

# **Structure and dynamics of outer membrane lectin involved in the surface expression of Group 1 capsular polysaccharides**

**Sanjana Anilkumar Nair**

**BO12M1005**

A Dissertation Submitted to  
Indian Institute of Technology Hyderabad  
In Partial Fulfillment of the Requirements for  
The Degree of Master of Technology



**भारतीय प्रौद्योगिकी संस्थान हैदराबाद**  
**Indian Institute of Technology Hyderabad**

**Department of Biotechnology**

**July, 2014**

## **Declaration**

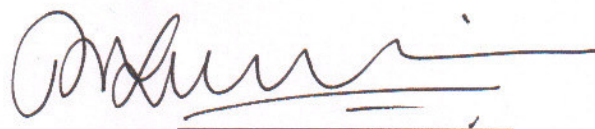
I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

Sanjana Anilkumar Nair

BO12M1005

## Approval Sheet

This thesis entitled “Structure and dynamics of outer membrane lectin involved in the surface expression of Group 1 capsular polysaccharides” by Sanjana Anilkumar Nair is approved for the degree of Master of Technology from IIT Hyderabad.



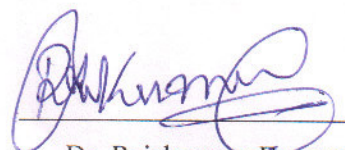
Dr. Thenmalarchelvi Rathinavelan  
(Thesis advisor)  
Department of Biotechnology  
IIT Hyderabad



Dr. D.S. Sharada  
Department of Chemistry  
IIT Hyderabad



Dr. Anindya Roy  
Department of Biotechnology  
IIT Hyderabad



Dr. Rajakumara Eerappa  
Department of Biotechnology  
IIT Hyderabad

## **Acknowledgements**

I would like to extend my gratitude to all individuals who have helped me with their kind support and guidance.

I am highly indebted to my advisor, Dr. Thenmalarchelvi Rathinavelan, Department of Biotechnology, IIT Hyderabad, for her valuable guidance and support in completing the project. I thank all the committee members for their kind cooperation and encouragement which helped me in completion of this project.

Furthermore, I extend my gratitude to my lab members for their help and support for completion of this work.

## Abstract

Multi-drug resistance of Gram-negative bacteria such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* is the major cause for human mortality worldwide. Bacterial capsular polysaccharides (CPS) are important virulence determinants of such Gram-negative bacteria, helping in colonization and offering antibacterial drug resistance. Though the outer membrane protein Wzi has been identified as responsible for surface anchorage of the Group 1 capsule of *E. coli*, the mechanism involved remains unknown. In this context, we carry out molecular dynamics simulations of Wzi protein and its mutants. Our results for the first time reveal, the water conducting property of Wzi protein. Two residues, namely Tyr380 and Trp39 at the extracellular and periplasmic faces respectively are identified to play an important role in regulating water conduction. A potassium ion binding pocket is observed on the extracellular face of the protein, coinciding with the water entry region. We hypothesize that this may be the interacting region for CPS onto the bacterial surface through non-covalent interactions. We also model the extracellular loop L5, which is absent in the crystal structure. Simulations show that it plays a major role in anchoring the protein to the lipid membrane. In addition, a hexameric repeating unit of Group 1 capsular polysaccharide,  $\rightarrow 2)$ - $\alpha$ -D-Manp-(1 $\rightarrow$ 3)- $\beta$ -D-Galp-(1 $\rightarrow$  chain carrying a  $\beta$ -D-GlcUAp-(1 $\rightarrow$ 3)- $\alpha$ -D-Galp-(1 $\rightarrow$  branch at position 3 of the mannose is also be simulated and is found to exist in two major conformations, namely bent and extended. The outcome of this study may facilitate the design of antibacterial drugs that can specifically alter the water conductance of Wzi. This may further perturb CPS binding to the bacterial surface and thus lead to decreased virulence and higher antibiotic susceptibility.

## Abbreviations

ABC transporter – ATP Binding Cassette transporter  
ABNR – Adopted basis Newton Raphson  
BLAST – Basic Local Alignment Search Tool  
CHARMM – Chemistry at HARvard Molecular Mechanics  
CPS – Capsular Polysaccharide  
CPT – Constant Pressure and Temperature  
Gal – Galactose  
Gal-1-P – Galactose-1-phosphate  
GlcA – Glucuronic acid  
IBC – Intracellular Biofilm like communities  
LJ – Lennard Jones  
LPS – Lipopolysaccharide  
Man – Mannose  
MD – Molecular Dynamics  
MDR – Multiple Drug Resistance  
PBC – Periodic Boundary Conditions  
PME – Particle Mesh Ewald  
POPE – Palmitoyloleoylphosphatidylethanolamine  
POPG – Palmitoyloleoylphosphatidylglycerol  
SD – Steepest Descent  
UB – Urey Bradley  
UDP- Galactose – Uridine Diphosphate Galactose  
Und-pp – Undecaprenyl diphosphate  
UPEC – Uropathogenic *E. coli*  
UTI – Urinary Tract Infections  
WHO – World Health Organization

## Preface

Capsular polysaccharides are major virulence determinants of Gram-negative bacteria, aiding in colonization and protecting the bacteria from the host immune system. In the present study, we focus on the group 1 K30 capsule of *E. coli*. K30 CPS is synthesized by a Wzy-dependent mechanism, in which polysaccharide monomers formed in the cytoplasm are polymerized in the periplasmic space and transported to the exterior of the cell, where they are anchored to the cell surface. The outermembrane protein, Wzi, has been implicated in the surface anchorage of CPS, but the mechanism involved is unknown. The basic goal of the present investigation is to explore the mechanism of surface anchorage of K30 CPS to the bacterial surface. To this end, we carry out molecular dynamics simulations of Wzi protein and its mutants. We also carry out molecular dynamics simulations of a hexameric unit of K30 CPS, to understand its structural preference. The rationale behind this approach is that once the binding mechanism is understood, “Wzi-blocker” molecules can be designed, to perturb CPS attachment and concomitantly decrease bacterial virulence and increase their susceptibility to antibacterial drugs.

The crystal structure of Wzi has 15 residues missing from an extracellular loop, L5. Bioinformatics analysis shows these residues to be highly conserved across several *E. coli* and *Klebsiella* species. We model these residues and carry out simulations of wild type Wzi. Comparison of simulations of the wild type protein and an L5 deletion mutant indicates the main role of L5 to be membrane insertion.

From the simulations of wild type Wzi protein, we observe for the first time, the water conducting property of Wzi protein. Though the beta barrel is occluded at both ends, it conducts water through a specific path, dictated by a hydrophobic plug. The presence of water molecules within the beta barrel of the Wzi crystal structure validates this observation. Tyr380 on the extracellular side is identified as a vital residue in facilitating water entry into the channel, with its point mutant Y380W showing a marked decrease in water conduction. We hypothesize that as the CPS translocated to the cell exterior accumulates, the concentration gradient created must drive water molecules from the periplasm to the extracellular space through Wzi. Such a hydrophilic environment within the protein may provide a suitable environment for the highly hydrated CPS to interact with it. Thus, the water conductance of Wzi is of high biological relevance.

Trp39, which is part of helix H1 at the periplasmic side, is also important in regulating water entry into the channel. Two helix substitution mutants, in which this residue is mutated to glycine show an increased influx of water from the periplasmic end. We hypothesize that the increased transport of water to the cell exterior may hinder binding of the CPS rendering the protein non-functional. These findings are supported by previous mutational studies by Bushell and co-workers in 2013, where the two helix substitution mutants were found to be non-functional. The reason behind this was unknown at the time.

We also observe the presence of a potassium ion binding pocket on the extracellular side. A single potassium ion enters into the pocket at around 3 ns and remains stably associated with it upto the end of the simulation timescale of 50 ns. This positively charged ion may facilitate the attachment of the negatively charged CPS to the protein. From simulations of wild type Wzi protein carried out at a range of transmembrane voltages we conclude that it does not exhibit any voltage dependent properties.

Simulations of a hexameric unit of K30 CPS show two main structural conformations, namely an extended form and a bent form. Preferred values of ( $\psi$ ,  $\phi$ ) for each of the four linkages over the simulation timescale are also determined.

Chapter 1 is a general introduction to the subject area. A brief introduction on the relevance of this area of study with emphasis on the emergence of multi-drug resistant strains is provided. The importance of the bacterial capsule, its classification and biogenesis is also discussed. Group 1 capsule biosynthetic mechanism and the outer membrane lectin involved in surface anchorage of CPS, which forms the focus of this study is discussed in detail. In Chapter 2, an introduction to computational techniques used in this study and the CHARMM force field is provided, followed by a detailed explanation of system setup and simulations carried out as part of this study. In Chapters 3 and 4, the results of MD simulations of Wzi protein and K30 CPS respectively are elucidated.

### **Poster presentation**

“Dual role revealed: *E. coli* outer membrane lectin is also an osmotic pressure regulator” in “*The first annual TCIS summer research symposium*”, TIFR-TCIS Hyderabad, June 26, 2014.



# Contents

Declaration.....	i
Approval Sheet.....	ii
Acknowledgements.....	iii
Abstract.....	iv
Abbreviations.....	v
Preface.....	vi
<b>1 Introduction.....</b>	<b>1</b>
1.1 Gram-negative bacteria.....	4
1.1.1 Polysaccharides produced by Gram-negative bacteria.....	5
1.2 Bacterial capsular polysaccharide.....	6
1.2.1 Functions of the capsule.....	6
1.2.2 E. coli as a model organism.....	8
1.2.3 E. coli serotyping.....	9
1.2.4 Classification of E. coli capsules.....	9
1.3 Group 1 capsules.....	11
1.3.1 Structure of K30 CPS.....	12
1.3.2 Method of K30 capsule biogenesis.....	13
1.3.2.1 Wzy dependent supramolecular assembly.....	15
1.4 Structure of Wzi.....	17
1.5 Challenges in the study of membrane proteins.....	18
1.6 Scope of the study.....	20
1.7 References.....	21
<b>2 Methods.....</b>	<b>23</b>
2.1 Historical perspective on classical MD simulations.....	24
2.2 Force fields.....	24
2.3 CHARMM.....	25
2.3.1 Potential energy function of CHARMM.....	26
2.3.2 Computation of non-bonded interactions.....	27
2.3.2.1 Lennard-Jones interactions.....	28
2.3.2.2 Electrostatic interactions.....	29
2.3.3 Solvation of the system.....	29
2.3.4 Periodic boundary conditions in CHARMM.....	30
2.3.5 Energy minimization.....	31

2.3.6	Underlying principles of molecular dynamics .....	31
2.3.7	Numerical integrators in CHARMM .....	32
2.3.8	Thermostats in CHARMM.....	34
2.4	Generation of protein-bilayer models.....	34
2.5	MD simulations of Wzi protein.....	37
2.5.1	Loop modelling of Wzi .....	37
2.5.2	Wzi models for simulation .....	38
2.5.3	Protocol followed during equilibration .....	40
2.5.4	Protocol followed for production run .....	41
2.5.5	Application of transmembrane voltage .....	41
2.6	MD simulations of K30 CPS .....	42
2.7	References.....	43
<b>3</b>	<b>Structure and dynamics of K30 Capsular polysaccharide .....</b>	<b>45</b>
3.1	Introduction.....	46
3.2	Dynamics of K30 CPS.....	47
3.3	Conclusions.....	49
3.4	References.....	50
<b>4</b>	<b>Structural characterization of Wzi .....</b>	<b>51</b>
4.1	Introduction.....	52
4.2	Water conducting property of Wzi protein .....	55
4.2.1	Comparison with crystal structure of Wzi.....	56
4.2.2	Role of hydrophobic plug in regulating water conduction.....	58
4.2.3	Entry of water from the extracellular side.....	58
4.2.3.1	Simulations of Y380 mutants .....	59
4.2.4	Entry of water from the periplasmic side .....	61
4.2.5	Secondary structural changes .....	62
4.2.6	Pore opening .....	62
4.3	Ion binding pocket of Wzi .....	63
4.4	Importance of extracellular loop 5 .....	68
4.4.1	Membrane insertion .....	68
4.4.2	Bioinformatics analysis .....	71
4.4.3	L5 deletion simulations .....	73
4.5	Other membrane protein interactions of Wzi.....	75
4.6	Helix mutants of Wzi.....	76

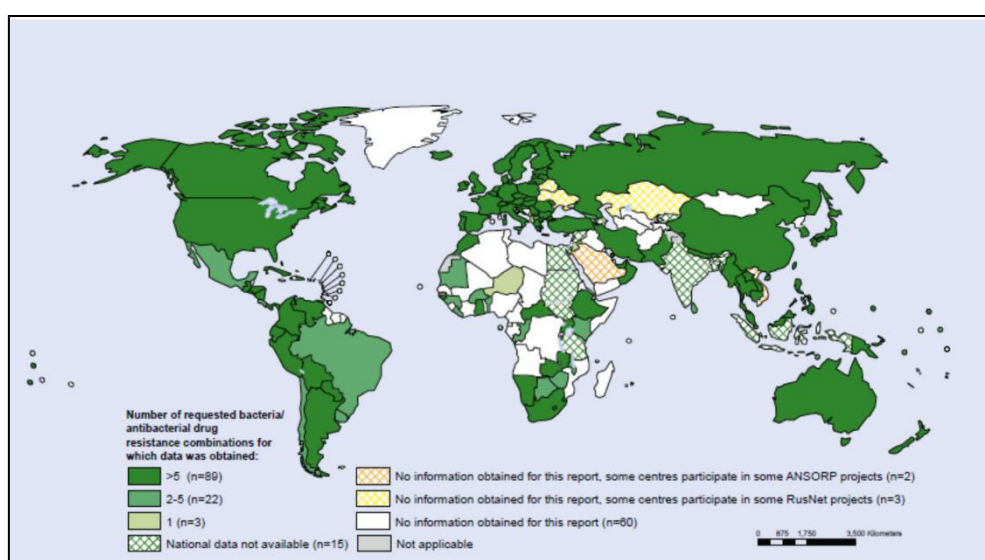
4.6.1	Secondary structure.....	77
4.6.2	Water conduction.....	79
4.6.3	Role of Trp39 in water regulation at the periplasmic end.....	80
4.7	Voltage independent water conduction of Wzi.....	82
4.8	Conclusions.....	84
4.9	References.....	85

# **CHAPTER 1**

## **Introduction**

## CHAPTER 1 : Introduction

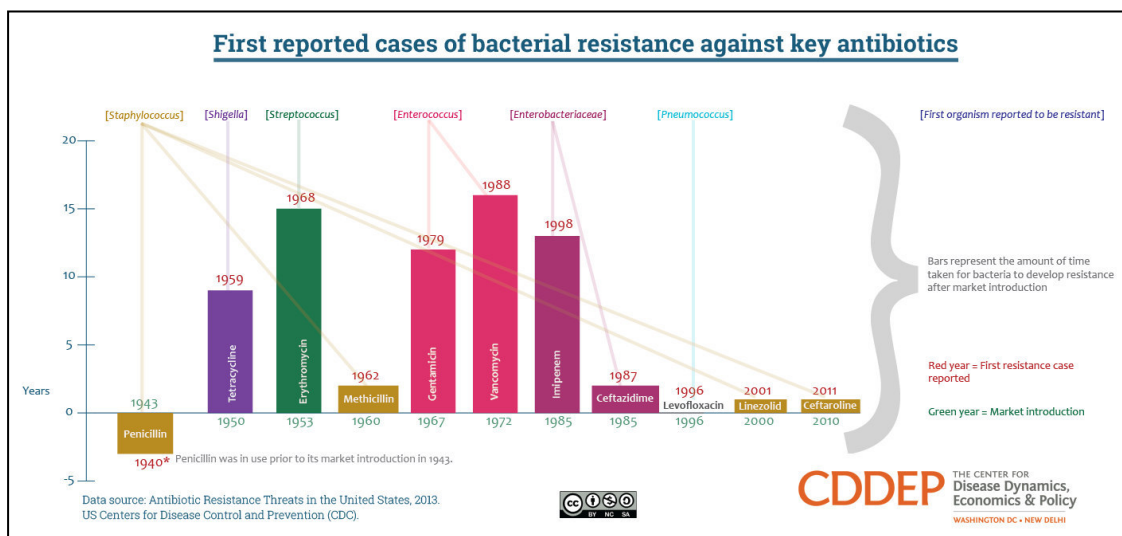
World Health Organization (WHO) had identified Anti-Microbial Resistance (AMR) , viz., bacterial, fungal, viral and parasitic drug resistance, as one of the major problems for human health (Bassetti et al., 2011). Bacterial infectious diseases are one of the leading causes of human death and disability across the globe, due to the emergence of new drug resistance mechanisms. In fact, WHO has identified antibiotic resistance as “no longer a prediction for the future, but happening right now across the world.” (*Antimicrobial resistance, global report on surveillance* (2014) by WHO). Figure 1.1 depicts the emergence of multi drug resistant bacterial strains across the globe. The extremely high distribution of drug resistant strains emphasizes the danger they pose to the human society.



**Figure 1.1** Number of reported bacterial-antibacterial drug resistant combinations across the globe, 2013. *E. coli* resistant to cephalosporins and fluoroquinolones, *K. pneumoniae* resistant to cephalosporins, carbapenems, Methicillin resistant *S.aureus*, *S. pneumoniae* resistant to penicillin, *Salmonella* and *Shigella* species resistant to fluoroquinolones and *N. gonorrhoeae* resistant to cephalosporins are considered (Adopted from *Antimicrobial resistance, global report on surveillance* (2014) by WHO).

Emergence of Multi Drug Resistant (MDR) strains of bacteria and the lack of novel vaccines (prevention) and antibacterial drugs (treatment) pose a significant threat. Bacteria responsible for majority of the MDR infections caused today have been grouped together as the ‘ESKAPE’ pathogens, namely, *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species

(Boucher et al., 2009). The most severe threat posed by these “superbugs” is in nosocomial settings, where they are responsible for urinary tract infections (UTI’s), lower respiratory tract infections, pneumonia and bacteremia in immunocompromised patients (Bassetti et al., 2011). In fact, the risk of emergence of MDR strains is very high in developing countries, where hospital practises are not standardized and several antibiotics can be easily purchased off the counter without a medical prescription. The emergence of resistance to key antibiotics is depicted as a timeline in Figure 1.2. The problem of MDR is even more severe in case of Gram-negative pathogens as discussed in the following section.



**Figure 1.2 Emergence of bacterial resistance to key antibiotics.** Starting from 1963 (the year penicillin was introduced in the market), the emergence of resistance to key antibiotics by various bacterial strains is traced. The X axis depicts the year of market entry of the antibiotic and the Y axis depicts the number of years from market introduction at which the first case of resistance was reported (Adopted from <http://www.cddep.org/>).

At present, the infections caused by Gram-negative bacteria pose a much higher risk. In some parts of the world, more than half of the infections are caused by Gram-negative bacteria, such as *E. coli* and *Klebsiella pneumoniae* that are resistant to carbapenem, a last resort antibiotic (*Nature news* doi:10.1038/nature.2014.15135) (<http://www.nature.com/news/who-warns-against-post-antibiotic-era-1.15135>). Notably, four of the six ‘ESKAPE’ pathogens, namely *Klebsiella*, *Acinetobacter*, *Pseudomonas* and *Enterobacter* species are Gram-negative pathogens. These Gram-negative MDR strains have developed multiple mechanisms of resistance to the antibiotics used as listed in Table 1.1.

