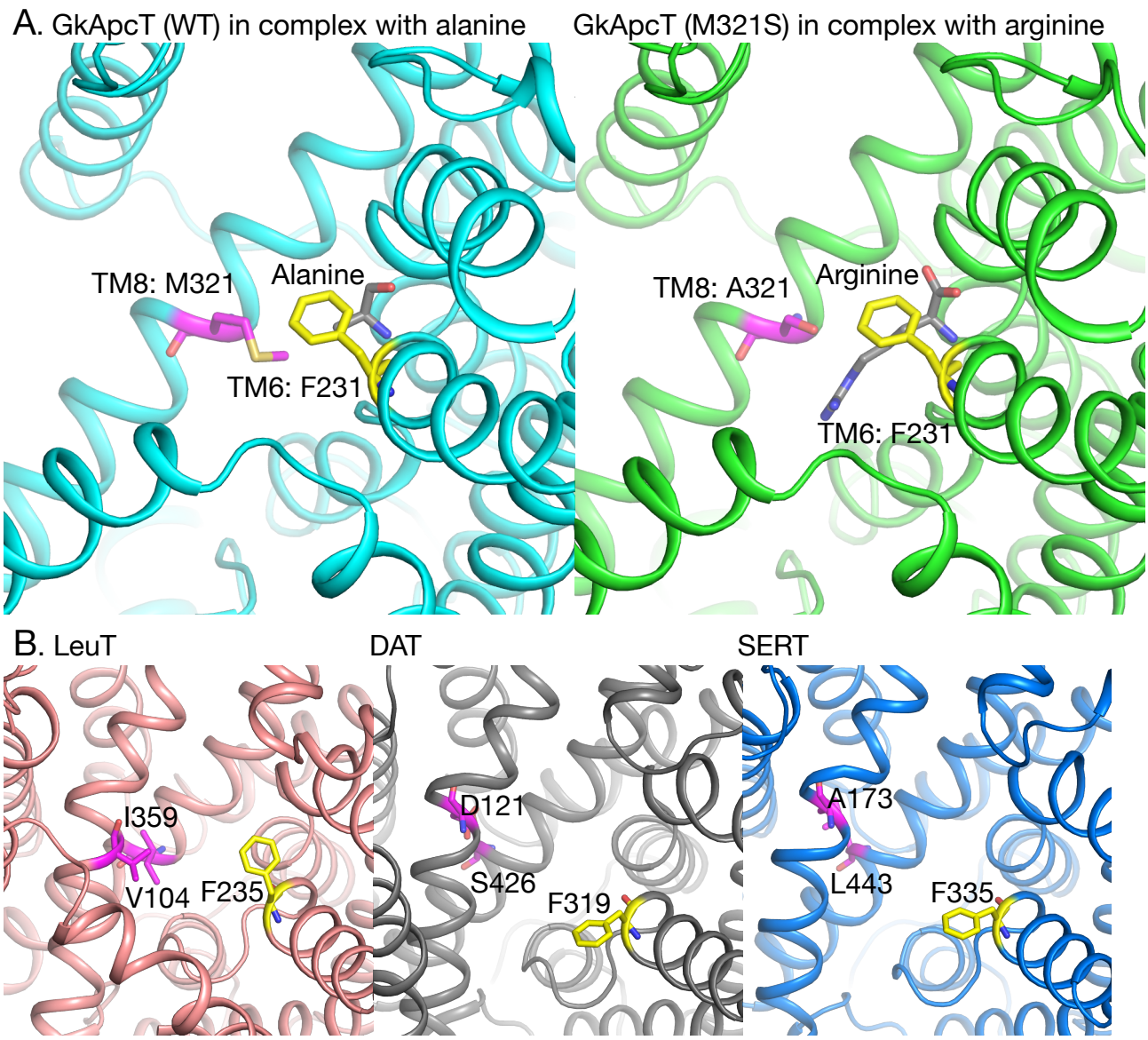


***Figure 1.13**: Subregion B/C in the binding pocket of the LeuT structural fold. A. In GkApcT, subregion B determines the substrate side-chain, whereas M321 (left magenta) would only permit substrate with small side chain such as alanine (left grey sticks), while for bulkier substrate such as arginine (right grey sticks), the residue has to change to serine (M321S). B. A key phenylalanine (yellow sticks) marks subregion C and is observed in GkApcT, LeuT, DAT and SERT. Similar to the GkApcT, LeuT, DAT and SERT all have different subregion B (magenta sticks) so as to differentiate substrates.*

The subregion C is the unwound part of the TM1 and TM6. From the three outward facing structures of LeuT in complex with Leucine (44) and AdiC in complex with arginine (40, 41), it seems that the α-amino carboxyl moiety of the substrate is always recognised by the backbone of the subregion C. A similar binding pose has also been observed in GkApcT (4). The subregion C is also marked by a conserved phenylalanine (F319 in DAT; F253 in LeuT; F335 in SERT) and was found to act as a lid to close solvent access to the binding pocket (Figure 1.12B). For transporters that transport substrate together with other ions, these ions were found to bind the subregion C as well.

Competitive inhibitors could bind the same place as the substrate in the primary binding pocket. By outcompeting the endogenous substrate, competitive inhibitors such as antidepressants and psychostimulants could slow down the transport of neurotransmitter (51). The competitive inhibitor could block the transport cycle by steric clash to prevent the closing up of the extracellular gate (52).

1.5.2 The Allosteric Binding Pocket

It has been suggested that the transport function could be regulated by substrate binding to an allosteric site. The antidepressants clomipramine, when bound to LeuT, block the binding of substrate or prevent the unbinding of the substrate in the primary binding pocket located in the unwinded region of TM1 and TM6 (53). Furthermore, the binding of leucine in the allosteric site has been proposed to be necessary for the release of the substrate from the main binding pocket (54). Computer simulations have shown that cocaine transiently binds the allosteric site before moving to the primary binding site (55). Though there are many physiological reports suggesting that there can be a non-competitive binder that influences transporter function, structural evidence for this allosteric site has been scarce. There has been argument that the specific allosteric site

might not exist and the allosteric effect is due to the non-specific binding. The reason that these non-specific binding sites are absent from the crystal structure could be that they are occupied by detergent in crystal structures (56), where the polar head of the detergent mimics the polar interaction between the protein and the N- or C-terminus of the leucine while the hydrophobic tail mimic the hydrophobic interaction between the hydrophobic side chain of leucine and the protein.

The only example of a concrete evidence for the allosteric site comes from S-citalopram in SERT (92). Biochemical evidence suggests that S-citalopram binds and inhibits SERT in a non-competitive fashion (57). Crystal structure shows that S-citalopram binds in an allosteric site coordinated by TM1b, 6a, 10 and 11 (58). S-citalopram sits between the extracellular bulk and the binding pocket by inserting the large benzofurancarbonitrile group between R104 (TM1b) and F335 (TM6a), which prevents the binding or the unbinding of the substrate. The R104 (TM1b), being unable to form salt bridge with E493 in the presence of S-citalopram, means that the conformational change of closing the extracellular gate is prohibited (59).

1.5.3 The Conformational Change

After the binding of the substrate, comes the closing of the extracellular gate. The extracellular vestibule is accessible from the extracellular solution and is lined with TM1b, TM6a and TM3, 8 and 10. The closing up of the extraceullar gate seems to be the coordinated motion of extracellular loops 2, 3 and 4 moving against the scaffold domain (60). For the transporters with an extracellular loop 4 (EL4), the EL4 is thought to act as a lid on the extracellular vestibule. The surface of the extracellular vestibule is mostly negatively charged to guide the cationic amino acid into the binding site (48).

The extracellular gate is thought to be gated by a salt bridge joining the TM1b and TM10. In LeuT, the salt bridge is formed by R30 and D404, which are R52 and D475 in DAT and R104 and E493 in SERT. Though this salt bridge is observed in the crystal structure of the LeuT (44) and SERT (58), the salt bridge is not observed in the DAT crystal structure (51). However, the crystal structure indicates that the residues are close enough that the salt bridge might readily form under some conditions.

Upon the binding of the substrate, the TM1b and TM6a moves inward which draws the EL4 in to close the bulk access to the extracellular vestibule. In SERT, the coupled motion between EL4 and TM1b is mediated by the pi-pi stacking between W103 (TM1b) and P403 (EL4). Computer simulations have shown that upon the binding of the dopamine to the binding pocket, the salt bridge between R52 and D475, which is not observed in the crystal structure, does form in the DAT (61). The coordinated movement of TM1b and TM6a moves the F319 to directly cover the extracellular side of the binding pocket. The hydrophobic nature of F319 is thought to be the gate which shuts the water access to the binding pocket.

Though it is widely accepted that the rearrangement of TM6b and TM1a opens the intracellular gate, there is still a question regarding which helix plays the dominant role.
In GkApcT, it is proposed that TM6b swings open and part of the project is to validate this hypothesis. However, in LeuT, it is found that the TM1a swings out by 40° forming a major helix rearrangement. In the occluded state, it is found that the tryptophan near the N-terminus (W8) interacts with Y265 and Y268 to stop the substrate from accessing the cytoplasm in the inward-occluded state (62). The displacement of Y95 as a result of TM1a tilting, opens the binding pocket and allows the solvent to access the binding pocket. This outward swing of the TM1a is observed in both LeuT (42) and SERT (63). In SERT,

the outward movement of the TM1a is facilitated by the unwinding of TM5. The intracellular gate is gated by W82 (TM1a), F88 (TM1a) and Y350 (TM6b) and, thus, the movement of W282 in TM5 allows the opening up of hydrophobic gate to allow the intracellular gate to open and the solvent to access the binding pocket (63). The unwinding of TM5 and the outward swing of TM1a is also observed in MhsT (64). For LeuT, partial unwinding of TM7 and TM6a are also observed along with the swing of TM1a and the unwinding of TM5.

1.5.4 Factors Determining Conformational Change

Numerous factors have been proposed to regulate the kinetics of transition and water is thought to be a key player. The shift from inward-occluded state to inward-open state is marked by the penetration of water from the intracellular side into the binding pocket (64). For LeuT, the release of the sodium and the reconfiguration of the intracellular gate is found to be coupled by microsecond-scale transient water penetration (65). During this process, the rearrangement of the sodium binding residue D421 is crucial (66). In a computational model of inward-facing LeuT, the sodium exiting as well as water penetration is induced by TM1 bending (67). This change of water network was found to be crucially related to W8. For DAT, computer simulation using Markov State Models shows that the conformational change has crucial correlation with the water penetration (68). Lipids such as the phosphatidylinositol 4,5-biphosphate (PIP_2) is thought to play a role in this connection network rearrangement (69).

Many other factors could influence the kinetics and equilibrium of transition between the inward-facing and the outward-facing state. Though GkApcT was found to be proton-coupled and the transport efficiency is enhanced by the proton gradient, the mammalian CATs are not found to operate differently in different pH conditions. However, there has

been a report that the plant CATs might exhibit pH-dependent behaviour (70). This echos a phylogenetic study where the SLC7 family are distantly related to the SLC36 proton coupled amino acid transporters family (71). Most of the LeuT fold transporters are otherwise influenced by the sodium concentration. For leucine transporter, at saturating sodium concentration, the inward-occluded state was found to be stabilised in the absence of the substrate (72). On the other hand, molecular dynamics simulation results show that the presence of the sodium stabilised the outward facing state and hinders the transition between outward facing and inward facing state. Thus, allowing the substrate to bind more conveniently when the sodium concentration gradient is present (73).

The presence of substrate is also found to play a key role in investigating the kinetics and equilibrium of different states. The kinetics of transition to open or occluded state are substrate dependent. For LeuT, substrate phenylalanine was found to accelerate the transition from occluded state to open state or from occluded state back to open state compared with substrate alanine (60).

The equilibrium between the inward-facing and the outward-facing state could also be influenced by various factors. The presence of substrate as well as the low sodium concentration seem to stabilise the inward-facing partially open state (72). This state could be converted to the occluded state via the addition of sodium. At saturating concentration, substrate phenylalanine was found to freeze the LeuT in an inward-open state compared with alanine (60).

Lipids have also been shown to modulate transporter function. For example, it is well known that cholesterol plays a key role in mediating transporter function (74). In the GkApcT crystal structure, cholesterol was found to bind to the transporter and influence

its thermostability (4). Similar binding of cholesterol has also been observed in the outward-open structure of SERT (63). Cholesterol was also found to stabilise the outward-open inhibitor bound state of DAT (75). Two cholesterol sites have been found in the DAT. In the first site, the cholesterol binds in the interface between TM5 and TM7 and close to TM1a (51). Interestingly, a cholesterol was found to bind to the same site in SERT (63). Close to the first site, cholesterol can also bind to the groove made by TM2, 7 and 10. In addition to these two sites found in the DAT, an additional site in SERT where cholesterol binds to TM12, can be found in the crystal structure (58). Two more cholesterol binding sites has been suggested by coarse-grain simulations (76).

The binding of cholesterol could also make some conformational states more energetically favourable. Amongst all the cholesterol sites discussed previously, the shared cholesterol binding site as is observed in SERT and DAT is found to be the most likely cholesterol binding site via computer simulations (77). If cholesterol were bound to this site, the movement of TM1a is thought to be limited due to steric clash. Thus, the binding of cholesterol could stabilise the inward-facing state and is demonstrated both biochemically (78) and computationally (76). The conformational change towards inward-facing state without cholesterol is thought to be artifactual due to the unwinding of TM5 (64).

1.5.5 The Role of the Termini

The N- and C-terminus of the transporter could also play an interesting role in the transport cycle. For the GkApcT transporter, preceding the TM1, there exists a cytosolic helix, which is not observed in other LeuT fold crystal structures. However, considering that the regions before the TM1 are usually truncated for the sake of crystallography, it is possible that this helix could also exist in other transporters. From computational

modelling, it is known that the N- and C-terminus could exist as α-helical and β-strand structures, as is shown in SERT (79) and hDAT (80). There is also physiological significance to the N- and C- tails. For example, the N-terminus of hDAT is known to interact with PIP_2 and could aid the amphetamine induced efflux (81). Furthermore, the activation of Ca^{2+}/calmodulin-dependent protein kinase II (CamKII) could result in the phosphorylation of the N-terminus of hDAT and enhance the amphetamine induced efflux (82). The N-terminus could also act as a tag which allows DAT, NET, SERT and GAT to interact with syntaxin-1A as a part of SNARE mediated exocytosis (83). Alternatively, the N-terminus could be phosphorylated by protein kinase C, which resulted in endocytosis (84). Among different crystal structures of the N-terminus in inward facing state, some inconsistency is noted. In the MhsT, the N-terminus binds to a tryptophan which is located in a region similar to the W8 in LeuT (64). This binding of the N-terminus to a tryptophan is also observed in the outward facing LeuT. However, in the inward facing state, the N-terminus cannot be resolved from the electron density, signalling that the N-terminus is disordered and plays no significant structural role.

The transport function could also be regulated by glycosylation. In LeuT, the loop between TM3 and TM4 is extraordinarily long and contains various glycosylation sites (58). For DATs, this loop could even function as a zinc binding site (85) and the binding of zinc is shown to inhibit the transport (86). Since GkApcT is a bacterial transporter, bacteria cannot provide the necessary machinery for glycosylation.

Oligomerization is also thought to play a role in the regulation of the LeuT type transporter. The dopamine transporter was found to form a trimer to modulate the dopamine signalling (87).

1.6 Protein Trafficking in the Cell

1.6.1 Cellular Distribution

Within the cell, many organelles exists, such as the nucleus, the endoplasmic reticulum (ER) and the Golgi apparatus. Each of these organelles has their own assigned function sand, thus, has their own set of proteins to facilitate them performing these functions. The distribution of organelles and the composition of their membranes can quickly become asymmetric under scenarios such as nutrient-depletion based reposition of lysosomes for autophagy (88), polarisation of immune cells for secretory immune response (89) or cell migration toward ‘attractive clue’ (90). It is critical that, in these processes, the proteins crucial for the functioning of these organelles are localised in their corresponding locations during this dynamic reorganisation.

For secretory cells, the synthesis of new proteins involves the functioning of multiple organelles. DNA in the nucleus are transcribed to RNA. The RNA is then translated to nascent peptides by ribosomes and translocated to the ER. In the ER, various chaperones helped the folding of the nascent peptide chain into proteins. The correctly folded proteins are then transported to the Golgi apparatus to perform post-translational modifications, which are then secreted either to the plasma membrane or to their corresponding intracellular locations. During this bulk flow, the organelles are continuously undergoing exchange with other organelles via various vesicular and tubular transport systems. It is estimated that over one-fifth of all the mammalian proteins originates from the ER and is transported to their destination via various vesicular systems (91).

The nascent peptide chain translated by the ribosome enters the ER via translocation. This process is mediated by Sec61 channel where the signalling peptides from the nascent peptide chain are recognised by the signal recognition particle (SRP) receptor, which target the peptide chain to the Sec61 channel. This process is facilitated by the multi-subunit ER membrane protein complex (EMC), which helps the insertion of transmembrane helices into the ER membrane. The driving force for this insertion into the ER is provided by binding immunoglobulin protein (BiP) after the insertion of the signalling peptide into the translocon. Other peptides where the signalling peptide is at the tail, are inserted into the ER via the GET/TRC complex. The signalling transmembrane helix is recognised by the Get3, which is known as TRC40 in mammals. The complex binds the Get1-Get2 receptor complex, (known as WRB/CAML in mammals) and is then inserted into the ER membrane. Though membrane proteins are usually co-translated into the ER with the help of Sec61 channel, other proteins are translocated into the ER with the help of proteins such as calmodulin, HSP71 family chaperones and small glutamine-rich tetratricopeptide repeat-containing protein alpha (SGTA). These cytoplasmic chaperones prevents the folding of the peptide chain before the insertion into the ER to prevent misfolding events.

1.6.2 ER-Golgi Trafficking

The role of the ER is to make sure that only correctly-folded proteins can be exported to the Golgi for further modification. The misfolded proteins are targeted for ubiquitination and degradation or ER retrotranslocation. Proteins which have passed the ER quality control are exported to the Golgi via coated protein complex II (COPII) transport vesicles. One example of such quality control is the calnexin/calreticulin cycle. For the protein which is not in the correctly folded form, the N-terminal of the protein is glycosylated by UDP-glucose/glycoprotein glucosyl transferase (UGGT). The glycosylation is recognised

by the lectins calnexin and calreticulin, which form a complex with the unfolded protein and hold the unfolded protein in the ER. The glucosidase II will gradually trim the N-terminal glycosylation and when the innermost glucose is being trimmed, the protein is released from the calnexin/calreticulin. The correctly folded protein is exported to the Golgi while the incorrectly folded protein is glycosylated again to enter the cycle. The terminal mannose in the proteins which are persistently incorrectly folded are trimmed by ER degradation-enhancing α-mannosidase-like protein (EDEM), while leads to the protein being targeted for ER associated degradation (ERAD) (Figure 1.14) (92).

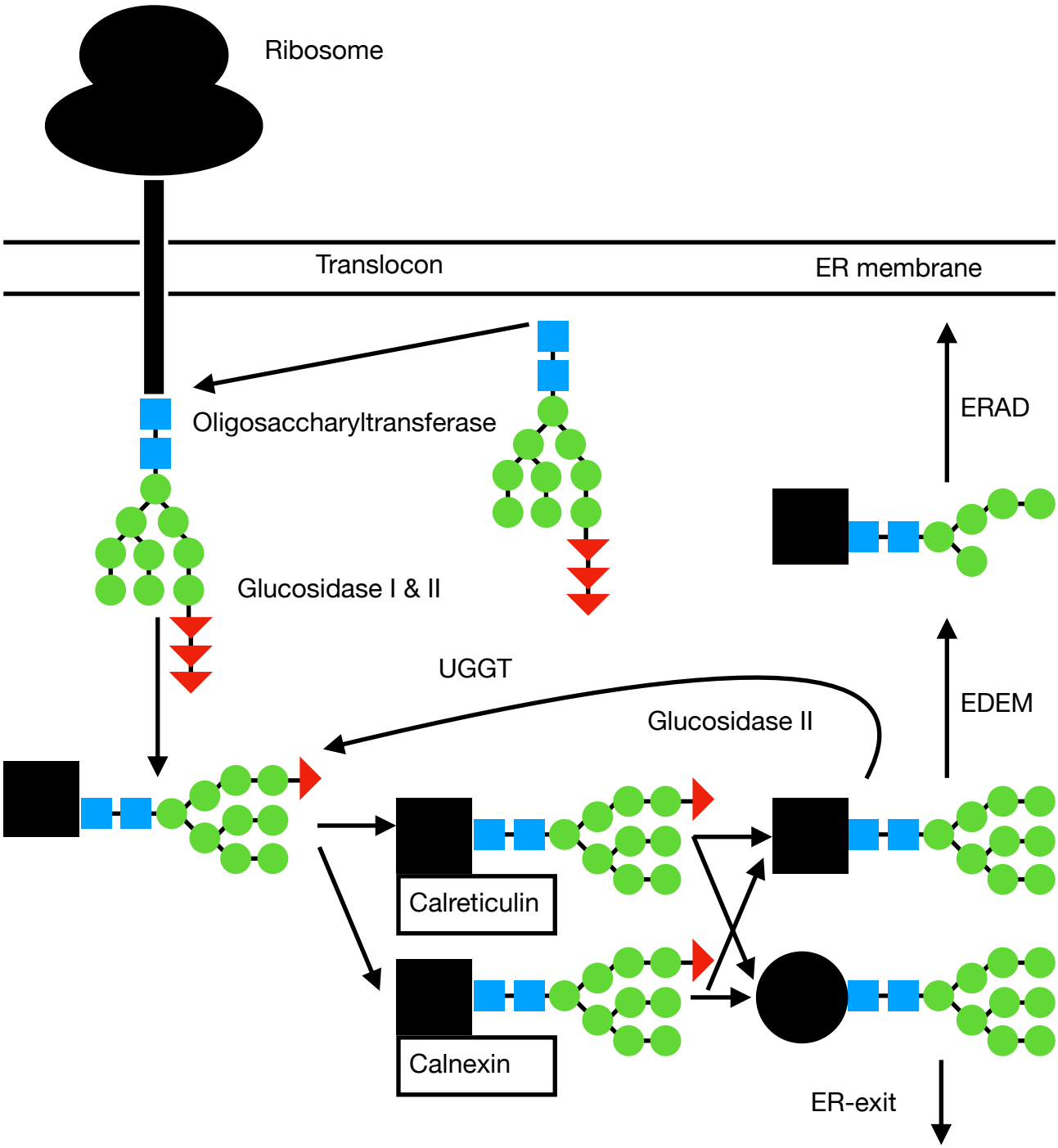


***Figure 1.14**: The calnexin/calreticulin cycle of protein quality control. The nascent peptide chain are glycosylated by oligosaccharyltransferase, where the glucose (red triangle) are trimmed by glucosidase I/II. During the folding process, the incorrectly folded protein (black square) forms a complex with calnexin or calreticulin. After the final glucose is trimmed by glucosidase II, the correctly folded protein (black circles) are sent to the ER, while the incorrectly folded protein are reglycosylated by UGGT. Alternatively, the terminal mannose (green circle) are being trimmed by EDEM, which lead to ERAD. Figure adapted from Philipp Bräuer's thesis.*

The COPII vesicles mediates the ER to the Golgi trafficking and the ER to the ER-Golgi intermediate compartment (ERGIC) trafficking. The COPII vesicles has a double coated structure, where the outer coat is the SEC13-SEC31 and the inner coat is the SEC23-SEC24. The initiation of the COPII vesicle is mediated by the SAR1-Guanine nucleotide Exchange Factor (GEF) (SEC12), which activates SAR1 in the ER exit sites (ERES). The activated GTP-bound SAR1 recruits the SEC23-24, which further recruits SEC13-31. The COPII vesicle recognises various sorting signals including the di-acidic ([DE]x[DE]), the di-hydrophobic and di-aromatic signal, which binds to SEC24.

During this bulk flow, a significant portion of the ER are moved to the Golgi via the COPII vesicles. It is estimated that 1.5~5% of the ER volume are lost to the Golgi, which accounts for 5-10% of the Golgi volume every minute (93). In addition to the protein synthesis, the lipid synthesis is also an important role of the ER, where the majority of cholesterol is synthesised. However, it is noted that the cholesterol concentration in the ER is relatively low (94). During this massive flow, crucial ER resident proteins such as the ER chaperones can escape into the Golgi.

1.6.3 The KDEL Receptor

During the ER-Golgi trafficking some ER proteins will follow the bulk flow and enter the ER. Some of these ER resident proteins are retrieved back to the ER via the KDEL receptor (95). The KDEL receptor was initially found in the yeast as one of two ER retention deficient mutants (96). The ER chaperones, intended to be recognised by the KDEL receptor, hosted an ER retention signal (ERS) at their C-terminus. In mammalian cells, this ER retention signal has the sequence of lysine, aspartate, glutamate and leucine and thus, is usually known as the KDEL tag. Attaching a KDEL tag to the non-ER resident protein such as cathepsin D, allows cathepsin D to be retrieved back to the ER (97). It is,

however, recognised that other sequences can also be recognised by the KDEL receptor as well, such as the calreticulin which has an HDEL tag (98), where the lysine is substituted by a histidine. It has been argued that the different ERS (KDEL/HDEL) might illicit different responses.

In *Homo sapiens*, there exists three different KDEL receptors, namely KDELR1/2/3. The KDELR1 preferentially binds the KDEL tag, while the KDELR2 binds preferentially to HXEL where X can be any hydrophobic amino acid and KDELR3 binds preferentially to HDEL (95, 99). KDELR1 is expressed most abundantly, while KDELR2 is transcribed in a higher quantity compared with KDELR3 (99). KDEL receptor is also found in other organisms as well and preferentially binds HDEL in *Saccharomyces cerevisiae* (100) and yeast (96). Other KDEL-like tags such as the RTDL in mesencephalic astrocyte-derived neurotrophic factor and [Q/K]TEL in cerebral dopamine neurotrophic factor can be found in neuron cells (101).

Though it is well accepted that the last four residues play a key role in the binding of protein to the KDEL receptor, it has also been argued that the residues preceding the last four residues (HDEL or KDEL), such as the -5 and -6 position could also affect the binding affinity of the ERS to the KDEL receptor (102) and the sequence FEHDEL was found to be strongest binder to the yeast KDEL receptor (103). The specificity of the -5 and -6 position is thought to be mediated by hydrophobic interactions on the surface of the KDEL receptor (104). There are also claims that aside from the several residues in the C-terminus of the protein, other residues in the protein might play a role in the interactions with the KDEL receptor as well. Human myeloid-derived growth factor (hMYDGF) was found to have shape complementarity with the KDEL receptor and is hypothesised to interact with the KDEL receptor though the interactions in the protein-protein interface

(105). However, it is also argued that the hMYDGF is an outlier as hMYDGF hosted RTEL in the C-terminus instead of the normal KDEL and HDEL. The deviation from these cardinal sequences are then compensated with protein-protein interactions, as a short peptide was found to be enough to achieve tight binding of the KDEL receptor and the retrieval to the ER (106).

1.6.4 The Coat Protein Complex I (COPI) Vesicles

The KDEL-receptor dependent retrieval of proteins hosting the ERS to the ER is mediated by the COPI vesicles (Figure 1.15). At the periphery of Golgi cisternae, the bundling of the KDEL receptor to the COPI vesicle begins with the recruitment of cytosolic COPI complex to the membrane by GTPase ADP ribosylation factor 1 (Arf1). Arf1 is activated when the GDP in the Arf1 is replaced with GTP by Arf guanine nucleotide exchange factors (GEF) (107). The activated Arf1 inserts its myristoylated amphipathic N-terminal helix into the lipid bilayer to anchor itself to the lipid membrane (108). The COPI complex is a heptameric coatomer, which is consisted of seven subunits (α-COP, β'-COP, ε-COP, β-COP, δ-COP, γ-COP and ζ-COP), where the γ-COP was found to interact directly with ArfGEF guanine nucleotide exchange factor 1 (GBF1) (109). Since the ArfGEF GBF1 is localised in the cis-Golgi, ArfGEF GBF1 assembles the activated Arf1 as well as the heptameric coatomer COPI in cis-Golgi (109).

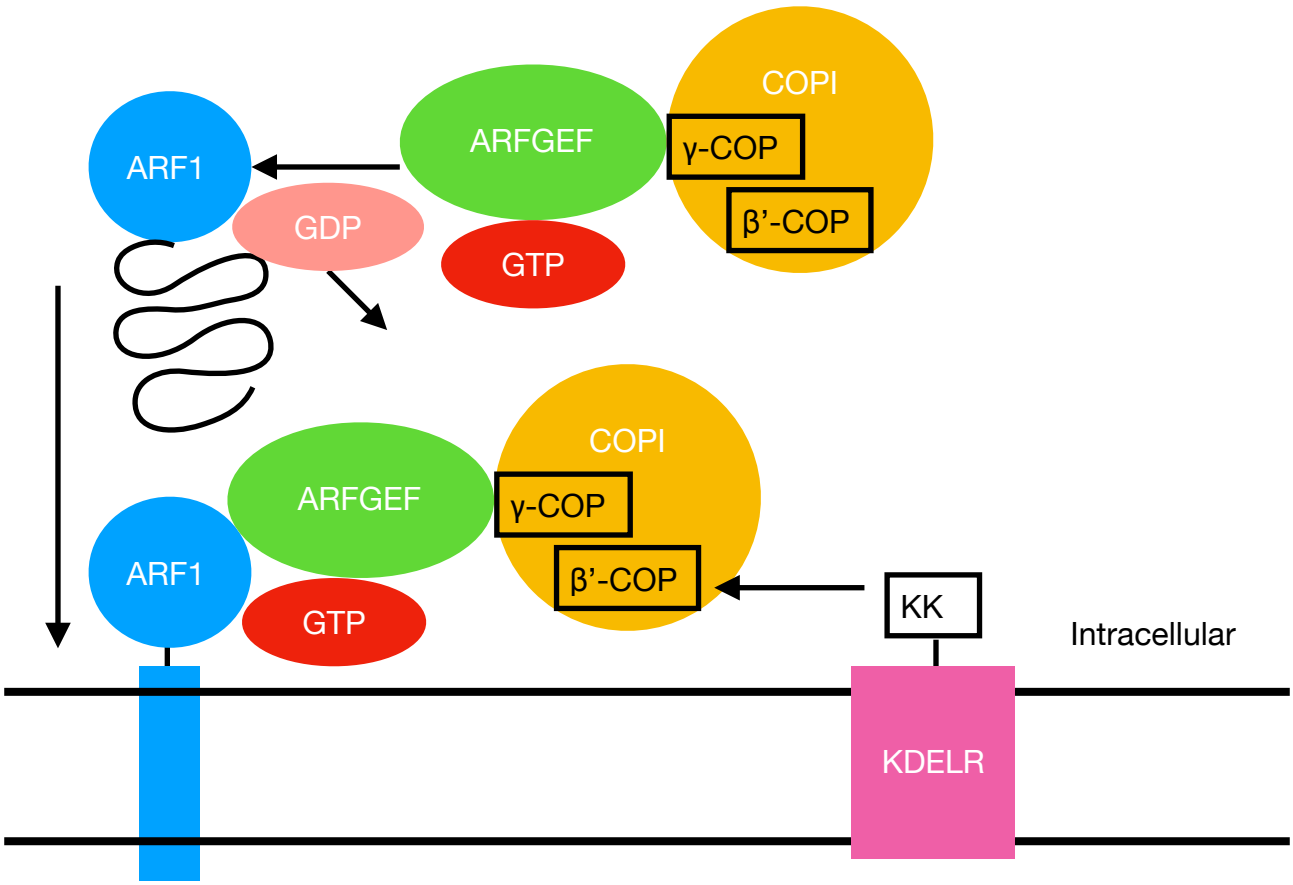


***Figure 1.15**: The recruitment of KDELR to the COPI. Upon the binding of ARFGEF (green), the GDP-bound (pink) ARF1 (blue) becomes GTP-bound (red) ARF1, which allow the ARF1 to form a TM that are embedded into the membrane. Thus, the COPI (orange) which is in complex with the ARFGEF via the γ-COP are bought close to the membrane. The membrane protein KDELR (pink) is then able to interact with COPI via the β'-COP through the di-lysine tag (YY).*

When the COPI heptameric coatomer is brought close to the membrane by the activated Arf1, COPI can bind various membrane proteins. One of the signature binding sequences is the di-lysine ER retrieval signal KKxx or KxKxx motif, as is seen in many ER proteins. The di-lysine ER retrieval signal binds the propeller domain of α-COP or β'-COP of the COPI complex (110). There are also proteins which host the di-lysine sequence but are not themselves ER resident proteins. These proteins act as adaptors to link the ER resident protein to the COPI vesicles for retrograde ER transport (e.g. Rer1, Vps74 and the Erv41-46 complex (111)). Among which, an important example is the KDEL receptor, which was found to be important in the retrieval of ER protein from Golgi. The β'-COP propeller can also recognise other tags such as SNARE Snc1 (112) or arginine-based ER retrieval signals ϕRxR where ϕ represents any hydrophobic amino acid (113). P24 protein family, which hosted the FFxxBB(x)n motif is another binding partner of the COPI vesicle and was found to regulate the interplay between Arf1 and COPI (114).

After the COPI heptameric coatomer binds the "cargo" proteins, COPI and Arf1 induces the curvature of the membrane to allow the vesicle to bundle off. COPI polymerises by creating flexibly linked units (115). Each unit is composed of three heptamers forming a triad (116). The units are linked together via either dimeric or trimetric interactions (117). The two-dimensional lattice of COPI triad generates positive member curvature (118). However, it was found that for the vesicle formation, Arf1 and COPI are both required (119). The completion of a COPI vesicle is aided by Arf1 dimerization (120), which scissors the COPI vesicle from the Golgi.

At the end of the vesicle formation, the COPI heptameric coatomer and Arf1 needs to be dissociated from the vesicle, so that they can facilitate the formation of the next COPI vesicle. Similar to the association of heptameric coatomer to the membrane, the

dissociation of heptameric coatomer from the membrane is also mediated by Arf1. Since the GTP is required for the Arf1 to attach to the membrane, a bound GDP will allow Arf1 to dissociate from the membrane (121). The activation of the GTPase activity of Arf1 is regulated by Arf GTPase-activating proteins (GAPs) (122), which are recruited by COPI vesicles (108) to catalyse the GTP hydrolysis (117). To ensure the GTPase activity is only activated upon the formation of a complete vesicle, ArfGAP1 hosted a lipid-packing sensor (123), which stimulates the GTPase activity in response to the increase in membrane curvature (124). The KDEL receptor was also found to affect the activity of ArfGAP1 (125).

1.6.5 The Structure and Function of KDEL Receptor

The cargo proteins are then release after the fusion of COPI vesicles with ER and the KDEL receptor is then transported back to the Golgi via COPII vesicles (126). Thus, the KDEL receptor takes up the cargo protein in the Golgi and releases them in the ER. It is hypothesised that this functional differentiation is based on the different pH between the Golgi and the ER. The pH in the cis-Golgi of the mammalian cell is considered to be 6.5 (127). In the ER, the pH is considered to be 7.2-7.4 (128). The binding of the KDEL ligand was also found to be pH dependent, as maximum binding was found to occur at pH 6 (129). The pH dependence is thought to be modulated by a key histidine in the binding pocket of the KDEL receptor (130). However, it is also argued that the residues in the protein bearing the KDEL tag can affect this pH dependence. For example, hMYDGF hosted some protonatable residues at the protein-protein interface. The different protonation states of these residues under different pH might explain the pH dependent binding (131). On the other hand, it is also argued that the pH difference (proton gradient) between Golgi and ER is more important than the absolute pH of the Golgi or the ER (132).

The binding of the KDEL tag activates the KDEL receptor. A conformational change is observed upon the binding where the Helix 7 changes from the straightened conformation to the kinked conformation (133) (Figure 1.16). This helix 7 hosted a key residue D193, where the D193N mutant was found to affect the G_S protein and cAMP-dependent protein kinase A activity and impair the COPI vesicle dependant transport (125). The activated KDEL receptor is also thought to undergo oligomerisation upon KDEL binding (134), which exposes the di-lysine tag so that the complex could interact with the COPI vesicle.

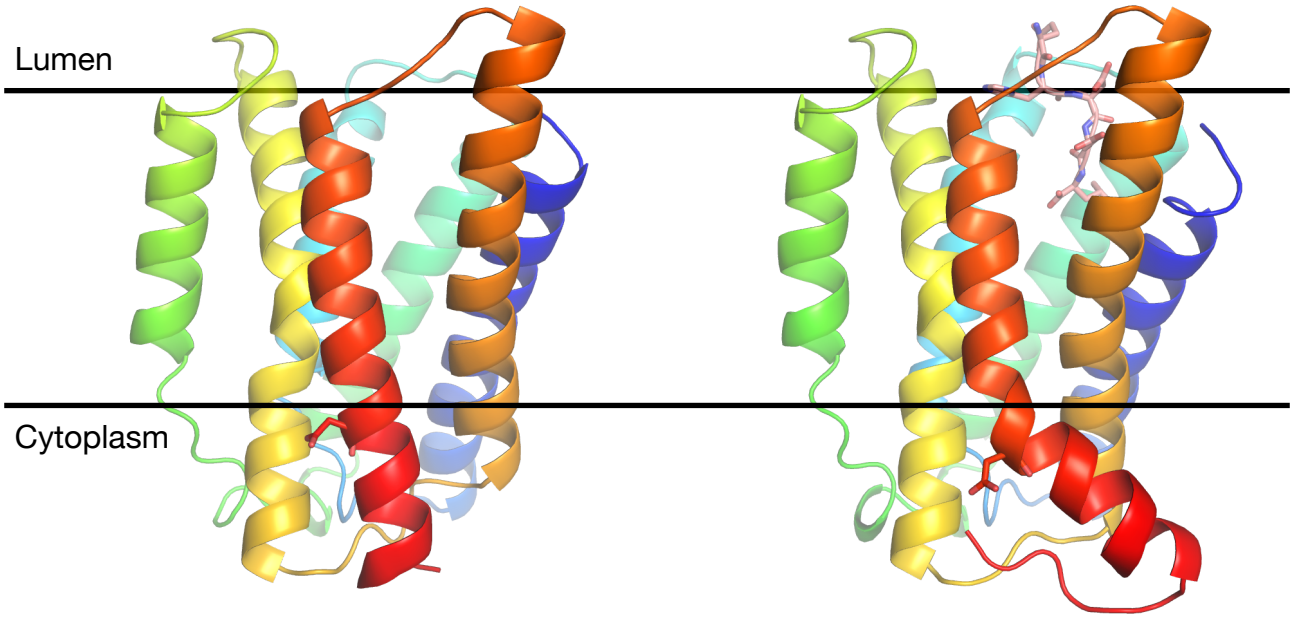


***Figure 1.16**: Ligand binding induced conformational change. Upon the binding of the KDEL ligand (salmon sticks), the helix 7 (red helix) underwent a conformational change from the straight form (left: apo state (PDB: 6I6B)) to the kinked form (right: KDEL-bound form (PDB: 6I6H)). The residue 193 (red sticks) is thought to play an important role in this conformational change.*

Under steady state, the KDEL receptor is thought to be concentrated in the cis-Golgi (135), though there are reports that KDEL receptors can be observed in ER-Golgi-Intermediate compartment (ERGIC), ER (136, 137) and plasma membrane (138). The presence of the KDELR in the plasma membrane was found to be an important entry point for viral infection, where functional KDELR is required for the uptake of A/B toxins such as cholera toxin (138), pseudomonas exotoxin (PE) and killer toxin K28. Interestingly, though cholera toxin and killer toxin K28 both contain the KDEL sequence in the C-terminus, the killer toxin K28 contains REDLK in the C-terminus (138) and, thus, renders it unable to bind the KDELR. It is hypothesised that the final residue is cleaved in the blood plasma, which enables it to enter the cell through KDELR (139). Replacing the REDLK sequence with the KDEL sequence is found to enhance the toxicity (140).

High abundance of the KDEL receptor in the ER can only be observed under artificial conditions, such as the over-expression of the KDEL receptor (136), the KDEL ligand (141) or excess doping with proteins (e.g. cholera toxin) with KDEL tag (142). This localisation was found to be dependent upon a functional C-terminus, which hosts the di-lysine tag (143). During the bulk flow, the KDEL receptor might follow the bulk flow from the cis-Golgi to the trans-Golgi. It is hypothesised that the COPI vesicle mediated retrograde transport can also bring the KDEL receptor back to the cis-Golgi in a fashion similar to the retrieval of glycosyl transferases. This process of retrograde transport of the proteins from trans-Golgi to cis-Golgi is thought to be a feature of the Golgi maturation (144).

The functioning of the KDEL receptor is mediated by many proteins, such as the P24 protein family (145). P24 protein family are involved in the regulation of protein trafficking and was found to be packed in the COPI- and COPII-vesicles together with the KDEL receptor (146). The loss of P24 protein was also found to be associated with reduced

KDEL receptor functioning (147), whereas the increase in the function of P24 protein allows more KDEL receptor to be found in the ER (148). Further examples include the protein-tyrosine kinase (SRC). When the SRC is activated, it localises to the Golgi (149) and can disperse the KDEL receptor leading to reduced KDEL-dependent retrograde transport (150). The Gαo-subunit of the heterotrimeric G proteins has also been found to be localised in the Golgi and interacts with the KDEL receptor, which is thought to affect the Rab1/3 signalling process (151).

In additional to fetching the ER proteins back to the ER, the KDEL receptor also mediates many crucial homeostasis functions, such as the unfolded protein response (UPR). When ER is under stress such as hypoxia, ischemia, malnutrition or mutation, the folding process might be affected and the incorrectly-folded or unfolded protein might escape into the Golgi (152). To ensure that all unfolded proteins produced during the ER stress are properly dealt with, the immunoglobulin heavy chain-binding protein (BiP) proteins are secreted from ER and Golgi to bind the unfolded protein in the ER to prevent peptide aggregation (153). During the ER stress, unfolded or misfolded protein escaped to the Golgi with their bound BiP. Since BiP hosted the KDEL tag, the unfolded or misfolded protein in the Golgi are recognised by the KDEL receptor through the KDEL tag in the BiP. Thus, the unfolded or misfolded protein are brought back to the ER via COPI vesicles to be degraded (154).

The KDELR also performs the role of returning unassembled oligomeric proteins back to the ER. Unassembled TCRα (T-cell receptor) can sometimes escape to the Golgi, where it should be targeted for ER-associated degradation (ERAD). Since these proteins are not ER resident and do not host the KDEL tag, their return is facilitated by the ERp44. Similar to the KDELR, the ERp44 binds proteins in a pH dependent manner and itself hosts a

KDEL tag (RDEL). Thus, the ERp44 acts as an adaptor to transport unassembled oligomeric proteins back to the ER (155).

The KDEL receptor mediated UPR response extends beyond simply bringing the misfolded and unfolded protein back to the ER and includes increasing the ER chaperones to help folding (156), suppressing the synthesis of new proteins (157) and catching the unfolded protein to degrade them (130). During the UPR, the KDELR2 and KDELR3 are upregulated by XBP1 to cope with the increased demand for KDEL functionality (158). Furthermore, KDEL receptor could also regulate the UPR by interacting with PP1 and EIF2A. Defective KDELR interaction with PP1 was found to lead to prolonged phosphorylation of EIF2A, which would result in apoptosis (159). The activation of the KDEL receptor could lead to the activation of $G\alpha_{q/11}$ proteins (160), leading to the activation of Src kinase and cell division cycle 42 (CDC42) and eventually the enhancement of the retrograde transport (161). To compensate for the loss of Golgi membrane to the ER, Src activation was also found to enhance the anterograde transport via G_q mediated activation (160). This enhancement of the anterograde transport via SRC activation is also mediated by prohibitin-1 (PHB). PHB is also found to be important for the anterograde transport of KDEL receptor back to the Golgi, whereas lack of PHB resulted in KDEL receptor concentrated in the ER (162). KDEL receptor was also found to regulate the activity of lysosome to facilitate the degradation of unfolded protein (163).

