

ABSTRACT

Title of Document: COMPUTATIONAL STUDIES ON LACTOSE PERMEASE OF *E. COLI* AS A MODEL FOR MEMBRANE TRANSPORT PROTEINS

Pushkar Yashvant Pendse, Doctor of Philosophy, 2013

Directed By: Jeffery B Klauda, Assistant Professor, Department of Chemical and Biomolecular Engineering

Membrane transport proteins actively transport ions, metabolites, drug molecules, and others across the amphiphilic cell membrane. The Major Facilitator Superfamily (MFS) is an important class of membrane transporters whose members are found in almost all types of organisms. The MFS proteins transport a diverse range of molecules including sugars, peptides, and drug molecules. In this work, lactose permease (LacY) of *E. coli*, which transports galactosides across the plasma membrane by the proton symport mechanism, is studied as a model for the MFS proteins. First, a hybrid two-step simulation approach was developed to determine the unknown periplasmic-open structure of LacY starting with the cytoplasmic-open x-ray crystal structure. A periplasmic-open model for LacY that agrees with several indirect experimental measurements was obtained. Sugar binding and protonation of Glu269 were found to be the triggers for LacY's structural change from the

cytoplasmic- to the periplasmic-open state. Mutations in residues Asn245, Ser41, Glu374, Lys42 and Gln242 to prevent cross-domain hydrogen bonding were proposed that might aid in crystallizing LacY in the periplasmic-open state. The second focus of this dissertation was a comprehensive study on the binding of high- and low-affinity binders as well as non-binders to LacY in its cytoplasmic- and periplasmic-open states. A possible pathway for the substrate translocation, which involves aromatic stacking interactions of substrates with Phe354 and Tyr350, was suggested. Binding free energy values calculated using the alchemical free energy perturbation method agree with the experimental data. The differences in binding affinities result from dissimilarities in the binding structures of different sugars. The binding free energy values as well as the binding structures in the cytoplasmic- and the periplasmic-open states of LacY provide a quantitative proof of the alternating access mechanism in LacY. Lastly, hybrid quantum mechanics/molecular mechanics (QM/MM) studies on LacY were performed, which are the first QM/MM studies of proton transport in an MFS protein. These QM/MM studies suggest an important role of water molecules in proton transfer. A hydronium ion intermediate was observed during proton transfer from Glu325 to His322, which is consistent with the experimental hypothesis. The transfer of proton from His322 to Glu269 was found to be the rate-determining reaction. More extensive QM/MM calculations on LacY are proposed to fully probe proton translocation in this protein. This dissertation has touched all the important aspects of LacY's transport cycle and results from this study can be beneficial for understanding the mechanism of other MFS proteins such as the drug efflux protein EmrD, which shares its structural and functional motif with LacY.

COMPUTATIONAL STUDIES ON LACTOSE PERMEASE OF *E.COLI* AS A
MODEL FOR MEMBRANE TRANSPORT PROTEINS.

By

Pushkar Yashvant Pendse

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctorate of Philosophy
2013

Advisory Committee:

Jeffery B Klauda, Assistant Professor, Chemical and Biomolecular
Engineering, Chair

Srinivasa Raghavan, Professor, Chemical and Biomolecular
Engineering

Ganesh Sriram, Assistant Professor, Chemical and Biomolecular
Engineering

Santiago Solares, Associate Professor, Mechanical Engineering

Panagiotis Dimitrakopoulos, Associate Professor, Chemical and
Biomolecular Engineering

© Copyright of Chapters 1, 2, 4, 5, and 6 by
Pushkar Yashvant Pendse
2013

© Copyright of Chapter 3 by
Pushkar Yashvant Pendse, Bernard R. Brooks
Jeffery Klauda

2010

Dedication

I dedicate my Ph.D. dissertation to my soulmate Shilpa

I dedicate this work to my parents for their incredible support, love and
encouragement

Acknowledgements

With utmost gratitude, I thank Dr. Jeffrey Klauda for his guidance and support throughout my Ph.D. Under his mentorship I received critical and timely guidance and also developed qualities of an independent thinker. For this, I sincerely thank Dr. Klauda.

I also thank Brent Rogaski and Kunal Pandit for their support in the lab and frequent value adding discussion. Special thanks to Kunal for providing me with the *E. coli* membrane coordinates.

I also thank Vladimir Kasho and Ron Kaback for sharing with us raw experimental DEER data and a model for LacY.

Special thanks to the CHARMM community for solving my queries about CHARMM directly or indirectly from time to time.

I also thank HPCC at University of Maryland for helping me out with several compilations and installations.

I would like to thank the Department of Chemical and Biomolecular Engineering at University of Maryland and my doctoral committee members for their thoughtful insights and discussions on my research work.

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
Chapter 1: Introduction.....	1
1.1. Role of membrane transport systems in biology.....	1
1.2. Classification of membrane transport proteins.....	2
1.3. Major Facilitator Superfamily.....	3
1.3.1. Multidrug efflux proteins in MFS.....	4
1.3.2. Structural and functional similarities in MFS proteins.....	4
1.4. Lactose permease: paradigm for MFS proteins.....	5
1.4.1. Structure of LacY.....	6
1.4.2. Alternating access mechanism in LacY.....	7
1.4.3. Sugar binding in LacY.....	8
1.4.4. Proton translocation in LacY.....	9
Chapter 2: Computational Methods.....	11
2.1. Molecular Dynamics Simulations.....	11
2.1.1. Central idea.....	11
2.1.2. Potential energy and force calculations.....	13
2.1.3. Integration schemes.....	14
2.1.4. Periodic boundary conditions.....	14
2.1.5. Ensembles and control of system variables.....	16
2.1.6. Molecular simulation programs.....	17
2.2. Binding Free Energy calculations.....	19
2.2.1. Ligand-protein association.....	20
2.2.2. The double decoupling scheme.....	22
2.2.3. Contributions from multiple protein conformations.....	23
2.2.4. Orientational decomposition.....	24
2.2.5. Use of restraints and constraints.....	25
2.2.6. Reduced atom boundary potentials.....	28
2.3. Hybrid QM/MM simulations.....	29
2.3.1. Central idea behind QM/MM.....	29
2.3.2. The QM region.....	30
2.3.3. The MM region.....	31
2.3.4. Treatment of boundary between QM and MM regions.....	32
Chapter 3: Study of conformational changes in LacY.....	33
3.1. Background.....	33
3.2. Computational design and methods.....	35
3.2.1. Two-step hybrid simulation approach: IM-EX.....	35
3.2.2. Simulation setup.....	36
3.2.3. Analysis methods.....	39
3.3. Results.....	41
3.3.1. Conformational changes in LacY.....	41

3.3.2. Description of the transitions to the Outward-facing State of LacY	49
3.3.3. Comparison with experimental data	52
3.3.4. Comparison with other computational models of periplasmic-open LacY	59
3.3.5. Sugar binding	59
3.4. Discussion	61
3.5. Insights into LacY's transport mechanism	66
Chapter 4: Thermodynamics of sugar binding to LacY	69
4.1. Introduction	69
4.2. Computational design and Methods	74
4.2.1. Binding structure determination	74
4.2.2. Absolute binding free energy calculations	77
4.3. Results	80
4.3.1. Sugar binding conformations	80
4.3.2. Quantification of binding affinities	88
4.4. Discussion	92
Chapter 5: Proton translocation pathways in LacY	99
5.1. Introduction	99
5.2. Computational design and methods	102
5.2.1. MD simulations	102
5.2.2. The QM/MM simulations	103
5.3. Results	104
5.3.1. Equilibrium MD simulations	104
5.3.2. Proton translocation from Glu325 to His322	106
5.3.3. Proton translocation from His322 to Glu269	108
5.3.4. Proton translocation from His322 to Glu325	109
5.4. Discussion	111
Chapter 6: Conclusions and Future Directions	114
Future directions	117
References:	120

Chapter 1: Introduction

This dissertation describes computational studies on the transport mechanism of lactose permease (LacY) of *Escherichia coli*. LacY is a membrane protein that carries out stoichiometric transport of protons and different sugar molecules across the plasma membrane of *E. coli*. In the introduction chapter, I will begin with a brief discussion on the importance of membrane transport proteins in biological systems and their classification based on energetics and mode of transport, before focusing on the energetics, structure and function of LacY.

1.1. Role of membrane transport systems in biology

In biological cells, the interior is enclosed by a membrane, which separates it from the surroundings. The cell membranes are formed as assemblies of lipid molecules, e.g., phospholipids with a hydrophilic head and two hydrophobic tails. As a consequence of their amphiphilic nature, phospholipids undergo self-assembly in an aqueous milieu to form a bilayer with the hydrophobic tails on average pointed towards the center and the hydrophilic heads exposed to water. The lipid bilayer is nearly impermeable to ions and most polar organic molecules; therefore cells require transport protein systems for their transport.

Membrane transport proteins play critical roles in human physiology, including those responsible for the transport of glucose and amino acids into cells, transport of ions and organic molecules by the renal tubules, transport of Ca^{2+} and Na^+ out of cells, uptake of neurotransmitters and neurotransmitter precursors into

nerve terminals and vesicles, and transporters involved in multidrug resistance. Malfunctions in membrane transporters are responsible for a range of human diseases [1]. Various neurological and cardiac diseases are caused by defective ion channels [2]. Cystic fibrosis is a result of mutations in a chloride transporter that functions in the lungs [3]. Membrane transport proteins also act as drug targets, such as efflux proteins [4].

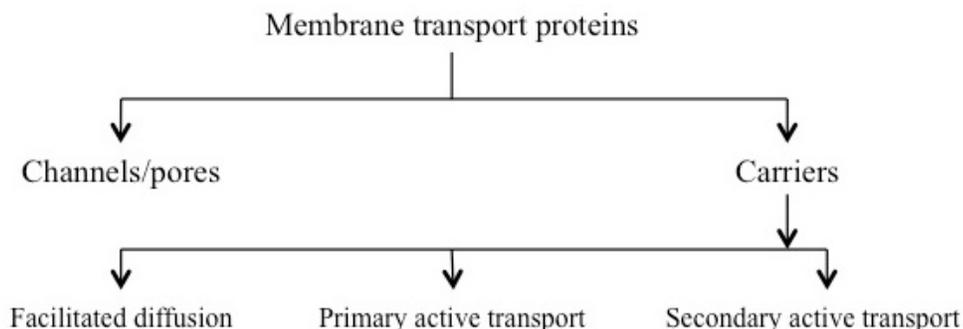


Figure 1.1: Classification of membrane transport proteins based on the transport mechanism

1.2. Classification of membrane transport proteins

Based on their transport mechanisms, membrane transport proteins can be broadly categorized as channels/pores and carriers (Figure 1.1). Channel proteins are transmembrane spanners that act as selective pores for the transport of substrates down their electrochemical gradient through an energy-independent process. While carrier proteins perform mediated transport of substrates, either by facilitated diffusion or by active transport. In facilitated diffusion, specific binding between substrate and protein accelerates its transport along the electrochemical potential gradient. Active transporters can transport substrates against their concentration

gradients in an energetic process. Primary active transporters use a primary source of energy such as the chemical energy of ATP or light. Secondary active transporters use electrochemical potential gradient of ions generated by primary active transport as an energy source for accumulating substrates against their concentration gradients.

According to the transporter classification (TC) system [5, 6] adopted by International Union of Biochemistry and Molecular Biology (IUBMB) which categories proteins based on both functional and phylogenetic information, there are seven classes of transport proteins including: 1) Channels/pores, 2) Electrochemical potential-driven transporters 3) Primary active transporters. LacY, a secondary active transporter belongs to the class of electrochemical potential driven transporters.

1.3. Major Facilitator Superfamily

Based on the sequence similarities, transporters are grouped into more than 300 families. Barring a few exceptions, members of a particular family employ the same mode of transport and energy coupling mechanism [7]. The major facilitator superfamily (MFS), which belongs to class 2 membrane transporters, is of interest for my research. MFS is the largest family of electrochemical potential-driven transporters with more than 10000 sequenced members identified [8, 9]. The MFS proteins are very diverse in terms of substrate transport and catalyze transport of nearly all types of small to moderately sized molecules (sugars, drugs, metabolites, anions) by uniport (transport of a single molecule by facilitated diffusion or by membrane potential dependent process for a charged solute), symport (transport of two or more molecules in the same direction) or antiport (transport of two or more

molecules in opposite directions). Transport proteins from this superfamily are found in almost all types of organisms, from bacteria (e.g. LacY and EmrD from *E. coli*) to humans (e.g. glucose-transporters GLUTs) [9].

1.3.1. Multidrug efflux proteins in MFS

Some MFS proteins are considered to be responsible for the antibiotic resistance in bacteria. Antibiotic resistance in bacteria is achieved by four different mechanisms: (i) inactivation of antibiotics by chemical modification, (ii) modification of the antibiotic target by mutation or enzymatic modification, (iii) inhibition of drug entry into the cell, and (iv) active efflux of drugs [10]. Some transport proteins are involved in active efflux of a specific drug or a class of drugs, whereas multidrug resistance (MDR) transporters can discard various structurally unrelated compounds. Several MFS proteins, EmrD from *E. coli*, NorA from *Staphylococcus Aureus*, LmrP from *Lactococcus lactis*, FlorR from *S. enteric*, act as MDR transporters [10, 11].

1.3.2. Structural and functional similarities in MFS proteins

Majority of MFS proteins are 400-600 amino acid residues long and have 12 transmembrane α -helices, with the exception of a few proteins with 14 and 24 α -helices [8, 9]. An x-ray crystal structure has been determined EmrD [12], which acts as an efflux pump for various structurally unrelated uncouplers of oxidative phosphorylation as well as detergents such as benzalkonium and sodium dodecylsulfate. Atomic level structures have been determined for five other MFS

proteins, lactose permease [13-15] (a galactoside transporter), GlpT [16] (a Glycerol-3-Phosphate transporter), FucP [17] (a fucose transporter), Xyle [18] (bacterial homolog of GLUT1-4), and a peptide transporter [19]. All these proteins have twelve transmembrane α -helices forming two domains of six helices each and a central cavity for substrate binding (Figure 1.2). Moreover, the transport of substrates in all these proteins is believed to involve alternating access of the central binding cavity to either sides of the membrane, but none of the MFS proteins has been crystalized in both of its open states. LacY, transporter of sugar molecules in *E. coli*, is one of the most extensively studied transmembrane proteins and it was investigated as a model for MFS proteins for the purpose of this dissertation.

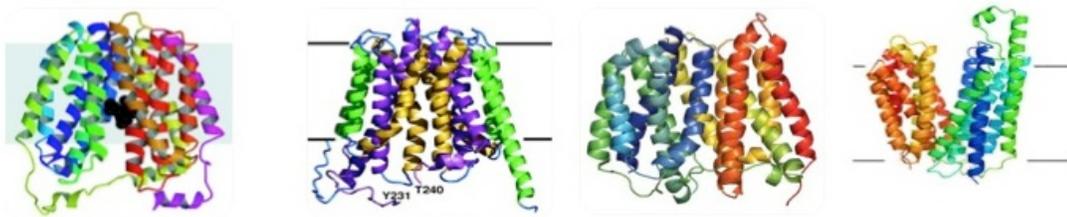


Figure 1.2: Illustrations of MFS protein crystal structures for (a) LacY [13] - galactoside transporter (b) GlpT [16] - glycerol-3-phosphate transporter (c) EmrD [12]- multidrug transporter (d) FucP [17] - fucose transporter

1.4. Lactose permease: paradigm for MFS proteins

LacY is the most comprehensively studied MFS protein with plethora of biochemical and biophysical experimental data available. LacY catalyzes coupled stoichiometric transport of a galactoside with a proton (H^+ /galactoside symport) in *E. coli* [20]. In presence of a proton electrochemical gradient ($\Delta\mu_{H^+}$, interior negative and/or alkaline), it uses free energy released from downhill movement of a proton to

transport a galactoside against its concentration gradient. In absence of $\Delta\mu_{H^+}$, it couples energetically downhill movement of a galactoside with an uphill movement of a proton with the generation of $\Delta\mu_{H^+}$ [20].

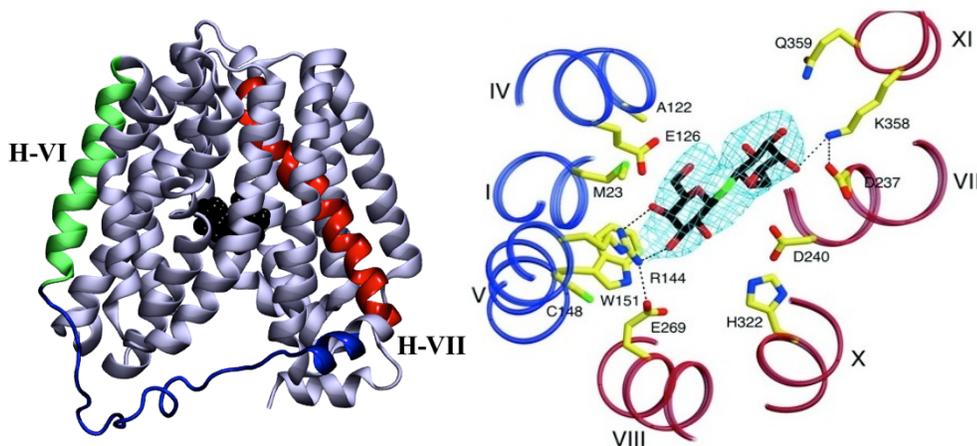


Figure 1.3: (left) Tertiary structure of LacY consisting of 12 α -helices that are divided in to two domains connected by a long loop (shown in blue color) between helix-VI (green) and helix VII (red) (right) Substrate binding in LacY-C154G mutant [13]. The picture on the right is adopted from Figure 4(a) of the review article by Guan and Kaback [20].

1.4.1. Structure of *LacY*

LacY acts as a monomer and its tertiary structure consists of 12 transmembrane α -helices that form two pseudo-symmetric domains, which are connected by a long loop between helices VI and VII (Figure 1.3 (left)) [20]. The x-ray crystal structure was first determined for the C154G mutant of LacY that was arrested in the cytoplasmic-open conformation (PDB ID: 1PV7) [13]. Subsequently crystal structure of the wild-type protein was determined [14] followed by structures of several LacY mutants [21, 22]. In all the crystal structure, LacY's lumen is open to the cytoplasm and tightly closed to the periplasm preventing substrate movement to

the periplasm. The pseudo-two-domain symmetry suggests that these domains could rotate and form the periplasmic-open structure of LacY. Attempts to crystallize the periplasmic-open state were unsuccessful although abundant experimental data indicates a mechanism where LacY alternates between cytoplasmic- and periplasmic-open states during substrate transport [23].

1.4.2. Alternating access mechanism in LacY

Since atomic-level descriptions of the periplasmic-open state of LacY were not determined, various biochemical experiments were used to indirectly measure how LacY structurally changes to allow for sugar transport. The double electron-electron resonance (DEER) [24] and single-molecule fluorescence resonance energy transfer (sm-FRET) [25] experiments that use labeled residue pairs to measure interhelical distances indicated a decrease in cytoplasmic interhelical distances and an increase in periplasmic interhelical distances for the wild-type LacY (LacY-wt) compared to the cytoplasmic-open state. Site-directed alkylation experiments [26, 27] that measure changes in reactivity for single-Cys replacement mutants for nearly all the 417 residues in LacY indicated an increase in the reactivity on the periplasmic side upon sugar binding. Thiol cross-linking experiments [28] also showed that long and flexible cross-linking agents exhibit sugar transport, while no transport activity was observed for short cross-linking agents. All these experiments indicate a mechanism that involves significant opening and closing of the hydrophilic cavity on either side of the membrane. The atomic-level details of the periplasmic-open state of LacY are thus critical in understanding the overall mechanism of sugar/proton

symport by LacY. As a part of this dissertation, I have determined a model for periplasmic-open structure of LacY and suggested mechanisms that trigger its conformational change from the cytoplasmic- to the periplasmic open state.

1.4.3. Sugar binding in LacY

LacY transports various galactosides across the plasma membrane. Monosaccharide galactose is the most specific substrate for LacY, but it has very low affinity [21, 29]. LacY's specificity towards galactopyranosides is believed to be originating from its interactions with C4-OH group on galactopyranoside moiety and substitutions on the galactopyranosides ring results in increase the binding affinity [29, 30]. LacY's putative binding site consists of Glu126, Arg144, Trp151 and Glu269 (Figure 1.3 (right)). In absence of a substrate, the residues in the binding site are not in the correct configuration to bind to the substrate. Sugar recognition in LacY takes place via hydrophobic stacking interactions between galactopyranosyl and indole rings of Trp151. On substrate recognition by Trp151, the residues in the binding pocket are rearranged to form a binding site [13]. LacY shares this binding motif with other sugar transporters such as the sucrose transporter CscB [29], where Asp129, Arg147, Tyr154, and Glu270 are believed to play roles similar to corresponding LacY residues in substrate binding. Although LacY shows specificity towards galactopyranosides, it binds to an α -anomer of a sugar with greater affinity compared to β [31, 32]. This could be a result of either the different binding sites for different anomers or the differences in the interactions between the sugar and specific protein residues. Only a subtle stereochemical alteration in the substrate results in a

significant change in its affinity towards the protein. The mechanism of substrate binding in LacY is studied in detail through binding free energy calculations (Chapter 4).

1.4.4. Proton translocation in LacY

Based on the experimental findings, a mechanism for the sugar/proton transport by LacY has been postulated [20]. The proton-coupled influx of a sugar consists of the following six steps (Figure 1.4): (1) LacY in the periplasmic-open state is protonated; (2) sugar enters LacY from periplasmic side; (3) protein undergoes conformational change to the cytoplasmic-open state; (4) sugar is then released on the cytoplasmic side; (5) proton is released on the cytoplasmic side; (6) LacY transitions back to the periplasmic-open state.

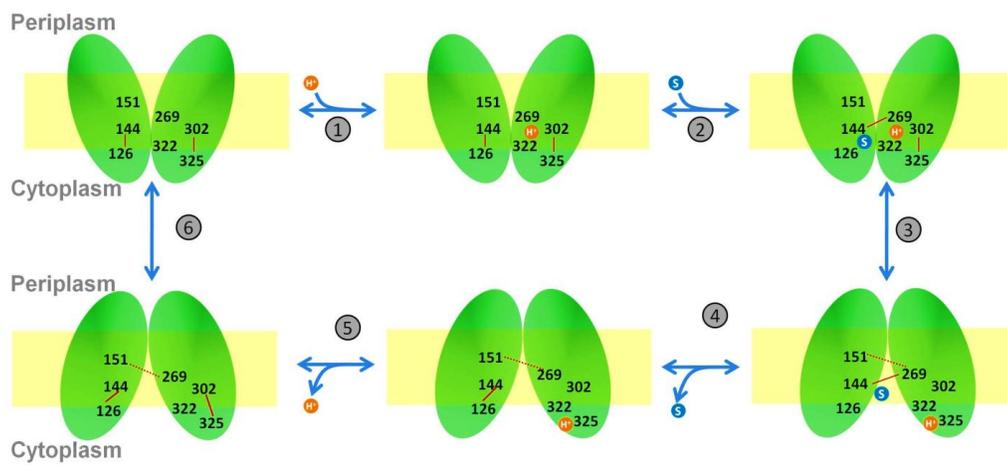


Figure 1.4: Mechanism of substrate transport by LacY based on the experimental findings [20].

It has been established that protonation of LacY, in other words binding of a proton, is a necessary precursor to sugar binding [33]. LacY then couples

translocation of protons with the transport of sugar molecules but the mechanism of this coupling is unclear. It has been suggested that the residues Glu325, Arg302, Tyr236, His322, which lie approximately at the center of the protein are critical for protonation and deprotonation of the protein while Glu269, which is also believed to be involved in sugar binding, couples the stoichiometric proton/sugar transport processes [33, 34]. Proton translocation in LacY was investigated using a hybrid QM/MM approach. The knowledge of proton translocation and its energetic coupling with substrate translocation will have far reaching benefits in understanding the functioning of other MFS proteins such as those involved in multidrug efflux.

Chapter 2: Computational Methods

In this chapter, I will give an overview of various computational techniques that are used in this dissertation. I will begin with a brief discussion about the theoretical foundations of molecular dynamics (MD) simulations and general practices for simulation setup and analysis. This will be followed by a discussion about the protocols used for calculating protein-ligand binding free energies. In the final section, hybrid quantum mechanics/molecular mechanics (QM/MM) protocol used in proton translocation studies will be reviewed.

2.1. Molecular Dynamics Simulations

2.1.1. Central idea

Complete understanding of how a biomacromolecule works requires knowledge of its structure and dynamics. An x-ray crystal structure of a protein can provide a static picture in atomic-level details, while several experimental approaches such as site-directed mutagenesis can offer abundant insights into its functioning and dynamics but lack atomic-level details. All-atom MD simulations provide a powerful tool to bridge these experimental techniques and obtain a complete picture. Since their introduction by Alder and Wainwright in 1959 [35] and first application on biomolecules by McCammon et al. in 1977 [36], MD simulations have seen tremendous growth in terms of their ability and applicability, thanks to a rapid improvement in computational power and emergence of new techniques. Present day

MD simulations are capable of reaching microsecond timescales for hundreds of thousands of atoms making them a critical part of the biomolecular study.

In MD simulations all atoms in the system are treated classically and Newton's equations are solved to generate successive configurations of the system. In this section, I will briefly discuss the theoretical foundations and common practices in MD simulations. Interested readers are encouraged to refer to several texts that discuss these topics more comprehensively [37-40].

Newton's equations of motion for a set of N particles is written as:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.1)$$

where

$$\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = -\nabla_{\mathbf{r}_i} V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.2)$$

Here is \mathbf{F}_i the total force acting on the particle i at position \mathbf{r}_i and V is the total potential energy. Combining these two equations, we get:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r}_i} V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.3)$$

Because of the deterministic nature of equations of motion, if initial positions and velocities of all the particles in the system are known, a trajectory of subsequent positions and velocities can be calculated by solving equation (2.3)

2.1.2. Potential energy and force calculations

A molecular dynamics program calculates potential energy and forces using a force field. A force field consists of a potential energy function and a set of parameters that define how the potential energy is related to atom positions. Equation (2.4) represents the potential energy function in a typical CHARMM force field [41].

$$\begin{aligned}
 V = & \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi (1 + \cos(n\Phi - \delta)) \\
 & \sum_{Urey-Bradley} k_{UB} (S - S_0)^2 + \sum_{impropers} k_\omega (\omega - \omega_0)^2 + \sum_{non-bonded} \left\{ \epsilon_{ij} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon r_{ij}} \right\} \\
 & + \sum_{residues} V_{CMAP}(\Phi, \Psi)
 \end{aligned}
 \tag{2.4}$$

First three terms that represent bond stretching, angle bending and rotation around a bond (Figure 2.1), respectively, constitute the bonded contribution to the potential energy function. The non-bonded potential energy has two components, the van der Waals interactions and the electrostatic interactions. Van der Waals interactions are described by a 6-12 Lennard-Jones (LJ) potential. In addition to these commonly used energy terms, CHARMM force fields have additional terms for distance between atoms separate by two bonds (Urey-Bradley), improper dihedral to maintain chirality and planarity in molecules and a cross-term correction (CMAP) for backbone dihedral angle values which improves conformational properties of protein backbones.

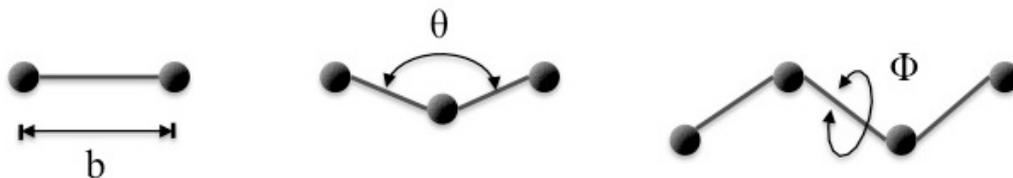


Figure 2.1: Bonded interactions (a) Bond stretching (b) Angle bending (c) Rotation around a bond

2.1.3 Integration schemes

Various algorithms are available in MD packages to integrate the equations of motion. The most commonly used ones are the Verlet, Velocity Verlet, and Leapfrog Verlet. In all these integration schemes, positions, velocities and acceleration are approximated by Taylor series expansion. In the Velocity Verlet algorithm, positions and velocities are calculated at the same values of time, in contrast to the Verlet algorithms where velocities are not calculated and leapfrog algorithm where positions are calculate at time t whereas velocities are calculated at time $t+\delta t$. Leapfrog verlet is the preferred integration scheme in MD simulations because of its high accuracy and ability to print the Hamiltonian on the go.

2.1.4. Periodic boundary conditions

Although the purpose of MD simulations is generally to study the macroscopic behavior, due to their high computational demands only a limited number of degrees of freedom can be handled by computers. Consider a system consisting of a protein embedded in a hydrated membrane (Figure 2.2 (left)). Water

and lipid molecules near the edge of the simulation box would experience surface artifacts with vacuum on the periphery. By applying the periodic boundary conditions (PBC), replicas of the simulation box are created as can be seen in 2D in Figure 2.2 (right). This allows for the particles at the surface of the simulation box to interact with the particles in its corresponding replicas, thus removing any surface artifacts. When a particle from the central simulation box escapes from its left surface, an image of the particle enters the box from its replica on the right, thus maintaining the particle density inside the box.