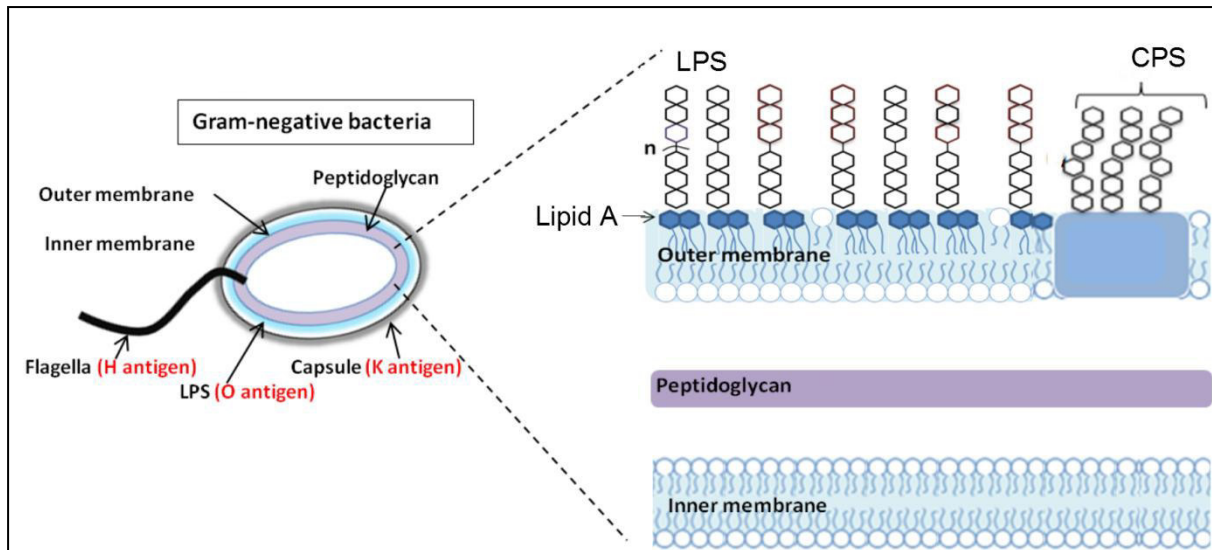
Gram-negative bacteria are found to be more resistant to antibiotics and chemotherapeutic agents compared to Gram-positive species, mainly due to the presence of the outer membrane (Figure 1.3) which acts as a physical barrier (Nikaido, 1998; Silhavy et al., 2010). They are capable of upregulating or acquiring genes that code for antibiotic resistance with very high efficiency. Consequently, the increase in resistance of Gram-negative bacteria to drugs is much faster compared to that in Gram-positive bacteria (Peleg and Hooper, 2010). In addition, there are very few new antibiotics that are active against Gram-negative strains. Some of the drugs are still at a very early stage of development and will not be able to provide sufficient therapeutic cover in the near future (Kollef et al., 2011; Kumarasamy et al., 2010).

Table 1.1 MDR Gram-negative bacteria and mechanisms of resistance (Kanj and Kanafani, 2011; Peleg and Hooper, 2010)		
Pathogenic bacteria	Antibiotics in use	Mechanisms of resistance
<i>Pseudomonas aeruginosa</i>	Beta lactams	Cephalosporinases, Extended spectrum
<i>Enterobacter</i>	Penicillins	beta-lactamases, Carbapenemases
<i>Klebsiella</i>	Cephalosporins	
<i>Acinetobacter</i>	Carbapenems	
<i>P.aeruginosa</i>	Quinolone/	Mutations in target of antibiotic
<i>Acinetobacter</i>	Fluoroquinolones	
<i>E. coli</i>		
<i>P.aeruginosa</i>		Modification of aminoglycoside by
<i>Acinetobacter</i>	Aminoglycosides	bacterial enzymes

Thus, the emergence of such Gram-negative MDR strains and the lack of treatment options is effectively catapulting us back to a ‘pre-antibiotic’ era and there is a need to develop novel strategies and targets to overcome such infections.

## 1.1 Gram-negative bacteria

Figure 1.3 shows a schematic representation of the surface of a Gram-negative bacterium depicting the major surface polysaccharides which also function as important surface virulence determinants and form the basis for its serological classification.



**Figure 1.3 Schematic representation of a Gram-negative bacterium (left) and a cross section of outer membrane (right).** (Left) Inner membrane (IM), peptidoglycan (PG) layer and outer membrane (OM) along with the O, K and H antigens are marked. (Right) Surface associated LPS and CPS virulent polysaccharides are shown. Refer text for details.

### 1.1.1 Polysaccharides produced by Gram-negative bacteria

Bacteria convert various carbon sources into polymers with distinct properties. The extracellular polymers produced by bacteria can be classified as polysaccharides, polyesters, polyamides and inorganic polyanhydrides. Polysaccharides are the most abundant of all Gram-negative bacterial polymers. They can be intracellular (like glycogen, which is the only intracellular storage polysaccharide found in bacteria and archaea), part of the cell membrane (as the Lipopolysaccharide (LPS) structures in the outer membrane), tightly associated with the surface (like Capsular polysaccharides (CPS)) or loosely associated with the cell surface (such as Exopolysaccharide (EPS) or ‘slime polysaccharide’) (Figure 1.3 (right))(Rehm, 2010). CPS and LPS are major virulence determinants of Gram-negative bacteria (Rehm, 2010). Yet another virulence determinant of Gram-negative bacteria is the H antigen, associated with the flagellar subunit of the bacteria. The LPS, CPS and H antigen form the basis for classification of bacteria. Combinations of these antigens on the bacterial surface give rise to unique ‘Serotypes’. Several studies are being carried out to understand the pathogenic mechanism of Gram-negative bacteria mediated by virulent capsular polysaccharides.



## 1.2 Bacterial capsular polysaccharide

Successful pathogens have evolved over the course of time and developed a range of strategies to help them survive and proliferate. One such factor that plays a major role in bacterial virulence is the capsule. The functions of the bacterial capsule are wide and varied, ranging from acting as a physical barrier to helping in colonization, and most importantly in evasion of immune response (discussed in more detail below and Figure 1.4). It mediates the interaction between the bacteria and host in the initial stage of the infection. The capsule is thus an important target to develop new drug molecules against bacteria. Some of the clinically important encapsulated bacteria and the associated diseases are shown in Table 1.2 (the capsular nomenclature is different in case of *Haemophilus influenzae* alone (a - f), whereas in all other cases the capsules are named as K antigens).

**Table 1.2 Clinically significant encapsulated bacteria & associated diseases (Nwodo et al., 2012)**

Bacterial species	Pathogenic serotypes	Capsular antigen nomenclature	Associated clinical disease
<i>Escherichia coli</i>	>80	K antigen	Diarrhoea, Neonatal meningitis , Urinary tract infection
<i>Haemophilus influenzae</i>	>6	a-f	Meningitis , Epiglottitis, Septicemia , Pneumonia
<i>Neisseria meningitides</i>	>10	K antigen	Meningitis , Meningococemia
<i>Klebsiella pneumoniae</i>	>80	K antigen	Pneumonia , Bacteremia , Thrombophlebitis , Urinary tract infection , Diarrhoea , Upper respiratory tract infection , wound infection , Osteomyelitis , Meningitis and Pyogenic liver abscess

### 1.2.1 Functions of the Capsule

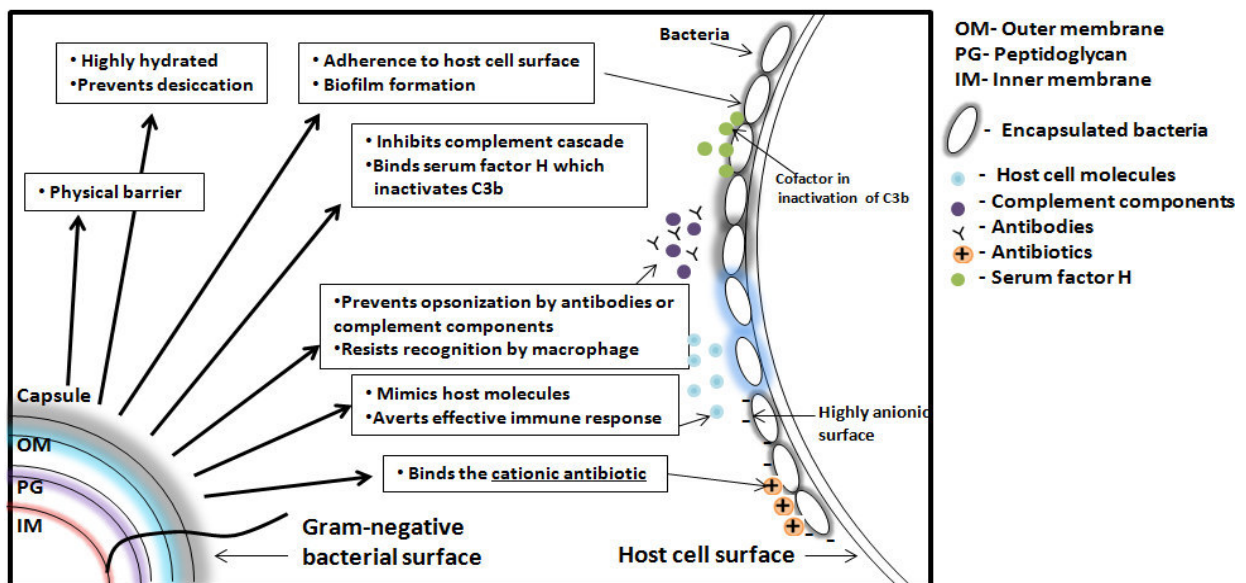
The bacterial capsule plays several roles in bacterial virulence and survival as discussed below.

#### (i) Prevention of desiccation

CPS forms a hydrated gel surface surrounding the bacterial outer membrane and acts as a physical barrier to circumvent desiccation. It also aids the bacterial transmission to the host and, thus, its survival.

## (ii) Adherence and biofilm formation

The capsule aids in adherence of bacteria to surfaces for colonization and also in the adherence of bacteria to each other to promote biofilm formation. The hydrophilic nature of the capsule aids in adherence to surfaces. For example, UPEC strains form intracellular biofilm-like communities (IBC's) within the bladder epithelial cell with the help of type 1 pili and the capsular K antigen. Intracellular signalling and metabolism of the bacterial community within the IBC may lead to release of the bacteria to reinitiate an infection cycle (Goller and Seed, 2010).



**Figure 1.4 Schematic representation of the function of Gram-negative capsular polysaccharide (Legend shown on right).** Serum factor H is shown in green, antibiotics in orange, host cell molecules in light blue, complement components in purple and encapsulated bacteria shown with a grey outline and antibiotics in the characteristic Y shape. OM, IM and PG stands for outer membrane, inner membrane and peptidoglycan respectively.

## (iii) Resistance to non-specific host immunity

The capsule acts as a permeability barrier to the components of the complement cascade. It also provides resistance to complement mediated opsonophagocytosis, the degree of resistance being proportional to the charge on the CPS. For instance, K1 capsule binds to serum factor H (a cofactor that inactivates C3b), thus inhibiting the alternate complement cascade. The capsule can trigger release of cytokines and hinder the coordination of cell

mediated immunity. CPS also confers resistance to antibacterial peptides like lactoferrin (Campos et al., 2004).

**(iv) Resistance to specific host immunity / Molecular mimicry**

Capsules with structural similarity to host cell polysaccharides are poorly immunogenic. For example, the K1 CPS is structurally identical to a carbohydrate found in the Neural cell-adhesion molecule (N-CAM) and K5 polysaccharide is similar to an intermediate in the synthesis of heparin. Such strains are not recognized as foreign by the host immune system are resistant to host cell specific immunity and are hence, very dangerous pathogens (Roberts, 1996).

**(v) Antibiotic resistance**

In addition to acting as a physical barrier, the highly anionic capsule can attract and trap the cationic aminoglycoside antibiotics. Treatment with Kanamycin/Streptomycin results in increased K30 CPS production in *E. coli*. This in turn causes increased antibiotic resistance (Iman Naimi et al., 2009).

*E. coli* is one of the best understood organisms in terms of physiology and biochemistry of CPS and its relevance in bacterial virulence. Thus it is justified to consider *E. coli* as a model organism in the current investigation and further explore the mechanisms of CPS mediated bacterial virulence.

### **1.2.2 *E. coli* as a model organism**

*E. coli* belongs to the *Enterobacteriaceae* family, that includes plant pathogens like *Erwinia*, *Rhizobium*, *Xanthomonas*, human and animal pathogens like *Yersinia*, *Salmonella* and *Klebsiella* (Drummel-Smith and Whitfield, 2000). Various strains of *E. coli* are highly virulent, causing urinary tract infections (UTI), septicemia, meningitis, etc. Pathogenic *E. coli* can be divided into 2 main groups : (i) the Extra-intestinal pathogenic *E. coli* (ExPEC), which include uropathogenic *E. coli* (UPEC) and meningitis associated *E. coli* (MNEC), and (ii) Intestinal pathogenic *E. coli*, which include 6 well described categories, namely, Enteropathogenic *E. coli* (EPEC), Enterohaemorrhagic *E. coli* (EHEC), Enterotoxigenic *E. coli* (ETEC), Enteroaggregative *E. coli* (EAEC), Enteroinvasive *E. coli* (EIEC) and diffusively adherent *E. coli* (DAEC) (Kaper et al., 2004). Thus, it is important to understand the mechanism of virulence exerted by *E. coli*.

### 1.2.3 *E. coli* Serotyping

As mentioned earlier, LPS (O antigen), CPS (K antigen) and H antigens are major virulent polysaccharides of Gram-negative bacteria, based on which, the serotyping of *E. coli* is defined. Serotyping helps distinguish between pathogenic and non-pathogenic strains as specific serotypes are found to be consistently associated with particular diseases.

One of the earliest classification systems of *E. coli* was put forward by Kauffman in 1940 and is still in use in a modified form. The classification is based on O, K and H antigens. As mentioned previously, the O antigen forms part of the Lipopolysaccharide (LPS) endotoxin. It consists of a specific polysaccharide attached to a lipid A-core unit in the outer membrane. K antigen is the ‘Capsular polysaccharide’, which forms a high molecular weight capsule associated with the outer membrane of the bacteria. H antigen is associated with the flagellar subunit of the bacteria. Here, ‘O’ stands for *ohne hauch* [Ger.”without huff” or “without film”], ‘H’ stands for *Hauch*, and ‘K’ for *Kapsel*. The O, K, and H antigens associated with a Gram-negative bacterium have been shown in Figure 1.1. There are about 173 O antigens, 80 K antigens and 56 H antigens found in *E. coli* making its serotyping highly complex. Different combinations of these O, K, and H antigens give rise to more than 50,000 different serotypes (Orskov et al., 1977). As mentioned before, the current focus will be on *E. coli* CPS as it is one of the major virulence factors.

### 1.2.4 Classification of *E. coli* capsules

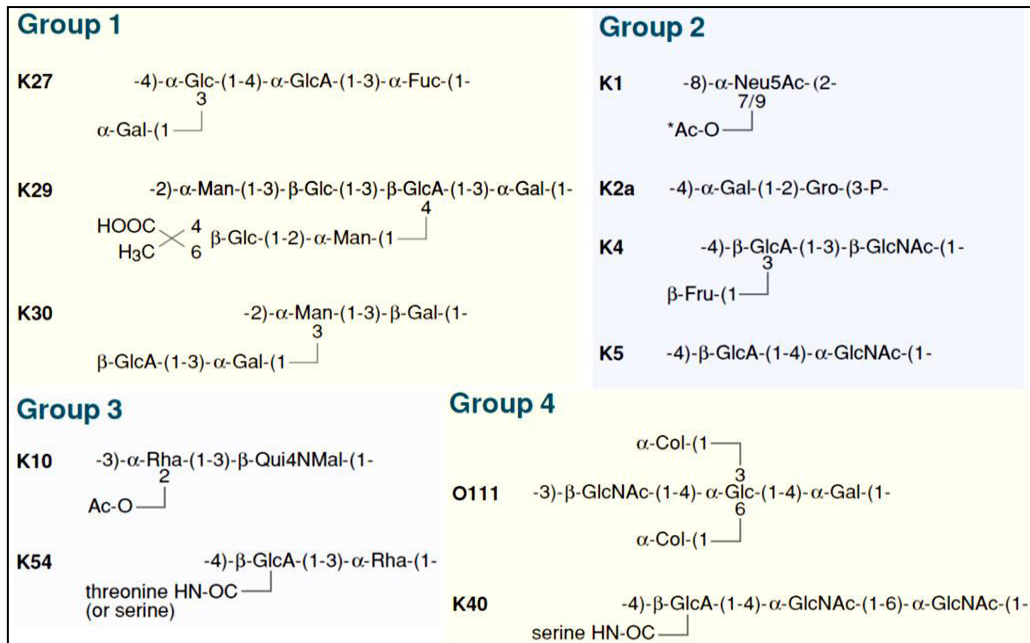
Since *E. coli* is the model system for our present study, the classification of its CPS is discussed here. As mentioned previously, the earliest serological classification of *E. coli* was by Kauffman in the 1940’s. This classification divided the K antigen into three groups (L, A & B) based on thermostability in serotyping analyses. Groups L and B are thermolabile, ‘envelope’ antigens while group A consisted of thermostable ‘capsular antigen’. A second important classification was put forward in 1997, where K antigen was divided into Group 1 (with subdivisions 1a and 1b) and Group 2, mainly based on serological characteristics, along with a few biochemical and genetic criteria.

The most recent classification scheme takes into account several biochemical and genetic criteria (as shown in Table 1.3) and divided *E. coli* into 4 groups. This differs from older schemes as it does not rely on serological assignments or polysaccharide structure as predictive criteria (Whitfield, 2006).

Table 1.3. Latest classification of *E. coli* capsules (Whitfield and Roberts, 1999).

Characteristics	Group			
	1	2	3	4
Thermostability	Yes	No	No	Yes
Polymerization system	Wzy dependent	ABC transporter dependent	ABC transporter dependent	Wzy dependent
Genetic locus	<i>cps</i> near <i>his</i>	<i>kps</i> near <i>ser A</i>	<i>kps</i> near <i>ser A</i>	Near <i>his</i>
Coexpression with O serogroups	Limited (O8,O9,O20,O101)	Many	Many	Often O8,O9; occasionally none
Coexpression with colanic acid	No	Yes	Yes	Yes
Terminal lipid moiety	Lipid A core in K <sub>LPS</sub> ; Unknown for capsular K antigen	α-Glycerophosphate	α-Glycerophosphate	Lipid A core in K <sub>LPS</sub> ; Unknown for capsular K antigen
Serotype of model system	K30	K1, K5	K10, K54	K40, O111
Similarity to capsules of other organisms	<i>Klebsiella</i> , <i>Erwinia</i>	<i>Neisseria</i> , <i>Haemophilus</i>	<i>Neisseria</i> , <i>Haemophilus</i>	Many genera

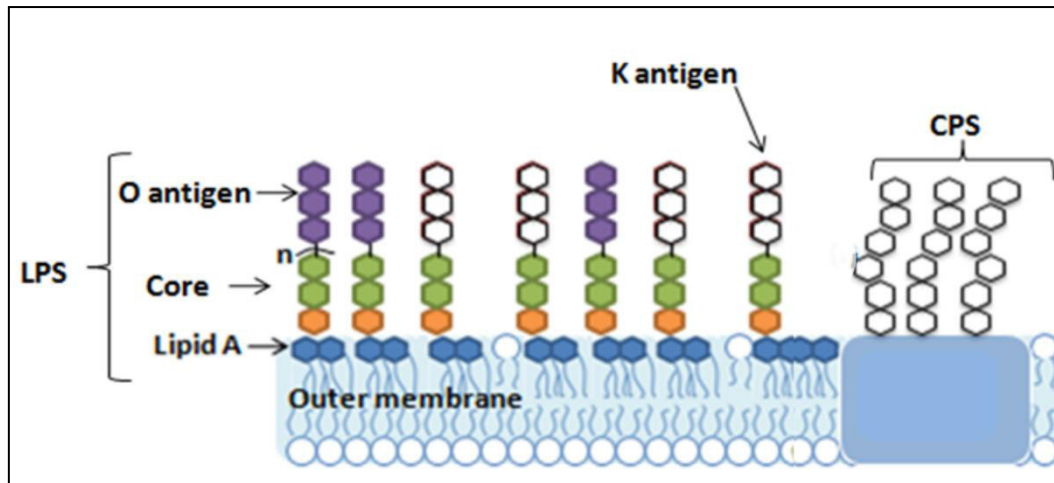
Among the four groups mentioned above, *E. coli* group 1 and 4 capsules possess several similarities. Both are thermostable, are synthesized by a Wzy dependent pathway (discussed below) and have a similar genetic locus. Likewise, group 2 and 3 capsules possess similarities, both being heat sensitive, synthesized by an ATP binding cassette (ABC) transporter dependent pathway and with a similar genetic locus. The structures of the CPS repeating units of *E. Coli* belonging to each of the four groups is depicted in Figure 1.5.