The KDEL receptor can also regulate protein trafficking in the sense that it could allow the ER resident proteins to be exported to the plasma membrane in a controlled fashion. For example, the BiP/HSPA5 complex was found to be secreted to the cell surface in response to viral entry (164). Other ER chaperones such as calreticulin and protein

disulfide isomerase (PDI) were found to be mediating the immune response and cell to cell adhesion (165)(166).

The importance of the KDEL receptor is recognised not only in physiological conditions. In the field of recombinant protein production, it is common to overexpress the ER chaperones (e.g. calreticulin/calnexin (167), PDI family member A1 (168), binding immunoglobulin protein (BiP) (169)) to increase the yield (170). These over-expressed ER chaperones, if left unchecked, will migrate to the cell surface and dampen the effect of over expression (171). Thus, a functional KDEL receptor, which ensures the localisation of these ER chaperones, is required for bioengineering purposes as well. Since the presence of the KDEL tag will enable the KDEL-tagged protein to concentrate in the ER, the KDEL tag is used widely when one wants to target an arbitrary protein to the ER. The interaction between the chemokine receptor CXCR4 and the ligand SDF-1 is thought to be important for the migration, invasion, and metastasis of cancer cells. Knocking out the wild-type SDF-1 and replacing with the KDEL-tagged SDF-1 (172) was found to block the metastasis of cancer cells (173).

1.6.6 Structural Similarity and Evolution relationship

Evolutionarily, the KDEL receptor belongs to the PQ-loop family, which is a large and diverse family of membrane proteins and is thought to play a key role in nutrient transport, lysosomal transport and tissue development. The PQ-loop family contains amino acid transporters such as cystinosin (a lysosomal cystine proton symporter) and PQLC2 (a lysosomal cationic amino acid exporter). Further examples include receptors in protein trafficking such as the KDEL receptor. The cystinosin is an interesting example of the PQ-loop family as it contains both amino acid transport activity and is thought to play a role in protein trafficking (174). Similar examples included Human SWEET1, which could

transport glucose and has a similar role in protein trafficking (28). In the PQ-loop family, there are also “transporters” which don’t transport amino acids, such as the phospholipid flippase Cfs1p (175).

Considering the diverse nature of the proteins in the PQ-loop family, one thing that is shared by all the proteins in the PQ-loop family is that they are all localised in the organelle membrane instead of the plasma membrane (176). Thus, it has been hypothesised that the role of the PQ-loop family proteins are all related to inter-organelle protein or amino acid or lipid trafficking. However, it is still unclear as to the role of the PQ-loop in these functions.

Though there are some orphan genes that share higher sequence similarity to the KDELR, the closest protein with known function is the SWEET sugar transporters, which share 25% sequence identity to KDELR (116). The structure of a bacterial homolog of the SWEET has been solved and is also structurally similar to the KDELR (28). The KDELR and SWEET both belong to the transporter-opsin-G protein-coupled receptor (TOG) superfamily, which includes the the PQ-loop family, lysosomal cystine transporter family, nicotinamide ribonucleoside uptake permease family and the GPCRs (177). Proteins in this superfamily are topologically similar in that they all contain 7-8 transmembrane helices. The bacterial members of the TOG family usually have the structurally similar 6 TM fold, where the first 3 TMs are the repeat of the next 3 TMs. The eukaryotic members on the other hand, have one more TM between the 3 TM repeats and thus, have the 3+1+3 topology, where the first 3 TMs are the repeat of the last 3 TMs. This 3 TM repeat structural fold has also been observed for the structurally similar KDELR and the SWEETs (Figure 1.17).

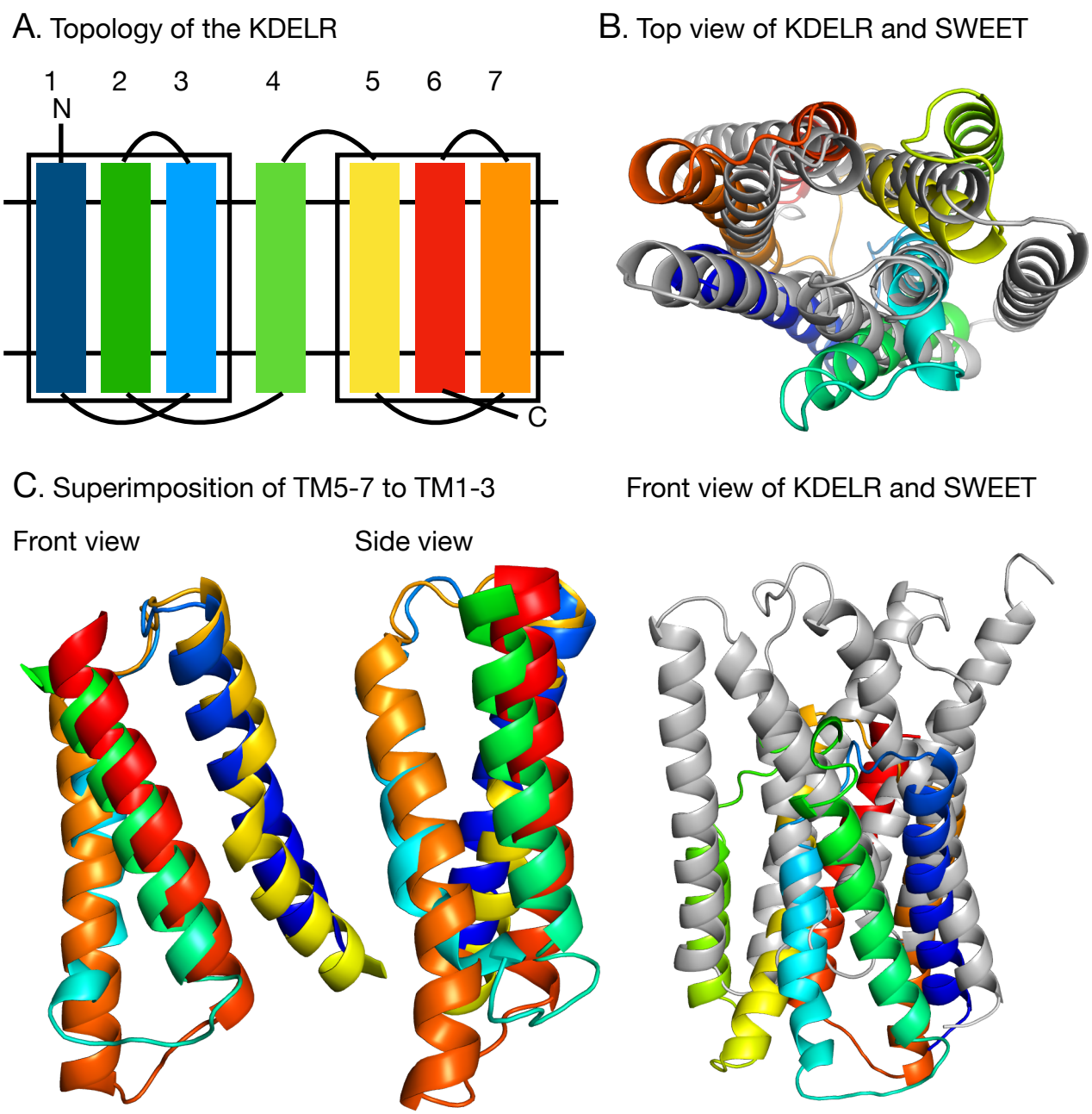


***Figure 1.17**: Structure comparison of KDELR and SWEET. A. The general topology of the 7 TMs KDELR follows the 3+1+3 pattern, where the first three helices TM 1-3 is the inverted repeat of the last three TMs (TM 5-7). B. The key helices of the KDELR (rainbow) could be superimposed on to the SWEET (grey), which shows structural similarity. C. Internal structure similarity is also being observed between TM1-3 and TM5-7.*

1.7 Aims and Hypothesis

Protons play a key role in amino acid transport and protein trafficking. However, the atomic details of how protons trigger these processes remain elusive. In this thesis, I'm interested in elucidating the molecular mechanism behind the proton-coupling of the GkApcT and KDEL receptor.

For GkApcT, the key question has been what role protons have in the transport cycle. This question is addressed by the following small themes.

- Which residue is responsible for proton binding?
- How does proton binding (or releasing) trigger conformational change?
- How is substrate binding affected by proton binding?

The key question for KDELR has been what role do protons play in the processes from KDEL binding to COPI-vesicle mediated transport? This will be addressed in two chapters. The short hydrogen bond above the KDEL binding pocket is purposed to be the key factor in proton-coupled KDEL binding and the first chapter will address the following questions:.

- What is the nature of the short hydrogen bond and its relationship to the proton?
- How would this short hydrogen bond affect the binding of the KDEL?
- How would the short hydrogen bond and the KDEL binding lead to the conformational change?

Oligomerisation has been proposed to trigger the COPI-vesicle mediated transport. this leads to the following set of questions:

- How does KDELR oligomerise?
- Is the oligomerisation process conformation dependent?

- Is the oligomerisation process pH dependent?

2. Methods

2.1 Molecular Dynamics

2.1.1 Equations of Motion for Molecular Systems

The key question that we are trying to answer using molecular dynamics is how would a system of N particles move under the interacting potential V, which can be written down as the Lagrangian equation of motion.

$$\frac{d}{dt}\left(\partial L/\partial \dot{q}_k\right) - \left(\partial L/\partial q_k\right) = 0 \tag{2.1}$$

Where the potential of the system is $L = T - V$, where T and V are the kinetic and potential energy of the system respectively. q_k and $\dot{q}_k$ are the coordinate of the particle and the time derivative of the coordinates. For a molecular system, where particles are atoms, the evolution of the particle coordinates with respect to time can be written as:

$$m_i \ddot{r}_i = f_i \tag{2.2}$$

Where m_i is the mass of the atoms, r_i is the Cartesian coordinates and f_i is the force acting on the atom, which is related to the potential V as $f_i = -\bigtriangledown_{r_i} V$.

Since for many-body systems, an analytical solution to the Lagrangian equation of motion does not exist, finite-difference methods are usually used to numerically integrate the equations. A popular integration algorithm is call the “leapfrog” scheme and is written as:

$$v(t + \frac{1}{2}\delta t) = v(t - \frac{1}{2}\delta t) + \delta t a(t) \tag{2.3}$$

$$r(t + \delta t) = r(t) + \delta t\, v(t + \frac{1}{2}\delta t) \tag{2.4}$$

where velocity v and coordinate r are updated at the same time interval δt but velocity v is lagged half step compared with coordinate r.

2.1.2 Force Computation

To numerically solve the Lagrangian equation of motion, in addition to the mass and coordinates of the particle, the force on each particle is required as well. The force is computed as the derivative of the potential of the system, which is usually written as:

$$V = \sum_i v_i(r_i) + \sum_i \sum_{j>i} v_i(r_i, r_j) + \sum_i \sum_{j>i} \sum_{k>j} v_i(r_i, r_j, r_k) + \cdots \tag{2.5}$$

Where $\sum_i v_i(r_i)$ is the potential of particle i with respect to its coordinate. In molecular systems, this term describes the external potential such as the magnetic field.

$\sum_i \sum_{j>i} v_i(r_i, r_j)$ is the pairwise interaction and is the only interaction which is considered in normal computer simulations. The three-body interactions as well as the higher-order interactions are usually ignored to speed up the calculation. Since the pairwise

interactions depend on the separation between r_i and r_j, the interaction can also be written as $v_2(d_{i,j})$. The interactions between two atoms can be classified into three categories, namely electron hybridisation related interactions, charge interactions, and Van der Waals interactions.

Charge interactions are described by the Coulomb equation,

$$v^{qq}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r} \tag{2.6}$$

Where q_i and q_j is the charge of both particles and the ε_0 is the permittivity of free space.

For molecular systems, the charge of a molecule is distributed across all the atoms in the form of “partial charge”. Partial charges are derived from quantum chemistry calculations to mimic the electrostatic potential at the molecular surface.

Van der Waals interactions can be separated into two phases according to its nature. The repulsive force due to Pauli repulsion and attractive force due to dipole induced instantaneous dipole moment. The repulsive force decay exponentially with respect to the distance, while the decay of attractive force with respect to the distance can be fitted with a power of -6. Thus, for computational efficiency, the repulsive force is computed as the square of the attractive force. The interaction is termed Lennard-Jones potential and has the form:

$$v^{LJ}(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^{6}\right] \tag{2.7}$$

Where σ is the point that potential energy is zero and ε is the height of the energy minimum.

Chemical bonds due to electron hybridisation require explicit treatment of electrons. Thus, the bond, angle as well as dihedral potentials were introduced to mimic the local chemical environment. AMBER force field, for example, uses the following form to represent bonded interactions.

$$v^{Bonded}(r,\theta,\omega) = \sum_i \frac{k_i}{2}\left(r_i - r_{i,0}\right)^2 + \sum_i \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_i \frac{V_n}{2}\left(1 + \cos(n\omega - \gamma)\right)$$

(2.8)

Where both bond ($\sum_i \frac{k_i}{2}\left(r_i - r_{i,0}\right)^2$) and angle ($\sum_i \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2$) are modelled as

harmonic oscillator with a force constant of k_i centred around the equilibrium bond length $r_{i,0}$ and angle $\theta_{i,0}$. The dihedral potential is modelled as $\sum_i \frac{V_n}{2}\left(1 + \cos(n\omega - \gamma)\right)$,

where V_n is the strength of the potential, n is the multiplicity, γ is the phase factor, which is the dihedral angle of the minimum potential. Since the dihedral potential usually acts as an “error absorber” to allow molecular mechanic potentials to reflect quantum mechanic potentials, the dihedral potential can also be expressed as:

$$v(\omega) = \sum_{n=0}^{N} C_n \cos(\omega)^n \qquad (2.9)$$

to allow more flexibility and is named Ryckaert-Bellemans potential.

2.1.3 Time steps and Constrains

For numerically integrating the Lagrangian equation of motion, the step size is determined by the speed with which the forces change. In molecular systems, where all atoms are treated explicitly, the fastest motion is the bond vibration between a hydrogen atom and the associate non-hydrogen atom. Given that this motion has a vibration frequency of 10 fs, to achieve a stable integration, the time steps has to be one tenth of the vibration frequency and, thus, 1 fs time step is necessary to achieve stable integration.

However, in the fastest motion between hydrogen atom and the associate non-hydrogen atom, the average bond length between the hydrogen atom and the associate non-hydrogen atom is usually invariant with respect to the environment and can be considered as a constant. Thus, this vibration can be removed from the system by using constraints which constrain the bond length during the simulation. This allows a larger time step of 2 fs to be used. Since the vibration frequency depends on the mass of the atoms involved, one way of further increase the time step would be increase the mass of the corresponding atoms. Hydrogen mass repartition, which redistribute the mass from the heavy atom to the bounded hydrogens, allows a large time steps of 4 fs to be used.

2.1.4 Coarse-grained Model

Another way of increasing the time steps would be grouping particles to a larger particle. The Martini force field (178) allows the mapping of roughly four non-hydrogen atoms into a Martini bead. Thus, the vibration between bonded beads are slowed down significantly and a longer timestep of 20 fs can be used.

The Martini force field classifies different groups of atoms into a few types distinctive beads to reflect their chemical property. Four classes were defined, which are charged

(Q), polar (P), nonpolar (N), and apolar (C), where the Q and N class are further subdivided three types according to their hydrogen bonding capability. The P and C class are also further divided into five classes to reflect their hydrophobicity. The pairwise LJ interactions are individually tuned to fit the experimental interaction strength between different pair of beads. Since the charge of several atoms are concatenated into very small number of beads, the explicit screening effect provided by the partial charge on each atom no longer exists. Thus, all the coulombic interactions were computed with an implicit screening factor of $\varepsilon = 15$.

The bonded parameters in the Martini force field is determined in a fashion similar to the atomistic force field to reflect the local conformation. However, it has since been shown that the topology created based on residue type and secondary structure alone cannot enable a stable conformation (179). Thus, an empirical elastic network is usually applied to restraint the protein to its original conformation.

2.2 Long-range Forces

In contrast to the bond, angle, dihedral restraint where only a small number of pairwise interactions are considered, the LJ interactions and charge interactions should be computed between every pair of atoms and their infinite periodic neighbours at each step. Due to its fast decay of $1/r^6$, where r is the distance between two particles, the LJ interactions are negligible beyond 1 nm. Thus, a spherical cutoff of 1 nm is usually applied to only take into account of the LJ interactions when the separation is less than 1 nm. The charge interactions, on the other hand, has a decay rate of $1/r$. Thus, hundreds of nanometers are required for the interaction to become negligible.

2.2.1 Reaction Field Method

If one were to assume that the interactions beyond the 1 nm is of no importance and decided to use a simple cutoff, similar to the treatment of LJ interactions, it is quite often that the sum of the charge within the sphere is not neutral. For example, in ionic solutions, the presence of an ion phasing in and out of the boundary will create an artificial oscillation. To smooth the boundary effect while at the same time retain the computational efficiency of having short cutoff, reaction field is being purposed. In the reaction field method, the space beyond cutoff is represented as a polarisable medium, which has a predefined dielectric constant. The charge within the sphere creates a dipole which polarised the medium outside the sphere. The polarised medium in turn acting back on the particles within the sphere, mimicking the damping effect of the solution acting on the boundary. For example, when an ion enters the spherical cutoff, it initiates a strong interaction to the centre atom. The ion, due to its closeness to the boundary, creates a strong dipole which polarises the medium. The polarised medium, thus, initiates an interaction of similar scale but opposite sign to the centre atom. Thus, damping the effect of the ion.

2.2.2 Particle Mesh Ewald (PME)

Though the reaction field method has smoothened the boundary effect significantly, it is worth noting that for many systems, the medium outside the cutoff is not a homogenous polarisable medium. For example, the protein, due to its structural rigidity, won't polarise in response to the charge in the cutoff. Thus, the electrostatic contribution from these point charges beyond the cutoff can have significant impact on the behaviour of the atoms in the cutoff. The Particle Mesh Ewald (PME) method was introduced to account for all the partial charges across the infinite system, which means all of the atoms in the box as well as the periodic repeats.

Since the difficulty of the direct computation of the charge-charge interaction according to the coulombic rule lies within the slow decay with respect to the distance, by applying another Gaussian charge of the opposite sign at the same position, the sum of the electrostatic potential from point charge and the Gaussian charge of the opposite sign at the same position, decays much faster and is negligible at the distance of one nanometer. Thus, the total electrostatic potential can be written as:

$$V(r) = (V_{point}(r) + V_{Gauss}(r)) - V_{Gauss}(r) \tag{2.10}$$

Though $(V_{point}(r) + V_{Gauss}(r))$ needs to computed explicitly in real space, due to the fast decay with respect to the distance, a small spherical cutoff of 1 nm can be used. $V_{Gauss}(r)$ is computed in the reciprocal space and can be accelerated by gridding the space and put charge in pre-defined grids to allow fast Fast Fourier transform to be performed.

2.3 Isothermal-isobaric Ensemble

The original Lagrangian equation of motion assumes that the system is in the micro-canonical ensemble, where the energy of the system is fixed. However, the simulations are set up to reproduce physiological conditions, where the temperature as well as the pressure is kept constant. Thus, thermostats and barostats are used to allow the system to sample the isothermal-isobaric ensemble.

The instantaneous temperature (T) of a system with N particles and mass m and velocity v at time t can be expressed as:

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f} \tag{2.11}$$

Where $k_B N$ is the Boltzmann's constant and N_f is the degree of freedom of the system and is $N_f = 3N - 3$ of a system with fixed total momentum. Due to the large fluctuation in the instantaneous temperature, the temperature of the system is expressed as the time-averaged temperature according to the virial theorem. To adjust the system temperature to a pre-defined value, the simplest way is to scale the velocity of all the particles by a factor of $\sqrt{T_1/T_2}$, where T_1 is the target temperature and T_2 is the current temperature of the system. However, it has been shown that this crude treatment would drain the internal degrees of motion and create the flying ice-cube effect (180).

Thus, more complex treatments are usually required. Deterministic methods exists for both thermostats and barostats. Both of them function in similar manner, where the thermostat (e.g. Nose-hoover) couples the system to an external thermal bath as an extra degree of freedom and the barostat (e.g. Parrinello-Rahman) couples the box volume to an external "piston", also as an extra degree of freedom . For thermostats, the energy is allow to flow back and forth between the system and the reservoir. The temperature of the external heat bath is controlled with a thermal inertia so that it is fixed to the desired temperature. For barostats, the external "piston" compresses or expands the system according to the difference between the pressure of the system and the desired pressure. The fluctuation of the volume as a result of the movement of the "piston" is damped by the "piston mass".

In addition to the deterministic thermostats, stochastic thermostat such as Langevin thermostat can also be used to restrain the system to a pre-defined temperature.

Langevin thermostat modifies the kinetic energy of the system to mimic the coupling of the system to a thermal bath. The motion of the system is modified by the thermostat in two steps. A stochastic step where random forces are added to the system mimicking the random collision between the thermal bath and the system. This is followed by a deterministic step where frictional force proportional to the particle velocity are also added. The relationship between the these two steps and the desired temperature are linked via fluctuation–dissipation theorem.

2.4 Grand Canonical Ensemble and Monte Carlo Sampling

In the isothermal-isobaric ensemble, the system has fixed number of atoms, fixed temperature and fixed pressure. Though this ensemble reproduces the experimental condition, where the pressure and temperature is fixed, the content of interest may still exchange matter with the outside environment. Thus, the fixed number of atoms can be replaced with constant chemical potential which gives grand canonical ensemble, where the system can exchange matter with the external environment as long as the chemical potential stays constant.

Though grand canonical ensemble can faithfully reproduce the experimental environment, the move of adding and removing atoms cannot be sampled with the ordinary MD method. Thus, a Monte Carlo (MC) sampling is suggested to circumstances this problem. During MC sampling, the chemical potential of the system, chemical potential of the system plus a new molecule or the chemical potential of the system minus a molecule are computed. If the addition or the deletion of a molecule will lower the chemical potential, the move is accepted and the addition or deletion is performed. However, if the chemical

potential after addition or deletion is higher than the original system, metropolis criteria is adopted, where the move has a certain chance of being accepted.

2.5 Energy Minimisation and Equilibrium Detection

Before starting the MD runs, the system has to be energy minimised. When constructing the system, the system is in a high energy state, where there might be a pair of atoms that are very close to each other. The pauli repulsion between this pair of atoms will generate a strong force to push this pair of atoms apart. However, the repulsive force decays exponentially with respect to the distance between a pair of atoms. Thus, during leap-frog integration, where the coordinates are updated for a fixed time interval (e.g. 2 fs), the repulsion might overshoot and destabilise the system. One way of fixing this problem is to use steepest descent integrator, where change of the coordinate is capped to a certain distance (e.g. 0.01 nm).

After the energy minimisation, the system needs to be fully equilibrated before the production run, where it is assumed that the system is at equilibrium. The system is said to be equilibrated, when the potential, temperature and pressure of the system have converged to a fixed value. For membrane protein, the presence of the membrane adds an extra degree of complexity, where the area per lipid has to be converged as well, which could be examined by checking if the box area in the xy-plane has converged.

2.6 Ligand Binding Free Energy

2.6.1 The Definition of Binding Free Energy

One thing molecular dynamics could do is to compute the binding free energy of the ligand to the protein under a certain concentration. The binding event could be expressed

as the interconversion between Protein-Ligand (LP) complex and Ligand (L) Protein (P) in solution.

$$L + P \rightleftharpoons LP \tag{2.12}$$

Thus, the free energy of binding can be defined as:

$$\Delta G_b^\circ = -\, k_B T \, \ln K_b^\circ = -\, k_B T \, \ln \left(c^\circ \frac{[PL]}{[P]\,[L]} \right) \tag{2.13}$$

Where k_B is the boltzmann constant, T is the temperature, c° is the standard concentration and $[PL]/[P]/[L]$ are the concentration of protein-ligand complex, protein and ligand. However, since the time scale of simulating the binding or unbinding and the spacial requirement of simulating a box with an ordinary concentration is very large, direct computation is beyond the reach of normal simulations. To make this problem tractable, we write the free energy of binding in terms of the binding probability:

$$\Delta G_b^\circ = -\, k_B T \, \ln K_b^\circ = -\, k_B T \, \ln \frac{P_1}{P_0} = -\, k_B T \, \ln \frac{Q_1}{Q_0} \tag{2.14}$$

Where P_1 and P_0 is the probability of the bound and the unbound state, which is determined by their partition function Q_1 and Q_0, which, in the Isothermal-isobaric ensemble, is written as:

$$Q = \frac{1}{N!h^{3N}} \iiint e^{-\beta\left[H(x,p) + pV\right]} dV dx dp \tag{2.15}$$

Where N is the number of atoms, h is the Planck's constant, $\beta = 1/k_B T$, $H(x, p)$ is the Hamiltonian of the system based on atom coordinates x and momenta p, V is the volume of the system.

Since it is usually assumed that the temperature and the volume of the system remain unchanged after the binding, the binding free energy can be approximated as the potential difference between unbound and bound state.

$$\Delta G_b \cong \Delta A_b = -k_B T \ln \frac{\int e^{-\beta U_1(x))} dx}{\int e^{-\beta U_2(x))} dx} \tag{2.16}$$

One way of computing the free energy difference is via thermodynamic Integration, which is written as:

$$\Delta G_{0,1} = \int_{\lambda=0}^{\lambda=1} \frac{\partial U(\lambda, x)}{\partial \lambda}_{\lambda} \tag{2.17}$$

Where λ is a continuous variable between 0 and 1, which links the bound state $U(\lambda = 1, x)$ to the unbound state $U(\lambda = 0, x)$. Thus, by obtaining the the gradient at various lambda values, the free energy difference can be commuted by trapezoidal rule.

2.4.2 Thermodynamic cycle

However, such a λ, which directly connects the bound state to the unbound state, is hard to find. Thermodynamic cycles, where an indirect route composed of several stages, are more often constructed. The final free energy difference is the sum of all the free energy

differences from several stages. For the calculation of ligand binding to the protein, a thermodynamic cycle is shown is Figure 2.1.

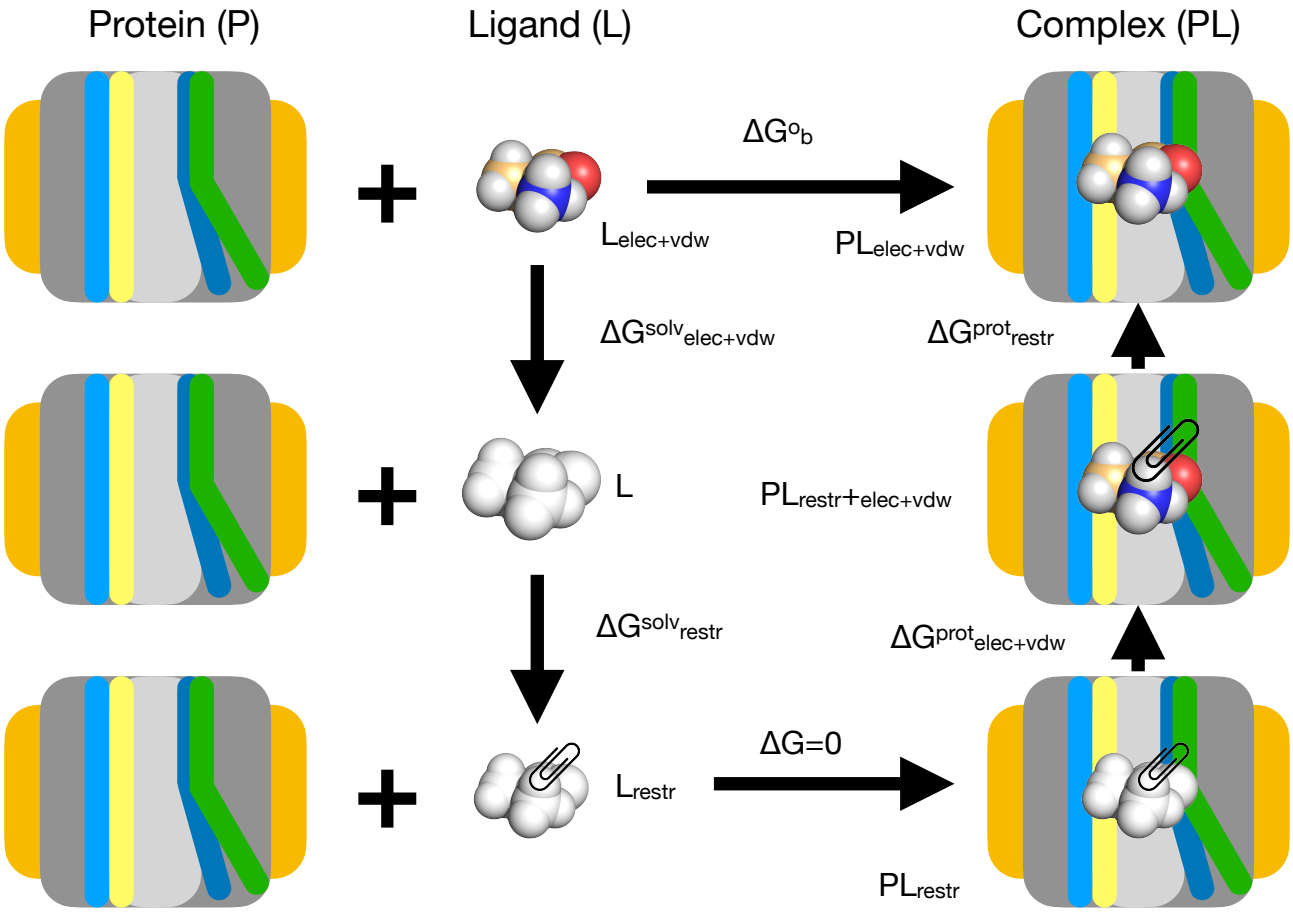


Figure 2.1. *Using thermodynamic cycle to compute the binding free energy. Since the direct computation of binding free energy* ΔG_b° *is not feasible, thermodynamic cycle is usually constructed to compute it indirectly. The ligand is first desolvated from the solution* $\Delta G_{elec+vdw}^{solv}$. *A analytical restraint is then applied to connect the non-interacting ligand and protein* ΔG_{restr}^{solv}. *The interactions between the ligand and protein (electrostatic and van der Waal interactions) were then switched on* $\Delta G_{elec+vdw}^{prot}$. *The restraint between the interacting protein and ligand is then removed* ΔG_{restr}^{prot} *to return to the physiological state of ligand in the protein binding pocket.*

The ligand binding free energy is usually expressed as the sum of decoupling the ligand from the protein and coupling the ligand to the bulk solution.

$$\Delta G_b^\circ = \Delta G_{elec+vdW}^{prot} - \Delta G_{elec+vdW}^{solv} \tag{2.18}$$

The process of decoupling is the removal of interactions between the ligand and the surrounding environment, namely the electrostatic and the Van der Waals interactions. $\Delta G_{elec+vdW}^{prot}$ is the free energy to decouple the electrostatic and Van der Waals interactions between the ligand and the protein, while $\Delta G_{elec+vdW}^{solv}$ is the free energy of decoupling the ligand from solution. The interactions within the ligand, unlike annihilation, are maintained. Thus, the decoupling process mimics the moving of ligand from the environment to the vacuum.

It is straightforward to compute the decoupling of the ligand from the solution. The decoupling of the ligand from the protein is, however, more problematic. When the interactions between the ligand and the environment is gradually decoupled, the ligand is no longer confined to the binding pocket and can freely explore the whole space. However, since the protein box is not a homogeneous medium like the ligand in solution, the time that is required for the ligand to fully explore the box would be too long to achieve convergence.

To facilitate the convergence, a restraint is imposed to restraint the ligand to the binding pocket. Thus, the process of of decoupling the ligand from the protein is separated into three steps (181).

$$\Delta G^{prot}_{elec+vdW} = \Delta G^{prot}_{rest} + \Delta G^{prot,rest}_{elec+vdW} - \Delta G^{solv}_{rest} \tag{2.19}$$

A restraint ΔG^{prot}_{rest} is added to restrain the ligand to the protein. The interactions between the ligand and the protein is then removed in the presence of the restraint $\Delta G^{prot,rest}_{elec+vdW}$. The restraint between the non-interacting protein and ligand are then removed ΔG^{solv}_{rest}.

Several criteria need to be satisfied when finding the correct restraint. The computation of the restraint free energy between both interacting ligand and protein and non-interacting ligand and protein shouldn't have a too high computation cost. Furthermore, when releasing the restraint between the non-interacting protein and ligand, the influence on the free energy from the volume which the ligand can sample should be be evaluated analytically. Thus, the set of restraints purposed by Boresch et al. (182) is used. This set of restraint is composed of one bond, two angle and three dihedrals as is shown in Figure 2.2. Though the free energy of applying the restraint to the interacting ligand and protein need to be calculated explicitly. The free energy of releasing the restraint between the non-interacting ligand and protein can be computed analytically with respect to the concentration and restraint strength:

$$\Delta G^{solv}_{rest} = k_B T \ln \left[\frac{8\pi^2 V^\circ}{r^2_{aA} \cdot \sin\theta_a \cdot \sin\theta_A} \frac{\left(k_{r_{aA}} k_{\theta_a} k_{\theta_A} k_{\phi_{ba}} k_{\phi_{aA}} k_{\phi_{AB}} \right)^{1/2}}{(2\pi k_B T)^2} \right] + k_B T \ln \frac{V}{V^\circ}$$

(2.20)

Where V° is the standard volume of 1666 Å^3, which is the inverse of Avogadro constant times one litre. r_{aA} and $k_{r_{aA}}$ is the equilibrium length and the force constant of the bond restraint, θ_a, θ_A and k_{θ_a}, k_{θ_A} are the equilibrium angles and the force constants of the two angle restraints. $k_{\phi_{ba}}$, $k_{\phi_{aA}}$ and $k_{\phi_{AB}}$ are the three force constants of the dihedral restraints.

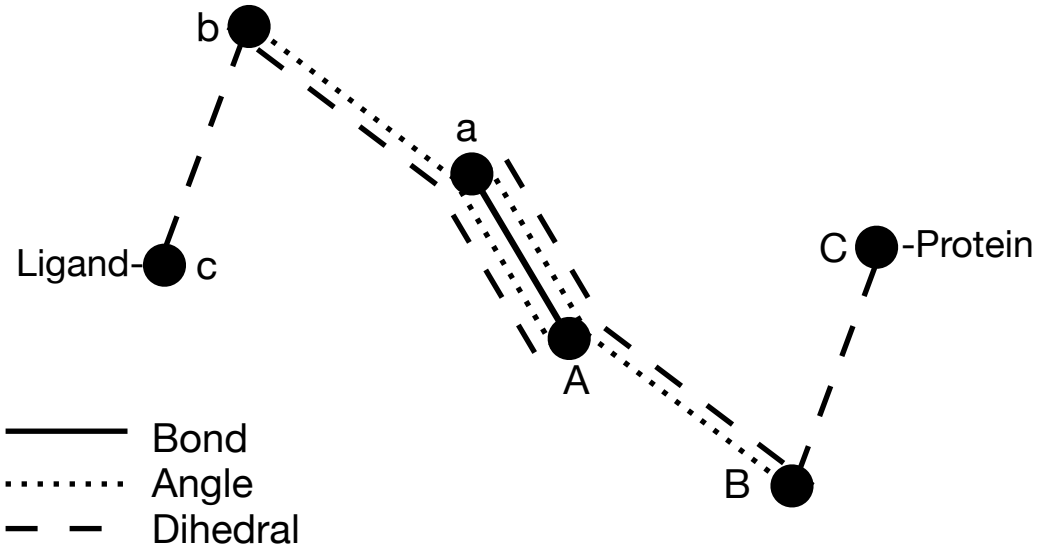


***Figure 2.2**. Using restraint to restraint the relative orientation of ligand to protein. A set of restraints were chosen based on three atoms in the ligand (a,b,c) and another three atoms in the protein (A,B,C). The restraint set composes of a distance restraint (solid line), two angle restraints (dotted line) and three dihedral restraints (dash line).*

2.7 Charge Correction

When computing the ligand binding free energy, the electrostatic interaction between the ligand and the protein is very important. Thus, the electrostatic interactions within the system is usually evaluated by PME to account for the long range interactions. However, in PME, it is assumed that the sum of the charge in the simulation box is neutral. When decoupling a charged ligand from the simulation box, the sum of the charge within the box without the ligand will be non-zero. To allow the PME to work, a neutralising plasma is distributed equally within the box to keep the box neutral. However, it has been shown that this treatment will have size dependent effect on the system (183-186). To account for this problem, Rocklin *et al.* (187) has purposed a correction that changes the ligand binding free energy in periodic systems into the ligand binding free energy in an infinite non-periodic box. The correction is expressed as:

$$\Delta G_{MD,NBC} = \Delta G_{MD,PBC}(L) + \Delta\Delta G_{ANA}(L) + \Delta\Delta G_{DSC}(L) \tag{2.21}$$

Where $\Delta G_{MD,NBC}$ is the corrected binding free energy under no periodical boundary. $\Delta G_{MD,PBC}(L)$ is the binding free energy for ligand L under periodic boundary conditions. $\Delta\Delta G_{ANA}(L)$ is the analytic correction scheme. $\Delta\Delta G_{DSC}(L)$ is a correction for discrete solvent effects.

The discrete solvent correction accounts for the effect of the neutralising plasma on the quadrupole-moment of water. For N_S number of solvents, it is evaluated as:

$$\Delta\Delta G_{DSC}(N_S, L) = -\frac{\gamma_S Q_L}{6\varepsilon_0}\frac{N_S}{L^3} \tag{2.22}$$

γ_S is the quadrupole-moment trace of the solvent model relative to its single van der Waals interaction site. For TIP3P, it is 0.00764 e nm². Q_L is the charge of the ligand and ε_0 is the permittivity of vacuum. L is the length of the box and L^3 is the volume of the box.

The analytic correction scheme $\Delta\Delta G_{ANA}(L)$ consists of four components as is shown in equation.

$$\Delta\Delta G_{ANA}(L) = \Delta\Delta G_{NET}(L) + \Delta\Delta G_{USV}(L) + \Delta\Delta G_{RIP}(L) + \Delta\Delta G_{EMP}(L) \quad (2.23)$$

$\Delta\Delta G_{NET}(L)$ is the correction for periodicity-induced net-charge interactions. $\Delta\Delta G_{USV}(L)$ is the correction for periodicity-induced undersolvation. $\Delta\Delta G_{RIP}(L)$ is the correction for residual integrated potential effects and $\Delta\Delta G_{EMP}(L)$ is the additional empirical correction term.

$\Delta\Delta G_{NET}(L)$ accounts for the superfluous interaction between the charged ligand with its periodic neighbour in vacuum and is written as:

$$\Delta\Delta G_{NET}(L) = -\frac{\xi_{LS}}{8\pi\varepsilon_0}\left[\left(Q_P + Q_L\right)^2 - {Q_P}^2\right]\frac{1}{L} \quad (2.24)$$

ξ_{LS} is the cubic lattice-sum (Wigner) integration constant = –2.837297 and Q_P is the charge of the rest of the system.

$\Delta\Delta G_{USV}(L)$ accounts for the disruption of the solvation shell around the charged ligand by the edge of the simulation box and is, thus, partially cancelled by $\Delta\Delta G_{NET}(L)$:

$$\Delta\Delta G_{NET}(L) + \Delta\Delta G_{USV}(L) = \Delta\Delta G_{NET}(L) \times \varepsilon_S{}^{-1} \tag{2.25}$$

ε_S is the static relative dielectric permittivity of the solvent medium, for TIP3P water is 97.

The $\Delta\Delta G_{RIP}(L)$ accounts for the different between the $\Delta\Delta G_{USV}(L)$, which assumes that the ligand is a point charge in the centre of the box surrounded by a homogeneous medium and the real simulation box.

$$\Delta\Delta G_{RIP}(L) = \left[\left(I_P + I_L\right)\left(Q_P + Q_L\right) - I_P Q_P\right]\frac{1}{L^3} \tag{2.26}$$

I_P and I_L are the residual integrated potentials of protein and ligand. I_L is calculated from poisson–boltzmann equations in a a heterogeneous dielectric environment where interior permittivity set to $\varepsilon_I = 1$, exterior permittivity set to $\varepsilon_S = 97$.
Since the correction for a single ion in pure solvent can be computed analytically, an additional term $\Delta\Delta G_{EMP}(L)$ is derived to make this correction equal to the analytical answer.

$$\Delta\Delta G_{EMP}(L) = -\frac{1}{8\pi\varepsilon_0}\frac{16\pi^2}{45}\left(1 - \frac{1}{\varepsilon_S}\right) \times \left[\left(Q_P + Q_L\right)^2 - {Q_P}^2\right]\frac{{R_L}^5}{L^6} \tag{2.27}$$

R_L is the effective solvation radius for the ligand, which is computed as:

$$R_L = \left\{\left[\frac{1}{8\pi\varepsilon_0}\frac{4\pi}{3}\left(1 - \frac{1}{\varepsilon_S}\right)Q_L\right]^{-1} I_{L,SLV}\right\}^{1/2} \tag{2.28}$$

$I_{L,SLV}$ represents the solvation contribution to the residual integrated potential I_L of the ligand.

$$I_{L,SLV} = I_L - I_{L_\mathrm{hom}} \tag{2.29}$$

I_{L_hom} is calculated in a homogenous dielectric environment where both exterior and interior permittivity are the same as vacuum.

2.8 Attach, Pull, Release (APR) Scheme

In the previous section, it has been shown that in order to obtain the binding free energy, a route from the bound state, where the ligand is interacting with the protein to the unbound state where the ligand is not interacting with the protein, needs to be defined. Thus, one route would be gradually scaling down the strength of interaction via an unphysical route. The free energy difference between the ligand bound state and the unbound state can also be defined as the work which is required to pull the ligand out of the binding pocket. In this case, the interactions between the ligand and the protein are scaled down via a physical route, which is via the increase in the distance between ligand and the protein. Thus, in this case, λ is defined as the distance between the ligand and the protein. The principle of thermodynamic integration can, thus, be defined in a similar fashion.

$$W = \int \frac{\partial U(\lambda, x)}{\partial \lambda} \tag{2.30}$$

Where W is the total amount of work and $\frac{\partial U(\lambda, x)}{\partial \lambda}$ is the average force between the ligand and the protein when they are separated by a distance of λ. One example of such method is the attach, pull, release (APR) scheme (188). The simplest scheme would be gradually increasing the distance between the binding pocket and the ligand. The average force at each distance could then be integrated to obtain the binding free energy. However, as the two bodies are moving further away from each other, the volume that the system is able to sample increases at the rate of O(n^3), as is shown in Figure 2.3. Thus, to tackle this problem, we applied a restraint so that the volume which the ligand is able to sample, is invariant to the distance between the ligand and the protein.