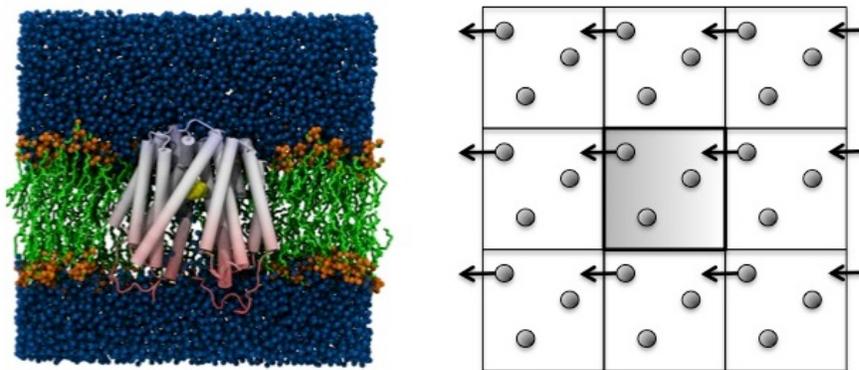


Figure 2.2: (left) A hydrated membrane-LacY complex. (right) A 2D implementation of a PBC. Shaded box in middle is the simulation cell surrounded by eight replicas. When a particle leaves the simulation cell, corresponding particle from a replica enters the cell from the opposite replica

In most of our protein-membrane simulations, we employed the more commonly used P1 boundary condition. In P1, simulation box replicas are created using translation alone. This approach (Figure 2.3 (left)) resembles the previously

described representation of PBC. In $P2_1$ boundary condition, the simulation box which is an asymmetric unit is replicated by rotational symmetry so as to create a continuous surface [42]. If a lipid molecule from the top leaflet escapes the simulation box, its image from the corresponding replica enters the box from the bottom leaflet (Figure 2.3(right)). This approach is particularly useful when a conically shaped protein like LacY undergoes significant conformational change. The $P2_1$ boundary condition allows for better equilibration of the membrane and therefore was used for our studies on conformational changes in LacY.

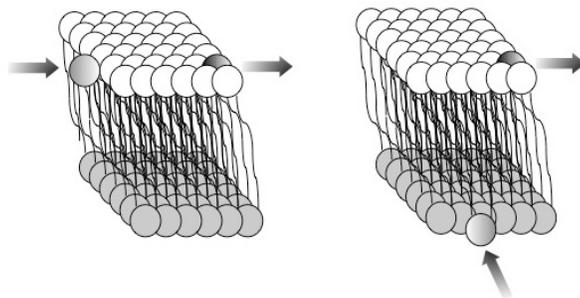


Figure 2.3: When a lipid from the top leaflet leaves the simulation cell in (left) P1 PBC, a lipid enters the top leaflet from the corresponding replica (right) $P2_1$ PBC; a lipid enters the bottom leaflet from the corresponding replica

2.1.5. Ensembles and control of system variables

In statistical mechanics, an ensemble is a collection of all the systems that are in an identical macroscopic state but in different microscopic states. The discussion in the previous sections dealt with MD simulations as a means for generating time evolution of a system where number of particles in the system (N), simulation volume (V) and the total energy (E) is conserved. This ensemble is known as microcanonical

or constant- NVE ensemble. But for practical purposes, it is generally more desirable to perform MD simulations in other ensembles such as:

- I. NVT or Canonical ensemble: The macroscopic state is characterized by number of particles (N), simulation volume (V), and temperature (T).
- II. NPT or Isobaric-Isothermal ensemble: The macroscopic state is characterized by number of particles (N), pressure (P), and temperature (T).
- III. $NPAT$ ensemble: An ensemble used for lipid membrane simulations where the macroscopic state is defined by the area per lipid (A) in addition to N , P , and T .

Several temperature and pressure control schemes are available in molecular simulation packages to allow for carrying out MD simulations in different ensembles by controlling different system variables including Berendsen thermostat and barostat, Nosé-Hoover thermostat and Parinello-Rahman barostat, Langevin thermostat.

2.1.6. Molecular simulation programs

All the MD simulations as a part of this dissertation were carried out using the simulations packages CHARMM (Chemistry at HARvard Molecule Mechanics) [41, 43, 44] and NAMD [45]. CHARMM is a very versatile simulation package with abundant functionalities. Wide-range of functionalities including various integration schemes, temperature and pressure control schemes, free energy methods, QM methods and large number of analysis tools have been implemented in CHARMM. We used CHARMM for molecular modeling, performing complex simulations, and

for analyzing trajectories. NAMD is specifically designed to carry out high-performance parallelized simulations of large biomolecular systems and was used to perform most production runs.

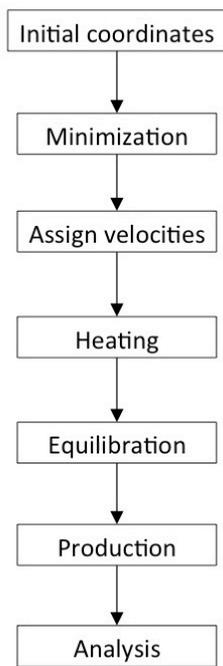


Figure 2.4: Flowchart representing the steps an MD simulation of a membrane-protein complex

Figure 2.4 represents the flow chart of a typical MD simulation performed for the purpose of this dissertation.

- I. Initialization: Initial coordinates for the protein are obtained from an x-ray crystal structure obtained from the protein databank (www.pdb.org) or screened from previous simulations. The protein structure is then inserted in a

pre-equilibrated membrane or a membrane-protein complex is built using CHARMM-GUI [46-48], a web based graphical user interface.

- II. Minimization: An energy minimization is carried out prior to starting MD simulations to remove any high-energy contacts that may exist.
- III. Heating dynamics: Initial velocities are assigned by randomly selecting from a Maxwell-Boltzmann distribution at a temperature lower than the temperature of interest. The velocities are gradually scaled up to a higher temperature followed by equilibration till the desired temperature is reached.
- IV. Equilibration: Once the desired temperature is reached, MD simulations are performed to verify that the system variables have reached an equilibration. If there is considerable drop or increase in the temperature, velocities can be reassigned at the desired temperature.
- V. Production and analysis: MD simulations are then performed on the equilibrated system. If constant pressure, temperature or both are desired in the ensemble, then the appropriate barostats and thermostats are used. The trajectories from the production phase are analyzed to collect the desired information.

2.2. Binding Free Energy calculations

In this section, I will describe the FEP/MD protocol used for binding free energy calculations. The details of simulation design can be found out in Chapter 4.

2.2.1. Ligand-protein association

Non-covalent association of a ligand (L) to a protein (P) can be represented by the process [49]:



$$\Delta G_b = -k_B T \ln(K_{eq}) \quad (2.6)$$

where K_{eq} and ΔG_b are the equilibrium constant and free change associated with ligand binding, respectively. Here k_B is the Boltzmann constant and T is the temperature. The equilibrium constant can be written as a function of the concentrations of the three species as:

$$K_{eq} = \frac{[PL]}{[P][L]} \quad (2.7)$$

Following the classical statistical mechanics according the formalism used by Woo and Roux [50], at low ligand concentrations K_{eq} can be written as:

$$K_{eq} = \frac{\int d\mathbf{L} \int d\mathbf{R} \exp(-U / k_B T)}{\int_{solv} d\mathbf{L} \delta(\mathbf{r}_L - \mathbf{r}^*) \int d\mathbf{R} \exp(-U / k_B T)} \quad (2.8)$$

where U represents the total potential energy of the system, \mathbf{L} and \mathbf{R} respectively represent the coordinates of the ligand and the other remaining atoms including the protein, membrane, waters and ions, \mathbf{r}_L represents the coordinates of center of mass of the ligand and \mathbf{r}^* represents the coordinates of an arbitrary point in the bulk region. Since the bulk around the solute is isotropic, the delta term is introduced in the denominator. This equation represents the ratio of configurational integrals for a ligand bound to the protein in the numerator and a ligand in the bulk solution with its

center of mass held at \mathbf{r}^* . In principle, this ratio can be calculated from a single MD trajectory in which one starts with a ligand bound to the protein, follows the process of dissociation and ends up with an unbound ligand in the bulk phase far away from the protein. But this approach is practically impossible due to the limited conformational sampling of MD simulations.

To make this calculation computationally feasible, free energy perturbation (FEP) formalism can be used. According the FEP formalism [49], the difference in free energies between states A and B is given by:

$$G_B - G_A = -k_B T \ln \langle (-\Delta U / k_B T) \rangle_A \quad (2.9)$$

where ΔU is the difference in the potential free energy functions between the two states and the angle brackets represent the Boltzmann average taken in the state A . Equation (2.9) converges well only when the perturbation (ΔU) is small enough. Several intermediates between bound and solvated states of the ligand can be constructed in such a way that the perturbation (ΔU) between the neighboring intermediates is small enough, which can be represented as:

$$K_{eq} = \frac{\int d\mathbf{L} \int d\mathbf{R} \exp(-U / k_B T)}{Z_1} \times \frac{Z_1}{Z_2} \times \dots \times \frac{Z_{n-1}}{Z_n} \times \frac{Z_n}{\int_{solv} d\mathbf{L} \delta(\mathbf{r}_L - \mathbf{r}^*) \int d\mathbf{R} \exp(-U / k_B T)} \quad (2.10)$$

Where Z_i 's represent the configurational integrals associated with the intermediates.

2.2.2. The double decoupling scheme

Since free energy is a state function, the intermediates can be chosen based on the computational convenience. The “double decoupling scheme” [51] is used to construct the intermediates. Here the interactions between the ligand and its surrounding in the bound state as are gradually turned off until it is completely decoupled from the surrounding and are subsequently restored in the bulk solution thus completing the thermodynamic cycle (Figure 2.5).

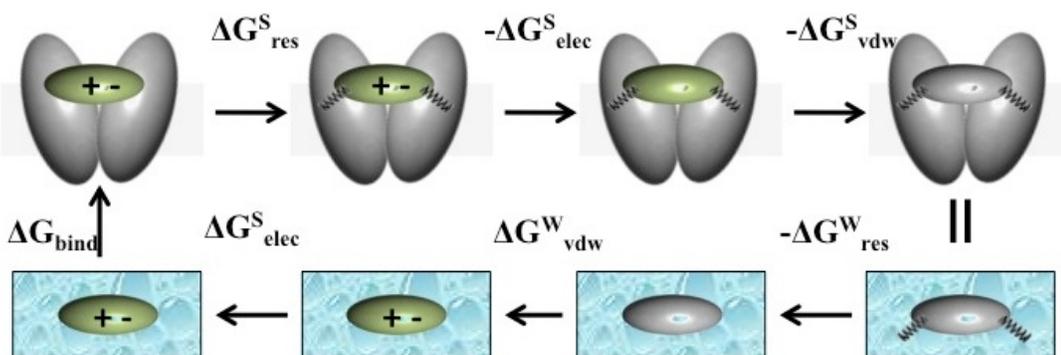


Figure 2.5: Alchemical thermodynamic cycle for substrate binding to LacY. On the upper left is a completely interacting substrate in the bound state and on the lower left is the completely the substrate in bulk solution. The electrostatic (+ -) and LJ (color green) interactions are gradually turned off and on in the bound are solvated states, respectively.

The interaction free energy between the ligand and its surrounding is divided into electrostatic and Lennard-Jones (LJ) contributions. The LJ contribution is further divided into dispersive and repulsive interaction energies using the Weeks-Chandler-Andersen (WCA) scheme [52].

$$\Delta G_{\text{int}}^X = \Delta G_{\text{elec}}^X + \Delta G_{\text{disp}}^X + \Delta G_{\text{repu}}^X \quad (2.11)$$

The free energy contributions for different interactions are calculated using alchemical FEP formulation by separately turning off the electrostatic, dispersive and repulsive interactions. A coupling parameter ‘ λ ’ is introduced to turn on and off the interactions between ligand and its surrounding. When is λ equal to 0, the specific interaction between the substrate and its surrounding is completely turned off and when λ is equal to 1, the interaction is completely turned on. The potential energy can be expressed as a function of as:

$$U(\lambda_j) = U^0 + U_j(\lambda_j) \quad \dots j = \text{disp, elec, repu} \quad (2.12)$$

Here U^0 is the total potential energy when the interaction between ligand and its surrounding is completely turned off and $U_j(\lambda_j)$ is the contribution of interaction ‘ j ’ to the potential energy scaled by λ_j . Values of λ_j are gradually varied from 0 to 1 and free energy change associated with each step is calculated using equation (2.9).

2.2.3. Contributions from multiple protein conformations

Since LacY undergoes conformational transformation on sugar binding, the overall sugar binding affinity will have contributions from the cytoplasmic- as well as periplasmic-open states. The contributions of binding conformations from individual open states of LacY towards the overall binding affinity will be proportional to the population of LacY in cytoplasmic- (P^{cyto}) and periplasmic-open (P^{peri}) states. Following the formalism by Gallicchio and Levy [53], the overall binding free energy can be expressed as:

$$\Delta G_b^0 = -k_B T \ln [P^{cyto} e^{-\beta \Delta G_b^0(cyto)} + P^{peri} e^{-\beta \Delta G_b^0(peri)}] \quad (2.13)$$

where $\Delta G_b^0(peri)$ and $\Delta G_b^0(cyto)$ are the standard binding free energies calculated from binding conformations in the periplasmic- and the cytoplasmic-open states of LacY, respectively. To get the relative populations of the two open states of LacY, we used the areas-under-curve of the distance distribution plots from the double electron-electron resonance (DEER) experiments by Smirnova et al. [24].

2.2.4. *Orientational decomposition*

Our studies on conformational changes in LacY (details in Chapter 3) showed that LacY binds to the substrate molecules in multiple stable binding conformations in both open states. Therefore to obtain meaningful $\Delta G_b^0(cyto)$ and $\Delta G_b^0(peri)$, it is necessary for the ligand to sample through all the possible binding conformations while calculating the interaction free energies in the bound state (ΔG_{int}^{Bound}). It is impractical to sample through all the relevant points in the configurational space from the MD simulation windows. To overcome this problem, overall free energy of binding is decomposed into individual contributions from different binding conformations according to the orientational decomposition scheme by Mobley et al. [54].

According to the notations used by Gilson et al. [55] and Mobley et al. [54], free energy of binding after neglecting the contribution due to change in equilibrium volume on ligand binding, can be expressed as:

$$\Delta G^0 = -\beta^{-1} \ln \left[\frac{C^0}{8\pi^2} \frac{\sigma_P \sigma_L}{\sigma_{PL}} \frac{Z_{PL} Z_0}{Z_P Z_L} \right] \quad (2.14)$$

If we decompose the overall bound state into n distinct orientations, equation (2.14) becomes:

$$\Delta G_{bind}^0 = -\beta^{-1} \ln \left[\frac{C^0}{8\pi^2} \frac{\sigma_P \sigma_L}{\sigma_{PL}} \frac{(Z_{1,PL} + Z_{2,PL} + \dots + Z_{n,PL}) Z_0}{Z_P Z_L} \right] \quad (2.15)$$

On orientational decomposition, the binding free energy in orientation ‘ i ’ is given by:

$$\Delta G_i^0 = -\beta^{-1} \ln \left[\frac{C^0}{8\pi^2} \frac{\sigma_P \sigma_L}{\sigma_{PL}} \frac{Z_{i,PL} Z_0}{Z_P Z_L} \right] \quad (2.16)$$

ΔG_i^0 can be calculated separately using the double decoupling scheme by keeping the ligand in the state defined by $Z_{i,PL}$. By combining equations (2.15) and (2.16) we can get the overall free energy of binding as:

$$\Delta G_{bind}^0 = -k_B T \ln \sum_{i=1}^n \exp \left(\frac{-\Delta G_i^0}{k_B T} \right) \quad (2.17)$$

2.2.5. Use of restraints and constraints

It is important to be able to define the bound state to be able to apply orientational decomposition. The translational and rotational restraints are applied on the ligand to define different binding modes using the formalism developed by Boresch et al. [51]. Besides defining the bound state, translational and rotational restraints considerably improve convergence in FEP/MD simulations. Three points on the protein (P1, P2, P3) and three points on the ligand (L1, L2, L3) are used to

formulate the restraints. Although, the choice of these points can be arbitrary, centers of mass for the protein (P1) and ligand (L1) were chosen as the two points common to all our FEP/MD simulations. The translational degrees of freedom of the ligand relative to the protein are defined by the distance (r), angle (θ), and the dihedral (ϕ) as shown in Figure 2.6. Similarly, the rotational degrees of freedom are defined by the angle (Θ) and the dihedrals (Ψ and Φ).

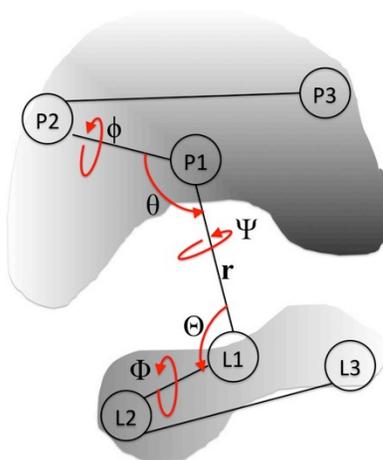


Figure 2.6: Three translational (r , θ , ϕ) and three rotational degrees of freedom (Θ , Φ , Ψ) of ligand relative to protein are defined based on three protein points (P1, P2, P3) and three ligand points (L1, L2, L3) that can be chosen arbitrarily. In our FEP/MD simulations we have chosen P1 and L1 as the centers of mass of protein and ligand, respectively.

Translational and rotational restraints define the position of center of mass and the orientation of ligand relative to the protein, respective. The potential energy functions for the translational (u_t) and rotational (u_r) restraints have the following forms:

$$u_t(r, \theta, \phi) = \frac{1}{2}k_r(r - r_0)^2 + \frac{1}{2}k_\theta(\theta - \theta_0)^2 + \frac{1}{2}k_\phi(\phi - \phi_0)^2 \quad (2.18)$$

$$u_r(\Theta, \Psi, \Phi) = \frac{1}{2}k_\Theta(\Theta - \Theta_0)^2 + \frac{1}{2}k_\Psi(\Psi - \Psi_0)^2 + \frac{1}{2}k_\Phi(\Phi - \Phi_0)^2 \quad (2.19)$$

where r , θ , and ϕ are the translational and Θ , Ψ , and Φ are the rotational degrees of freedom of the ligand relative to the protein. k_r , k_θ , k_ϕ , k_Θ , k_Ψ , and k_Φ are the corresponding force constants, and r_0 , θ_0 , ϕ_0 , Θ_0 , Ψ_0 , and Φ_0 are the corresponding reference values.

The free energy change associated with application of restraints in the bound state is calculated from equation (2.9) by gradually increasing the restraint strength using a coupling parameter that varies from 0 (No restraints) to 1 (fully active restraints) The free energy change for removing the translational and rotational restraints the bulk solution is obtained from the translational (F_t) and rotational (F_r) factors, respectively. F_t and F_r are calculated by integrating restraint potentials over the translational and rotational degrees of freedom, respectively.

$$F_t = \int_0^\infty dr r^2 \int_0^\pi d\theta \sin\theta \int_{-\pi}^\pi d\phi \exp[-\beta u_t(r, \theta, \phi)] \quad (2.20)$$

$$F_r = \frac{1}{8\pi^2} \int_0^\pi d\Theta \sin\Theta \int_0^\pi d\Psi \int_{-\pi}^\pi d\Phi \exp[-\beta u_r(\Theta, \Psi, \Phi)] \quad (2.21)$$

To improve convergence, ligand is constrained near its bound conformation using a harmonic root-mean-standard deviation (RMSD) potential, $u_c = k_c(\xi - \xi^0)^2$.

The free energy contributions associated with the configurational constraints (ΔG_c^{site} and ΔG_c^{solv}) are calculated using an umbrella sampling scheme [56]. The binding free

energy (ΔG_i^0) for the binding conformation 'i' is obtained by summation of interaction, conformational and restraint free energy contributions as:

$$\Delta G_1^0 = \Delta G_{\text{int}}^{\text{site}} - \Delta G_{\text{int}}^{\text{solv}} - \Delta G_{\text{conf}}^{\text{site}} + \Delta G_{\text{conf}}^{\text{solv}} - \Delta G_{\text{trans}}^{\text{site}} - k_B T \ln(-F_i C^0) - \Delta G_{\text{rot}}^{\text{site}} - k_B T \ln(-F_r) \quad (2.22)$$

The binding free energy contributions for the two open states of LacY can be calculated from equation (2.22) and the overall value can be obtained from equation (2.13).

2.2.6. Reduced atom boundary potentials

Although the use of RMSD constraints and translational and rotational restraints improve conformational sampling for FEP/MD calculations, the computational cost of simulating a membrane-protein complex is significantly high. To overcome this problem, a reduced atom potential approach was used named generalized solvent boundary potential (GSBP) [57] to simulate substrate dynamics in the bound state. In this formulation, only a small subset of the system is treated explicitly using molecular mechanics, while an effective boundary potential represents the effect of rest of the system. GSBP was originally applied for a soluble protein system, but was later tested successfully to study substrate binding and calculation for substrate binding affinities in LeuT, a membrane transport protein [58]. The application of GSBP reduces the system size from ~100,000 atoms in an explicit membrane-LacY assembly to ~6000 atoms. The spherically symmetric boundary potential (SSBP) [59] was used to simulate substrate dynamics in water.

2.3. Hybrid QM/MM simulations

This section deals with the idea behind the hybrid QM/MM approach and the selection of a QM/MM protocol for our proton binding studies. The details about the computational design of QM/MM simulations and calculation of energetics of proton translocation are described in Chapter 5.

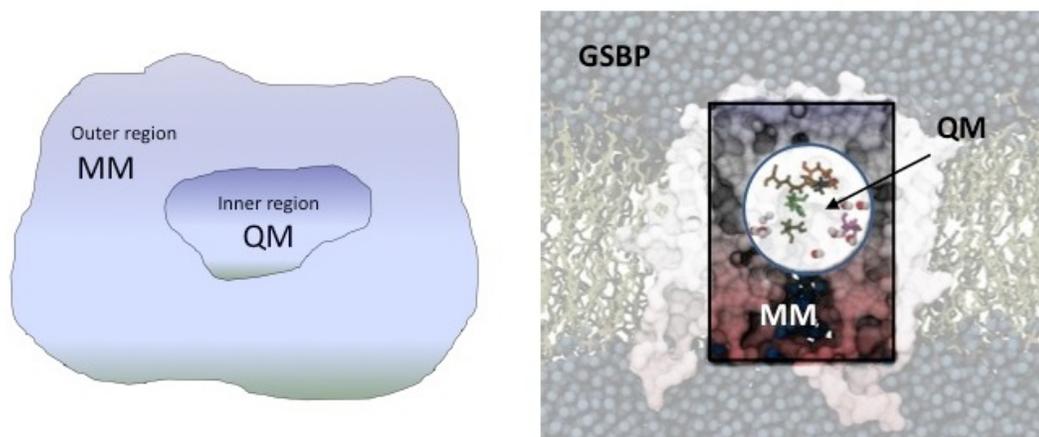


Figure 2.7: (left) A hybrid QM/MM scheme where system is divided into an inner region treated with QM and an outer region treated with MM (Right) SCC-DFTB/MM-GSBP scheme [60]. The system is divided into inner and outer regions while the rest of the bulk is taken care of by GSBP.

2.3.1. Central idea behind QM/MM

To simulate transport cycle of LacY that involves proton translocation that is believed to involve certain residues in approximate center of LacY and several structural water molecules, the system needs to be divided into two subsystems (Figure 2.7 (left)). The inner region that participates in proton translocation undergoes changes in the electronic structures and needs to be treated using quantum mechanical

(QM) methods. The applicability of present QM methods is limited to a few hundred atoms therefore the rest of the protein, which undergoes conformational changes in response to the proton translocation, and the hydrated membrane in which the protein is embedded can be treated using classical molecular mechanical (MM) methods.

The formulation of a QM/MM approach requires selection of the QM method, the MM methodology and the scheme for treatment of the transition between the QM and MM regions. We used the SCC-DFTB/MM-GSBP as the QM/MM protocol [60] (Figure 2.7(Right)).

2.3.2. *The QM region*

Selection of the QM method depends on the desired level of accuracy, computational cost, and specifics of the region of interest. In principle, post-Hartree–Fock *ab initio* electron-correlation methods, such as those based on Møller–Plesset perturbation theory or coupled-cluster theory (e.g., CCSD), offer very high-level of accuracy. But high computational costs associated with these methods limit their use to fixed-structure or single-point energy calculations. The other popular QM method for biomolecular QM/MM applications is the density functional theory (DFT) on account of its high accuracy/computational cost ratio. The QM method used will be the self-consistent-tight-binding density-functional-theory (SCC-DFTB). The SCC-DFTB method is an approximate QM method based on DFT. In contrast to high-level QM methods, SCC-DFTB has greater degree of conformational sampling ability without compromising on accuracy that is required for studying proton translocation. The SCC-DFTB method has been used successfully in many biological applications

involving QM regions larger than 100 atoms [60-63] and has been shown to be very effective in the proton binding and translocation reactions [60, 62]. The third order expansion of SCC-DFTB method instead of the default second order expansion will be used as it has shown significant improvements in proton binding affinities [64].

SCC-DFTB has been implemented in CHARMM to perform QM/MM calculations. The total energy of the of the system ($E_{QM/MM}^{tot}$) is written as [65]:

$$E_{QM/MM}^{tot} = \langle \Psi | H_{SCC-DFTB} + H_{QM/MM}^{elec} | \Psi \rangle + U_{QM/MM}^{vdW} + U_{QM/MM}^{bonded} + U_{MM} + U_{QM/MM}^{cons} \quad (2.23)$$

Here Ψ represents the wave function of the QM region, and $H_{SCC-DFTB}$ and $H_{QM/MM}^{elec}$ represent the Hamiltonian operators of the QM region and electrostatic interactions between QM and MM regions, respectively. $U_{QM/MM}^{vdW}$, $U_{QM/MM}^{bonded}$, U_{MM} , $U_{QM/MM}^{cons}$ are the energy contributions from QM/MM van der Waals interactions, bonded interactions, potential energy of MM region and contributions from constraints, respectively.

2.3.3. The MM region

Remaining atoms in the MM region of the system will be treated using CHARMM MM force fields. The GSBP method [57, 66] will be used to describe the electrostatic contributions far away from the active site. The GSBP method combined with the QM/MM simulations has been shown to produce results in agreement with the experimental data for carbonic anhydrase II [60].

2.3.4. Treatment of boundary between QM and MM regions

The link-atom approach [67] was used to treat boundary between QM/MM regions across a covalent bond. In this approach, an additional atomic center (usually a hydrogen atom) is introduced between at the boundary between QM and MM regions. It is covalently bound to the interface MM atom thus satisfying its free valency. The link atom is included in the QM regions thus allowing QM calculations on an electronically saturated system.

Chapter 3: Study of conformational changes in LacY

3.1. Background

Although all the crystal structures of LacY have been determined in the cytoplasmic-open state, several indirect biochemical experiments (see Review by Smirnova et al. [23] for reference) indicate presence of a periplasmic-open state and a possible intermediate state where LacY is closed towards both periplasmic- and cytoplasmic ends. LacY alternates between the two open states via a possible intermediate state during its transport cycle. Other MFS proteins were crystallized in the periplasmic-open state thus corroborating the suggestion that MSF proteins follow the alternating access mechanism of transport. Several computational studies were performed on LacY and other MFS proteins to throw light on their transport mechanism. These were focused on helix dynamics and stability [68], substrate binding [69-72], and substrate transport pathway[70, 72, 73]. Yin et al. [70] were the first to report any structural changes in LacY using a simulation approach. They carried out molecular dynamics (MD) simulations of LacY embedded in a fully hydrated lipid bilayer to study the conformational changes in LacY on translocation of a proton from Glu325 to Glu269. In the Glu269 protonated state, slight closing was observed in the cytoplasmic half during the 10-ns simulation. Similar cytoplasmic closure was observed during the 50-ns simulations by Holyoake et al. [73] with Glu325 protonated. This suggests that closing of the cytoplasmic half is not dependent on the protonation state of LacY. Steered molecular dynamics (SMD) simulations by Jensen et al. [74] in which sugar was pulled through the protein lumen

resulted in small periplasmic conformational changes. Although transport of the disaccharide from the binding pocket to the periplasm was observed, the periplasmic pore did not open in a manner consistent with experimental studies [75-77]. This indicates that the observed structural changes could be an artifact of the SMD approach, which may not allow enough time for the protein to respond to the substrate being pulled through the lumen. MD simulations on an MFS protein, GlpT, also resulted in partial closure of the cytoplasmic half while no significant periplasmic changes were observed [78]. Recently, a structural model for the periplasmic-open state of LacY was determined by swapping the conformations of inverted topology repeats identified in LacY's N- and C-terminal domains [79]. Although, this model gives a static picture of the periplasmic-open LacY, the analysis is entirely based on an energy-minimized structure and lacks the details about the dynamics of the proposed structure in a cell-like environment. Andersson et al. [80] performed MD simulations with a focus on understanding the proton coupled conformational dynamics in LacY and lipid-protein interactions. No periplasmic opening of LacY was observed during the ~100 ns simulations.

The focus of this work was to understand the atomic-level details of the outward-facing state of LacY and the mechanism for structural changes between the inward- and outward-facing states following the dynamics of the protein in a cell-like environment. This research was published in the *Journal of Molecular Biology* (cover article) [81].