**Figure 1.5 Representative chemical structures corresponding to K antigen repeat units of the four groups of *E. coli* capsules** (Adapted from Whitfield, 2006). K27, K29 and K30 from Group 1, K1, K2a, K4 and K5 from Group 2, K10 and K54 from Group 3 and K40 and O111 from Group 4 are shown. Note that in Group 4, isolates producing CPS as the only serotype specific polysaccharide are given O antigen status (such as O111 shown in figure).

### 1.3 Group 1 capsules

Group 1 K antigens are expressed on the bacterial surface in two forms: one with shorter oligosaccharides that linked to the lipid A core (similar to LPS structures), known as  $K_{LPS}$  and the other with longer polysaccharides, forming the high molecular weight structure. It is noteworthy that the repeating unit is same for both the former and latter. LPS,  $K_{LPS}$  and CPS structures attached to the Gram-negative bacteria outer membrane have been represented in Figure 1.6.



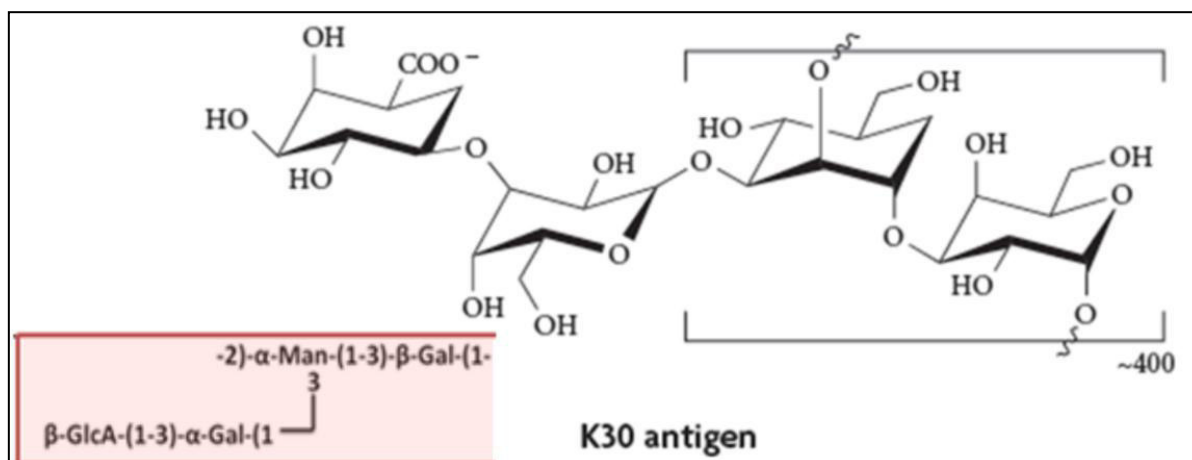
**Figure 1.6 Surface associated polysaccharides in a Gram-negative bacterium.** LPS that consists of O antigen (purple) repeating unit covalently linked to a lipid A–core unit (Lipid A is shown in blue and the outer and inner core units are depicted in green and orange respectively) in the outer membrane,  $K_{LPS}$  that consists of a few repeats of the K30 antigen (white) linked to a lipid A core and K30 CPS (white) that is non-covalently attached to the surface are shown.

*E. coli* and *Klebsiella* strains that use Group 1 antigen are responsible for several diseases (Table 1.2) and the presence of these capsules is found to be essential for virulence in the strains producing them (Rahn and Whitfield, 2003). The chemical structure of K30 (*E. coli*) is given in Figure 1.7. The method of group 1 capsule biogenesis, namely, the Wzy dependent pathway is widespread among both Gram-positive and Gram-negative bacteria. In addition, the outer membrane proteins involved in Group 1 capsule biogenesis have homologs in a very wide range of bacteria, some of which are highly virulent. Group 1 and 4 bacteria are similar to each other in various aspects, as shown in table 1.3. These groups are found mainly in the *E. coli* strains belonging to the EPEC, ETEC and EHEC categories, which as explained earlier, are known to cause intestinal infections. Group1 K30 capsule, which is a well characterized antigen, is prevalently found in UPEC *E. coli* strains (Goller and Seed, 2010) and is the primary focus of this study. The other K antigen groups are not discussed, as it is beyond the scope of this study.

### 1.3.1 Structure of K30 CPS

Group 1 capsules consist of several different capsular phenotypes such as K27, K29 and K30. They are acidic polysaccharides and typically contain uronic acids. All group 1 capsules tend to be similar in structure. In the present investigation, the model system for group 1 capsular

polysaccharides is the K30 serotype of *E. coli*. The repeating unit of the K30 antigen is shown in Figure 1.7 (Chakraborty et al., 1980).



**Figure 1.7 Chemical diagram of K30 CPS monomeric unit.** K30 monomer consists of an  $\rightarrow 2)$ - $\alpha$ -D-Man-(1 $\rightarrow$ 3)- $\beta$ -D-Gal-(1 $\rightarrow$ 3)- $\beta$ -D-GlcUA-(1 $\rightarrow$ 3)- $\alpha$ -D-Gal-(1 $\rightarrow$ 3) chain carrying a  $\beta$ -D-GlcUA-(1 $\rightarrow$ 3)- $\alpha$ -D-Gal-(1 $\rightarrow$ 3) branch at position 3 of the mannose. Inset, schematic diagram of the monomeric unit.

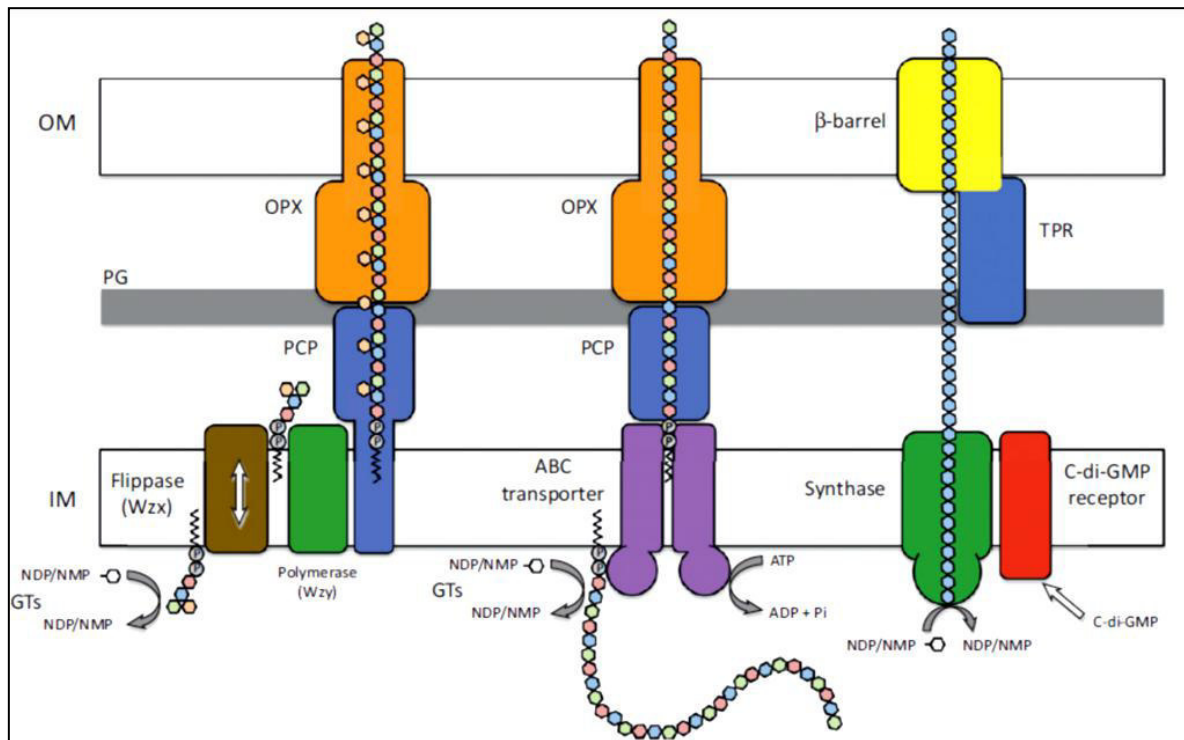
The mechanism of K30 CPS biosynthesis and the proteins which constitute the supramolecular assembly for capsule biosynthesis and surface expression are discussed in the following sections. Note that the terms Group 1 CPS and K30 CPS will be used synonymously henceforth.

### 1.3.2 Method of K30 capsule biosynthesis

Even though the number of capsular phenotypes found in bacteria is extremely large, capsular biosynthetic pathways are limited to the Wzy dependent pathway, ABC transporter dependent pathway and in some cases, a synthase dependent pathway (summarized in Figure 1.8). The Wzy dependent pathway involves formation of a carrier lipid linked CPS repeat unit in the periplasm, followed by its block-wise polymerization in the inner leaflet of the inner membrane and finally, the transport across the outer membrane (Figure 1.8 (left)). In the ABC transporter dependent pathway, a full length CPS polymer is first synthesized in the cytoplasm, followed by its ATP-binding cassette (ABC) dependent transport across the inner membrane and concomitant transport across the outer membrane (Figure 1.8 (middle)). In the synthase dependent pathway, a synthase protein is responsible for initiation of capsule polymerization and its export across the inner membrane. The polymer is then transported across the outer membrane (Figure 1.8 (right)) (Whitney and Howell, 2013).



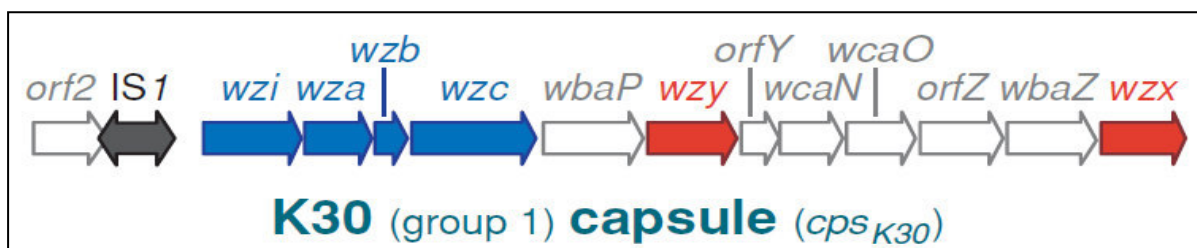
As K30 CPS uses the Wzy dependent pathway for its synthesis, export and surface expression, it is discussed in detail in the following section.



**Figure 1.8 Methods of capsule biogenesis** (adopted from Whitney and Howell, 2013): (Left) Wzy dependent-, (middle) ABC transporter dependent and (right) synthase dependent pathways are depicted. In the **Wzy dependent pathway (left)** (explained in detail in the next section), the monomeric units attach to a lipid carrier in the inner membrane and are flipped across the membrane to the inside, where they are polymerized. Transport to the exterior is regulated by a combination of PCP and OPX proteins. In the **ABC transporter dependent system (middle)**, the pre-polymerized CPS is transported across the inner membrane by an ABC transporter facilitates in a process energized by ATPase, while OPX and PCP proteins transport it across the outer membrane. In the **synthase-dependent pathway (right)**, polysaccharide polymerization and transport across the inner membrane is carried out by a synthase protein that is post-translationally regulated by a C-di-GMP receptor. Transport across the outer membrane is facilitated by a periplasmic TPR containing protein and an integral OM beta barrel protein. The repeating sugar units of CPS are depicted as white circles. Note that the following abbreviations are used: IM-Inner membrane, OM-Outer membrane, PG-peptidoglycan, OPX-Outer membrane polysaccharide exporting proteins, PCP-Polysaccharide Copolymerase, TPR-tetratricopeptide, GT-Glycosyltransferase, ABC-ATP binding cassette.

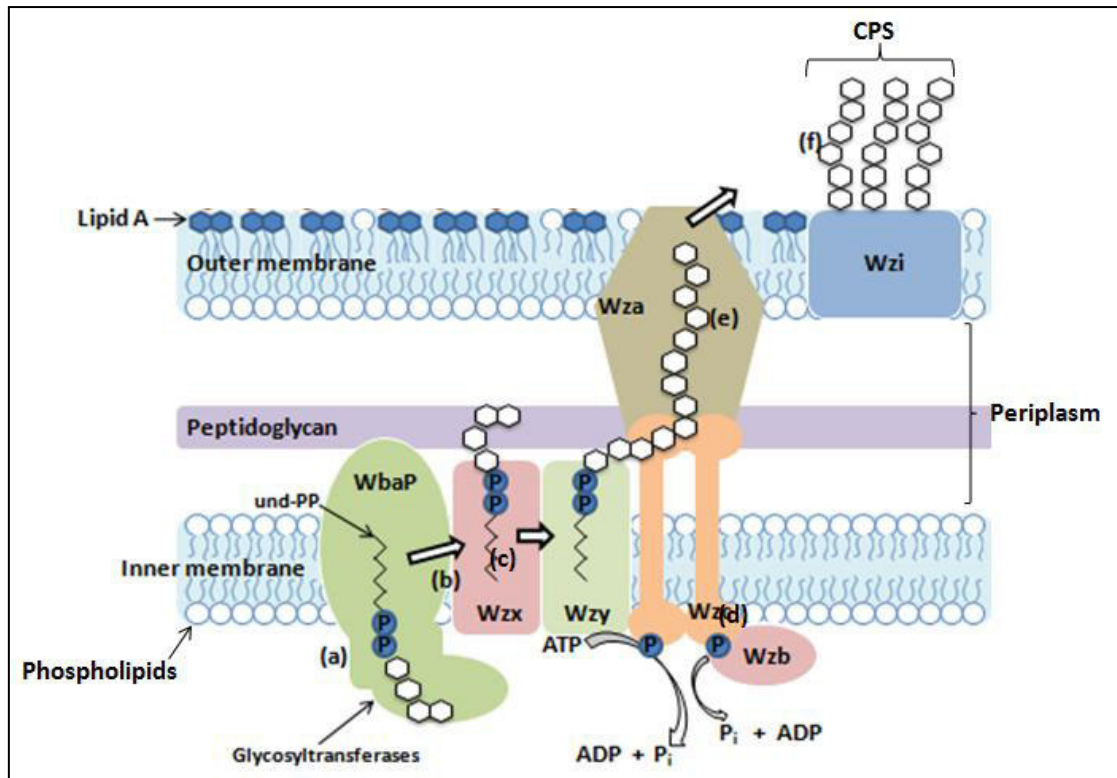
### 1.3.2.1 Wzy dependent supramolecular assembly

The Wzy dependent supramolecular assembly comprises of several proteins, ranging from cytoplasmic to inner and outer membrane proteins. The genes require for K30 CPS assembly are located near *his* (the histidine biosynthetic cluster) and the locus for expression of LPS(Drummel-Smith and Whitfield, 1999). The 16kb locus (Figure 1.9) consists of 12 open reading frames and is divided into two regions by a stem-loop transcriptional attenuator. The 5' end of the locus consists of a highly conserved block of four genes, *wzi*, *wza*, *wzb* and *wzc*. Three of the conserved gene products (Wza, Wzb and Wzc) are involved in the polymerization and translocation of CPS and function independent of capsule type. The 3' region of the locus is serotype specific and codes for enzymes producing sugar precursors for capsule synthesis, glycosyltransferases, and integral membrane proteins Wzx and Wzy. It is the presence of the conserved genes on the 5' end that distinguish this locus from that for O antigens.



**Figure 1.9 Genetic locus for Group 1 capsule** The characteristic *wxz* and *wzy* genes are highlighted in red, and known genes involved in regulation of high-level polymerization and translocation and surface expression, *wzi*, *wza*, *wzb* and *wzc* are in blue. *wbaP*, *wcaN*, *wcaO* and *wbaZ* gene products are periplasmic glycosyltransferases (Adopted from Whitfield, 2006).

The Wzy dependent pathway is explained in detail in Figure 1.10. The biosynthesis of CPS starts in the periplasmic region, catalyzed by several glycosyltransferases, including WbaP. A single monomer of K30 CPS, linked to a lipid carrier, is flipped across the inner membrane by Wzx where it undergoes polymerization by the combined action of Wzy and the linked Wzc autokinase-Wzb phosphatase proteins. Translocation to the exterior is aided by a Wza translocase and surface expression is achieved by interaction with the outer membrane lectin, Wzi.

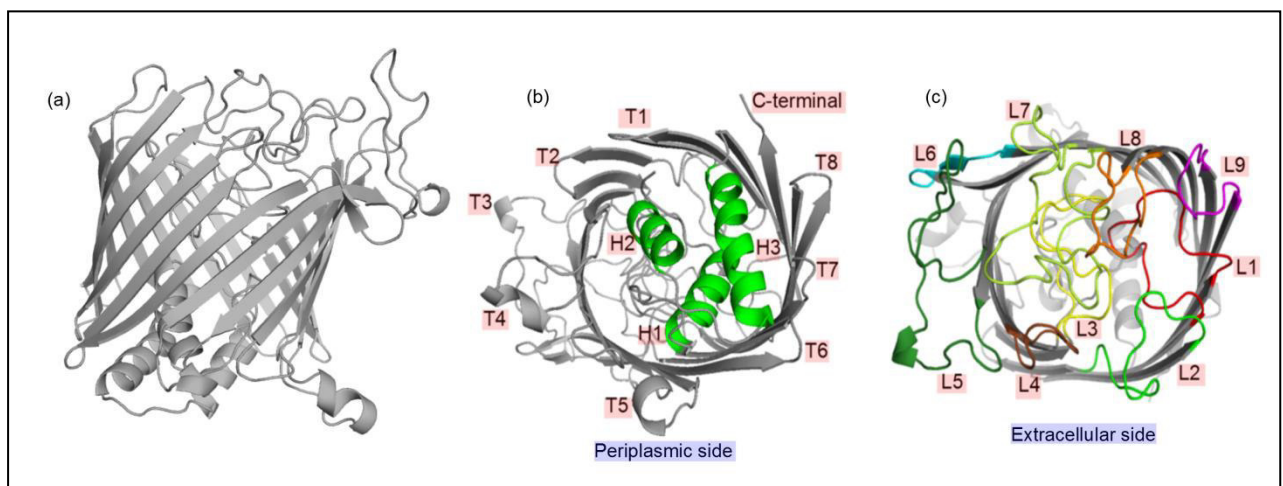


**Figure 1.10 Wzy dependent pathway for Group 1 K30 capsule biosynthesis and surface expression.** (a) The pathway is initiated by the transfer of galactose-1-Phosphate (Gal-1-P) from a UDP-Galactose (Uridine Diphosphate Galactose) unit to the Und-pp (Undecaprenyl diphosphate) lipid carrier by a glycosyltransferase of the inner membrane, WbaP. Other glycosyltransferases add additional glycoses, namely mannose and glucuronic acid, to the Und-pp linked intermediate. (b) The Und-PP linked intermediate is flipped across the inner membrane by a flippase protein, Wzx. (c) Polymerization takes place in the periplasmic region, facilitated by the Wzy polymerase. (d) A polysaccharide copolymerase, Wzc, also aids in polymerization and plays a role in length control of the polymer. It is a tyrosine autokinase and its activity depends on the degree of autophosphorylation. Wzb, a tyrosine phosphatase, is the cytoplasmic associate of Wzc, facilitating Wzc dephosphorylation. Repeated phosphorylation and dephosphorylation cycles are required for capsule production to take place. (e) Wzc associates with the outer membrane translocon, Wza, to form an outer membrane channel to export K30 CPS to the cell exterior. (f) K30 CPS (white hexagons) associates with an outer membrane lectin, Wzi, for surface expression (Collins et al., 2006; Cuthbertson et al., 2009; Nesper et al., 2003; Silhavy et al., 2010; Whitfield and Larue, 2008).