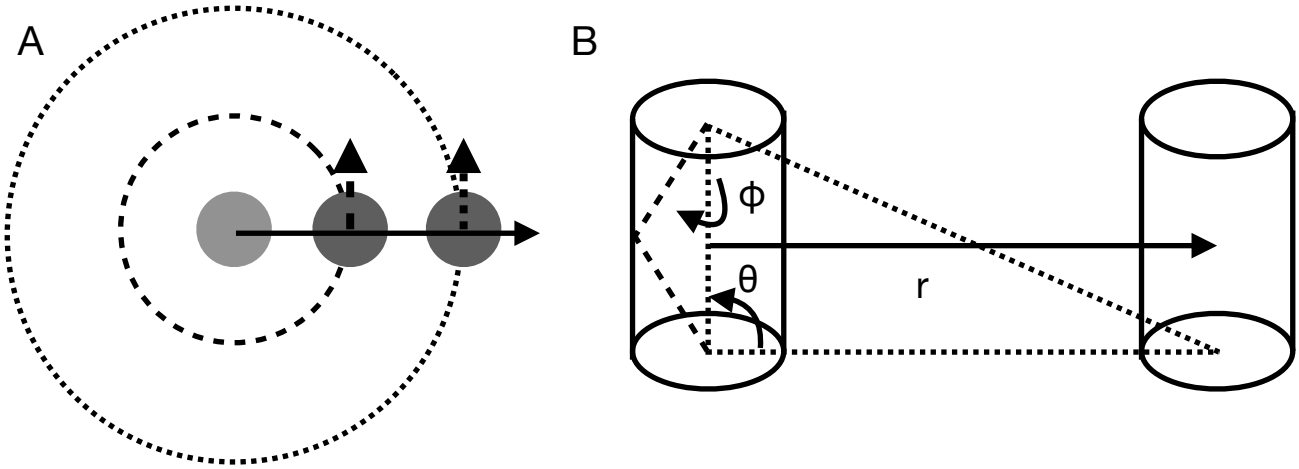


***Figure 2.3**: Restraints for the Attach, Pull, Release scheme. A. The configuration space increase at a rate of O(n^3) as the two bodies are further apart. When the dark grey boy is pulled away from the light grey body (solid arrow), the configuration space orthogonal to the pulling direction (dashed arrow and dotted arrow) increase rapidly as the distance between the two bodies increases from dashed circle to the dotted circle. B. A distance r is defined between the two bodies. a torsion is defined as the angle between the plane of one body (dashed triangle) and another plane between two bodies (dotted triangle). The angle is defined in the three atoms which define the plane of the two bodies.*

Various versions of APR scheme exists and the restraints are used to confine the degree of freedom in three directions. A distance (r) was applied to maintain the separation between the two bodies and a torsion (φ) that limits the orientation of the ligand to the protein in the x-y plane and finally an angle restraint (θ) that restrains the degree of freedom in the z axis.

To compute the absolute binding free energy, a similar thermodynamic cycle is required and the whole process will be separated into three phases. The first phase is the attach phase, where the restraint is attached between the two interacting bodies. The second phase is the pulling phase, where the two bodies are pulled apart while maintaining the restraint between the two bodies. The final step is the release step, where the restraint is released when the two bodies are distant enough to be regarded as not interacting.

$$\Delta G^{\circ}_{bind} = -\left(W_{attach} + W_{pull} + W_{release}\right) \tag{2.31}$$

Since the system now includes non-physical restraints, the chemical potential of the system is redefined as:

$$\mu^{\circ} = -RT\ln\left(\frac{8\pi^2}{C^{\circ}}\right) - RT\ln\left(\int e^{-\beta\left[U(x) + \sum_{n=1}^{N}\lambda k_n (r_n - r_{0,n})^2\right]} dx\right) \tag{2.32}$$

Where C° is the standard concentration and $\beta = 1/RT$, $U(x)$ is the potential energy of the original system with respect to the atomic coordinates. N non-physical harmonic restraints (distance, angle and dihedral) were added where k is the restraint strength and

x_0 is the centre of the restraint. λ is the scaling factor which allows the restraint to be turned on and off.

During the first phase, the system transforms from the vanilla state, where no non-physical restraint is applied ($\lambda = 0$), to the state where all the restraints are on ($\lambda = 1$). The force during this process is the the partial derivative of chemical potential μ° with respective to λ.

$$\langle F \rangle_\lambda = \frac{\partial \mu^\circ}{\partial \lambda} = \sum_{n=1}^{N} k_n \left\langle \left(x_n - x_{0,n} \right)^2 \right\rangle_\lambda \tag{2.33}$$

The angle bracket indicates the average force at scaling factor lambda. Thus, the total work can be computed by integral over the force curve:

$$W_{attach} = \int_0^1 \sum_{n=1}^{N} k_n \left\langle \left(x_n - x_{0,n} \right)^2 \right\rangle_\lambda d\lambda \tag{2.34}$$

The second phase is where all other restraints were kept the same and only the distance restraint ($x_{0,1}$) is altered. To compute the average force, the chemical potential μ° is we partial differentiated with respective to ($x_{0,1}$).

$$\langle F \rangle_{x_{0,1}} = \frac{\partial \mu^\circ}{\partial x_{0,1}} = -2k_1 \left\langle \left(x_1 - x_{0,1} \right) \right\rangle \tag{2.35}$$

To obtain the overall work of pulling the system from complex state to the non-interacting state, a similar integration of the force curve is required as well. For a system which prefers to stay in complex state, the overall work of pulling is positive

$$W_{pull} = \int_{attached}^{pulled} -2k_1 \left\langle \left(x_1 - x_{0,1}\right) \right\rangle dx_{0,1} \tag{2.36}$$

The last phase is the removal of the restraint, which is the exact opposite of the first phase, where the restraint is released and system returns to the standard state of two non-interacting bodies.

$$W_{release} = \mu_A^\circ + \mu_B^\circ - \mu_{AB}^\circ$$

$$= -RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta E(X_A)} dX_A\right) - RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta E(X_B)} dX_B\right)$$

$$+RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta\left[E(X) + \sum_{n=1}^{N} \lambda k_n (x_n - x_{0,n})^2\right]} dX\right) \tag{2.37}$$

Since it is assume that at the end of the pulling phase, the two bodies are not interacting anymore.

$$RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta E(X_A)} dX_A\right) + RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta E(X_B)} dX_B\right) = RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta E(X)} dX\right)$$

(2.38)

The release free energy could be semi-analytically computed, where r is the distance between the two bodies, θ is the angle and φ is the dihedral.

$$W_{release} = RT \ln\left(\frac{8\pi^2}{C^\circ}\int e^{-\beta\sum_{n=1}^{N}\lambda k_n(x_n-x_{0,n})^2}dX\right)$$

$$= RT \ln\left(\frac{8\pi^2}{C^\circ}\right) + RT \ln\left(\int_0^\infty r^2 e^{-\beta k_r(r-r_0)^2}dr\right) + RT \ln\left(\int_0^{2\pi} e^{-\beta k_\varphi(\varphi-\varphi_0)^2}d\varphi\right)$$

$$+RT \ln\left(\int_0^\pi \sin(\theta)e^{-\beta k_\theta(\theta-\theta_0)^2}d\theta\right) \quad (2.39)$$

2.9 Well-tempered Metadynamics

Though the free energy separating two bodies could be computed by the attach-pull-release scheme, the free energy of pulling the ligand out of the protein can be sampled more efficiently via enhanced sampling methods. The reason for using an enhanced sampling method for computing the free energy of binding is due to the long residence times when the system gets stuck in an energy minimum. Thus, if the energy minimum can be “filled up” to flatten the energy landscape, the system can move rapidly between the bound and unbound state. The energy which is used to flatten the energy landscape is the inverse of the original energy landscape. Well-tempered Metadynamics is a method aiming to obtain this energy.

The time that the system stays in one state usually correlates well with the depth of the free energy well. The longer the system stays in one state, the more negative this free

energy well is. Thus, one way of estimating the amount of bias that is required to fill up the energy well is via the computation of the residue time within this well. Metadynamics adds a fixed sized Gaussian bias to the current state every fixed time interval. Thus, due to the longer residue time in an energy well, more Gaussians will be deposited. The addition of the Gaussians only stops when the amount of Guassians deposited at the current point is the same as the depth of the energy well. Thus, since the energy well no longer exists, the system leaves the current state to explore other states.

However, after all the original energy wells are flattened, Metadynamics continues to add Gaussians. Thus, the total bias always oscillates around the inverse of the free energy landscape. One way of solving this problem is via Well-tempered Metadynamics, where the newly added Gaussians were scaled down gradually. Thus, though the sum of Gaussians is still oscillating around the inverse of the free energy landscape, the oscillation amplitude is gradually damped due to the diminishing height of the newly added Gaussians. However, it is worth noting that since the amplitude of the newly added Gaussians is always diminishing, the system might not be able to escape the energy well as efficiently as regular Metadynamics.

2.10 Quantum Mechanics

2.10.1 Solving Schrodinger Equation

Though Molecular Dynamics has been proven to be an excellent tool at investigating the motion of proteins, the bonded parameters, partial charge and LJ parameters of a residue are invariant to its surroundings. Thus, in the cases where these factors are changing, molecular mechanics might be inappropriate and one might need quantum mechanics methods to accurately describe the property of the electrons. The key to quantum

chemistry calculations has been trying to find the correct wave function for describing the property of the electrons. The wave function of the system obeys the Schrödinger equation, which for a single particle where the external potential is invariant to time, can be written as:

$$E\Psi = -\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + V\Psi \tag{2.39}$$

Where E is the energy, Ψ is the wave function, $\hbar$ is the reduced Planck constant, m is the particle mass, x is the particle coordinate and V is the external potential. To solve the Schrödinger equation for a molecular system, we need to define the theoretical method and the basis set, where theoretical methods describe how the interactions between the electrons are computed and a basis set describes the level of detail required to describe the electrons. The combination of theoretical methods and basis sets is called the level of theory and will determine the accuracy of the results.

An example of the theoretical method is Hartree-Fock. The Hartree-Fork method defines the Fock operator, which transforms the Schrödinger equation into the Hartree-Fork equation.

$$\hat{F}\varphi\left(r_i\right) = \varepsilon\varphi\left(r_i\right) \tag{2.40}$$

Where $\varphi\left(r_i\right)$ is the Hartree-Fork orbital and ε is the Hartree-Fork energy and $\hat{F}$ is the Fock operator:

$$\hat{F}_i = -\frac{1}{2}\bigtriangledown_i^2 - \sum_A \frac{Z_A}{r_{iA}} + V(r_i) \tag{2.41}$$

Where $V(r_i)$ is the mean field potential from other electrons and is written as

$$V(r_i) = \int \frac{\rho(r')}{\left|r_i - r'\right|} dr' + \xi(r_i) \tag{2.42}$$

Where $\int \frac{\rho(r')}{\left|r_i - r'\right|} dr'$ is the coulombic potential from other electrons and $\xi(r_i)$ is the

exchange potential.

When solving the Hartree-Fork equation, the molecular orbital $\varphi\left(r_i\right)$ is expressed as a

linear combination of the basis functions.

$$\varphi\left(r_i\right) = C_{1,i}\chi_1(r) + C_{2,i}\chi_2(r) + \cdots + C_{n,i}\chi_n(r) \tag{2.42}$$

The basis function is expressed as Gaussian type functions (GTF), which can be written as:

$$\phi(r) = Nx^l y^m z^n e^{-\alpha(r-R_A)^2} \tag{2.43}$$

where N is the normalisation factor, α is the exponential of the Gaussian function, R_A is the coordinate of the core, x,y,z is the projection of the $(r - R_A)$ on the Cartesian coordinate system, i,m,n determines the type of the GTO and the sum of i,m,u is the angular moment of the orbital.

2.10.2 Wave Function Analysis

To have a detailed understanding of the properties of chemical bonds, the wave function as well as the electron density obtained from QM calculations can be used. One key question that is interesting to solve is estimating the type and strength of the interactions between two atoms or between two fragments.

2.10.2.1 Bond Order Analysis

When investigating the existence and strength of the covalent interactions between two atoms, the bond order can be a useful descriptor. The classical description of the bond order is the Lewis description, which classifies the bond as single, double or triple bond. However, many chemical systems have conjugated bonds, which cannot be classified into this integer bond system. Furthermore, single bonds in similar systems might have different bond strength, which is also not captured by the Lewis description.

The Mayer bond order (189) is one solution to these problems. Mayer bond order computes the number of electron pairs shared between two atoms. Similar to the Lewis model, the carbon carbon bond in ethane, ethene and acetylene has a Mayer bond order of 1.0, 2.0 and 3.0. For highly conjugated system such as benzene, the carbon-carbon bond order is 1.5. Within the same type of bonds, the Mayer bond order correlates well with the bond strength. Thus, the Mayer bond order can also be used to compare the strength of the interaction between two atoms.

2.10.2.2 Atom-in-molecule (AIM) Analysis

Bader purposed the quantum theory of Atom-in-molecule (AIM) topology analysis based on the electron density, which defines the bond critical points (BCP), that can be used to infer the strength and type of the interactions. The AIM topology analysis computes the hessian matrix of the electron density and attempts to find the critical points, which is defined as the point where the gradient norm is zero. The critical points can then classified based on the number of eigenvalues that are negative (Figure 2.4 A).

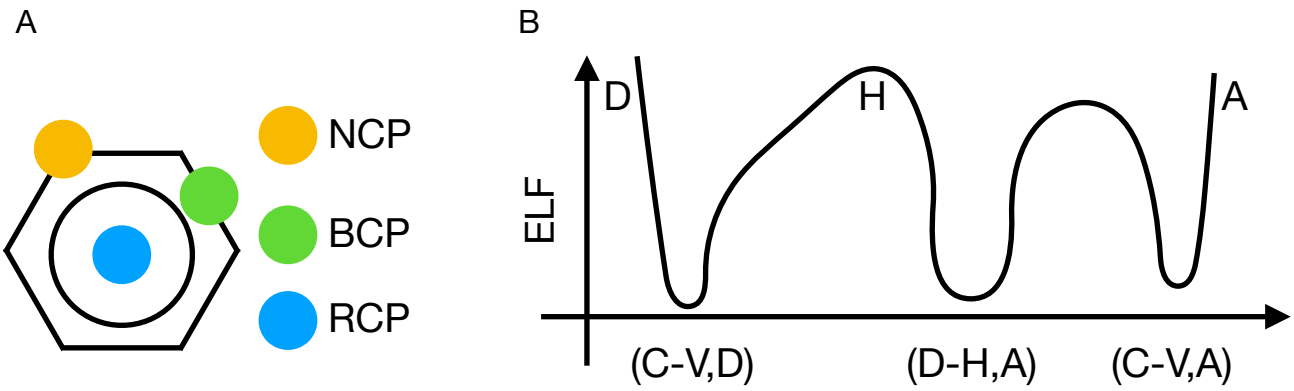


Figure 2.4. *Critical points in AIM analysis and ELF analysis.*
A. Three types of critical points could be observed in benzene during AIM analysis. The nuclear critical point (NCP) coincide with the position of the nucleus. The bond critical point (BCP) sits in the middle of chemical bond, while the ring critical point (RCP) sits in the middle of a ring structure. B. Three local minimum points could be identified along the path from hydrogen bond donor heavy atom (D) to hydrogen bond acceptor (A). The point between the core and valance of the hydrogen bond donor heavy atom (C-V, D). The point between the core of the hydrogen (H) and the valance of the acceptor (D-H, A) and the point between the core and valance of the acceptor (C-V, A). CVBi is defined as the difference between the ELF at (C-V,D) and at (DH-A).

When all three of the eigenvalues are negative, the critical point is the local maximum and is named as type (3, -3). Type (-3,3) critical point occurs only in the nuclear and is, thus, called the nuclear critical point (NCP). The number of NCPs are the same as the number of atoms in the system.

The Bond critical point (BCP) is the point where two out of the three eigenvalues are negative (3, -1) and is the point of interest for weak interactions. A (3, -1) point is the second-order saddle point and is found between a pair of atoms attractive to each other. At BCP, the value of electron density, the laplacian of the electron density, as well as the potential energy density were found to be correlated to the bonding strength (190), or the strength of hydrogen bond (191). The maximal gradient path joining the BCP with NCPs are called the bond path and indicates how two atoms are joined by a bond.

When only one of the eigenvalues is negative, the critical point is termed as a ring critical point (RCP), as it is usually found in the middle of a ring suggesting steric repulsion. The RCP is the first-order saddle point and is termed as (3, +1) in topology. When none of the eigenvalues are negative (3, +3), the point is the local minimum. Since point (3, +3) usually appear in the middle of the cage, it is also termed the cage critical point (CCP).

2.8.2.3 Electron Localisation Function (ELF) Analysis

Besides the AIM analysis, the electron localisation function (ELF) Analysis is also commonly used when investigating weak interactions that are electrostatic in nature such as the hydrogen bond. ELF is a variable ranging from 0 to 1. A value close to 1 indicates that there is a very high likelihood of finding electrons of the same spin while a value close to zero signals that an electron of the opposite spin is more likely to be found. ELF can be

analysed using topology methods similar to the AIM analysis. One method of evaluating the strength of hydrogen bond is core-valence bifurcation index (CVBi) (192).

Hydrogen bond systems can written in the form of $D - H \cdots A$, where D is the hydrogen bond donor, H is the hydrogen bonding to the hydrogen bond donor and A is the hydrogen bond acceptor. Under topology analysis, the plane of ELF ($D - H \cdots A$) can be divided into several types of basins (Figure 2.4 B). The nuclear of the hydrogen bond donor heavy atom (D) and the hydrogen bond acceptor (A) shows up as the core of the basin ($C(D)$ and $C(A)$). The valance electrons of the hydrogen bond acceptor form another type of basin $V(A)$. The last type of basin $V(D, H)$ composes of the valance electrons of the hydrogen bond donor and the hydrogen bonding to the hydrogen bond donor. A bifurcation point, which corresponds to the (3,-1) critical point, can be found between two basins. The CVB index can be expressed as:

$$CVB_i = ELF(C - V) - ELF(DH - A) \tag{2.44}$$

Where $ELF(C - V)$ is the ELF value of the bifurcation point between the core $C(D)$ and valance $V(D, H)$ basins and $ELF(DH - A)$ is the bifurcation point between the hydrogen bond donor $V(D, H)$ and hydrogen bond acceptor$V(A)$. The value of CVBi correlates well with the hydrogen bond strength (192), where the higher the CVBi value, the stronger the hydrogen bond.

2.8.2.4 Electrostatic Potential Analysis

Aside from using ELF function to investigate the electrostatic interactions, the electrostatic potential could be analysed directly to investigate the electrostatic

interactions. The electrostatic potential is the amount of work that is required to move a unit of positive electric charge from the reference point to a point in the space within the electric field. For a point charge Q, the electrostatic potential V_E at distance r is defined as:

$$V_E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} \tag{2.45}$$

In quantum chemistry calculation, where the electron is represented as electron density, the electrostatic potential with respect to the electron density could be computed as:

$$V_E = \frac{1}{4\pi\epsilon_0}\int \frac{\rho(r_0)}{r} dr_0 \tag{2.45}$$

Where $\rho(r_0)$ is the electron density at point r_0, which has a distance of r from the reference point.

The sign as well as the value of the electrostatic potential on the surface of molecule, which is usually defined as the contour at a certain electron density, can be used to deduce the type and the magnitude of the electrostatic interaction that the molecule is likely to perform.

2.11 QM/MM calculations

Though QM gives good description of the behaviour of the electrons, it is too costly for a system beyond a hundred atoms. The MM description, on the other hand, can be used to describe a system with a hundred thousand atoms, but cannot give correct description for the behaviour that is not parametrised. One way of describing a large system where the behaviour is not parametrised by the MM force field is though the QM/MM technique,

where the system of interest and is not well described by the MM force field is described at the QM level.

In QM/MM calculations, at every time step, the forces are computed for both QM and MM atoms and the coordinates are updated accordingly. The key to the QM/MM calculations is how the MM region is affected by the QM region and how would the QM region affects the MM region. These effects are mainly manifested as non-bounded interactions, which are van der waals interactions and electrostatic interactions. The van der waals interactions within the QM region is represented in a native QM fashion through exchange and correlation functionals, while the interactions between the MM region and the QM region and within the MM region are all calculated using the LJ potential according to the force-field definition. The electrostatic interactions are more complicated and, in this project, the electrostatic embedding scheme is used. In this scheme, the partial charge of the system is represented as point charge in the QM calculations, during which the effect of the QM system on the point charge are computed. The electrostatic interactions within the MM region is computed using the simple coulomb's law.

Since the QM calculation is repeated million times during the QM/MM calculation, a theoretical level that is fast to compute is required. In this project, the composite method B97-3c (193) is used. B97-3c is based on the pure functional B97 (194), such that the computation load is small. DFT-D3(BJ) (195) is used to empirically correct for the long-range dispersion that is not present in the functional. Pure functional are known to overestimate the bond length and thus, short-ranged basis set errors is corrected (196). The Ahlrichs triple-zeta basis set is modified so that polarisation on hydrogen atoms is reduced and additional polarisation functions are added to oxygen atoms, such that the description of the strong hydrogen bonds are improved.

3. Proton Control of Transitions in an Amino Acid Transporter

3.1 Introduction

Amino acids are important players in regulating the internal environment of the cell. The intracellular amino acid level is regulated by various transporters that utilise different mechanisms. To move a substrate against its concentration gradient, transport is sometimes coupled to another electrochemical gradient such as the proton electrochemical gradient. One example would be the solute carrier (SLC) 36 family of proton-coupled amino acid transporters, which are involved in regulating cellular metabolism through interacting with the mechanistic targets of rapamycin (mTORC) kinase signaling complex (197). In eukaryotic cells, the SLC36 transporters are also involved in lysosome-mediated amino acid export (198). In other cases, the transport can be coupled to the sodium concentration gradient (199). Some transporters can also act as a facilitated diffuser, which transports the substrate down its concentration gradient (200), or exchanges the substrate for its metabolised product. This exchange activity can be regulated by pH, where the transporter is only active in acidic pH (201). However, the detailed mechanisms of these transporters and how transport activity is regulated by protons, sodium or pH remain elusive.

Many amino acid transporters share a similar overall structure, characterised by the leucine transporter (LeuT). The structural fold of LeuT is comprised of an inverted repeat of a five helices bundle and two extra helices in the C terminal (chapter 1.4.4). It is hypothesised that transporters sharing this structural fold share similar transport mechanisms (Figure 3.1A) (202). The binding of the substrate to the outward-open state

leads to the closure of the extracellular gate giving rise to the outward-occluded state. After that, the transporter enters a state where the extracellular gate is firmly closed and the intracellular gate is about to open. This state is termed the inward-occluded state and is followed by the opening of the intracellular gate and the release of the substrate, where the transporter is in the inward-open state and the intracellular gate is fully open.

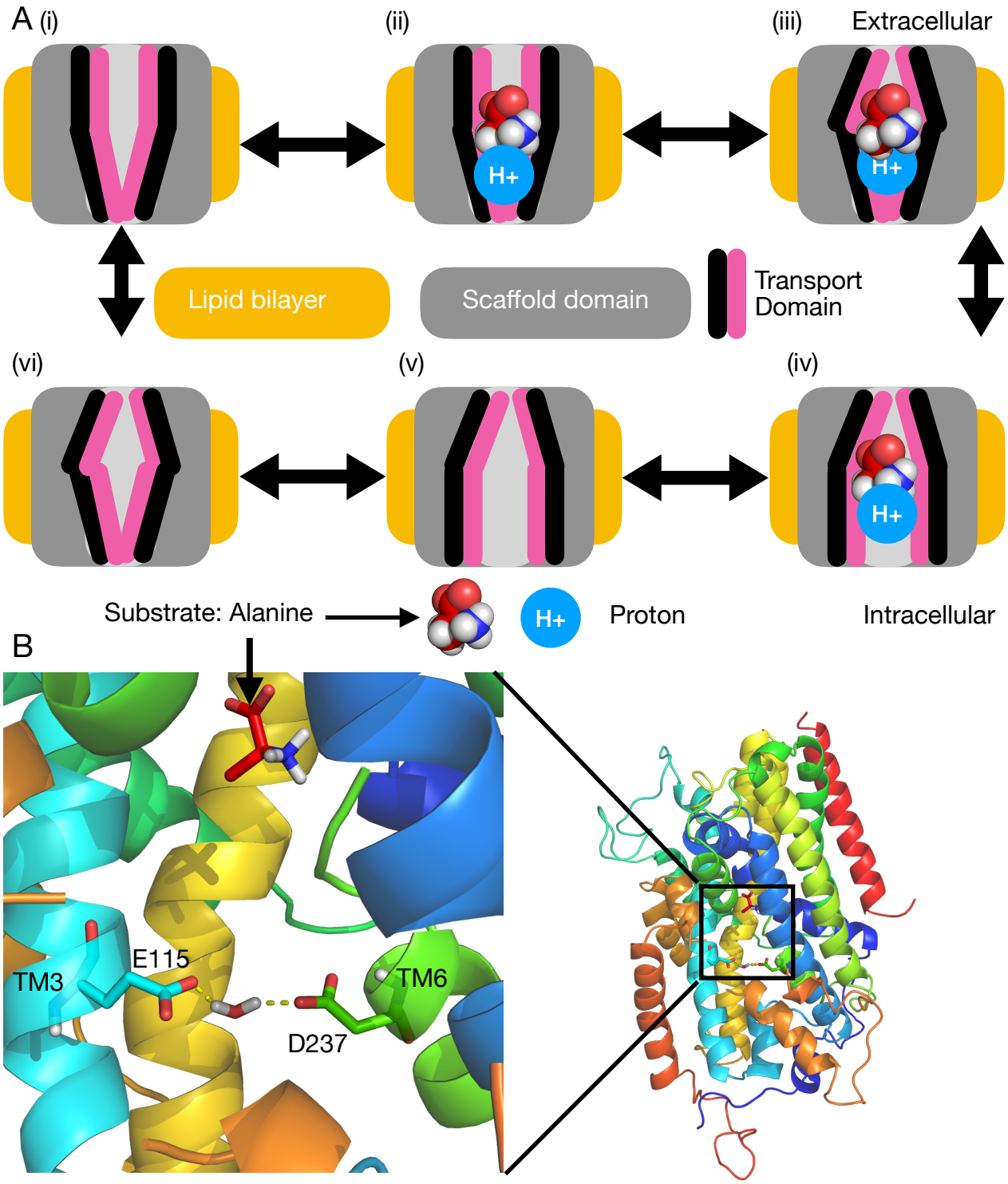


Figure 3.1. Structure and mechanism of GkApcT (4). (A) In the outward-open state (i), the substrate and proton enters the binding site via the extracellular gate (ii). The binding of substrate triggers the closing of the extracellular gate (iii) and the opening of the intracellular gate (iv), which leads to the release of the proton and the substrate (v). Transporter goes back to outward-open state through an occluded state (vi) (B) GkApcT in the inward-occluded state with alanine (red sticks) in the binding pocket. The zoomed out view of the binding pocket (right black square) is shown on the left. The binding pocket is gated by E115 from TM3 (aqua stick) and D237 from TM6 (green sticks), which are joined by a water molecule.

A proton-coupled amino acid transporter with the LeuT fold from *Geobacillus kaustophilus* (GkApcT) was recently solved in an inward-occluded state (4). GkApcT is related to SLC7 family and distantly related to the SLC36 family. Thus, this structure is an interesting model for elucidating the mechanistic detail of the proton-coupled transport mechanism. The crystal structure suggested that a glutamate residue gating the intracellular side of the binding pocket (TM3: E115) might be the key to the opening of the transporter and the ejection of the substrate. The intracellular gate under the binding pocket is formed by this glutamate residue (E115) as well as an aspartate residue (TM6: D237) (Figure 3.1B). Our hypothesis is that in the inward-occluded conformation, the intracellular gate is stabilised by protonated E115 (supported by the calculated pKa value of 8.22) interacting with D237 via a linking water. Once the E115 is deprotonated, the negatively charged glutamate residue repels the negatively charged aspartate residue, which opens the intracellular gate allowing the transporter to enter the inward-open state and release the substrate. To investigate the plausibility of this hypothesis, we have used computational methods to examine the role of protons, specifically the protonation of E115 and its influence on the free energy landscape of the inward-occluded to inward-open transition.

3.2 Material and Methods

3.2.1 System Preparation

The conformations of the two GkApcT constructs were taken from the crystal structure (PDB: wild-type: 5OQT (resolution: 2.86 Å); M321S mutant: 6F34 (resolution: 3.13 Å)) (4). Human CATs, which are the human homology of GkApcT, is know to transport arginine (34). However, liposome-based transport assay showed that wild-type GkApcT cannot transport arginine. Mutating the methionine in 321 position to the serine as is seen in the

hCATs would allow GkApcT to transport arginine (4). Thus, the M321S mutant was made by sculpting the binding pocket which allowed the transporter to transport arginine, mimicking the behaviour of hCATs. The crystallographic binding pose of alanine to wild-type was used. The binding pose of alanine to M321S mutant was generated by truncating the arginine side chain atoms beyond the Cβ atom. This truncated binding pose, when compared with the binding pose that alanine adapted in the wild-type structure, exhibited very high similarity (RMSD < 1 Å). The missing residues were modelled using modeller v9.20 (203). A thousand models were generated and the model with the best QMEANDisCo (QMEAN with distance constraints) score (204) was selected. All organic molecules except cholesterol were removed. During prototypic runs, the buried pocket between the C-terminus of the ligand and the protein was not sampled by the TIP3P water. However, a crystal water was observed in this pocket in the crystal structure. Thus, this crystal water which bridges the ligand to the protein was preserved due to its importance for the ligand recognition. All other crystal waters were removed.

3.3.2 Bilayer Setup

The GkApcT, YneM helix and cholesterol obtained from X-ray crystallography were converted to a coarse-grained (CG) representation. The YneM helix is a single transmembrane helix native to *E. coli* and is part of the lac operon. The YneM helix was co-crystallised with GkApcT and was found to structurally stabilise the protein. The conversion was performed using the MARTINI v2.2 coarse-grain forcefield (178), where 4 atomistic particles were represented as one “bead”. The self-assembly process was initiated by adding 500 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) lipid molecules to the simulation box. Coarse-grained simulation was then performed for 25 ns in order for the lipids to form a bilayer around the protein.

The POPC bilayer was converted to a mixture of POPE and POPG (205) in a 3:1 ratio in both upper and lower leaflet to match the experimental conditions in liposome transport assay (4) and bacterial membrane (206). To maximise computational efficiency, the original simulation box with 500 lipids was truncated to be twice as small as the bounding box of the protein in all x, y, z dimensions. The bounding box of the protein is defined as the minimal box which encloses all the atoms in the protein. A 1 μs equilibration run was performed to allow POPE and POPG to equilibrate and the final frame was converted to atomistic resolution using the Backwards tool from the MARTINI toolkit (207).

The crystallographic coordinates of the GkApcT, YneM helix, cholesterol and substrate arginine or alanine were superimposed to the backward mapped POPE/POPG bilayer to replace their course-grain counterpart. The resulting bilayer system was solvated with TIP3P water (208) in a cubic box with periodic boundary conditions and neutralised with NaCl to reach a final ionic concentration of 150 mM. The resulting atomistic system was energy minimised using the steepest descents method and allowed to equilibrate for 25 ns in NPT ensemble with position restraint applied to GkApcT, YneM helix and cholesterol.

3.2.3 All-atom Molecular Dynamics

Two replicates of 500 ns simulation were performed with E115 protonated and deprotonated and with alanine or arginine bound. The AMBER ff99SB-ILDN forcefield (209) was used to describe protein and the Slipids forcefield (205) was used for the lipid molecules. The topology files for zwitterion arginine and alanine were obtained from AMBER parameter database (210). Protonation states of E115 were assigned using the pdb2gmx tool from Gromacs.

The timestep was 2 fs. The LINCS algorithm was used to constrain H-bond lengths and the long-range electrostatic interactions were calculated with the particle mesh Ewald (PME) method. The system was heated in NVT ensemble with V-rescale thermostat for 200 ps to reach a final temperature of 310 K. The system was then equilibrated in NPT assemble with Nosé-Hoover thermostat at 310 K and the Berendsen barostat for 5 ns. During the equilibration phase, all heavy atoms were position restrained at 1000kJ/mol/nm. The production runs were performed under the NPT ensemble, and the temperature was kept at 310 K using the Nose-Hoover thermostat and the pressure at one bar using the Parrinello-Rahman barostat. The equilibrated system was checked to make sure that the potential energy of the system, the box volume as well as the box area parallel to the x-y plane is stable at the end of the equilibration phase.

3.2.4 Reparameterisation of Protonated Glutamate

Consistent with Plamen Dobrev *et al.*'s (211) finding, when the protonated glutamate is represented with AMBER99SB-ILDN force field and solvated with TIP3P water, the protonated glutamate favours the syn orientation (Figure 3.2A) compared with the trans orientation (Figure 3.2B), contrary to what one would have expected. To examine the reason behind this behaviour, the potential energy of the C-C-O-H dihedral in protonated glutamate have been computed in vacuum using AMBER99SB-ILDN force field and at the level of HF/6-31G*, where both of them gave comparable results and have the trans orientation as the low energy form. Thus, to ensure that the correct (trans) orientation is favoured when protonated glutamate is solvated with TIP3P water, the model potential energy surface of the C-C-O-H dihedral in water was computed at the level of HF/6-31G* with polarizable continuum (IEFPCM) model water solvent (using Gaussian 09). As one would expect, the trans orientation was favoured compared to the cis orientation. Thus, we have reparameterised the C-C-O-H dihedral in protonated glutamate so that the

potential energy landscape obtained with explicit TIP3P solvent is consistent with the model potential energy landscape.

The protonated glutamate fragment was capped at both N- and C- terminus with acetyl and methyl group to represent the amino acid in the peptide chain. The capped glutamate was then solvated in a box of 4*4*4 nm^3 and neutralised with a sodium ion. To minimise the interaction between the capped glutamate and the sodium ion, the Cα of the glutamate was restrained to the centre of the box while the sodium ion was restrained to the edge of the box at 1000 kJ/mol/nm. Due to the instantaneous fluctuation of the water molecules, the potential energy of a certain C-C-O-H dihedral value was computed as the average potential energy of a 100 ns simulation, where the dihedral angle was restrained at 3000 kJ/mol/rad. During the parameterisation, the potential energy surface without the wrong dihedral term was computed by setting the existing dihedral potential of the C-C-O-H to zero. The resulting free energy landscape was fitted with a three-term Ryckaert-Bellemans potential to match the QM potential energy surface. The resulting dihedral with the new parameters were sampled again to check the fitting of the QM energy landscape (Figure 3.2).

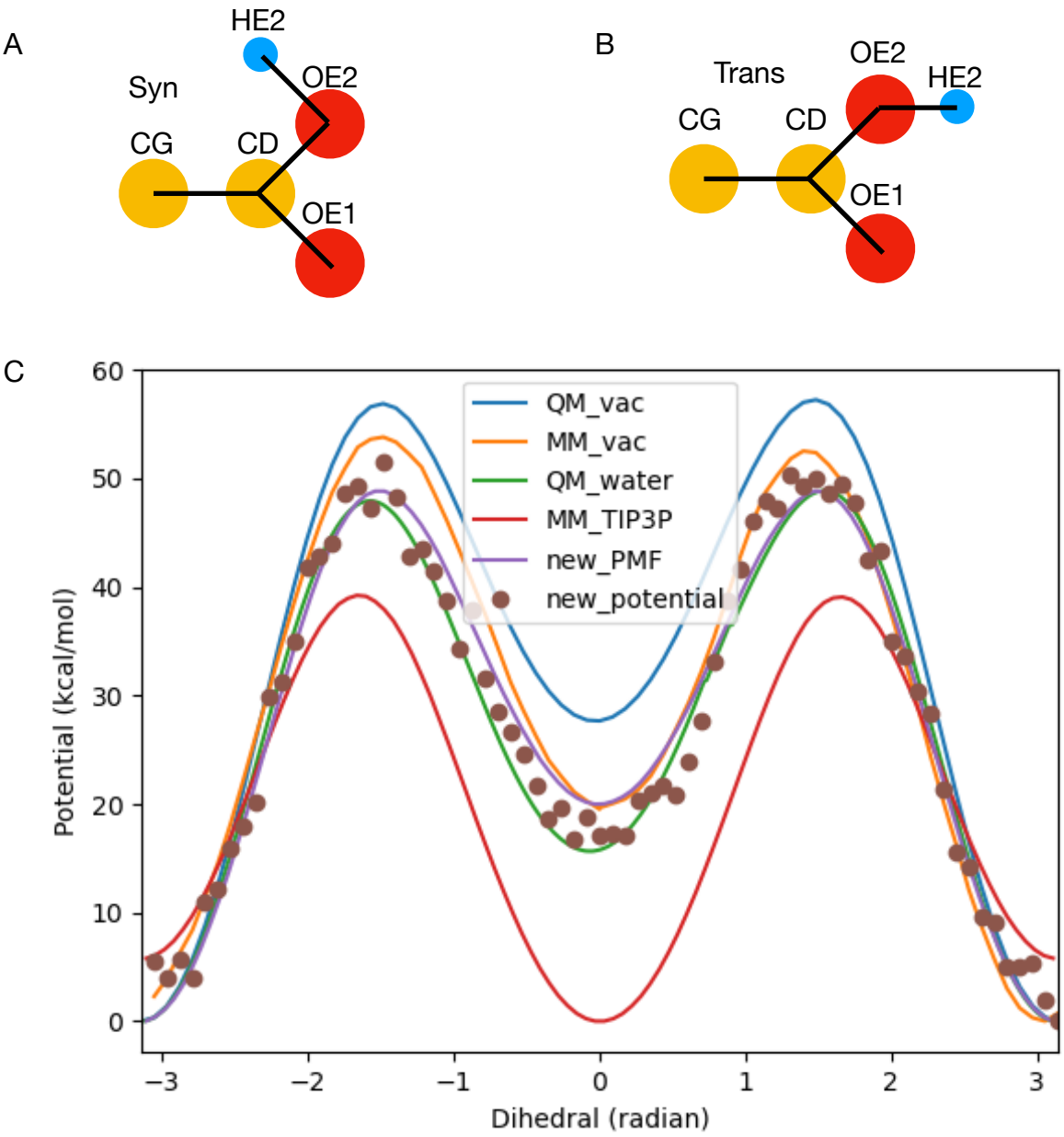


Figure 3.2*. Potential energy surface of the C-C-O-H dihedral of protonated glutamate. A.The Syn orientation of the C-C-O-H dihedral (atom name: CG-CD-OE2-HE2). B. The trans orientation of the C-C-O-H dihedral. C. Though the potential landscape of the dihedral (atom name: CG-CD-OE2-HE2) in vacuum is very similar between QM (blue line: HF/6-31G*) and MM (orange line: Amber99SB-ILDN force field), the QM (green line) and MM (red line) energy surface in solvent water is quite different. A new Ryckaert-Bellemans dihedral has been fitted to the dihedral so that the resulting energy landscape (purple for PMF profile and brown as the average potential energy) is consistent with QM result (green line) derived with implicit water.*

3.2.5 Umbrella Sampling Simulations and Potential of Mean Force Calculations

The collective variables used as reaction coordinate for umbrella sampling were defined as the distance between E115 and D237. The coordinate of E115 was defined as the centre of the Cα of residue number 113 to 117. The coordinate of D237 was defined as the centre of the Cα of residue number 235 to 239. A total of 25 windows were created from 13 Å to 19 Å with a step of 0.25 Å. The initial frames for umbrella sampling were extracted from the unbiased simulation of M321S GkApcT in complex with arginine with E115 deprotonated. The frame which was closest to the specified CV for each window was chosen as the initial frame for that window.

The umbrella sampling was performed in gromacs 5.1.2 (212) patched with plumed 2.3 (213). The bias applied to the collective variables was 1000 kJ/mol. The PMF profiles were generated using weighted histogram analysis method (WHAM) (214). The error associated with the free energy landscape was estimated using the block averaging method, where the error was defined as the standard deviation of the free energy landscape computed with the last 50%, 40%, 30%, 20% and 10% of the data.

3.2.6 Absolute Binding Free Energy Calculations

The calculation was carried out in Gromacs 2018. The free energy cycle is constructed as was described in chapter 2.4.2. The partial charges was annihilated with $\Delta\lambda = 0.1$, whereas 15 nonuniformly distributed λ values (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0) were used for the decoupling of van der Waals interactions. Similarly, to restrain the substrate to the protein, 10 nonuniformly distributed λ values (0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.35, 0.5, 0.75, 1.0) were used. Thus, a total of 36

windows for the substrate and protein complex simulation and substrate in solution simulation were employed.

During the charge elimination and restraint free energy calculations, an ordinary timestep of 2 fs was used. The Boresch restraint was found to introduce a high frequency motion to the system. This motion was normally shadowed by the much stronger electrostatic and van der Waals (vdw) interactions. However, in the later stages of the decoupling of the vdw interactions, since both electrostatic and vdw interactions were mostly turned off, this high frequency motion became the predominant player in the system. To accurately account for this high frequency motion, a smaller timestep was required. Thus, during the vdw interaction decoupling stage, a 1 fs timestep was used for the window (λ = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75) and a timestep fo 0.5 fs was used for the rest of the vdw windows.

To restrain the protein in the inward-open or inward-occluded state, a harmonic distance restraint was applied to the distance between E115 and D237 as was described in the previous section 3.2.5. The centre of this restraint was set according to the energy minimum as determined in the PMF profile.

After equilibration, $dH/d\lambda$ data were collected from 30 ns production run. The relative position and orientation of the substrate to the protein were restrained using one distance (G44:N, A:C), two angles (T43:C, G44:N, A:C and G44:N, A:C, A:CA) and three dihedrals (T43:CA, T43:C, G44:N, A:C; T43:C, G44:N, A:C, A:CA and G44:N, A:C, A:CA, A:N). The centre of the restraints were set based on the average value of these distance, angles and dihedrals according to a 100 ns unbiased equilibrium run. The force constant of these restraints were set to 10 kcal/mol/Å and 10 kcal/mol/radian accordingly. The restraint free

energy between the decoupled ligand and the protein was evaluated analytically (182). The collected dh/dλ was processed using alchemical analysis (215) where the first 5 ns of data were discarded as equilibrium time. To obtain the free energy difference, the integration of the dh/dλ was done using multistate Bennet acceptance ratio (MBAR) method and the error was estimated from the MBAR error with error prorogation.

3.2.7 Protonation Free Energy Calculations

The free energy of protonation or deprotonated was done in a fashion similar to the previous section 3.2.6. The protonated glutamate was alchemically changed to the deprotonated glutamate. The topology for the alchemical transformation was generated using pmx (216). The morphing of the partial charge was done with uniform lambda difference of $\Delta\lambda = 0.1$ and the morphing of the rest of the interactions (bonds, angles, dihedrals and vdws) were done with a uniform lambda difference of $\Delta\lambda = 0.25$. Thus, a total of 16 windows were run for 30 ns and the dh/dλ was analysed in a similar fashion using alchemical analysis (215), where the first 5 ns were discarded as equilibrium.

As is described in section 2.5, when the electrostatic interactions in the system were computed using the PME method, the system has to stay neutral. However, when computing the protonation free energy, the total charge of the system cannot stay neutral as the protonated glutamate and the deprotonated glutamate have different total charges. To solve this problem, an alchemical chloride was brought in to compensate for this change in total charge. Thus, when the protonated glutamate was changed into deprotonated glutamate, a chloride ion was also changed from having -1 charge to neutral. To prevent the interaction between the key glutamate and the alchemical chloride ion, the chloride ion was position restrained to the corner of the box, where the chloride ion was around 8 nm away from the key glutamate.

The free energy of decharging the chloride ion was then deducted from the final result. The free energy of the chloride ion charge annihilation was computed by simulating a single chloride ion in a cubic box of water. The error introduced due to the non-neutral box was corrected with the Rocklin correction (217) (as is described in method section 2.5). A range of box sizes with edge length from 30 to 100 Å with a step of 10 Å were tested to ensure that the final result didn't exhibit any dependency on the size of the box (Figure 3.3). For each box size, thirty repeats were run and the free energy of the chloride ion charge annihilation was taken as the average of the eight box size times three repeats.