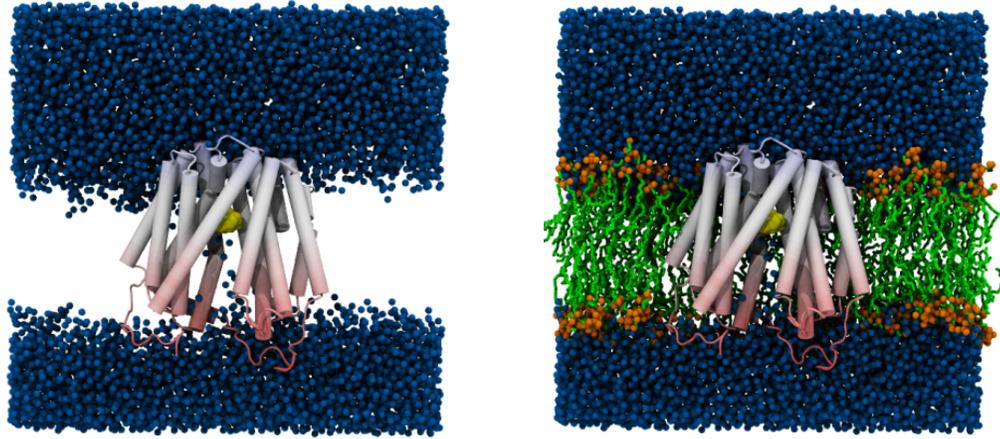


Figure 3.1: Snapshots of LacY embedded in an implicit membrane (left) and an explicit membrane (right)

3.2. Computational design and methods

3.2.1. Two-step hybrid simulation approach: IM-EX

A two-step hybrid simulation approach was used. In the first step, self guided Langevin dynamics (SGLD) [82] simulations were performed with LacY in an implicit bilayer (Figure 3.1 (left)) to enhance the conformational changes that MD cannot attain. SGLD uses local averaged momentum, which is averaged over the neighboring conformational space, to guide the conformational transition [82]. The local averaging time and the guiding factor determine the low frequency motions to be enhanced and the strength of guiding, respectively [82]. Previously, SGLD has been shown to significantly improve the ability to search protein conformations without the loss of accuracy in peptide folding [82] and various conformational changes in aqueous proteins [83, 84]. This method has been tested extensively in

small molecule and peptide folding simulations with the density of states, energies (kinetic and potential), and energy fluctuations all unaltered compared to long unguided simulations. A reasonable choice in SGLD parameters maintains this stability while significantly enhancing conformational sampling [82]. In the second step, MD simulations of structures screened from the first step were carried out in a fully explicit and hydrated lipid bilayer (Figure 3.1(right)) to obtain the final LacY conformations in a more realistic environment.

3.2.2. *Simulation setup*

The CHARMM simulation package [85, 86] was used with a simulation time step of 2 fs. The CHARMM family of force fields are used to describe the atomic interactions of the lipids [87, 88], protein (C22-CMAP) [89-91], and disaccharide [92]. The TIP3P water model was used (the version consistent with CHARMM parameters [93, 94]). Long range electrostatic interactions were calculated with the particle-mesh Ewald (PME) [95] method. The screening parameter (κ) was set at 0.45 \AA^{-1} and a fast Fourier transform grid density of $\sim 1 \text{ \AA}^{-1}$. Lennard-Jones interactions were switched smoothly to zero between 8–10 \AA [85]. The SHAKE algorithm [96] was used to constrain hydrogen atoms. Extended system formalism is used to maintain the pressure with a barostat [97, 98] and the temperature constant (EX simulations only) with the Nosé-Hoover method [99] at 310.15 K. Protein structures simulated had the Glu269 residue protonated, which has been suggested to be the preferred protonated state open to the periplasm [70, 100]. All simulation coordinates were saved every picosecond for analysis purposes.

A new two step hybrid simulation method, referred to here as IM-EX (implicit-explicit membrane), was used to probe conformational changes in LacY because fully explicit membrane MD simulations are limited to timescales approaching 1 μ s [101]. For the first step (IM) in our hybrid simulation procedure, an implicit membrane was used reduce computational demand and enhance conformational changes of the protein. Generalized Born solvation models would be the ideal choice for an implicit membrane, such as those developed by Feig et al [102-104]. However, our benchmark simulations with LacY and the heterogeneous dielectric generalized Born method[103] required simulation times that were equal to or greater than a fully explicit bilayer. Instead, an external hydrophobic potential based on the method of Edholm and Jähnig [105] is used to represent the lipid/water interface. This potential is assumed to vary exponentially across the membrane surfaces,

$$V(z) = \begin{cases} \frac{1}{2} \sum_{\forall i} h_i e^{-(|z|-z_0)/\lambda} & \text{for } |z| \geq z_0 \\ \frac{1}{2} \sum_{\forall i} h_i \left[2 - e^{(|z|-z_0)/\lambda} \right] & \text{for } |z| < z_0 \end{cases}$$

where the membrane interface is at $\pm z_0$ (assumed to be 15Å), the decay length (λ) is 2, and h_i is the hydrophobic energy of atoms taken from reference [105]. An explicit water interface was included in all implicit lipid membrane simulations (Figure 3.1(left)). The lumen of LacY is filled with water [69, 70, 73] and not including explicit water may result in an unphysical closing of the cytoplasmic opening. Consequently, the implicit membrane only acts on the protein. At the membrane interface, a thermodynamically stable vapor-liquid interface is formed which prevents water from flooding into the membrane region.

SGLD was used to further enhance protein motions [82]. This method has been shown to enhance low-frequency motions and conformational searching of proteins [82-84]. The following SGLD parameters were used in the implicit simulations: local averaging time $t_L = 0.25$ ps, guiding temperature of 1 K, and a friction factor $\gamma = 1 \text{ ps}^{-1}$. The local averaging time enhances all motions with periods larger than t_L , while the guiding temperature limits the perturbation to the system. This method has been shown to maintain the correct ensemble averages and fluctuations [82].

The implicit bilayer simulations (IM) were performed with three residues (Glu269, His322 and Glu325) separately protonated because of their involvement in the proton translocation mechanism. The initial protein structures for the IM simulations are from certain snapshots of previous explicit membrane (EX) simulations [69]. The 5-ns snapshot of EX-wt/ $\beta\beta$ -(Galp)₂ (run 2) in Klauda and Brooks [69] was used for the initial conformations of IM-r1 (run 1 with SGLD), IM-*apo* (simulation without sugar but using SGLD), IM-322^H (simulations with His322 protonated), IM-325^H (simulations with Glu325 protonated) and IM-MD (simulation with MD instead of SGLD). The 5-ns snapshot was chosen because it has a slight conformational change in the periplasmic half of LacY compared to the x-ray crystal structure.[13] A single disaccharide, $\beta\beta$ -(Galp)₂, is used for all but the *apo* simulation because sugar binding is known to induce structural changes toward the outward-facing state [76]. Two different protein coordinates were used as initial structures for the IM-r2 and IM-r3 runs. IM-r2 used a protein snapshot from the EX-wt/ $\beta\beta$ -(Galp)₂ (run 2) in reference [69] at 12 ns (some conformational change in the periplasmic

half) and IM-r3 used the final coordinate of this previous run (no changes in the periplasmic half). Four IM simulations with SGLD were run for 10 ns each and the IM-MD simulation was run for 20 ns. Each simulation consisted of 35,370 total atoms with 9,563 water molecules.

For the second step (EX), two independent simulations were run for each of the two screened conformations from the previous IM step with different initial velocities. All EX simulations were run with a time step of 2 fs. An explicit palmitoyloleoylphosphatidylethanolamine (POPE) bilayer was used to model the plasma membrane, which is known to be important for the function and topology of LacY [106-108]. Simulations were carried out in the NPAT ensemble at physiological temperature of 310.15 K and an experimentally suggested area of $65.2 \text{ \AA}^2/\text{lipid}$ [109]. The details about the simulation time and system size for each run are given in Table 3.1. The $P2_1$ periodic boundary condition [110] was used which allows for the redistribution of lipids between the leaflets, which is important for any significant structural change in LacY. For EX-r1, the average number of lipids in the periplasmic leaflet reduced from 142.0 ± 1.7 in the first ns to 133.1 ± 1.9 in the last ns.

3.2.3. Analysis methods

Pore radius analysis calculations were done using the program HOLE [111] based on 0.5-ns blocks. The distances between all the helices on the periplasmic and cytoplasmic halves of the protein are used as a measure of protein structural change and helix packing. Helix-helix distance maps were generated by calculating average distances between the helix centers on periplasmic and cytoplasmic sides every

nanosecond. The helix centers are calculated based on the center of mass of backbone atoms between $z=10\pm 1$ and -10 ± 1 Å for periplasmic and cytoplasmic sides respectively, as used previously [73]. The molecular surface areas are calculated using Lee and Richards algorithm [112] that uses a sphere of radius 1.6 Å to probe the molecular surface. Visual molecular dynamics (VMD) was used to make all molecular figures [113].

Table 3.1: System sizes and simulation times for the lactose permease simulations. MD was used in the EX and IM-MD simulations and SGLD for the other IM simulations. Unless otherwise noted, the simulations are with Glu269 protonated. ^aFor this conformation three simulations were run with Glu269, His322, and Glu325 protonated.

Run name	Simulation Time (ns)	Initial structure	Total atoms	Water molecules	Lipid molecules
IM-MD	20ns				
IM-r1 ^a	10ns	Snapshot from Klauda	35,370	9,563	-
IM-r2	10ns	& Brooks,			
IM-r3	10ns	2007[69]			
EX-r1	25ns	3.4ns snapshot	91,471	17,055	269
EX-r2	20ns	from IM-r1			
EX-r3	20ns	5.5ns snapshot	94,876	18,440	263
EX-r4	40ns	from IM-r1			

3.3. Results

This section is divided in five subsections. The first subsection explains the rationale behind the use of this hybrid simulation approach (IM-EX) for obtaining structural changes in LacY. The details of this approach are given in the Methods section, but the nomenclature, system size and simulation times are in Table 3.1. Structural changes are described by means of Root Mean Square Deviation (RMSD) from 1PV7 crystal structure and pore radius analysis. Structural information about the proposed outward-facing LacY is described in the second subsection, In third and fourth subsections, the data obtained from the simulations is compared with the experimental results [75] [76] [77] [114] and other computational models, respectively. The final subsection describes the binding of a sugar molecule in the inward- and outward-facing states.

3.3.1. Conformational changes in LacY

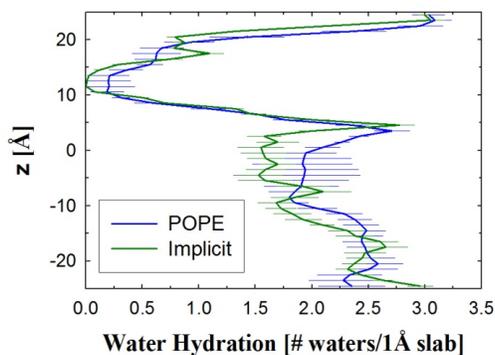


Figure 3.2: The hydration of LacY's lumen for the average of six simulations in an explicit POPE membrane (POPE) [69] and the implicit membrane simulation with MD (Implicit). The number of water molecules within bins of 1 Å is reported with the periplasmic half as positive z.

A. Implicit membrane simulations

Initially, a MD implicit membrane simulation and a previous explicit simulation [69] were compared to verify that the implicit membrane did not influence the protein structure and that of the water in the lumen. (Unless otherwise noted IM simulations are with Glu269 protonated.) Since the implicit membrane only acts on the protein, waters that are in the protein lumen are not directly influenced by this implicit membrane. If there is no significant conformational change in the protein compared to previous explicit simulations [69], then the hydration of the protein lumen should be similar for the EX and IM simulations. As described below, the IM-MD simulations resulted in no significant change in protein structure and can be compared with previous EX simulations. The water profile is nearly identical between the EX and IM-MD simulations (Figure 3.2), which suggests a negligible influence on waters in the protein lumen.

Conformational changes on the periplasmic half of LacY in the IM-MD simulation were negligible compared to previous EX simulations and the x-ray crystal structure. The distances between helices at the membrane interface only showed slight changes from the x-ray structure and consistent with fluctuations seen in previous EX simulations [69]. The pore radius at the periplasmic constriction (Figure 3.3) did not change from the x-ray crystal structure, and for the cytoplasmic half, it decreased by ~ 2 Å which is consistent with previous simulations [69, 70]. Multiple simulations will be required to determine if the IM-MD simulations result in the conformational changes similar to EX simulations, which is not the focus of present

study. The details of the helix-helix distance and pore radius calculations are given in the Methods.

Since MD was not successful in obtaining conformational changes in the periplasmic half of LacY, SGLD was used to enhance conformational sampling. Since we are starting from a state open to the cytoplasm, the mechanism [20] in Figure 1.4 suggest that proton translocation from Glu325 to Glu269 facilitates this structural change. Therefore, there are two controls in our simulations that should not result in periplasmic conformational changes if protonation of Glu269 triggers such a change, i.e., IM-r1 simulations with Glu325 and His322 protonated. In addition, simulations were carried out with *apo*-LacY (without sugar) to understand the effect of sugar binding on conformational changes [76, 115] as a likely third control.

All simulations resulted in a significant closure of the cytoplasmic half but only the simulation with Glu269 protonated resulted in some opening on the periplasmic side (data not shown here). The cytoplasmic closure was observed to be independent of the protonation states of Glu269, His322 and Glu325 which is in accordance with the findings by Holyoake et al. [73] All SGLD simulations with Glu269 protonated resulted in conformational changes in the periplasmic half of LacY (Figure 3.3) indicating that protonation of Glu269 and binding of sugar triggers opening of LacY on the periplasmic side.

The remaining presentation of the results will be based on simulations with Glu269 protonated and a disaccharide bound because these were the only to show periplasmic conformational changes. IM-r1 resulted in the largest conformational change with an increase in pore radius ~ 3 Å (Figure 3.3, bottom). For IM-r2 and -r3,

the opening of the periplasmic half is reduced compared to IM-r1. Consistent for all these runs is that H-V/VI and H-IX/X remain packed with only minor changes in helix-helix distances.

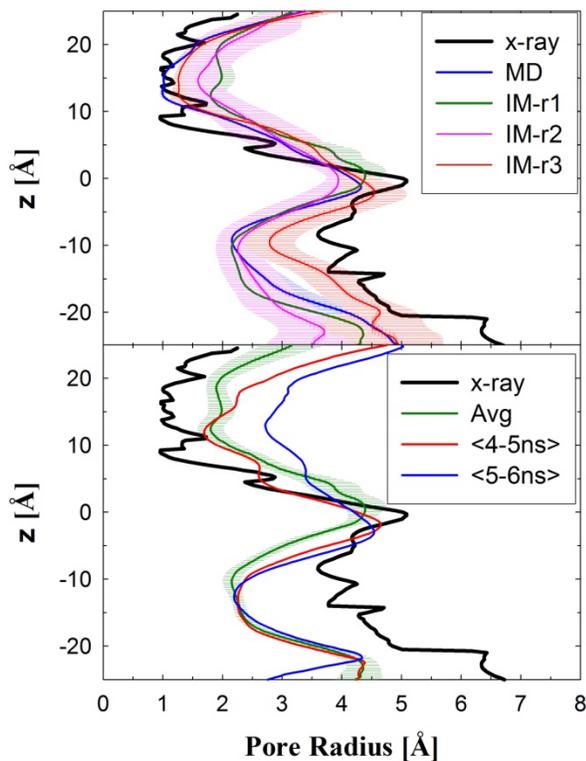


Figure 3.3: The pore radius profile of LacY's lumen is reported for the implicit membrane simulations, where the periplasmic half is for $z > 0$. The top panel is a comparison of several IM simulations with the x-ray structure pore radius. The pore radius profiles reported in the top panel are calculated based on the averages over the entire simulation described in Table 3.1. The bottom panel is for IM-r1 for the average over the simulation of 10 ns and averages of the 5 and 6 one-nanosecond blocks

The cytoplasmic half of the protein remained partially closed with a pore radius of ~ 2.25 Å for each run except IM-r3. Similar to the periplasmic half, this

involved a decrease in helix-helix distances between the C- and N-terminus domains of LacY. The largest change from the x-ray crystal structure involved H-V. Overall, H-IV and VIII had the largest RMSD for the runs with SGLD and the disaccharide, i.e., 2.1 ± 0.2 and 1.9 ± 0.4 Å, respectively. Moreover, the C-terminus domain of LacY had a larger RMSD (5.4 ± 0.5 Å) compared to the N-terminus domain (3.3 ± 0.3 Å).

Based on the results above, significant conformational changes in LacY exist for the cyto- and periplasmic halves of the protein with the SGLD implicit membrane simulations with Glu269 protonated and a disaccharide bound. Since the implicit membrane was used to enhance simulation time scales, two protein conformations (Figure 3.4) were selected and inserted into an explicit POPE membrane to obtain the final conformations of LacY. The protein is simulated in an explicit membrane to prevent any possible artifacts due to the implicit membrane by obtaining structures in a more realistic environment (see Methods). These protein conformations were obtained from the 3.4 and 5.5 ns snapshots of the IM-r1 simulation and represent an intermediate and periplasmic-open state based on the lumen radii, respectively. These were chosen carefully based on the comparison with the distances measured from the DEER experiments [76] to see if initial states in the explicit membrane are important to obtain final protein conformations open to the periplasm.

B. Explicit membrane simulations

The EX-r1 and EX-r2 simulations were carried out starting with the intermediate state and EX-r3 and EX-r4 were carried out starting with the periplasmic-open state. Protein backbone RMSDs from 1PV7 were calculated over

the course of the trajectories. For EX-r4, RMSD of the N- and C-terminal domains were 4.1 ± 0.2 and 5.1 ± 0.2 Å, respectively, which is consistent with the IM simulations in which the C-terminal domain shows more structural deviation from the crystal structure than the N-terminal domain. Locally, H-IV with Glu126 and H-VIII with the protonated Glu269 showed the maximum structural drift with average RMSDs of 1.5 ± 0.1 Å and 1.9 ± 0.2 Å, respectively.

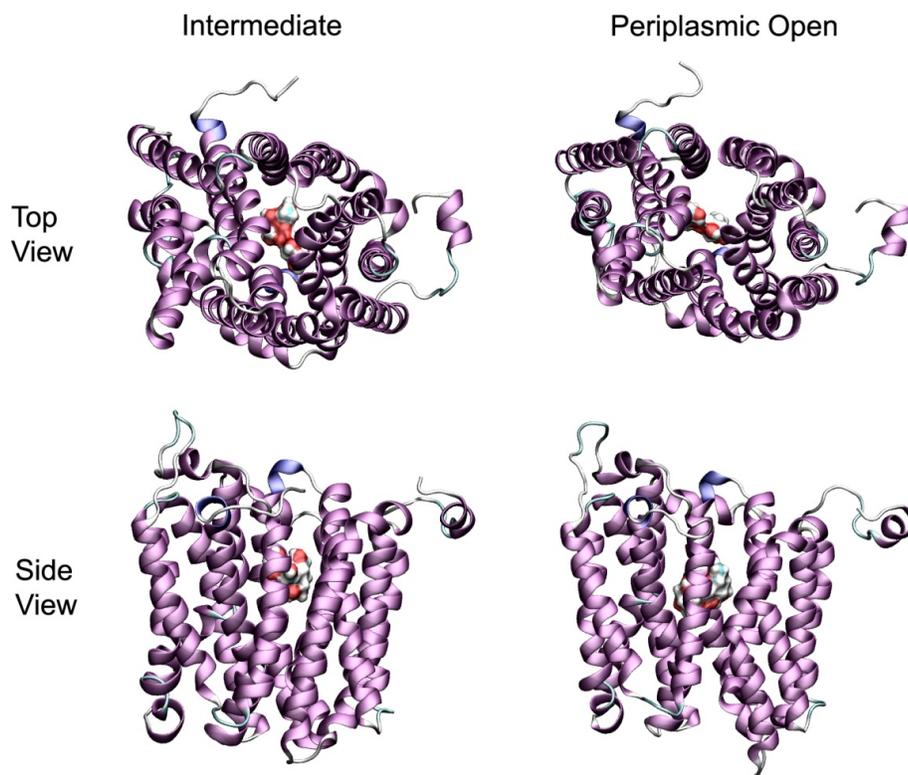


Figure 3.4: Screened structures used for the initial configurations of the explicit membrane simulations. The intermediate and periplasmic open structures were taken from the 3.4 and 5.5 ns coordinates of IM-r1, respectively.

All the simulation runs showed a decrease in the pore radius on the cytoplasmic side by ~ 2 Å from the crystal structure (Figure 3.5, top), which is similar to the implicit simulations. The pore radius increase on the periplasmic side was 1–

1.5 and 2.5–3 Å for EX-r1 and -r4, respectively. The average pore radius profiles for EX-r3 and EX-r4 were compared with the profile of the periplasmic-open state screened from IM-r1 (Figure 3.5, bottom). EX-r4 showed an increase of 1–1.2 Å in the pore radius while EX-r3 showed a decrease of ~1 Å from the periplasmic-open snapshot structure. Similarly, EX-r1 showed opening beyond the intermediate snapshot while EX-r2 showed closing (data not shown). In the inward-facing state of LacY (1PV7), the pore radius in the tightly closed periplasmic half is ~1 Å. For EX-r1 and EX-r4, the pore radius at the narrowest section in the cytoplasmic half is ~1.5 and ~2 Å, respectively. For these runs, the cytoplasmic side is not as tightly packed as the periplasmic side in the inward-facing crystal structure, but is sufficient to prevent the disaccharide, which has a minimum span of ~7 Å, to transport to the cytoplasm. Only water can transport from the cytoplasm to the protein lumen in this state and this occurs rarely for this nearly closed cytoplasmic structure.

On the cytoplasmic side, H-V showed the largest change in separation for all EX runs (Figure 3.6) and is consistent with the IM simulations. For EX-r1 and -r4, several helices showed significant increases for the periplasmic half of the protein. For EX-r2 and -r3, no significant change in the helix-helix separation was observed on the periplasmic side. For these runs, H-IV was the only helix to show some increase in separation with respect to the helices in C-terminal domain.

To summarize, for EX-r1 and EX-r4, the protein underwent major conformational changes beyond IM-r1 resulting in a significant periplasmic opening (Figure 3.7, left). The protein lumen closed on the cytoplasmic side (Figure 3.7, right)

and blocked the access of sugar on the cytoplasmic side, which is in accordance with the well-established alternating access model[115].

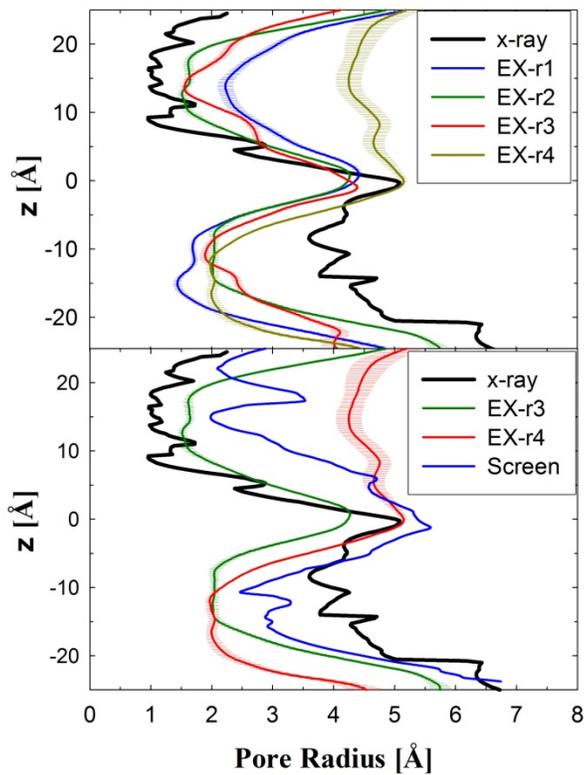


Figure 3.5: The pore radius profile of LacY lumen is reported for the explicit membrane simulations. The top panel is the comparison of the average pore radius profiles for the four explicit simulation runs with the x-ray structure pore radius. Bottom panel is the comparison of the pore radius profiles of EX-r3 and EX-r4 with the pore radius profiles of the periplasmic-open screened structure and the x-ray crystal structure.

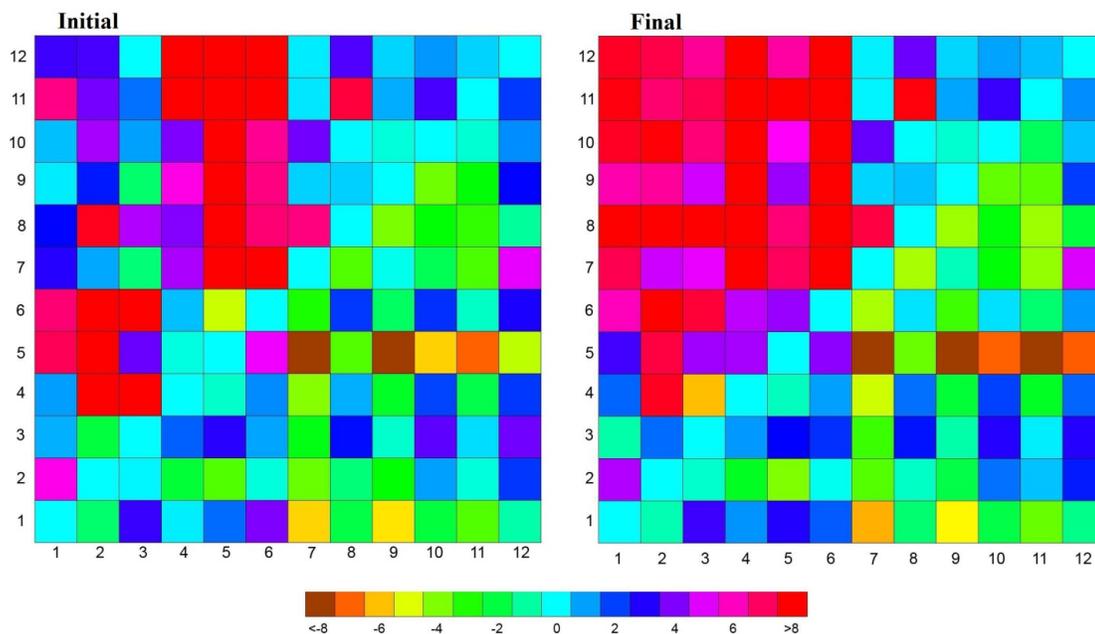


Figure 3.6: Average helix center distance deviation from the x-ray crystal structure (1PV7) for (left) first and (right) last nanosecond from the trajectory of EX-r4. All helix-helix pairs (# in rows and columns) are shown with the distances below the diagonal corresponding to the cytoplasmic half and periplasmic above. The legend or colorbar is shown on the bottom and is in Å.

3.3.2. Description of the transitions to the Outward-facing State of LacY

A full description of the of transition between the open states of this protein (Figure 1.4, step 3) is left for the discussion section but important structural and stabilizing interactions are briefly presented in this section. The pore radius profiles clearly indicate that two out of four explicit simulation runs (EX-r1 and -r4) resulted in opening of the protein structure on the periplasmic side (movie in supplementary material), while the other runs lacked any significant periplasmic opening. Figure 3.8 (top) shows the separation transmembrane helices of the periplasmic half ($z>0$) and opening of the protein lumen for EX-r4. Although the conformational changes of

LacY involve a mechanism more complex than rigid-body rotation of the two domains [20], the angle between the principal axis vector of the entire domain obtained from simulations and that from the crystal structure is calculated to quantify the pseudo-rotation of the domain. The principal axes of the N- and C-terminal domain are calculated based on the backbone atoms of the first six and the last six helices, respectively. For EX-r1, principal axis vectors of the N- and C-terminal domains made $11.3 \pm 1.6^\circ$ and $11.4 \pm 1.6^\circ$ angles with the corresponding crystal structure vectors. For EX-r4, the N-terminal domain rotated by $\sim 11^\circ$, whereas the C-terminal domain rotated by $\sim 20^\circ$ from the crystal structure. The domains rotate in directions opposite to each other so that the protein lumen closes on the cytoplasmic side and opens on the periplasmic side.

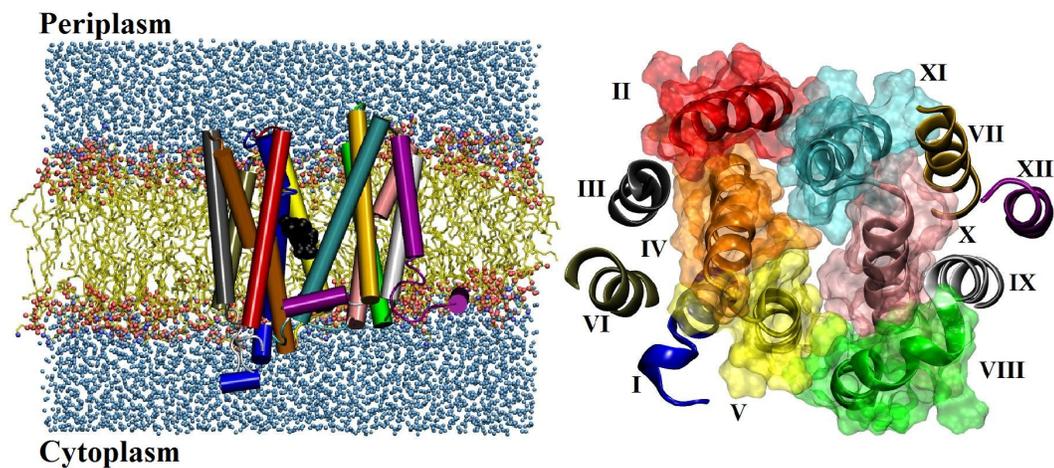


Figure 3.7: The outward-facing structure of LacY from EX-r4. (left) Snapshot of LacY embedded in a lipid bilayer. (right) The cytoplasmic halves of the transmembrane helices viewed from the cytoplasmic end. The residues on the helices forming the lumen are represented as molecular surfaces based on the van der Waals radii of the individual atoms. The color-coding of the helices is shown in right.