In the present study, we focus on one of the outer membrane protein involved in the Wzy dependent polymerization of K30 CPS, namely Wzi. Wzi is involved in surface expression of K30 CPS and is thus an important therapeutic target.

## 1.4 Structure of Wzi

The crystal structure of Wzi was first solved at 2.6Å<sup>o</sup> resolution in 2013. Wza is a monomeric, outer membrane protein with a molecular weight of 50 KDa. It forms an 18 stranded anti-parallel beta barrel and has a circular cross section with a diameter of 36 Å (Figure 1.11(a)). The N and C termini of the protein are located in the periplasmic region. The extracellular region of Wzi consists of nine loops (L1-L9) connecting the adjacent antiparallel beta strands while the periplasmic region consists of 8 turns (T1-T8), joining the ends of the beta strands. In the periplasmic region, there is an alpha helical bundle that consists of three helices (H1, H2 and H3) which occludes the pore from the periplasmic end (Figure 1.11(b)). Some of the extracellular loops such as L3, L7, L8 & L1 fold into the barrel to completely occlude the pore on the extracellular side, while other loops like L5, L6 splay outwards (Figure 1.11(c)). Strand 6 of the barrel is shorter than the other strands, forming a triangular notch like region at the extracellular phase of the barrel. Though several beta-barrel structures are known, Wzi protein is unique in that it has a circular cross section unlike the elliptical cross section found in other proteins such as porins.



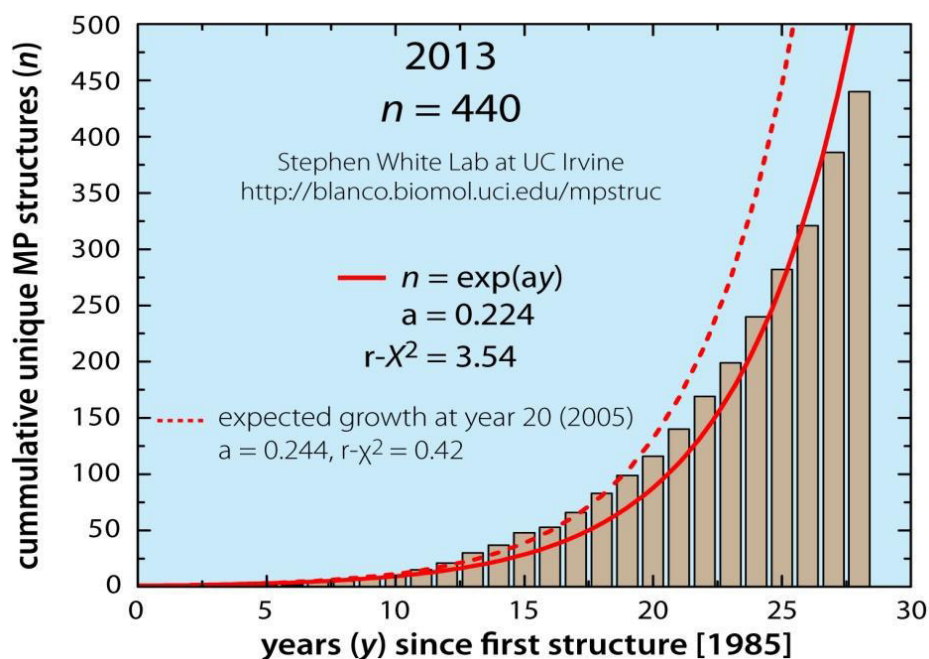
**Figure 1.11** Cartoon representation of Wzi [PDBID: 2YNK] (S R Bushell et al., 2013). (a)View perpendicular to the axis of the beta barrel, (b) View from the periplasmic side with helical bundle (green) and turns (labelled T1 to T8) shown, (c) View from the extracellular side. Extracellular loops (labelled L1 to L9) are shown.

Wzi homologs are found only in those Gram-negative bacteria which have the group 1 type capsule structure; it is absent in EPS secretion pathways such as colanic acid (from *E. coli*). Homologs of the protein can be found in several species, including *Klebsiella pneumoniae* (which is also classified under group 1 Gram-negative bacteria; refer table 1.3 ), *Providentia spp.*, *Serratia spp.*, *Shewanella spp.*, *Rhodobacter spp.*, *Psychrobacter spp.* and also in the highly virulent *Acinetobacter baumannii* (S R Bushell et al., 2013). The presence of Wzi homologs across such a wide range of species highlights the importance of studying its dynamics and its role in the surface expression of CPS and the resulting bacterial pathogenicity.

## **1.5 Challenges in the study of membrane proteins**

The major obstacles in the study of membrane proteins is the lack of availability of crystal and NMR structures, owing to the technical problems at all levels, including expression, solubilisation, purification, crystallization, solubilisation, data collection and finally structure solution. Further, experimental structure determination methods would require the protein to be in its natural lipid bilayer environment. This poses problems in sample preparation for X-ray crystallography and solution NMR. Another major problem is that these proteins are insoluble in aqueous solutions and detergents have to be used to solubilise them. If the concentration increases beyond a point, it may even denature the protein. Thus, the rate at which membrane protein structures have been solved is extremely slow. Even though membrane proteins constitute around 20-30% of the proteome, the number of solved, non-redundant structures available constitutes just about 2% of the structures deposited in databases (White, 2004).

Figure 1.12 depicts the cumulative number of membrane protein structures reported each year since the first membrane protein structure discovery in 1985. The data is from a curated database of membrane proteins of known structure called mpstruc, which is part of the Structural Biology Knowledge Base (SBKB).



**Figure 1.12 Rate of discovery of crystal structures of membrane proteins** (Adopted from [http://blanco.biomol.uci.edu/Membrane\\_Proteins\\_xtal.html](http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html)). The bars represent the cumulative number ( $n$ ) of structures plotted against the number of years ( $y$ ) since the first structure was reported. The solid curve is the best fit to the equation  $n = \exp(ay)$ , where  $a = 0.23$ ; the reduced  $\chi^2$  of the fit is 0.6. (Only unique structures are included in the statistics. Proteins of the same type from different species are included, but structures of mutagenized versions of proteins are excluded, as are proteins that differ only in terms of substrate bound or physiological state).

Fortunately, the crystal structure of Wzi, the protein of our interest, was determined recently. The crystal structure, however, is static and does not offer any information on the conformational dynamics of the protein. Further, there might be significant crystal packing effects that affect the protein conformation. As mentioned previously, studying the dynamics of these proteins using NMR also poses significant challenges, since a membrane environment has to be provided. MD simulation is thus an appropriate tool to explore the conformational dynamics of membrane proteins.

Thus in the present investigation, the crystal structure of Wzi [PDBID: 2YNK.pdb] of *E. coli* is considered as an initial model for molecular dynamics simulations in an explicit solvent-membrane environment to understand its conformational dynamics.

## 1.6 Scope of the study

Capsular polysaccharides are major virulence determinants of bacteria, helping in colonization and providing resistance against the host immune system, but little is known at present about how it translocates to the exterior of the bacteria or associates with the outer membrane. In this context, we study the outer membrane lectin Wzi, using the Molecular Dynamics (MD) simulation technique in a lipid bilayer considering group 1 K30 *E. coli* as a model system. We aim to understand various properties including the dynamics, electrostatics, ionic affinity, channel conduction properties of Wzi and its interaction with the lipid bilayer. Simulations of six substitution mutants of Wzi will be carried out in addition to the wild type protein. Three of the mutants are helix H1 substitution mutants, two are point substitution mutants in the extracellular loop region and the sixth one is a deletion mutant of Wzi with 15 residues deleted from extracellular loop 5. Further, wild type simulations will be carried out at a range of transmembrane voltages to mimic the natural system more closely and improve the sampling of ion crossing events, if any. In addition, a hexameric repeating unit of K30 CPS,  $\rightarrow 2$ - $\alpha$ -D-Manp-(1 $\rightarrow$ 3)- $\beta$ -D-Galp-(1 $\rightarrow$  chain carrying a  $\beta$ -D-GlcUAp-(1 $\rightarrow$ 3)- $\alpha$ -D-Galp-(1 $\rightarrow$  branch at position 3 of the mannose, will also be simulated to understand its structural preference. Homologs of Wzi of K30 *E. coli* are found in several highly virulent strains such as *Klebsiella* and *Acinetobacter*, thus warranting the need to understand the conformational dynamics of Wzi and K30 CPS. We expect that the simulation results will provide a better understanding of the attachment of CPS to the bacterial surface. This would eventually facilitate the design of “Wzi blockers” that perturb the attachment of CPS to the bacterial surface and concomitantly decrease bacterial virulence and increase antibiotic susceptibility. The results obtained from this study would be a step towards realizing the goal of preventing bacterial infectious diseases and combating the emerging MDR strains.

## 1.7 References

- Bassetti, M., Ginocchio, F., Mikulska, M., 2011. New treatment options against gram-negative organisms. *Crit. Care Lond. Engl.* 15, 215. doi:10.1186/cc9997
- Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, J.E., Gilbert, D., Rice, L.B., Scheld, M., Spellberg, B., Bartlett, J., 2009. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 48, 1–12. doi:10.1086/595011
- Bushell, S.R., Mainprize, I.L., Wear, M.A., Lou, H., Whitfield, C., Naismith, J.H., 2013. Wzi is an outer membrane lectin that underpins group 1 capsule assembly in *Escherichia coli*. *Structure* 21, 844–853. doi:10.1016/j.str.2013.03.010
- Campos, M.A., Vargas, M.A., Regueiro, V., Llompert, C.M., Albertí, S., Bengoechea, J.A., 2004. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun.* 72, 7107–7114. doi:10.1128/IAI.72.12.7107-7114.2004
- Chakraborty, A.K., Friebolin, H., Stirm, S., 1980. Primary structure of the *Escherichia coli* serotype K30 capsular polysaccharide. *J. Bacteriol.* 141, 971–972.
- Collins, R.F., Beis, K., Clarke, B.R., Ford, R.C., Hulley, M., Naismith, J.H., Whitfield, C., 2006. Periplasmic protein-protein contacts in the inner membrane protein Wzc form a tetrameric complex required for the assembly of *Escherichia coli* group 1 capsules. *J. Biol. Chem.* 281, 2144–2150. doi:10.1074/jbc.M508078200
- Cuthbertson, L., Mainprize, I.L., Naismith, J.H., Whitfield, C., 2009. Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. *Microbiol. Mol. Biol. Rev. MMBR* 73, 155–177. doi:10.1128/MMBR.00024-08
- Drummelsmith, J., Whitfield, C., 1999. Gene products required for surface expression of the capsular form of the group 1 K antigen in *Escherichia coli* (O9a:K30). *Mol. Microbiol.* 31, 1321–1332.
- Drummelsmith, J., Whitfield, C., 2000. Translocation of group 1 capsular polysaccharide to the surface of *Escherichia coli* requires a multimeric complex in the outer membrane. *EMBO J.* 19, 57–66. doi:10.1093/emboj/19.1.57
- Goller, C.C., Seed, P.C., 2010. Revisiting the *Escherichia coli* polysaccharide capsule as a virulence factor during urinary tract infection: contribution to intracellular biofilm development. *Virulence* 1, 333–337. doi:10.4161/viru.1.4.12388
- Iman Naimi et al., n.d. The Role of wza in Extracellular Capsular Polysaccharide Levels During Exposure to Sublethal Doses of Streptomycin.
- Kanj, S.S., Kanafani, Z.A., 2011. Current concepts in antimicrobial therapy against resistant gram-negative organisms: extended-spectrum beta-lactamase-producing Enterobacteriaceae, carbapenem-resistant Enterobacteriaceae, and multidrug-resistant *Pseudomonas aeruginosa*. *Mayo Clin. Proc. Mayo Clin.* 86, 250–259. doi:10.4065/mcp.2010.0674
- Kaper, J.B., Nataro, J.P., Mobley, H.L., 2004. Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 2, 123–140. doi:10.1038/nrmicro818
- Kollef, M.H., Golan, Y., Micek, S.T., Shorr, A.F., Restrepo, M.I., 2011. Appraising contemporary strategies to combat multidrug resistant gram-negative bacterial infections--proceedings and data from the Gram-Negative Resistance Summit. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 53 Suppl 2, S33–55; quiz S56–58. doi:10.1093/cid/cir475
- Kumarasamy, K.K., Toleman, M.A., Walsh, T.R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C.G., Irfan, S., Krishnan, P., Kumar, A.V., Maharjan, S., Mushtaq, S., Noorie, T., Paterson, D.L., Pearson, A., Perry, C., Pike, R.,



- Rao, B., Ray, U., Sarma, J.B., Sharma, M., Sheridan, E., Thirunarayan, M.A., Turton, J., Upadhyay, S., Warner, M., Welfare, W., Livermore, D.M., Woodford, N., 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *Lancet Infect. Dis.* 10, 597–602. doi:10.1016/S1473-3099(10)70143-2
- Nesper, J., Hill, C.M.D., Paiment, A., Harauz, G., Beis, K., Naismith, J.H., Whitfield, C., 2003. Translocation of group 1 capsular polysaccharide in *Escherichia coli* serotype K30. Structural and functional analysis of the outer membrane lipoprotein Wza. *J. Biol. Chem.* 278, 49763–49772. doi:10.1074/jbc.M308775200
- Nikaido, H., 1998. Antibiotic resistance caused by gram-negative multidrug efflux pumps. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 27 Suppl 1, S32–41.
- Nwodo, U.U., Green, E., Okoh, A.I., 2012. Bacterial exopolysaccharides: functionality and prospects. *Int. J. Mol. Sci.* 13, 14002–14015. doi:10.3390/ijms131114002
- Peleg, A.Y., Hooper, D.C., 2010. Hospital-acquired infections due to gram-negative bacteria. *N. Engl. J. Med.* 362, 1804–1813. doi:10.1056/NEJMra0904124
- Rahn, A., Whitfield, C., 2003. Transcriptional organization and regulation of the *Escherichia coli* K30 group 1 capsule biosynthesis (cps) gene cluster. *Mol. Microbiol.* 47, 1045–1060.
- Rehm, B.H.A., 2010. Bacterial polymers: biosynthesis, modifications and applications. *Nat. Rev. Microbiol.* 8, 578–592. doi:10.1038/nrmicro2354
- Roberts, I.S., 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu. Rev. Microbiol.* 50, 285–315. doi:10.1146/annurev.micro.50.1.285
- Silhavy, T.J., Kahne, D., Walker, S., 2010. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2, a000414. doi:10.1101/cshperspect.a000414
- White, S.H., 2004. The progress of membrane protein structure determination. *Protein Sci.* 13, 1948–9. doi:10.1110/ps.04712004
- Whitfield, C., 2006. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* 75, 39–68. doi:10.1146/annurev.biochem.75.103004.142545
- Whitfield, C., Larue, K., 2008. Stop and go: regulation of chain length in the biosynthesis of bacterial polysaccharides. *Nat. Struct. Mol. Biol.* 15, 121–123. doi:10.1038/nsmb0208-121
- Whitney, J.C., Howell, P.L., 2013. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. *Trends Microbiol.* 21, 63–72. doi:10.1016/j.tim.2012.10.001

# **CHAPTER 2**

## **Methods**

## Chapter 2 : Methods

NMR and X-ray crystallography are the two most important experimental techniques that provide atomistic details about molecules. However, x-ray crystallography provides only limited information about the molecule. It provides only “snapshots” of the structure and does not offer time-dependent conformational dynamics for the molecule. Further, there might be significant crystal packing effects on the conformation of the molecule. Though NMR is capable of providing the dynamic behaviour of a molecule, it is highly limited to molecules of smaller size and poses significant challenges when a membrane environment has to be provided. Handling highly flexible molecules like oligosaccharides by x-ray and NMR is also highly challenging. In this context, group 1 outer membrane protein Wzi as well as K30 CPS are studied by the molecular dynamics simulation technique.

### 2.1 Historical perspective on classical MD simulations

The first MD simulations were carried out by B J Alder and T E Wainbright in 1957 to study the interactions of hard spheres. This was followed by the first simulation using a realistic potential for liquid Argon in 1964 by Rahman and the first simulation of a realistic system in 1974 by Rahman and Stillinger where they simulated liquid water. The first protein simulation was carried out in 1977 by McCammon J A and Martin Karplus , where they simulated the bovine pancreatic trypsin inhibitor (BPTI) (Adcock and McCammon, 2006; Durrant and McCammon, 2011; Karplus and McCammon, 2002). This was the first time a macromolecule of biological interest was simulated. The BPTI system consisted of 458 atoms and ran for a total of 9.2 ps. The first nanosecond simulation was reported much later , around 1998. With the tremendous increase in computational resource and theoretical and methodological developments, MD simulations can now be run for much larger systems (sizes >100000 atoms) for microsecond timescales and beyond, allowing us to study macromolecules at biologically relevant timescales.

### 2.2 Force fields

In MD simulations, the interactions of particles in a system are defined by a potential energy function. The potential energy function gives an idea of the force acting on each particle in the system and subsequently, these forces are used to determine the dynamics of the system

using classical equations of motion. This potential energy function along with a set of empirical parameters forms the ‘force field’.

One fundamental assumption in the force-field approach is the ‘Born-Oppenheimer approximation’, which separates the nuclear and electronic motions of particles and allows the energy of a system to be expressed as a function of nuclear coordinates alone. Force-fields are also assumed to be ‘additive’ and ‘transferable’. Additivity implies that the net energy of the system at any point is the sum of the individual energy contributions (from bonded and nonbonded interactions). Transferability on the other hand implies that the Hamiltonian (Potential energy function) developed based on a small set of particles can be extended to calculate the energy for a larger system, provided they possess the same chemical groups (Monticelli and Tieleman, 2013).

The earliest force-fields were developed in the 1960’s, such as the MM (Molecular Mechanics) potentials developed by Allinger *et.al.* for organic molecules and the consistent force field (CFF) developments by Lifson *et.al.* for proteins and nucleic acids. These force-fields were initially used in case of small molecules and later extended to deal with larger and more complex systems (Levitt, 2001; Ponder and Case, 2003). Some of the major force-fields that are in use today for biomolecular simulations are CHARMM, AMBER, OPLS and GROMOS. Each force-field has its own strengths and weaknesses and the choice of a specific force field depends on the system being studied. All simulations in this study are performed using CHARMM forcefield (version C36b2).