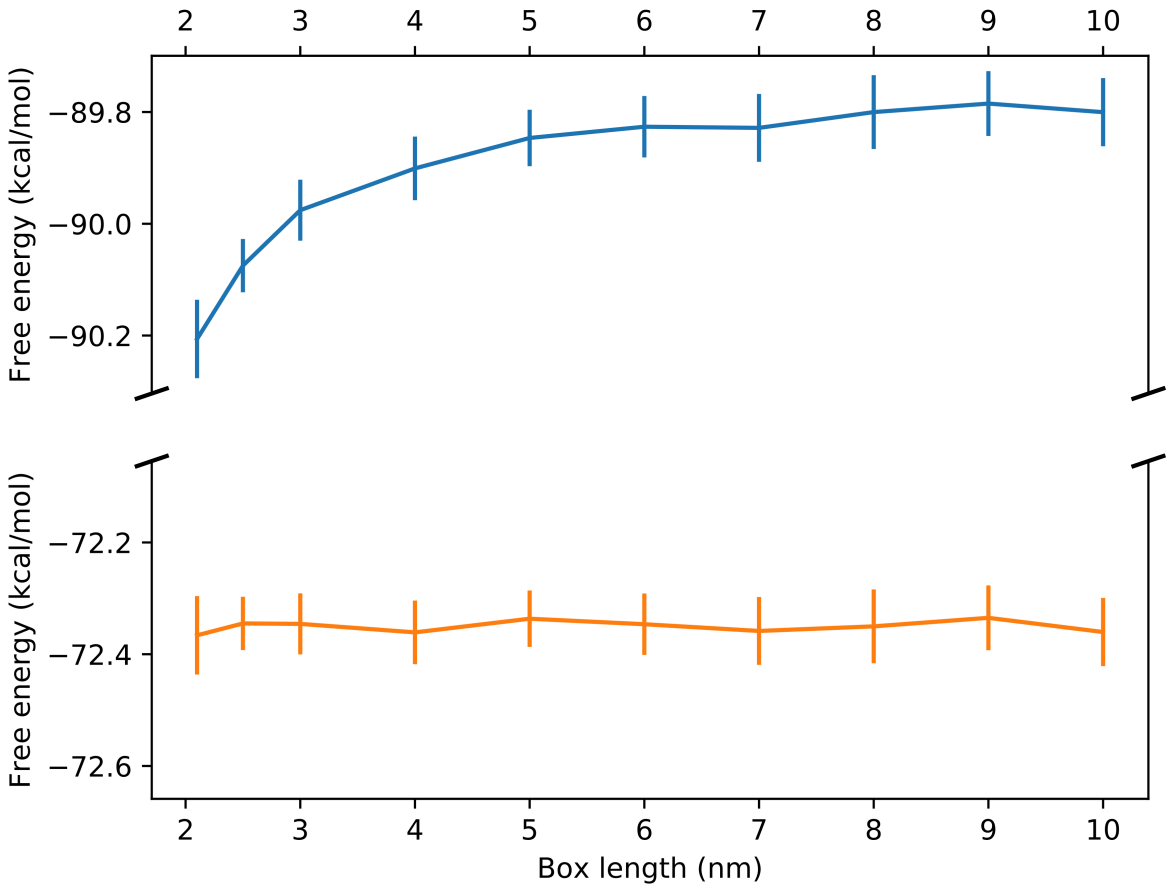


***Figure 3.3**. Finite size correction for the chloride ion charge annihilation free energy. The free energy of the charge annihilation of the chloride ion shows significant dependency on the box size (blue line). However, upon using the Rocklin correction, the size dependency is diminished (orange line). The final charge annihilation free energy is taken as the average of between box 3 nm to box 10 nm.*

3.3 Results

3.3.1 Protonated Glutamate Stabilised the Transporter in the Inward-occluded State

Functional assay showed that GkApcT couples the transport of the amino acid to the proton electrochemical gradient (4). However, the mechanistic detail of this proton-coupling is unknown. The crystal structure of the GkApcT was solved in the inward-occluded state, where the extracellular gate is closed by the crossing over helices and, thus, tightly sealed. The intracellular gate, on the other hand, is partially opened. However, the close proximity of the residue side chains from TM3 and TM6 indicates that the intracellular gate is occluded enough to prevent the release of the substrate. To allow the eventual exit of the substrate, conformational change from the inward-occluded state to the inward-open state would require the separation of TM3 and TM6. Two keys residues, the E115 on TM3 and the D237 on TM6 were found to be directly opposing each other in the crystal structure. PropKa (218) suggested that the glutamate E115 have a pka of 8.2 compared with 4.38 for D237. Thus, under neutral pH, It is assumed that the E115 is protonated and it is postulated that the conformational change from the inward-occluded state to the inward-open state is triggered by the deprotonation of the E115.

To explore this hypothesis, 500 ns of unbiased molecular dynamics simulations were performed for GkApcT embedded in a mixed lipid bilayer of POPE/POPG in 3:1 ratio (Figure 3.4A). Different substrates (arginine or alanine) and different protonation state (E115 protonated and deprotonated) were tested. When E115 was deprotonated, a significant expansion between TM3 and TM6 was observed (Figure 3.4BC), which was especially predominant between E115 and D237. The same expansion was not observed

when the E115 was protonated. After the expansion, the intracellular gate is significant widened and is consistent with the description of an inward-open state.

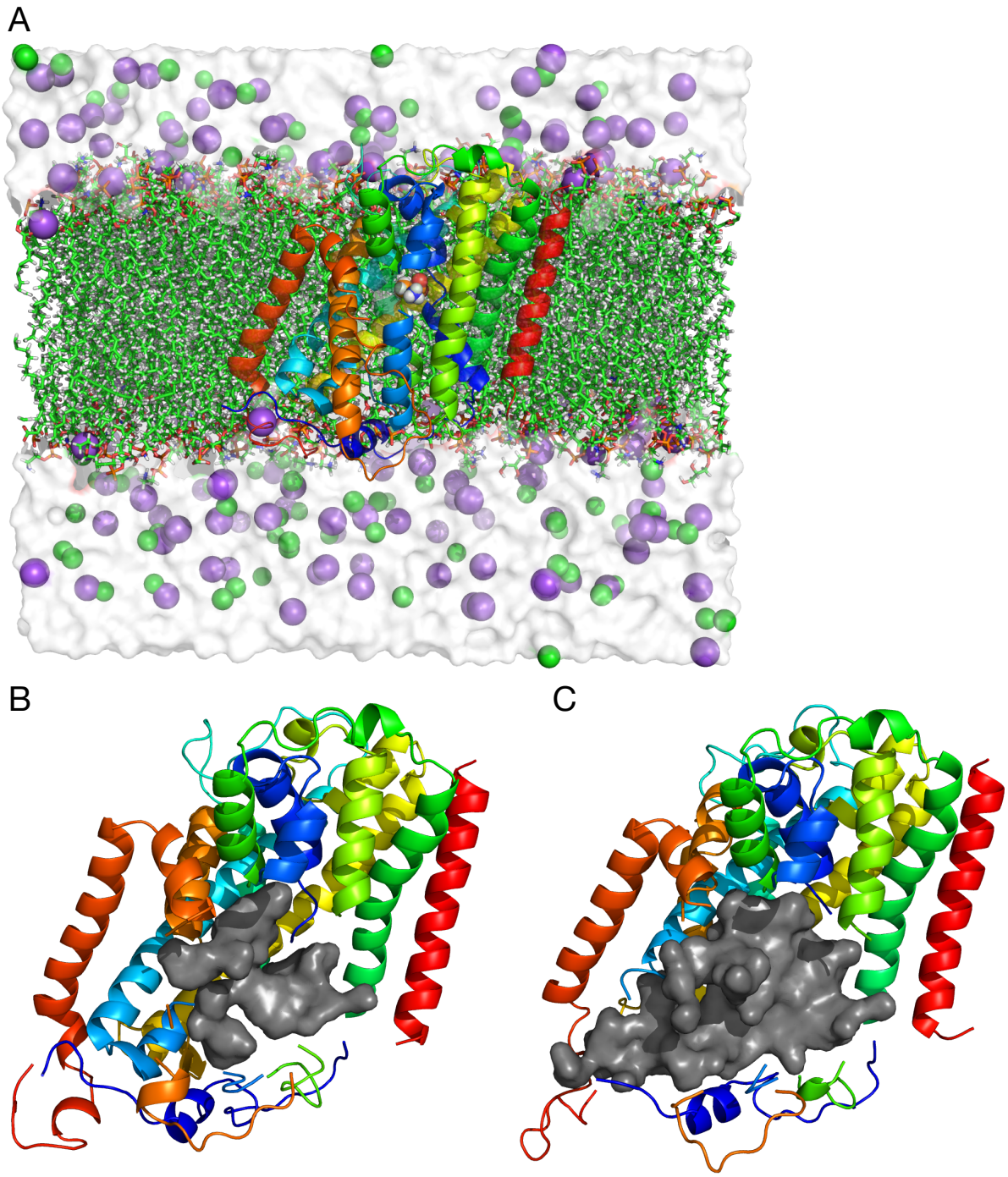


***Figure 3.4**. Unbiased simulation of GkApcT in a lipid bilayer. (A) A snapshot of the simulation system, where the protein shown as cartoon is embedded in the lipid bilayer of a mixture of POPE/POPG. The system is solvated with NaCl (purple and green spheres). The substrate alanine is shown as spheres in the binding pocket. (B) The crystallographic inward-occluded state. (C) The inward-open state obtained in the unbiased MD simulation. The binding cavity (grey surface) is massively expanded when transition from inward-occluded to inward-open state. In the inward-open state, the binding cavity is linked to the bulk allowing the substrate to exit.*

3.3.2 The Inward-open State Permits Substrate Exit

The inward-open state is characterised by the opening of the intracellular gate and the releasing of the substrate. To test if the inward-open state obtained from the unbiased simulation fits this description, the substrate exit path was examined to test if any further energy barrier exists. Starting from the inward-open state, the alanine was pulled away from the binding pocket using steered MD where a moving restraint was applied to increase the distance between the alanine (defined as the Cα atom) and the binding pocket defined as the centre of the Cα atoms of residues I40, T43, G44, F231, and I234. These residues were chosen as they form hydrogen bonds with the substrate alanine. As is shown in Figure 3.5, after the initial energetically costly breaking of the hydrogen bonds between the binding pocket and the substrate, the energy landscape is flat. Thus, in the inward-open state, after the initial dissociation from the binding pocket, no more energy barrier was encountered when moving the substrate to the bulk.

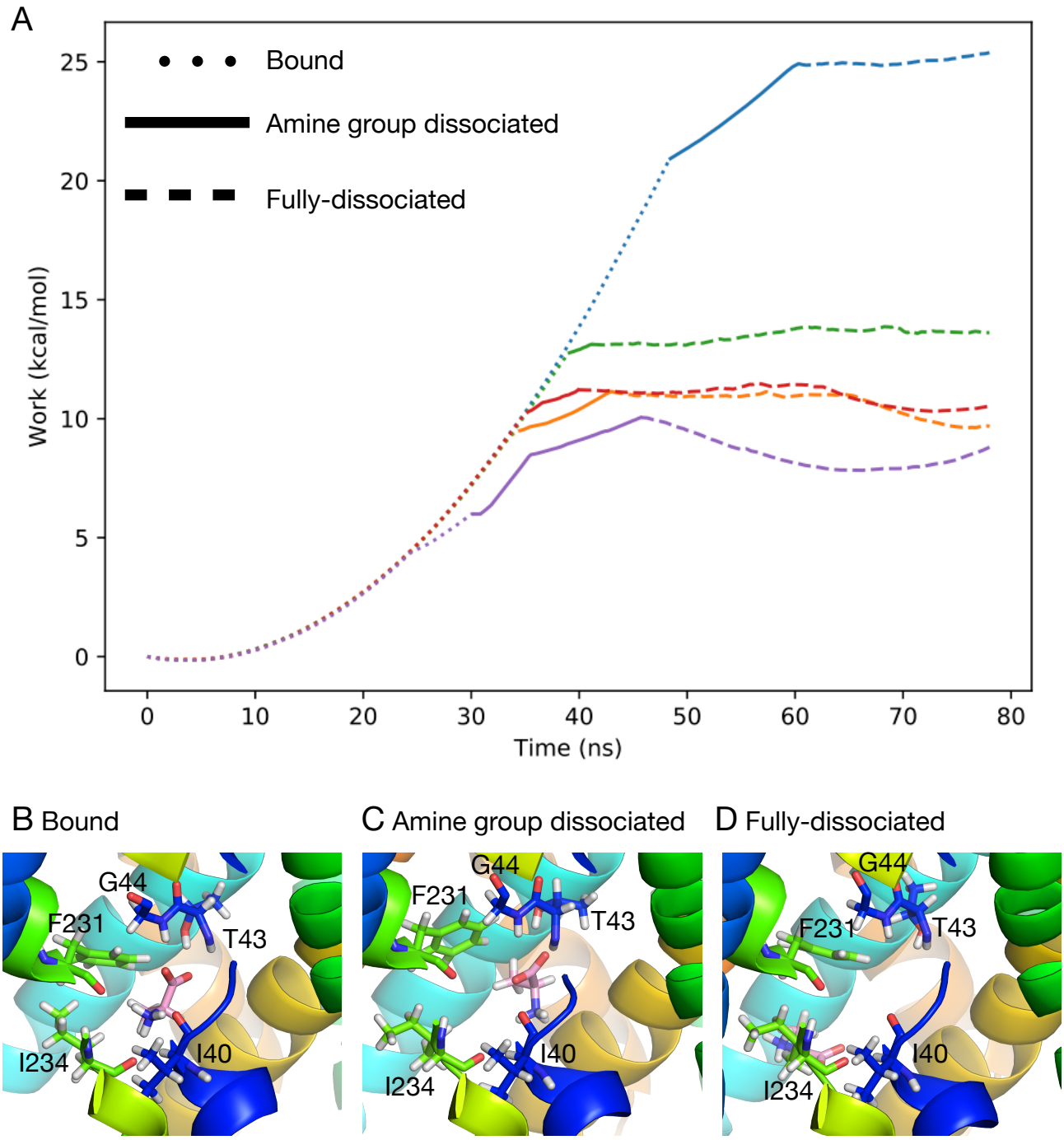


***Figure 3.5**. Work done as a function of time during the steered MD. The dotted line represents the fully bound state (B), where both the carboxyl group and the amine group are forming hydrogen bonds with the protein. The solid line represents the state where the hydrogen bonds between the amine group and protein are broken but the interactions between the carboxyl group and the protein are intact (C). The dashed line represents the fully dissociated state (D). After dissociation, no more work is required as is shown by the plateaus in the profile. Each colour represents a different repeat.*

To further test the hypothesis that the substrate could freely dissociate in the postulated inward-open state, the restraint between the substrate and the binding pocket was turned off from the last frame of one steered MD. Thus, the alanine was allowed to freely explore the intracellular vestibule. After 180 ns of simulation, the alanine was found to make its way back to the binding pocket and exhibited the same binding pose as the crystal structure (Figure 3.6BD). The RMSD of the alanine between the last 35 ns of the simulation (Figure 3.6D) and the crystal structure (Figure 3.6C) has a mean of 0.8 Å and a standard deviation of 0.2 Å. To understand the relationship between the conformational change and the exit of the substrate, the crystallographic inward-occluded state was compared with the MD-derived inward-open state and it is found that the separation between the A238 in TM6 and V328 in the TM8 is much larger in the inward-open state. These two residues form the hydrophobic gate which occludes the alanine exit pathway in the inward-occluded state (Figure 3.6A). In the inward-open state, the separation between these two residues increased to a minimum of 6.5 Å (Figure 3.6B) and, thus, allowing the effortless exit of the alanine.

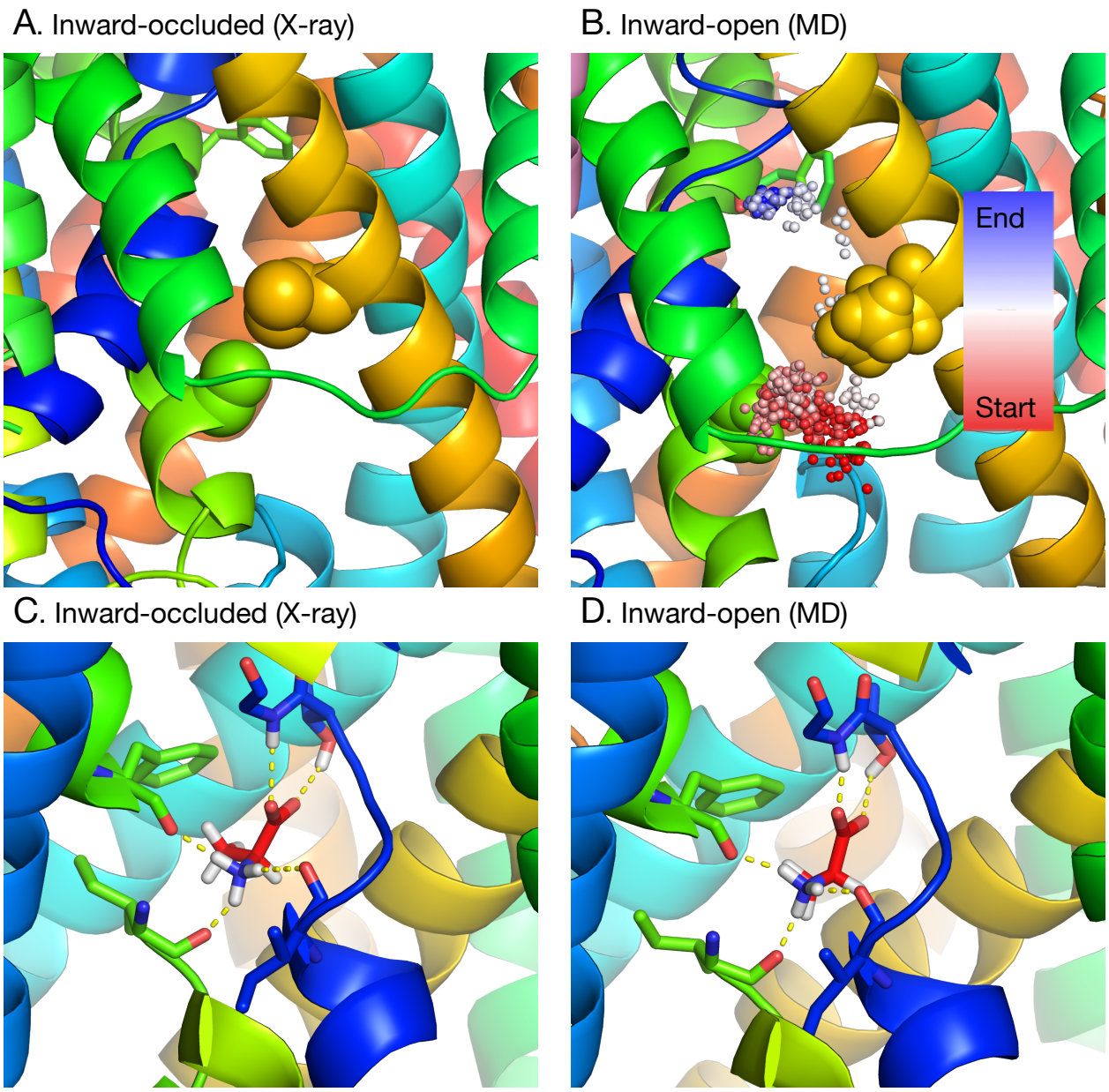
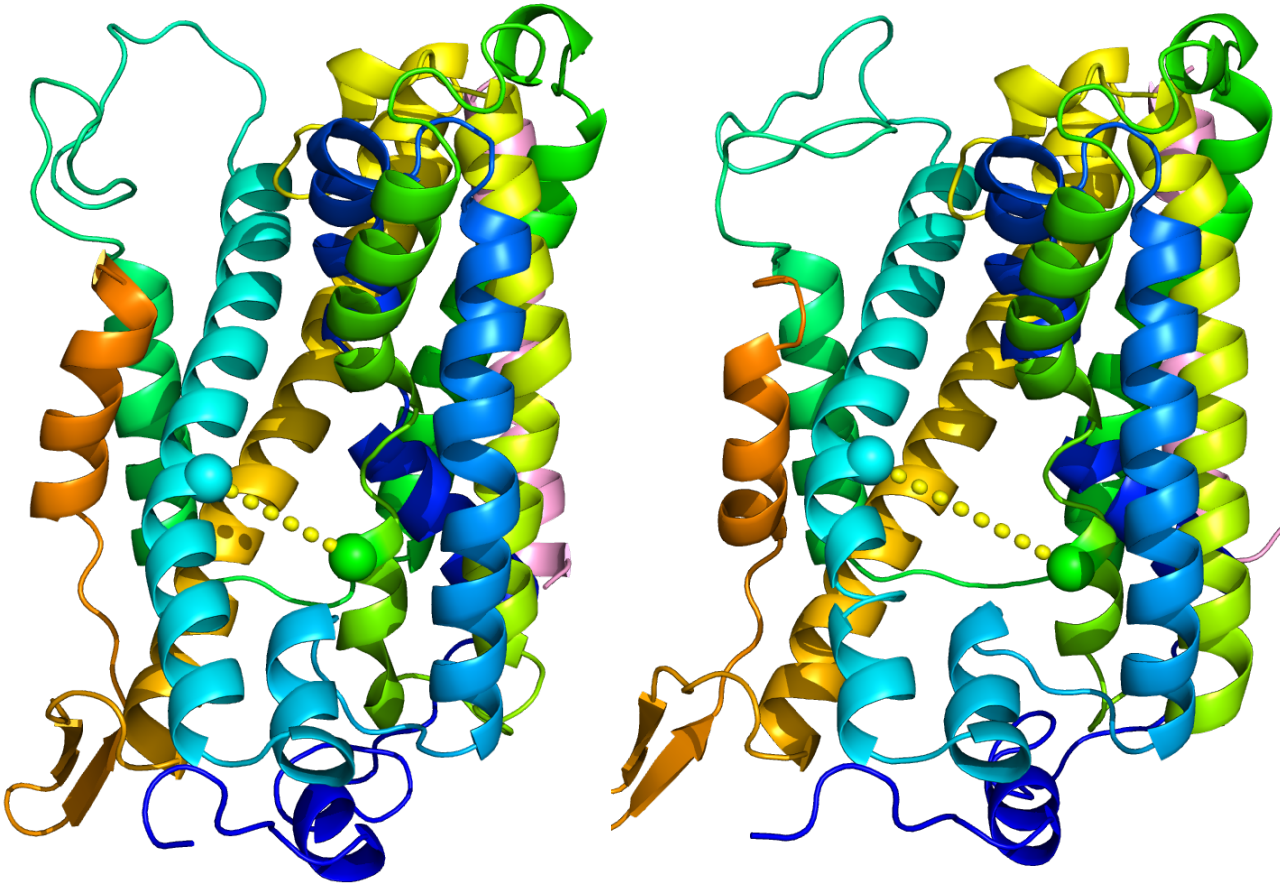


Figure 3.6*. The rebinding of alanine to the binding site. In the inward-occluded state (A), the substrate exit path is obstructed the hydrophobic gated formed by A238 (green sphere) and V320 (yellow sphere). The obstruction is alleviated in the inward-open state (B), where the substrate could pass this gate during the rebinding process (the trajectory of the CA is shown as small dot from red to blue). After the rebinding, the substrate alanine in the inward-open state (D) adopts the same binding pose as the crystal structure (C), where all the hydrogen bonds between alanine and the protein are present.*

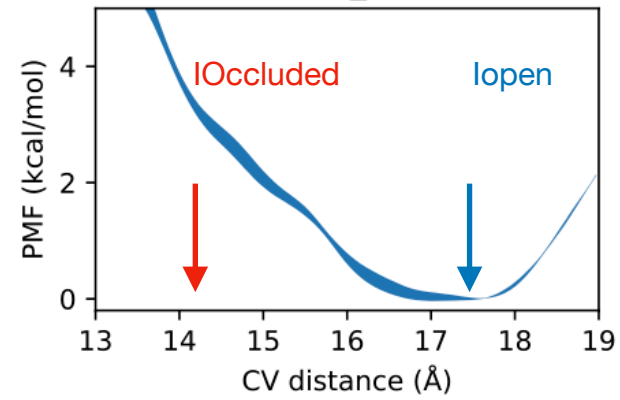
A. Inward-occluded state

B. Inward-open state

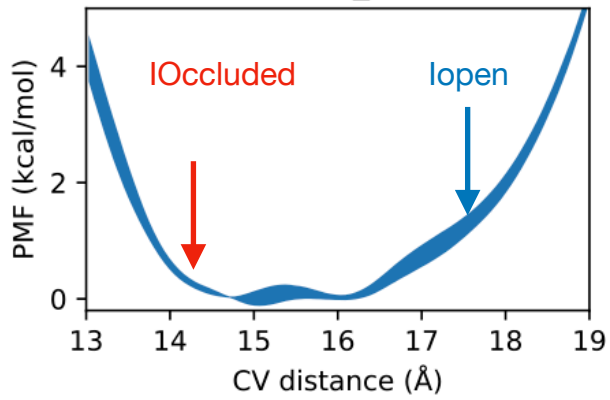
Out



C.

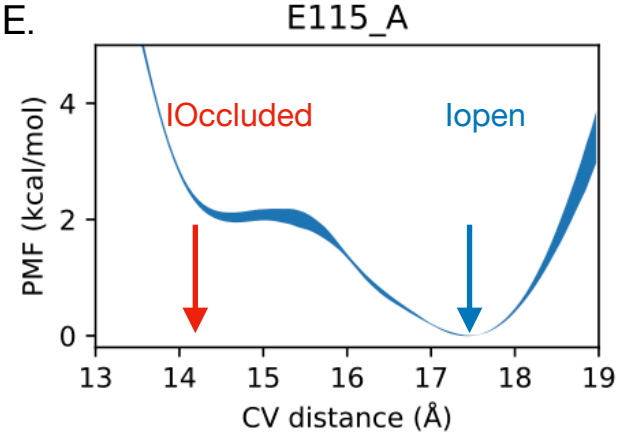


D.



In

E.



F.

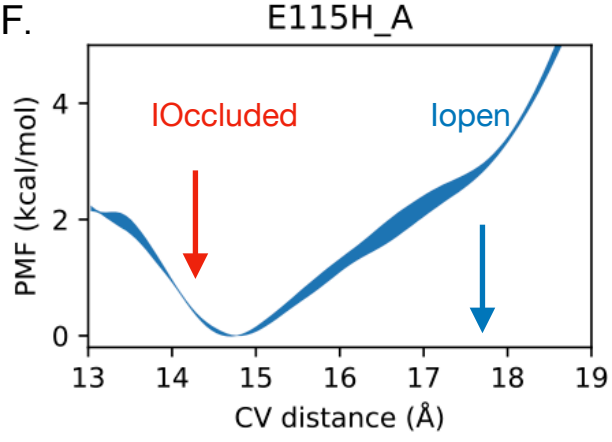


Figure 3.7. A. The transition from inward-occluded state (A) to the inward-open state (B) is reflected by the increase in the separation between E115 (cyan sphere) and D237 (green sphere). The X axis of the PMF profile (C-F) is the collective variable, which is the distance between the centre of the Cα atoms of residues 113–117 and the centre of the Cα atoms of residues 235–237. The location of the inward-occluded state and inward-open state are indicated by red and blue arrow respectively. Four conditions: apo state with deprotonated E115 (C), apo state with protonated E115 (D), alanine-bound state with deprotonated E115 (E), alanine-bound state with protonated E115 (F) are shown and the error is computed using block averaging method and is represented as the width of the line. (Method 3.2.5)

Both the steered MD simulations and the rebinding of the alanine showed that no steric barrier exists in the inward-open state to prevent the release of the substrate and our inward-open state is likely to be the true inward-open state.

3.3.3 PMF Calculations Showed that E115 Plays a Key Role in the Proton-coupling of the Transport Mechanism

To investigate the effect of the protonation state of the E115 on the transition from inward-occluded state and the MD-derived inward-open state, the free energy difference between these two states was estimated using the umbrella sampling method (chapter 3.2.5). The simulations were run for four conditions, where the E115 was protonated or deprotonated and the transporter was in the apo state or the alanine bound state. From the unbiased simulations, the distance between the E115 and D237 was found to be a good indicator of the intracellular cavity. Thus, the collective variable for the PMF profile was chosen as the distance between the E115, which is defined as the centre of the Cα atoms of residues from 113 to 117 (TM3) and the D237, which is defined as the centre of the Cα atoms of residues from 235 to 239 (TM6) (Figure 3.7AB). The initial frames for the umbrella sampling were taken from the unbiased simulations.

When E115 was deprotonated and alanine was bound, the inward-open state was approximately 2.2 kcal/mol lower than the inward-occluded state (Figure 3.7E) and is consistent with our hypothesis that the deprotonation of E115 triggers the conformational change from the inward-occluded state to the inward-open state. Interestingly, when E115 is deprotonated, the local energy minimum at the inward-occluded state (Figure 3.7E) is not observed in the apo state (Figure 3.7C), signalling that the binding of the substrate stabilised the inward-occluded state. The protonation of E115, on the other hand, had a similar effect as the binding of the substrate, where the PMF profile in the presence of protonated E115 and alanine displayed a single energy well in the inward-occluded state (Figure 3.7F).

3.3.4 The Protonation Free Energy Confirmed that Protonation of the E115 Stabilised the Inward-occluded State

After using PMF calculations to show that the protonation state of E115 would affect the equilibrium between the inward-occluded and the inward-open state, the result was independently verified by computing the protonation free energy at different conformations. Since deprotonated E115 stabilised the inward-open state, the deprotonation of E115 would be energetically favourable in the inward-open state. The reverse should be observed for the inward-occluded state, where the deprotonation of E115 would be energetically unfavourable. As is shown in Table 3.1, consistent with our hypothesis and the PMF calculations, the free energy of the protonating of E115 is positive in inward-open state and negative in the inward-occluded state.

Table 3.1. Summary of free energy to protonate the E115 sidechain. All energies in kcal/mol.

	Inward-occluded	Inward-open
Apo	- 3.62 ± 0.2	+ 2.72 ± 0.1
Alanine bound	- 6.01 ± 0.2	+ 7.01 ± 0.2

Errors are estimated by MBAR under error propagation (see methods).

3.3.5 The Binding Affinity of Alanine is Constant Across Different Conditions

As is described in section 3.3.2, the conformational change from the inward-occluded state to the inward-opens state eliminated the potential steric clash which might prohibit the release of the substate. It is also interesting to investigate if the conformational change to the inward-open state would loosen the binding of the alanine and, thus, facilitates the egress of the alanine. However, as is shown in Table 3.2, though, numerically, the binding affinity is decreased slightly by 0.2 kcal/mol in inward-open state compared with the inward-occluded state, it could be concluded that the binding of the alanine is independent of the conformation or the protonation state of E115.

Table 3.2. Absolute free energies for alanine. All energies in kcal/mol ± standard

	Alanine in water	Alanine in E115_Inward-occluded	Alanine in E115_Inward-open	Alanine in E115H_Inward-occluded	Alanine in E115H_Inward-open
∆G	-47.73 ± 0.01	-5.89 ± 0.04	-5.55 ± 0.07	-5.67 ± 0.08	-5.27 ± 0.14

*Errors are estimated by MBAR under error propagation (see methods).

3.4 Discussion

In this chapter, the hypothesis that the deprotonation of E115 would facilitate the opening of the intracellular gate as is suggested by Jungnickel *et al.* (4), was tested. The PMF profile is consistent with the hypothesis in that the deprotonation of E115 did destabilise the inward-occluded state and allowed the protein to perform the conformational change from the inward-occluded state to the inward-open state. Though it is shown that the steric clash in the inward-occluded state, which prevented the egress of substate, was eliminated in the MD-derived inward-open state, it is worth mentioning that a partial open state was recently found for LeuT by single-molecule fluorescence resonance energy transfer (72). It is argued that this partial open state plays an important physiological role in the transport cycle and the possibility that there might be a more open state compared with the inward-open state cannot be ruled out .

The protonation free energy of the E115 supports the PMF results in that the favourable protonation state of E115 for each conformation is the protonation state which favoured that conformation. The trend that protonated E115 stabilised the inward-occluded state and deprotonated E115 stabilised the inward-open state, held regardless of the the presence of the ligand. However, it is worth noting that the absolute protonation free energy calculated based on fixed charged model can be problematic (219, 220) and certainly more work is required in this area. However, in this project, the relative free energy difference instead of the absolute free energy is of interest to me and the result is consistent with my hypothesis.

It is a bit surprising that the binding affinity for the alanine was not reduced upon the conformational change. However, it is worth noting that the binding pocket hosting the alanine is deeply buried inside the transporter. The conformational change from the

inward-occluded state to the inward-open state has brought very little change to the binding pocket, where all the key interactions between the alanine and the binding pocket are still intact. Thus, it is expected that the binding free energy stayed the same during the conformational change. It has been argued that PMF calculations could also be used to compute the binding free energy as is shown in the binding of xylopyranose to XylE (221). However, no substrate dissociation pathway can be found for the inward-occluded state and, thus, PMF calculations cannot be used. The ABFE method, which does not require an exit pathway, was thus, used in this project.

In this project, the MD simulations were used to understand the transport cycle of the GkApcT. Similar methods has been used for other transporters in the APC superfamily such as the sodium-coupled LeuT (222, 223). However, it has also been noted that the local conformational change observed in short MD simulations might not represent the postulated global conformational change (224). Examples such as the semiSWEET transporter (225) or the GluT1 transporter (226) showed that very long simulations are necessary to understand global conformational change. However, even very long MD simulations are short if the whole transport cycle were to be considered. To understand the whole transport cycle, approaches such as Markov state modelling have been shown to be helpful as was shown for the $PepT_{so}$ transporter (227).

This project has only dealt with part of the transport cycle (from inward-occluded state to the inward-open state). In terms of the rest of the transport cycle, where the proton went after leaving the E115 is a very relevant question. However, due to the physical and chemical property of the proton, analysing a single proton traversing a complex biological system is very difficult, though there has been some promising work in this direction (228). It is worth noting that the exit of the proton is likely to involve water. The water

penetration into the binding pocket has been shown be crucial for the conformational change as well as the release of the substrates as was shown in the DAT (69), SERT (229), XylE (230) and LeuT (231) (as is discussed in Section 1.5.4). Furthermore, constant pH simulations has also been shown to be a useful tool in studying proton mediated transport in the sodium-proton antiporter (232).

Understanding how GkApcT works allows us to probe the transport mechanism of a SLC36 family proton-coupled amino acid transporter, which has been shown to shuttle between the plasma membrane and the lysosome and use the proton gradient to facilitate many important biological processes. Elucidating the transport cycle of the proton-coupled transport that could also facilitate the understanding of transporters sharing the same structure fold but use different electrochemical gradient (e.g. sodium gradient) and the understanding of how secondary-active transporter works in general.

3.5 Conclusion

In this project, mechanistic insight into the role that protons play in the transport cycle have been provided. A model of explaining the release of the proton and the substrate from the inward-occluded state has also been provided. The results show that the deprotonated E115 triggers the opening of the intracellular gate that allows the transition from the inward-occluded state to the inward-open state and the exit of the substrate.

4. The KDEL Trafficking Receptor Exploits pH to Tune the Strength of an Unusual Short Hydrogen Bond

4.1 Introduction

KDELR belongs to the PQ-loop superfamily, which plays a key role in nutrient transport, lysosomal transporter and tissue development (233). KDELR is located in the cis-Golgi (234) which has a pH of 6.2, where it binds proteins with a C-terminal KDEL tag. This binding event is pH dependent and maximum binding occurs at pH below 6 (129). Thus, it is hypothesised that KDELR binds ER protein at pH 6 (Golgi) and release them at a higher pH such as that found in the ER where the pH is 7.2 to 7.4 (128).

The molecular mechanism of this pH-dependent binding remains elusive. Recently the crystal structure of the KDEL receptor has been solved in both KDEL-bound (pH 6.0) and apo state (pH 9.0) (Figure 4.1). The binding pocket in the KDEL-bound state and the apo state are very similar, whereas the most significant difference is that in the KDEL-bound state, a short hydrogen bond (~ 2.5 Å) was formed between the oxygen of the carboxyl group of E127 and the oxygen in the hydroxyl group of Y158, which expands to 2.8 Å in the apo state. Both cellular retrieval and *in vitro* binding assays (133) highlighted the importance of a conserved histidine (H12), which is located directly above E127 and Y158. Being a residue that could adopt different protonation states between the ER and the Golgi, it is proposed that the H12 may function as a pH sensor for the KDEL retrieval system (133).

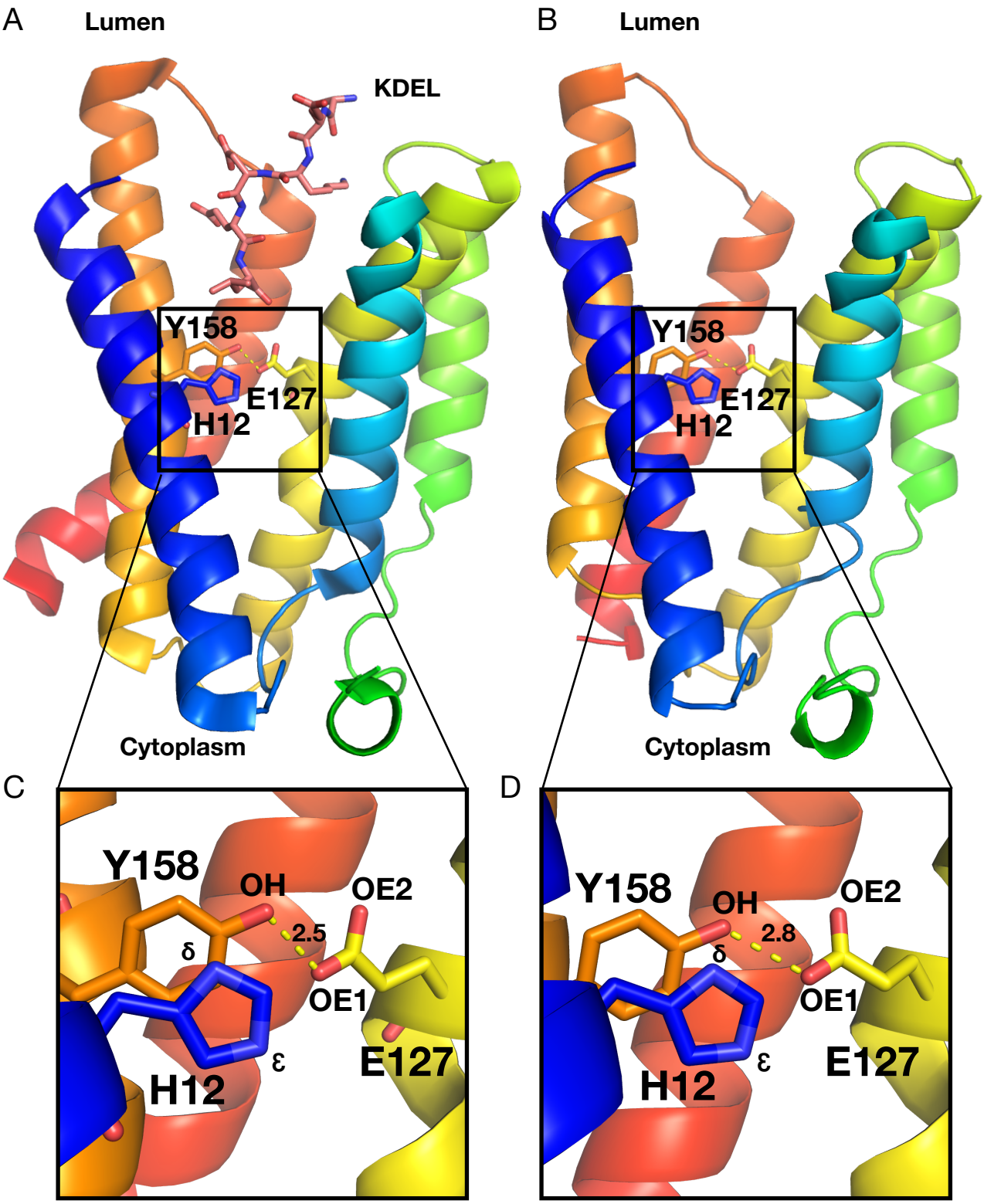


Figure 4.1. *Structures of the KDEL receptor used in this study. (A) and (C) KDEL-peptide bound (PDB: 6i6h) and (B, D) the apo (PDB: 6i6b) state. The hydrogen bond between E127 and Y158 is slightly shorter when peptide is bound at 2.5 Å (C) compared to 2.8 Å in the apo form (D). Atom labels used within this study are also shown along with the position of His12, which has previously be suggested to act as the pH sensor.*

Low-barrier hydrogen bonds (LBHB) have been found to play an important role in enzymatic catalysis reactions (1). Compared with ordinary hydrogen bonds, the proton in the LBHB is shared more between the donor and the acceptor and interacts with both donor and acceptor covalently instead of tightly bonded to the donor and interacts electrostatically with the acceptor. In some enzymatic reactions, the enzyme alters the pKa of the substrate and some specific residues (e.g. histidine) so that they have a matching pka which allows a LBHB to form and triggers the catalysis (235, 236). Experimentally, deuterium enrichment is observed compared with the normal hydrogen bond (for example, the OH bond in water) in NMR experiments (237), showing that the hydrogen bond is stronger in a LBHB. Computationally, the LBHB can be tested via quantum mechanical methods, where one criterion is that the adiabatic energy barrier of transferring the proton from hydrogen bond donor to acceptor is smaller than the zero-point energy (ZPE) of the protons (236).

The hypothesis here is that in the Golgi, the protonation of H12 matches the pKa of Y158 and E127 and allows a LBHB to form, which holds TM5 (E127) and TM6 (Y158) firmly together, and, thus, ensures the tight binding of the KDEL sequence. When KDELR is in the ER, H12 is deprotonated in the sightly higher pH, which disrupts this short hydrogen bond by having an unmatched pKa and releases the KDEL. To explore this hypothesis, a combination of molecular mechanics and quantum mechanics calculations were performed. The result shows that the protonation of the key histidine (H12) is essential for the formation of the LBHB and the stabilisation of the active state of the receptor.

4.2 Methods

4.2.1 QM/MM Simulations

The KDEL-bound KDEL receptor and the apo KDEL receptor were obtained from the PDB (apo: 6I6B; KDEL-bound: 6I6H) (133) and embedded in a DMPC bilayer as described in the previous chapter (3.3.2). The QM region was defined as the three residues (H12, Y159 and E127) and was described at the level of B97-3c (193). As it has been shown that the inclusion of the charge residues from the whole protein is crucial for the QM/MM calculations (238). The QM region cutoff was set to 3 nm so that the partial charge of the whole protein would be taken into account. The force field amber ff14SB (239) was used to describe the protein and LIPID17 (240) was used to describe the lipid. The simulation was run with ambertools19 (241) sander module in conjunction with ORCA 4.2.0 (242). For the free energy calculations, the collective variable was chosen as the distance between the oxygen in the hydroxyl group of the Y158 and the closest oxygen in the carboxyl group of E127. 11 equally spaced windows from 2.4 Å to 3.4 Å with a restraint strength of 25 kcal/mol/Å were run with 1 fs timestep for 20 ps. The free energy profile was constructed with WHAM 2.0.9.1 (243).