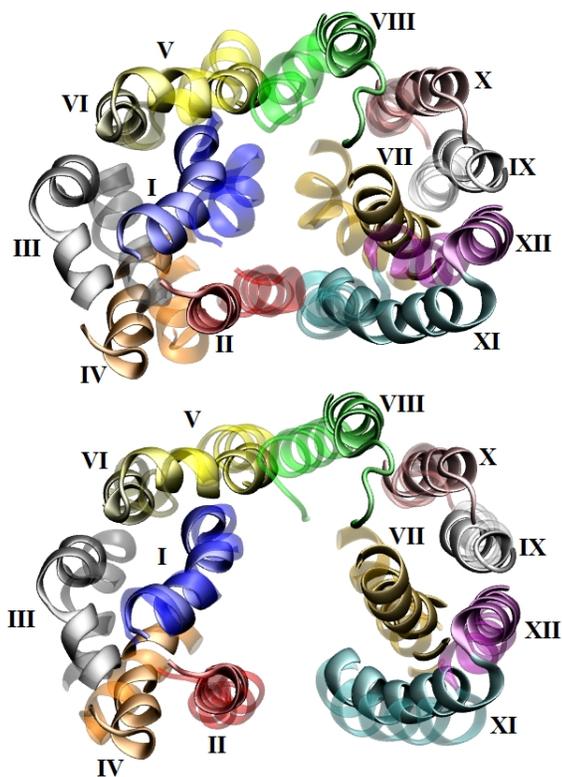


Figure 3.8: The periplasmic halves of the transmembrane helices ($z > 0$ from the COM of the protein) viewed from the periplasmic end. Solid helices indicate the periplasmic-open structure from EX-r4. Transparent helices represent (top) the crystal structure (1PV7) and (bottom) the structure from EX-r1.

Packing of some helices on the periplasmic side was stabilized by hydrogen bond interactions, which are listed in the **Table 3.2**. These interdomain hydrogen bonds are broken in EX-r1 and EX-r4 resulting in the opening of LacY on the periplasmic side. The backbone oxygen of Pro31 was involved in hydrogen bonding in EX-r2 and EX-r3 indicating the existence of close tertiary packing between H-I/H-VII near the periplasmic end. The other residue involved in hydrogen binding with Pro31, Asn245, has been shown to play an important role in gating the periplasmic pathway of LacY.[116]

Table 3.2: Hydrogen bond (HB) pairs that stabilize the cytoplasmic-open state. These hydrogen bonds existed for the runs that were periplasmic-closed (EX-r2/r3) and destabilized in the runs with a periplasmic-open state (EX-r1/r4). (SC-side chain, BB-backbone)

Residue Pairs	Donor-Acceptor Atoms	Helices
N245-P31	N _{SC} -O _{BB}	H-VII/II
S41-E374	O _{SC} -O _{BB}	H-II/XI
K42-Q242	N _{BB} -O _{SC}	H-II/VII
	N _{SC} -O _{BB}	

3.3.3. Comparison with experimental data

A. DEER and FRET Experiments

Previous DEER experiments were carried out on Cys mutants of LacY labeled with (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethiosulfonate (MTSL). The distances between the spin labels measured in the experiments were compared with the corresponding C_{α} - C_{α} distances calculated from the simulations. The correspondence between the two sets of distances is not one to one. However, the comparison throws light on cytoplasmic and periplasmic structural changes and distance changes between cytoplasmic- and periplasmic-open states may cancel out the effect of this bulky spin label if the orientation remains constant. The C_{α} - C_{α} distance changes from the cytoplasmic-open (x-ray structure) to the periplasmic-open state (DEER-based) are compared with the LacY simulations in Figure 3.9. The values reported are based on Gaussian fits to distance distributions of the pairs.

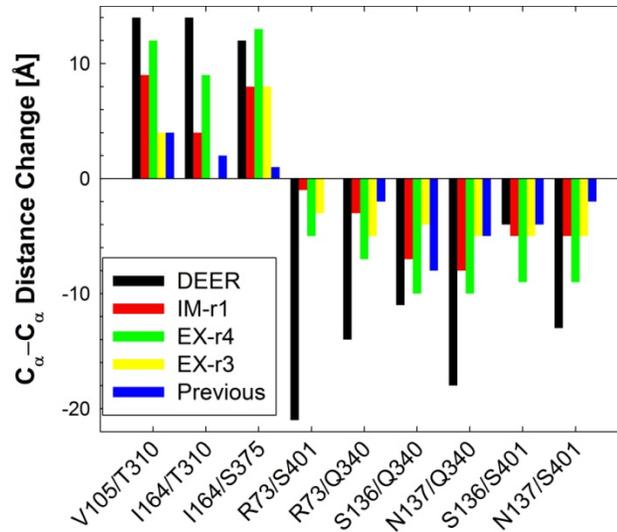


Figure 3.9: The C_{α} - C_{α} difference in distance between residue pairs relative to the x-ray crystal structure. Negative values indicate that the helix-helix distance has decreased. The distance is measured as the mean of the Gaussian peak fit to the distance distributions

On the periplasmic side, the distances increased significantly from IM-r1 for EX-r1 and EX-r4 (Figure 3.9). For EX-r4, increase in the distances from IM-r1 was 3-5 Å. The pairs Val105/Thr310 (H-IV/H-IX) and Ile164/Ser375 (H-V/H-XI) showed the maximum separation from the x-ray crystal structure by 12-13 Å. These two distances match considerably well with the DEER experiments. The Ile164/Thr310 distance (H-V/H-X) increased over the IM-r1 run by ~3 Å and ~7 Å from the cytoplasmic-open state but was slightly lower than experiment. For EX-r2 and EX-r3, the distances remained the same or decreased from IM-r1 for all periplasmic pairs.

On the cytoplasmic side, the distance between the pairs decreased for EX-r1 and EX-r4 compared to IM-r1. For EX-r4, this decrease was 2-5 Å for all the pairs. The distances Ser136/Gln340 and Asn137/Gln340 (both H-IV/H-X) showed the maximum decrease of 10 Å from the crystal structure. The distance change for Arg73/Ser401 (H-III/H-XII) was 5 Å, which was considerably smaller than the

experimental value of 21 Å, but all other distances from EX-r4 compared favorably with the experimental data. For EX-r3, most distances increased from IM-r1, but a few show no significant change or a slight decrease (Figure 3.9).

These results also qualitatively comply with the findings from the FRET experiments in which the cytoplasmic ends of H-III and H-XII move closer while the periplasmic ends of H-V and H-XI move apart upon sugar binding [75].

B. Cross-linking Experiments

Three paired double-Cys mutants (I40C/N245C, T45C/N245C and I32C/N245C) located at the interface of the N- and C-terminal domains near the periplasmic end were used in previous cross-linking experiments [77]. Homobifunctional thiol cross-linking reagents of different lengths and flexibilities were used to test the influence of cross-linking on the transport activity of a disaccharide, TDG. For the I40C/N245C double-Cys mutant, transport activity of sugar was almost entirely blocked with cross-linking reagents of length less than ~15 Å. With the flexible reagents MTS-14-O4-MTS (~17 Å) and MTS-17-O5-MTS (~22 Å), full transport activity of the disaccharide was observed. These cross-linkers suggest that the periplasmic opening of the protein should be between 15 – 17 Å for the outward-facing state.

To determine the size of periplasmic opening, C_{β} - C_{β} distances between the three residue pairs (I40/N245, T45/N245 and I32/N245) were calculated from our simulations. The experimental periplasmic open distances are the spacer arm distances of the cross-linking reagent and not the distance between C_{β} atoms. The

distance between the C_{β} atoms will be higher than the spacer arm distance and the equilibrium bond distance between C_{β} -S of cysteine (1.8 Å) is used as an offset. Therefore, C_{β} - C_{β} distances between the residue pairs should be between 18.6 and 20.6 Å. The results (Table 3.3) show that for EX-r1 and EX-r4, the separations for the residue pairs I32/N245, T45/N245 and I32/N245 agree favorably with the size of the opening indicated by the cross-linking experiments. These cross-linking experiments also indicate that the inter-thiol distance between I32C/N245C is close to ~6 Å in absence of the ligand, which is similar to C_{β} - C_{β} distance for the inward-facing crystal structure (1PV7) [13]. For EX-r2 and EX-r3 (periplasmic-closed runs), the mean values for C_{β} - C_{β} distance between I32/N245 are 5.8 Å and 6.5 Å, respectively. Also, the mean C_{β} - C_{β} distances between the other two pairs are close to the crystal structure.

Table 3.3: Mean C_{β} - C_{β} distances between residues (Å) used in the double-Cys mutant experimental studies.[77] The suggested C_{β} - C_{β} distance in the outward-facing state from cross-linking experiments are 18.6-20.6 Å.

Residue Pairs	Calculated mean distances				1PV7
	EX-r1	EX-r2	EX-r3	EX-r4	
I40-N245	20.5	10.4	12.1	21.2	8.9
T45-N245	17.0	8.1	9.7	17.7	8.1
I32-N245	17.2	5.8	6.5	18.3	6.9

C. Accessibility and Reactivity Experiments

Site directed alkylation of single-Cys mutants of LacY at various positions with small membrane-permeant alkylating reagents, N-ethylmaleimide (NEM) [115, 117-122] and tetra-methylrhodamine-5-maleimide (TMRM) [114], was previously measured to investigate the conformational changes in LacY upon sugar binding by quantifying the accessibility and/or reactivity at those residues. The reactivity of a Cys residue can be limited by close contacts between transmembrane helices in its surroundings and a change in the reactivity on sugar binding indicates an alteration in the environment at that position in the sequence [114, 115]. Sugar binding was shown to result in an increase in the reactivity of Cys-mutants that predominantly lie in the periplasmic half while a decrease in the reactivity was observed at positions predominantly lying in the cytoplasmic half. For the protein conformations obtained from our simulations, the contact surface areas were calculated for the residues at these positions as a measure of the close contacts (see Methods). These areas were compared with the contact surface areas of the respective residues in the crystal structure to assess the change in environment of these residues (Table 3.4). As the experiments were performed on the Cys-mutants, the bulky side chain atoms beyond the C_β were neglected when calculating the surface areas.

On the periplasmic side, most (85%) of the residue positions that had an increase in Cys reactivity [114, 115] also had an increase in the contact surface area for the EX-r1 and -r4 (Table 3.4). On the cytoplasmic side, some residues that had a decreased Cys reactivity did not show a decrease in the contact surface area. This may be a result of incomplete cytoplasmic closure observed during the simulations.

Table 3.4: Changes in contact surface areas compared to 1PV7 for positions that showed significant change in NEM and TMRM activity on sugar binding.

Helix	Increased surface area compared to 1PV7		Decreased surface area compared to 1PV7		Increased reactivity on sugar binding	Decreased reactivity on sugar binding
	EX-r1	EX-r4	EX-r1	EX-r4		
I	8, 12, 14, 17, 24, 28 – 32, 11, 15, 21, 22, 27	8, 12, 24, 28 – 32, 21, 22, 27	25	11, 14, 15, 17, 25	8, 12, 14, 17, 24, 25, 28 - 32	11, 15, 21, 22, 27
II	42, 44, 45, 49, 53	42, 44, 45, 49, 53	60, 70, 71	60, 70, 71	42, 44, 45, 49, 53, 70, 71	34, 60
III	96	96	100	100	96, 100	81, 84, 86, 87, 88
IV	122	122, 136	136	-	136	122
V	158-160	158-160	141, 145, 148, 157	141, 145, 148, 157	157-161	141, 145, 148
VII	241, 242, 244 - 246, 248	241, 242, 244 - 246, 248	-	-	241, 242, 244 - 246, 248	-
VIII	264, 265	264, 265, 272	272	-	265	264, 268, 272
IX	308	308	291, 295, 298,	291, 295, 298,	291, 295, 298, 308	-
X	315, 329	315, 329	327, 331	327, 331	315	327, 329, 331
XI	359, 362-364	359, 362-364	356, 357, 361	357, 361	359, 361-364	356, 357

Site specific labeling with methanethiosulfonate ethylsulfonate (MTSES), a small hydrophilic, membrane-impermeant thiol reagent was used to determine the change in water accessibility at various positions on sugar binding [117-123]. The positions that showed a change in water accessibility were determined. To compare with the MTSES accessibility studies, water accessible surface areas of various residues were calculated from our simulations and were compared with the corresponding surface area values for 1PV7 (**Table 3.5**). For EX-r1 and EX-r4, there is near perfect agreement for the positions that measured an increase or decrease in accessibility on sugar binding.

Table 3.5: Changes in water accessible surface areas compared to the crystal structure (1PV7) for positions that showed change in MTSES accessibility on sugar binding.

Increased surface area compared to 1PV7		Decreased surface area compared to 1PV7		Increased MTSES accessibility	Decreased MTSES accessibility
EX-r1	EX-r4	EX-r1	EX-r4		
18, 32,	18, 32,	60,326	60	18, 32, 38,	60, 326
38, 42-44,	38, 42-44,			42-44, 241-	
241-246,	241-246,			246, 248,	
248, 249,	248, 249,			249, 265, 327	
265, 327	326, 265, 327				

3.3.4. Comparison with other computational models of periplasmic-open LacY

Our model was also contrasted against a periplasmic-facing model obtained by swapping the conformations of inverted topology repeats identified in LacY's N- and C-terminal domains [79] (Figure 3.10). This approach by Radestock and Forrest exploits an inherent symmetry in LacY's two domains to generate a periplasmic-open model based on the cytoplasmic-open crystal structure of LacY-C154G mutant [13]. They also generated a LacY model based on the periplasmic-open crystal structure of FucP [17]. Protein backbone RMSDs of our model from these two models are 4.97 and 4.87 respectively. The cytoplasmic half of our model showed greater deviation (5.12 and 4.94 respectively) compared to the periplasmic half (4.31 and 4.94 respectively). This is consistent with our hypothesis of an incomplete cytoplasmic-closure in our simulations. Although this comparison is not exact because stability of the inverted-topology and FucP based models is not tested in a membrane environment.

3.3.5. Sugar binding

Since conformational changes of the protein are coupled to sugar binding [76, 115] in the final section of the result we briefly describe sugar binding. Although these results are based on several explicit membrane simulations, more detailed and comprehensive studies on binding are needed to obtain a complete picture of sugar binding in the periplasmic open state and presented here may not be the entire story. The interactions between the disaccharide, $\beta\beta$ -(Galp)₂, with LacY are similar for all EX runs. The disaccharide binding pocket in EX-r1 and -r4 (outward-facing LacY) is

similar to the IM simulations. The disaccharide formed hydrogen bonds with Glu126, Arg144, Asp237, Glu269, Lys358, Phe118, Asn119, Asp240, His322 and Tyr350, which is consistent with our previous disaccharide binding studies [69]. Separate from these well-known sugar-protein interactions, the disaccharide formed hydrogen bonds with Gln60 and Phe354. Both these residues have been shown to be important in sugar transport [117, 120, 124, 125]. For EX-r1, $\beta\beta$ -(Galp)₂ escaped the binding pocket during the last 5 ns of simulation. It migrated ~ 7 Å towards the periplasm after the protein started opening on the periplasmic side and formed hydrogen bonds with Ser53 and Asn119 in this binding conformation with the importance of the former in the transport cycle confirmed by experiments [120]. Similarly for Ex-r2, where LacY started opening on the cytoplasmic side, the disaccharide escaped the binding pocket and migrated towards the cytoplasm. The disaccharide was found to be undergoing a change in its orientation in order to escape on either sides of the membrane. In the binding pocket, the two sugar rings are aligned nearly parallel to the membrane, whereas the rings become perpendicular to the membrane when the disaccharide starts to escape the protein lumen as observed by Jensen et al. [74] Despite the differences in sugar binding when comparing the cytoplasmic- and periplasmic-open states, our simulations agree with the alternating access model [115] in which the same binding site is believed to be accessible to the sugar molecules in inward and outward-facing states. Moreover, in the outward-facing state, the protein lumen is closed on the cytoplasmic side preventing the sugar from escaping to the cytoplasm.

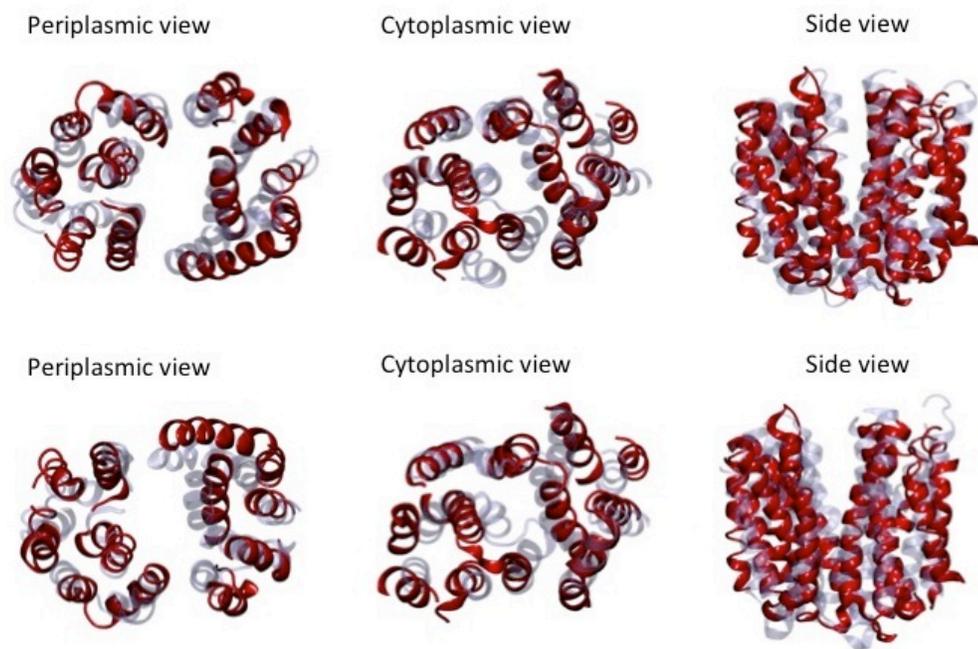


Figure 3.10: Comparison of our model for periplasmic-open LacY with a model based on inverted topology repeats (top panel) and a model based on the crystal structure for periplasmic-open MFS protein FucP (bottom panel). Our model is represented in solid red while other two models are shown in transparent color. In the periplasmic half, there is a good match between our model and these two models although the cytoplasmic closures of these two models are greater than model.

3.4. Discussion

LacY undergoes widespread conformational changes during substrate transport based on various biochemical experiments [20]. Previous MD simulations have resulted in significant conformational changes on the cytoplasmic side compared to the crystal structure. A decrease in pore radius was observed in the cytoplasmic half for LacY with Glu269 protonated in a 10-ns simulation [70] and with Glu325

protonated in 50-ns simulation. Our previous EX simulations with Glu325 protonated [69] also resulted in similar decrease in the cytoplasmic pore radius by ~ 2 Å. However, negligible pore radius changes in the periplasmic half were observed during these simulations.

Several control simulations were performed to verify that our methods open the protein on the periplasmic side only when a sugar is bound and the proper protonation of a titratable residue. Only simulations with Glu269 protonated and a sugar bound resulted in conformational changes in the periplasmic half. Simulations without the sugar (*apo*), Glu325 protonated, or His322 protonated, all maintained a conformation prevent sugar transport to the periplasm. Although this was expected for simulations without the sugar or Glu325 protonated (Figure 1.4) [20], our simulations suggest that protonation of His322 does not trigger a conformational change.

All IM simulations with Glu269 protonated and a sugar bound resulted in greater periplasmic conformational changes in LacY compared to previous EX simulations because of enhanced sampling. The structural changes were enhanced by the use of SGLD instead of MD and an implicit bilayer. The enhancement alone appears to allow for structural changes and sugar binding is the main trigger for protein structural changes because LacY-*apo* remained closed on the periplasmic side. Structures were screened from the IM simulations and inserted into an explicit bilayer so that a more realistic environment can be used to obtain a final set of conformations (Figure 3.8, bottom). Essentially, it was our aim to use the implicit membrane simulations with SGLD to overcome a transition state or energy barrier

between the two open states in a reasonable amount of time. If the screened structures lie in the metastable transition state region in the free energy profile, then there would be an equal probability to go back to the inward-facing state or further open to the outward-facing state assuming a single transitional barrier. For each of the screened structures, one simulation resulted in a fully periplasmic-open state, while another started closing on the periplasmic side and opening on the cytoplasmic side. Certainly more runs would be required to determine if the screened structures are at or near a transition state, but these results do indicate the ability of this approach to lead to a periplasmic-open state and also return to a conformation closed to the periplasm.

The pore radius profile of LacY's lumen (Figures 3.3 and 3.5) and helix-helix separation distance maps (Figure 3.6) clearly demonstrate a protein structural change. An interesting point to be noted is the average pore radius profile for EX-r4 (Figure 3.5, top). The cytoplasmic half of this profile resembles greatly with the periplasmic half of the pore radius profile of the crystal structure and vice versa. Besides the similarities in the two halves, the two profiles are analogous in the sugar binding region ($\sim z = 0 \text{ \AA}$).

The helix-helix maps (Figures 3.6) indicate that on the periplasmic side structure changes involve significant intradomain as well as interdomain helix movements, whereas only the interdomain displacements are dominant on the cytoplasmic side. The periplasmic conformational changes indicate a more complex mechanism involving significant intradomain changes to yield a final outward-facing state. However, if the initial and final structure of LacY is only considered, then the rigid body rotation of LacY is consistent with the helix map. Therefore, intradomain

changes are important as the protein changes between open states, but in the end these collapse back to similar helix packing. More specifically, H-V showed the largest change in the separation on the periplasmic and cytoplasmic sides. The C154G mutant of LacY, which is trapped in an inward-facing conformation, has Gly154 (H-V) at the domain interface. Gly154 is believed to cause tighter packing of H-V near the interface hence hindering the movement of H-V which was found to be significant during the conformational transitions from our simulations.

The C_{α} - C_{α} distances calculated from EX-r1 and EX-r4 agree favorably with the DEER experiments (Figure 3.9) [76]. On the periplasmic side, the distance changes for pairs V105/T210 and I164/S375 are in agreement with the experimental values. On the cytoplasmic side, the distance changes differ from some of the experimental values indicating a possible partial incomplete cytoplasmic closure. However, the cytoplasmic half is closed to the extent that it prevents and transport of sugar to the cytoplasm (Figure 3.7, right). The largest discrepancy between the simulations and the experiments is for the cytoplasmic pair R73/S401 suggesting that H-II/XII is the last helix pair to close. Moreover, the distances calculated from the simulations are C_{α} - C_{α} distances while those measured in the experiments are the distances between the spin labels (MTSL). The approximate distance between the label and the C_{α} atom is 8 Å, which may result in the maximum deviation of ± 16 Å. For example, the distance changes reported from the DEER experiments differ considerably for the pairs S136/S401 and N137/S401 even though there is little distance separating the two residues (Ser136 and Asn137). Similar difference is present between the pairs S136/Q340 and N137/Q340. It was suggested that these

striking differences may be a result of significant rotation of the helices including winding and unwinding at the ends [76]. No such conformational changes were observed during our simulations. The distances calculated from EX-r4 are close to the mean of the DEER measurements for the S136(N137)/Q340 and S136(N137)/S401 pairs. This suggests that the orientation of the spin label may play a role in the DEER measurements. To explicitly investigate the effect of spin labels, simulations with labeled residue pairs will be required and are planned as future work.

In addition to DEER measurements, our simulations are in excellent agreement with the amount of the periplasmic opening measured using cross-linking experiments [77] (Table 3.3) and site-directed alkylation with alkylating agents of various sizes [126]. Based on these helix distance measurements, it can be concluded that the periplasmic opening observed from EX-r1 and -r4 closely matches the outward-facing state.

The periplasmic opening of LacY obtained from our simulations is also in compliance with NEM [115], TMRM [114], and MTSES [117-122] labeling experiments. 85% of residues that had a significant increase in NEM and TMRM activity on the periplasmic side showed an increase contact surface area in EX-r1 and -r4. On the cytoplasmic side, the closure obtained from our simulations may not be complete thus preventing the expected decrease in the surface area values for some residues, especially those on H-III and -IX (Table 3.4). However, we suggest that the steric effect of sugar binding at Gln60 may be a reason behind the reduced NEM [120] and TMRM [114] reactivity observed with the Cys replacement at that position. Although the comparisons with DEER and NEM/TMRM accessibility studies

indicate a possible incomplete cytoplasmic closure for our outward-facing LacY model, the pore radius profiles indicate that the closure is sufficient to block the access of sugar to the cytoplasm as per the alternating access model [115]. Inaccessibility of the disaccharide from the cytoplasmic end in the outward-facing state can be clearly seen in Figure 3.7 (right) where the protein lumen is closed on the cytoplasmic side.

The interdomain hydrogen bonds N245/P31, S41/E374 and K42/Q242 are destabilized in EX-r1 and -r4 while they are maintained in EX-r2 and -r3 (Table 3.2). Moreover, these hydrogen bonds stabilized the helix-packing in previous EX simulations [69] as well as in the initial parts of our IM simulations. These interdomain hydrogen bond interactions appear to be crucial in stabilizing the inward-facing state. Side chains of Asn245, Ser41, Glu374, Lys42 and Gln242 are involved in stabilizing the inward-facing structure and mutations in these residues may destabilize the inward-facing state. Moreover, breaking of these hydrogen bonding interactions in the outward-facing state explains the increase in NEM[115] and TMRM[114] labeling at residues 31, 42 and 245 on sugar binding. In addition, any change in the interactions between H-VII and the surrounding helices at residue 242 was found to inhibit the transport activity which is consistent with our findings [127].

3.5. Insights into LacY's transport mechanism

The findings from our simulations are summarized below to give mechanistic insights into the sugar transport cycle and conformational changes of LacY from the

outward to inward facing state (Figure 1.4, step 3). This is also shown as a cartoon in Figure 3.11.

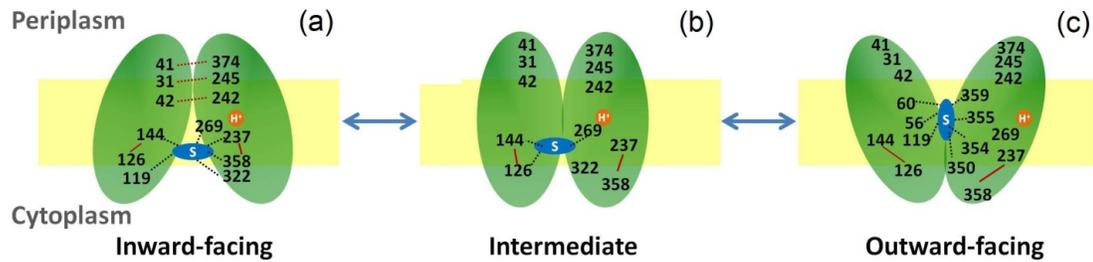


Figure 3.11: A proposed mechanism for the conformational transition to the outward-facing state starting with the inward-facing crystal structure (1PV7) [13]. Important residues are labeled. Interactions between protein residues are represented in red and protein-sugar interactions are represented in black. Salt bridges and hydrogen bonds are represented as solid and dotted lines respectively.

- (a) Binding of a disaccharide and translocation of proton to Glu269 is necessary to trigger the conformational transformation of LacY. The interdomain hydrogen bonds N245/P31, S41/E374 and K42/Q242 stabilize the inward-facing structure of LacY. In the binding pocket, two rings of the disaccharide are oriented nearly parallel to the membrane.
- (b) In the intermediate state, the disaccharide lies in the binding pocket but is weakly bound to the protein at Glu126, Arg144 and Glu269. The structure stabilizing interdomain hydrogen-bond interactions is broken. The salt bridges E126/R144 and D237/K358 remain intact during the transition
- (c) The disaccharide remains in the binding pocket during the conformational change which is in accordance with the alternating access model [115] but undergoes an orientational change such that the two rings are nearly

perpendicular to the membrane. It interacts with Ser56, Gln60, Asn119, Tyr350, Phe354, Cys355, and Gln359 in this binding conformation, thus indicating there are multiple sugar-LacY binding conformations.

In summary, we were able to obtain significant structural changes on the periplasmic as well as the cytoplasmic side of LacY. The periplasmic changes resulted in the opening of the protein lumen to outside the cell. The outward-facing state from our simulations agrees favorably with a multitude of experimental data. The inward-facing state is stabilized by certain interhelical and interdomain interactions and destabilizing these interactions may help crystallize the outward-facing state. Our simulations are in compliance with the alternating access model [115] for the sugar binding but some additional residues that are important in sugar binding are determined. The importance of these residues in sugar binding and conformational change was confirmed by experiments.

Chapter 4: Thermodynamics of sugar binding to LacY

4.1. Introduction

Large numbers of experimental studies have been conducted to study substrate binding in LacY prior to and after the determination of LacY's x-ray crystal structure. Indirect experimental studies including NEM labeling [20] and Cysteine scanning mutagenesis [128] identified LacY's residues that participate in sugar binding. These experiments revealed that the carboxyl group of Glu126 (H-IV) and guanidine group of Arg144 (H-V) are essential for LacY's function. Any neutral replacements at these positions abolish substrate transport, therefore it was suggested that they form a charged pair in the *apo* state of LacY [13, 20]. Trp151 (H-V) also plays a critical role in sugar binding and aromatic mutations at this position (W151Y and W151F) do not alter the substrate transport kinetics. Moreover, the steady-state fluorescence spectrum of Trp151 and fluorescence quenching experiments [30] indicated that in presence of a substrate, Trp151 becomes less accessible to the solvent and possibly forms a stacking interaction with the substrate. Site-specific mutations at Glu269 indicated that carboxyl group at this position is critical for substrate binding and Glu269 may be involved in coupling of proton and sugar translocation [20].

The x-ray crystal structure of LacY was determined first for a mutant C154G (PDB ID: 1PV7) that was arrested in the cytoplasmic-open state in presence of a substrate TDG [13]. In the crystal structure (Figure 4.1(left)), Arg144 forms a bidentate hydrogen bond (H-bond) with O3 and O4 of the galactoside. There are no direct interactions between Glu126 and the sugar molecule but it possibly forms

water-mediated H-bonds with O6 atom of the galactoside. Asp237 (H-VII) and Lys358 (H-XI), which form a charged pair, are also involved in sugar binding. It was suggested that although Asp237 and Lys358 assist in substrate binding, they are not essential for substrate transport [20, 129].