### **2.3 CHARMM (Chemistry at HARvard using Molecular Mechanics)**

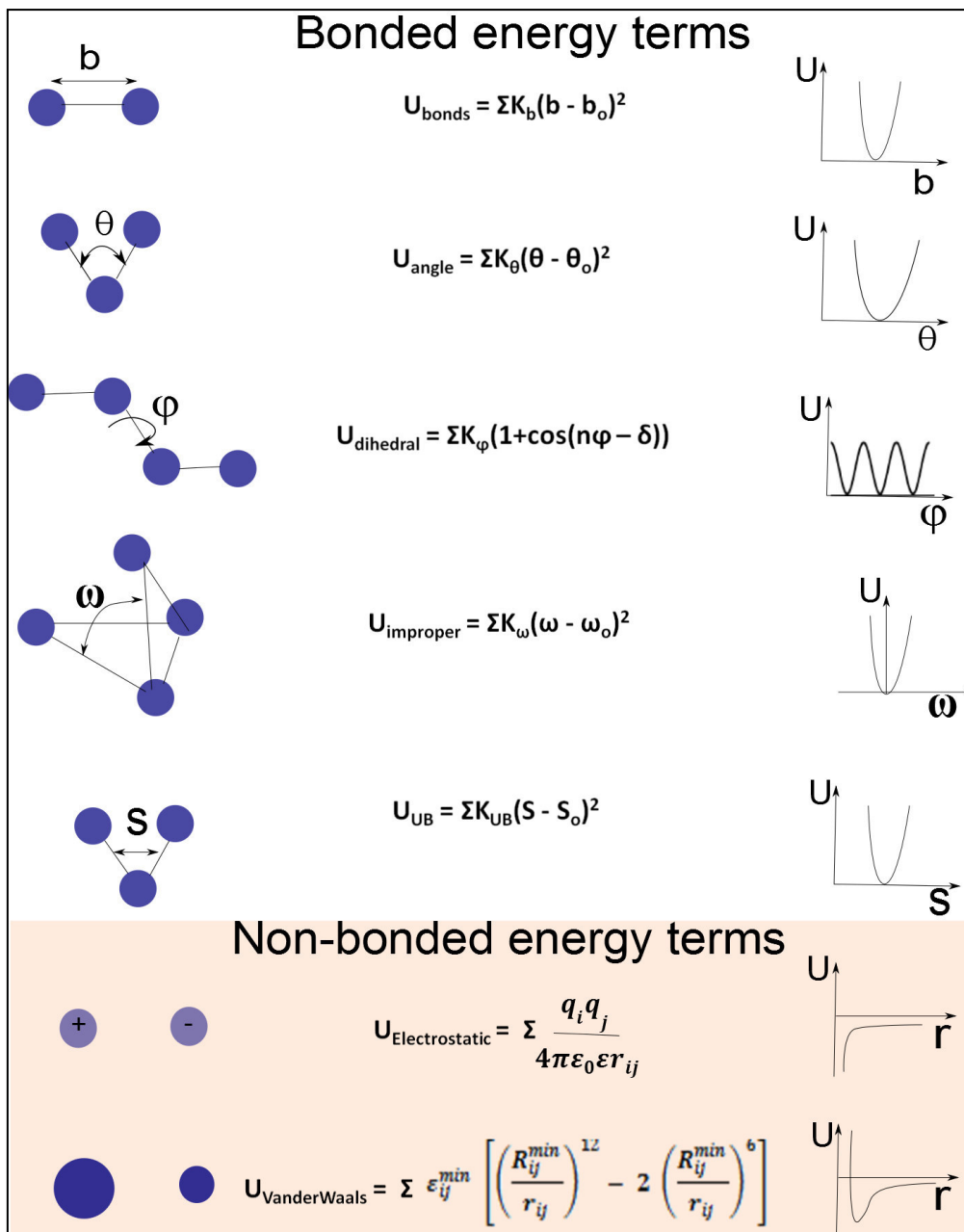
CHARMM was developed in the early 1980’s and was first described in the seminal paper by Karplus and co-workers in 1983. It is one of the most successful force-fields in the simulations of biomolecules and has been used in many landmark MD simulations. For instance, the first MD simulation to cross the one million atoms barrier was carried out by Schulten *et.al.* in 2006 and the first all atom computational model of a life form (using the satellite tobacco mosaic virus) by the same group in 2008 were both carried out using the CHARMM force field (Zhu et al., 2012).

### 2.3.1 Potential energy function of CHARMM

The potential energy function of CHARMM force field,  $U(\vec{R})$ , can be represented as:

$$\begin{aligned}
 U(\vec{R}) = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi (1 + \cos(n\phi - \delta)) \\
 & + \sum_{impropers} K_\omega (\omega - \omega_0)^2 + \sum_{Urey-Bradley} K_{UB} (S - S_0)^2 \\
 & + \sum_{residues} U_{CMAP}(\phi, \psi) \\
 & + \sum_{non-bonded\ pairs} \left\{ \epsilon_{ij}^{min} \left[ \left( \frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \right\} \dots\dots\dots(i)
 \end{aligned}$$

Here,  $U(\vec{R})$  is the potential energy of the given system. The first term in equation (i) accounts for changes in bond length ( $b$ ) from the equilibrium value ( $b_0$ ), where  $K_b$  is the bond force constant. The second term represents changes in the bond angle ( $\theta$ ), from the equilibrium value ( $\theta_0$ ), where  $K_\theta$  is the angle force constant. The third term is for changes in dihedral angles or torsion angles ( $\phi$ ), where  $n$  represents periodicity of the dihedral angle,  $\delta$  represents the phase shift and  $K_\phi$  is the corresponding force constant. The fourth term, impropers, accounts for out-of-plane bending, where  $K_\omega$  is the force constant,  $\omega$  is the improper angle and  $\omega_0$  its equilibrium value. The fifth term (Urey-Bradley), is a cross term that accounts for angle bending using 1,3 non-bonded interactions. It is a quadratic function of distance ‘S’ between the first and third atoms, in a series of three bonded atoms.  $S_0$  represents equilibrium value of the distance and  $K_{UB}$  is the respective force constant. The sixth term (CMAP) is a dihedral energy correction term for the protein backbone, where  $\phi$ ,  $\psi$  are the protein backbone dihedral angles. The last two terms represent the non-bonded interactions between a pair of atoms  $i$  and  $j$  and include the Lennard Jones (LJ) 6-12 term (for core-core repulsion and attractive van der Waals dispersion interaction) as well as the coulombic interactions between the atoms of charges  $q_i$  and  $q_j$  respectively.  $\epsilon_{ij}^{min}$  represents well depth,  $r_{ij}$  is the interatomic distance and  $R_{ij}^{min}$  is the distance at which the LJ term has its minimum. The relative dielectric constant,  $\epsilon$ , is set to 1 for calculations with explicit solvent models corresponding to the permittivity of vacuum,  $\epsilon_0$  (Brooks et al., 2009).



**Figure 2.1** Illustration of the bonded and non-bonded energy terms in CHARMM potential energy function (equation (i)). A schematic representation of the bonded and non-bonded energy terms is shown in the first column with the corresponding equations in the second column. The third column illustrates the energy terms graphically. All terms have been explained in text.

### 2.3.2 Computation of non-bonded interactions :

Non-bonded interactions (van-der Waals and electrostatic interactions) are defined on atom pairs and require number of atoms squared calculations (i.e., N number of atoms would require N x N calculations). Unlike bonded interactions which involve only the next

neighbours, non-bonded interactions constitute a computationally demanding part of energy calculations. One way to overcome this problem and reduce computational cost is by introducing a cutoff distance beyond which interactions are ignored. This method is effective in case of computation of LJ forces, as these forces decay rapidly over large distances. However, electrostatic forces decay much slower (proportional to  $1/r$ ). Hence, applying a cutoff distance to electrostatic calculations would be highly inaccurate. A different method, called particle mesh ewald method is used to calculate electrostatic instead. A detailed description of both these methods is provided in the following paragraphs.

### 2.3.2.1 LJ interactions

To simplify calculations, a nonbonded atoms list is generated first, followed by actual energy calculations. However, generating the non-bonded list is still a time consuming process and to simplify it switching/shifting functions are used, where a cutoff value (CTOFnb) is defined, beyond which interactions and forces are not considered. Within CTOFnb, a smaller radius (CTONnb) is defined, within which the interactions are meaningful and energies are calculated. Beyond CTONnb, the energy decreases sigmoidally upto CTOFnb. In case of switching function, the energy and interactions beyond CTOFnb drop to zero immediately, while in case of shifting functions, it continuously decreases to zero. Also, in case of shifting functions, only the CTOFnb value is meaningful; the CTONnb value is neglected. During MD simulations, the atoms are constantly in motion. Thus, if any one atom taking part in an interaction shifts slightly such that it moves out of the CTOFnb radius, the interaction energy would not be considered. To circumvent this problem, an additional cutoff radius is defined (CUTnb), which is used to generate the nonbonded list. However, energies are calculated only for a smaller fraction of atom pairs lying within the CTONnb radius. Generally,  $CTOFnb = CUTnb - 2$  (Angstrom) and  $CTONnb = CTOFnb - 2$  (Angstrom).

Three different algorithms are used in CHARMM to generate the nonbonded list, namely BYGROUPS, BYCUBES and BYCC (By-Cluster-In-Cubes) algorithms. BYGROUPS uses CHARMM standard atomic groupings and first creates a group-group pair and calculates the non-bonded atoms list from it. BYCUBES divides the system into cubical spatial regions (with sidelength equal to outer non-bonded cutoff value) and tests for interactions only between atoms in the same cube or in adjacent cubes. BYCC makes use of grouping as well as partitioning techniques and is the fastest of the three algorithms. However, it has markedly higher memory requirements as well. BYCBIM extends the BYCUBES algorithm to systems

with images or periodic boundaries and is the most efficient listbuilder in CHARMM for calculations in systems with image atoms.

### 2.3.2.2 Electrostatic interactions

Particle mesh Ewald (PME) method is used for calculation of electrostatics of a system in an infinite lattice or in case of periodic boundary conditions. The energy can be expressed as a lattice sum over all pair interactions and over all lattice vectors excluding the  $i=j$  term in the primary box. Basically, the Ewald sum splits electrostatic calculations to two rapidly convergent sums, one in real space and the second in reciprocal space.

Consider a system of point charges in the primary cell. The charge density is given by the equation:  $r_o(r) = \sum_i q_i \Delta(r - r_i)$ , which in turn can be split into two separate sums,

$$r_{o_1}(r) = \sum_i q_i (\Delta(r - r_i) - f(r - r_i)) \text{ and } r_{o_2}(r) = \sum_i q_i f(r - r_i)$$

Here,  $f(r - r_i)$  represents a spherical distribution (usually Gaussian) and the width of the Gaussian is specified by  $\kappa$ .  $r_{o_1}(r)$  represents the short range potential, calculated as a direct, real-space summation, while  $r_{o_2}(r)$  is a smoothly varying function of  $r$  and is expanded as a Fourier series. It is obtained by solving the Poisson equation (Ulrich Essmann et al., 1995).

### 2.3.3 Solvation of the system

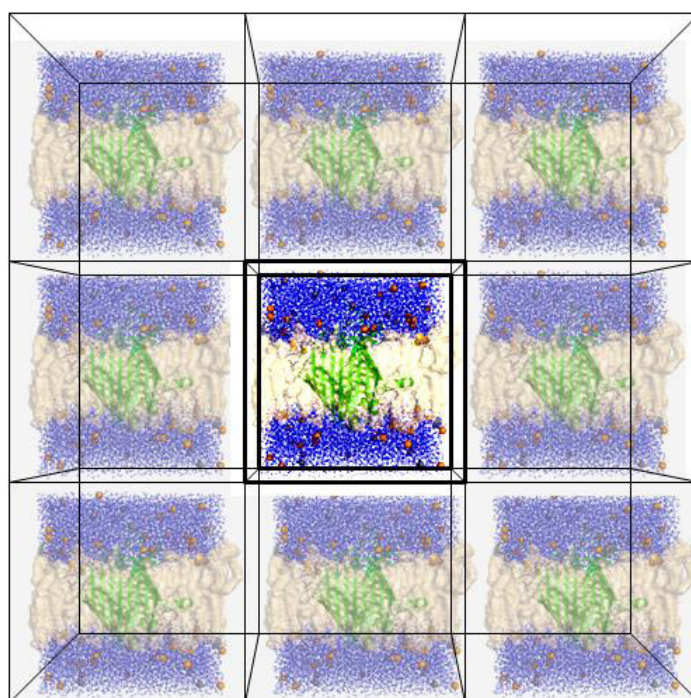
In order to mimic natural conditions during simulations, the system must be completely solvated. The type of solvent system used for simulations depends on several factors, such as system size, available computational resource, the type of data being generated and required accuracy of the results. In the present study, all simulations are performed in explicit solvent systems. Explicit solvent models involve adding discrete solvent molecules to the system (hundreds to several thousands). It results in a very large system size and is computationally demanding. The water model used in CHARMM is TIP3P, derived from the TIP3 model (William L. Jorgensen et al., 1983). It is a 3-site explicit water model (i.e., each individual water molecule has 3 sites of interaction, one corresponding to each atom of the water molecule). Each atom is assigned a point charge. In the TIP3 model, all 3 atoms are also assigned an LJ parameter, while in the derived TIP3P model, only the oxygen atom is assigned an LJ parameter.



### 2.3.4 Periodic Boundary Conditions in CHARMM

Periodic Boundary Conditions (PBC) are used to model an infinite system with a finite volume. i.e., a system containing relatively smaller number of particles can be simulated such that the particles experience forces as though they were in bulk solution. In order to study the bulk properties of the system, a solvent must also be introduced along with certain boundary conditions. If the boundary is assumed to be a rigid wall, surface effects would come into picture, thus affecting the bulk dynamics of the system. In order to remove these surface effects from the system, we use Periodic Boundary Conditions (Figure 2.2).

PBC is achieved by an infinite tiling/replication of the cubical simulation box in 3 dimensional space to create a lattice. Thus, when an atom moves in the primary cell, its images in all other boxes would move in the same way, thus maintaining the total number of atoms in the primary cell as constant. Further, it is not required to store the coordinates of all atoms in the lattice during the simulation, only those atoms in the central box are considered.



**Figure 2.2 Periodic Boundary Conditions (PBC).** The central cell (dark shaded), with a finite number of particles is surrounded by infinitely replicated periodic images of itself (light shaded), thus creating an infinite volume system. Note that an outer membrane protein (green) embedded in a lipid bilayer (shown in surface representation) and solvated in explicit solvent (blue) and with neutralizing ions added (orange) is considered to illustrate the PBC.

### 2.3.5 Energy minimization

Minimization is carried out to relieve strain in the initial model using equation (ii) to determine the energetically stable conformation. If the starting model used for MD simulations has a very high potential energy, it would heat up to an unreasonable temperature very fast. Thus, the potential energy of the system is minimized to the lowest possible value prior to MD simulations. CHARMM's minimization algorithms check the first and second derivatives of the potential energy function and adjust system coordinates to achieve minimum energy.

$$\frac{\partial E}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial^2 E}{\partial x_i^2} > 0 \dots(\text{ii})$$

CHARMM provides various minimization algorithms, namely, steepest descent (SD), conjugate gradient (CONJ), adopted basis Newton Raphson (ABNR), Newton Raphson (NRAP), Powell (POWE) and truncated Newton method (TNPACK). SD is the simplest method, in each step of which the coordinates are adjusted in the negative direction of the gradient. CONJ has better convergence than SD and checks the previous minimization history and current gradient to determine the next step. NRAP involves calculation of the derivative of the gradient to identify an energy minimum, ie., where the derivative is zero. In ABNR, energy minimization is based on a Newton Raphsons algorithm. POWE is generally used when ABNR cannot be applied. TNPACK is based on a linear conjugate-gradient technique for solving Newton equations.

### 2.3.6 Underlying principles of molecular dynamics

In MD simulations, the macroscopic properties of a system are studied using microscopic simulations with the help of statistical mechanics principles. One of the cornerstones of MD simulations is the '*Ergodic hypothesis*' which states that the time averages equals the ensemble average or  $\langle A \rangle_{\text{ensemble}} = \langle A \rangle_{\text{time}}$ . An 'ensemble' is nothing but a collection of all possible systems which have different microscopic states but identical macroscopic/thermodynamic states. The Ergodic hypothesis thus states that if a system is allowed to evolve in time it will eventually pass through all possible states. By carrying out sufficiently long MD runs, one can thus arrive at the minimum energy state/preferred state of the system being studied.

Commonly used thermodynamic ensembles in MD simulations are NVE (microcanonical ensemble : N, V , E constant) , NVT (canonical ensemble : N, V, T constant) and NPT (isothermal isobaric ensemble : N, P,T constant ), where N is the total number of particles in the system, P is the pressure, T is the temperature and V is the total volume of the system.

MD simulations are based on numerical integration of Newton’s second law of motion. Newton’s second law states that that the acceleration ( $a_i$ ) of a particle i , depends directly on the net force ( $F_i$ ) acting on it and inversely on the mass ( $m_i$ ) of the particle, i.e.,

$$F_i = m_i \times a_i \dots\dots\dots(iii)$$

However, the force can also be expressed as a gradient of the potential energy ( $E$ ), as

$$F_i = -\Delta_i E \dots\dots\dots(iv)$$

Equating equations (iii) and (iv), we get,

$$\frac{dE}{dr_i} = -m_i a_i = -m_i \frac{d^2r}{dt^2} \dots\dots\dots(v)$$

The solution of the differential equation (v) gives us the trajectory of particle i . However, the potential energy term is a function of the positions of all particles of the system over the simulation time and is highly complex. Thus, the solution of differential equation (v) is not computable. Instead, its value is approximated by numerical integration algorithms that approximate positions, velocities and accelerations of all the particles in the system being simulated. All integrators assume that the position, velocity and acceleration of particles can be approximated by a Taylor’s series expansion :

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots \dots\dots(vi)$$

Where, r is the position, v is the velocity (first derivative with respect to time), a is the acceleration (second derivative with respect to time) and so on.

### 2.3.7 Numerical integrators in CHARMM

As mentioned earlier, MD simulations require numerical integration of Newton’s second law of motion and involve approximations by various numerical integration algorithms. The most commonly used integration scheme is the ‘Verlet’ integration scheme which can be derived from equation (vi) as follows :

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots \dots\dots(\text{eqn vi})$$

$$r(t - \delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots \dots\dots(\text{vii})$$

$$(\text{iv}) + (\text{v}) \Rightarrow r(t + \delta t) = 2 r(t) - r(t - \delta t) + a(t)\delta t^2 \dots\dots\dots(\text{viii})$$

CHARMM supports five main integrators, namely ORIG, LEAP, VVER, VER4 and VV2. All these integrators are based on the Verlet scheme. ORIG is the oldest integrator, based on a lower precision verlet three step method. LEAP is the most commonly used and is based on the Verlet leapfrog algorithm. Velocity verlet integrator (VVER) is a high precision integrator and supports multiple-timestep methods, but is limited by a lack of pressure calculations. VV2 is one of the newer algorithms with improved temperature and pressure controls. VER4 can only be used for enhanced conformational sampling in 4 dimensional MD. All the above mentioned integrators are compatible with the SHAKE constraint. Leapfrog and VVER are used for the simulations in this study and are discussed below.