4.2.2 QM Geometry Optimisation

The crystallographic coordinates of residues H12, E127 and Y158 were taken from both apo and KDEL-bound states as described in the previous section. To ensure consistent atom naming between these two structures, the oxygen in the carboxyl group of the E127 which is closest to the Y158 is named as OE1, whereas the other oxygen is named as OE2. The N-termini of the three residues were capped with an acetyl group and the C-termini were capped with N-methyl group. Hydrogen atoms were added to the crystal structure and different protonation states of the three residues were assigned using

Maestro. The structure of the H12, E127 and Y158 complex, as well as each capped amino acid monomer and all the possible dimers (H12/E127; H12/Y158) were independently optimised with non-hydrogen atoms constrained using B3LYP together with def2-TZVP (244) basis set with DFT-D4 (245) for empirical dispersion correction, with vibrational frequency done at the end of optimisation using ORCA 4.2.1. Due to a known bug in the ORCA frequency calculations, the following flags (%freq Hess2elFlags[0] 0 Hess2elFlags[3] 0 end) were added as is suggested by the developer (https://orcaforum.kofo.mpg.de/viewtopic.php?f=26&t=5566). The CVBi (246) and AIM analysis were performed with Multiwfn 3.7 (247).

4.2.3 QM Interaction Energies

Single point energies were computed on the optimised structure with DLPNO-CCSD(T)/aug-cc-pVTZ (248) at the level of normalPNO, with aug-cc-pVTZ/JK for RIJK acceleration and aug-cc-pVTZ/C (249) for electron correlation calculations. The thermal correction for Gibbs free energy was computed from the frequency calculations and corrected with quasi-harmonic correction (250) and a ZPE scaling factor or 0.985 (251) using Shermo (252) to 310 K. For complexes with more than one amino acid, the energy is corrected for basis set superposition error (BSSE). BSSE is defined as half (253) of the BSSE energy. To obtain the interaction between the E127(E) and Y158(Y) under the influence of the H12(H), the interaction free energy is defined as:

$$G = G(EYH) - G(EH) - G(YH) + G(H) \tag{4.1}$$

Where $G(EYH)$, $G(EH)$, $G(YH)$ and $G(H)$ are the free energy of the geometry optimised E127, Y158, H12 complex, geometry optimised energy of the H12 in complex with E127 or Y158 and the energy of geometry optimised histidine, all with counterpoise corrected.

4.2.4 QM Proton Transfer Energy

The energy of transferring a proton from tyrosine to glutamate was computed at the level of PWPB95 (254)-D4/def2-QZVPP//B3LYP-D4/def2-TZVP. The calculation was performed in the presence of three different protonation states of the histidine. During the geometry optimisation, the positions of all the non-hydrogen atoms were constrained and the bond between the proton and the closest oxygen was also constrained. The vibration mode corresponding to the motion of the key proton was taken as the mode with the highest infrared intensity and confirmed with visual inspection using Gaussview 6.0. The ZPE corresponding to this mode was computed using Shermo (252).

4.2.5 MM Protonation Free Energy Calculations.

A double bilayer box with KDEL-bound and apo protein were constructed. The two proteins were placed 3 nm apart in a box of 8 nm * 8 nm * 17 nm and solvated with TIP3P water. The simulation box was neutralised with NaCl and topped up to 150 mM NaCl concentration. The short hydrogen bond was maintained with a harmonic restraint joining the two oxygens in E127 and Y158 with restraint strength of 10,000 kj/mol/nm. The equilibrium length was taken from the crystal structure. The topology for alchemical transformation was generated with pmx (255), where the histidine in KDEL-bound structure was transformed from HIP to HID, while the same histidine in apo state was transformed from HID to HIP. 21 equally spaced lambda windows were used during the transformation, where the bonded, charge and vdw parameters were changed linearly at the same time. The simulations were run with Gromacs 2019.1 (212), with parameters taken from (256). After the initial energy minimisation, a 200 ps simulation was conducted in the NVT ensemble, which was followed by a 1 ns run in the NPT ensemble. The production run was performed with hamiltonian replica exchange at an interval of 1 ps for

30 ns. The resulting energy files were analysed with alchemical analysis (215), with the first 5 ns discarded as equilibrium time.

4.2.6 Grand Canonical Monte Carlo (GCMC) Simulations

The GCMC simulation were run with ProtoMS 3.4.0 (257). 32 equally spaced Adam windows from -17.0 to -1.5 were used. Five repeats of 200 millions steps of production run were preceded with 15 million steps of insertion and deletion only equilibration and 5 million steps of full equilibration.

4.3 Results

4.3.1 Positioning the Hydrogen in the Protein

In order to investigate the hydrogen bond, it is necessary that the precise locations of the hydrogens are known. It is well known that X-ray diffraction experiments often cannot resolve the precise location of hydrogen atoms and indeed for this reason they are often omitted from the refined models. To characterise the nature of the hydrogen bond between E127 and Y158 and the influence of the nearby H12, It is first needed to obtain precise positions of the hydrogen that would, under most conditions, be strongly associated with the hydroxyl oxygen of the tyrosine (Y158). Since, in this case the working hypothesis is that there is a particularly short hydrogen bond between the E127 and Y158, It is not known, a *priori*, which atom the hydrogen could be bonded to and there are three candidate oxygen atoms to consider; the hydroxyl oxygen of Y158 (labelled OH), and the two carboxyl oxygens of E127 (labelled OE1 and OE2, where OE1 is the oxygen closest to atom OH. Figure 4.1). The geometry optimisations (see section 4.2.2) were performed where a hydrogen atom was positioned next to one of these oxygens. This approach at least allows the “correct from the quantum mechanical point of view” position of the hydrogens to be located. In the KDEL-bound conformation, when

the hydrogen is placed next to E127:OE1, which is the oxygen closest to Y158, the position of the hydrogen converged toward the oxygen of Y158 during the optimisation and indeed overlapped with the geometry optimised position of the hydrogen when it is initially placed next to the oxygen of Y158 (Y158:OH). Contrary to this, if the hydrogen is positioned closed to E127:OE2, the hydrogen remains bound to that oxygen. In the apo receptor, all hydrogens remain close to their original position after geometry optimisation. Thus, there are clear differences in energetic landscape of the hydrogen in these different conformations.

To obtain the most probable location of the hydrogen, the free energy of the geometry optimised structures (Table 4.1) were computed. Unsurprisingly, protonated Y158 in complex with the deprotonated E127 (E127···H—Y158) has the lowest free energy in both KDEL-bound and apo state compared with the two protonated form of E127 (E127—HE1···Y158 and E127—HE2···Y158). To understand the reason for this preference, we computed the electrostatic potential (contoured at 0.001 a.u. electron density (258)). As is shown in Figure 4.2, only E127···H—Y158 shows very good charge complementarity, whereas both E127—HE1···Y158 and E127—HE2···Y158 have significant overlap of negative charge, which render them energetically less favourable.

Table 4.1. Free energy and geometry of the hydrogen bonds

		Free energy (Hatree)	R(D-H) (Å)	R(H···A) (Å)	A(D-H···A)
	KDEL-bound				
(HID12)	E127···H—Y158	-2245.33	1.04	1.50	174
(HID12)	E127—HE2···Y158	-2245.26	1.00	2.15	145
(HIE12)	E127···H—Y158	-2245.33	1.03	1.51	174
(HIE12)	E127—HE2···Y158	-2245.25	1.00	2.15	146
(HIP12)	E127···H—Y158	-2245.79	1.06	1.48	173
(HIP12)	E127—HE2···Y158	-2245.74	0.99	2.17	144
	Apo				
(HID12)	E127···H—Y158	-2245.33	1.00	1.86	164
(HID12)	E127—HE1···Y158	-2245.31	1.04	1.80	174
(HID12)	E127—HE2···Y158	-2245.27	0.99	2.21	145
(HIE12)	E127···H—Y158	-2245.33	1.00	1.87	164
(HIE12)	E127—HE1···Y158	-2245.31	1.05	1.79	175
(HIE12)	E127—HE2···Y158	-2245.26	0.99	2.21	145
(HIP12)	E127···H—Y158	-2245.81	1.01	1.87	162
(HIP12)	E127—HE1···Y158	-2245.79	1.03	1.81	171
(HIP12)	E127—HE2···Y158	-2245.76	0.98	2.22	144

Free energy is computed at the level of D3LYP-D4/def2-TZVP. R(D-H) is the intramolecular distance between the donor heavy atom and the hydrogen. Two significant figures were preserved as 0.01 Hatree is 6.28 kcal/mol. Hatree is used instead of kcal/mol as the reference state is difficult to select. R(H· · ·A) is the intermolecular distance between the hydrogen and the acceptor heavy atom. A(D-H· · ·A) is the angle of the hydrogen bond along the donor heavy atom, the hydrogen and the acceptor heavy atom. HID12 indicates the system is optimised under a histidine which was protonated on the delta nitrogen, whereas HIE12 indicates that the histidine was protonated on the epsilon nitrogen. HIP12 means that both nitrogens were protonated. E127···H—Y158 stands for the hydrogen bond system between protonated Y158 and deprotonated E127, while E127—HE1···Y158 and E127—HE2···Y158 signifies the deprotonated Y158 hydrogen bonding to the E127 protonated on the E127:OE1 and E127:OE2 oxygen atoms respectively.

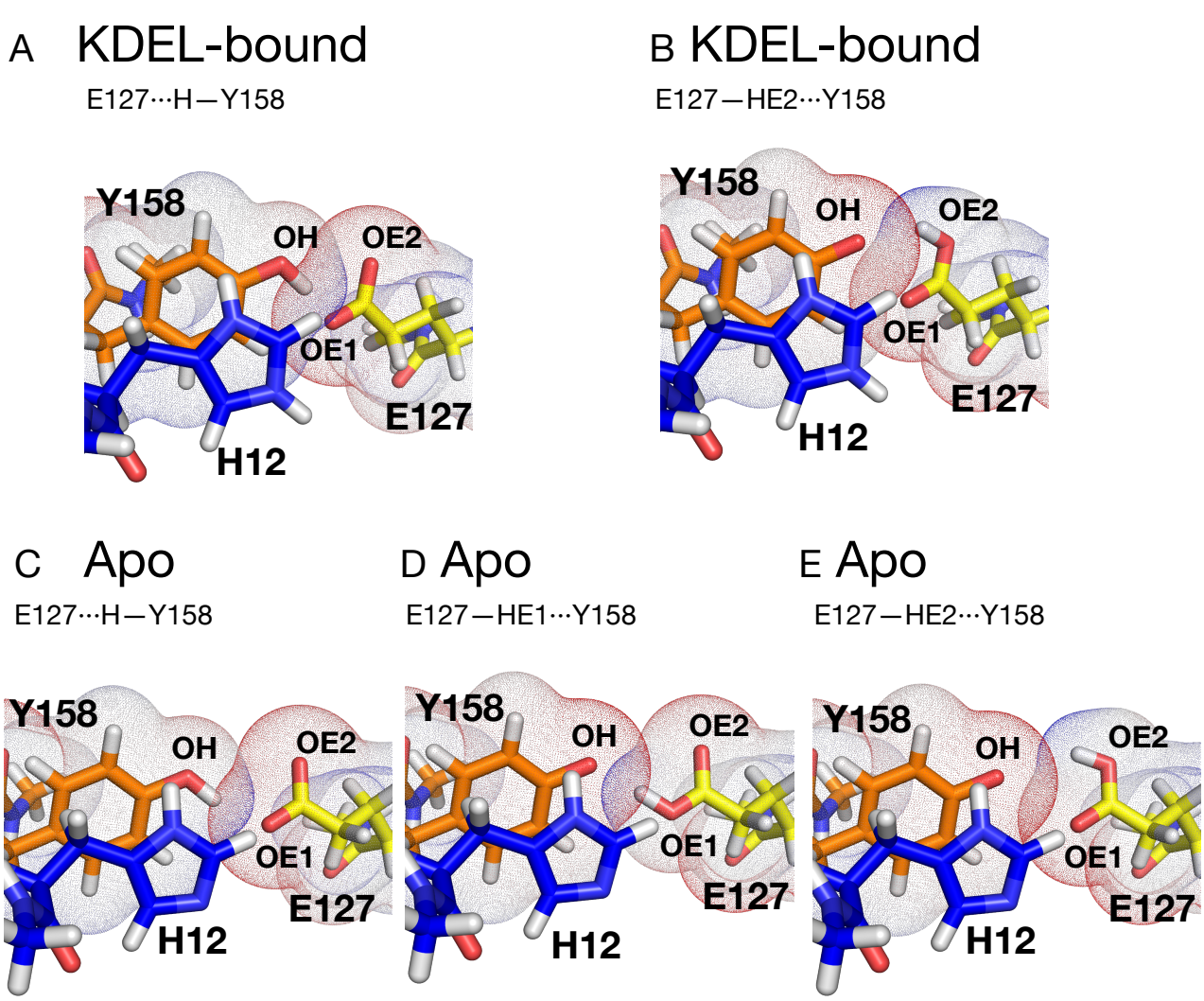


Figure. 4.2. *The electrostatic potentials on the 0.001 a.u. contour of the molecular electron density. The region of negative electrostatic potential is given in red, and the region of positive electrostatic potential is given in blue. In the KDEL-bound conformation, E127···H—Y158 has complementary charge overlap during the region of the hydrogen bond, whereas the region of overlap in both contributing areas are negatively charged in the E127—HE1···Y158 system. Favourable overlap is also observed in E127···H—Y158 in apo conformation.*

Thus, both electrostatic potential calculations and free energy calculations showed that the most energetically favourable position of the proton is close to the Y158.

4.3.2 The Strength of the E127···H—Y158 H-bond Depends on the Protonation State of H12

Given the potential role of H12 in mediating the pH dependent KDEL binding, it is postulated that the protonation of histidine could strengthen the E127···H—Y158 hydrogen bond, which would allow the KDEL receptor to lock the KDEL tag in the binding pocket in acidic pH. This hypothesis was tested by computing the free energy difference of the transition from the hydrogen bonding system in the KDEL-bound state to the same system in the apo state under different protonation state of the H12. As is shown in Table 4.2, the free energy difference of this transition is higher when H12 is protonated (HIP: -15.83 - (-31.37)= 15.54 kcal/mol) compared with the free energy difference when H12 is deprotonated (HID: -20.55-(-30.33)= 9.78 kcal/mol; HIE: -17.4-(-27.30)= 9.9 kcal/mol). Thus, the protonation of H12 would make the transition more energetically unfavourable by 5.7 kcal/mol.

Table 4.2. Properties of the hydrogen bond

	KDEL-bound			Apo		
	HID	**HIE**	**HIP**	**HID**	**HIE**	**HIP**
	Interaction Free energy between E127 and Y159 (kcal/mol)					
ΔG	-30.3	-27.3	-31.4	-20.6	-17.4	-15.8
	The Core-Valence Bifurcation Index (CVBi)					
CVBi	-0.23	-0.20	-0.26	-0.03	-0.02	-0.04
	Quantum Theory of Atoms in Molecules (QTAIM)					
H(R)	-0.031	-0.028	-0.034	-0.002	-0.001	-0.002
V(R)/G(R)	-1.53	-1.48	-1.58	-1.09	-1.06	-1.09
H(R)/P(R)	-0.39	-0.36	-0.41	-0.06	-0.05	-0.07

The interaction Free energy, CVBi and AIM metrics, where a more negative value represents a stronger interaction. Both AIM and CVBi displayed the same trend, where the interactions in the KDEL-bound is stronger than that in the apo state. In the KDEL-bound state, the protonated histidine (HIP) gives a stronger interaction compared with deprotonated histidine (HID/HIE). Such protonation state dependency is not observed in the apo state.

After showing that the protonation of the histidine would incur a free energy penalty on the transition, the influence of the protonation state on the hydrogen bond between Y158 and E127 was investigated. A crude metric of estimating the hydrogen bond strength is the length of the covalent bond between the proton and the hydrogen bond donor R(D-H) (259). A stronger hydrogen bond usually results from a stronger sharing of the electrons between the proton and the hydrogen bond acceptor. A consequence of this sharing of electrons is the stretching of the R(D-H). As is shown in Table 4.1, all of the R(D-H) in the KDEL-bound state are longer than the apo state, suggesting a stronger hydrogen bond in the KDEL-bound state compared with the apo state. Furthermore, the longest R(D-H) is observed in the KDEL-bound state when the histidine is protonated, suggesting that the protonation of histidine would strengthen the hydrogen bond. Similar protonation state dependent elongation of R(D-H) is not observed in the apo state.

A more accurate estimate of the hydrogen bond strength could be obtained by analysing the wave function from the QM calculations. According to Bader's quantum theory of atoms in molecules (QTAIM), topology analysis of the electron density (ρ(r)) and its Laplacian can be used to give a quantitative description of the bond strength. For a bond to be present, an atomic bond path has to link the two atoms which form the bond and a bond critical point (BCP) has to be present on this path (258, 260). In all hydrogen bond systems, an atomic bond path was found between the proton and the hydrogen bond acceptor and a BCP was present on this path. Table 2 listed some important topological parameters for this BCP, including the total energy density H(r) (261) which is the sum of Lagrangian kinetic energy G(r) and Potential energy density V(r) (262), the V(r)/G(r) and H(r)/ρ(r). The strength of the hydrogen bond can be inferred from H(r), where Rozas et al (263) proposed that weak bonds will have a H(r) close to zero, while strong bonds will display a negative value. The hydrogen bonds formed in the KDEL-bound states between the Y158H and E127 could be classified as a strong hydrogen bond, while in the apo state, it was classified as a weak hydrogen bond. The strongest hydrogen bond was observed when the histidine is protonated (HIP). Espinosa et al. (262) purposed that both H(r)/ρ(r) and V(r)/G(r) can serve as indicators of the bond strength. H(r)/ρ(r) is also commonly referred to as bond degree (BD) and a smaller value indicates a stronger bond. Similar to BD, the smaller the V(r)/G(r), the stronger the bond. The ranks of hydrogen bond strength obtained from H(r)/ρ(r) and V(r)/G(r) were consistent with result obtained from H(r). For estimating hydrogen bond strength, specific metrics such as the Core-Valence Bifurcation Index (CVBi) (246) are proposed, where a more negative value corresponds to a stronger hydrogen bond. All of these methods gave the same result where the hydrogen bond is stronger in the KDEL-bound state compared with the apo state. The hydrogen bond could be further enhanced by the protonation of the histidine in the KDEL-bound state but is not affected in the apo state.

Though it has been shown that in the KDEL-bound state, the protonation of the histidine strengthens the short hydrogen bond, it is unclear as to the reason behind this protonation state dependent enhancement. The short hydrogen bond, as is seen in the KDEL-bound state, is indicative of a low-barrier hydrogen bond (LBHB). A property of the LBHB is that the energy barrier of proton transferring from the hydrogen bond donor to the acceptor is smaller than the zero-point energy (ZPE) of the proton (264). To test if this hydrogen bond is indeed a LBHB, the adiabatic energy surface of transferring the proton from the hydrogen bond donor (Y158) to the hydrogen bond acceptor (E127) is computed. As is shown in Figure 4.3A, a large (~20 kcal/mol) energy barrier is encountered during the proton transfer in the apo state, which is much higher than the ZPE (~4.3 kcal/mol). The protonation state of the H12 had very little effect on the energy profile. On the other hand, the energy landscape for the KDEL-bound state is very different in that protonation of the histidine had a significant effect (Figure 4.3B). When the histidine is deprotonated, the energy difference between the global energy minimum (proton bound to the hydrogen bond donor) and the local energy minimum (proton bound to the hydrogen bond acceptor) is quite large (~10 kcal/mol for HIE and ~7 kcal/mol for HID) and is still higher than the ZPE (~3.7 kcal/mol). However, when the histidine is protonated the energy difference shrinks to only 3 kcal/mol and is smaller than the ZPE (~3.1 kcal/mol). Thus, in the KDEL-bound state, the hydrogen bond system can be considered as a LBHB when the histidine is protonated and is consistent with the hypothesis that protonated histidine strengthens the hydrogen bond in the KDEL-bound state.

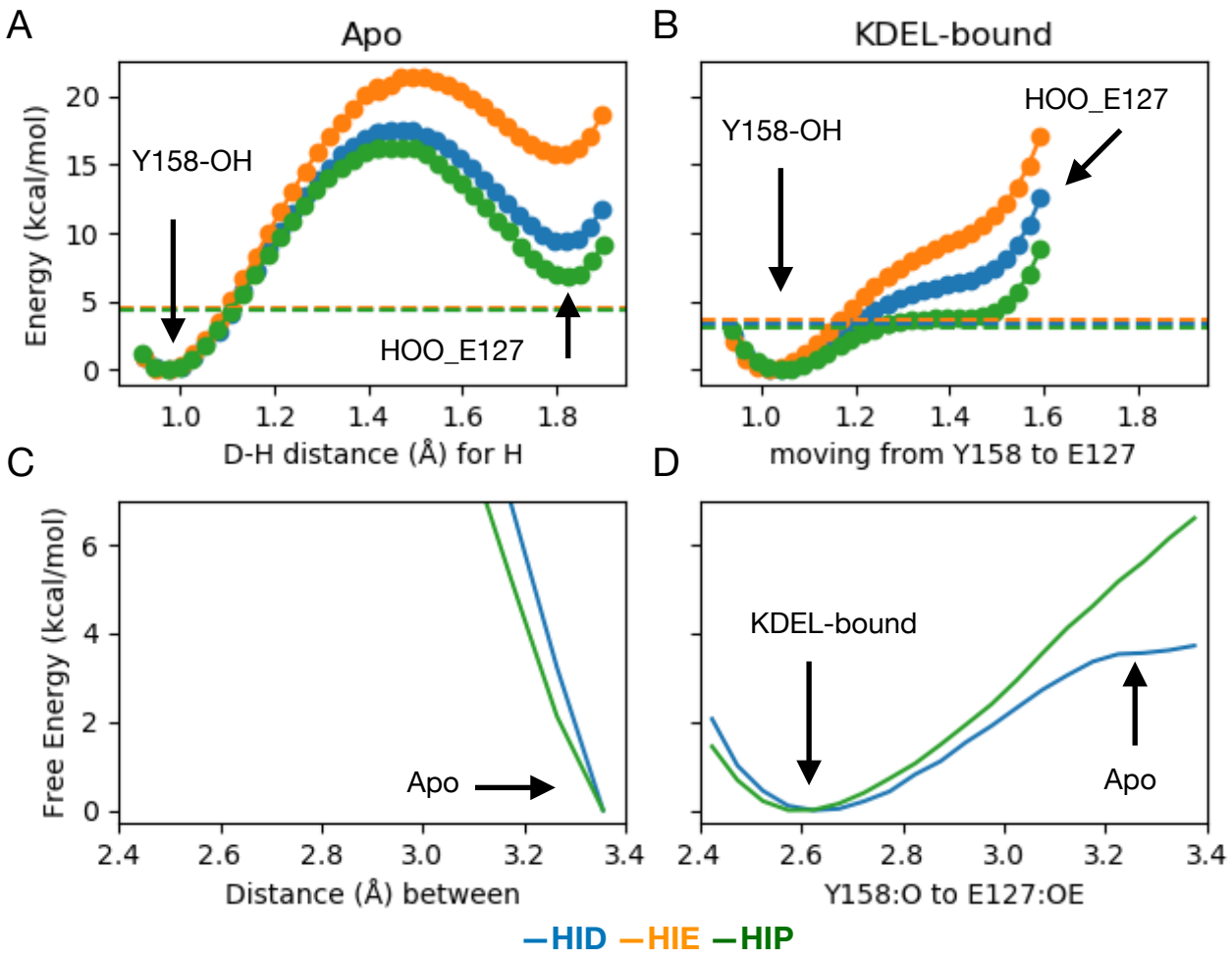


Figure. 4.3. *Energy of proton transfer from the Y158 to E127. A. In the apo state, a very large barrier (~20 kcal/mol) is observed during the proton transfer, which is not observed in the KDEL-bound state. B. In the KDEL-bound state, the energy barrier for the proton movement is smaller than the ZPE (dashed line) when the histidine is protonated (green). The energy difference is greatly increased when histidine is deprotonated (orange and blue). C. The energy landscape of separating the hydrogen bond. In the apo state, short hydrogen bond is prohibited by the larger energy barrier. D. In the KDEL-bound state, 6.5 kcal/mol of energy is required to break the hydrogen bond when histidine is protonated (green line), while only 3.5 kcal/mol is required when histidine is deprotonated (blue line). The error is computed using the block average method and is thinner than the line width.*

Though the QM calculations show that in vacuum, the short hydrogen bond is a LBHB when histidine is protonated, it is unclear as to the strength of this hydrogen bond in the protein in physiological conditions. QM/MM calculations were thus performed to compute the free energy penalty of stretching the hydrogen bond. As is shown in Figure 4.3C, in the apo state, the formation of short hydrogen bond is not energetically favourable and this result is not affected by the protonation state of the histidine. In the KDEL-bound state, however, the short hydrogen bond is energetically favourable to form as is shown by the energy minimum around 2.6 Å (Figure 4.3D). The energy that is required to break this hydrogen bond is significantly enhanced (from ~3.5 kcal/mol to 6.6 kcal/mol) when the histidine is protonated, which is consistent with the hypothesis that protonation of histidine trapped the KDEL tag in the binding pocket via the formation of the strong hydrogen bond.

4.3.3 The Short Hydrogen Bond Influences the Affinity of H12 for Protons

Since the short hydrogen bond in the KDEL-bond state is stabilised by the protonation of the histidine, the reciprocal effect where the short hydrogen bond should stabilise the protonated histidine in the KDEL-bound state but not in the apo state, should also be true. To test this hypothesis, a double bilayer system was constructed to examine the proton affinity difference between the KDEL-bound state and the apo state (Figure 4.4A). Alchemical transformations of deprotonating the histidine in the KDEL-bound state and protonating the same histidine in the apo state were performed simultaneously to estimate the proton affinity difference. To test the effect of the short hydrogen bond, two conditions were tested where in one condition, the short hydrogen bond was imposed as a harmonic restraint between the corresponding atoms and in the other condition, no artificial restraint was imposed. The calculations show that when the short hydrogen bond

was imposed, the protonation of the histidine (H12) was ~3.5 kcal/mol more favourable in the KDEL-bound state (Table 4.3) and thus, consistent with the hypothesis. However, when the short hydrogen bond was not being restrained, the free energy difference dropped to ~2.0 kcal/mol highlighting the significance of the short hydrogen bond in stabilising the protonated histidine. Calculations done on the histidines in other parts of the protein (H90 and H111) were served as control and showed that only H12 displayed different proton affinity in different conformations.

Table 4.3. ΔΔG of (in kcal/mol) of deprotonation* of H12 in the KDEL-bound and protonation** of H12 in apo state under different conditions.

Condition	**Run 1**	**Run 2**	**Run 3**
Short H-bond imposed	3.6	3.8	3.2
No H-bond restraint	1.9	2.0	2.2
H90 (control – no restraint)	0.8		
H11 (control – no restraint)	0.4		

*to or from** a neutral histidine with the proton on the δ carbon (HID).

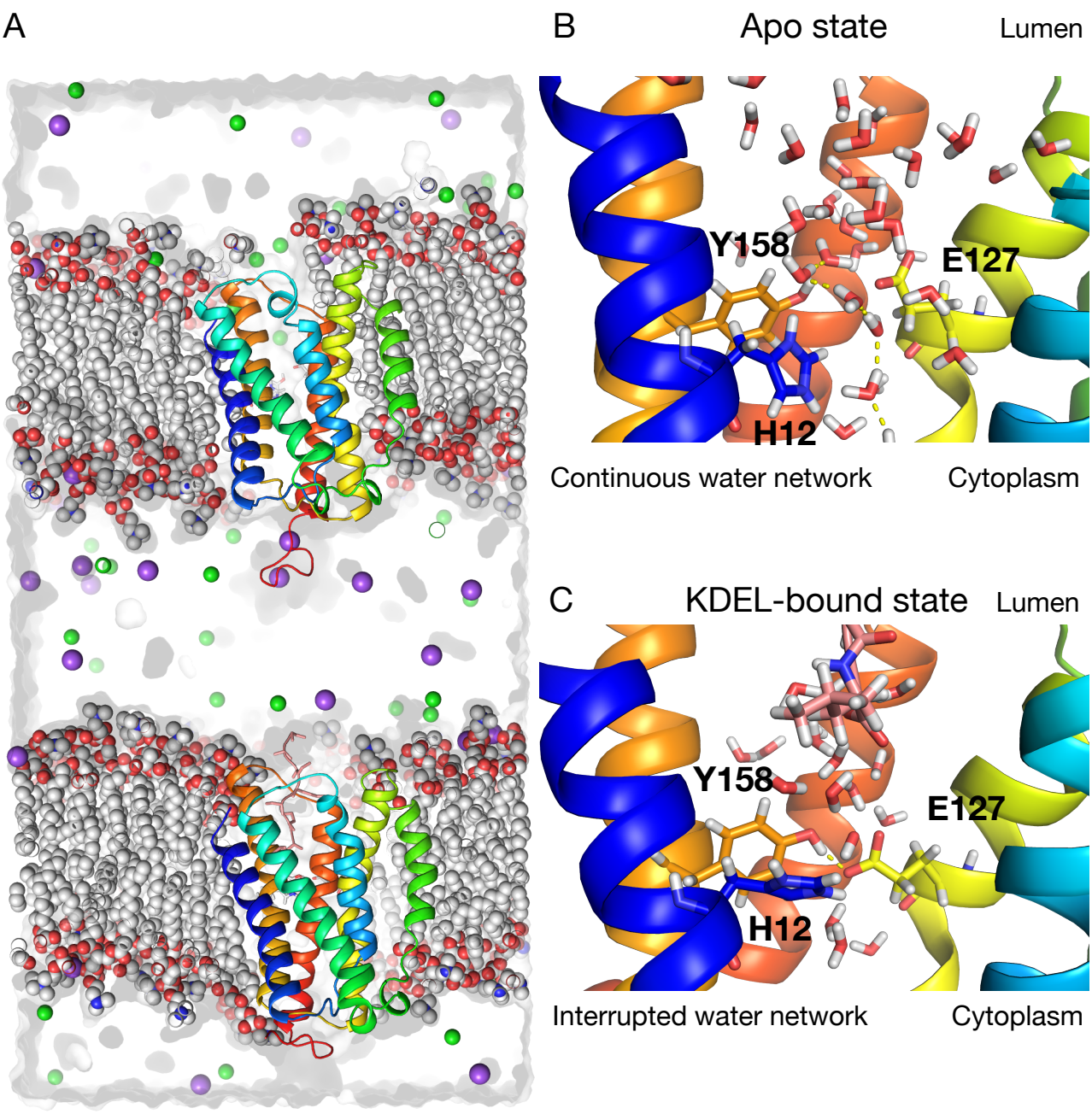


Figure 4.4. *Influence of the short H-bond on H12 affinity. (A) shows the simulation box used, which is a double bilayer system to simultaneously add/remove a proton from H12 in each KDEL receptor. Na^+ and Cl^- ions are represented as purple and green spheres respectively and the lipid bilayer is drawn as vdW spheres. (B) In the apo state, a continuous water network linking the lumenal and cytoplasmic pockets is observed as highlighted by yellow dashed lines. (C) In the KDEL-bound state, the water network does not reform as expected. H12 (blue), E127 (yellow) and Y158 (orange) are drawn as sticks in all figure parts.*

Though it is reassuring that the absence of the short hydrogen bond destabilised the protonated histidine, it is interesting to investigate the factors causing this destabilised effect. The KDEL receptor resembles a transporter with both lumenal and cytoplasmic pocket topologically (see chapter 1.5.6) (133). The short hydrogen bond acted as a gate between these two pockets. In the apo state, the two water pockets are joined by a water path allowing water exchange between the two pockets (Figure 4.4B). In the KDEL-bound state, on the other hand, the formation of the short hydrogen bond prohibited the water exchange between the two water pockets (Figure 4.4C). In the absence of the short hydrogen bond, transient water wire could form and marked this state as an intermediate state between the apo and the KDEL-bound state, and thus, resulting in a smaller proton affinity difference. Overall, the protonation free energy calculations give results consistent with the hypothesis in that there is a mutual strengthening effect between the short hydrogen bond and the protonated histidine.

4.3.4 Protonated Histidine Stabilised More Water in the Intracellular Pocket

Given that the short hydrogen bond which prohibits the water exchange between lumenal pocket and the cytoplasmic pocket, could stabilise the protonated histidine, the effect of the protonation state of the histidine on water is being investigated. As is shown in Figure 4.5A, GCMC calculations show that the protonation of the histidine allows the cytoplasmic pocket to host more waters, where the energy minimum is shifted from 8 waters when histidine is deprotonated to 10 waters when histidine is protonated. In the apo state, the change in the number of water as a result of the change of the protonation state could be mitigated by the water exchange with the lumenal pocket. In the KDEL-bound state, however, this water exchange is prohibited by the short hydrogen bond, which resulted in the different protonation free energy between two states. When the

short hydrogen bond is not being imposed, occasional water exchange between the cytoplasmic pocket and the lumenal pocket could dampen this difference.

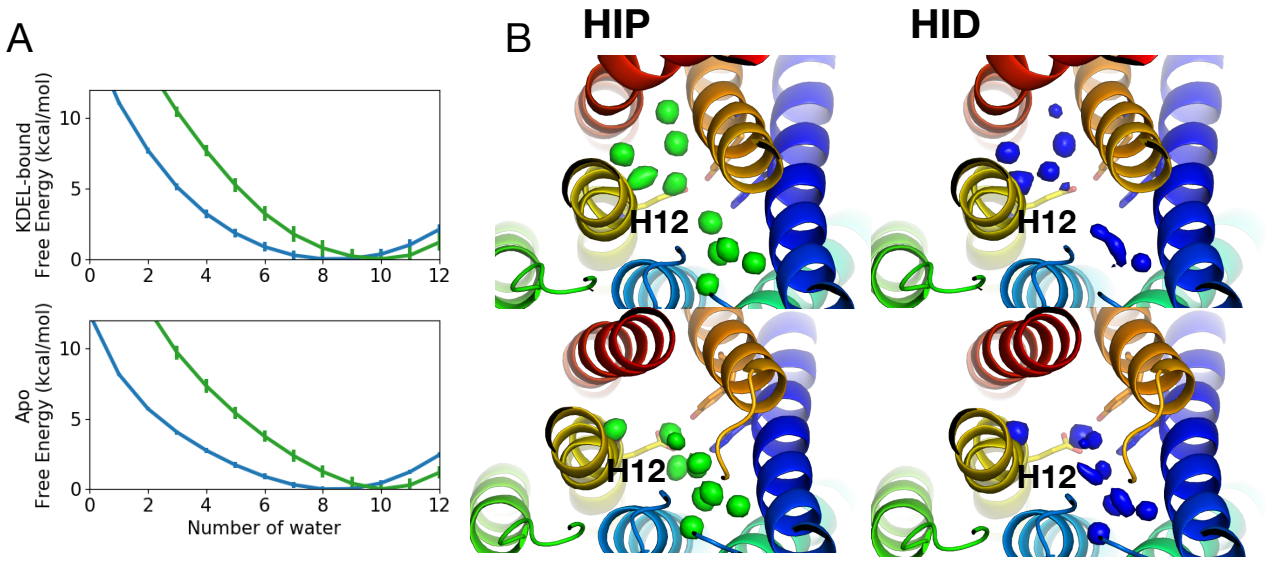


Figure 4.5. *The water network in the cytoplasmic water pocket. A. In both KDEL-bound and apo state, protonated histidine would stabilise two more water in the cytoplasmic pocket compared with the apo state. B. Though the position of the water differ between the KDEL-bound state and the apo state, the two additional waters are both located close to the histidine (blue sticks).*

4.4 Discussion

In this chapter, it has been shown that the short distance between the E127 and Y158 is consistent with the description of a short hydrogen bond, which can also found in the active site of an enzyme (265). A key feature of these short hydrogen bonds is that the distance between the hydrogen bond donor heavy atom and hydrogen bond acceptor distance is usually below 2.7 Å and the energy barrier for the proton to shuttle from hydrogen bond donor to the hydrogen bond acceptor is lower than ZPE (266). Thus, the proton, instead of being covalently bound to the hydrogen bond donor, is considered to be delocalised between the donor and the acceptor atoms and the resulting hydrogen bond is predicted to be stronger than an ordinary hydrogen bond by 10~20 kcal/mol (236). The results in this chapter confirmed that short hydrogen bond observed in the crystal structure is indeed a short hydrogen bond. In the apo state, the potential energy landscape of the proton transfer (Figure 4.3A) is consistent with an ordinary hydrogen bond, where a large energy barrier is expected when moving the proton away from the preferred position. In the KDEL-bound state, however, this energy barrier is much diminished compared with the apo state and the energy profile is consistent with other hydrogen bonds with this short distance (~2.5 Å) as is recently reviewed (267). The result also shows that the energy barrier is further reduced upon the protonation of the histidine and when the histidine is protonated, the energy barrier for the proton transfer is lower than the ZPE suggesting the presence of a LBHB. Interestingly, the participation of tyrosine in the formation of short hydrogen bond is also observed in another study (268) suggesting the participation of tyrosine in this type of interaction is not a rare case.

This key histidine (H12) is identified as the proton sensor (133), which would be protonated in the Golgi, where the KDEL receptor engages the KDEL ligand. The above calculations suggest that the LBHB could exist only when the key histidine (H12) is

protonated. The reciprocal effect of the hydrogen bond on the protonation state of the histidine is also being investigated. In the KDEL-bound state, the presence of the short hydrogen bond favours protonated histidine and in the absence of the short hydrogen bond, the protonated histidine is still preferred but to a much less extent. Further investigation shows that this preference is linked to the dynamics of the surrounding water molecules, where the trajectory shows that the short H-bound would cut off the water exchange between the cytoplasmic pocket and the lumenal pocket. GCMC calculations further confirmed that different protonation states of the histidine favours different number of waters in the cytoplasmic pocket and the cutoff of water exchange would hinder the protonation state dependent water movement.

Based on the results, the mechanism of the KDEL receptor could be summarised as shown in Figure 4.6. In the Golgi, where the pH is low, the pH sensor H12 is protonated, which stabilise 10 waters in the cytoplasmic pocket and the cytoplasmic pocket could exchange water with the lumenal pocket. Upon the binding of the KDEL-tag, the water exchange is hindered by the short hydrogen bond and the short hydrogen bond tightens up the KDEL-tag to the complex. In the ER, where the pH is much higher, the key histidine is deprotonated and the cytoplasmic pocket can only host 8 waters. Furthermore, the short hydrogen bond is weakened by the deprotonated histidine, which resulted in the loosening of the KDEL-tag. The combined force of expelling water from the cytoplasmic pocket and the loss of the short hydrogen bond, enables the exit of the KDEL tag.

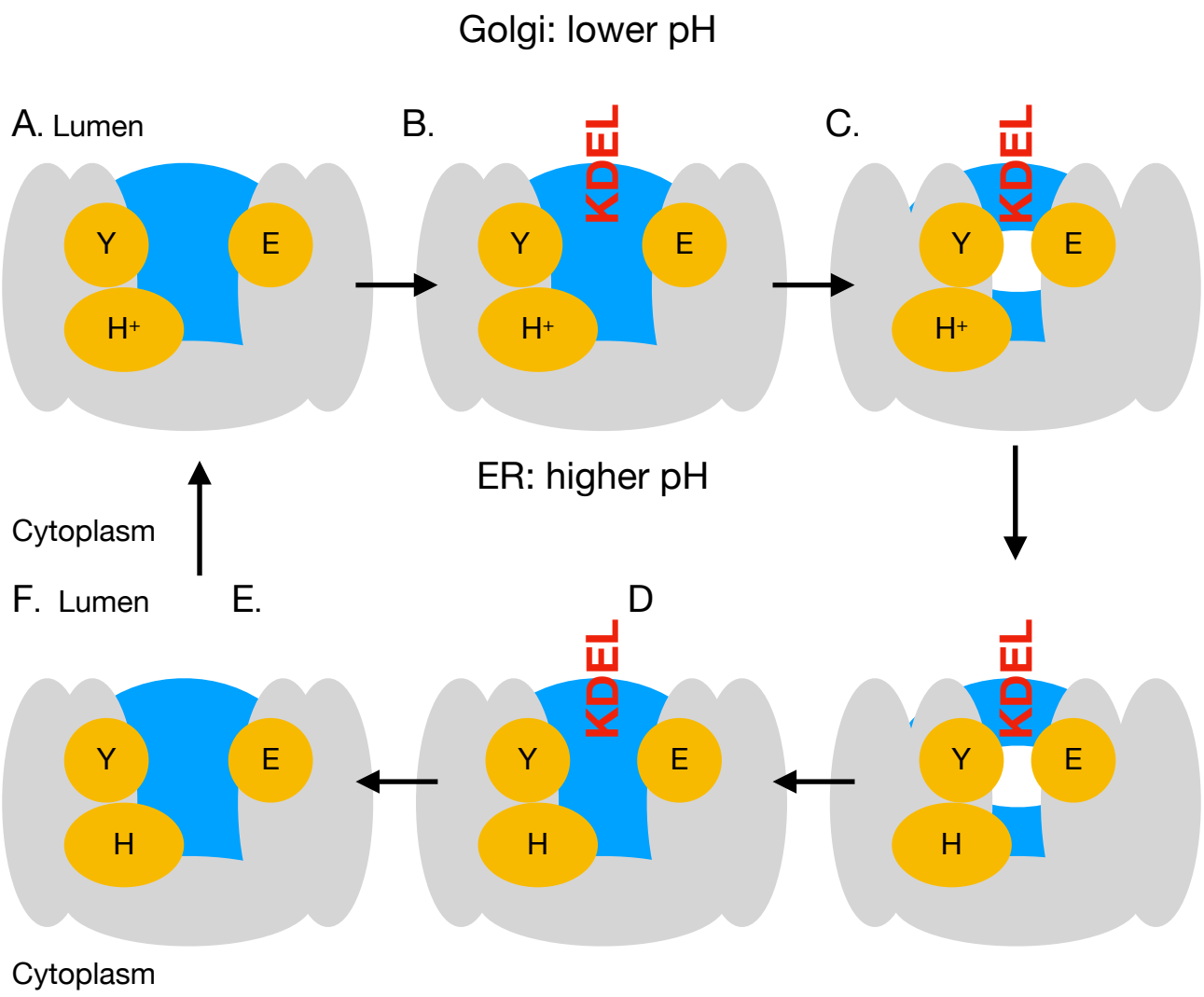


Figure 4.6. *Purposed scheme of KDEL-mediated ER-Golgi trafficking. In the Golgi, the KDEL receptor is initially in its apo state, with the key histidine protonated and water pockets on both sides of the E127/Y158 (A). The KDEL-tagged substrate binds the receptor trapping waters in a pocket that is no longer exposed to lumenal bulk (B). The binding is completed by a change in receptor conformation and the formation of the short hydrogen bond to go into the crystallographically observed KDEL-bound conformation (C). The KDEL receptor/KDEL-tagged substrate complex is then transferred to the ER where the high pH deprotonates the histidine (D). Without the protonated histidine, the short hydrogen bond is no longer maintained and the KDEL-tagged substrate dissociates (E). The apo configuration as is observed in the crystal structure (F).*

4.5 Conclusion

In this chapter, extensive calculations were made that support the existence of the short hydrogen bond in the KDEL receptor. Furthermore, the calculations have shown that the strength of the short hydrogen bond is drastically affected by the protonation state of the histidine. Through investigation of the reciprocal effect of the short hydrogen bond on the the protonation state of the histidine, it is found that the short hydrogen bond disrupted a hydrogen-bonded water network, which might play a key role in the functioning of the receptor.