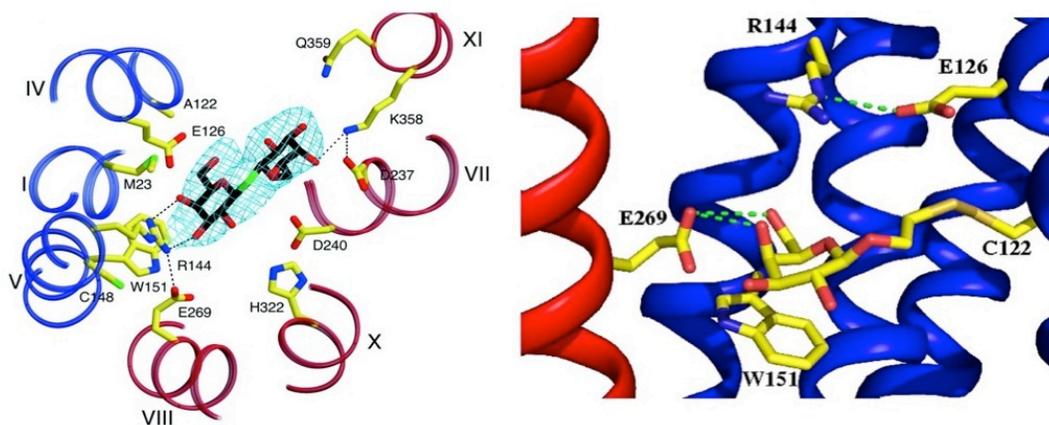


Figure 4.1: Substrate binding conformations in LacY's x-ray crystal structures. (Left) TDG forms H-bonds with Arg44 and Lys358. Glu126, Glu269, and Trp151 are in the vicinity of the sugar. The picture is adopted from Figure 4 from Guan and Kaback [20]. (Right) MTS-gal covalently bound to Cys122 of LacY-A122C mutant. MTS-gal can be seen forming a H-bond with Glu269 and an aromatic stacking interaction with Trp151. The picture is adopted from Figure 3 from Chaptal et al. [21].

A crystal structure of LacY was later determined for a single-Cys mutant at 122 (PDB ID: 2Y5Y) [21]. This mutant was inactivated by a suicide substrate methanethiosulfonyl-galactopyranosides (MTS-gal) that covalently binds to Cys122. The stacking interaction between Trp151 and the galactoside can be clearly seen in the crystal structure (Figure 4.1 (right)). Glu269 is in close proximity of Arg144 and Trp151 thus indicating a possible energetic linking between the N- and C- terminal domains of LacY. It was suggested that the galactoside position in this crystal

structure occurs when the substrate is either entering or leaving the protein lumen. Moreover, there were no direct H-bonds between the galactoside and Asp237 or Lys358 supporting the claim that these residues merely assist in substrate binding rather than playing critical roles.

MD simulations by Klauda and Brooks [130] on the binding of $\alpha\beta$ - and $\beta\beta$ -(Galp)₂ indicated that besides the residues observed in the two crystal structures (1PV7 and 2Y5Y), Phe118, Asn119, Asn240, His322, Glu325, and Tyr350 were involved in sugar binding. These studies reported higher binding enthalpy for $\alpha\beta$ -(Galp)₂ compared to $\beta\beta$ -(Galp)₂, which was consistent with the experimental findings that the binding affinity of an α -anomer is greater than β [131, 132]. Only a subtle alteration of the anomeric state of sugar resulted in significant change in its binding enthalpy. There were peculiar differences in the binding structures of the two anomers. Glu126 and Glu269 were observed to have stronger interactions with $\alpha\beta$ -(Galp)₂, whereas Phe118 had stronger interactions with $\beta\beta$ -(Galp)₂. There were multiple stable binding conformations for the two sugars that were different from the crystal structures. In the IM-EX simulations, focused on conformational changes in LacY [81], $\beta\beta$ -(Galp)₂ was observed to be forming H-bonds with residues that were not previously reported including Phe354, Ser56, and Ser53 (Figure 4.2), thus corroborating the hypothesis that sugar molecules bind to LacY in multiple stable binding conformations.

Isothermal titration calorimetry studies were performed on LacY-wt and LacY-C154G to describe the thermodynamics of sugar binding [133]. Binding free energy values of thio-di-galactoside (TDG) and p-nitrophenyl- α -D-galactopyranoside

(NPG) to LacY-wt and LacY-C154G were similar but binding of sugars to LacY-wt was shown to be entropically driven in contrast to binding to LacY-C154G that was enthalpically driven. The C154G mutant of LacY is severely restricted in the cytoplasmic-open state [13, 133] and allows for a tighter binding of substrate whereas LacY-wt is highly dynamic in nature [20, 133] for which the major contribution to binding free energy comes from multiple substrate-bound conformational states [133].

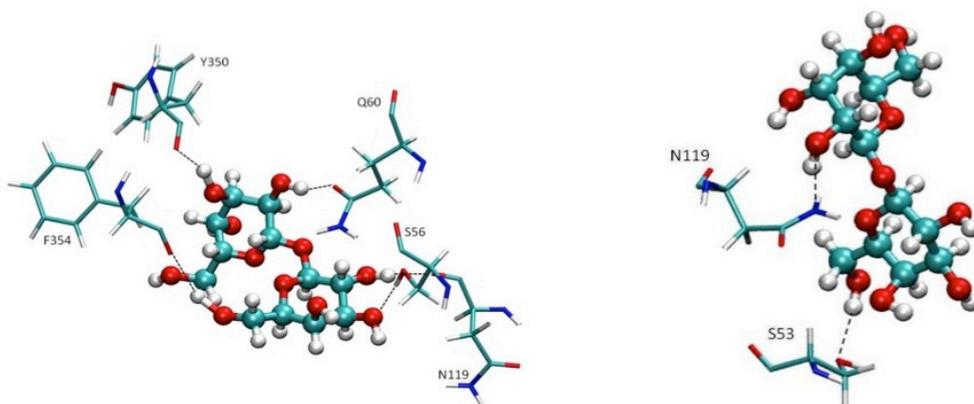


Figure 4.2: Binding conformations of in the periplasmic-open state. Sugar is interacting with Phe354, Tyr350, Gln60, Ser56, Asn119, and Ser53. The figure is adopted from Pendse et al. (2010) [81].

The focus of this work was to quantify binding affinities of different sugars in the cytoplasmic- and periplasmic-open LacY to gain insights into the sugar binding mechanism at atomic level details. To allow for a direct comparison with the experiments on anomeric binding [31], studies were performed on α - and β -anomers of methyl-D-galactopyranoside. We also studied binding of NPG, $\alpha\beta$ -(Galp)₂ and $\beta\beta$ -(Galp)₂, which is an analog of TDG in which glycosidic sulfur from TDG is replaced

with an oxygen. NPG and TDG have been used as high-affinity binders in several studies on LacY [24, 133]. Binding of α/β -lactose (a natural substrate of LacY) and α/β -maltose (a non-binder) was also studied. The illustrations of the sugars studied are shown in Figure 4.3.

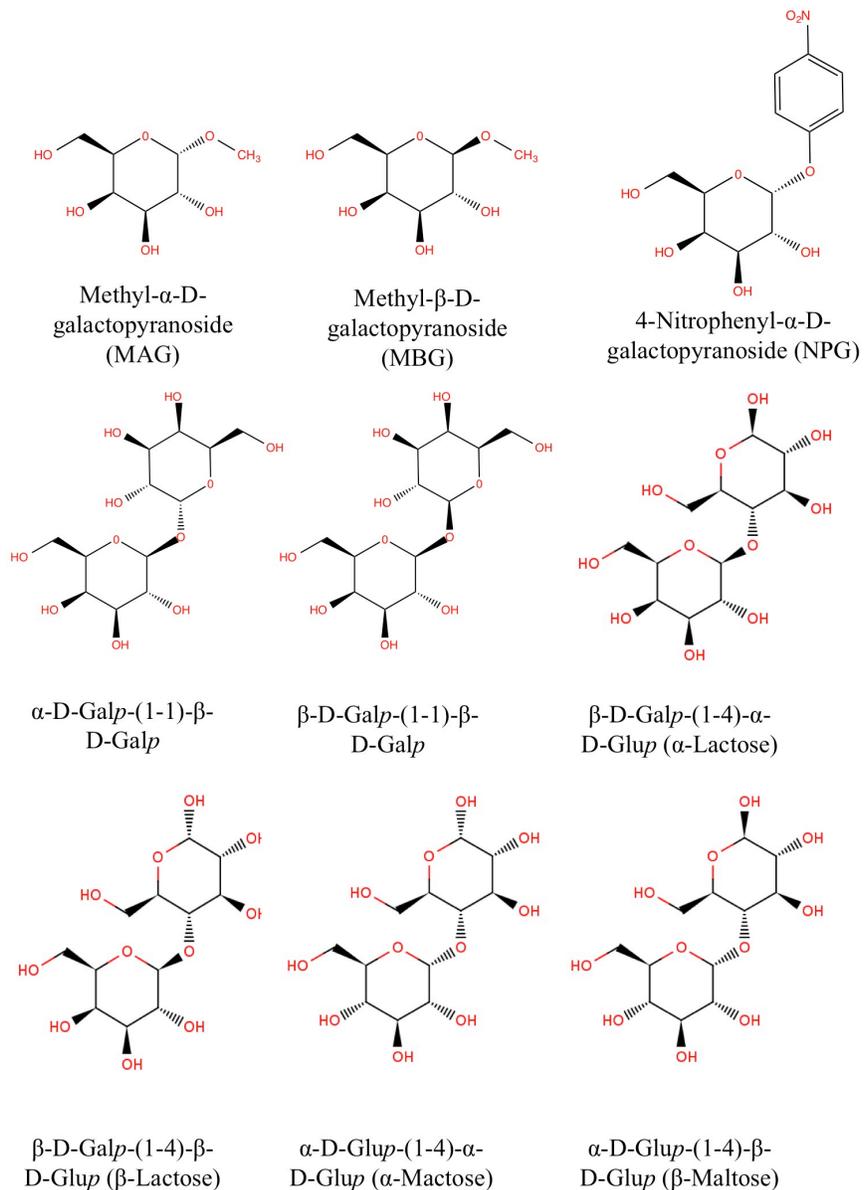


Figure 4.3: Illustrations of the sugars studied in this work.

4.2. Computational design and Methods

4.2.1. Binding structure determination

A. Molecular models for LacY and sugars

Initial protein structures for determining the substrate binding conformations in the cytoplasmic-open state and periplasmic-open states were obtained from the x-ray crystal structure (PDB ID: 2V8N) [14] and the periplasmic-open model for LacY [81], respectively. Molecular models for the natural sugars (lactose and maltose) were obtained from the ZINC database (<http://zinc.docking.org/>). Models for $\alpha\beta$ and $\beta\beta$ - $(Galp)_2$ were built using glycam carbohydrate builder facility (<http://glycam.ccruc.uga.edu/ccrc/>). To build NPG, the coordinates for galactoside and nitrophenyl moieties were obtained from glycam carbohydrate builder and from the work of Klauda and Brooks [134], respectively. The parameters for the nitrophenyl group in NPG were obtained from Klauda and Brooks [134]. The missing dihedral parameters were obtained by fitting dihedral profiles to MP2/cc-pVDZ.

B. MD simulations of the sugar-protein complexes

Since sugar molecules bind to LacY in multiple conformations, it is important to acquire all the possible binding modes to obtain meaningful binding free energy values. Equilibrium MD simulations were performed for each sugar molecule separately. Docking approach was used initially to determine the starting points for MD simulations. SWISSDOCK, a CHARMM force field based docking web server [135, 136] was used to dock ligands to initial protein structures in the cytoplasmic- and periplasmic-open states. Independent MD simulations with the docked

conformations of lactose and maltose converged to binding conformations similar to that observed in the x-ray crystal structure (1PV7) [13]. The alternative binding regions that were observed from our previous simulations (Chapter 3) were not captured during these simulations. My hypothesis is that since substrate binding to LacY involves an induced-fit mechanism [22], the initial guesses obtained from the docking calculations were not able to sample the entire binding region effectively. Therefore sugar molecules were randomly placed inside the protein lumen in distinct translational and rotational orientations relative to the protein to form the starting conformations for MD simulations. Sufficiently long relaxation of these structures should sample the entire binding region.

To achieve greater computational speeds, MD simulations in the cytoplasmic-open state were performed using an implicit membrane model, which was used successfully in our previous work. A hydrophobic potential [105], which varies exponentially across the protein surface, was used similar to our previous work [81]. The TIP3P explicit water model, which is consistent with the CHARMM parameters [93, 94], was used since water mediated protein-sugar and protein-protein interactions are crucial in LacY's function [69, 70, 73]. The sugar-protein system embedded in an implicit membrane is shown in Figure 4.4 (a).

All the implicit membrane simulations were performed in CHARMM simulation package [85, 86] with a time step of 2 fs and temperature of 293.15 K. The CHARMM force fields were used to describe the protein and sugars [89-91, 137, 138]. Production runs were performed for 10-15 ns for every starting structure.

Protein structures in the cytoplasmic-open state had Glu325 protonated since that is believed to be the protonation state of LacY in the cytoplasmic-open state [20].

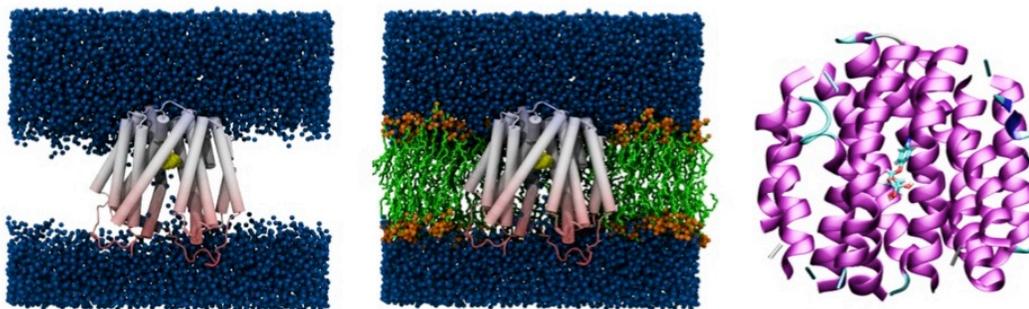


Figure 4.4: (a) LacY inserted in an implicit membrane. (b) LacY inserted in a hydrated explicit membrane. (c) LacY in a GSBP setup, where only the atoms in the 15Å radius sphere centered at the ligand are represented explicitly. Remaining atoms of the system are treated using continuum electrostatics.

MD simulations of periplasmic-open LacY were performed using an explicit membrane, as these were first MD studies on substrate binding in the periplasmic-open state of LacY. A six-lipid complex model for the *E. coli* membrane (Top6) consisting of lipids containing a cyclopropane moiety was used for these simulations [139]. The Top6 membrane model represents the *E. coli* membrane more accurately compared to single or two lipid membrane models in terms of its hydrophobic thickness, density and rigidity [139]. The computational speeds for explicit membrane simulations in NAMD were equivalent to the implicit membrane simulations in CHARMM. The LacY-membrane complexes (Figure 4.4 (b)) were built using the membrane builder facility on CHARMM-GUI [46-48]. All the explicit membrane simulations, after inserting sugars inside the protein lumen, were performed in the NAMD simulation package [62]. The explicit simulations were performed in the NPT

ensemble at pressure of 1 bar and temperature of 293.15K. The production runs were carried out for 15-30 ns. Protein structures in the periplasmic-open state had Glu269 protonated, which has been suggested to be the preferred protonated state open to the periplasm [70, 100].

C. Screening of binding structures:

The hydrogen bonding patterns of the sugar molecules and all the participating protein residues were analyzed for all production runs. The binding conformations that were kinetically distinct were chosen. In each distinct binding conformation, different sets of protein residues bind to the sugar molecule. The distinctness of binding conformations was confirmed by ensuring that they have sufficiently different translational positions relative to protein. Only the conformations that were stable for at least 2-3 ns were considered for further free energy calculations.

4.1.2. Absolute binding free energy calculations

The alchemical free energy perturbation/molecular dynamics (FEP/MD) protocol as described in Chapter 2 was used for the binding free energy calculations.

A. Reduced atom boundary potentials

To reduce computational cost, only the atoms in 15Å radius of the bound sugars were treated explicitly. Remaining atoms in the system were represented by

the generalized solvent boundary potential (GSBP) [57]. The use of GSBP reduces the size of the system to ~6000 atoms compared ~100,000 atoms in the explicit membrane system (Figure 4.4(c)). For the FEP calculations in the solvent, the sugars were solvated in a sphere of radius ~18 Å. Remaining bulk is represented by the spherically symmetric boundary potential (SSBP) [59].

B. Use of translational, rotational and conformational restraints

Three translational and three rotational degrees of freedom of the substrate relative to the protein were restrained using a harmonic potential as described in Chapter 2. The force constants used for the distance and angle/dihedral restraints were 10 kcal/mol/Å² and 200 kcal/mol/rad², respectively. The ligand was constrained near its bound conformation by applying harmonic RMSD restraint with a force constant of 10 kcal/mol/Å².

Table 4.1: Values of the coupling parameter (λ) used for calculation of different free energy contributions.

Contribution	Coupling parameter (λ)
Electrostatic	0.0, 0.1, 0.2, 0.3, 0.4, 0.4, 0.6, 0.7, 0.8, 0.9, 1.0
Dispersive	0.0, 0.25, 0.5, 0.75, 1.0
Repulsive	0.0, 0.2, 0.3, 0.4, 0.4, 0.6, 0.7, 0.8, 0.9, 1.0
Restraints	0, 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0

C. Double decoupling

A coupling parameter (λ) was used to turn ON /OFF the specific interactions and restraints on the substrate. The coupling parameter values used for each free energy component are given in Table 4.1. For each value, system was minimized and equilibrated, following which production runs were performed to collect the sampling data. The details of equilibration and production times are given in Table 4.2. To validate the timescales used, longer timescale calculations were performed on NPG, where 0.1 ns of equilibration was followed by 0.4 ns of production. The values differ only by ~ 0.5 -1 kcal/mol between the two sets of calculations. The sampling window of 20-40 ps has been prescribed by several protocols [50, 140] and has been shown to be sufficient when translational and rotational restraints are applied. The data from production simulations was post-processed with the weighted histogram analysis method (WHAM) [141] to obtain the free energy contribution of each component. The free energy contributions from restraining the ligand in its bound state conformation were calculated by integrating the Boltzmann factor of PMFs [50]. The PMFs as a function of RMSD were obtained from umbrella sampling simulations by applying a harmonic biasing potential centered at successive values of RMSD set at [0, 0.2, 0.4, ..., 4.0]. The reaction coordinate (RMSD) data was post-processed with WHAM [141] to obtain the PMFs.

The FEP/MD simulations for each sugar conformations were performed in three blocks and the statistical uncertainty for each individual free energy component as well as the binding free energy is reported based on the standard deviation of the three blocks. Propagation of error was calculated from the Taylor series expansion.

Table 4.2: Equilibration and sampling simulation times in ps for different free energy components. Last column represents the number of runs which is a product of number of windows and number of runs per window.

Interaction	Binding site (ps)		Water (ps)		# Runs
	Eqlb.	Prod.	Eqlb.	Prod.	
Electrostatic	40	40	40	40	11
Dispersive	40	40	40	40	5
Repulsive	50	20	50	20	36
Restraints	10	20	-	-	15

4.3. Results

This section is divided into two subsections. In the first subsection, I will describe substrate binding in the cytoplasmic- and the periplasmic-open states, followed by a discussion on the sugar translocation pathway. The computed binding free energy values are reported in the second subsection.

4.3.1. Sugar binding conformations

A. Sugar binding in cytoplasmic open state

I. Binding of methyl- α -galactopyranoside and methyl- β -galactopyranoside

Methyl- α -D-galactopyranoside (MAG) was shown to have twenty times higher affinity to LacY compared to methyl- β -D-galactopyranoside (MBG). Therefore binding of these two sugars was studied to allow for a direct comparison with the experiments. MBG binds to LacY in a conformation equivalent to the x-ray

crystal structure of LacY-C154G (PDB ID: 1PV7) [13] with a bound disaccharide (TDG). Arg144 forms bidentate H-bonds with 4- and 3-hydroxyl groups of the galactopyranoside moiety (Figure 4.5 (right)). O3 and O4 are believed to be important for the recognition of galactoside in LacY. Glu126 forms a direct H-bond with the 6-hydroxyl group of galactopyranoside. The glycosidic oxygen atom establishes a van der Waals interaction with the S atom of Met23 with an average distance of $\sim 4.5\text{\AA}$. This finding was consistent with the x-ray crystal of LacY-C154G with a bound disaccharide molecule. In a similar binding conformation, MAG forms H-bonds with Arg144 and Glu126 (Figure 4.5 (left)). The charged pair between Glu126 and Arg144 is maintained. The O6 atom of MAG forms a van der Waals interaction with Met23 as opposed to O1 atom for MBG. Although the interacting protein side chains are similar for both the anomers, there are subtle differences in their binding orientations. In alternative binding conformations, the two anomers interact with residues mainly from the C-terminal domain of LacY. The binding structures of these sugars can be divided into two conformational families. Table 4.3 summarizes the proteins residues involved in binding of the two sugars.

Table 4.3: The conformational families observed for MAG and MBS in the cytoplasmic-open state of LacY.

Sugar	Conformation #	Residues
MAG	1	Glu130, Glu126, Arg144, Asp237, Met23
	2	Tyr236, Gln60, Glu269, Lys358, Arg144, Gln359
MBG	1	Glu126, Arg144, Tyr350, Asn119, Met23
	2	Asp237, Lys358, His322, Glu126

II. Binding of α/β -Lactose and α/β -Maltose

A summary of LacY's residues involved in binding of α/β -Lactose α/β -Maltose is given in Table 4.4. Both α and β -Lactose reproduce the crystal-like structures where Glu126 and Glu144 are involved in sugar binding. Glu130 also plays an important role in binding of these sugars. In an alternative binding conformation these sugars form H-bonds with Gln60, Asn119, Asp237, Lys358, and Tyr350 (Figure 4.6). Both anomeric forms of Maltose form two binding conformations similar to Lactose, although the interactions of Maltose are relatively weak with fewer number of protein residues participating in binding (Figure 4.7). The galactoside moiety, which is specifically recognized by LacY [20], is absent in α - and β -Maltose. This can explain its weaker H-bonding interactions with LacY.

Table 4.4: The protein residues involved in binding of α/β -Lactose and α/β -Maltose. The binding structures of the sugars can be divided into two conformational families.

Sugar	Conformation #	Residues
α -Lactose	1	Asp237, Gln60, Asn119, Lys358, Phe354, Tyr350
	2	Glu130, Glu126, Arg144, Asp68, Gly64, Glu269, Glu272
β -Lactose	1	Gln60, Lys358, 359, Asn119, Arg144, His322, Asp237
	2	Glu126, Glu130, Asn272, His322, Tyr350
α -Maltose	1	Asp237, Asn119, Asp240, Lys358
	2	Asn119, Glu126, Arg151, Arg144
β -Maltose	1	Lys358, Asp237, Asn119, Gln359, His322
	2	Glu126, Asn119, Tyr350, Arg144

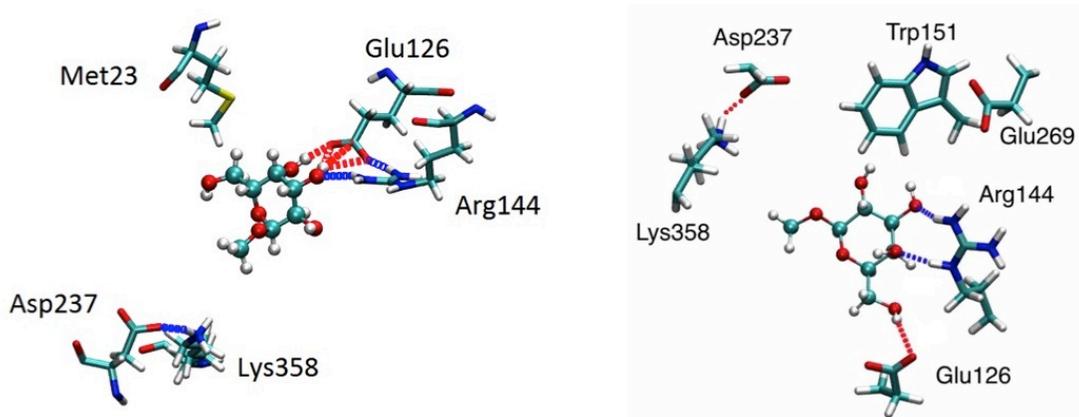


Figure 4.5: Binding conformations in x-ray crystal-like (PDB ID: 1PV7 [13]) conformations for MAG (left) and MBG (right).

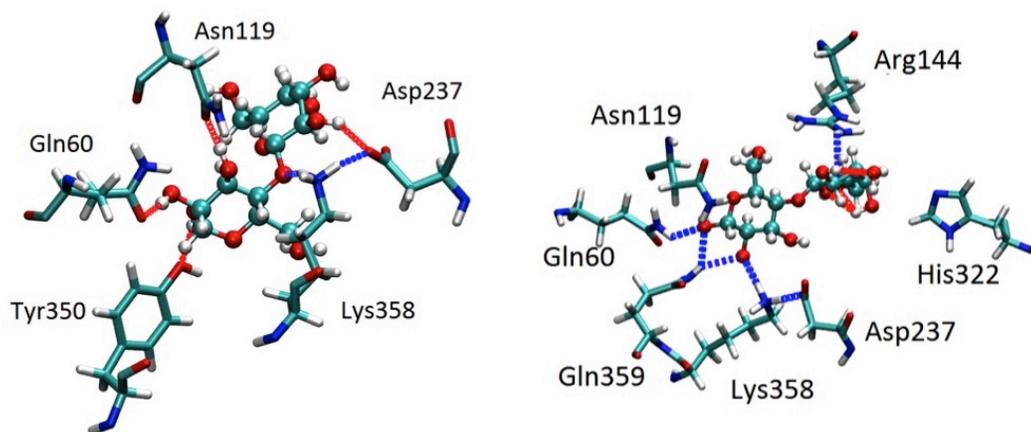


Figure 4.6: Binding conformations for α -lactose (left) and β -lactose (right).

III. Binding of high-affinity binders

The $\alpha\beta$ -(Galp)₂ formed strong H-bond interactions with the residues involved in proton translocation (Glu325, His322, and Glu269). This indicates a possible role of sugar in proton translocation and can provide insights into the energetic coupling

mechanism. The disaccharides, $\alpha\beta$ - and $\beta\beta$ -(Galp)₂ demonstrated strong H-bonding in both of their binding conformations.

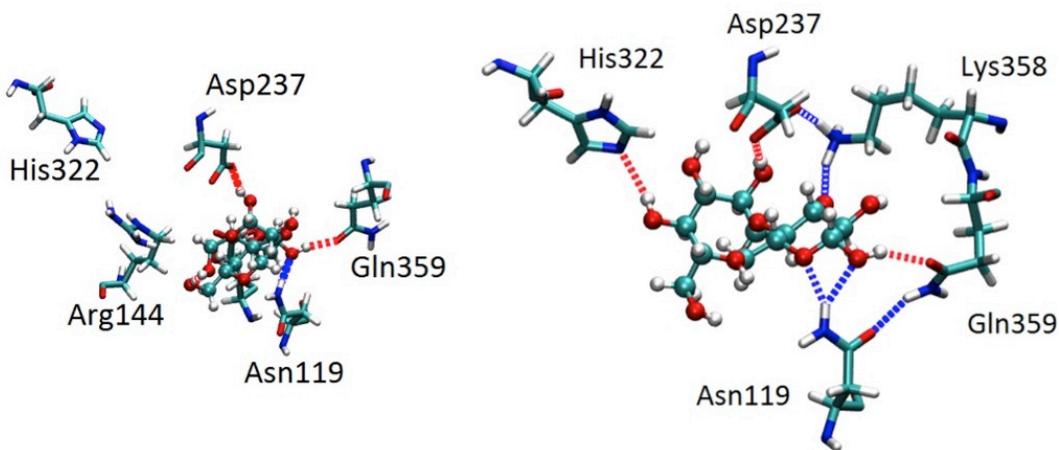


Figure 4.7: Binding conformations of α -maltose (left) and β -maltose (right).

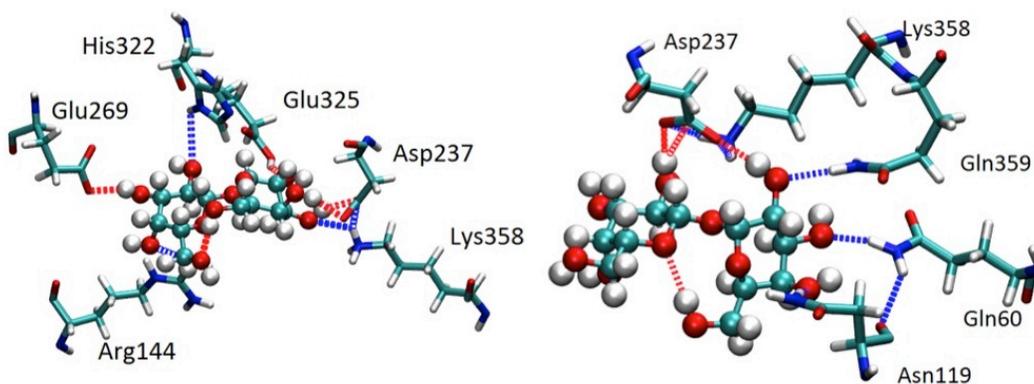


Figure 4.8: Binding conformations of $\alpha\beta$ -(Galp)₂ (left) and $\beta\beta$ -(Galp)₂ (right). Residues involved in proton translocation are seen H-bonding with $\alpha\beta$ -(Galp)₂.