In the leap frog algorithm, velocities are calculated  $(t + \frac{1}{2}dt)$  and this is used to calculate positions (r values) at  $(t + dt)$ , i.e., the velocities ‘leap’ over positions and vice versa. The major disadvantage of the method is that velocities are not calculated at the same times as the position values. The velocity at time t is given by :

$$v(t) = \frac{1}{2} \left[ v \left( t - \frac{1}{2} \delta t \right) + v \left( t + \frac{1}{2} \delta t \right) \right] \dots\dots\dots(\text{ix})$$

VVER calculates velocities, positions and accelerations at time t as :

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \dots\dots\dots(\text{x})$$

$$v(t)\delta t = v(t) + \frac{1}{2} [a(t) + a(t + \delta t)]\delta t \dots\dots\dots(\text{xi})$$

CHARMM also supports non-Verlet integrators such as in the case of Langevin dynamics (LD) simulations, which are also used in the current investigation. In LD, coordinates of the particles at any point of time is determined by numerical integration of the Langevin equation instead of Newton’s equation of motion.

$$\text{Thus we have, } F_i = m_i a_i = -\nabla V_i - \gamma_i + R(t) \dots\dots\dots(\text{xii})$$

where,  $F_i$ ,  $m_i$ ,  $a_i$  are the force acting on particle  $i$ , mass of particle  $i$  and acceleration of particle  $i$  respectively.  $-\nabla V$  is the negative gradient of the system,  $\gamma$  is the frictional drag on the system and  $R(t)$  represents the jostling forces at time  $t$ . LD is similar to normal MD simulations, the major difference being that frictional drag is taken into account.

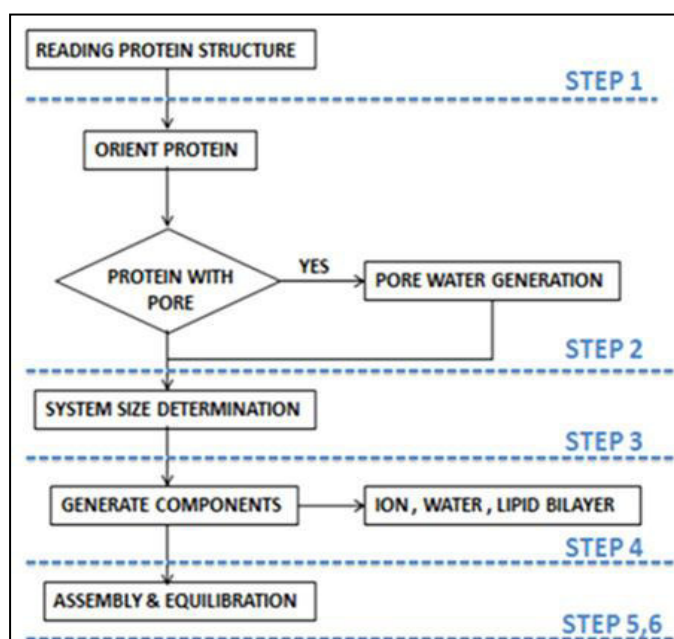
### 2.3.8 Thermostats in CHARMM

Berendsen and Nose Hoover thermostats are the two main thermostats used in CHARMM to maintain constant temperature during the simulation. The Langevin heatbath can be used to maintain constant temperature during LD simulations. Nose Hoover thermostat is part of CPT (Constant Pressure and Temperature) dynamics and is the best available thermostat. It is generally compatible with VER and VVER integrator. The system can be coupled to a single heat bath or to multiple heat baths using the Nose-Hoover thermostat.

The major advantage of Nose-Hoover over older methods is that it provides continuous dynamics with well defined conserved quantities. Methods such as Berendsen's thermostat have discontinuous dynamics, lower accuracy and are generally not preferred. Thus, we use Nose-Hoover thermostat in all simulations.

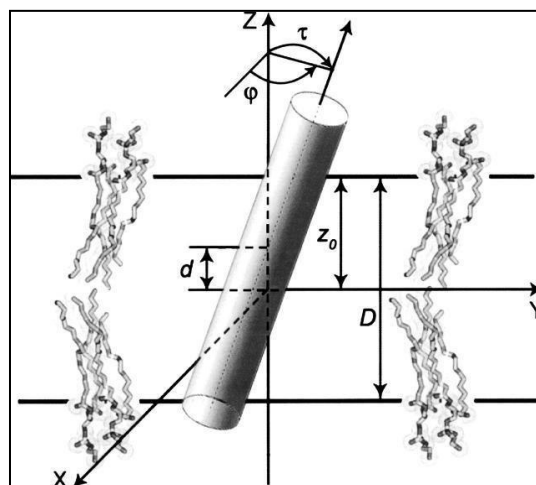
## 2.4 Generation of protein – bilayer models

A graphical web-server of CHARMM is used to generate the membrane embedded Wzi protein system. A schematic representation of the steps involved in building a model using the membrane builder module of CHARMM-GUI is illustrated in Figure 2.3.



**Figure 2.3 Building a protein – lipid membrane complex system using Membrane builder.** The steps involved in generating a model of a protein in a lipid bilayer are explained in text.

As shown in Figure 2.3, in **Step 1**, the coordinates of the protein is read using PDB reader. Coordinates of pre-oriented molecules can be read directly from the OPM database. The method used to orient proteins in OPM is shown in Figure 2.4. **Step 2** orients the protein along the Z axis with its centre at  $z = 0$ . This step is not required in case of structures from the OPM database.

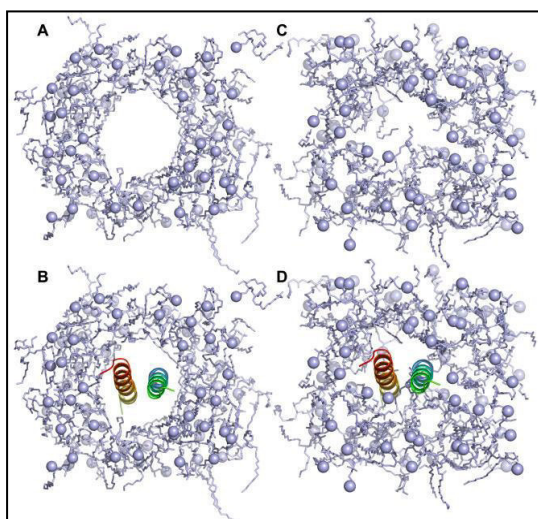


**Figure 2.4 OPM (Orientation of Proteins in Membranes) database :** The optimal spatial arrangement of a protein is determined by minimizing its transfer energy from water to a hydrophobic slab with a decadiene like property. (a) Schematic representation of a protein in a hydrophobic slab (Figure adopted from A. L. Lomize et al., 2006) . Four parameters are used to define the orientation of the protein in the membrane :  $d$ , shift along the bilayer normal;  $\tau$ , tilt angle;  $\phi$ , rotational angle;  $D (=2z_0)$ , the hydrophobic thickness of the protein (A. L. Lomize et al., 2006; M. A. Lomize et al., 2006). The orientation of 2436 membrane proteins in pdb have been calculated by this method and deposited in the OPM database (<http://opm.phar.umich.edu/>).

Step 2 also provides an option for generation of pore water and calculation of pore area (if required). For generation of pore water, first, the transmembrane region of the protein is solvated. The protein is kept fixed and planar restraint potentials are applied above and below the pore to prevent water within the protein pore from moving out. This is followed by high temperature dynamics at around 5000 K. Water molecules inside the pore remain while those outside get evaporated at the high temperature. A final refinement step is carried out to remove water molecules that may still be present, interacting strongly with the protein exterior.

In **step 3**, the system size is determined. The shape of the system (box type) can be specified as rectangular or hexagonal. The water thickness, i.e., the minimum height of the water layer above and below the system is also specified (default value of 20 Å is used in this study), based on which the Z extent of the system is estimated. The x and y dimensions of the system depend on the type of lipid selected to build the bilayer. Membrane builder supports many different kinds of lipid molecules, grouped under separate heads such as sterols, PA (Phosphatidic Acid) lipids, PC (Phosphatidylcholine) lipids, PE (Phosphatidylethanolamine) lipids, PG (Phosphatidylglycerolipids), PS (Phosphatidylserine) lipids, PUFA (PolyUnsaturated Fatty Acid) lipids and other bacterial lipids.

In **step 4**, the individual components such as lipid membrane, water and ions are built, based on the system size determined in the previous step. There are 2 methods to generate the lipid bilayer, namely, the Insertion method and the Replacement method. In the Insertion method, a hole of required size is created in the lipid bilayer and the membrane protein is inserted into it. Several pre-equilibrated bilayer libraries are available in CHARMM-GUI with holes of sizes varying from 0 to 45 Å for membranes of different sizes. This method is generally used in case of proteins with a regular and cylindrical shape. It cannot be used in case of heterogeneous lipid bilayers. The second method is the Replacement method. It can be used for any type of protein, irrespective of shape and size. In this method, lipid-like pseudo atoms are distributed around the protein and each pseudo atom is then replaced by a lipid molecule selected from the available libraries. The lipid molecules thus wrap around the protein. While the insertion method is faster and produces an equilibrated system, it has limitations based on system shape and size. The second method, on the other hand is slower, but can be used for any system (Jo et al., 2007).



**Figure 2.5 DMPC lipid bilayers generated by (A and B) the insertion method and (C and D) the replacement method for PDB:2HAC (Adopted from Jo et al., 2007).**

Since the simulations will be carried out in explicit solvent, the system is solvated with pre-equilibrated cubical TIP3 waterboxes of length 18.8560 Å. Based on the X, Y and Z extent of the system, the number of water boxes to be added to generate a planar layer of TIP3 is determined. The final waterbox is generated by stacking multiple planar layers one above the other. Ions are added to neutralize the system. Finally, 0.15 M concentration of KCl is added to neutralize the system. **Step 5** involves assembly of the generated components, i.e., lipid, protein, water and ions together to form the final system, which serves as input for further simulations.

The generated model is subsequently subjected to several rounds of equilibration runs with gradually decreasing constraints on the system components, in order to relax the system. This is followed by the production run. Details of the equilibration and production run protocols are discussed in the latter part of the chapter.

## **2.5 MD simulation of Wzi protein**

The X-ray crystallographic structure of the 50 KDa Wzi protein (determined at 2.6 Å resolution) (pdbid : 2YNK.pdb) is monomeric and consists of residues 24 – 479 (the first 23 residues form a cleavable signal sequence). Residues 275-289, (Leu-Thr-Gly-Lys-Asp-Asn-Thr-Ala-Ala-Asn-Asp-Pro-Asn-Glu-Pro), which form part of the extracellular loop, L5, of Wzi, are absent in the crystal structure. Thus, the missing residues are modelled by the Modloop server, prior to MD simulations.

### **2.5.1 Loop modelling of Wzi**

Loop modelling is done using the ModLoop web server which in turn makes use of the Modeller software, designed to refine X-ray crystallographic structures. Loop prediction by Modeller involves optimization of the position of all non hydrogen atoms of the loop residues in a fixed environment. The optimization protocol includes conjugate gradient minimization and MD with simulated annealing. The energy function for the minimization step uses the CHARMM force field potential energy function to restrain bonds, angles, dihedrals and impropers. In addition, restraints by statistical potentials, based on known protein structures, for mainchain and sidechain dihedrals and non-bonded atom pairs are used to model the loop.



## Optimization protocol for loop modelling

The optimization protocol consists of three steps : (i) Conjugate gradient minimization, (ii) MD with simulated annealing and (iii) a second Conjugate gradient minimization step. The **first step** is carried out to relax the system and consists of 5 successive minimization steps (upto 200 steps each) with a gradually increasing scaling factor,  $\xi$ , for non-bonded restraints (0, 0.01, 0.1, 0.5 and 1 respectively for the 5 steps). The atoms are allowed to pass very close to each other without creating large energy barriers and non-bonded atom pairs are generated only for the loop atoms, i.e., the environment is neglected. This helps lower the final energy values. The **second step** consists of two hundred 4 fs steps in which the systems are heated up very fast to 150, 250, 400, 700 and 1000 K during MD simulations. This is followed by the main optimization step, where the systems are cooled down from 1000, 800, 600, 500, 400 and 300 K respectively in six hundred steps of 4 fs each. The **third step** involves conjugate gradient minimization of upto 1000 steps (Fiser and Sali, 2003; Fiser et al., 2000).

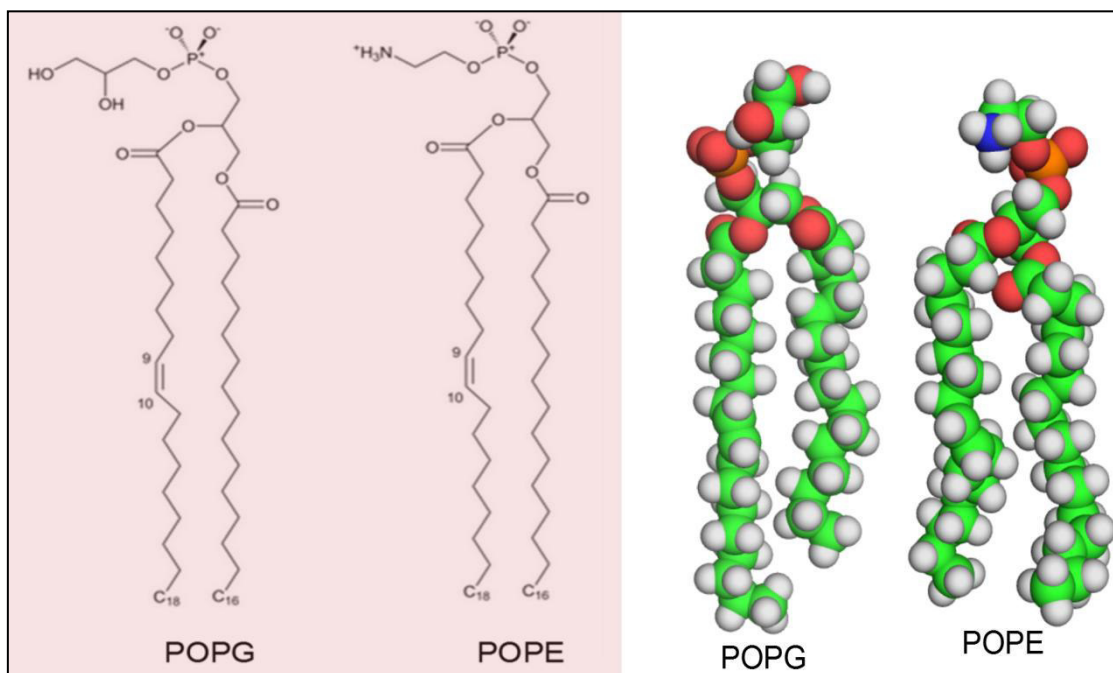
The maximum number of residues that can be modelled is limited to 20 to avoid any significant error. The protocol followed generates upto 300 independent optimized conformations from random initial conformations. The final loop prediction obtained is an optimized conformation with the lowest energy value.

### 2.5.2 Wzi models for simulation

The protein structure obtained from ModLoop is used as the starting structure for wild type Wzi simulations. Apart from the wild type protein, six mutants of Wzi are also subjected to MD simulations, as listed below.

- (i) Wild type Wzi obtained after loop modelling from ModLoop (Wzi<sup>WT</sup>)
- (ii) Wzi with a deletion of 15 residues in the extracellular loop L5, which is absent from the crystal structure (Wzi<sup>ΔL5</sup>)
- (iii) Wzi<sup>mut1</sup> - residues 32-36 (Asp-Leu-Arg-Asn-Asp) of helix H1 mutated to glycine
- (iv) Wzi<sup>mut2</sup> - residues 35-39 (Asn-Asp-Leu-Ala-Trp) of helix H1 mutated to glycine
- (v) Wzi<sup>mut3</sup> - residues 39-43 (Trp-Leu-Ser-Asp-Arg) of helix H1 mutated to glycine
- (vi) Wzi<sup>mut4</sup> – residue Tyr380 mutated to Trp [Wzi(y380w)]
- (vii) Wzi<sup>mut5</sup> – residue Tyr380 mutated to Ala [Wzi(y380a)]

From the wild type structure, six mutants are generated using CHARMM. For the Wzi<sup>ΔL5</sup> model, the crystal structure of Wzi is directly considered. Subsequently, the protein is embedded in a lipid bilayer and hydrated with water and neutralized with 0.15 M KCl as detailed previously. A heterogenous lipid bilayer consisting of Palmitoyloleoylphosphatidylglycerol (POPG) and Palmitoyloleoylphosphatidylethanolamine (POPE) in the ratio 1:3 is used to mimic the composition of the *E. coli* outer membrane (Chugunov et al., 2013). Structures of POPE and POPG are shown in Figure 2.6. The composition of each leaflet of the lipid bilayer is taken as 25 POPG and 75 POPE units to satisfy the 1:3 ratio. The replacement method is used to generate all the Wzi models. All systems are neutralized by the addition of 0.15 M KCl. The total system size (in terms of number of atoms), the number of water molecules, K<sup>+</sup> and Cl<sup>-</sup> ions added for each system is detailed in table 2.1.



**Figure 2.6. Chemical structures of lipids POPG and POPE used to generate the lipid bilayer.** Left, schematic representation of POPG and POPE. Right, POPG and POPE in sphere representation. Carbon, hydrogen, oxygen and nitrogen atoms are shown as green, white, red and blue spheres respectively. Both POPG and POPE possess a single C-C double bond between C9 and C10.

Models	Total system size (no. of atoms)	No. of water molecules	No. of K <sup>+</sup> ions	No. of Cl <sup>-</sup> ions	Simulation time (ns)
Wzi wild type	74982	14282	70	10	50
Wzi ( $\Delta$ L5)	73608	13893	68	10	50
Wzi (mut1H1)	75022	14311	69	10	50
Wzi (mut2H1)	74742	14217	69	10	50
Wzi (mut3H1)	74954	14291	70	10	50
Wzi Y380A	75252	14394	70	10	10
Wzi Y380W	75148	14353	70	10	20

### 2.5.3 Protocol followed during equilibration

All the seven Wzi system generated are subjected to six equilibration steps with gradually decreasing restraints, in order to relax the systems prior to the production run. The details of the equilibration steps are summarized in Table 2.2. Non-bonded interactions are calculated with standard values of ctonnb 10.0, ctofnb 12.0 and cutnb 16.0. The cutim value is taken as 16.0. Rectangular periodic box is considered for the simulations.

Langevin dynamics is implemented in the first two steps of equilibration (ie., for the first 50 ps) during which the temperature is maintained at 303.15 K. In the following four steps of equilibration, CPT dynamics is followed, where the temperature is again maintained at 303.15 K by the Nose-Hoover thermostat. As shown in Table 2.2, the force constants are gradually reduced and the system is relaxed. A total of 450 ps equilibration is carried out, of which the system is completely relaxed with a very low force constant on protein backbone alone in the last 100 ps.