5. Conformational change of the KDEL receptor drives different oligomerization profile

5.1 Introduction

After binding ER proteins hosting the KDEL tag in the Golgi, the KDEL receptor needs to be trafficked back to the ER. However, the detailed molecular mechanism of how the KDEL receptor manages to localise in the cis-Golgi (269) and only move to the ER upon the binding of a protein with a KDEL tag, remains elusive. Lectin Mannose Binding Protein 1 (LMAN1) (270) is a protein that ensures the exclusion of ER chaperones from the departing ER vesicles (271). Similar to the KDEL receptor, LMAN1 was also found to shuttle between the ER and the cis-Golgi (270). The C-terminus of LMAN1 ends with KKFF, where the di-phenylalanine tail is found to interact with Sec24 for exporting from the ER to the Golgi (272) (273) (Figure 5.1A). The di-lysine tag, on the other hand, similar to the di-lysine tag in the KDEL receptor, mediates the interaction with COPI and allows the retrieval from the Golgi back to the ER (274) (Figure 5.1B). LMAN1 was found to form dimers or hexamers via disulphide bonds. The formation of the dimer was found to be required for the binding of COPII (275). However, the formation of the hexamer is not required for the ER exit as the LMAN1 without the disulfide bond that is crucial for the hexamer formation could be exported to ER normally (276) (Figure 5.1C). The KDEL receptor and LMAN1 are quite similar in the sense that both of them recognise their corresponding cargo in a pH-dependent manner via a key histidine (133, 277) and both shuttle cargo between the ER and the Golgi. Thus, it is likely that dimerisation and oligomerisation also plays a key role in the normal functioning of the KDEL receptor.

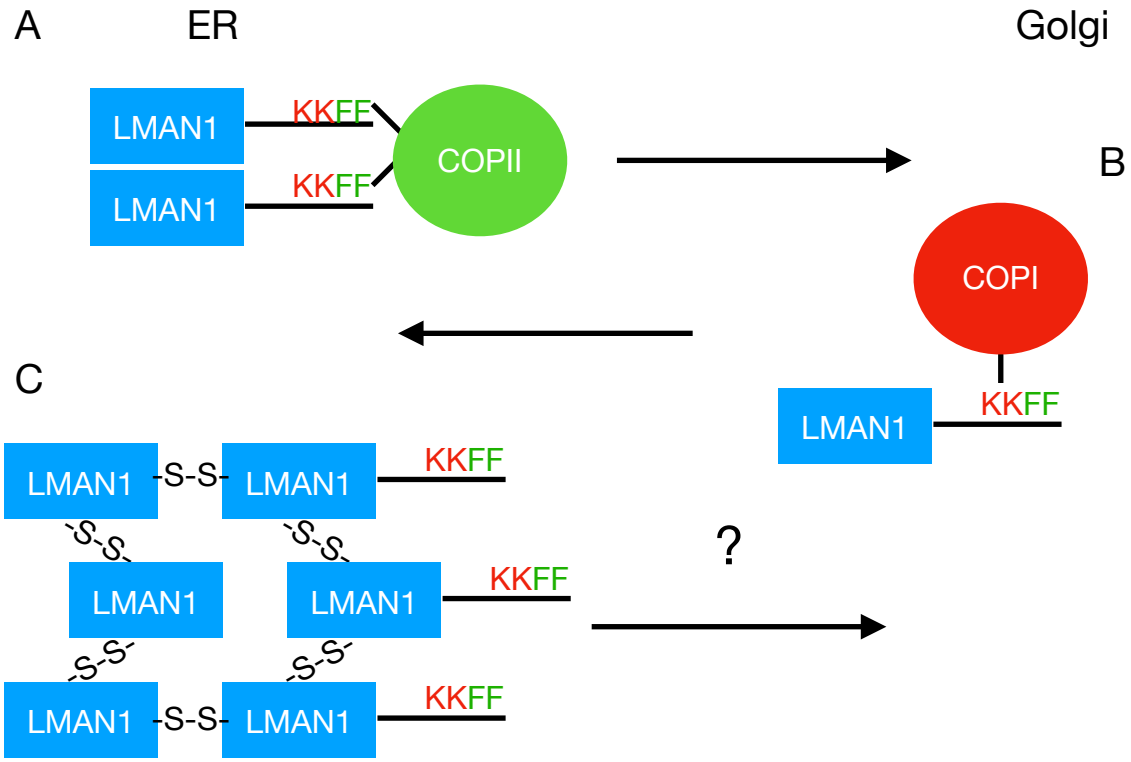


Figure 5.1. *A model of the trafficking of LMAN1 between ER and Golgi. The formation of a LMAN1 dimer is required for the trafficking from ER to the Golgi through the interactions between the di-phenylalanine tag and sec24 of COPII vesicle. The trafficking from Golgi back to the ER is mediated by the interaction between the di-lysine tag and COPI vesicle. Though LMAN1 could form hexamer through di-sulfide bond, the formation of hexamer is not required for the ER to Golgi export and its function is unclear.*

In addition to bringing the escaped ER chaperones back to the ER, the KDEL receptor can also function as a G protein-coupled receptor (GPCR) as is discussed in chapter 1.5.5. The activation of a GPCR can happen through a diverse set of mechanisms, including activation by dimerisation (278). Notably, the activation of β2-adrenergic receptor has been suggested to require the formation of a homodimer (279). The KDEL receptor is structurally very similar to a eukaryotic SWEET transporter (280) and it has also been shown that the Mammalian Sweet Taste Receptors T1R2 and T1R3 need to form dimers to function (281). Thus, it is also possible that the GPCR function of KDEL receptor is activated via dimerisation.

Recently, the crystal structure of the KDEL receptor has been solved in the KDEL-bound and apo conformation (133). One striking difference between the two structures is the kinked helix 7 in the KDEL-bound structure, which is straight in the apo state (Figure 5.2A), while the rest of the receptor is conformationally identical. In helix 7, at the turning point located a strictly conserved acidic residue (D193), mutation of this aspartate to asparagine was found to retain the KDEL binding capability but abolished the trafficking from Golgi to the ER (282). Furthermore, the di-lysine ER retrieval tag is located after helix 7, which is resolved in the KDEL-bound crystal structure but not in the apo state (Figure 5.2A) (133). The ordered interactions between the di-lysine tag and the β'-COP is only achievable when the KDEL receptor is in the KDEL-bound state. Thus, It is hypothesised that the kinked helix 7 enables the KDEL-bound and apo structure to have different Golgi-ER trafficking profile, which might be a result from different dimerisation or oligomerisation properties.

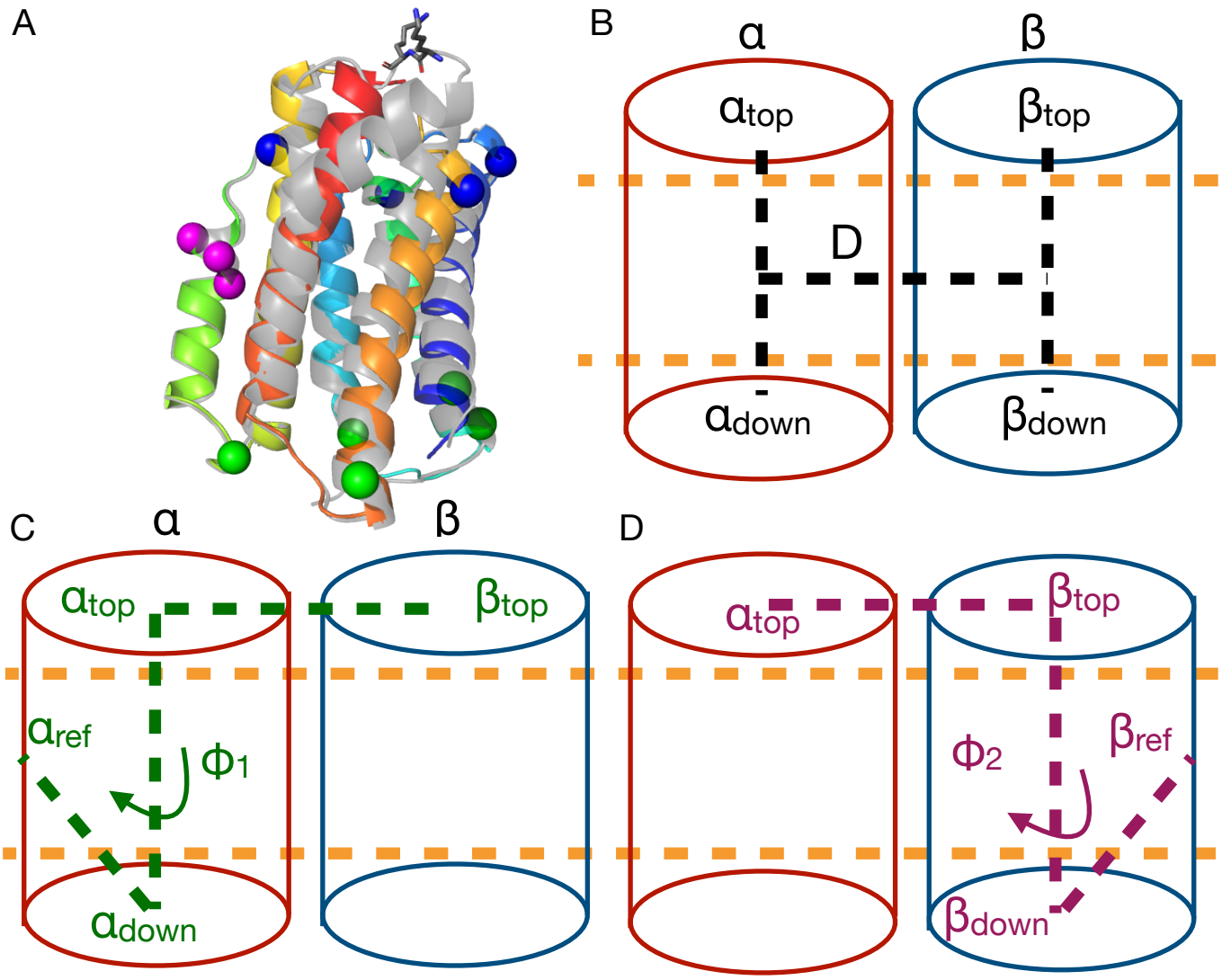


***Figure 5.2**. Two dihedrals characterising the dimer interface. A. The Apo (rainbow) and KDEL-bound (grey) state are differed only by the kinked helix 7 (Apo: red). The di-lysine tag (grey sticks) is shown at the end of helix 7. The protein is orientated with three points (top, down and ref). The upper centre (top) is defined as the centre of five backbone particles from residue number 25, 80, 136 and 150 (blue). The lower centre (down) is defined as the centre of five backbone particles from residue number 6, 50, 59, 111 and 171 (C: green). The reference point (ref) is defined as the centre of three backbone particles from residue id 97 to 99 (magenta). B. α (red) and β (blue) are two proteins represented as cylinders in the lipid membrane (brown). The distance separating the protein is defined as the distance between the midpoints of the upper and lower centre of the two proteins. The dimer interface is defined by the two dihedrals ϕ1 and ϕ2. Dihedral ϕ1 (C) is defined using four coordinates α_{ref}-α_{down}-α_{top}-β_{top} and dihedral ϕ2 (D) using β_{ref}-β_{down}-β_{top}-α_{top}.*

However, studying the dimerisation and oligomerisation properties of a protein is not a simple feat. Early experiments depended on the over-expression of the protein and there was concern over artefacts (283). Fluorescence Resonance Energy Transfer (FRET) and Bioluminescence Resonance Energy Transfer (BRET) are two experimental measures commonly used for studying oligomerisation. However, FRET and BRET reflect the proximity in distance rather than physiological meaningful association and can lead to false-positive results due to nonspecific proximity (284, 285). Though time-resolved FRET (TR-FRET) (286) and the proximity ligation in situ assay (P-LISA) (287) are said to resolve these issues, they still lack the detailed characterisation of the dimer interface and require modification of the protein.

Computational methods such as molecular dynamics simulations, on the other hand, can provide excellent spatial and temporal resolution that allow the protein to be studied in its natural state. Coarse-grained simulations (288), which have been used to study the dimerisation of a single helix such as the EphA1 Receptor (289), ErbB receptor (290), Glycophorin (291) and poly-alanine peptides (292) or larger proteins such as GPCR (293) or HIV capsid (294), is the tool which is used here to study the dimerisation and oligomerisation property of the KDEL receptor.

5.2 Methods

5.2.1 Characterisation of the Dimer Interface

To understand the oligomerisation property of the KDEL receptor in the apo and the KDEL-bound conformations, the interface between the two proteins is defined by two dihedral angles (ϕ1, ϕ2) (42) as is shown in Figure 5.2CD. Since for homodimer, the dimer interface (ϕ1, ϕ2) and (ϕ2, ϕ1) are essentially the same, the dimer interface will be labelled as (ϕ1, ϕ2), where ϕ1 < ϕ2 to avoid ambiguity. Another collective variable used

to describe dimer is the distance between the two monomers, which is defined as a single distance between the centre of both proteins (Figure 5.2B).

5.2.2 System Setup and Simulations Details

The initial coordinates of the KDEL-bound and the apo KDEL receptor were obtained from the PDB (PDB: 6I6B (2.59 Å) and 6I6H (2.00 Å) (133)). The C-terminus of the apo structure was modelled with modeller 9.21 (295) to the same length as the KDEL-bound crystal structure. 100 models were created and the one with the best DOPE score were selected (296). The coarse-grain representation of the protein was generated with the martinize script in a roughly 4:1 mapping to force field MARTINI 3.0 (version 3.0.b.3.2). The side chain dihedral corrections were added using vmd (297) and addDihedral.tcl (298) with reference to the crystal structure. As for the C-terminal residues in the Apo structure which were added via modeller, the elastic network and the side chain dihedral corrections were removed. Since the protein tertiary conformation was fixed with the elastic network, the ligand KDEL in the KDEL-bound structure were removed, as it was found that they were not stable in the simulations. The protein embedding was done using the self-assembly protocol (chapter 3.3.2) with 500 DLPC lipids for 100 ns. A final box where the box edges were 1.5 nm away from the protein were sculptured from the patch which includes 186 DLPC lipids and had a NaCl concentration of 150 mM and was equilibrated for 25 ns.

The simulations were run with GROMACS 2018 (212) patched with plumed 2.5 (213). The simulation input parameters were set according to the new-rf.mdp (299), where a timestep of 20 fs was used. Verlet algorithm (300) was used for neighbour list updating and a cutoff of 1.1 nm was used for Lennard-Jones and Coulomb interactions. Reaction-field potential was used to describe the Coulombic interactions with a ε=15 and ε_{rf}=0. The

temperature of protein-lipid complex and the water-ions complex were coupled separately with velocity rescale (301) with a time constant of 1 ps at 320 K. The pressure of the system was maintained at 1 bar semiisotropicly with a compressibility of 3e-4 bar^{-1} and a time constant of 12 ps. The berendsen barostat (302) was used during the equilibration phase and the parrinello-rahman barostat (303) was used in the production run.

5.2.3 Unbiased Simulations

The equilibrated single protein box was replicated 6 times in both x and y axes resulting in a lipid membrane patch with 36 proteins. To break the symmetry in protein orientation, a 100 ns equilibration run was performed with position restraints on the centre of the protein (backbone particle of residue id 159). The restraint strength was set to 1000 kJ/(mol.nm) on both x and y axes. Three independent equilibration runs were then performed with different initial velocities. The position restraint was alchemically switched off at the end of 100 ns run. This was then followed by 10 μs production runs and the trajectory was then analysed with MDAnalysis (304).

To study dimerisation, the single protein box was replicated 2 times in both the x- and y-axis resulting in a lipid membrane patch with 4 proteins. Two proteins were removed and the resulting system was equilibrated for 100 ns with a distance restraint of 1000 kJ/(mol.nm) at 5 nm separating the remaining two proteins. A further 49 equilibration runs were initiated where the two dihedrals (ϕ_1, ϕ_2) determining the orientation of the two proteins with regard to each other were restrained at seven different values from -3 rad to 3 rad with a step of 1 rad at 1000 kJ/(mol.rad). 49 production runs, where all restraints were removed, were started from the 49 equilibrated system and ran for 10 μs.

5.2.4 Two-dimensional Potential of Mean Force Calculations

To obtained a converged free energy landscape of the dimerisation interface in the two-dimensional (2D) dihedral space (ϕ_1, ϕ_2), the enhance sampling technique of umbrella sampling was used. The whole dihedral space was separated into 3969 windows, where both ϕ_1 and ϕ_2 were sampled from -3.1 to 3.1 with a step of 0.1 radian. The initial frames for these windows were generated using steered MD in three stages. In the first stage, the system was equilibrated such that both ϕ_1 and ϕ_2 were restrained to 0.0 with a restraint strength of 1000 kJ/(mol.rad). To prevent dissociation, a loose distance restraint (10 kJ/(mol.nm)) was imposed between the two proteins at 4 nm. In the second stage, ϕ_2 was kept constant, while the restraint of ϕ_1 was changed from 0.0 radian to 7.0 radian (Figure 5.3A) in 7 μs, where the first 0.7 μs was discarded. 63 frames from -3.1 to 3.1 with a step of 0.1 were taken from the simulation, which were feed into the third stage equilibration, where ϕ_1 was kept constant and ϕ_2 was changed from 0.0 radian to 7.0 radian in 7 μs (Figure 5.3B). With the first 0.7 μs was discarded, 3969 frames were extracted from these 63 simulations. For those windows where a LINCS error occurred, the equilibrium run was performed from 0.0 to -7.0. A production run of 4 μs were performed for these 3969 windows, where the two dihedrals were restrained at 250 kj/(mol.rad), with a one-sided distance restraint of 1000 kj/(mol.rad) at 5 nm. The resulting dihedrals were analysis with WHAM (243) to generate the 2D PMF profile.

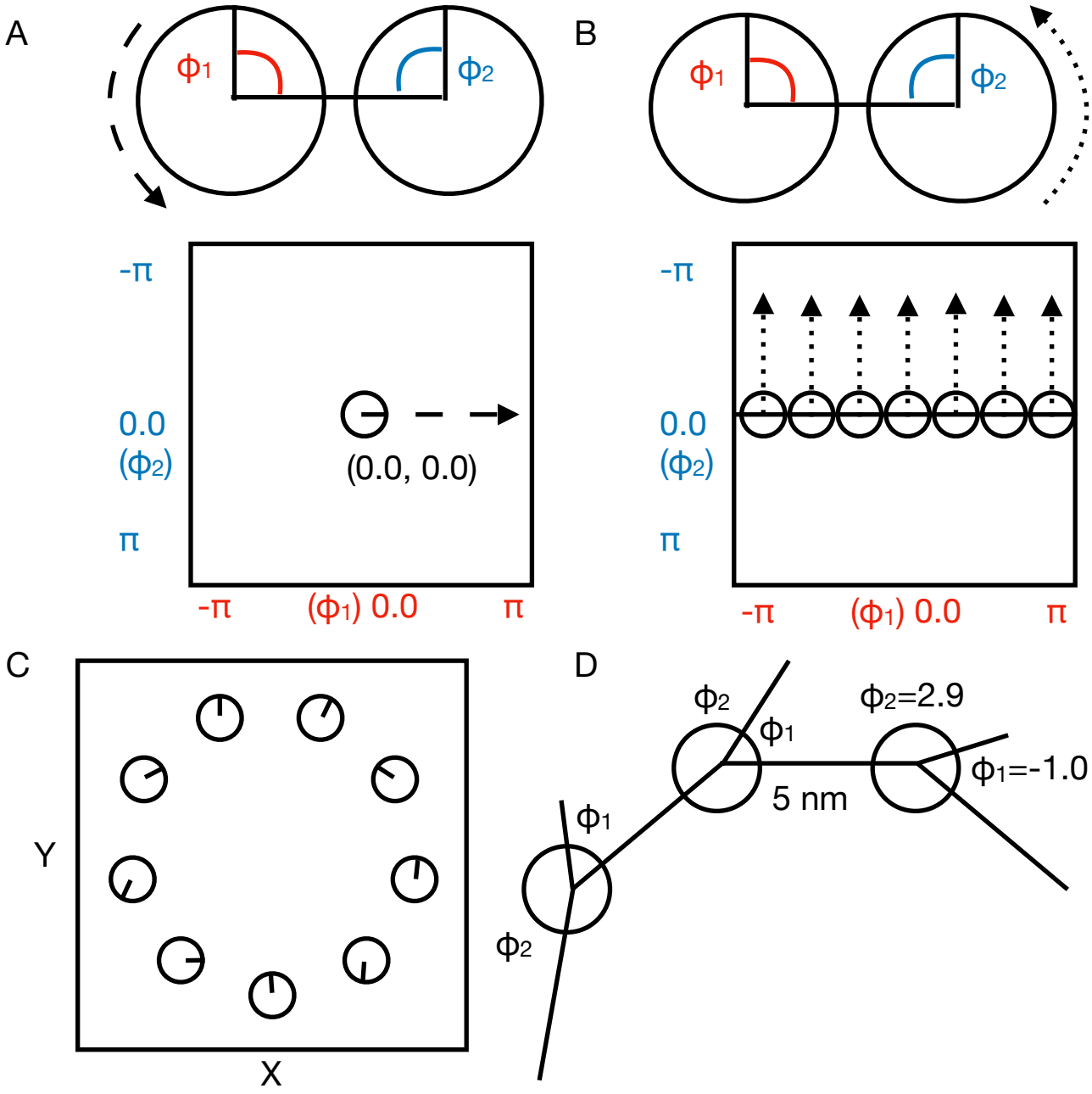


***Figure 5.3**. Equilibration stages of the two-dimensional potential of mean force calculations and the nonamer formation. The two dihedrals ϕ1 and ϕ2 are used to describe the relative orientation of both proteins with respect to each other. A. Starting from ϕ1 =0 and ϕ2 =0, ϕ2 was kept constant while ϕ1 was moved from 0 to π and from -π to 0. C. 63 simulations were then started where ϕ1 was fixed and ϕ2 was moved from 0 to π and from -π to 0. D. Nine copies of randomly orientated proteins were position-restrained to precomputed coordinates which form an idealised nonagon with an edge of 5 nm. E. The position restraints were replaced with distance restraints of 5 nm between the centre of the proteins and two dihedral restraints where ϕ1 = -1.0 and ϕ2 = 2.9.*

5.2.5 Semi-restrained Nonamer Formation

To test if it is energetically favourable for the nonamer to form, one protein system was replicated in the x and y dimensions three times to make a nine protein patch. The proteins were position restrained to the vertexes of an idealised nonagon (edge = 5 nm) at 1000 kj/(mol.nm) (Figure 5.3C) through an 100 ns equilibration run. After verifying that all proteins were in their corresponding positions, the position restraint was discarded and a distance restraint of 5 nm between each pair of neighbouring proteins were applied with a force constant of 1000 kj/(mol.nm). A dihedral restraint with a force constant of 1000 kj/(mol.rad) was also applied at the same time to ensure that the adjacent proteins were orientated with two dihedral angles being -0.99 and 2.87 radian (Figure 5.3D). The system was then equilibrated for 200 ns, before a further 10 μs production run where the distance restraint was replaced with one-side distance restraint. After the formation of nonamer, a further 10 μs production run was performed with all restraints removed to check the stability of the nonamer.

5.2.6 Attach Pull Release Calculations

To independently assess the dimerisation free energy with dihedral pair (-1.0, -1.0)/ (-1.0, 2.9) and the potential impact from protonation states, the attached-pull-release scheme (305) was used. The scheme consisted of three stages, where in the attached phase, two dihedral restraint (ϕ_1, ϕ_2) as well as a distance restraint were gradually added to the system. 14 lambda windows ranging from 0, 0.004, 0.008, 0.016, 0.024, 0.04, 0.055, 0.0865, 0.118, 0.181, 0.244, 0.37, 0.496, 0.748 and 1.0 were used to switch the restraint strength from 0 to 100 kj/mol.rad for dihedral restraints and 1000 kj/mol.nm for distance restraint. The pull phase pulls the two proteins apart while maintaining the dihedral restraints. The umbrella windows were setup from the equilibrium dimer distance to 5 nm with a step of 0.04 nm. The release phase releases the dihedral restraint while keeping the

distance restraint intact. 14 windows similar to the attached phase were used to scale the restraint strength to 0. All of the windows of the attach, pull and release phases were run for 2 μs. The free energy calculation for the attach and the release phase were done using Multistate Bennett Acceptance Ratio with alchemical analysis (215), while WHAM is used for pull phase. A semi-analytical correction is used to release distance restraint as well as correct the concentration to one molar.

$$W_{release} = RT \ln\left(\frac{C^{\circ}}{8\pi^2}\right) + RT \ln\left(\int_0^{\infty} r^2 e^{-\beta k_r (r-r_0)^2} dr\right) \quad (5.1)$$

C° is the standard concentration, r_0 is the distance restraint (5 nm), k_r is the restraint strength (1000 kj/mol/nm). $W_{release}$ evaluates to -3.696 kcal/mol under 320 K.

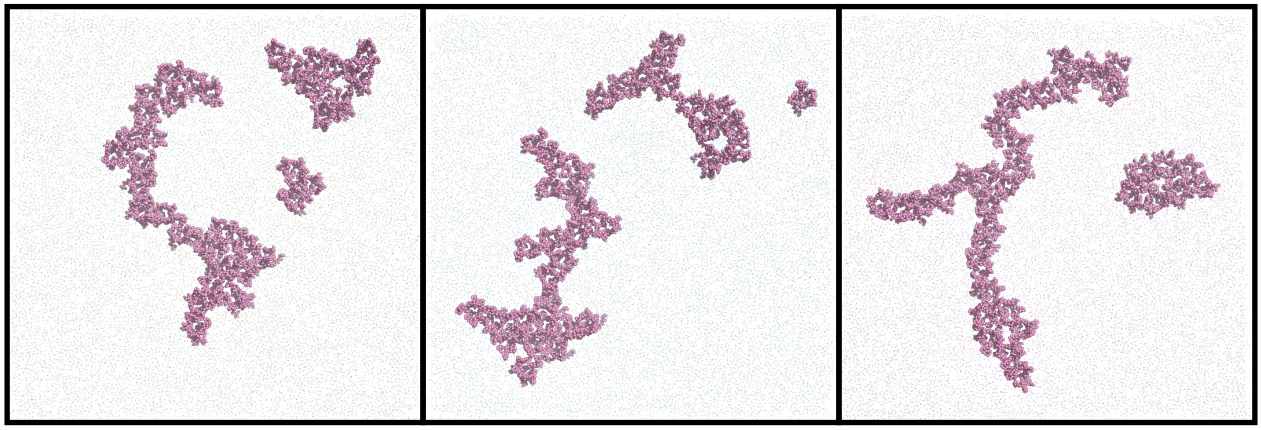
5.3 Results

5.3.1 Unbiased Simulations Showed Multiple Possible Dimerisation Interfaces

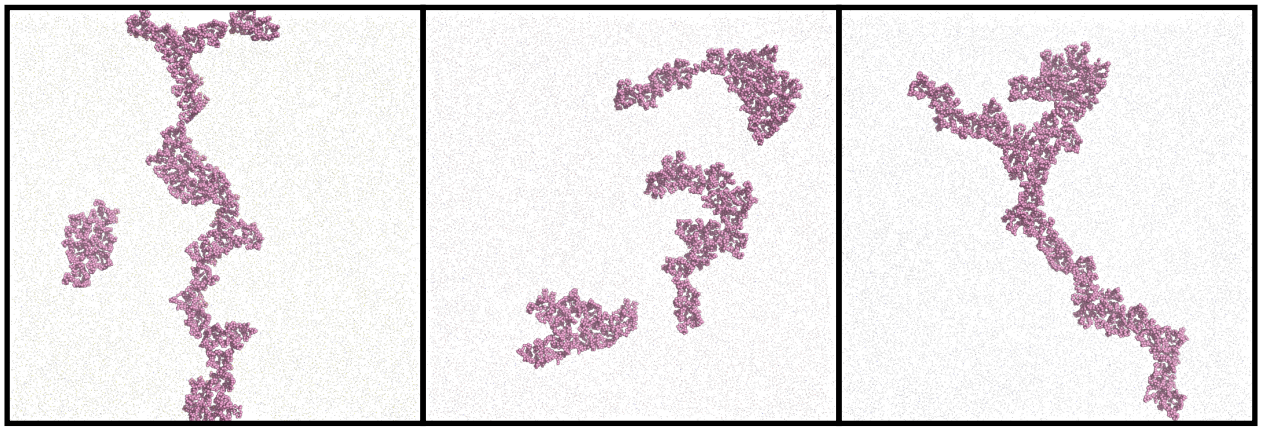
To observe the oligomerisation process and to test if the KDEL-bound structure has a different oligomerisation tendency compared with the apo structure, unbiased simulations have been run on a membrane patch containing 36 KDEL receptors. At the end of the 10 μs simulations, all three repeats show different degrees of oligomerisation, as shown in Figure 5.4AB. For the KDEL-bound structure, the dimer interface characterised by the dihedral pairs shows up in the two-dimensional dihedral space as very few distinctive isolated hot spots (notably (-1.0, -1.0) and (-0.5, -2.5)) and a long strip close to (-1, 2) (Figure 5.4C). It is worth noting that the dihedral pair (-1.0, -1.0) is the only pair that shows up in more than one repeat. As for the Apo structure, more hot spots can be observed in the dihedral space (Figure 5.4D) and the hot spots are more diffuse compared with their

counterpart in the KDEL-bound structure. There are huge variations between different repeats and the only hot spot that has been observed in more than one repeat is at (-2,1). It is clear that the preferential dimer interface is not converged in the dihedral space. Considering the long dissociation time of protein dimers, extending the simulation beyond 10 μs is not likely to help convergence. Though running more repeats might alleviate this problem, the ultra-high protein concentration (36 copies of the protein in a box of 23,000 nm^3 ≈ 50 mg/ml), which is a concentration where proteins will normally precipitate, means that artificial protein-protein interfaces can happen.

A. KDEL-bound structure



B Apo structure



C

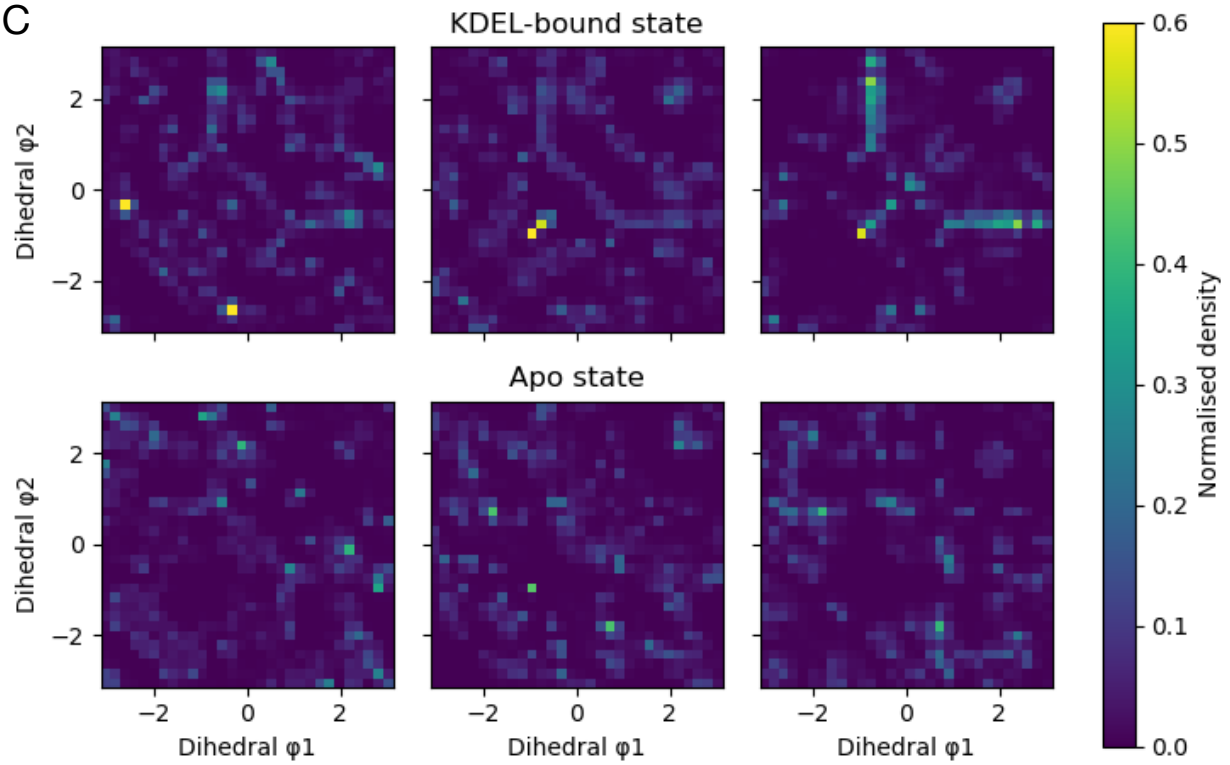


Figure 5.4*. Unbiased oligomerisation simulations. The three repeats of the unbiased simulations of KDEL-bound conformation (A) and Apo conformation (B) all yield very different final frames. (C). Few distinct hotspots were observed in dihedral space for KDEL-bound structure with (-1.0,-1.0) being the only one observed in two repeats. More diffuse hotspots were observed for apo structure with (-2,1) being the recurring hotspot. Note: density is normalised where the sum of bin area times bin height is one.*

To observe the dimer interface in a more physiological concentration, a membrane patch containing two proteins was set up (Figure 5.5A) and 49 repeats were performed for both structures for 10 μs. As is shown in Figure 5.5B, the two proteins were able to explore different dihedral spaces before eventually forming a dimer. After the dimer formation, the two proteins will usually stay tightly bound to each other for the rest of the simulation. For the KDEL-bound structure (Figure 5.5C), the stable dimer interfaces (stay bound for more than 4 μs) fell roughly in the same area as is observed in the 36 proteins oligomerisation simulations. Notably, 6 out of the 49 simulations end up with a dimer interface of (-1.0, -1.0), which is the recurring hotspot observed in the oligomerisation simulations. Similarly, four simulations (three for (-1.2, 2.5) and one for (-1.3, 2.7)) ended up in the long strip of (-1, 1-3), which is also consistent with the oligomerisation simulations. For apo structures (Figure 5.5D), different dimer interfaces were observed in the dimer simulations versus the oligomerisation simulations. The recurring hot spot of (-2, 1) was not found in any of the 49 repeats. Some highly populated hotspots (-1.0, 2.8) (n=3), (-1.2, 1.8) (n=3), (-1.8, 2.5) (n=4) and (-2.3, 0.3) (n=3) were not observed at all in the oligomerisation simulations. The only hotspot that is shared between the dimer simulations and oligomerisation simulations seems to be (-1.0, -1.0) (n=4). However, unlike the case with KDEL-bound structure where there are few well-defined hotspots in a very narrow region in the dihedral space, for apo structure, hotspots are relatively diffuse and cover the whole dihedral space, which is a feature shared between the dimer simulations and the oligomerisation simulations.

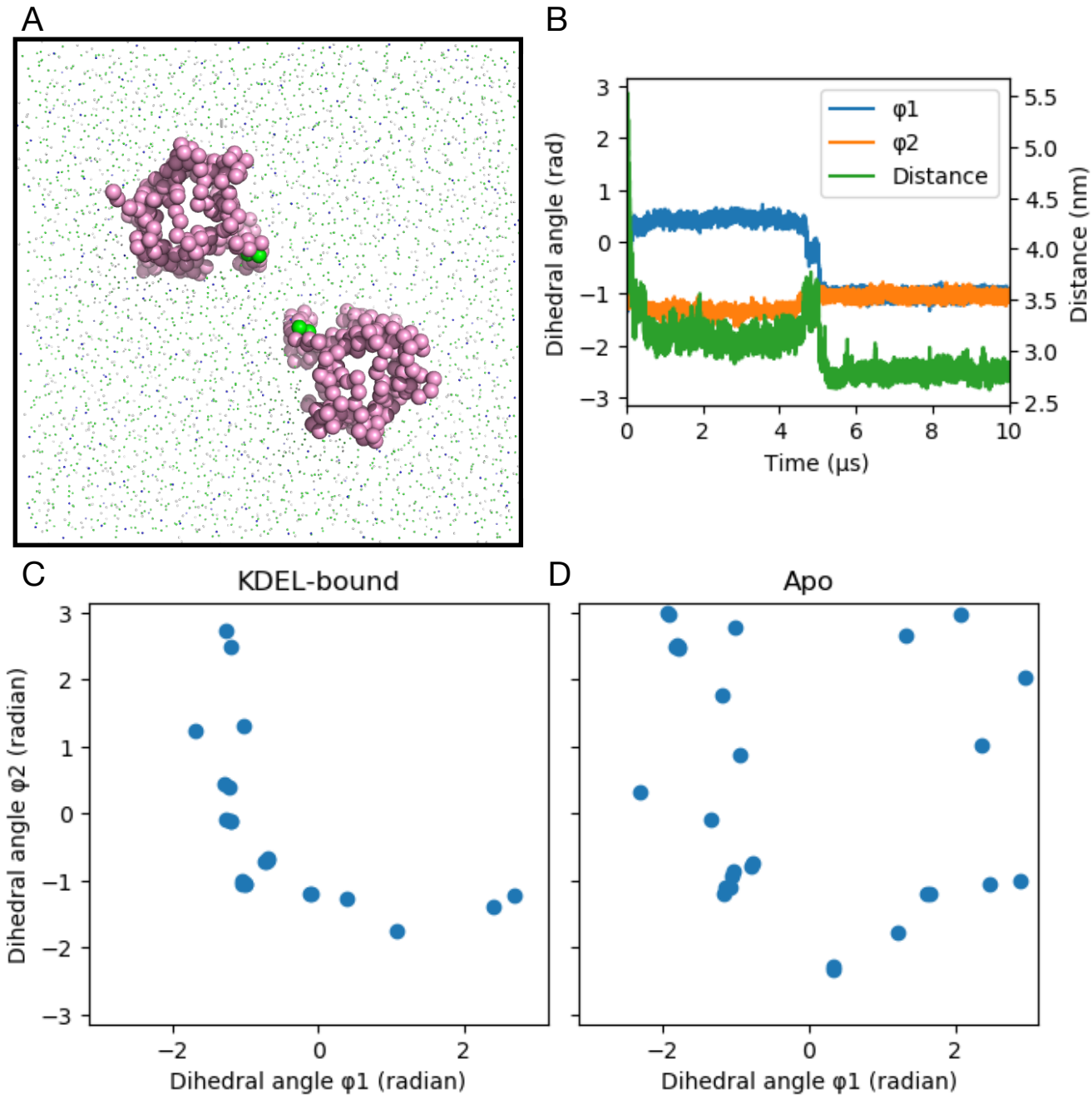


Figure 5.5*. Unbiased dimerisation simulations. A. The two proteins were separated with 5 nm in a membrane patch with the reference point painted as Green. B. The two dihedrals (blue and orange) fluctuated a bit before reaching the final (-1.0, -1.0) pair. The distance (blue) between two proteins dropped readily at the start of the simulation and then stabilised at 2.8 nm. C. The dimer interfaces which KDEL-bound structure preferentially populates is similar to the oligomerisation simulations, with many repeats dimerise as (-1.0, -1.0). D. The apo structure forms a stable dimer in a more widespread region compared with the KDEL-bound structure. The (-1.0, -1.0) is still a hotspot for apo structure.*

5.3.2 Two-dimensional Free Energy Calculations Reveal the Most Populated Dimer Interface

To obtain a converged view of the preferred dimer interface as well as their relative population, two-dimensional (2D) potential of mean force (PMF) calculations on both KDEL-bound and apo state were performed. For a homo-dimer, the free energy landscape should be symmetric across the diagonal line. Though no symmetry has been imposed, the free energy landscape are pretty symmetrical as is shown in Figure 5.6CD with a mean unsigned difference of 0.3 kcal/mol (KDEL-bound) (Figure 5.6C) and 0.9 kcal/mol (Apo) (Figure 5.6D) between the symmetry mates, illustrating a good convergence. Though the occupancy plot of the apo structure are not fully symmetrical (Figure 5.6B), where the hotspots (-1.8, 2.5) and (3.0, 2.1) were visible but their symmetrical pair (2.5, -1.8) and (2.1, 3.0) were not, it can be seen that in the free energy plot, both (-1.8, 2.5)/(2.5, -1.8) pair and (3.0, 2.1)/(2.1, 3.0) pair have quite similar free energy. Since the occupancy is exponentially related to the free energy, it is not surprising that a tiny bit of noise in the free energy landscape will result in an asymmetric occupancy plot.

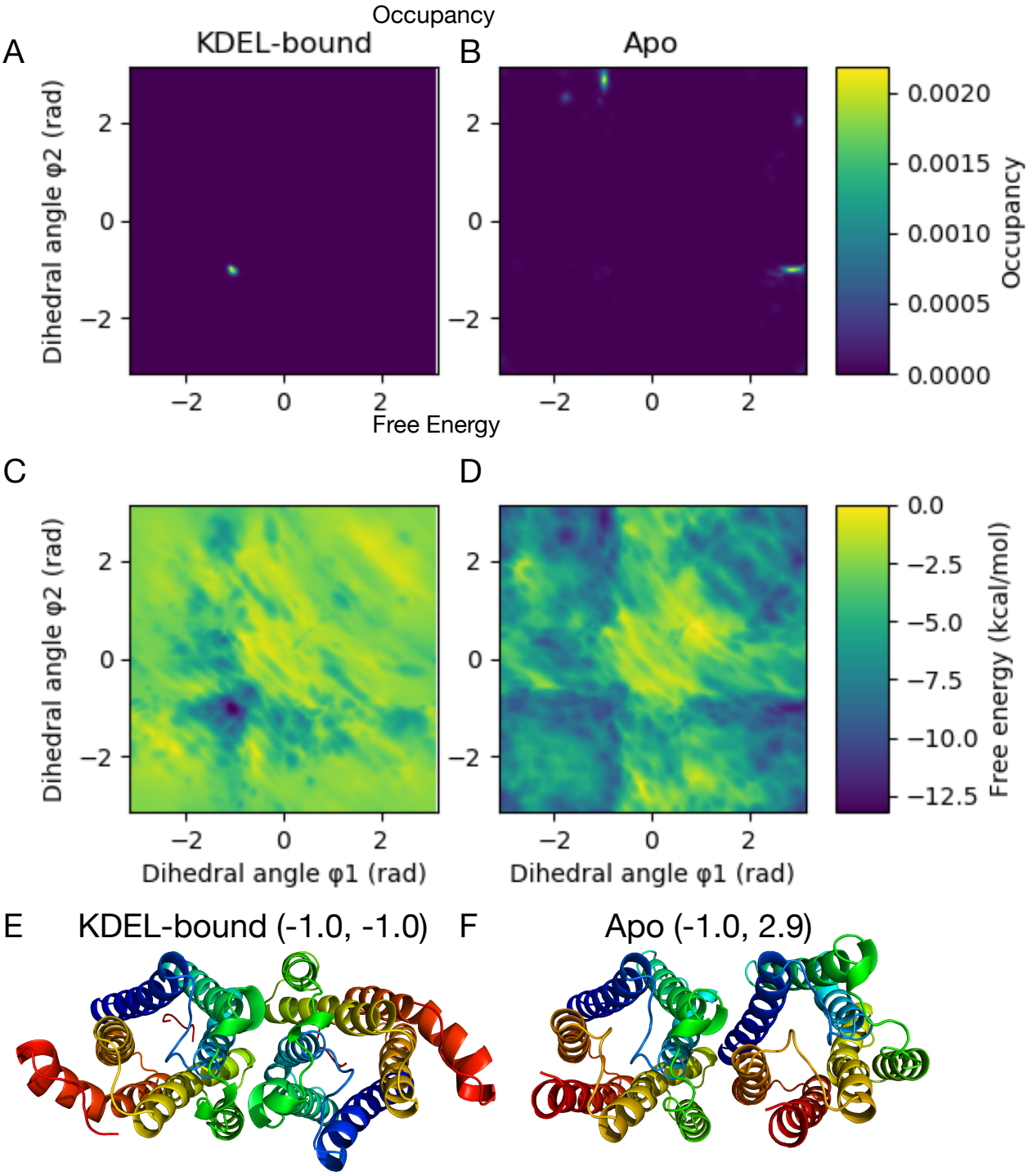


***Figure 5.6**. Two-dimensional occupancy and free energy landscape. The KDEL-bound structure (AC) favours overwhelming the (-1.0, -1.0) dimerisation state (E), whereas most other dimerisation states are unfavourable. On the other hand, the apo structure (BD) enables many favourable dimerisation states with the (-1.0, 2.9) (F) being the most favoured one. Both free energy landscapes (CD) show good convergence across the diagonal axis. Note: Occupancy sums up to one.*

The 2D-PMF agrees well with the dimerisation simulations and oligomerisation simulations for KDEL-bound proteins. When the KDEL-bound protein forms a dimer, the majority of them will have a dimer interface of (-1.0, -1.0) (Figure 5.6E), whose free energy of dimerisation is the lowest in the whole dihedral space (Figure 5.6C) and it is the solo point that is visible from the occupancy plot (Figure 5.6A). This dimer interface is the recurring hotspot in the oligomerisation simulations and is the spot where 6 out of the 49 dimerisation simulations end up with. Notably, most of the area in the dihedral space are not favourable where only 32% of the whole space has a free energy value under -3 kcal/mol. Furthermore, the region of relatively low free energy are all concentrated in the long strip with one dihedral being -1.0 and the other dihedral ranging from -1 to 3. This is consistent with dimerisation simulations where all dimer interfaces end up in this long strip region and the long strip of (-1.0, 1-3) as is found in the oligomerisation simulations.