The bidentate H-bonds of Arg144 were prominently observed in MAG and MB, were not observed either for the natural substrate (Lactose) or for the high-affinity binders (NPG and α/β -(Galp)₂). Disaccharides including α - and β -lactose did not form bidentate H-bonds with Arg144. For these sugars, stable binding conformations formed cross-domain interactions. Therefore, it is possible that the crystal structure of MTS-gal covalently bound to single-Cys mutant of LacY at 122 [21] may only match the binding structure of galactose or a galactopyranoside with a smaller functional group. For disaccharides or the galactopyranosides with a bulkier group, interactions of the functional group with protein side chains determine the binding structures.

B. Sugar binding in the Periplasmic-open state

In the periplasmic-open state of LacY, sugar interacts with the same set of residues as summarized in Table 4.5. In the periplasmic-open state, Trp151 is also prominently involved in H-bonding in addition to the aromatic stacking interactions (Table 4.5). The binding conformations in the periplasmic-open state confirm that the same protein residues recognize and ligate sugar molecules in either of the open states thus validating the alternating access model of transport in LacY.

Table 4.5: Summary of protein residues involved in binding of substrates in the periplasmic-open state of LacY. All sugars except for the high-affinity binders, show two conformational families. Same set of residues participate in substrate binding in cytoplasmic- and periplasmic states confirming the alternating access mechanism.

Sugar	Conformation	Residue
	#	
α-Lactose	1	Glu126, Arg144, Lys358, Asp237
	2	Trp151, Asn272, Glu269, Lys358, Asp237
β-Lactose	1	Glu325, Asn272, His322, Trp151, Glu269
	2	Glu126, Arg144, Asp237, Lys358, Trp151
α-Maltose	1	Asp237, Asn119, Asp240, Lys358, Cys148
	2	Glu126, Asn272, Glu269, Glu325
β-Maltose	1	Glu126, Phe354, Tyr350, Ser233
	2	Glu126, Arg144, Asp237, Lys358, Trp151
MAG	1	Glu126, Gln60, Arg144, Gln359
	2	Trp151, Asp237, His322, Glu325
MBG	1	Glu126, Tyr350, Phe354
	2	Glu325, Cys148, Trp151, His322
$\alpha\beta$-(Galp)₂	1	Glu126, Arg144, Trp151, Asp237, Lys358, Asn272,
$\beta\beta$-(Galp)₂	1	Glu126, Arg144, Asp237, Lys358, Asn272, Trp151
NPG	1	Glu126, Arg144, Lys358, Trp151, Asn272

C. Sugar translocation pathways

For one of the MD runs in the cytoplasmic-open state of LacY, α -lactose was placed at the cytoplasmic end of the lumen. The sugar was initially recognized through H-bond interactions with Glu126 and Glu130 and an aromatic stacking interaction

with Tyr350. Both galactoside and glucoside rings of α -lactose form aromatic stacking interactions with Tyr350 and Phe354. The distances between the center of mass of the interacting rings is in the range of 4-6Å (Figure 4.9). The sugar travels from the cytoplasmic end of the protein to the binding site in 10 ns, which in accordance with the findings that initial sugar recognition is a rapid process followed by recognition by relatively slow recognition by Trp151. Residues Tyr350 and Phe354 in combination with the irreplaceable residue Glu126 are possibly involved in initial rapid sugar recognition and facilitate sugar movement inside the protein lumen by ensuring its correct positioning and orientation through stacking interactions.

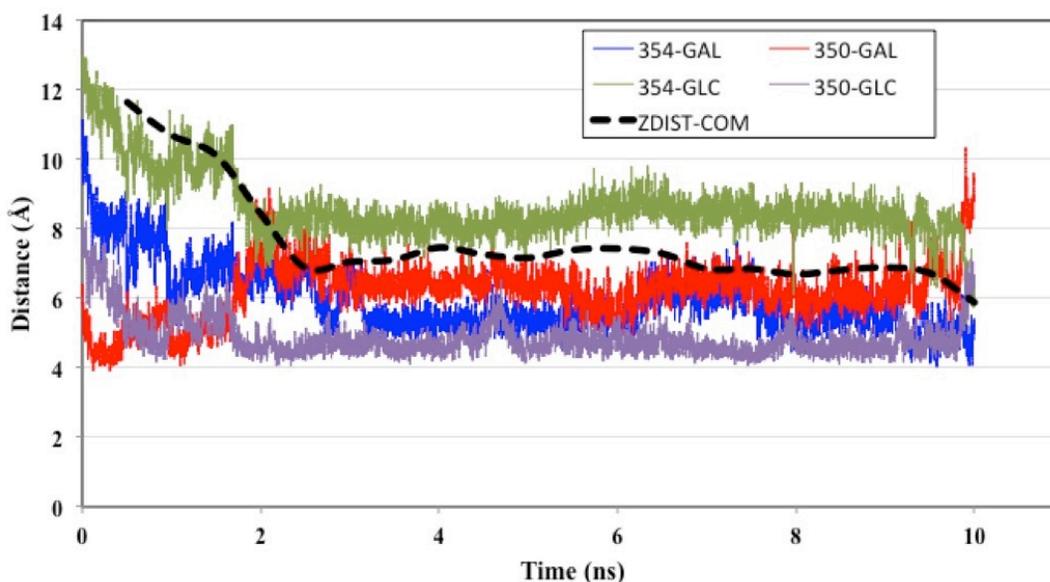


Figure 4.9: Distances between Phe354/Tyr350 and galactoside/glucoside rings of α -Lactose. The distances reported are based on the centers of mass (COM) of the heavy atoms. The black dotted line shows the distance between COMs of sugar and protein in the Z-direction (direction perpendicular to the membrane). The sugar movement from the cytoplasmic-end towards the center of the protein is coordinated through aromatic stacking interactions (distances \sim 4-6Å).

4.3.2. Quantification of binding affinities

A. Binding affinities in the cytoplasmic-open state

Table 4.6: Binding free energy values (kcal/mol) in the cytoplasmic-open state

Sugar	ΔG^0_b (kcal/mol)		$\Delta G^0_{\text{bind (cyto)}}$ (kcal/mol)
	Conformation 1	Conformation 2	
α -Lactose	-5.59 ± 0.33	-7.37 ± 0.48	-7.39 ± 0.96
β -Lactose	1.42 ± 0.67	-3.47 ± 0.56	-3.48 ± 1.14
α -Maltose	6.92 ± 1.67	-2.19 ± 0.81	-2.19 ± 1.69
β -Maltose	0.04 ± 1.41	0.26 ± 0.55	-0.26 ± 1.80
MAG	-4.12 ± 0.38	-3.99 ± 0.41	-4.46 ± 0.58
MBG	-4.43 ± 0.36	-4.13 ± 0.53	-4.70 ± 0.63
NPG-C154G	-8.27 ± 0.49	-8.70 ± 0.94	-8.93 ± 1.37
NPG-LacY-wt	-10.14 ± 1.79	-10.49 ± 1.27	-10.74 ± 2.16
$\alpha\beta$ -(Galp) ₂	-11.91 ± 1.15	-13.34 ± 1.18	-13.39 ± 2.28
$\beta\beta$ -(Galp) ₂	-12.29 ± 1.60	-11.10 ± 0.91	-12.36 ± 2.97

Binding free energy values of all the sugars in the cytoplasmic open state are reported in Table 4.6. For α -Maltose, the binding conformation where substrate interacts with Glu126, Asn272, Glu269, Glu325 resulted in a slightly favorable binding affinity while the other conformation resulted in highly unfavorable binding, while for β -Maltose both binding conformations show close to neutral binding affinity. The binding affinities of MAG and MBG (-4.46 ± 0.58 and -4.70 ± 0.63) are statistically equivalent, which is contrary to the findings from the NEM labeling experiments [31]. Although, it should be noted that the experimental binding affinity

values are based on the complete transport cycle while values reported in Table 4.5 are based on only the cytoplasmic-open structures. The binding of NPG to LacY-wt is slightly more favorable compared to its binding to NPG-C154G (-10.74 ± 2.16 and -8.93 ± 1.37 respectively). For $\alpha\beta$ and $\beta\beta$ -(Galp)₂, the binding free energy values are -13.39 ± 2.28 and -12.36 ± 2.97 . This yields a binding affinity ratio of 5.87 with the $\alpha\beta$ -anomer being the more favorable binder.

B. Binding affinities in the periplasmic-open state

Binding free energies in the periplasmic open state are reported in Table 4.7. In the periplasmic-open state, binding of MAG is considerably more favorable compared to MBG. The ratio of dissociation constants (K_d) is 64 with binding of the α -anomer being more favorable. For $\alpha\beta$ and $\beta\beta$ -(Galp)₂, the binding free energy values are -13.13 ± 1.9 and -9.70 ± 1.1 , respectively. The ratio of binding affinities (inverse of the ratio of K_d) is 363.

Figure 4.10 gives a comparison of the cytoplasmic- and periplasmic-open binding free energy values for all the sugars. In both the open states, binding free energies are statistically very close to each other except β -Lactose and MAG. For these sugars, binding in the periplasmic-open state was more favorable. In summary, the results shows here indicate that there is no significant change the binding affinities during the conformational turnover of LacY but for β -Lactose and MAG. This is in accordance with the argument that the differential binding affinities in the two-states is not a driving mechanism of sugar release at the end of the transport cycle [20, 142].

Table 4.7: Binding free values (kcal/mol) in the periplasmic-open state

Sugar	ΔG_b^0 (kcal/mol)		$\Delta G_{\text{bind (peri)}}^0$ (kcal/mol)
	Conformation 1	Conformation 2	
α -Lactose	-3.2 ± 0.7	-6.56 ± 0.6	-6.56 ± 1.71
β -Lactose	6.99 ± 0.8	-5.68 ± 0.6	-5.7 ± 1.80
α -Maltose	0.31 ± 4	1.12 ± 0.2	0.18 ± 0.93
β -Maltose	0.56 ± 1.2	0.01 ± 0.6	-0.18 ± 1.69
MAG	-7.16 ± 0.3	-2.2 ± 0.3	-7.18 ± 1.03
MBG	0.09 ± 0.3	-4.74 ± 0.9	-4.76 ± 2.69
NPG-LacY-wt	-9.15 ± 0.9	-9.4 ± 1.4	-9.72 ± 2.76
$\alpha\beta$ -(Galp) ₂	-13.09 ± 1.9	—	-13.13 ± 1.9
$\beta\beta$ -(Galp) ₂	-9.67 ± 1.1	—	-9.70 ± 1.1

C. Overall binding free energies

The overall binding free energy values reported in Table 4.7 were obtained from the weighted sum of contributions from the two open states. The weighing factors used for the cytoplasmic- and periplasmic-open states were 0.52 and 0.48, respectively. The weighing factor values were obtained from the double electron-resonance (DEER) data of the relative populations of the two states in presence of a sugar (TDG) for the double-Cys mutant of LacY (R73C/S401C) with attached spin labels [24].

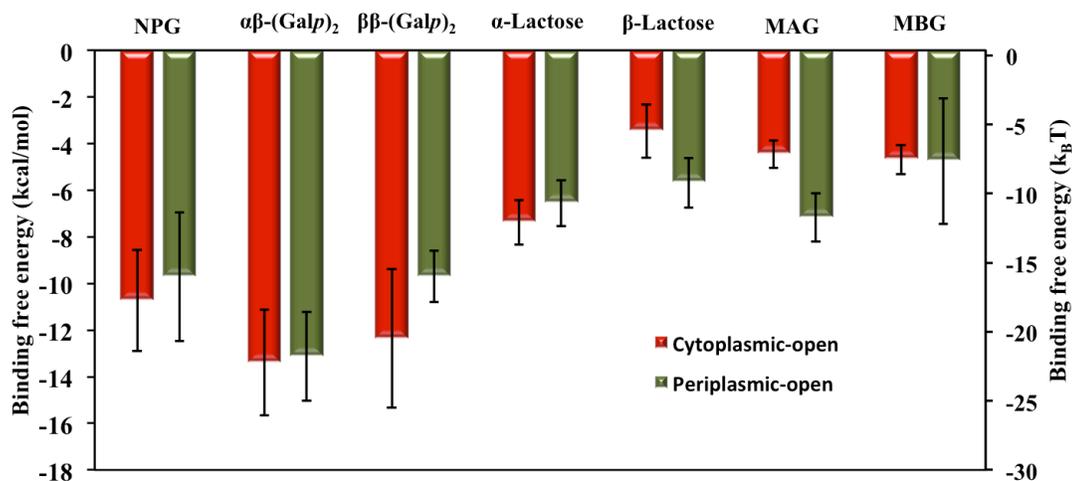


Figure 4.10: Comparison of binding free energies of different sugars in the cytoplasmic- and periplasmic-open states of LacY. Values for the cytoplasmic- and periplasmic-open states are statistically equivalent for all sugars except for MAG and β -lactose. This observation for sugars except MAG and β -lactose is in agreement with the alternating access model [23] of LacY's sugar transport.

Figure 4.11 shows a comparison of computed values with the experimental data. The binding free energies of NPG to LacY-wt and LacY-C154G are obtained from the isothermal calorimetry experiments while the values for MAG and MBG are obtained from the NEM labeling experiments. The values obtained from our simulations are generally more favorable than the experimental values with an average of ~ 2.42 kcal/mol. The dissociation constant (K_d) values from our computations for MAG and MBG are 0.05 mM and 0.3 mM, which yields the binding affinity ratio of 6. Although, this is lower than the experimental value of 16 [30], the FEP/MD protocol was able to correctly obtain the higher affinity binder of the two anomers.

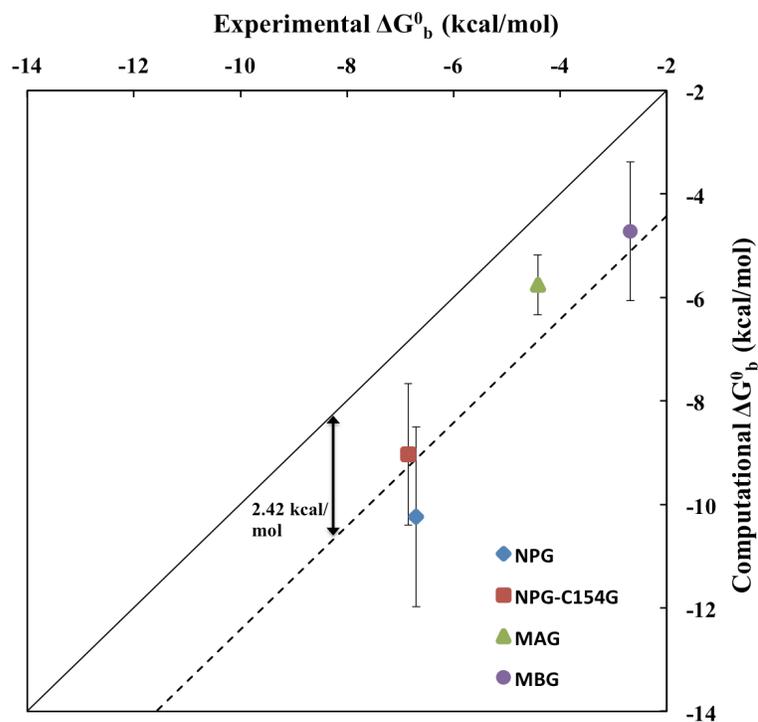


Figure 4.11: Comparison between computational and experimental binding free energies. The computational values reported are the standard binding free energies reported in Table 4.7. The experimental values for NPG are from the isothermal calorimetry experiments [133] and those for MAG and MBG are from studies on NEM labeling of Cys148 [31]. The computational values are more favorable than the experimental values by ~ 2.42 kcal/mol.

4.4. Discussion

For methyl- β -D-galactopyranosides, stable-binding conformations matched with the crystal structure of LacYC154G [13]. Arg144 formed a bidentate interaction with O3 and O4 atoms of the galactopyranoside moiety. The O6 atom formed a direct H-bond interaction with Glu126 as opposed to the water-mediated interaction suggested from the x-ray crystal [13]. The van der Waals interactions with the S atom of Met23 was also observed with the average distance of ~ 4 Å. Methyl- α -D-galactopyranosides also showed formed a similar binding conformation where

Arg144 and Glu126 formed H-bonds with O3 and O4 atoms of the galactopyranoside ring. The binding structures of these sugars are very similar to the suggested structures for galactose binding, which is the most specific substrate for LacY [30]. For disaccharides and NPG, stable binding conformations differed considerably than the structures observed for methyl- α/β galactopyranosides. This indicates that the functional group attached to the galactopyranoside ring greatly affects its binding structures in LacY's binding site. It is also possible that the binding structures seen in the crystal structures represent positioning of the sugar near the two ends of its transport cycle, i.e. when it is entering or exiting the protein lumen as suggested by Chaptal and coworkers [21].

Table 4.8: Overall binding free energy values in kcal/mol. The values reported are weighted sum of binding free energy values in the cytoplasmic- and periplasmic-open states. The weighing factors were obtained from relative populations of the two states reported by Smirnova et al. (2007) [24].

Sugar	ΔG^0_b (kcal/mol)
α -Lactose	-7.00 ± 0.96
β -Lactose	-4.54 ± 1.04
MAG	-5.76 ± 0.58
MBG	-4.72 ± 1.33
NPG-LacY-wt	-10.24 ± 1.74
$\alpha\beta$ -(Galp) ₂	-13.24 ± 1.49
$\beta\beta$ -(Galp) ₂	-11.06 ± 1.62

The alternative binding structure observed for almost all the sugar molecules involve mainly the C-terminal domain residues including Asp237, Lys358, Gln359, Tyr350, Phe354, and Asn119. The x-ray crystal structure of LacY-C154G (Figure 4.1) [13] with a bound TDG molecule possibly indicates an intermediate structure when sugar is transferred from the N-terminal binding site (involving Glu126, Arg144, Trp151) to the alternative binding site.

Figure 4.12 shows contributions of various components of binding free energy for $\beta\beta$ -(Galp)₂ (a high-affinity binder), α -Lactose (a low-affinity natural substrate), and α -Maltose (a non-binder). The results shown in Figure 4.12 correspond to the most favorable binding conformation of each sugar. Although, the decomposition into separate components is path dependent and values of individual components depend on the choice of restraints, force constants, and ligand reference state, it can be used to gain insights into the binding mechanisms of sugars. The contributions from the RMSD constraints are of the order of 0.5 kcal/mol for all the sugars. The effect of orientational restraints is also limited indicating that configurational entropy does not have a significant effect of differential binding. The dispersive and electrostatic contributions are most favorable for $\beta\beta$ -(Galp)₂ and least favorable for α -Maltose. This along with the conformational analysis suggests that all the sugars have different binding structures in LacY's binding pocket that result in differences in their binding affinities.

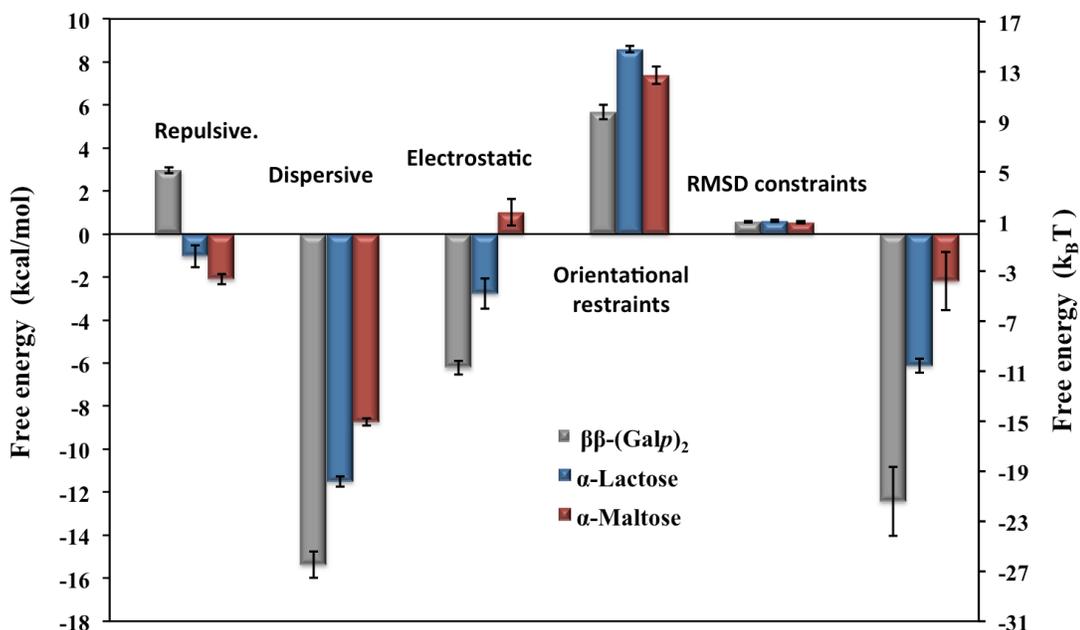


Figure 4.12: Binding free energy contributions from repulsive, dispersive, and electrostatic interactions and orientational and RMSD restraints for $\beta\beta\text{-(Galp)}_2$ (a high-affinity binder), $\alpha\text{-Lactose}$ (natural substrate), $\alpha\text{-Maltose}$ (non-binder). The values represented are from the most favorable conformation for each sugar.

The binding free energies are statistically similar in either of the open states for all sugars except $\beta\text{-Lactose}$ and MAG, which is in accordance with the alternating access model [20, 142]. For MAG, binding free energies for the two configurations in the cytoplasmic-open state are -4.12 ± 0.38 and -3.99 ± 0.41 whereas for the periplasmic-open state the most favorable conformation has the binding free energy of -7.16 ± 0.3 . For $\beta\text{-Lactose}$, only one favorable binding conformation exists for the cytoplasmic- and periplasmic-open states with binding free energies of -3.47 ± 0.56 and -5.68 ± 0.6 , respectively. Multiple binding conformations of MAG in the cytoplasmic-open state of LacY do not result in an entropic contribution enough to exceed the binding free energy of the single favorable configuration in the

periplasmic-open state. For β -Lactose, the interaction free energy change values are -10.47 ± 0.87 and -12.33 ± 1.11 and the free energy changes from constraints are 7.01 ± 0.17 and 6.66 ± 0.49 , respectively for the cytoplasmic- and periplasmic-open states. The major differences in the two states arise from the interaction free energies. These results indicate that the higher affinity in the periplasmic-open state for the two sugars may arise from enthalpic contributions.

The computed binding free energy values are consistently more favorable than the values obtained from the isothermal titration calorimetry and NEM labeling experiments, but agree with the experiments in terms of relative affinities of different anomers of same sugar. The electrostatic component of free energy is crucially important in determining the of computational binding free energy values. Zhao et al. have used an approach where partial charges on the atoms of the substrate were assigned using quantum mechanical calculations [58]. Although, this approach is believed to produce more accurate results, it is computational very demanding. CHARMM force parameters [89-91, 137, 138] have been extensively tested and although the binding free energy values obtained using CHARMM force fields have been shown to be more favorable than the experimental binding free energies by 2-3 kcal/mol [140], they have proven to be useful for comparative analysis of different ligands [58].

Glu325 and Glu269 were protonated in the cytoplasmic- and periplasmic-open states of LacY, respectively, which is in agreement with the experimental and computational findings [20, 81]. Although, it has been proposed that a hydronium ion intermediate, which is stabilized by the ionizable residues in the proton binding site,

is formed during LacY's turnover. Since proton binding residues are also involved in substrate binding, wrong protonation states of important protein residues could lead significant errors in binding free energy calculations. A comprehensive understanding of proton translocation in LacY is important for accurate binding affinity predictions.

Another important factor that may affect the computed free energy values is the protein conformational change. LacY undergoes widespread conformation changes between two states that are separated by a large kinetic barrier (Chapter 3) [81]. Therefore the two open states of LacY can be considered independently and an overall free energy can be calculated as a weighted sum. Although, there are increasing evidences suggesting that an intermediate state of LacY exists which is closed to both sides of the membrane [81, 143]. The free energy contributions from this state need to be accounted for accurate determination of binding affinities in LacY.

The findings from this work can be summarized as follows:

- I. Sugar binding structures can be generally divided into two conformational families in either of LacY's open state, one involving the irreplaceable residues Glu126 and Arg144 and the other involving the C-terminal residues including Asp237, Lys358 and for some sugars the residues involved in proton translocation (Glu269, His322 and Glu325).
- II. In the cytoplasmic-open state of LacY, the substrate is recognized by through hydrogen-bonding interactions with Glu126 and Glu130 and aromatic stacking interactions with Tyr350 and Phe354. The aromatic amino acid residues ensure proper positioning and orientation of the sugar

along its pathway to the binding site of the protein. This suggests a mechanism for rapid initial binding of substrate observed in experiments [144].

- III. The differences in binding affinities of different sugars may arise from the differences in their binding structures to LacY. The binding structures for methyl- α -galactopyranoside and methyl- β -galactopyranoside were not observed for the galactosides with larger functional groups.
- IV. The disaccharides $\alpha\beta$ - and $\beta\beta$ -(Galp)₂ showed considerably tighter binding to LacY. The residues that are believed to be involved in proton translocation (Glu269, His322, Glu325) participate in binding of these sugar molecules indicating a possible mechanism of energetic coupling for the stoichiometric symport.

Chapter 5: Proton translocation pathways in LacY

5.1. Introduction

One of the fundamental problems in understanding the function of LacY is the detection of the residues involved in proton binding and the mechanism of coupling between proton and sugar translocation. It was observed that the sugar binding affinity to LacY decreases for alkaline pH with a pK_a of ~ 10.5 [145]. It was also shown that binding of a sugar to LacY induces no change in pH whereas under similar conditions ligand binding to EmrE results in release of a proton [146]. Based on these findings, it was established that protonation of LacY is a necessary precursor to sugar binding.

Site-directed mutagenesis at almost every position in LacY was used to identify LacY's residues that are critically important for its function [34, 128]. Tyr236 (H-VII), His322 (H-X), and Glu269 (H-VIII) are believed to be involved in the coupled proton/substrate translocation and Arg302 (H-IX) and Glu325 (H-X) are involved in proton binding and translocation [33]. Earlier it was suggested that either His322 or Glu269 gets protonated and the protonation state of these residues is determined by the relative orientation of Tyr236 [20]. However, double mutants at 322 and 269 do not abolish active transport completely [34].

Residues Tyr236 (H-VII), Asp240 (H-VII), Arg302 (H-IX), and His-322 (H-X) form a strong H-bond/salt bridge network approximately in the middle of the protein (Figure 5.1 (left)) and any mutations in these residues result in decreased sugar binding affinity by directly affecting proton binding [33]. Lys319 and Glu325

lie on the periplasmic- and cytoplasmic-sides of this network, respectively, and neutral replacements at these positions abolish sugar transport completely while maintaining the ability of the transporter to carry out equilibrium exchange [33]. The high pK_a of sugar binding (~ 10.5) could not be assigned to any single residue in the central binding cluster, therefore it was suggested that a water molecule in the central binding pocket forms a hydronium ion intermediate that is stabilized by the H-bond network in the central cluster [33, 34]. Neutral replacements at 319 and 325 prevent deprotonation of the hydronium ion intermediate, thus affecting the active transport [34]. From the electronic density maps for the amino acids in the crystal structure of LacY-C154G [13], two structural water molecules were found to be present (Figure 5.1 (right)), one close to the central binding cluster and other forming H-bonds with Lys319 and Thr265 [147]. Thr265 is only one turn away from Glu269, which is believed to be involved in energetic coupling of proton and sugar transport. This theory is analogous to a previously proposed model according to which several residues in the proton binding pocket form a coupling sensor and couple the sugar/proton transport without essentially being involved directly [148].

MD simulations in the inward-facing state of LacY in presence of a sugar molecule revealed a tightly connected H-bond network involving Tyr236, Asp240, Arg302, Lys319, His322, and Glu325 (unpublished data). Direct as well as water-mediated H-bonds were observed throughout the length of the simulations, which indicate an important role of water molecules in the proton binding and translocation.

The focus of this portion of my dissertation is to study the role of important LacY residues in proton translocation and determine the mechanism for coupling of

sugar and proton transport reactions. A hybrid quantum mechanics/molecular mechanics (QM/MM) protocol with self-consistent-charge density-functional-tight-binding (SCC-DFTB) as the QM method was applied to investigate several possible proton transfer pathways involving Glu325, His322, and Glu269. The calculations reveal a central role of waters in proton transfer reaction.

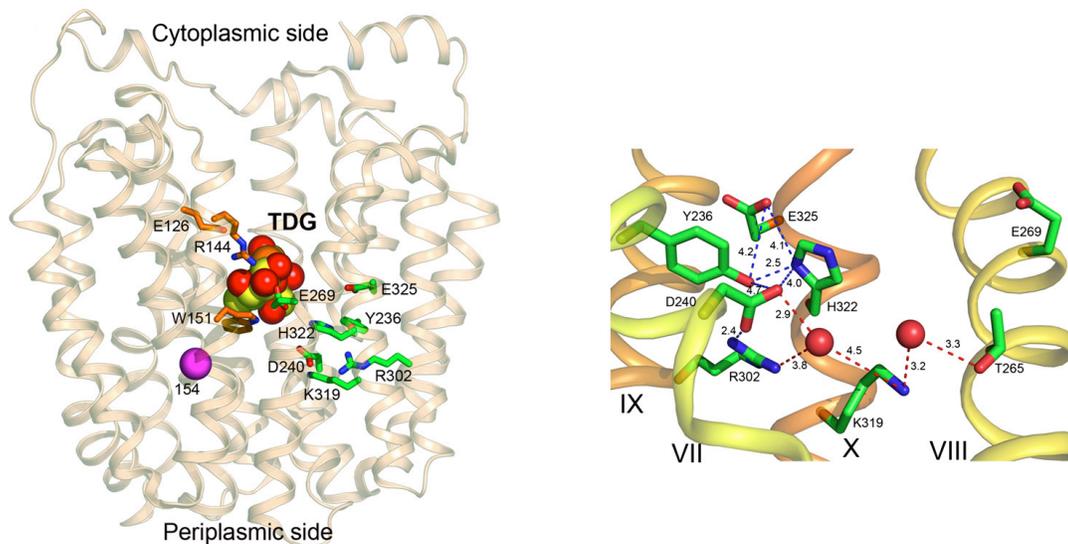


Figure 5.1: The x-ray crystal structure of LacY-C154G [16] with important protein residues. (Left) Residues involved sugar transport Glu126, Arg144, and Trp151 lie in the N-terminal domain, while residues involved in proton transfer Glu325, His322, Arg302, Asn319, Tyr236, and Glu269 lie in the C-terminal domain approximately at the center of the protein. (Right) The residues in the proton binding site form a tight H-bond/salt-bridge network. Electron density maps reveal existence of two water molecules, one being close to the proton binding cluster and the other linking Asn319 and Thr265. The pictures are adopted from Figure S1 from Smirnova et al. [147].