Equilibration step	Restraining force constants (Kcal/mol · Å <sup>2</sup> )								Minimization (no. of steps)		Langevin dynamics	CPT dynamics	No. of steps x time = Time (ps)
	BB	SC	wforce	tforce	mforce	fcis	fc2	lon	SD	ABNR			
Step 6.1	10.0	5.0	2.5	2.5	2.5	250	250	10.0	1500	1500	✓	-	25000 x 0.001 = 25
Step 6.2	5.0	2.5	2.5	2.5	2.5	100	100	0.0	-	-	✓	-	25000 x 0.001 = 25
Step 6.3	2.5	1.0	1.0	1.0	1.0	50	50	0.0	-	-	-	✓	50000 x 0.002 = 100
Step 6.4	1.0	0.5	0.5	0.5	0.5	50	50	0.0	-	-	-	✓	50000 x 0.002 = 100
Step 6.5	0.5	0.1	0.1	0.1	0.1	25	25	0.0	-	-	-	✓	50000 x 0.002 = 100
Step 6.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	✓	50000 x 0.002 = 100

Here, BB and SC are the force constants used to restrain the protein Backbone and side chain respectively. wforce is the force constant used to keep water molecules away from the hydrophobic core, tforce is the force constant used to keep the lipid tail below +/-% and mforce is the force constant to keep lipid head groups close to target values. fcis and fc2 are dihedral restraint force constants to keep the cis double bond and c2 chirality respectively. Ion is the force constant applied on all the ions in the system.

#### **2.5.4 Protocol followed for production run simulations**

After completion of the equilibration steps, each Wzi system is subject to a production run under CPT dynamics. The system temperature is maintained at 303.15 K by the Nose-Hoover thermostat. Wzi wild type protein, loop 5 deletion mutant and the three helix substitution mutants are simulated over a 50 ns timescale. Simulations of y380a and y380w are ongoing and 10 ns and 20 ns trajectories respectively, have been analyzed as part of this study. The non-bonded cutoffs used are same as that in the equilibration steps.

#### **2.5.5 Application of transmembrane voltage**

The ultimate goal when carrying out MD simulation is to study the given system under conditions which most closely mimic natural conditions. In case of membrane protein simulations, one of the major factors to be taken into consideration is the cell transmembrane potential. Thus, in order to mimic natural conditions more closely and improve ion sampling events, a range of transmembrane potentials (from -100 to +100 mV) are applied to the system during the simulation.

This is achieved computationally by the application of a uniform electric field 'E' perpendicular to the membrane, such that the resulting applied membrane potential V equals the product of E and the length of the periodic cell in the direction perpendicular to the membrane, also known as the 'Constant electric field approach' (Gumbart et al., 2012). Consider a protein-bilayer system with the axis of the protein oriented along Z direction. The applied membrane potential V is then given by the product of  $E_z$ , the electric field applied along Z (perpendicular to the membrane) and  $L_z$ , the length of the entire periodic cell in the direction of applied electric field. i.e.,  $V = E_z \times L_z$

From the Wzi<sup>WT</sup> system obtained after 100 ps of production run (at 0 mV), eight new systems are derived with applied transmembrane voltages of -100 mV, -75 mV, -50 mV, -25 mV, +25 mV, +50 mV, +75 mV and +100 mV. These systems are simulated for a total of

25 ns under the CPT dynamics scheme with temperature maintained at 303.15 K using the Nose-Hoover thermostat. Above mentioned production run protocol is followed here as well.

## 2.6 MD simulations of K30 CPS

**Glycam Biomolecular builder** is used to build the starting model of K30 CPS ([http://glycam.ccruc.uga.edu/CCRC/biombuilder/biomb\\_index.jsp](http://glycam.ccruc.uga.edu/CCRC/biombuilder/biomb_index.jsp)). The individual sugars of K30 CPS, their conformations (alpha/beta) and the linkages between the sugars are specified to generate a hexameric model (consisting of six K30 CPS repeating units). The final structure obtained from Glycam biomolecular builder is a model based on the input sequence and linkages provided and does not represent the favoured conformation of the biomolecule. Thus, to understand the secondary structural preference and dynamics, MD simulations of the hexameric unit are carried out with CHARMM forcefield. Being highly flexible, the K30 CPS hexameric unit is expected to adopt several conformations at different populations, which have to be further analyzed in order to understand the most stable conformation. Subsequently, glycan reader of CHARMM-GUI is used to generate input files to perform MD simulations (Jo et al., 2011).

The K30 CPS hexamer is solvated with 11472 TIP3 molecules and neutralized with 36 potassium ions & 30 chloride ions. The system is subjected to 50 steps of minimization by the SD method followed by another 50 steps of minimization by ABNR. As before, the equilibration is carried out in several rounds. A short equilibration step of 100 ps is carried out using the VVER integrator followed by 400 ps equilibration using CPT dynamics. The system temperature is maintained at 300.0 K by Nose-Hoover thermostat. Non bonded cutoffs are the same as used in previous systems. Following equilibration, a 100 ns production run is carried out with CPT dynamics. The same nonbonded options are used as in the equilibration steps and temperature is retained at 300 K with the help of the Nose-Hoover thermostat.

Analysis of the various Wzi systems and K30 systems are carried out using CHARMM. The results obtained are discussed in the following chapters.

## 2.7 References

- Adcock, S.A., McCammon, J.A., 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi:10.1021/cr040426m
- Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr, Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M., Karplus, M., 2009. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi:10.1002/jcc.21287
- Chugunov, A., Pyrkova, D., Nolde, D., Polyansky, A., Pentkovsky, V., Efremov, R., 2013. Lipid-II forms potential “landing terrain” for lantibiotics in simulated bacterial membrane. *Sci. Rep.* 3, 1678. doi:10.1038/srep01678
- Durrant, J.D., McCammon, J.A., 2011. Molecular dynamics simulations and drug discovery. *BMC Biol.* 9, 71. doi:10.1186/1741-7007-9-71
- Fiser, A., Do, R.K., Sali, A., 2000. Modeling of loops in protein structures. *Protein Sci. Publ. Protein Soc.* 9, 1753–1773. doi:10.1110/ps.9.9.1753
- Fiser, A., Sali, A., 2003. ModLoop: automated modeling of loops in protein structures. *Bioinforma. Oxf. Engl.* 19, 2500–2501.
- Gumbart, J., Khalili-Araghi, F., Sotomayor, M., Roux, B., 2012. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochim. Biophys. Acta* 1818, 294–302. doi:10.1016/j.bbamem.2011.09.030
- Jo, S., Kim, T., Im, W., 2007. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PloS One* 2, e880. doi:10.1371/journal.pone.0000880
- Jo, S., Song, K.C., Desaire, H., MacKerell, A.D., Im, W., 2011. Glycan Reader: automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* 32, 3135–3141. doi:10.1002/jcc.21886
- Karplus, M., McCammon, J.A., 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi:10.1038/nsb0902-646
- Levitt, M., 2001. The birth of computational structural biology. *Nat. Struct. Biol.* 8, 392–393. doi:10.1038/87545
- Lomize, A.L., Pogozheva, I.D., Lomize, M.A., Mosberg, H.I., 2006. Positioning of proteins in membranes: a computational approach. *Protein Sci. Publ. Protein Soc.* 15, 1318–1333. doi:10.1110/ps.062126106
- Lomize, M.A., Lomize, A.L., Pogozheva, I.D., Mosberg, H.I., 2006. OPM: orientations of proteins in membranes database. *Bioinforma. Oxf. Engl.* 22, 623–625. doi:10.1093/bioinformatics/btk023
- Monticelli, L., Tieleman, D.P., 2013. Force fields for classical molecular dynamics. *Methods Mol. Biol. Clifton NJ* 924, 197–213. doi:10.1007/978-1-62703-017-5\_8
- Ponder, J.W., Case, D.A., 2003. Force fields for protein simulations. *Adv. Protein Chem.* 66, 27–85.
- Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, Lee G. Pedersen, 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593.

- William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, Michael L. Klein, 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- Zhu, X., Lopes, P.E.M., Mackerell, A.D., Jr, 2012. Recent Developments and Applications of the CHARMM force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2, 167–185. doi:10.1002/wcms.74

## **CHAPTER 3**

# **Structure and dynamics of K30 Capsular polysaccharide**

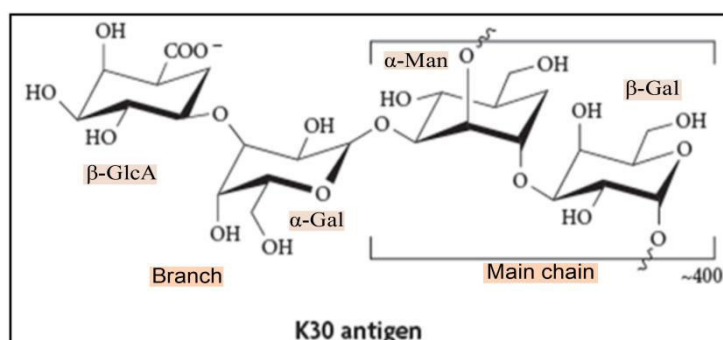


## CHAPTER 3:

### Structure and dynamics of K30 Capsular polysaccharide

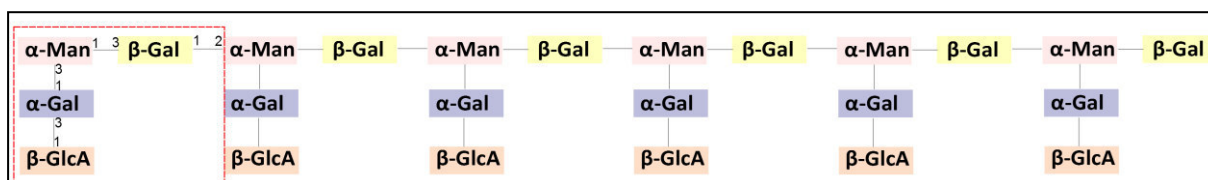
#### 3.1 Introduction

K30 CPS belongs to Group 1 of capsular polysaccharides of *E. coli*. Though the structure of the K30 repeat unit was determined in 1980 (Chakraborty et al., 1980) (see Figure 3.1), its structural preference remains unknown to date. Here, we carry out MD simulations of a hexameric unit of K30 CPS (Figure 3.2) to understand its dynamics and conformational preferences.

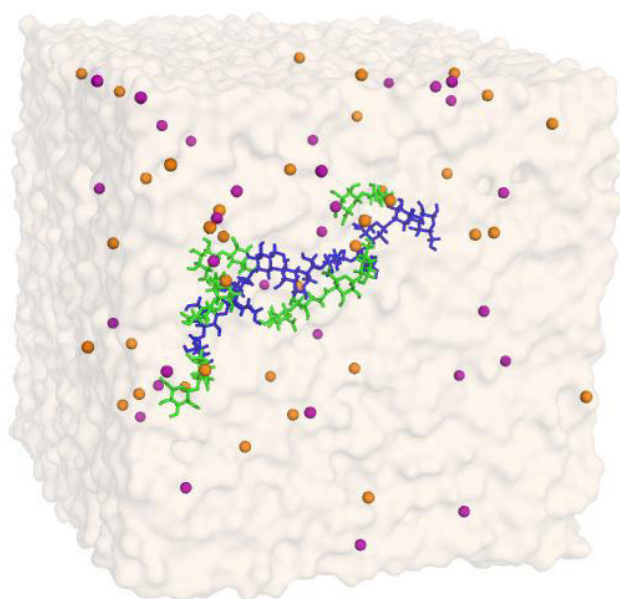


**Figure 3.1 Primary structure of K30 CPS.** K30 CPS repeating unit is a tetramer (Chakraborty et al., 1980). The main chain consisting of alternating  $\alpha$ -Man and  $\beta$ -Gal residues and branches at each mannose residue consisting of  $\alpha$ -Gal and  $\beta$ -GlcA units are marked.

Figure 3.2 illustrates the schematic representation of the hexameric unit simulated in the present study (A single repeating is indicated in a red box) and the entire system considered for simulations is depicted in Figure 4.3.



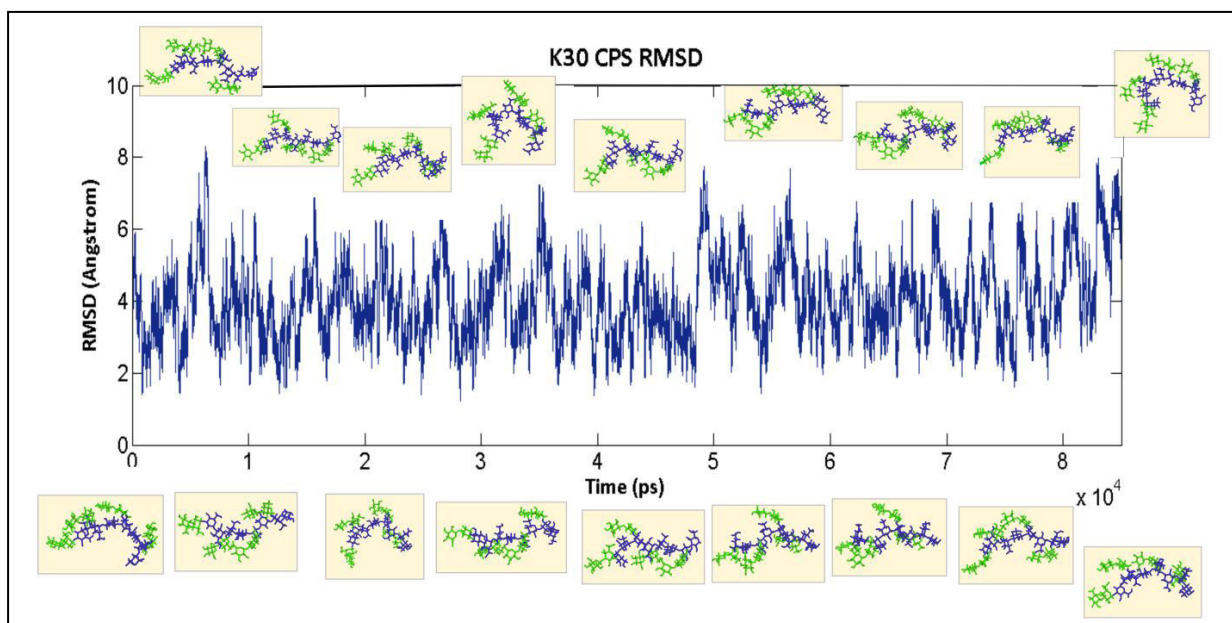
**Figure 3.2 Schematic diagram of K30 CPS hexameric unit used for simulations.** A single monomeric unit of K30 is shown in a red box. Connectivity between the residues is also shown.  $\alpha$ -mannose is shown in pink,  $\beta$ -galactose in yellow,  $\alpha$ -galactose in purple &  $\beta$ -glucuronic acid in orange.



**Figure 3.3 K30 CPS hexameric unit** (main chain shown in blue and branches in green) surrounded by water (shown as surface) with ions  $K^+$  and  $Cl^-$  (shown as orange and purple spheres respectively).

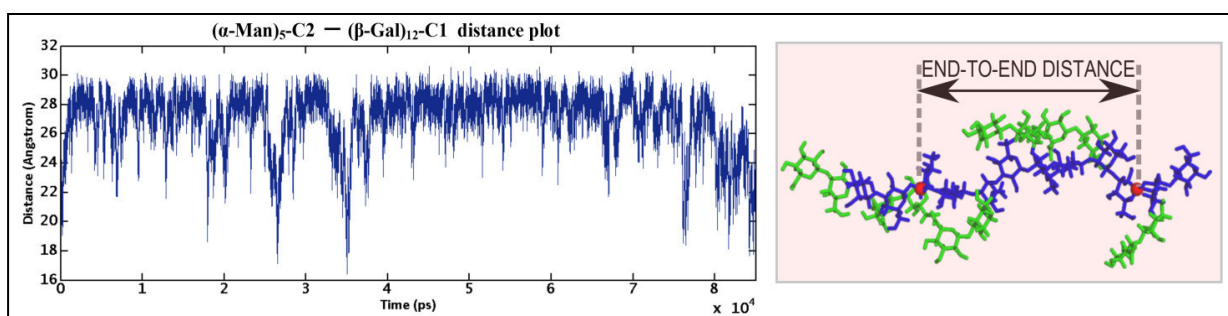
### 3.2 Dynamics of K30 CPS

We simulate the K30 CPS hexamer (Figure 3.2) upto 85 ns. RMSD of the structure plotted with respect to time is depicted in Figure 3.4. Snapshots of K30 structure at every 5 ns are also shown. Note that the terminal repeats are not considered for analysis due to end-fraying effect.



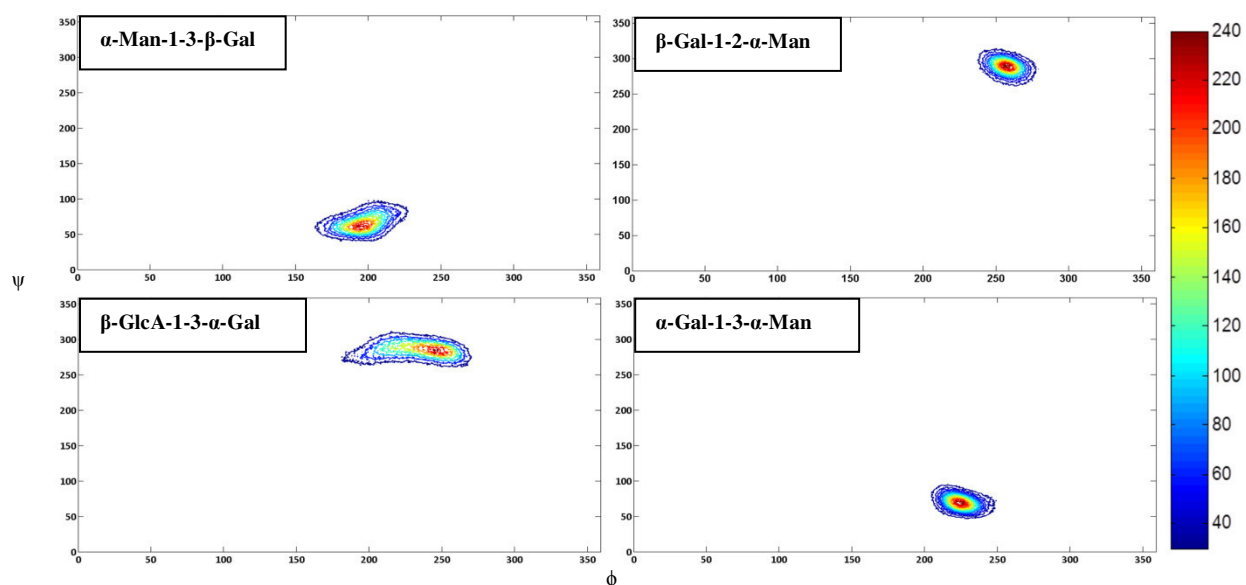
**Figure 3.4 RMSD plot of K30 CPS plotted with respect to time.** Snapshots of CPS are shown at 5 ns intervals. The main chain of CPS is shown in purple and the branches in green.

From Figure 3.4, it can be seen that there are mainly two preferred populations of K30 CPS hexamer, namely an extended form and a bent form. We also calculate the end-to-end distance of the K30 CPS unit. End-to-end distance calculated between C2 of  $\alpha$ -Man (of the first repeating unit) and C1 of  $\beta$ -Gal (of the last repeating unit) plotted with respect to time (Figure 3.5) shows the initial distance of  $\sim 20$  Å getting extended to  $\sim 30$  Å. This extended conformation is retained over a major part of the simulations, with occasional dips to 16-18 Å. Towards the end of the simulation, the end-to-end distance shows a slow downward trend. Longer simulations of larger K30 units may be required to accurately understand its structural preference.



**Figure 3.5 End-to-end distance of K30 CPS.** (Left), End-to-end distance plotted with respect to time, (right), Structure of K30 CPS, with the main chain shown in blue and branches in green. The atoms used to measure the end-to-end distance, namely,  $\alpha$ -Man C1 and  $\beta$ -Gal C2 are shown as red spheres.