The 2D-PMF also explains the inconsistency between dimerisation simulations and oligomerisation simulations for the apo protein. In contrast to the small percentage of low free energy region (32%) for the KDEL-bound structure, the majority (86%) of the dihedral space has a free energy of under -3 kcal/mol, suggesting that the protein can dimerise and adopt a wide range of dimer interfaces. Considering that in the timescale accessible to the technique that it being used, the presence of numerous local minimum means that unbiased simulations are bound to get trapped in these local minima and are unlikely to sample the whole dihedral space without considerably more repeats. Thus, the inconsistency between the dimerisation and oligomerisation simulations as well as the inconsistency between different repeats can be explained by insufficient sampling. However, the 2D-PMF do suggest that the protein dimer are unlikely to form when both dihedral angles are within the range of -1.0 to 1.0. This is consistent with the dimerisation simulations and the oligomerisation simulations. It is also worth noting that all of the

preferred dimer interfaces for the KDEL-bound structure are also preferred for apo structure. The conformational change from apo structure to KDEL-bound structure seems to make many dimer interfaces unfavourable, such as (-1.0, 2.9) (Figure 5.6F).

5.3.3 Apo Structure Permits a Nonamer Formation

Interesting, the energy minimum of the apo structure has a dihedral pair of (-1.0, 2.9) radian or (-56.47, 164.42) in degrees, which would allow an angle of 139.11 degree to form between a trimer linked from head to tail. The angle 139.11 is very close to 140, which would suggest that the apo structure could potentially form a nonamer, whereas the KDEL-bound structure can only form a dimer.

A system with nine copies of KDEL receptors were arranged in an idealised nonagon and allowed to associate with the orientation fixed at (-1.0, 2.9) (Figure 5.7A). Within 10 μs, the Apo proteins were able to form the predicted nonamer (Figure 5.7BD), whereas the KDEL-bound structure struggled to even form a dimer within the same time frame (Figure 5.7C). To test the stability of the nonamer, the apo nonamer were run for a further 10 μs without any restraint and all the protein pairs stayed tightly bound across the 10 μs simulation (Data not shown).

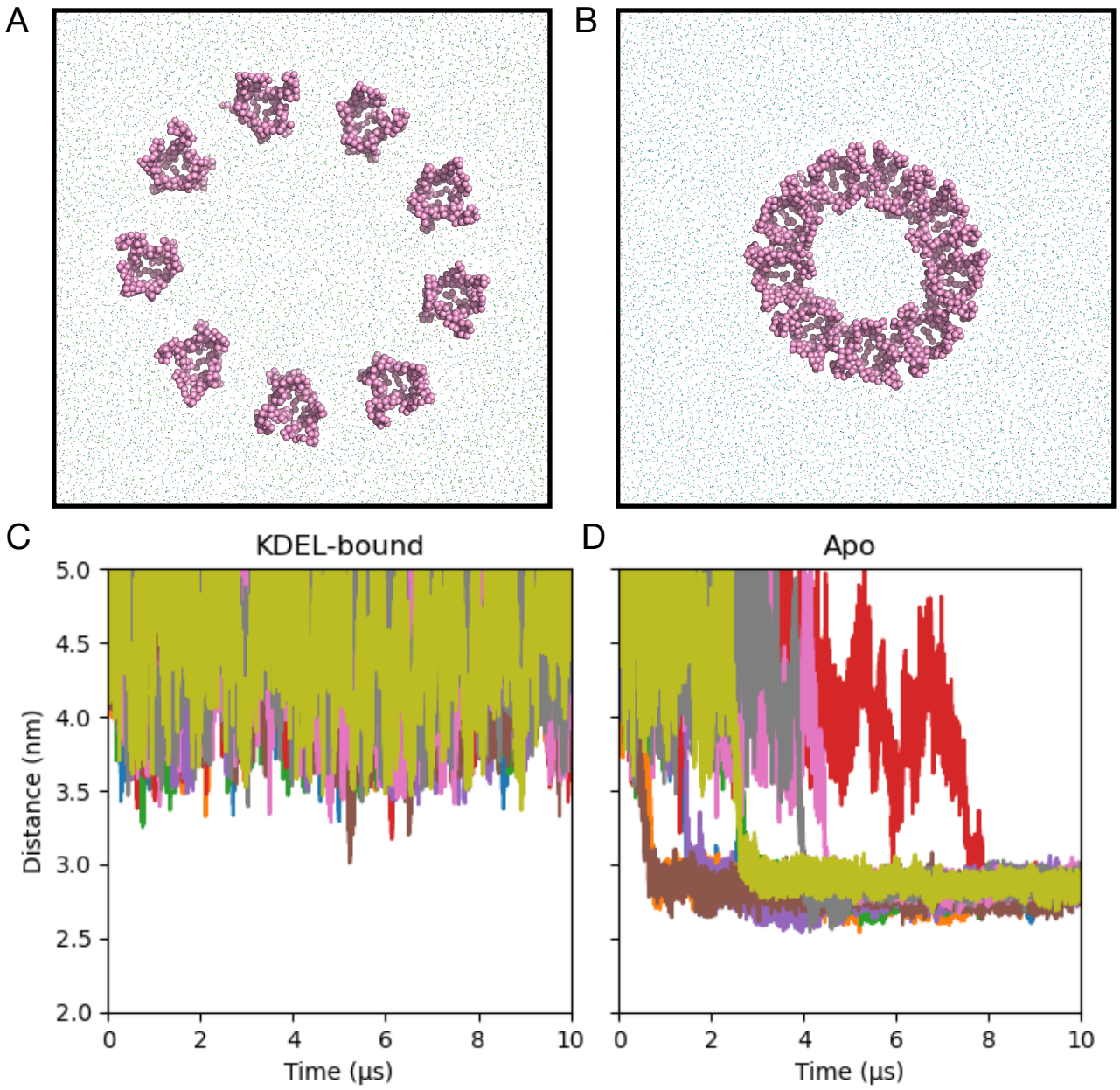


***Figure 5.7**. Nonamer formation. Nine copies of the apo structure (A) will readily form a nonamer (B). For the Apo state, between all pairs of the neighbouring proteins, the distance drop to below 3 nm within 10 μs (D). The monomer complex has stayed stable for another 10 μs, where all the protein pairs are in the tightly bound position (data not shown). The KDEL-bound structure, on the other hand, never form a dimer in the same timescale (C).*

5.3.4 Protonation of Histidine Destabilised the Apo Dimer

In the 2D-PMF calculations, it has been shown that the apo conformation favours the (-1.0, 2.9) dimer. Interestingly, a protonatable residue (H150) is observed in the dimer interface (Figure 5.8B). To investigate the effect of this key histidine, the absolute dimerisation free energy has been computed for both the apo and the KDEL-bound dimer with H150 in both protonated and deprotonated state (Figure 5.8AB). As is shown in Figure 5.8C, in the KDEL-bound state, since H150 is far from the dimerisation interface, different protonation states have very little effect on the dimerisation free energy. In the apo state, on the other hand, the dimerisation becomes less favourable when the histidine is protonated.

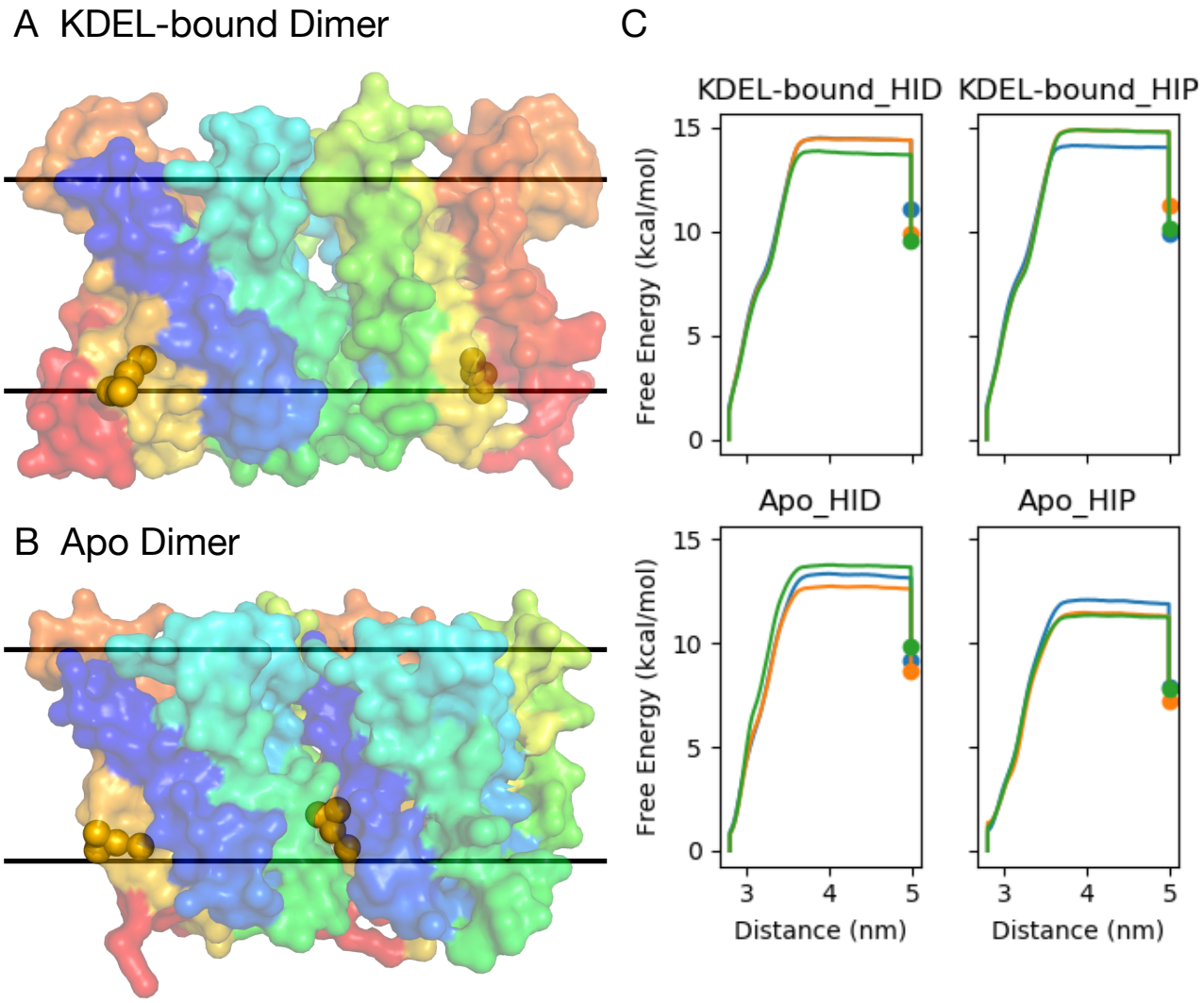


Figure 5.8. Protonation state of the H150 influences the dimerisation process. When the KDEL-bound KDEL receptor forms a dimer, the histidine (H150: Brown sphere) is located in the peripheral region (A) and the protonation state of the histidine has no effect on the dimerisation (C). When the apo state KDEL receptors form a dimer, H150 (H150: Brown sphere) is in the dimer interface (B) and the protonation of the histidine destabilised the dimer (C).

5.4 Discussion

In this study, three repeats of the 36 protein oligomerisation simulations and 49 repeats of dimerisation simulations have been performed to characterise the dimerisation and oligomerisation properties of both KDEL-bound and apo conformation of the KDEL receptor. Though some preferred dimer interfaces have been identified, significant difference between different repeats was observed, which is consistent with Wassenaar *et al.*'s study, where 500 repeats would still result in insufficient sampling (306). To obtain the most populated dimer, 2D-PMF calculations which cover all potential dimer interfaces were performed. For the KDEL-bound state, forming a dimer is only energetically favourable with dimerisation interface being (-1.0, -1.0). For the apo structure, on the other hand, many dimer interfaces are energetically favourable with (-1.0, -1.0) being one of them. Among these dimer interfaces, the (-1.0, 2.9) pair is the most stable one. Furthermore, the dimer interface of (-1.0, 2.9) permits the formation of a nonamer, where the KDEL-bound structure only allows the dimer formation.

In the Golgi, when a protein with the KDEL-tag binds the apo KDEL receptor, conformational change happens, which allows the Helix 7 to adopt a kinked position (133). Majoul *et al.*'s study shows that the oligomerisation level of the KDEL receptor increases sharply after the binding of the KDEL tag (307). This is consistent with the results that in the Golgi where the H150 is protonated, the KDEL-bound dimer (-1.0, -1.0) is 2.8 kcal/mol more stable compared with the apo dimer (-1.0, 2.9). This means that the conformational change would lead to an increase in the amount of KDEL-bound dimer.

After the binding of the KDEL-tag, the KDEL receptor/KDEL-tag complex will interact with the ARFGAP, ARF1 complex to form the COPI vesicle and eventually get transported to the ER. It was found that both the over-expression of the KDEL receptor and the binding

of the KDEL tag enhanced the ARFGAP activity (308, 309). The results have also provided an explanation for this phenomenon assuming that the activation of ARFGAP is governed by the “correct” dimerisation of KDELR. Since the KDEL-bound KDELR dimerisation occurs mainly in the (-1.0, -1.0) fashion, it could be hypothesised that the formation of (-1.0, -1.0) dimer is crucial to ARFGAP activation. The apo state, on the other hand, permits both (-1.0, -1.0) and (-1.0, 2.9) dimers. Since the apo (-1.0, -1.0) dimer is less stable compared with the apo (-1.0, 2.9) dimer, the population of (-1.0, -1.0) dimer is very rare compared with the (-1.0, 2.9) dimer resulting in little ARFGAP activation. Through the over-expression of the KDEL receptor, the greater number of available KDEL receptors means a greater number of apo proteins would adopt the (-1.0, -1.0) dimer and, thus, leads to the increase in the activation of ARFGAP.

Having shown that the different dimerisation behaviour of the apo and KDEL-bound receptor could explain the export of KDEL-bound KDELR to the ER, one question remains unanswered - how does the apo KDELR in the ER get trafficked back to the Golgi?
It has been shown that the KDEL receptor travels extremely fast (much faster than the bulk flow) in the ER-Golgi pathway and the process is mediated by the two key leucines (143). One possible answer to this question is that since the apo (-1.0, 2.9) dimer is more energetically favourable when H150 is deprotonated (HID: -7.6; HIP: -9.2 kcal/mol), the nonamer is prevalent in the ER compared with the Golgi which has a lower pH. The formation of nonamer allows a much faster retrieval of the apo KDELR from the ER to Golgi.

Based on these results, a model has been purposed to explain the shuttling of the KDEL receptor between the ER and the Golgi as is summarised in Figure 5.9. Upon the binding of the KDEL-tagged protein, the KDEL receptor performs conformational change, which

allow the KDEL receptor to form the (-1.0, -1.0) dimer, which is recognised by COPI and is trafficked to the ER. The KDEL-tagged protein is released in the ER, which allows the KDEL receptor to go back to the apo state. The conformational change allows the receptor to switch from the (-1.0, -1.0) dimer to the (-1.0, 2.9) dimer, which oligomerises to form a nonamer. The nonamer interacts with the COPII vesicle and is transported to the Golgi. In the Golgi, H150 is protonated, which leads to the dissociation of the nonamer and, thus, completes the cycle.

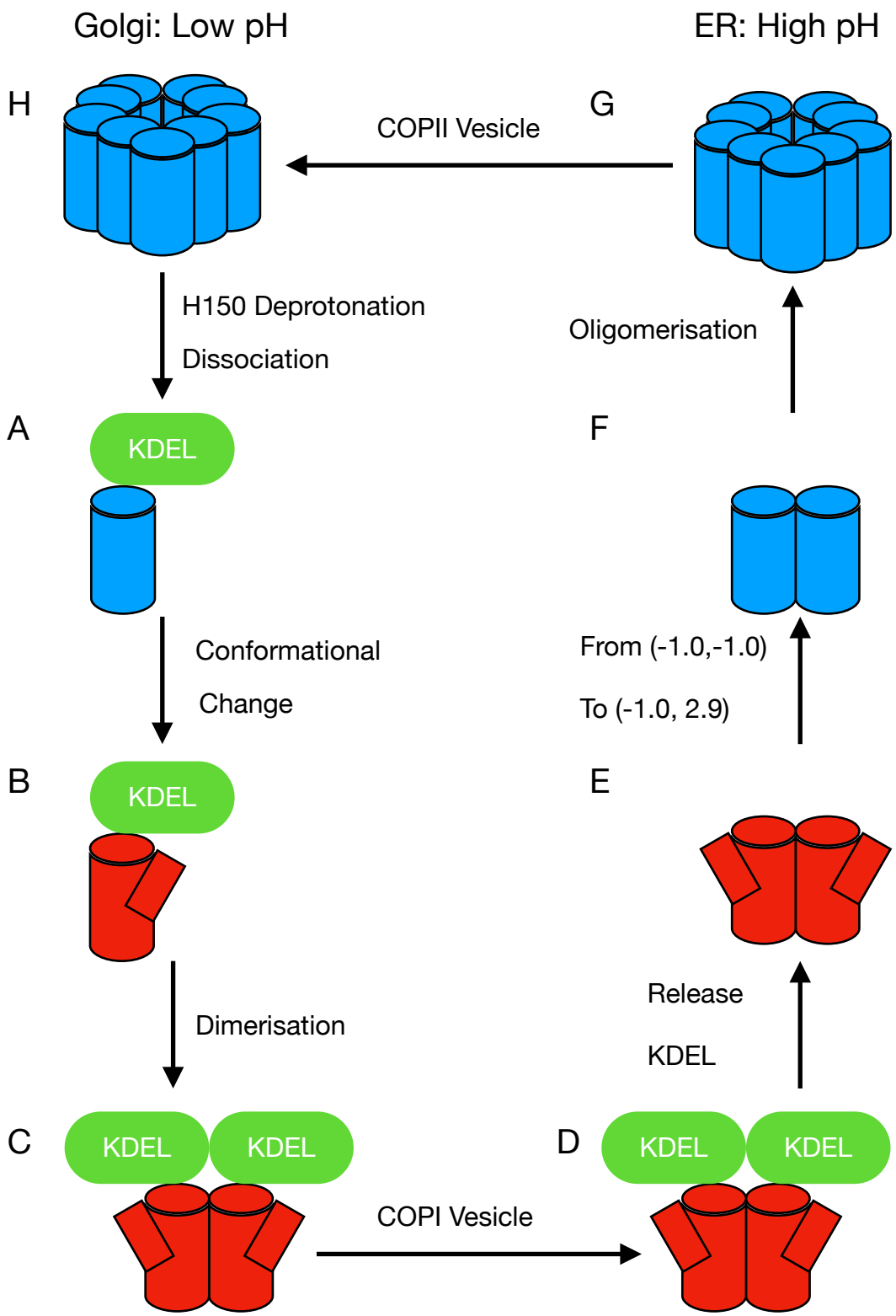


Figure 5.9. A schematics of the shuttling of the KDEL receptor between ER and Golgi. In the Golgi, the binding of the KDEL-tagged protein to KDEL receptor (A) triggers the conformational change (B), which lead to the formation of dimer (C). The dimer is transported to the ER (D) by COPI vesicle and the KDEL-tagged protein is released (E). In the absence of the ligand, the KDEL receptor adopted another dimer interface (F), which allows the formation nonamer (G). The nonamer is transported to the Golgi (H) and the deprotonation of H150 allows the nonamer to dissociate and completes the cycle.

However, these results are all built under the assumption that CG force field could faithfully reproduce the protein-protein interactions, which is a difficult feat even for atomistic simulations (310). The electrostatic interaction in CG force field are computed using reaction field method to achieve maximum efficiency. However, any charge interaction beyond the cutoff would not be counted in reaction field method. Furthermore, due to the large size of the martini beads, energy barrier for separating a pair of beads would be larger than atomistic simulations as the initial solvation is harder due to the large Martin bead size (311). Last but not least, the protein, when represented with Martin force field, cannot perform conformational change. This might affect the dimerisation free energy, as protein might dimerise through an induced-fit scheme.

5.5 Conclusion

In this chapter, It has been shown that the conformational change upon the binding of the KDEL-tagged protein changes the dimerisation and oligomerisation property of the KDEL receptor. It has also been shown that through 2D-PMF calculations, the whole dimerisation free energy landscape could be systematically sampled to obtain a converged result. The KDEL-bound structure can form a homodimer with a (-1.0, -1.0) interface, which could allow it to interact with the COPI vesicle. The Apo structure prefers the (-1.0, 2.9) dimer interface, which has the potential of forming a nonamer under neutral pH, which could explain the fast retrieval of apo KDEL receptor from the ER to the Golgi.

6 Conclusion and Future Work

In this thesis, it has been shown that protons can take part in various biological processes such as transporter substrate recognition, conformational change and oligomerisation. However, there are some pitfalls that still have not been addressed.

In chapter 3, part of the transport cycle for GkApcT, namely from the inward-occluded conformation to the inward-open conformation has been elucidated in detail. However, the detailed mechanism of the whole transport cycle is still elusive. One difficulty of addressing this question is that the timescale of the whole transport cycle exceeds the timescale achievable with normal MD. To accelerate this process, one could use different enhanced sampling techniques such as steered MD, metadynamics or the weighted ensemble method. However, the success of these techniques requires the definition of a collective variable that captures the conformational change. Though Chapter 3 has shown that the CV being used is not ideal in discriminating the inward-open from the inward-occluded state, it would still be interesting to see if the current CV could be used to drive the closing of the intracellular gate and allow the transporter to go from the inward-open to the occluded state.

Furthermore, though GkApcT is a bacterial homolog of hCAT transporters, only alanine has been examined in this thesis. It would be interesting to compute the transport-related conformational changes in the presence of other relevant amino acids, such as arginine. The computation of arginine binding free energy would involve a change in the total charge of the simulation box and it has been shown that though the binding free energy for the neutral species can be computed successfully, for charged species, this computation can be quite difficult (312). The difficulty may come from the finite size

problem (217),), different level of polarisation in protein and the water or inaccurate partial charge assignment due to the insufficiency at the level of the QM method usually applied (313). Further research is required to tackle all of these problems.

In Chapter 4, it has been shown that the protonation state of the key histidine affects the property of a short hydrogen bond. A small cluster model was computed at a high level of theory (CCSD(T)) and a full model with the whole protein embedded in a lipid membrane was computed with a low level of theory (B97-3c). Though CCSD(T) is a theory level that allows the accurate computation of the desired energy, the model of three residues alone in a vacuum is a limited representation of the real physiological system. On the other hand, though the QM/MM model is a relatively good representation of the physiological system. The energy difference between states that depend on the protonation of the key histidine of 3.5 kcal/mol is on the borderline of the accuracy of the B97-3c method. Furthermore, in the QM/MM system, only three residues are represented at the QM level and it is questionable as to whether the lack of polarisable water might affect the result. Thus, though QM methods have achieved great success when dealing with small molecules, further work is required for finding the best practice of using QM methods in protein systems.

In both Chapter 3 and Chapter 4, one key question that remains unanswered is where did the proton come from and where did the proton go? For the GkApcT, it is unclear as to how the proton binds the glutamate (E127) and how the proton is subsequently released. For the KDEL receptor, one could argue that in the apo state, histidine could be protonated via access to the bulk solvent. However, it is unclear as to how the histidine then gets deprotonated.

Though there has been some work on elucidating the pathway of proton movements in proteins (314), such rare cases are only feasible when the proton path is clearly resolved in the crystal structure (or established exhausitively via computation). For the situations where the destination of the proton is unknown, there isn't an easy solution for this question and considerably more effort is required. The native way of tracking the proton movement is via QM/MM calculations. However, including all the possible regions where the proton can explore would imply a very large QM area and thus, makes the calculation prohibitively expensive. Another way of tracking the movement of the proton might be a constant pH type of approach. However, cpHMD methods have trouble dealing with the coupling in protonation states between two residues, which can mean an artificially large energy barrier for proton transfer between residues.

Chapter 5 examines the oligomerisation property of the KDEL receptor and the influence of the protein conformation on the preferred dimer interface. The oligomerisation of the protein could be affected by the physical property of the lipid bilayer as well. It is known that from the ER to the Golgi, the lipid thickness level increases (315), which could potentially have an effect on the oligomerisation property of the KDEL receptor. Furthermore, though it is argued that the saturation level of the ER and the Golgi lipids are quite similar (316), there is evidence suggesting that the cholesterol concentration is quite different between ER and the Golgi (94), which might suggest a difference in lipid saturation level (317). Nevertheless, there is also argument that the different lipid head groups can affect oligomerisation properties (318).

However, investigation of the lipid effect on the oligomerisation process is not a simple feat. An immediate solution would be to repeat the same calculations with different lipids. However, it has been shown that when investigating the oligomerisation problem, the

current martini model might introduce an artificially larger energy barrier for separation due to the size of the martini bead (311). The more recently released martini3 forcefield may offer a remedy to this problem. On the other hand, due to the treatment of the electrostatics in martini model, the protein anionic lipids interactions, which can play a key role in protein functioning (319), are not taken into account properly in the current martini representation (320). Thus, solving the lipid protein interaction and the effect of different lipid composition on the protein dimerisation is still an ongoing problem.

Similar to the sodium, chloride, potassium and calcium ion concentration gradients, the proton gradient powers many transport processes in the human body. In this thesis, it has been shown that secondary active transporters use the proton gradient to transport substances from the extracellular solution into the intracellular environment and the KDEL receptor uses the proton gradient to transport the ER resident protein from the Golgi to the ER. Besides these two proteins, the role of the proton is also highlighted by the acid-sensing ion channels (321), where the proton controls the transport of the ions across the membrane. Similarly, protons play a role in controlling lipid transport through the lipid flippase (322). Likewise, many drugs are absorbed in the small intestine by the PEPT1 peptide transporter, which is another proton-coupled transporter that transports small peptides and drug molecules into the human body via utilising the proton gradient (323).

Though the importance of the proton in various transport systems has been widely recognised, experimental tools for accurately tracking the speed and amount of proton movement have been very limited (324). Thus, computational tools are regarded as a good complement thanks to the fine spatial and temporal resolution as is demonstrated in this thesis. However, these computational tools, such as molecular dynamics simulations are still plagued by the limited timescale of simulations and dynamic treatment of

protonation events. More effort from a combination of software engineering, hardware advancement and better sampling techniques will be required to resolve these problems.

7 Published Work

Chapter 3 (5) and chapter 4 (325) are both published.

Bibliography

1. W. W. Cleland, P. A. Frey, J. A. Gerlt, The Low Barrier Hydrogen Bond in Enzymatic Catalysis. *Journal of Biological Chemistry* **273**, 25529-25532 (1998).

2. L. Wang, S. D. Fried, S. G. Boxer, T. E. Markland, Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc Natl Acad Sci U S A* **111**, 18454-18459 (2014).

3. J. Grdadolnik, F. Merzel, F. Avbelj, Origin of hydrophobicity and enhanced water hydrogen bond strength near purely hydrophobic solutes. *Proceedings of the National Academy of Sciences* **114**, 322 (2017).

4. K. E. J. Jungnickel, J. L. Parker, S. Newstead, Structural basis for amino acid transport by the CAT family of SLC7 transporters. *Nature Communications* **9**, 550 (2018).

5. Z. Wu, I. Alibay, S. Newstead, P. C. Biggin, Proton Control of Transitions in an Amino Acid Transporter. *Biophysical Journal* **117**, 1342-1351 (2019).

6. M. Mihelič *et al.*, The mechanism behind the selection of two different cleavage sites in NAG-NAM polymers. *IUCrJ* **4**, 185-198 (2017).

7. J. Boyle, Molecular biology of the cell, 5th edition by B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Biochemistry and Molecular Biology Education* **36**, 317-318 (2008).

8. J. L. Parker *et al.*, Proton movement and coupling in the POT family of peptide transporters. *Proc Natl Acad Sci U S A* **114**, 13182-13187 (2017).

9. J. M. J. Swanson *et al.*, Proton solvation and transport in aqueous and biomolecular systems: insights from computer simulations. *J Phys Chem B* **111**, 4300-4314 (2007).

10. A. Bruce, *Molecular biology of the cell*, Ebook central (Garland Science, New York, New York, ed. Sixth, 2015).

11. H. Ishikita, K. Saito, Proton transfer reactions and hydrogen-bond networks in protein environments. *J R Soc Interface* **11**, 20130518-20130518 (2013).

12. M. A. Hediger, B. Clémençon, R. E. Burrier, E. A. Bruford, The ABCs of membrane transporters in health and disease (SLC series): Introduction. *Molecular Aspects of Medicine* **34**, 95-107 (2013).

13. D. L. Jack, I. T. Paulsen, M. H. Saier, The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* **146 (Pt 8)**, 1797-1814 (2000).

14. M. A. Hediger *et al.*, The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv* **447**, 465-468 (2004).

15. M. F. White, H. N. Christensen, Cationic amino acid transport into cultured animal cells. II. Transport system barely perceptible in ordinary hepatocytes, but active in hepatoma cell lines. *Journal of Biological Chemistry* **257**, 4450-4457 (1982).

16. N. Vékony, S. Wolf, J.-P. Boissel, K. Gnauert, E. I. Closs, Human Cationic Amino Acid Transporter hCAT-3 Is Preferentially Expressed in Peripheral Tissues. *Biochemistry* **40**, 12387-12394 (2001).

17. M. Pineda *et al.*, The amino acid transporter asc-1 is not involved in cystinuria. *Kidney International* **66**, 1453-1464 (2004).

18. M. S. Kilberg, B. R. Stevens, D. A. Novak, Recent Advances in Mammalian Amino Acid Transport. *Annual Review of Nutrition* **13**, 137-165 (1993).

19. J. Fort *et al.*, The Structure of Human 4F2hc Ectodomain Provides a Model for Homodimerization and Electrostatic Interaction with Plasma Membrane. *Journal of Biological Chemistry* **282**, 31444-31452 (2007).

20. M. Palacín, Y. Kanai, The ancillary proteins of HATs: SLC3 family of amino acid transporters. *Pflügers Archiv* **447**, 490-494 (2004).

21. J. Cantor *et al.*, CD98hc facilitates B cell proliferation and adaptive humoral immunity. *Nature Immunology* **10**, 412-419 (2009).

22. A. Takesono *et al.*, Solute carrier family 3 member 2 (Slc3a2) controls yolk syncytial layer (YSL) formation by regulating microtubule networks in the zebrafish embryo. *Proceedings of the National Academy of Sciences* **109**, 3371 (2012).

23. C. C. Feral *et al.*, CD98hc (SLC3A2) mediates integrin signaling. *Proc Natl Acad Sci U S A* **102**, 355 (2005).

24. Y. Kanai *et al.*, Transport Properties of a System y+L Neutral and Basic Amino Acid Transporter: INSIGHTS INTO THE MECHANISMS OF SUBSTRATE RECOGNITION. *Journal of Biological Chemistry* **275**, 20787-20793 (2000).

25. O. Yanagida *et al.*, Human L-type amino acid transporter 1 (LAT1): characterization of function and expression in tumor cell lines. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1514**, 291-302 (2001).

26. E. M. del Amo, A. Urtti, M. Yliperttula, Pharmacokinetic role of L-type amino acid transporters LAT1 and LAT2. *European Journal of Pharmaceutical Sciences* **35**, 161-174 (2008).

27. D. Drew, O. Boudker, Shared Molecular Mechanisms of Membrane Transporters. *Annual Review of Biochemistry* **85**, 543-572 (2016).

28. L.-Q. Chen *et al.*, Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527-532 (2010).

29. Y. Zhai, W. H. M. Heijne, D. W. Smith, M. H. Saier, Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predications for a putative fungal chaperone protein. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1511**, 206-223 (2001).

30. Y. Xu *et al.*, Structures of bacterial homologues of SWEET transporters in two distinct conformations. *Nature* **515**, 448-452 (2014).

31. Y. Lee, T. Nishizawa, K. Yamashita, R. Ishitani, O. Nureki, Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. *Nature Communications* **6**, 6112 (2015).

32. J. Wang *et al.*, Crystal structure of a bacterial homologue of SWEET transporters. *Cell Research* **24**, 1486-1489 (2014).

33. L. J. Y. M. Swier, A. Guskov, D. J. Slotboom, Structural insight in the toppling mechanism of an energy-coupling factor transporter. *Nature Communications* **7**, 11072 (2016).

34. E. I. Closs, J. P. Boissel, A. Habermeier, A. Rotmann, Structure and Function of Cationic Amino Acid Transporters (CATs). *The Journal of Membrane Biology* **213**, 67-77 (2006).

35. E. Gasol, M. Jiménez-Vidal, J. Chillarón, A. Zorzano, M. Palacín, Membrane Topology of System Xc- Light Subunit Reveals a Re-entrant Loop with Substrate-restricted Accessibility. *Journal of Biological Chemistry* **279**, 31228-31236 (2004).

36. I. A. Hansen *et al.*, AaCAT1 of the Yellow Fever Mosquito, Aedes aegypti: A NOVEL HISTIDINE-SPECIFIC AMINO ACID TRANSPORTER FROM THE SLC7 FAMILY. *Journal of Biological Chemistry* **286**, 10803-10813 (2011).

37. N. Reig *et al.*, Functional and Structural Characterization of the First Prokaryotic Member of the L-Amino Acid Transporter (LAT) Family: A MODEL FOR APC TRANSPORTERS. *Journal of Biological Chemistry* **282**, 13270-13281 (2007).

38. Y. Fang *et al.*, Structure of a prokaryotic virtual proton pump at 3.2Å resolution. *Nature* **460**, 1040-1043 (2009).

39. X. Gao *et al.*, Structure and Mechanism of an Amino Acid Antiporter. *Science* **324**, 1565 (2009).

40. X. Gao *et al.*, Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463**, 828-832 (2010).

41. L. Kowalczyk *et al.*, Molecular basis of substrate-induced permeation by an amino acid antiporter. *Proceedings of the National Academy of Sciences* **108**, 3935 (2011).

42. P. L. Shaffer, A. Goehring, A. Shankaranarayanan, E. Gouaux, Structure and Mechanism of a Na⁺-Independent Amino Acid Transporter. *Science* **325**, 1010 (2009).

43. D. Ma *et al.*, Structure and mechanism of a glutamate–GABA antiporter. *Nature* **483**, 632-636 (2012).

44. A. Yamashita, S. K. Singh, T. Kawate, Y. Jin, E. Gouaux, Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. *Nature* **437**, 215-223 (2005).

45. G. Jeschke, A comparative study of structures and structural transitions of secondary transporters with the LeuT fold. *European Biophysics Journal* **42**, 181-197 (2013).

46. P. Bartoccioni *et al.*, Role of Transmembrane Domain 8 in Substrate Selectivity and Translocation of SteT, a Member of the l-Amino Acid Transporter (LAT) Family. *Journal of Biological Chemistry* **285**, 28764-28776 (2010).

47. S. K. Singh, C. L. Piscitelli, A. Yamashita, E. Gouaux, A Competitive Inhibitor Traps LeuT in an Open-to-Out Conformation. *Science* **322**, 1655 (2008).

48. K. H. Wang, A. Penmatsa, E. Gouaux, Neurotransmitter and psychostimulant recognition by the dopamine transporter. *Nature* **521**, 322-327 (2015).

49. J. Andersen *et al.*, Molecular determinants for selective recognition of antidepressants in the human serotonin and norepinephrine transporters. *Proceedings of the National Academy of Sciences* **108**, 12137 (2011).

50. A. Penmatsa, K. H. Wang, E. Gouaux, X-ray structures of Drosophila dopamine transporter in complex with nisoxetine and reboxetine. *Nature Structural & Molecular Biology* **22**, 506-508 (2015).

51. A. Penmatsa, K. H. Wang, E. Gouaux, X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85-90 (2013).

52. S. Apparsundaram, D. J. Stockdale, R. A. Henningsen, M. E. Milla, R. S. Martin, Antidepressants Targeting the Serotonin Reuptake Transporter Act via a Competitive Mechanism. *Journal of Pharmacology and Experimental Therapeutics* **327**, 982 (2008).

53. S. K. Singh, A. Yamashita, E. Gouaux, Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448**, 952-956 (2007).

54. L. Shi, M. Quick, Y. Zhao, H. Weinstein, J. A. Javitch, The mechanism of a neurotransmitter:sodium symporter--inward release of Na+ and substrate is triggered by substrate in a second binding site. *Molecular cell* **30**, 667-677 (2008).

55. D. Joseph, S. Pidathala, A. K. Mallela, A. Penmatsa, Structure and Gating Dynamics of Na+/Cl– Coupled Neurotransmitter Transporters. *Frontiers in Molecular Biosciences* **6** (2019).

56. C. L. Piscitelli, H. Krishnamurthy, E. Gouaux, Neurotransmitter/sodium symporter orthologue LeuT has a single high-affinity substrate site. *Nature* **468**, 1129-1132 (2010).

57. F. Chen *et al.*, Characterization of an allosteric citalopram-binding site at the serotonin transporter. *Journal of Neurochemistry* **92**, 21-28 (2005).

58. J. A. Coleman, E. M. Green, E. Gouaux, X-ray structures and mechanism of the human serotonin transporter. *Nature* **532**, 334-339 (2016).

59. I. R. Möller *et al.*, Conformational dynamics of the human serotonin transporter during substrate and drug binding. *Nature Communications* **10**, 1687 (2019).

60. K. Gotfryd *et al.*, X-ray structure of LeuT in an inward-facing occluded conformation reveals mechanism of substrate release. *Nature Communications* **11**, 1005 (2020).

61. M. H. Cheng, I. Bahar, Molecular Mechanism of Dopamine Transport by Human Dopamine Transporter. *Structure* **23**, 2171-2181 (2015).

62. M. H. Cheng, I. Bahar, Complete Mapping of Substrate Translocation Highlights the Role of LeuT N-terminal Segment in Regulating Transport Cycle. *PLOS Computational Biology* **10**, e1003879 (2014).

63. J. A. Coleman *et al.*, Serotonin transporter–ibogaine complexes illuminate mechanisms of inhibition and transport. *Nature* **569**, 141-145 (2019).

64. L. Malinauskaite *et al.*, A mechanism for intracellular release of Na+ by neurotransmitter/sodium symporters. *Nature Structural & Molecular Biology* **21**, 1006-1012 (2014).

65. M. V. LeVine *et al.*, The allosteric mechanism of substrate-specific transport in SLC6 is mediated by a volumetric sensor. *Proceedings of the National Academy of Sciences* **116**, 15947 (2019).

66. M. V. LeVine, M. A. Cuendet, A. M. Razavi, G. Khelashvili, H. Weinstein, Thermodynamic Coupling Function Analysis of Allosteric Mechanisms in the Human Dopamine Transporter. *Biophysical Journal* **114**, 10-14 (2018).

67. S. A. Shaikh, E. Tajkhorshid, Modeling and Dynamics of the Inward-Facing State of a Na+/Cl− Dependent Neurotransmitter Transporter Homologue. *PLOS Computational Biology* **6**, e1000905 (2010).

68. A. M. Razavi, G. Khelashvili, H. Weinstein, A Markov State-based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter. *Scientific Reports* **7**, 40076 (2017).

69. G. Khelashvili *et al.*, Spontaneous Inward Opening of the Dopamine Transporter Is Triggered by PIP2-Regulated Dynamics of the N-Terminus. *ACS Chemical Neuroscience* **6**, 1825-1837 (2015).

70. Y.-H. Su, W. B. Frommer, U. Ludewig, Molecular and Functional Characterization of a Family of Amino Acid Transporters from Arabidopsis. *Plant Physiology* **136**, 3104 (2004).

71. A. Vastermark, S. Wollwage, M. E. Houle, R. Rio, M. H. Saier Jr, Expansion of the APC superfamily of secondary carriers. *Proteins: Structure, Function, and Bioinformatics* **82**, 2797-2811 (2014).

72. D. S. Terry *et al.*, A partially-open inward-facing intermediate conformation of LeuT is associated with Na+ release and substrate transport. *Nature Communications* **9**, 230 (2018).

73. J. Li, Z. Zhao, E. Tajkhorshid, Locking Two Rigid-body Bundles in an Outward-Facing Conformation: The Ion-coupling Mechanism in a LeuT-fold Transporter. *Scientific Reports* **9**, 19479 (2019).

74. S. M. Scanlon, D. C. Williams, P. Schloss, Membrane Cholesterol Modulates Serotonin Transporter Activity. *Biochemistry* **40**, 10507-10513 (2001).

75. W. C. Hong, S. G. Amara, Membrane cholesterol modulates the outward facing conformation of the dopamine transporter and alters cocaine binding. *The Journal of biological chemistry* **285**, 32616-32626 (2010).

76. T. Zeppelin, L. K. Ladefoged, S. Sinning, X. Periole, B. Schiøtt, A direct interaction of cholesterol with the dopamine transporter prevents its out-to-inward transition. *PLoS computational biology* **14**, e1005907-e1005907 (2018).

77. A. Schumann-Gillett, M. T. Blyth, M. L. O'Mara, Is protein structure enough? A review of the role of lipids in SLC6 transporter function. *Neuroscience Letters* **700**, 64-69 (2019).

78. L. Laursen *et al.*, Cholesterol binding to a conserved site modulates the conformation, pharmacology, and transport kinetics of the human serotonin transporter. *Journal of Biological Chemistry* **293**, 3510-3523 (2018).

79. C. Fenollar-Ferrer *et al.*, Structure and regulatory interactions of the cytoplasmic terminal domains of serotonin transporter. *Biochemistry* **53**, 5444-5460 (2014).

80. G. Khelashvili *et al.*, Computational modeling of the N-terminus of the human dopamine transporter and its interaction with PIP2 -containing membranes. *Proteins* **83**, 952-969 (2015).

81. P. J. Hamilton *et al.*, PIP2 regulates psychostimulant behaviors through its interaction with a membrane protein. *Nature Chemical Biology* **10**, 582-589 (2014).

82. J. U. Fog *et al.*, Calmodulin Kinase II Interacts with the Dopamine Transporter C Terminus to Regulate Amphetamine-Induced Reverse Transport. *Neuron* **51**, 417-429 (2006).