5.2. Computational design and methods

5.2.1. MD simulations

To obtain the conformations for QM/MM simulations, equilibrium MD simulations were performed starting with the x-ray crystal structure of wild-type LacY (PDB ID: 2V8N) [14]. Simulations were performed without a substrate molecule since protonation of LacY is a precursor to sugar binding. Independent MD simulations were performed in two protonation states, unprotonated (LacY^{up}) and (LacY^{325p}). The protein structures were inserted in a model *E. coli* membrane [139], which is a six-lipid model that has been shown to more accurately represent the *E. coli* membrane. The protein-membrane complexes were built using the membrane builder facility in CHARMM-GUI [46-48]. The simulations were performed in NAMD simulation package [45] at the physiological temperature 310.15 K with a timestep of 2 fs. The CHARMM force fields were used to represent the protein and lipids [89-91, 139, 149, 150]. The production runs were performed for 40 ns for both protonation states. The relative orientations of residues involved in proton translocation (Glu325, His322, and Glu2369) were analyzed from both sets of simulations and two structures corresponding to distinct orientations His322 relative to Glu325 and Glu269 were screened. The protonation states of the screened structures were altered to the protonated His322, and two independent MD runs (LacY^{322p-r1} and LacY^{322p-r2}) were performed for 40ns each following the same protocol described earlier. Protein structures in distinct protonation and conformational states were selected (Details in Results section) for further QM/MM simulations.

5.2.2. The QM/MM simulations

The SCCDFTB/MM-GSBP method [60] was used for QM/MM simulations as described in Chapter 2. The simulation protocol used here is similar to several previous SCCDFTB/MM-GSBP studies [61-63]. All the SCCDFTB simulations were performed in CHARMM [41, 43, 44] (version c37b1) with a full third-order extension (D3RD) of SCCDFTB [64]. The system was partitioned in a spherical inner region centered at the center of mass of the QM region with a radius of 20 Å. The QM region consists of Glu325, His322, and Glu269. For water-mediated proton transfer, the bridging water molecule is also added in the QM region. The remaining atoms in the system were treated using the GSBP method [57]. Weak GEO restraints were applied on all the waters to ensure that they do not leave the inner region. The extended electrostatics model [151] was used to treat the nonbonded interactions. All the QM/MM simulations were performed at the physiological temperature of 310.15K.

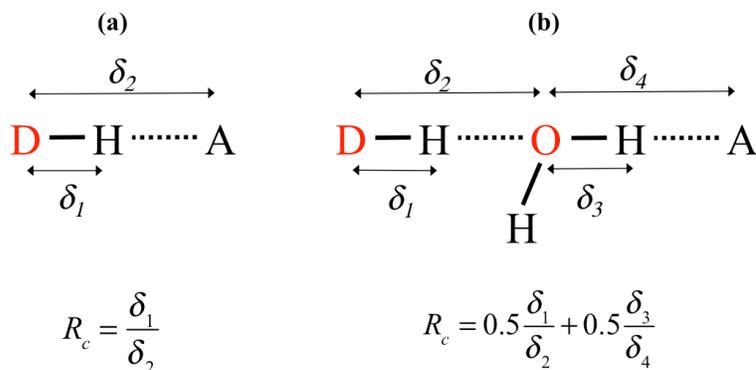


Figure 5.2: The reaction coordinates used in umbrella sampling calculations for single-step (left) and the two-step (right) proton transfers.

To study the energetics of proton translocation, potential of mean force (PMF) was calculated from umbrella sampling simulations. The ratio of donor-hydrogen distance and donor-acceptor distance (D-H/D-A) was chosen as the reaction coordinate (r_c) as shown in Figure 5.2 (left). For a water-mediated proton transfer, a combination of the two individual reaction coordinates (r_{c1} and r_{c2}) was chosen as r_c (Figure 5.2(right)). To ensure sufficient overlap between the neighboring umbrella sampling windows, the reaction space was divided into 16 windows centered at 0.2, 0.25, 0.3, ..., 0.95. At lower values of r_c , proton is bound to the donor, while for higher values of r_c , proton is bound to the acceptor. A harmonic biasing potential was used with the force constants ranging from 50 kcal/mol to 500 kcal/mol. For all umbrella sampling windows, equilibration for 30-50 ps was followed by 120-150 ps of data collection. The reaction coordinate data was then post-processed with the weighted histogram analysis method (WHAM) [141] to obtain the PMF.

5.3. Results

5.3.1. Equilibrium MD simulations

For LacY^{325p} and LacY^{up}, a stable H-bond interaction was observed between Glu325 and Tyr236 for ~40% of the simulation time (H-bond defined for donor-acceptor distance < 3.2 Å) with an average distance of ~3.5 Å. This is consistent with the distance of 2.9 Å in the x-ray crystal structure of wild-type LacY (PDB ID: 2V8N) [14]. Water mediated H-bond interactions were observed between Glu325-His322 and Tyr236-His322. This strongly indicates towards a critical role of waters in proton transport His322 flips between two configurations (Figure 5.4) where it

forms water-mediated interaction with Glu325 or Glu269 alternately, an observation consistent with Andersson et al. [80]. Relative orientation of His322 with respect to Glu325 and Glu269 (Figure 5.3) suggest that His322 or a water molecule coordinated by His322 may be involved in shuttling of protons from Glu325 to Glu269. One conformation was chosen each from LacY^{325apo} and LacY^{up}, where water-mediated interactions exist between Glu325-His322 and His322-269, respectively (Figure 5.4). These screened structures were used for subsequent QM/MM simulations.

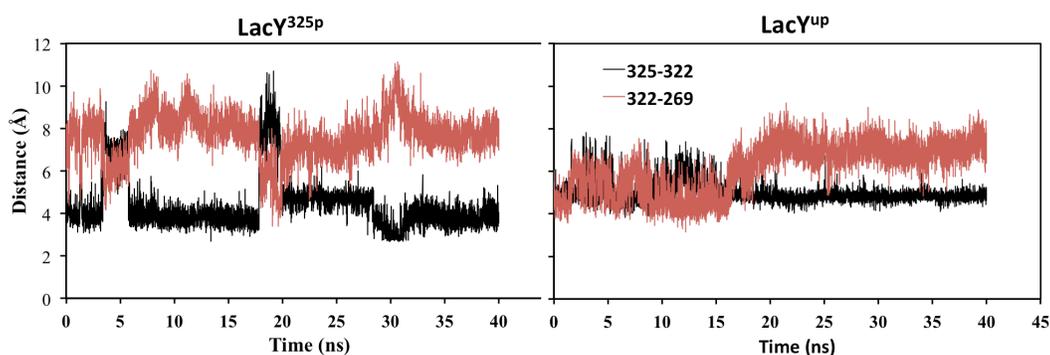


Figure 5.3: The minimum donor-acceptor distances the pairs Glu325-His322 and His322-Glu269 for LacY^{325p} (left) and LacY^{up} (right) simulations. His322 alternates two configurations forming water-mediated interactions with Glu325 and Glu269

For the screened structures (Figure 5.4), His322 was protonated and equilibrium MD simulations were performed (LacY^{322p-r1} and LacY^{322p-r2}) to further understand the role of His322 in proton shuttling. For LacY^{322p-r1}, a continuous network of H-bonds (Glu325-His322-Glu269) was observed (Figure 5.5(a)), suggesting the possibility of a direct proton translocation pathway between these residues. The configuration of His322 where it directly interacts with both Glu325 and Glu269 will be referred as 322^{center} configuration in further discussed. In addition

to this, two other structures (Figures 5.5(b) and (c)) were screened for further QM/MM calculations. The configuration where protonated His322 is oriented towards the proton binding cluster and formed an H-bond with Glu325 will be referred as 322^{IN} configuration (Figure 5.5(b)) and the other configuration where His322 is oriented away from the proton binding cluster forming an H-bond with Glu269 will be called as 322^{OUT} configuration (Figure 5.5(c)).

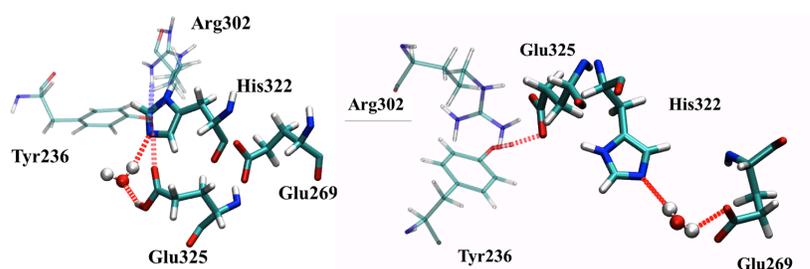


Figure 5.4: The structures screened from LacY^{325p} and LacY^{up} for QM/MM and LacY^{322p} simulations. Water-mediated interaction exists between Glu325-His322 on the left and Glu269-His322 on the right. Tyr236-Glu325 H-bond can be seen in both structures.

5.3.2. Proton translocation from Glu325 to His322

Proton translocation from Glu325 and His322 was studied starting with a structure screened from LacY^{325p} where a water-mediated H-bond interaction exists between Glu325 and His322. The system was divided into QM and MM regions as described in the Computational methods section. The QM region consisted of Glu325, His322, Glu269 and the bridging water molecule. The system minimization and heating was followed by 1 ns of equilibrium simulations. A stable water mediated interaction between Glu325 and His322 observed during the equilibrium QM/MM

run. The PMF curve for the two-step proton transfer from Glu325 to His322 via a water molecule is shown in Figure 5.6. The reaction coordinate, which is a combination of two single-step reaction coordinates, can be perceived as location of the excess charge with increasing values of the reaction coordinate correspond to the movement of an excess proton from Glu325 to His322. The two local minima at ~ 0.35 and ~ 0.85 represent the states where Glu325 and His322 are protonated, respectively. The minima at ~ 0.5 represents a state where the excess positive charge resides on water. The activation energy barrier from the protonated Glu325 to the hydronium ion state, and from the hydronium ion to protonated His322 are ~ 2.6 kcal/mol and ~ 2.7 kcal/mol, respectively. These energy barriers are relatively small and can be overcome through thermal fluctuations which is evident from the kinetic rate constants calculated from the transition state theory that are $9.44 \times 10^{10} \text{ s}^{-1}$ and $8.2 \times 10^{10} \text{ s}^{-1}$, respectively. These findings are consistent with the theory that a hydronium ion intermediate, which is stabilized by residues in the proton-binding cluster, exists during LacY's turnover [34, 147].

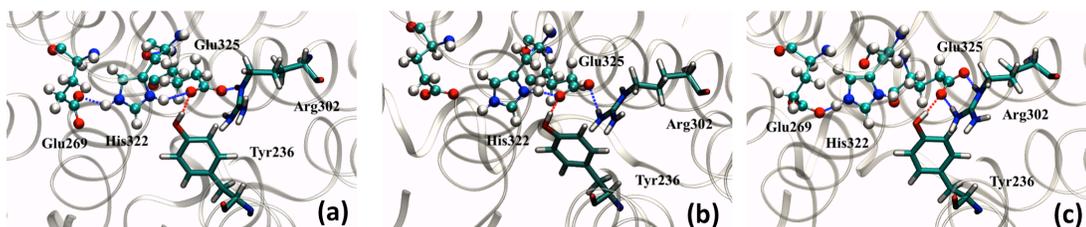


Figure 5.5: Structures screened from LacY325p runs where: (a) A direct H-bond network between Glu325-His322-Glu269 (b) His322 H-bonds with Glu325 and (c) His322 H-bonds with Glu269. H-bonding Interactions of Glu325 with Tyr236 and Arg302 can be seen in all three structures. These three configurations are referred as 322^{center} , 322^{IN} and 322^{OUT} , respectively in further discussion.

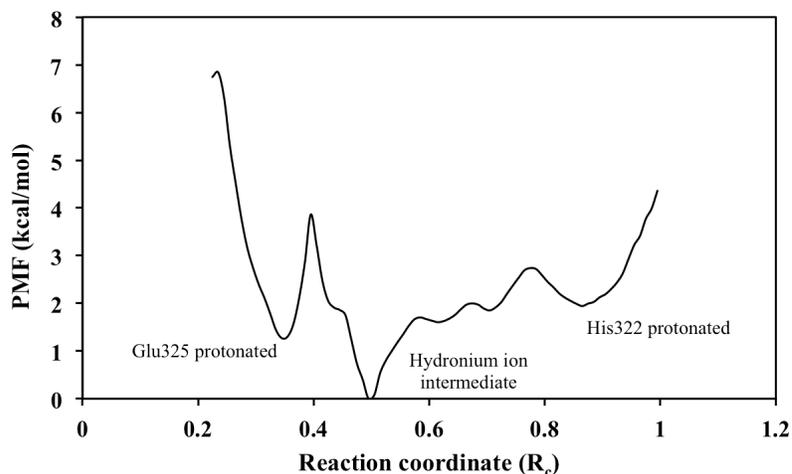


Figure 5.6: Potential mean transfer (PMF) for the water-mediated proton transfer between Glu325 and His322. A stable hydronium ion intermediate is formed at $R_c \sim 0.5$, which is in accordance with the experimental findings [33].

5.3.3. Proton translocation from His322 to Glu269

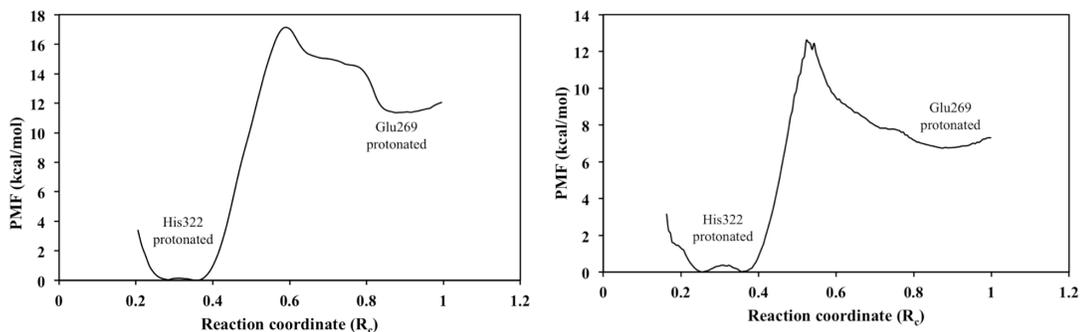


Figure 5.7: The potential mean force (PMF) curves for the proton transfer between His322 and Glu269 in 322^{center} (left) and 322^{OUT} (right) configurations. The energy barriers for the two reactions are ~ 17 kcal/mol and ~ 12.7 kcal/mol, respectively.

Figure 5.7 (left) and (right) represent the PMF curves for the proton transfer between His322 and Glu269 in 322^{center} and 322^{IN} configurations, respectively. The orientation of His322 affects the energetics of its deprotonation to Glu269. In the 322^{center} configuration PMF has an energy barrier of ~ 17 kcal/mol as opposed to the

322^{OUT} configuration where the barrier is ~ 12.7 kcal/mol. In the ‘center’ configuration, the H-bond with Glu325 is likely to induce an “H-bond pulling” effect on His322, thus increasing its pK_a . In the ‘OUT’ configuration (Figure 5.5 (a)) His322 is able to lose its proton relatively easily. The rate constants for the proton transfer reaction in the ‘IN’ and ‘OUT’ state of His322 are 6.7 s^{-1} and 7000 s^{-1} , respectively. The 322^{OUT} is clearly the more favored configuration for this proton transfer step based on our calculations.

5.3.4. Proton translocation from His322 to Glu325

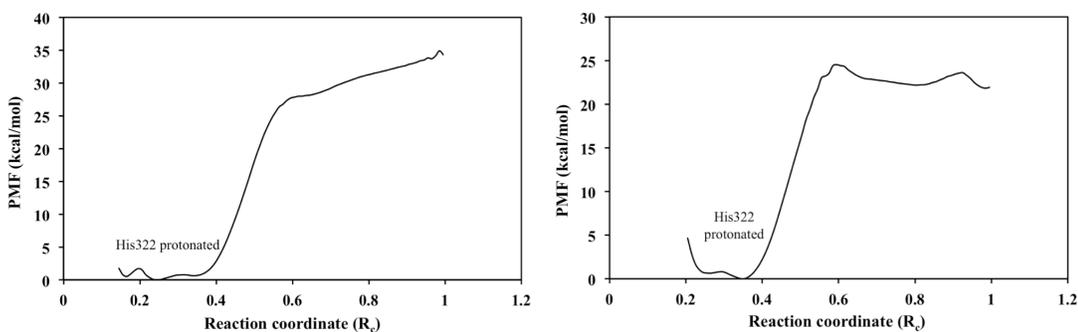


Figure 5.8: The potential mean force (PMF) curves for the proton transfer between His322 and Glu325 in 322^{IN} (left) and 322^{center} (right) configurations. For both reactions, PMF rises rapidly from $R_c \sim 0.4$ – $R_c \sim 0.6$, followed by a slow rise for the ‘IN’ and plateau for the ‘center’ configurations for His322, respectively.

The PMFs for the proton transfer from His322 to Glu325 in the 322^{IN} and 322^{center} configurations are shown in Figure 5.8. In the ‘IN’ configuration (Figure 5.8 (left)), the PMF rapidly rises from $R_c \sim 0.4$ till it reaches the barrier at ~ 27 kcal/mol at $R_c \sim 0.6$. This is followed by a slow increase after $R_c \sim 0.6$. In the ‘center’ configuration, similar rapid rise of the PMF was observed beyond $R_c \sim 0.4$ till a

plateau is reached at $R_c \sim 0.6$ (Figure 5.8 (b)). The broad barrier region is centered at ~ 27 kcal/mol. These PMF behaviors are indicative of the absence of a high pK_a proton acceptor. For higher values of R_c (approximately >0.7) in umbrella sampling simulations for the ‘IN’ and ‘center’ configurations of His322, Glu325 forms a strong salt bridge/H-bond interaction with Arg302 (Figure 5.9). Formation of a salt-bridge between Glu325 and Arg302 is believed to be a key step in deprotonation of Glu325. Neutral replacements at 302 result in mutants that are severely defective in active transport but perform equilibrium exchange, which does not involve net proton transfer [152]. It is believed that the salt-bridge formation with Arg302 significantly reduces the pK_a of Glu325, thus resulting in its deprotonation at the end of the transport cycle [152]. Andersson et al. used an empirical method for pK_a calculation to show that decrease in pK_a is analogous with decrease in the distance between amino group of His322 and hydroxyl group of Glu325 [80]. This phenomenon explains the plateau-like nature of the PMF curves.

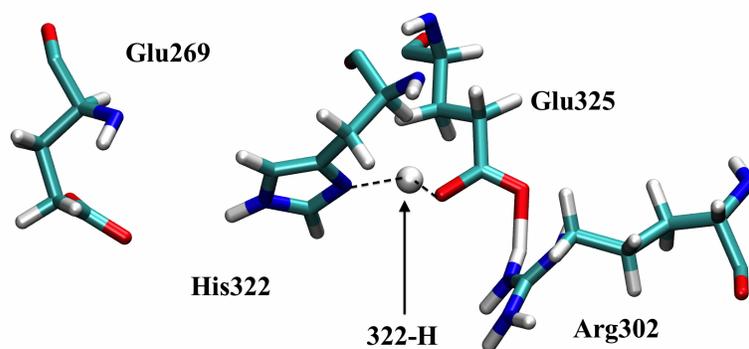


Figure 5.9: A salt-bridge/H-bond interaction between Arg302 and Glu325. The hydrogen atom of His322 is transferred on Glu325. The low pK_a of Glu325 resulting from its interaction with Arg302 causes plateau-like PMF curves shown in Figure 5.8.

5.4. Discussion

Proton transfer and its energetic coupling with substrate binding is one of the most inadequately understood aspects of LacY's mechanism. Although the residues involved in proton transfer have been identified, the exact nature of their contribution to the overall transport mechanism remains a challenge. There is significant evidence suggesting that Glu325 gets deprotonated at the end of active transport and gets protonated at the beginning of the efflux, none of the other residues in the proton binding pocket can be identified as an individual proton acceptor. Therefore a mechanism was proposed where a water molecule acts as a cofactor in proton binding, and an hydronium ion intermediate stabilized by surround residues in proton binding site exists during LacY's conformational turnover. Results from the MD simulations showed existence of water-mediated H-bond interactions between Glu325-His322, Tyr236-His322, and His322-Glu269 strongly suggesting the role of water molecules in proton transfer. For LacY^{325p} and LacY^{up} simulations, His322 alternates between two configurations where is it alternatingly forms water-mediated H-bonds with Glu325 and Glu269. His322 or a water molecule coordinated by His322 may acts as a shuttle for proton transfer between Glu325 and Glu269.

The PMF for water-mediated proton transfer between Glu325 and His322 clearly shows formation of a stabilized hydronium ion intermediate (Figure 5.6). When compared with the PMFs for His322-Glu269 (Figure 5.7) and His322-Glu325 proton transfers (Figure 5.8), the energy barrier for the water-mediated Glu325-His322 proton transfer is significantly less (2.7 kcal/mol vs. 12.7 kcal/mol and higher). These results imply that proton transfer to His322 may not be the rate-

limiting reaction. Proximity to Arg302 is believed to be responsible for lowering the pK_a of Glu325 [33, 80]. The Arg302-Glu325 salt-bridge/H-bonds interactions observed during the umbrella sampling simulation may explain the low free energy barrier for proton transfer from Glu325 to water as Glu325 readily loses its proton. Tyr236 also maintained H-bond interaction throughout the umbrella sampling simulations. The QM region included Glu325, His322, Glu269 and the bridging water molecule in the umbrella sampling runs for this proton transfer step. More productive QM/MM simulations that also include Arg302, Tyr236 and more water molecules in the vicinity will throw more light on the role of these residues and formation of the hydronium ion intermediate.

The energy barrier for proton transfer between His322-Glu269 is significantly lower for the 322^{OUT} configuration compared to 322^{center} (~12.7 kcal/mol vs. ~17 kcal/mol) (Figure 5.7). Flipping of His322 observed in LacY^{325p}, LacY^{up}, and LacY322p simulations is a possible mechanism for the change in His322 configuration. The PMFs for His322-Glu325 proton transfer reached a plateau-like region after a sharp rise (Figure 5.8) in both 322^{IN} and 322^{center} configurations. Protonation of Glu325 results in a very high-energy complex, indicating a low pK_a value for Glu325. Glu325 in this state forms a salt-bridge/H-bond interaction with Arg302, which has been postulated as being responsible for lowering the pK_a of Glu325 [33, 34]. More extensive simulations with Arg302, Tyr236 and neighboring waters included in the QM region will be required for a complete understanding of the proton transfer process.

In summary, the results from this work suggest a mechanism where Glu325 transfers the excess proton to His322 through a hydronium ion intermediate, followed by rotation of protonated His322 and transfer of proton to Glu269. And in light of the experimentally proposed model where Tyr236, His322, and Glu269 ligate a water molecule [34], we propose that His322 acting as the proton shuttle as one of the several possible proton transfer pathways.

Chapter 6: Conclusions and Future Directions

In this dissertation, lactose permease (LacY) of *E. coli* was investigated as a model for the major facilitator superfamily (MFS) proteins. This work focused on three key aspects on LacY's substrate transport mechanism. In the first part, a two-step hybrid simulation approach (IM-EX) was developed to efficiently sample conformational changes in secondary active proteins. This approach uses self-guided langevin dynamics (SGLD) simulations [82-84] in the first step to with an implicit membrane [105] to enhance conformational sampling. Molecular dynamics (MD) simulations are performed in the second step to obtain the final protein conformations in a more realistic environment. Starting with the x-ray crystal of LacY in the cytoplasmic-open state [13], a model for the unknown periplasmic-open state of LacY was determined which agrees with several indirect experimental measurements such as double electron-electron resonance (DEER) [24] and site-directed alkylation [26]. Three cross-domain H-bond pairs (Asn245/Pro31, Ser41/Glu374, and Lys42/Gln242) near the periplasmic-end of LacY were found stabilize the cytoplasmic-open state of LacY and destabilization of these interactions facilitated periplasmic-open structure of LacY. Mutations in these residues are suggested that may destabilize the cytoplasmic-open state and aid in crystallization of the periplasmic-open state. Protonation state of LacY affects the protein conformational change and protonation of Glu269 along with sugar binding was seen to trigger the periplasmic-opening of LacY. The IM-EX method developed as a part of this dissertation provides a

promising approach to determine conformational changes in secondary active proteins for which only one of the conformational states has been determined.

Determination of the periplasmic-open state at atomic-level details opened several avenues for exploring LacY's mechanism using computational techniques. In the second part of the dissertation, the periplasmic-open structure obtained from the first part was used along with the cytoplasmic-open x-ray crystal to study thermodynamics of sugar binding in detail. Major substrate binding protein residues were found to be the same in both open states. Binding free energy values obtained from the alchemical free energy perturbation simulations are statistically similar in both open states for all sugars except MAG and β -Lactose. These findings provide a quantitative validation of the alternating access mechanism. The binding affinity of the α -anomer of methyl-D-galactopyranoside was found to be greater than the β -anomer, which is consistent with the experimental findings [30]. Comparison of individual free energy components for different sugars shows that major contributions to the differential binding free energies come from dispersive and electrostatic interactions. Although the binding site for different binders in LacY is the same, subtle difference in binding structures are responsible for differential binding affinities. The disaccharides with two galactoside moieties showed tighter binding and considerably more favorable binding affinity. It needs be appreciated that subtle structural differences in galactoside and glucoside moieties are recognized by LacY resulting in highly specific binding. A possible pathway for substrate transport in the cytoplasmic-open state was proposed. It involves aromatic stacking interactions of saccharide rings with Tyr350 and Phe354. Aromatic stacking is important in sugar

recognition in many other sugar transporters [153] and Trp151 is also involved in sugar recognition in LacY. It is hypothesized that aromatic stacking interactions of sugars with Tyr350 and Phe354 are responsible for their rapid initial recognition. In one of the binding conformations of $\alpha\beta$ -(Galp)₂, important residues in proton transfer reaction (Glu325, His322, and Glu325) participate in H-bonding. Their relative configurations in this state are considerably different than *apo*-LacY configurations. This binding conformation presents opportunities for proton transfer studies in substrate bound LacY.

The results from the first part of this dissertation revealed the importance of proton transfer to Glu269 for LacY's conformational changes. In the third part of this dissertation, a hybrid QM/MM approach was applied to study proton translocation in LacY. A hydronium ion intermediate was observed during the proton transfer from Glu325 to His322, which is consistent with the experimental hypothesis. In the His322 protonated state, proton transfer to Glu325 resulted in a high-energy complex suggesting a low pK_a of Glu325 in this state. In this conformation, Glu325 formed a salt bridge/H-bond interaction with Arg302, which is consistent with the finding that interaction with Arg302 reduces the pK_a of Glu325 and causes its deprotonation [152]. Based on the results of this work, a mechanism of proton transfer is proposed. In the first step, a proton is transferred from Glu325 to His322 through a hydronium ion intermediate, which is followed by rotation of His322 away from Glu325 and towards Glu269. In the third step proton is transferred to Glu269. The agreement of the initial findings of proton translocation studies with the experiments is encouraging

and more extensive QM/MM studies are proposed to further our understanding of proton transfer in LacY.

Since LacY shares its structural and functional motif with other MFS transporters, results from the studies on conformational changes, sugar binding, and proton translocation can guide investigation of other MFS transporters such as the drug efflux protein EmrD. The information about the role of conserved residues from this work can potentially be helpful in structure-based drug design for EmrD. This dissertation touches all the important aspects of the mechanism of a secondary active transporter including the conformational changes, substrate binding and ion translocation. With the continuously evolving computational methods and improving computing power, this dissertation can serve as a template for future study of other secondary active transporters.

Future directions

The secondary active transporters (SATs) carry out coupled transport of a substrate and a cation by the alternating access mechanism, where protein alternates between multiple conformational states to allow for the passage of substrate across the membrane. The two-step hybrid simulation approach (IM-EX) [81] developed as a part of this dissertation provides an efficient route to investigate conformational changes in SATs. This method can be extended to determine the unknown conformational states in SATs where only one of the conformational states has been determined. For SAT proteins, this is the most common case except for the proteins of the LeuT fold e.g. LeuT [154] and Mhp1 [155].

The double electron-electron resonance (DEER) experiments on LacY measured distances between the spin labels attached to nine paired double-Cys mutants of LacY [24]. The C_{α} - C_{α} distances for the periplasmic-open conformations obtained from the IM-EX simulations (Chapter 3) agree fairly well with the corresponding spin label distances for some pairs but there are discrepancies between the experimental and simulation data for other pairs [24]. Moreover, the experimentally proposed unwinding helix dynamics was not observed during the IM-EX simulations [81], which could be an artifact of the attached spin labels. A long-timescale MD simulation study is currently being performed on LacY with explicitly attached spin labels to understand the spin label dynamics and conformational changes in LacY in presence of explicit spin labels. This study will enable understanding of the discrepancies between the experimental and simulation data and throw light on the postulated helix unwinding. This study is being performed with a complex six-lipid model that more accurately represents the *E. coli* membrane [139]. The lipid redistribution from the long-timescale simulations can also aid in understanding LacY-lipid interactions.