83. M. W. Quick, "The Role of SNARE Proteins in Trafficking and Function of Neurotransmitter Transporters" in Neurotransmitter Transporters, H. H. Sitte, M. Freissmuth, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006), 10.1007/3-540-29784-7_9, pp. 181-196.

84. R. A. Vaughan, Phosphorylation and Regulation of Psychostimulant-Sensitive Neurotransmitter Transporters. *Journal of Pharmacology and Experimental Therapeutics* **310**, 1 (2004).

85. T. Stockner *et al.*, Mutational analysis of the high-affinity zinc binding site validates a refined human dopamine transporter homology model. *PLoS computational biology* **9**, e1002909-e1002909 (2013).

86. L. Norregaard, D. Frederiksen, E. O. Nielsen, U. Gether, Delineation of an endogenous zinc-binding site in the human dopamine transporter. *The EMBO journal* **17**, 4266-4273 (1998).

87. M. H. Cheng *et al.*, Trimerization of dopamine transporter triggered by AIM-100 binding: Molecular mechanism and effect of mutations. *Neuropharmacology* **161**, 107676 (2019).

88. V. I. Korolchuk, D. C. Rubinsztein, Regulation of autophagy by lysosomal positioning. *Autophagy* **7**, 927-928 (2011).

89. M.-I. Yuseff, P. Pierobon, A. Reversat, A.-M. Lennon-Duménil, How B cells capture, process and present antigens: a crucial role for cell polarity. *Nature Reviews Immunology* **13**, 475-486 (2013).

90. G. G. Gundersen, H. J. Worman, Nuclear positioning. *Cell* **152**, 1376-1389 (2013).

91. C. Barlowe, A. Helenius, Cargo Capture and Bulk Flow in the Early Secretory Pathway. *Annual Review of Cell and Developmental Biology* **32**, 197-222 (2016).

92. S. S. Vembar, J. L. Brodsky, One step at a time: endoplasmic reticulum-associated degradation. *Nature Reviews Molecular Cell Biology* **9**, 944-957 (2008).

93. J. Cancino, J. E. Jung, A. Luini, Regulation of Golgi signaling and trafficking by the KDEL receptor. *Histochemistry and Cell Biology* **140**, 395-405 (2013).

94. G. van Meer, D. R. Voelker, G. W. Feigenson, Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* **9**, 112-124 (2008).

95. M. Capitani, M. Sallese, The KDEL receptor: New functions for an old protein. *FEBS Letters* **583**, 3863-3871 (2009).

96. K. G. Hardwick, M. J. Lewis, J. Semenza, N. Dean, H. R. Pelham, ERD1, a yeast gene required for the retention of luminal endoplasmic reticulum proteins, affects glycoprotein processing in the Golgi apparatus. *The EMBO journal* **9**, 623-630 (1990).

97. H. R. Pelham, Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. *The EMBO journal* **7**, 913-918 (1988).

98. A. J. Crofts *et al.*, Saturation of the endoplasmic reticulum retention machinery reveals anterograde bulk flow. *The Plant cell* **11**, 2233-2248 (1999).

99. I. Raykhel *et al.*, A molecular specificity code for the three mammalian KDEL receptors. *Journal of Cell Biology* **179**, 1193-1204 (2007).

100. S. Munro, H. R. Pelham, A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899-907 (1987).

101. J. Norisada, Y. Hirata, F. Amaya, K. Kiuchi, K. Oh-hashi, A Comparative Analysis of the Molecular Features of MANF and CDNF. *PLoS One* **11**, e0146923 (2016).

102. H. I. Alanen, I. B. Raykhel, M. J. Luukas, K. E. H. Salo, L. W. Ruddock, Beyond KDEL: The Role of Positions 5 and 6 in Determining ER Localization. *Journal of Molecular Biology* **409**, 291-297 (2011).

103. H. R. Pelham, K. G. Hardwick, M. J. Lewis, Sorting of soluble ER proteins in yeast. *The EMBO journal* **7**, 1757-1762 (1988).

104. M. Mei *et al.*, Characterization of aromatic residue–controlled protein retention in the endoplasmic reticulum of Saccharomyces cerevisiae. *Journal of Biological Chemistry* **292**, 20707-20719 (2017).

105. V. Bortnov *et al.*, Myeloid-derived growth factor is a resident endoplasmic reticulum protein. *Journal of Biological Chemistry* **293**, 13166-13175 (2018).

106. A. Tirla, P. Rivera-Fuentes, Peptide Targeting of an Intracellular Receptor of the Secretory Pathway. *Biochemistry* **58**, 1184-1187 (2019).

107. M. A. Gustafson, J. C. Fromme, Regulation of Arf activation occurs via distinct mechanisms at early and late Golgi compartments. *Molecular Biology of the Cell* **28**, 3660-3671 (2017).

108. S. Pasqualato, L. Renault, J. Cherfils, Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front–back' communication. *EMBO reports* **3**, 1035-1041 (2002).

109. Y. Deng, M.-P. Golinelli-Cohen, E. Smirnova, C. L. Jackson, A COPI coat subunit interacts directly with an early-Golgi localized Arf exchange factor. *EMBO reports* **10**, 58-64 (2009).

110. W. Ma, J. Goldberg, Rules for the recognition of dilysine retrieval motifs by coatomer. *The EMBO Journal* **32**, 926-937 (2013).

111. A. Shibuya, N. Margulis, R. Christiano, T. C. Walther, C. Barlowe, The Erv41–Erv46 complex serves as a retrograde receptor to retrieve escaped ER proteins. *Journal of Cell Biology* **208**, 197-209 (2015).

112. P. Xu *et al.*, COPI mediates recycling of an exocytic SNARE by recognition of a ubiquitin sorting signal. *eLife* **6**, e28342 (2017).

113. N. Zerangue *et al.*, Analysis of endoplasmic reticulum trafficking signals by combinatorial screening in mammalian cells. *Proceedings of the National Academy of Sciences* **98**, 2431 (2001).

114. J. Béthune *et al.*, Coatomer, the Coat Protein of COPI Transport Vesicles, Discriminates Endoplasmic Reticulum Residents from p24 Proteins. *Molecular and Cellular Biology* **26**, 8011 (2006).

115. M. Faini *et al.*, The Structures of COPI-Coated Vesicles Reveal Alternate Coatomer Conformations and Interactions. *Science* **336**, 1451 (2012).

116. S. O. Dodonova *et al.*, A structure of the COPI coat and the role of coat proteins in membrane vesicle assembly. *Science* **349**, 195 (2015).

117. S. O. Dodonova *et al.*, 9Å structure of the COPI coat reveals that the Arf1 GTPase occupies two contrasting molecular environments. *eLife* **6**, e26691 (2017).

118. S. Wang *et al.*, Structural characterization of coatomer in its cytosolic state. *Protein & Cell* **7**, 586-600 (2016).

119. R. Beck *et al.*, Membrane curvature induced by Arf1-GTP is essential for vesicle formation. *Proceedings of the National Academy of Sciences* **105**, 11731 (2008).

120. R. Beck *et al.*, Coatomer and dimeric ADP ribosylation factor 1 promote distinct steps in membrane scission. *Journal of Cell Biology* **194**, 765-777 (2011).

121. J. F. Presley *et al.*, Dissection of COPI and Arf1 dynamics in vivo and role in Golgi membrane transport. *Nature* **417**, 187-193 (2002).

122. C. Weimer *et al.*, Differential roles of ArfGAP1, ArfGAP2, and ArfGAP3 in COPI trafficking. *Journal of Cell Biology* **183**, 725-735 (2008).

123. J. Bigay, J.-F. Casella, G. Drin, B. Mesmin, B. Antonny, ArfGAP1 responds to membrane curvature through the folding of a lipid packing sensor motif. *The EMBO Journal* **24**, 2244-2253 (2005).

124. J. Bigay, P. Gounon, S. Robineau, B. Antonny, Lipid packing sensed by ArfGAP1 couples COPI coat disassembly to membrane bilayer curvature. *Nature* **426**, 563-566 (2003).

125. J. Cancino *et al.*, Control Systems of Membrane Transport at the Interface between the Endoplasmic Reticulum and the Golgi. *Developmental Cell* **30**, 280-294 (2014).

126. N. Gomez-Navarro, E. A. Miller, COP-coated vesicles. *Current Biology* **26**, R54-R57 (2016).

127. J. H. Lee, J. Kim, J.-H. Park, W. D. Heo, G. M. Lee, Analysis of Golgi pH in Chinese hamster ovary cells using ratiometric pH-sensitive fluorescent proteins. *Biotechnology and Bioengineering* **116**, 1006-1016 (2019).

128. M. M. Wu *et al.*, Organelle pH studies using targeted avidin and fluorescein–biotin. *Chemistry & Biology* **7**, 197-209 (2000).

129. D. W. Wilson, M. J. Lewis, H. R. Pelham, pH-dependent binding of KDEL to its receptor in vitro. *The Journal of biological chemistry* **268**, 7465-7468 (1993).

130. A. Almanza *et al.*, Endoplasmic reticulum stress signalling – from basic mechanisms to clinical applications. *The FEBS Journal* **286**, 241-278 (2019).

131. V. Bortnov *et al.*, Solution structure of human myeloid-derived growth factor suggests a conserved function in the endoplasmic reticulum. *Nature Communications* **10**, 5612 (2019).

132. J. C. Montesinos, N. Pastor-Cantizano, D. G. Robinson, M. J. Marcote, F. Aniento, Arabidopsis p24δ5 and p24δ9 facilitate Coat Protein I-dependent transport of the K/HDEL receptor ERD2 from the Golgi to the endoplasmic reticulum. *The Plant Journal* **80**, 1014-1030 (2014).

133. P. Brauer *et al.*, Structural basis for pH-dependent retrieval of ER proteins from the Golgi by the KDEL receptor. *Science* **363**, 1103-1107 (2019).

134. B. Becker *et al.*, Cargo binding promotes KDEL receptor clustering at the mammalian cell surface. *Scientific Reports* **6**, 28940 (2016).

135. G. Griffiths *et al.*, Localization of the Lys, Asp, Glu, Leu tetrapeptide receptor to the Golgi complex and the intermediate compartment in mammalian cells. *Journal of Cell Biology* **127**, 1557-1574 (1994).

136. B. L. Tang, S. H. Wong, X. L. Qi, S. H. Low, W. Hong, Molecular cloning, characterization, subcellular localization and dynamics of p23, the mammalian KDEL receptor. *Journal of Cell Biology* **120**, 325-338 (1993).

137. L. Orci *et al.*, Bidirectional Transport by Distinct Populations of COPI-Coated Vesicles. *Cell* **90**, 335-349 (1997).

138. B. Becker *et al.*, H/KDEL receptors mediate host cell intoxication by a viral A/B toxin in yeast. *Scientific Reports* **6**, 31105 (2016).

139. J. L. Hessler, R. J. Kreitman, An Early Step in Pseudomonas Exotoxin Action Is Removal of the Terminal Lysine Residue, Which Allows Binding to the KDEL Receptor. *Biochemistry* **36**, 14577-14582 (1997).

140. S. Seetharam, V. K. Chaudhary, D. FitzGerald, I. Pastan, Increased cytotoxic activity of Pseudomonas exotoxin and two chimeric toxins ending in KDEL. *Journal of Biological Chemistry* **266**, 17376-17381 (1991).

141. M. J. Lewis, H. R. B. Pelham, Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* **68**, 353-364 (1992).

142. I. Majoul *et al.*, KDEL receptor (Erd2p)-mediated retrograde transport of the cholera toxin A subunit from the Golgi involves COPI, p23, and the COOH terminus of Erd2p. *J Cell Biol* **143**, 601-612 (1998).

143. F. A. L. Silva-Alvim *et al.*, Predominant Golgi Residency of the Plant K/HDEL Receptor Is Essential for Its Function in Mediating ER Retention. *The Plant Cell* **30**, 2174 (2018).

144. Y. Ito, T. Uemura, A. Nakano, "Chapter Six - Formation and Maintenance of the Golgi Apparatus in Plant Cells" in International Review of Cell and Molecular Biology*,* K. W. Jeon, Ed. (Academic Press, 2014), vol. 310, pp. 221-287.

145. N. Pastor-Cantizano, J. C. Montesinos, C. Bernat-Silvestre, M. J. Marcote, F. Aniento, p24 family proteins: key players in the regulation of trafficking along the secretory pathway. *Protoplasma* **253**, 967-985 (2016).

146. F. Adolf *et al.*, Proteomic Profiling of Mammalian COPII and COPI Vesicles. *Cell Reports* **26**, 250-265.e255 (2019).

147. N. Pastor-Cantizano, C. Bernat-Silvestre, M. J. Marcote, F. Aniento, Loss of Arabidopsis p24 function affects ERD2 trafficking and Golgi structure, and activates the unfolded protein response. *Journal of Cell Science* **131**, jcs203802 (2018).

148. F. Aniento, M. J. Marcote, D. G. Robinson (Does the Plant K/HDEL Receptor Cycle Between ER and Golgi? (Am Soc Plant Biol).

149. P.-C. Chan, H.-C. Chen, p120RasGAP-Mediated Activation of c-Src Is Critical for Oncogenic Ras to Induce Tumor Invasion. *Cancer Research* **72**, 2405 (2012).

150. F. Bard, L. Mazelin, C. Péchoux-Longin, V. Malhotra, P. Jurdic, Src Regulates Golgi Structure and KDEL Receptor-dependent Retrograde Transport to the Endoplasmic Reticulum. *Journal of Biological Chemistry* **278**, 46601-46606 (2003).

151. G. P. Solis *et al.*, Golgi-Resident Gαo Promotes Protrusive Membrane Dynamics. *Cell* **170**, 939-955.e924 (2017).

152. C. Hetz, F. R. Papa, The Unfolded Protein Response and Cell Fate Control. *Molecular Cell* **69**, 169-181 (2018).

153. M. Wang, R. J. Kaufman, Protein misfolding in the endoplasmic reticulum as a conduit to human disease. *Nature* **529**, 326-335 (2016).

154. M. J. Lewis, H. R. B. Pelham, A human homologue of the yeast HDEL receptor. *Nature* **348**, 162-163 (1990).

155. S. Watanabe, M. Harayama, S. Kanemura, R. Sitia, K. Inaba, Structural basis of pH-dependent client binding by ERp44, a key regulator of protein secretion at the ER–Golgi interface. *Proceedings of the National Academy of Sciences* **114**, E3224 (2017).

156. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, XBP1 mRNA Is Induced by ATF6 and Spliced by IRE1 in Response to ER Stress to Produce a Highly Active Transcription Factor. *Cell* **107**, 881-891 (2001).

157. H. P. Harding, Y. Zhang, D. Ron, Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature* **397**, 271-274 (1999).

158. K. A. Trychta, S. Bäck, M. J. Henderson, B. K. Harvey, KDEL Receptors Are Differentially Regulated to Maintain the ER Proteome under Calcium Deficiency. *Cell Reports* **25**, 1829-1840.e1826 (2018).

159. K. Yamamoto *et al.*, The KDEL Receptor Modulates the Endoplasmic Reticulum Stress Response through Mitogen-activated Protein Kinase Signaling Cascades. *Journal of Biological Chemistry* **278**, 34525-34532 (2003).

160. M. Giannotta *et al.*, The KDEL receptor couples to Gαq/11 to activate Src kinases and regulate transport through the Golgi. *The EMBO journal* **31**, 2869-2881 (2012).

161. S.-Y. Park, J.-S. Yang, A. B. Schmider, R. J. Soberman, V. W. Hsu, Coordinated regulation of bidirectional COPI transport at the Golgi by CDC42. *Nature* **521**, 529-532 (2015).

162. M. Giannotta *et al.*, Prohibitin: A Novel Molecular Player in KDEL Receptor Signalling. *Biomed Res Int* **2015**, 319454-319454 (2015).

163. D. Tapia *et al.*, KDEL receptor regulates secretion by lysosome relocation- and autophagy-dependent modulation of lipid-droplet turnover. *Nature Communications* **10**, 735 (2019).

164. R. P. Shiu, J. Pouyssegur, I. Pastan, Glucose depletion accounts for the induction of two transformation-sensitive membrane proteinsin Rous sarcoma virus-transformed chick embryo fibroblasts. *Proc Natl Acad Sci U S A* **74**, 3840-3844 (1977).

165. B. Luo, A. S. Lee, The critical roles of endoplasmic reticulum chaperones and unfolded protein response in tumorigenesis and anticancer therapies. *Oncogene* **32**, 805-818 (2013).

166. V. R. Wiersma, M. Michalak, T. M. Abdullah, E. Bremer, P. Eggleton, Mechanisms of Translocation of ER Chaperones to the Cell Surface and Immunomodulatory Roles in Cancer and Autoimmunity. *Front Oncol* **5**, 7-7 (2015).

167. J. Y. Chung, S. W. Lim, Y. J. Hong, S. O. Hwang, G. M. Lee, Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant chinese hamster ovary cells. *Biotechnology and Bioengineering* **85**, 539-546 (2004).

168. C. Mohan, S. H. Park, J. Y. Chung, G. M. Lee, Effect of doxycycline-regulated protein disulfide isomerase expression on the specific productivity of recombinant CHO cells: Thrombopoietin and antibody. *Biotechnology and Bioengineering* **98**, 611-615 (2007).

169. N. Borth, D. Mattanovich, R. Kunert, H. Katinger, Effect of Increased Expression of Protein Disulfide Isomerase and Heavy Chain Binding Protein on Antibody Secretion in a Recombinant CHO Cell Line. *Biotechnology Progress* **21**, 106-111 (2005).

170. S. Xiao, J. Shiloach, M. J. Betenbaugh, Engineering cells to improve protein expression. *Current Opinion in Structural Biology* **26**, 32-38 (2014).

171. A. Samy, K. Kaneyoshi, T. Omasa, Improvement of Intracellular Traffic System by Overexpression of KDEL Receptor 1 in Antibody-Producing CHO Cells. *Biotechnology Journal* **n/a**, 1900352 (2020).

172. W. Zhang, X. Wang, K. Yue, S. Liu, X. Liu, Construction of a CXCL12-KDEL fusion gene to inhibit head and neck squamous cell carcinoma metastasis by intracellular sequestration of CXCR4. *Biomed Res Int* **2015**, 195828-195828 (2015).

173. H.-Y. Chen, E. S. Clayman, W.-F. Ma, Phenotypic Knockout of CXCR4 Expression by a Novel Intrakine Mutant hSDF-1α/54/KDEL Inhibits Breast Cancer Metastasis. *J Interferon Cytokine Res* **35**, 771-778 (2015).

174. J. Zhang *et al.*, Cystinosin, the small GTPase Rab11, and the Rab7 effector RILP regulate intracellular trafficking of the chaperone-mediated autophagy receptor LAMP2A. *Journal of Biological Chemistry* **292**, 10328-10346 (2017).

175. T. Yamamoto, K. Fujimura-Kamada, E. Shioji, R. Suzuki, K. Tanaka, Cfs1p, a Novel Membrane Protein in the PQ-Loop Family, Is Involved in Phospholipid Flippase Functions in Yeast. *G3: Genes|Genomes|Genetics* **7**, 179 (2017).

176. V. Saudek, Cystinosin, MPDU1, SWEETs and KDELR belong to a well-defined protein family with putative function of cargo receptors involved in vesicle trafficking. *PLoS One* **7**, e30876-e30876 (2012).

177. D. C. Yee *et al.*, The transporter–opsin–G protein-coupled receptor (TOG) superfamily. *The FEBS Journal* **280**, 5780-5800 (2013).

178. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **111**, 7812-7824 (2007).

179. X. Periole, M. Cavalli, S.-J. Marrink, M. A. Ceruso, Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of Chemical Theory and Computation* **5**, 2531-2543 (2009).

180. S. C. Harvey, R. K. Z. Tan, T. E. Cheatham Iii, The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *Journal of Computational Chemistry* **19**, 726-740 (1998).

181. M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, P. C. Biggin, Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *Journal of the American Chemical Society* **139**, 946-957 (2017).

182. S. Boresch, F. Tettinger, M. Leitgeb, M. Karplus, Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B* **107**, 9535-9551 (2003).

183. M. A. Kastenholz, P. H. Hünenberger, Influence of Artificial Periodicity and Ionic Strength in Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum Methods. *The Journal of Physical Chemistry B* **108**, 774-788 (2004).

184. M. A. Kastenholz, P. H. Hünenberger, Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids. *The Journal of Chemical Physics* **124**, 124106 (2006).

185. M. A. Kastenholz, P. H. Hünenberger, Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *The Journal of Chemical Physics* **124**, 224501 (2006).

186. M. M. Reif, V. Kräutler, M. A. Kastenholz, X. Daura, P. H. Hünenberger, Molecular Dynamics Simulations of a Reversibly Folding β-Heptapeptide in Methanol: Influence of the Treatment of Long-Range Electrostatic Interactions. *The Journal of Physical Chemistry B* **113**, 3112-3128 (2009).

187. G. J. Rocklin, D. L. Mobley, K. A. Dill, P. H. Hünenberger, Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *The Journal of chemical physics* **139**, 184103-184103 (2013).

188. J. Yin, N. M. Henriksen, D. R. Slochower, M. K. Gilson, The SAMPL5 host–guest challenge: computing binding free energies and enthalpies from explicit solvent simulations by the attach-pull-release (APR) method. *Journal of Computer-Aided Molecular Design* **31**, 133-145 (2017).

189. A. J. Bridgeman, G. Cavigliasso, L. R. Ireland, J. Rothery, The Mayer bond order as a tool in inorganic chemistry*Journal of the Chemical Society, Dalton Transactions* 10.1039/B102094N, 2095-2108 (2001).

190. M. A. Spackman, The Quantum Theory of Atoms in Molecules. From Solid State to DNA and Drug Design. Edited by Chérif F. Matta and Russell J. Boyd. *Angewandte Chemie International Edition* **46**, 6766-6767 (2007).

191. E. Espinosa, E. Molins, C. Lecomte, Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chemical Physics Letters* **285**, 170-173 (1998).

192. F. Fuster, B. Silvi, Does the topological approach characterize the hydrogen bond? *Theoretical Chemistry Accounts* **104**, 13-21 (2000).

193. J. G. Brandenburg, C. Bannwarth, A. Hansen, S. Grimme, B97-3c: A revised low-cost variant of the B97-D density functional method. *The Journal of Chemical Physics* **148**, 064104 (2018).

194. S. Grimme, Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **27**, 1787-1799 (2006).

195. S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **32**, 1456-1465 (2011).

196. R. Sure, S. Grimme, Corrected small basis set Hartree-Fock method for large systems. *Journal of Computational Chemistry* **34**, 1672-1685 (2013).

197. R. L. Wolfson, D. M. Sabatini, The Dawn of the Age of Amino Acid Sensors for the mTORC1 Pathway. *Cell Metabolism* **26**, 301-309 (2017).

198. S. Heublein *et al.*, Proton-assisted amino-acid transporters are conserved regulators of proliferation and amino-acid-dependent mTORC1 activation. *Oncogene* **29**, 4068-4079 (2010).

199. H. Krishnamurthy, E. Gouaux, X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* **481**, 469 (2012).

200. E. I. Closs, J. P. Boissel, A. Habermeier, A. Rotman, Structure and function of cationic amino acid transporters (CATs). *J Membr Biol* **213** (2006).

201. P. L. Shaffer, A. Goehring, A. Shankaranarayanan, E. Gouaux, Structure and mechanism of a Na+independent amino acid transporter. *Science (New York, N.Y.)* **325**, 1010-1014 (2009).

202. T. Beuming, L. Shi, J. A. Javitch, H. Weinstein, A Comprehensive Structure-Based Alignment of Prokaryotic and Eukaryotic Neurotransmitter/Na⁺ Symporters (NSS) Aids in the Use of the LeuT Structure to Probe NSS Structure and Function. *Molecular Pharmacology* **70**, 1630 (2006).

203. N. Eswar *et al.*, Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics* **15**, 5.6.1-5.6.30 (2006).

204. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46**, W296-W303 (2018).

205. J. P. M. Jämbeck, A. P. Lyubartsev, An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *Journal of Chemical Theory and Computation* **8**, 2938-2948 (2012).

206. G. Shahane, W. Ding, M. Palaiokostas, M. Orsi, Physical properties of model biological lipid bilayers: insights from all-atom molecular dynamics simulations. *Journal of Molecular Modeling* **25**, 76 (2019).

207. T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, D. P. Tieleman, Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation* **10**, 676-690 (2014).

208. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).

209. K. Lindorff-Larsen *et al.*, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958 (2010).

210. A. H. C. Horn, A consistent force field parameter set for zwitterionic amino acid residues. *Journal of Molecular Modeling* **20**, 2478 (2014).

211. P. Dobrev, S. Donnini, G. Groenhof, H. Grubmuller, Accurate Three States Model for Amino Acids with Two Chemically Coupled Titrating Sites in Explicit Solvent Atomistic Constant pH Simulations and pK(a) Calculations. *J Chem Theory Comput* **13**, 147-160 (2017).

212. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).

213. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604-613 (2014).

214. A. Grossfield (2003) WHAM: the weighted histogram analysis method.

215. P. V. Klimovich, M. R. Shirts, D. L. Mobley, Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **29**, 397-411 (2015).

216. D. Seeliger, B. L. de Groot, Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical Journal* **98**, 2309-2316 (2010).

217. G. J. Rocklin, D. L. Mobley, K. A. Dill, P. H. Hünenberger, Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *The Journal of Chemical Physics* **139**, 184103 (2013).

218. M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* **7**, 525-537 (2011).

219. A. Kyrychenko *et al.*, Refining Protein Penetration into the Lipid Bilayer Using Fluorescence Quenching and Molecular Dynamics Simulations: The Case of Diphtheria Toxin Translocation Domain. *J Membr Biol* **251**, 379-391 (2018).

220. Y. Matsunaga *et al.*, Energetics and conformational pathways of functional rotation in the multidrug transporter AcrB. *eLife* **7**, e31715 (2018).

221. T. O. Wambo, L. Y. Chen, C. Phelix, G. Perry, Affinity and path of binding xylopyranose unto E. coli xylose permease. *Biochemical and Biophysical Research Communications* **494**, 202-206 (2017).

222. S. A. Shaikh *et al.*, Visualizing Functional Motions of Membrane Transporters with Molecular Dynamics Simulations. *Biochemistry* **52**, 569-587 (2013).

223. M. Gur, E. Zomot, M. H. Cheng, I. Bahar, Energy landscape of LeuT from molecular simulations. *The Journal of Chemical Physics* **143**, 243134 (2015).

224. K. Immadisetty, J. Hettige, M. Moradi, What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT? *The Journal of Physical Chemistry B* **121**, 3644-3656 (2017).

225. N. R. Latorraca *et al.*, Mechanism of Substrate Translocation in an Alternating Access Transporter. *Cell* **169**, 96-107.e112 (2017).

226. T. Galochkina, M. Ng Fuk Chong, L. Challali, S. Abbar, C. Etchebest, New insights into GluT1 mechanics during glucose transfer. *Scientific Reports* **9**, 998 (2019).

227. B. Selvam, S. Mittal, D. Shukla, Free Energy Landscape of the Complete Transport Cycle in a Key Bacterial Transporter. *ACS Central Science* **4**, 1146-1154 (2018).

228. T. Lazaridis, G. Hummer, Classical Molecular Dynamics with Mobile Protons. *Journal of Chemical Information and Modeling* **57**, 2833-2845 (2017).

229. H. Koldsø, H. E. Autzen, J. Grouleff, B. Schiøtt, Ligand Induced Conformational Changes of the Human Serotonin Transporter Revealed by Molecular Dynamics Simulations. *PLOS ONE* **8**, e63635 (2013).

230. G. Wisedchaisri, M.-S. Park, M. G. Iadanza, H. Zheng, T. Gonen, Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nature Communications* **5**, 4521 (2014).

231. C. Zhao, S. Y. Noskov, The Role of Local Hydration and Hydrogen-Bonding Dynamics in Ion and Solute Release from Ion-Coupled Secondary Transporters. *Biochemistry* **50**, 1848-1856 (2011).

232. Y. Huang, W. Chen, D. L. Dotson, O. Beckstein, J. Shen, Mechanism of pH-dependent activation of the sodium-proton antiporter NhaA. *Nature Communications* **7**, 12940 (2016).

233. L. Feng, W. B. Frommer, Evolution of Transporters: The Relationship of SWEETs, PQ-loop, and PnuC Transporters. *Trends Biochem Sci* **41**, 118-119 (2016).

234. M. J. Lewis, H. R. Pelham, Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* **68**, 353-364 (1992).

235. P. A. Frey, S. A. Whitt, J. B. Tobin, A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **264**, 1927 (1994).

236. M. A. McAllister, Characterization of low-barrier hydrogen bonds. 3. Hydrogen maleate. An ab initio and DFT investigation. *Canadian Journal of Chemistry* **75**, 1195-1202 (1997).

237. M. M. Kreevoy, T. M. Liang, Structures and isotopic fractionation factors of complexes, A1HA2. *Journal of the American Chemical Society* **102**, 3315-3322 (1980).

238. S. Genheden, U. Ryde, Improving the Efficiency of Protein–Ligand Binding Free-Energy Calculations by System Truncation. *Journal of Chemical Theory and Computation* **8**, 1449-1458 (2012).

239. J. A. Maier *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696-3713 (2015).

240. I. R. Gould, Skjevik A.A., Dickson, C.J., Madej, B.D., Walker, R.C., Lipid17: A Comprehensive AMBER Force Field for the Simulation of Zwitterionic and Anionic Lipids. *in prep* (2018).

241. I. Y. B.-S. D.A. Case, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeG, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J.

Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018. *University of California, San Francisco.* (2018).

242. F. Neese, Software update: the ORCA program system, version 4.0. *WIREs Computational Molecular Science* **8**, e1327 (2018).

243. A. Grossfield, WHAM: the weighted histogram analysis method version 2.0.9.1.

244. F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **7**, 3297-3305 (2005).

245. E. Caldeweyher *et al.*, A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics* **150**, 154122 (2019).

246. F. Fuster, B. Silvi, Does the topological approach characterize the hydrogen bond? Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) **104**, 13-21 (2000).

247. T. Lu, F. Chen, Multiwfn: a multifunctional wavefunction analyzer. *J Comput Chem* **33**, 580-592 (2012).

248. R. A. Kendall, T. H. Dunning, R. J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **96**, 6796-6806 (1992).

249. F. Weigend, A. Köhn, C. Hättig, Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *The Journal of Chemical Physics* **116**, 3175-3183 (2002).

250. S. Grimme, Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chemistry – A European Journal* **18**, 9955-9964 (2012).

251. I. M. Alecu, J. Zheng, Y. Zhao, D. G. Truhlar, Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from

Electronic Model Chemistries. *Journal of Chemical Theory and Computation* **6**, 2872-2887 (2010).

252. T. Lu, Q. Chen, Shermo: A General Code for Calculating Molecular Thermochemistry Properties (2020), 10.26434/chemrxiv.12278801.

253. L. A. Burns, M. S. Marshall, C. D. Sherrill, Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *Journal of Chemical Theory and Computation* **10**, 49-57 (2014).

254. L. Goerigk, S. Grimme, Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **7**, 291-309 (2011).

255. V. Gapsys, S. Michielssens, D. Seeliger, B. L. de Groot, pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry* **36**, 348-354 (2015).

256. M. Aldeghi, B. L. de Groot, V. Gapsys, "Accurate Calculation of Free Energy Changes upon Amino Acid Mutation" in Computational Methods in Protein Evolution*,* T. Sikosek, Ed. (Springer New York, New York, NY, 2019), 10.1007/978-1-4939-8736-8_2, pp. 19-47.

257. M. Aldeghi *et al.*, Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Communications Chemistry* **1**, 19 (2018).

258. R. F. W. Bader, A quantum theory of molecular structure and its applications. *Chemical Reviews* **91**, 893-928 (1991).

259. C. Cuautli, J. Ireta, Theoretical investigations on the layer-anion interaction in Mg–Al layered double hydroxides: Influence of the anion nature and layer composition. *The Journal of Chemical Physics* **142**, 094704 (2015).

260. P. L. A. Popelier, *Atoms in Molecules. An Introduction* (Pearson Education, Harlow, Great Britain, 2000).

261. D. Cremer, E. Kraka, Chemical Bonds without Bonding Electron Density — Does the Difference Electron-Density Analysis Suffice for a Description of the Chemical Bond? *Angewandte Chemie International Edition in English* **23**, 627-628 (1984).

262. E. Espinosa, I. Alkorta, J. Elguero, E. Molins, From weak to strong interactions: A comprehensive analysis of the topological and energetic properties of the electron density distribution involving X–H···F–Y systems. *The Journal of Chemical Physics* **117**, 5529-5542 (2002).

263. I. Rozas, I. Alkorta, J. Elguero, Behavior of Ylides Containing N, O, and C Atoms as Hydrogen Bond Acceptors. *Journal of the American Chemical Society* **122**, 11154-11161 (2000).

264. W. W. Cleland, M. M. Kreevoy, Low-barrier hydrogen bonds and enzymic catalysis. *Science* **264**, 1887 (1994).

265. J. D. Graham, A. M. Buytendyk, D. Wang, K. H. Bowen, K. D. Collins, Strong, Low-Barrier Hydrogen Bonds May Be Available to Enzymes. *Biochemistry* **53**, 344-349 (2014).

266. J. P. Klinman, Low Barrier Hydrogen Bonds: Getting Close, but Not Sharing. *ACS Cent Sci* **1**, 115-116 (2015).

267. S. Zhou, L. Wang, Unraveling the structural and chemical features of biological short hydrogen bonds. *Chemical Science* **10**, 7734-7745 (2019).

268. H. W. Qi, H. J. Kulik, Evaluating Unexpectedly Short Non-covalent Distances in X-ray Crystal Structures of Proteins with Electronic Structure Analysis. *Journal of Chemical Information and Modeling* **59**, 2199-2211 (2019).

269. H. R. B. Pelham, Recycling of proteins between the endoplasmic reticulum and Golgi complex. *Current Opinion in Cell Biology* **3**, 585-591 (1991).

270. A. Schweizer, M. Ericsson, T. Bachi, G. Griffiths, H. P. Hauri, Characterization of a novel 63 kDa membrane protein. Implications for the organization of the ER-to-Golgi pathway. *Journal of Cell Science* **104**, 671 (1993).

271. W. Ma, E. Goldberg, J. Goldberg, ER retention is imposed by COPII protein sorting and attenuated by 4-phenylbutyrate. *eLife* **6**, e26624 (2017).

272. C. Appenzeller, H. Andersson, F. Kappeler, H.-P. Hauri, The lectin ERGIC-53 is a cargo transport receptor for glycoproteins. *Nature Cell Biology* **1**, 330-334 (1999).

273. M. Moussalli *et al.*, Mannose-dependent Endoplasmic Reticulum (ER)-Golgi Intermediate Compartment-53-mediated ER to Golgi Trafficking of Coagulation Factors V and VIII. *Journal of Biological Chemistry* **274**, 32539-32542 (1999).

274. H. Andersson, F. Kappeler, H.-P. Hauri, Protein Targeting to Endoplasmic Reticulum by Dilysine Signals Involves Direct Retention in Addition to Retrieval. *Journal of Biological Chemistry* **274**, 15080-15084 (1999).

275. C. Nie, H. Wang, R. Wang, D. Ginsburg, X.-W. Chen, Dimeric sorting code for concentrative cargo selection by the COPII coat. *Proceedings of the National Academy of Sciences* **115**, E3155 (2018).

276. E. P. A. Neve, U. Lahtinen, R. F. Pettersson, Oligomerization and Interacellular Localization of the Glycoprotein Receptor ERGIC-53 is Independent of Disulfide Bonds. *Journal of Molecular Biology* **354**, 556-568 (2005).

277. C. Zheng *et al.*, Structural Characterization of Carbohydrate Binding by LMAN1 Protein Provides New Insight into the Endoplasmic Reticulum Export of Factors V (FV) and VIII (FVIII). *Journal of Biological Chemistry* **288**, 20499-20509 (2013).

278. C. D. Rios, B. A. Jordan, I. Gomes, L. A. Devi, G-protein-coupled receptor dimerization: modulation of receptor function. *Pharmacology & Therapeutics* **92**, 71-87 (2001).

279. S. Angers *et al.*, Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET). *Proc Natl Acad Sci U S A* **97**, 3684-3689 (2000).

280. Y. Tao *et al.*, Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature* **527**, 259-263 (2015).

281. G. Nelson *et al.*, Mammalian Sweet Taste Receptors. *Cell* **106**, 381-390 (2001).

282. F. M. Townsley, D. W. Wilson, H. R. Pelham, Mutational analysis of the human KDEL receptor: distinct structural requirements for Golgi retention, ligand binding and retrograde transport. *The EMBO Journal* **12**, 2821-2829 (1993).

283. K. Salim *et al.*, Oligomerization of G-protein-coupled Receptors Shown by Selective Co-immunoprecipitation. *Journal of Biological Chemistry* **277**, 15482-15485 (2002).

284. T.-H. Lan, Q. Liu, C. Li, G. Wu, N. A. Lambert, Sensitive and high resolution localization and tracking of membrane proteins in live cells with BRET. *Traffic* **13**, 1450-1456 (2012).

285. N. A. Lambert, J. A. Javitch, CrossTalk opposing view: Weighing the evidence for class A GPCR dimers, the jury is still out. *J Physiol* **592**, 2443-2445 (2014).

286. L. Albizu *et al.*, Time-resolved FRET between GPCR ligands reveals oligomers in native tissues. *Nature Chemical Biology* **6**, 587-594 (2010).

287. J. Taura, V. Fernández-Dueñas, F. Ciruela, Visualizing G Protein-Coupled Receptor-Receptor Interactions in Brain Using Proximity Ligation In Situ Assay. *Current Protocols in Cell Biology* **67**, 17.17.11-17.17.16 (2015).

288. N. Basdevant, D. Borgis, T. Ha-Duong, A Coarse-Grained Protein−Protein Potential Derived from an All-Atom Force Field. *The Journal of Physical Chemistry B* **111**, 9390-9399 (2007).

289. M. Chavent, A. P. Chetwynd, P. J. Stansfeld, M. S. P. Sansom, Dimerization of the EphA1 receptor tyrosine kinase transmembrane domain: Insights into the mechanism of receptor activation. *Biochemistry* **53**, 6641-6652 (2014).

290. A. B. Pawar, D. Sengupta, Resolving the conformational dynamics of ErbB growth factor receptor dimers. *Journal of Structural Biology* **207**, 225-233 (2019).

291. N. Flinner, E. Schleiff, Dynamics of the Glycophorin A Dimer in Membranes of Native-Like Composition Uncovered by Coarse-Grained Molecular Dynamics Simulations. *PLoS One* **10**, e0133999-e0133999 (2015).

292. A. B. Pawar *et al.*, Thermodynamic and kinetic characterization of transmembrane helix association. *Physical Chemistry Chemical Physics* **17**, 1390-1398 (2015).

293. X. Prasanna, D. Sengupta, A. Chattopadhyay, Cholesterol-dependent Conformational Plasticity in GPCR Dimers. *Scientific reports* **6**, 31858-31858 (2016).

294. X. Qiao, J. Jeon, J. Weber, F. Zhu, B. Chen, Construction of a novel coarse grain model for simulations of HIV capsid assembly to capture the backbone structure and inter-domain motions in solution. *Data Brief* **5**, 506-512 (2015).

295. B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5.6.1-5.6.37 (2016).

296. M.-Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507-2524 (2006).

297. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38 (1996).

298. F. A. Herzog, L. Braun, I. Schoen, V. Vogel, Improved Side Chain Dynamics in MARTINI Simulations of Protein–Lipid Interfaces. *Journal of Chemical Theory and Computation* **12**, 2446-2458 (2016).

299. D. H. de Jong, S. Baoukina, H. I. Ingólfsson, S. J. Marrink, Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* **199**, 1-7 (2016).

300. L. Verlet, Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-103 (1967).

301. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).

302. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690 (1984).

303. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190 (1981).

304. R. Gowers *et al.* (2016) MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in *Proceedings of the 15th Python in Science Conference*, pp 98-105.

305. N. M. Henriksen, A. T. Fenley, M. K. Gilson, Computational Calorimetry: High-Precision Calculation of Host–Guest Binding Thermodynamics. *Journal of Chemical Theory and Computation* **11**, 4377-4394 (2015).

306. T. A. Wassenaar *et al.*, High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *Journal of Chemical Theory and Computation* **11**, 2278-2291 (2015).

307. I. Majoul, M. Straub, S. W. Hell, R. Duden, H.-D. Soeling, KDEL-Cargo Regulates Interactions between Proteins Involved in COPI Vesicle Traffic: Measurements in Living Cells Using FRET. *Developmental Cell* **1**, 139-153 (2001).

308. T. Aoe *et al.*, The KDEL receptor, ERD2, regulates intracellular traffic by recruiting a GTPase-activating protein for ARF1. *The EMBO journal* **16**, 7305-7316 (1997).

309. T. Aoe, A. J. Lee, E. van Donselaar, P. J. Peters, V. W. Hsu, Modulation of intracellular transport by transported proteins: insight from regulation of COPI-mediated transport. *Proc Natl Acad Sci U S A* **95**, 1624-1629 (1998).

310. P. Robustelli, S. Piana, D. E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115**, E4758 (2018).

311. R. Alessandri et al., Pitfalls of the Martini Model. Journal of Chemical Theory and Computation **15**, 5448-5460 (2019).

312. W. Chen *et al.*, Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *Journal of Chemical Theory and Computation* **14**, 6346-6358 (2018).

313. A. Zhou, M. Schauperl, P. S. Nerenberg, Benchmarking Electronic Structure Methods for Accurate Fixed-Charge Electrostatic Models. *Journal of Chemical Information and Modeling* **60**, 249-258 (2020).

314. J. L. Parker *et al.*, Proton movement and coupling in the POT family of peptide transporters. *Proceedings of the National Academy of Sciences* **114**, 13182 (2017).

315. Y. S. Bykov *et al.*, The structure of the COPI coat determined within the cell. *eLife* **6**, e32493 (2017).

316. J. H. Lorent *et al.*, Structural determinants and functional consequences of protein affinity for membrane rafts. *Nature Communications* **8**, 1219 (2017).

317. I. Levental, K. R. Levental, F. A. Heberle, Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends in Cell Biology* **30**, 341-353 (2020).

318. A. L. Duncan, R. A. Corey, M. S. P. Sansom, Defining how multiple lipid species interact with inward rectifier potassium (Kir2) channels. *Proceedings of the National Academy of Sciences* **117**, 7803 (2020).

319. I. Vorobyov, T. W. Allen, On the role of anionic lipids in charged protein interactions with membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 1673-1683 (2011).

320. H. M. Khan *et al.*, Capturing Choline–Aromatics Cation–π Interactions in the MARTINI Force Field. *Journal of Chemical Theory and Computation* **16**, 2550-2560 (2020).

321. M. L. Rook, M. Musgaard, D. M. MacLean, Coupling structure with function in acid-sensing ion channels: challenges in pursuit of proton sensors. *J Physiol* **n/a** (2020).

322. B. Zhang *et al.*, Structure of a proton-dependent lipid transporter involved in lipoteichoic acids biosynthesis. *Nature Structural & Molecular Biology* **27**, 561-569 (2020).

323. G. S. Minhas, S. Newstead, Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters. *Proceedings of the National Academy of Sciences* **116**, 804 (2019).

324. J. L. Parker, J. A. Mindell, S. Newstead, Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *eLife* **3**, e04273 (2014).

325. Z. Wu, S. Newstead, P. C. Biggin, The KDEL trafficking receptor exploits pH to tune the strength of an unusual short hydrogen bond. *Scientific Reports* **10**, 16903 (2020).