The initial QM/MM simulations on LacY resulted in a finding that corroborates an important role of waters in proton transfer (Chapter 5). Arg302 and Tyr236, which are within H-bonding distance from Glu325, are also important for proton transfer. Umbrella sampling simulations will be performed with Arg302, Tyr236 included in the QM region along with Glu325, His322, and Glu269. The proton binding region consists of \sim 8-10 water molecules. Although computations on a QM region of this size are computationally very expensive, preliminary benchmarks

of SCCDFTB parallelized over multiple processors in CHARMM (version c37b1) suggest that these computations are feasible within reasonable amount of time. The SCCDFTB/MM-GSBP protocol can be used to compute pK_a values of residues participating in proton transfer using the free energy perturbation protocol. Another possible route can be to study the long-range proton transfer from Glu325 to Glu269 via His322 and water molecules through umbrella sampling simulations with a collective variable defined by as the reaction coordinate. Sugar molecules interact with the residues involved in proton translocation (Chapter 4). These binding conformations possibly hold the key to the energetic coupling between sugar and proton translocation reactions. QM/MM simulations with a bound sugar may lead to an enhanced understanding of this coupling.

References:

1. Ravna, A., et al., *Membrane Transporters: Structure, Function and Targets for Drug Design*, in *Transporters as Targets for Drugs*, S. Napier and M. Bingham, Editors. 2009, Springer Berlin Heidelberg. p. 15-51.
2. Jentsch, T.J. and C.A. Hubner, *Ion channel diseases*. Human Molecular Genetics, 2002. **11**(20): p. 2435-2445.
3. Davis, P.B., *Cystic fibrosis since 1938*. American journal of respiratory and critical care medicine, 2006. **173**(5): p. 475-82.
4. Putman, M., H.W. van Veen, and W.N. Konings, *Molecular properties of bacterial multidrug transporters*. Microbiology and molecular biology reviews : MMBR, 2000. **64**(4): p. 672-93.
5. Saier, M.H., C.V. Tran, and R.D. Barabote, *TCDB: the Transporter Classification Database for membrane transport protein analyses and information*. Nucleic Acids Research, 2006. **34**(suppl 1): p. D181-D186.
6. Saier, M.H., et al., *The Transporter Classification Database: recent advances*. Nucleic Acids Research, 2009. **37**(suppl 1): p. D274-D278.
7. Chang, A.B., et al., *Phylogeny as a guide to structure and function of membrane transport proteins*. Molecular membrane biology, 2004. **21**(3): p. 171-81.
8. Pao, S.S., I.A.N.T. Paulsen, and M.H. Saier, *Major Facilitator Superfamily*. Microbiology, 1998. **62**(1): p. 1-34.
9. Reddy, V.S., et al., *The major facilitator superfamily (MFS) revisited*. FEBS Journal, 2012. **279**(11): p. 2022-2035.
10. Putman, M., H.W. van Veen, and W.N. Konings, *Molecular properties of bacterial multidrug transporters*. Microbiology and molecular biology reviews, 2000. **64**(4): p. 672-693.
11. Saidijam, M., et al., *Microbial drug efflux proteins of the major facilitator superfamily*. Current drug targets, 2006. **7**(7): p. 793-811.
12. Yin, Y., et al., *Structure of the multidrug transporter EmrD from Escherichia coli*. Science, 2006. **312**(5774): p. 741-4.
13. Abramson, J., et al., *Structure and mechanism of the lactose permease of Escherichia coli*. Science, 2003. **301**(5633): p. 610-615.
14. Guan, L., et al., *Structural determination of wild-type lactose permease*. Proceedings of the National Academy of Sciences, 2007. **104**(39): p. 15294-15298.
15. Vitrac, H., M. Bogdanov, and W. Dowhan, *Proper Fatty Acid Composition Rather than an Ionizable Lipid Amine Is Required for Full Transport Function of Lactose Permease from Escherichia coli*. Journal of Biological Chemistry, 2013. **288**(8): p. 5873-5885.
16. Huang, Y., et al., *Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli*. Science, 2003. **301**(5633): p. 616-620.
17. Dang, S., et al., *Structure of a fucose transporter in an outward-open conformation*. Nature. **467**(7316): p. 734-738.

18. Sun, L., et al., *Crystal structure of a bacterial homologue of glucose transporters GLUT1-4*. *Nature*, 2012. **490**(7420): p. 361-366.
19. Solcan, N., et al., *Alternating access mechanism in the POT family of oligopeptide transporters*. *EMBO J*, 2012. **31**(16): p. 3411-3421.
20. Guan, L. and H.R. Kaback, *Lessons from lactose permease*. *Annual review of biophysics and biomolecular structure*, 2006. **35**: p. 67-91.
21. Chaptal, V., et al., *Crystal structure of lactose permease in complex with an affinity inactivator yields unique insight into sugar recognition*. *Proceedings of the National Academy of Sciences*, 2011. **108**(23): p. 9361-9366.
22. Mirza, O., et al., *Structural evidence for induced fit and a mechanism for sugar/H⁺ symport in LacY*. *EMBO J*, 2006. **25**(6): p. 1177-1183.
23. Smirnova, I., V. Kasho, and H.R. Kaback, *Lactose Permease and the Alternating Access Mechanism*. *Biochemistry*, 2011. **50**(45): p. 9684-9693.
24. Smirnova, I., et al., *Sugar binding induces an outward facing conformation of LacY*. *Proceedings of the National Academy of Sciences of the United States of America*, 2007. **104**(42): p. 16504-16509.
25. Majumdar, D.S., et al., *Single-molecule FRET reveals sugar-induced conformational dynamics in LacY*. *Proceedings of the National Academy of Sciences*, 2007. **104**(31): p. 12640-12645.
26. Nie, Y., N. Ermolova, and H.R. Kaback, *Site-directed alkylation of LacY: effect of the proton electrochemical gradient*. *Journal of molecular biology*, 2007. **374**(2): p. 356-364.
27. Kaback, H.R., et al., *Site-directed alkylation and the alternating access model for LacY*. *Proceedings of the National Academy of Sciences*, 2007. **104**(2): p. 491-494.
28. Zhou, Y., et al., *Opening and closing of the periplasmic gate in lactose permease*. *Proceedings of the National Academy of Sciences*, 2008. **105**(10): p. 3774-3778.
29. Sugihara, J., et al., *Sugar Recognition by CscB and LacY*. *Biochemistry*, 2011. **50**(51): p. 11009-11014.
30. Vazquez-Ibar, J., et al., *Sugar recognition by the lactose permease of Escherichia coli*. *The Journal of biological chemistry*, 2004. **279**(47): p. 49214-21.
31. Sahin-Tóth, M., et al., *Ligand Recognition by the Lactose Permease of Escherichia coli: Specificity and Affinity Are Defined by Distinct Structural Elements of Galactopyranosides*. *Biochemistry*, 2000. **39**(17): p. 5097-5103.
32. Green, A.L., H.A. Hrodey, and R.J. Brooker, *Evidence for Structural Symmetry and Functional Asymmetry in the Lactose Permease of Escherichia coli*. *Biochemistry*, 2003. **42**(38): p. 11226-11233.
33. Smirnova, I., et al., *Residues in the H⁺ translocation site define the pKa for sugar binding to LacY*. *Biochemistry*, 2009. **48**(37): p. 8852-8860.
34. Zhou, Y., X. Jiang, and H.R. Kaback, *Role of the irreplaceable residues in the LacY alternating access mechanism*. *Proceedings of the National Academy of Sciences*, 2012. **109**(31): p. 12438-12442.
35. Alder, B.J. and T.E. Wainwright, *Studies in Molecular Dynamics. I. General Method*. *The Journal of Chemical Physics*, 1959. **31**(2): p. 459-466.

36. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**(5612): p. 585-590.
37. Rapaport, D.C., *The Art of Molecular Dynamics Simulation*. The Art of Molecular Dynamics Simulation, by D.~C.~Rapaport, pp.~564.~ISBN 0521825687.~Cambridge, UK: Cambridge University Press, April 2004. 2004: Cambridge University Press.
38. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 1989: Clarendon Press. 385.
39. Leach, A., *Molecular Modelling: Principles and Applications (2nd Edition)*. 2001: Prentice Hall.
40. Frenkel, D. and B. Smit, *Understanding Molecular Simulation*. 2001: Academic Press, Inc. 638.
41. MacKerell, A.D., et al., *CHARMM: The Energy Function and Its Parameterization*, in *Encyclopedia of Computational Chemistry*. 2002, John Wiley & Sons, Ltd.
42. Dolan, E.a., et al., *Simulations of membranes and other interfacial systems using P2(1) and Pc periodic boundary conditions*. Biophysical journal, 2002. **82**(5): p. 2317-25.
43. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
44. Brooks, B.R., et al., *CHARMM: The biomolecular simulation program*. Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
45. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**(16): p. 1781-1802.
46. Jo, S., T. Kim, and W. Im, *Automated builder and database of protein/membrane complexes for molecular dynamics simulations*. PloS one, 2007. **2**(9): p. e880-e880.
47. Jo, S., et al., *CHARMM-GUI: A web-based graphical user interface for CHARMM*. Journal of Computational Chemistry, 2008. **29**(11): p. 1859-1865.
48. Jo, S., et al., *CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes*. Biophysical journal, 2009. **97**(1): p. 50-58.
49. Chipot, C. and A. Pohorille, *Free Energy Calculations Theory and Applications in Chemistry and Biology*. Free Energy Calculations Theory and Applications in Chemistry and Biology, ed. C. Chipot and A. Pohorille. 2007: Springer Berlin Heidelberg.
50. Wang, J., Y. Deng, and B. Roux, *Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials*. Biophysical journal, 2006. **91**(8): p. 2798-2814.
51. Boresch, S., et al., *Absolute Binding Free Energies: A Quantitative Approach for Their Calculation*. The Journal of Physical Chemistry B, 2003. **107**(35): p. 9535-9551.
52. Weeks, J.D., D. Chandler, and H.C. Andersen, *Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids*. The Journal of Chemical Physics, 1971. **54**(12): p. 5237-5237.

53. Gallicchio, E. and R.M. Levy, *Recent theoretical and computational advances for modeling protein–ligand binding affinities*, in *Advances in Protein Chemistry and Structural Biology*, C. Christo, Editor. 2011, Academic Press. p. 27-80.
54. Mobley, D.L., J.D. Chodera, and K.A. Dill, *On the use of orientational restraints and symmetry corrections in alchemical free energy calculations*. *The Journal of Chemical Physics*, 2006. **125**(8): p. 084902-16.
55. Gilson, M.K., et al., *The statistical-thermodynamic basis for computation of binding affinities: a critical review*. *Biophysical journal*, 1997. **72**(3): p. 1047-1069.
56. Woo, H.-J. and B. Roux, *Calculation of absolute protein–ligand binding free energy from computer simulations*. *Proceedings of the National Academy of Sciences of the United States of America*, 2005. **102**(19): p. 6825-6830.
57. Im, W., S. Berneche, and B. Roux, *Generalized solvent boundary potential for computer simulations*. *The Journal of Chemical Physics*, 2001. **114**(7): p. 2924-2937.
58. Zhao, C., D.A. Caplan, and S.Y. Noskov, *Evaluations of the Absolute and Relative Free Energies for Antidepressant Binding to the Amino Acid Membrane Transporter LeuT with Free Energy Simulations*. *Journal of Chemical Theory and Computation*, 2010. **6**(6): p. 1900-1914.
59. Beglov, D. and B. Roux, *Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations*. *The Journal of Chemical Physics*, 1994. **100**(12): p. 9050-9063.
60. Riccardi, D., S. Yang, and Q. Cui, *Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations*. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2010. **1804**(2): p. 342-351.
61. Goyal, P., M. Elstner, and Q. Cui, *Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water*. *The Journal of Physical Chemistry B*, 2011. **115**(20): p. 6790-6805.
62. Goyal, P., et al., *Proton Storage Site in Bacteriorhodopsin: New Insights from Quantum Mechanics/Molecular Mechanics Simulations of Microscopic pKa and Infrared Spectra*. *Journal of the American Chemical Society*, 2011. **133**(38): p. 14981-14997.
63. Riccardi, D., et al., *Toward molecular models of proton pumping: Challenges, methods and relevant applications*. *Science China Chemistry*, 2012. **55**(1): p. 3-18.
64. Gaus, M., Q. Cui, and M. Elstner, *DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)*. *Journal of Chemical Theory and Computation*, 2011. **7**(4): p. 931-948.
65. Cui, Q., et al., *A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method*. *The Journal of Physical Chemistry B*, 2000. **105**(2): p. 569-585.
66. Schaefer, P., D. Riccardi, and Q. Cui, *Reliable treatment of electrostatics in combined QM/MM simulation of macromolecules*. *The Journal of Chemical Physics*, 2005. **123**(1): p. 014905-14.

67. König, P.H., et al., *A Critical Evaluation of Different QM/MM Frontier Treatments with SCC-DFTB as the QM Method*. The Journal of Physical Chemistry B, 2005. **109**(18): p. 9082-9095.
68. Bennett, M., et al., *Asymmetric stability among the transmembrane helices of lactose permease*. Biochemistry, 2006. **45**: p. 8088-8095.
69. Klauda, J.B. and B.R. Brooks, *Sugar binding in lactose permease: Anomeric state of a disaccharide influences binding structure*. Journal of Molecular Biology, 2007. **367**(5): p. 1523-1534.
70. Yin, Y., et al., *Sugar binding and protein conformational changes in lactose permease*. Biophysical Journal, 2006. **91**(11): p. 3972-3985.
71. Law, C.J., et al., *Salt-bridge Dynamics Control Substrate-induced Conformational Change in the Membrane Transporter GlpT*. Journal of Molecular Biology, 2008. **378**(4): p. 826-837.
72. Law, C.J., et al., *Structural Basis of Substrate Selectivity in the Glycerol-3-Phosphate: Phosphate Antiporter GlpT*. 2009. **97**(5): p. 1346-1353.
73. Holyoake, J. and M.S.P. Sansom, *Conformational change in an MFS protein: MD simulations of LacY*. Structure, 2007. **15**(7): p. 873-884.
74. Jensen, M.O., et al., *Sugar Transport across Lactose Permease Probed by Steered Molecular Dynamics*. Biophysical Journal, 2007. **93**: p. 92-102.
75. Majumdar, D.S., et al., *Single-molecular FRET reveals sugar-induced conformational dynamics in LacY*. Proceedings from the National Academy of Science, 2007. **104**(31): p. 12640-12645.
76. Smirnova, I., et al., *Sugar binding induces an outward facing conformation of LacY*. Proceedings of the National Academy of Sciences, 2007. **104**(42): p. 16504-16509.
77. Zhou, Y., et al., *Opening and closing of the periplasmic gate in lactose permease*. Proceedings from the National Academy of Science, 2008. **105**: p. 3774-3778.
78. Enkavi, G. and E. Tajkhorshid, *Simulation of Spontaneous Substrate Binding Revealing the Binding Pathway and Mechanism and Initial Conformational Response of GlpT*. Biochemistry, 2010. **49**: p. 1105-1114.
79. Radestock, S. and L.R. Forrest, *The Alternating-Access Mechanism of MFS Transporters Arises from Inverted-Topology Repeats*. Journal of Molecular Biology, 2011. **407**(5): p. 698-715.
80. Andersson, M., et al., *Proton-Coupled Dynamics in Lactose Permease*. Structure (London, England : 1993), 2012. **20**(11): p. 1893-1904.
81. Pendse, P.Y., B.R. Brooks, and J.B. Klauda, *Probing the Periplasmic-Open State of Lactose Permease in Response to Sugar Binding and Proton Translocation*. Journal of Molecular Biology, 2010. **404**(3): p. 506-521.
82. Wu, X.W. and B.R. Brooks, *Self-guided Langevin dynamics simulation method*. Chemical Physics Letters, 2003. **381**(3-4): p. 512-518.
83. Damjanovic, A., et al., *Backbone Relaxation Coupled to the Ionization of Internal Groups in Proteins: A Self-Guided Langevin Dynamics Study*. Biophysical Journal, 2008. **95**(9): p. 4091-4101.

84. Damjanovic, A., E.B. García-Moreno, and B.R. Brooks, *Self-guided Langevin dynamics study of regulatory interactions in NtrC*. Proteins: Structure, Function, and Bioinformatics, 2009. **76**(4): p. 1007-1019.
85. Brooks, B.R., et al., *CHARMM - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
86. Brooks, B.R., et al., *CHARMM: The biomolecular simulation program*. Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
87. Klauda, J.B., et al., *An Ab Initio Study on the Torsional Surface of Alkanes and its Effect on Molecular Simulations of Alkanes and a DPPC Bilayer*. Journal of Physical Chemistry B, 2005. **109**(11): p. 5300-5311.
88. Klauda, J.B., R.W. Pastor, and B.R. Brooks, *Adjacent gauche stabilization in linear alkanes: Implications for polymer models and conformational analysis*. Journal of Physical Chemistry B, 2005. **109**(33): p. 15684-15686.
89. Buck, M., et al., *Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme*. Biophysical Journal, 2006. **90**(4): p. L36-L38.
90. Feig, M., A.D. MacKerell, Jr., and C.L. Brooks, *Force field influence on the observation of π -helical protein structures in molecular dynamics simulations*. Journal of Physical Chemistry B, 2003. **107**(12): p. 2831-2836.
91. MacKerell, A.D., M. Feig, and C.L. Brooks, *Improved treatment of the protein backbone in empirical force fields*. Journal of the American Chemical Society, 2004. **126**(3): p. 698-699.
92. Kuttel, M., J.W. Brady, and K.J. Naidoo, *Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations*. Journal of Computational Chemistry, 2002. **23**(13): p. 1236-1243.
93. Durell, S.R., B.R. Brooks, and A. Bennaïm, *Solvent-Induced Forces between Two Hydrophilic Groups*. Journal of Physical Chemistry, 1994. **98**(8): p. 2198-2202.
94. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
95. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an NLog(N) Method for Ewald Sums in Large Systems*. Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
96. Ryckaert, J.P., G. Ciccotti, and H.J.C. Berendsen, *Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes*. Journal of Computational Physics, 1977. **23**: p. 327-341.
97. Nosé, S. and M.L. Klein, *A Study of Solid and Liquid Carbon Tetrafluoride Using the Constant Pressure Molecular-Dynamics Technique*. Journal of Chemical Physics, 1983. **78**(11): p. 6928-6939.
98. Andersen, H.C., *Molecular-Dynamics Simulations at Constant Pressure and/or Temperature*. Journal of Chemical Physics, 1980. **72**(4): p. 2384-2393.

99. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions*. Physical Review A, 1985. **31**(3): p. 1695-1697.
100. Abramson, J., S. Iwata, and H.R. Kaback, *Lactose permease as a paradigm for membrane transport proteins - (Review)*. Molecular Membrane Biology, 2004. **21**(4): p. 227-236.
101. Khalili-Araghi, F., et al., *Molecular dynamics simulations of membrane channels and transporters*. Current opinion in structural biology, 2009. **19**(2): p. 128-37.
102. Im, W., M. Feig, and C.L. Brooks, *An implicit membrane generalized born theory for the study of structure, stability, and interactions of membrane proteins*. Biophysical Journal, 2003. **85**(5): p. 2900-2918.
103. Tanizaki, S. and M. Feig, *A generalized Born formalism for heterogeneous dielectric environments: Application to the implicit modeling of biological membranes*. Journal of Chemical Physics, 2005. **122**(12).
104. Tanizaki, S. and M. Feig, *Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model*. Journal of Physical Chemistry B, 2006. **110**(1): p. 548-556.
105. Edholm, O. and F. Jahnig, *The Structure of A Membrane-Spanning Polypeptide Studied by Molecular-Dynamics*. Biophysical Chemistry, 1988. **30**(3): p. 279-292.
106. Xie, J., et al., *Phosphatidylethanolamine and monoglucosyldiacylglycerol are interchangeable in supporting topogenesis and function of the polytopic membrane protein lactose permease*. Journal of Biological Chemistry, 2006. **281**(28): p. 19172-19178.
107. Wang, X.Y., M. Bogdanov, and W. Dowhan, *Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition*. EMBO Journal, 2002. **21**(21): p. 5673-5681.
108. Bogdanov, M., P.N. Heacock, and W. Dowhan, *A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition*. EMBO Journal, 2002. **21**(9): p. 2107-2116.
109. Shaikh, S.R., et al., *Monounsaturated PE does not phase-separate from the lipid raft molecules sphingomyelin and cholesterol: Role for polyunsaturation?* Biochemistry, 2002. **41**(34): p. 10593-10602.
110. Dolan, E.A., et al., *Simulations of membranes and other interfacial systems using P2₁ and pc periodic boundary conditions*. Biophysical Journal, 2002. **82**(5): p. 2317-2325.
111. Smart, O.S., J.M. Goodfellow, and B.A. Wallace, *The Pore Dimensions of Gramicidin A*. Biophysical Journal, 1999. **65**: p. 2455.
112. Lee, B. and F.M. Richards, *The interpretation of protein structures: Estimation of static accessibility*. Journal of Molecular Biology, 1971. **55**(3): p. 379-400.
113. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.
114. Nie, Y., N. Ermolova, and H.R. Kaback, *Site-directed Alkylation of LacY: Effect of the Proton Electrochemical Gradient*. Journal of Molecular Biology, 2007. **374**: p. 356-364.

115. Kaback, H.R., et al., *Site-directed alkylation and the alternating access model for LacY*. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(2): p. 491-494.
116. Zhou, Y., Y. Nie, and H.R. Kaback, *Residues Gating the Periplasmic Pathway of LacY*. Journal of Molecular Biology, 2009. **394**: p. 219-225.
117. Ermolova, N., M.R. V., and H.R. Kaback, *Site-Directed Alkylation of Cysteine Replacements in the Lactose Permease of Escherichia coli: Helices I, III, VI and XI*. Biochemistry, 2006. **45**: p. 4182-4189.
118. Frillingos, S. and H.R. Kaback, *The Role of Helix VIII in the Lactose Permease of Escherichia coli: II. Site-Directed Sulphydryl Modification*. Protein Science, 1997. **6**(2): p. 438-443.
119. Kwaw, I., et al., *Site-Directed Sulphydryl Labeling of the Lactose Permease of Escherichia coli: Helices IV and V That Contain the Major Determinants for Substrate Binding*. Biochemistry, 2001. **40**: p. 10491-10499.
120. Venkatesan, P., Y. Hu, and H.R. Kaback, *Site-Directed Sulphydryl Labeling of the Lactose Permease of Escherichia coli: Helix X*. Biochemistry, 2000. **39**: p. 10656-10661.
121. Venkatesan, P., et al., *Site-Directed Sulphydryl Labeling of the Lactose Permease of Escherichia coli: Helix VII*. Biochemistry, 2000. **39**: p. 10641-10648.
122. Zhang, W., Y. Hu, and H.R. Kaback, *Site-Directed Sulphydryl Labeling of Helix IX in the Lactose Permease of Escherichia coli*. Biochemistry, 2003. **42**: p. 4904-4908.
123. Venkatesan, P., et al., *Site-Directed Sulphydryl Labeling of the Lactose Permease of Escherichia coli: N-Ethylmaleimide-Sensitive Face of Helix II*. Biochemistry, 2000. **39**: p. 10649-10655.
124. Green, A.L., E.J. Anderson, and R.J. Brooker, *A Revised Model for the Structure and Function of the Lactose Permease*. The journal of Biological Chemistry, 2000. **275**: p. 23240–23246.
125. Green, A.L., H.A. Hrodey, and R.J. Brooker, *Evidence for Structural Symmetry and Functional Asymmetry in the Lactose Permease of Escherichia coli*. Biochemistry, 2003. **42**: p. 11226-11233.
126. Nie, Y., F.E. Sabetfard, and H.R. Kaback, *The Cys154-->Gly mutation in LacY causes constitutive opening of the hydrophilic periplasmic pathway*. Journal of Molecular Biology, 2008. **379**: p. 695-703.
127. Nie, Y., Y. Zhou, and H.R. Kaback, *Clogging the Periplasmic Pathway in LacY*. Biochemistry, 2009. **48**: p. 738–743.
128. Frillingos, S., et al., *Cys-scanning mutagenesis: a novel approach to structure–function relationships in polytopic membrane proteins*. The FASEB Journal, 1998. **12**(13): p. 1281-1299.
129. Abramson, J., S. Iwata, and H.R. Kaback, *Lactose permease as a paradigm for membrane transport proteins (Review)*. Molecular membrane biology, 2004. **21**(4): p. 227-36.
130. Klauda, J.B. and B.R. Brooks, *Sugar binding in lactose permease: anomeric state of a disaccharide influences binding structure*. Journal of molecular biology, 2007. **367**(5): p. 1523-34.

131. Sahin-Toth, M., et al., *Ligand Recognition by the Lactose Permease of Escherichia coli: Specificity and Affinity Are Defined by Distinct Structural Elements of Galactopyranosides*. *Biochemistry*, 2000. **39**(17): p. 5097-5103.
132. Green, A.L., H.a. Hrodey, and R.J. Brooker, *Evidence for structural symmetry and functional asymmetry in the lactose permease of Escherichia coli*. *Biochemistry*, 2003. **42**(38): p. 11226-33.
133. Nie, Y., et al., *Energetics of ligand-induced conformational flexibility in the lactose permease of Escherichia coli*. *Journal of biological chemistry*, 2006. **281**(47): p. 35779-35784.
134. Klauda, J.B. and B.R. Brooks, *CHARMM Force Field Parameters for Nitroalkanes and Nitroarenes*. *Journal of Chemical Theory and Computation*, 2007. **4**(1): p. 107-115.
135. Grosdidier, A., V. Zoete, and O. Michielin, *SwissDock, a protein-small molecule docking web service based on EADock DSS*. *Nucleic Acids Research*, 2011. **39**(suppl 2): p. W270-W277.
136. Grosdidier, A., V. Zoete, and O. Michielin, *Fast docking using the CHARMM force field with EADock DSS*. *Journal of Computational Chemistry*, 2011. **32**(10): p. 2149-2159.
137. Guvench, O., et al., *Additive empirical force field for hexopyranose monosaccharides*. *Journal of Computational Chemistry*, 2008. **29**(15): p. 2543-2564.
138. Raman, E.P., O. Guvench, and A.D. MacKerell, *CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses*. *The Journal of Physical Chemistry B*, 2010. **114**(40): p. 12981-12994.
139. Pandit, K.R. and J.B. Klauda, *Membrane models of E. coli containing cyclic moieties in the aliphatic lipid chain*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 2012. **1818**(5): p. 1205-1210.
140. Deng, Y. and B. Roux, *Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant*. *Journal of Chemical Theory and Computation*, 2006. **2**(5): p. 1255-1273.
141. Kumar, S., et al., *The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method*. *Journal of Computational Chemistry*, 1992. **13**(8): p. 1011-1021.
142. Forrest, L.R., R. Krämer, and C. Ziegler, *The structural basis of secondary active transport mechanisms*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2011. **1807**(2): p. 167-188.
143. Jiang, X., et al., *Evidence for an intermediate conformational state of LacY*. *Proceedings of the National Academy of Sciences*, 2012.
144. Smirnova, I.N., V.N. Kasho, and H.R. Kaback, *Direct Sugar Binding to LacY Measured by Resonance Energy Transfer†*. *Biochemistry*, 2006. **45**(51): p. 15279-15287.
145. Smirnova, I.N., V. Kasho, and H.R. Kaback, *Protonation and sugar binding to LacY*. *Proceedings of the National Academy of Sciences*, 2008. **105**(26): p. 8896-8901.
146. Soskine, M., Y. Adam, and S. Schuldiner, *Direct evidence for substrate-induced proton release in detergent-solubilized EmrE, a multidrug*

- transporter*. The Journal of biological chemistry, 2004. **279**(11): p. 9951-9955.
147. Smirnova, I., et al., *Role of protons in sugar binding to LacY*. Proceedings of the National Academy of Sciences, 2012.
 148. Franco, P., et al., *A Suppressor Analysis of Residues Involved in Cation Transport in the Lactose Permease: Identification of a Coupling Sensor*. The Journal of Membrane Biology, 2006. **211**(2): p. 101-113.
 149. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields*. Journal of Computational Chemistry, 2010. **31**(4): p. 671-690.
 150. Klauda, J.B., et al., *Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types*. The Journal of Physical Chemistry B, 2010. **114**(23): p. 7830-7843.
 151. Steinbach, P.J. and B.R. Brooks, *New spherical-cutoff methods for long-range forces in macromolecular simulation*. J. Comput. Chem., 1994. **15**(7): p. 667-683.
 152. Sahin-Tóth, M. and H.R. Kaback, *Arg-302 facilitates deprotonation of Glu-325 in the transport mechanism of the lactose permease from Escherichia coli*. Proceedings of the National Academy of Sciences, 2001. **98**(11): p. 6068-6073.
 153. Jesús, J.-B., et al., *Protein-Carbohydrate Interactions: A Combined Theoretical and NMR Experimental Approach on Carbohydrate-Aromatic Interactions and on Pyranose Ring Distortion*, in *NMR Spectroscopy and Computer Modeling of Carbohydrates*. 2006, American Chemical Society. p. 60-80.
 154. Krishnamurthy, H. and E. Gouaux, *X-ray structures of LeuT in substrate-free outward-open and apo inward-open states*. Nature, 2012. **481**(7382): p. 469-474.
 155. Shimamura, T., et al., *Molecular Basis of Alternating Access Membrane Transport by the Sodium-Hydantoin Transporter Mhp1*. Science, 2010. **328**(5977): p. 470-473.