Torsion angles are calculated for ( $\alpha$ -Man)-1-3-( $\beta$ -Gal) & ( $\beta$ -Gal)-1-2-( $\alpha$ -Man) which form the main chain and ( $\alpha$ -Gal)-1-3-( $\alpha$ -Man) & ( $\beta$ -GlcA)-1-3-( $\alpha$ -Gal) of the branched chain residues and the preferred range of torsion angles is determined (Figure 3.6 and Table 3.1). The torsion angle values lie in the (gauche-, gauche-) region for three of the linkages, namely  $\beta$ -Gal-1-2- $\alpha$ -Man of the main chain and  $\beta$ -GlcA-1-3- $\alpha$ -Gal and  $\alpha$ -Gal-1-3- $\alpha$ -Man of the branched chain while it is (trans, gauche-) for the  $\alpha$ -Man-1-3- $\beta$ -Gal linkage of the main chain. Thus, the possibility of a helical conformation cannot be completely ruled out.



**Figure 3.6** Torsion angle contour plot of K30 CPS. (Top left) corresponds to  $\alpha$ -Man-1-3- $\beta$ -Gal, (top right)  $\beta$ -Gal-1-2- $\alpha$ -Man, (bottom left)  $\beta$ -GlcA-1-3- $\alpha$ -Gal and (bottom right)  $\alpha$ -Gal-1-3- $\alpha$ -Man linkages. Scaling used for the contour plot is shown on right. Colour is based on the density of points, red being the highest and blue the lowest.

**Table 3.1** Preferred values of torsion angle

Linkage	$\phi$		$\psi$	
	Value	Defining atoms	Value	Defining atoms
<b>(<math>\alpha</math>-Man)-1-3-(<math>\beta</math>-Gal)</b>	60	O5-C1-O3-C3	170	C1-O3-C3-C2
<b>(<math>\beta</math>-Gal)-1-2-(<math>\alpha</math>-Man)</b>	250	O5-C1-O2-C2	270	C1-O2-C2-C1
<b>(<math>\beta</math>-GlcA)-1-3-(<math>\alpha</math>-Gal)</b>	250	O5-C1-O3-C3	280	C1-O3-C3-C2
<b>(<math>\alpha</math>-Gal)-1-3-(<math>\alpha</math>-Man)</b>	225	O5-C1-O3-C3	70	C1-O3-C3-C2

### 3.3 Conclusion

From the simulations carried out, we infer that there are two preferred conformations of K30 CPS, namely a compact/bent form with an  $\alpha$ -Man C1 to  $\beta$ -Gal C2 (end-to-end) distance of  $\sim 16$ - $17$  Å and a more extended form with an end-to-end distance of nearly 30 Å. This, together with the torsion angle calculation offers the favourable conformation for CPS. The possibility of a helical conformation is revealed based on the preferred values of torsion angles. In order to understand the dynamics of the molecule and its preferred conformation in more detail, longer simulations are necessary.

### **3.4 References**

Chakraborty, A.K., Friebolin, H., Stirm, S., 1980. Primary structure of the *Escherichia coli* serotype K30 capsular polysaccharide. *J. Bacteriol.* 141, 971–972.

## **CHAPTER 4**

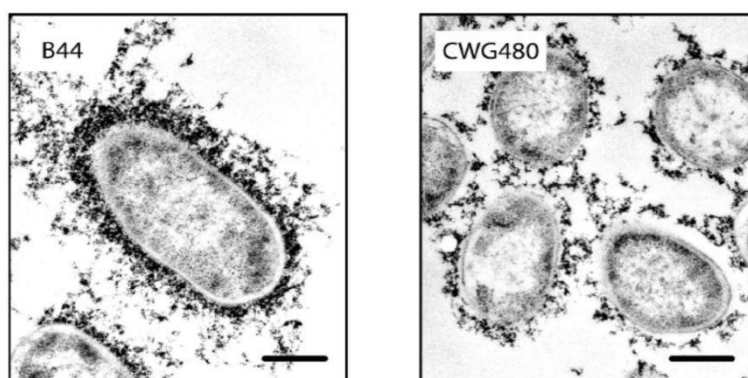
# **Structural characterization of Wzi**

## CHAPTER 4 :

### Structural characterization of Wzi

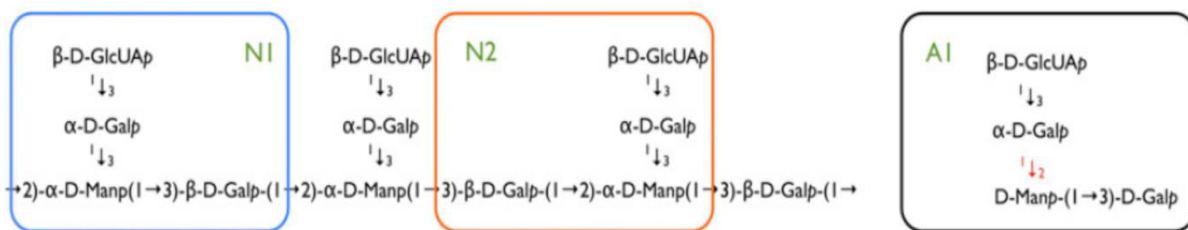
#### 4.1 Introduction

Wzi is a 50 KDa outer membrane beta barrel protein that is a part of the Wzy dependent supramolecular assembly which is responsible for capsular export and surface expression. It functions as a lectin and binds K30 CPS to the surface of the bacteria to form the high molecular weight capsule. Figure 4.1 shows the capsular morphology of a Wzi chromosomal insertion mutant, where the capsule is no longer tightly attached to the surface, but forms a diffuse layer similar to exopolysaccharide secretions.



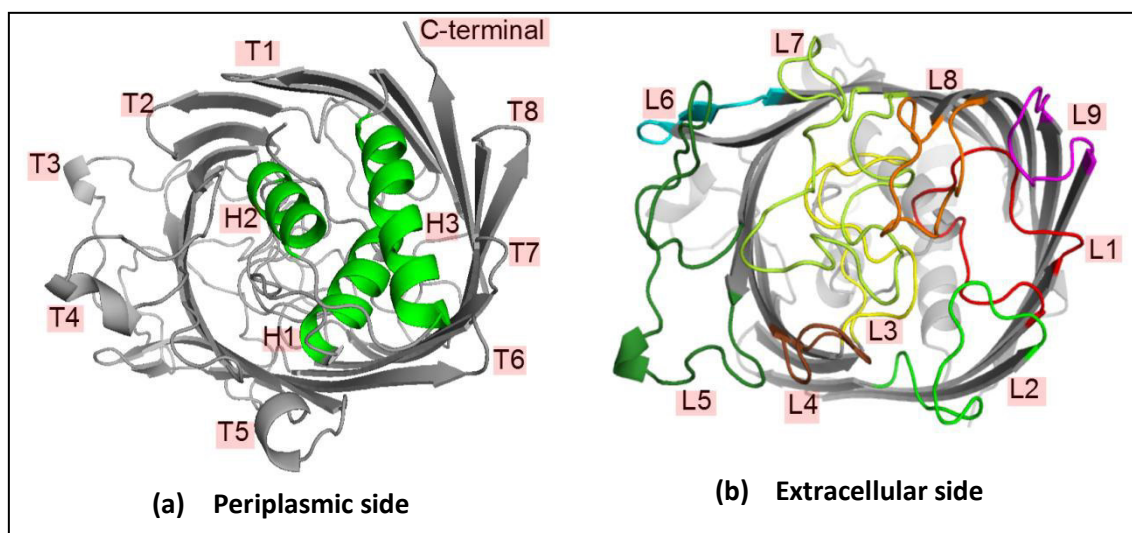
**Figure 4.1** Capsule morphology of *E.coli* B44 (O9:K30) (left) and its *wzi* chromosomal insertion mutant, CWG480 (right) labelled with cationized ferritin and visualized by electron microscopy. Surface association of the capsule is reduced in case of the mutant and it remains loosely attached to the cell, similar to EPS. Wzi protein thus plays an important role in surface anchorage of CPS. (Adopted from Rahn et al., 2003).

To understand the mode of interaction of Wzi with K30 CPS, binding of three different monomeric units (Figure 4.2) has been tested using Surface Plasmon Resonance (SPR) (S R Bushell et al., 2013). Two of the monomeric units are contained within the K30 CPS repeat unit and one is analogous to the K30 structure (refer Figure 4.3). Out of the three, binding is observed in case of N2 alone, indicating high specificity for K30 recognition by Wzi. However, the K30 CPS recognition site in the extracellular region of Wzi is still unknown and information about it would facilitate the design of drug molecules that target CPS binding.



**Figure 4.2** Monomeric repeating units of K30 CPS tested for binding to Wzi by SPR. N1 and N2 are segments of the K30 CPS unit and A1 is an analogous tetrasaccharide (note the difference in linkage at the branching point). Of the three, binding is observed only in case of N2 (Adopted from (S R Bushell et al., 2013)).

The X-ray structure of Wzi (2.6 Å resolution) was determined fairly recently by Naismith and co-workers (S R Bushell et al., 2013) and its detailed conformation in an explicit membrane environment is unknown. Though it is implicated in the surface anchorage of K30 CPS, based on mutagenesis studies and SPR, the exact mechanism involved is not precisely understood. In this context, we carry out MD simulation of wild type Wzi and its six mutants. Wild type Wzi simulations are also carried out at a range of transmembrane voltages to understand voltage dependent properties of the protein, if any.

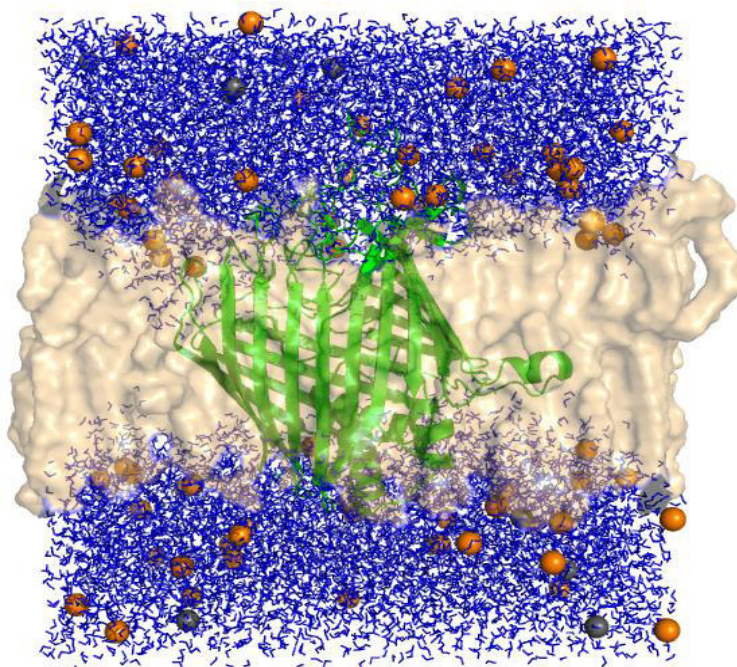


**Figure 4.3** (a) Periplasmic view of Wzi protein (shown in grey in cartoon representation) with the periplasmic helical bundle (encompassing three alpha helices, H1, H2 and H3) shown in green. Periplasmic turns T1 to T8 are also marked. (b) Extracellular view of Wzi protein, with loops L1 to L9 shown in different colours.



As detailed in chapter 2, Wzi beta barrel has nine extracellular loops and eight transmembrane turns (Figure 4.3). A helical bundle composed of three alpha helices occludes the barrel at the periplasmic end. Previous mutational studies have shown that individual deletion of loops L3, L6 or L7 of Wzi results in loss of K30 binding to Wzi. As these loops form a positive patch, it is suggested that they may interact with the negatively charged K30 CPS. However, deletion of L8 is found to have no effect. Another important and surprising observation is that substitution mutations of the middle five residues (residues 35-39) or last five residues (39-43) of helix H1 of Wzi to glycine lead to a loss of binding activity even though it does not interact directly with K30 CPS. However, mutation of the first five residues (32-36) of H1 to glycine is not found to have any effect (S R Bushell et al., 2013).

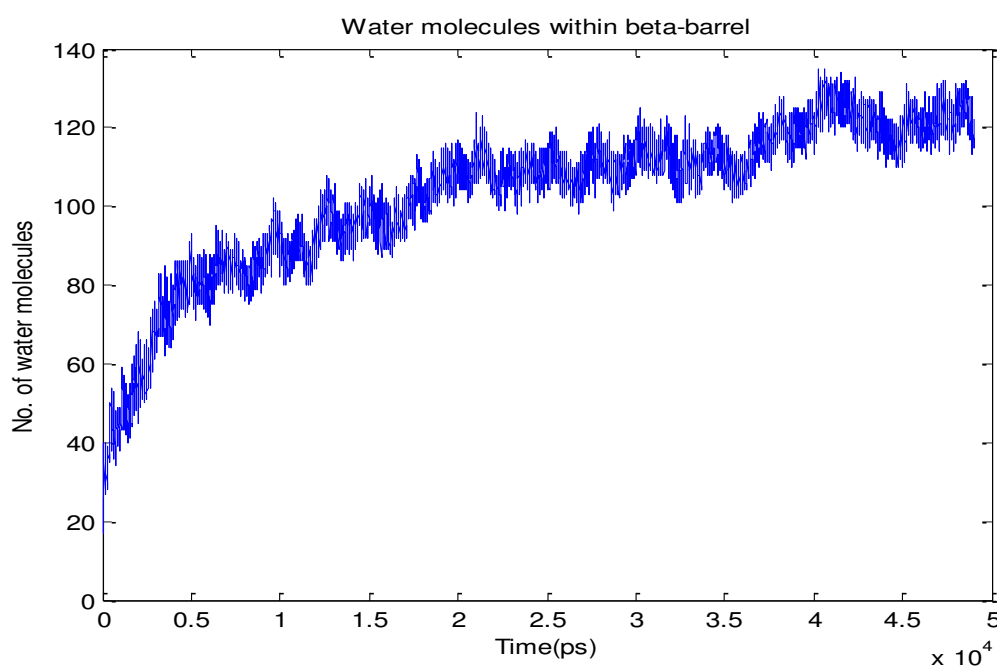
Thus, in the present study, we carry out MD simulations of wild type Wzi and its three helical mutants wherein residues of helix H1 are mutated to glycine. Two point mutants, Y380W and Y380A are also simulated to understand the role of conserved residue Y380 in the surface attachment of K30 to Wzi. In addition, wild type Wzi protein is simulated at eight different transmembrane voltages (ranging from -100 mV to +100 mV in intervals of 25 mV) to 25 ns each. The starting model for Wzi simulations is shown in Figure 4.4.



**Figure 4.4** Initial model of Wzi considered for MD simulations. Wzi is shown in green in cartoon representation, embedded in a lipid bilayer (shown in yellow in surface representation) and solvated with explicit water (shown in blue) and 0.15 M KCl added (potassium & chloride ions shown as orange and grey spheres respectively).

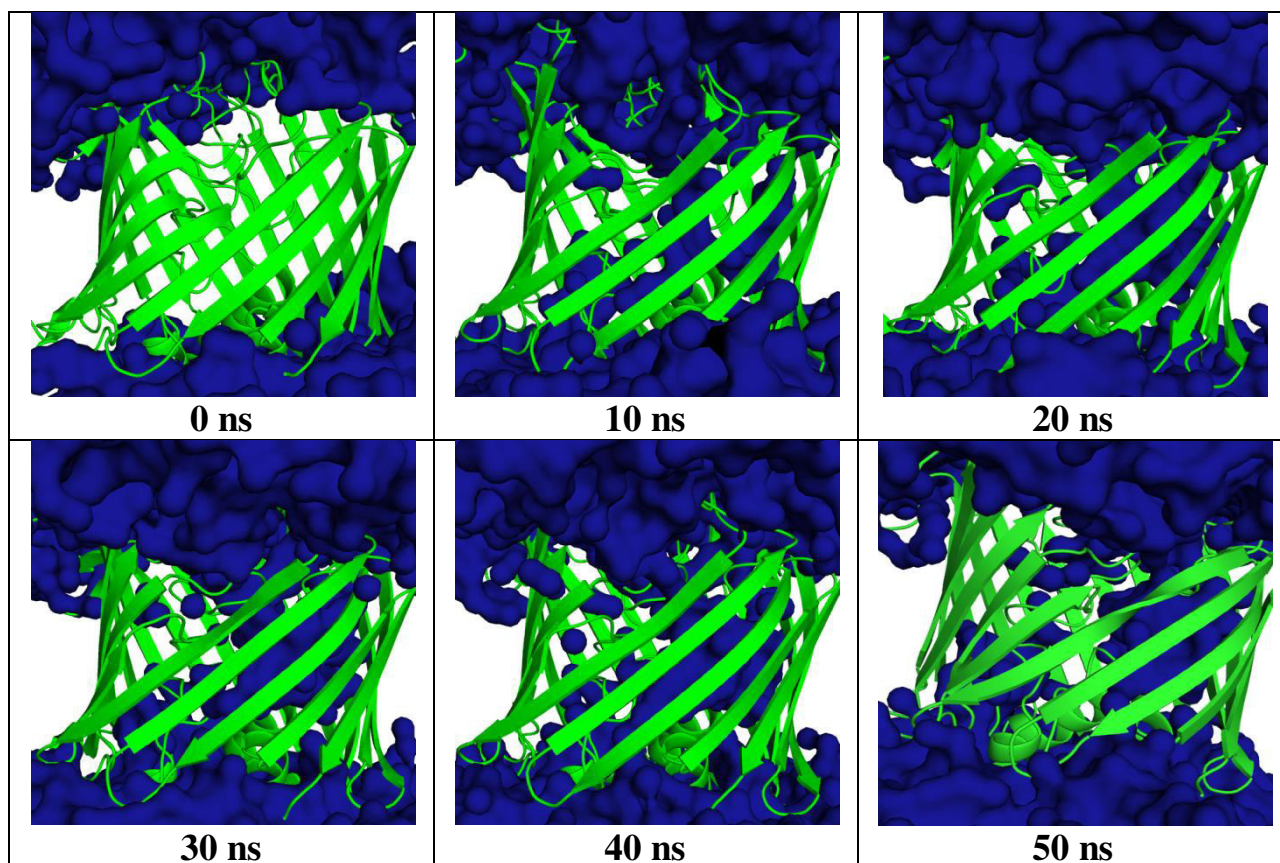
## 4.2 Water conducting property of Wzi protein

Our first major observation on simulation of wild type Wzi is its water conducting property. This is a new and surprising development as the inner surface of the protein is lined by a large number of hydrophobic residues and the extracellular loops and the periplasmic helical bundle plug the channel from both ends. We observe that the interior of the protein which is initially devoid of any water molecules gradually gets filled up with water, starting from ~2-3 ns (Figure 4.5). The channel gets completely filled with water by 6ns.



**Figure 4.5** Number of water molecules within the beta barrel of wild type Wzi plotted as a function of time. As the channel is oriented along the Z axis (see section 2.4), water molecules within -10 to 10 Å along the Z axis are alone considered. The total number of molecules increases from 20 to 120 over the course of a 50 ns simulation.

The number of water molecules present within the barrel over the course of simulation is plotted as a function of time (Figure 4.5). A steep rise is observed in the number of molecules over the first 5-6 ns of simulations, after which it increases gradually and reaches ~120 at 50 ns. Snapshots of the Wzi system at every 10 ns over the 50 ns trajectory are depicted in Figure 4.6. Water entering and filling up the channel that is empty initially can be clearly observed.

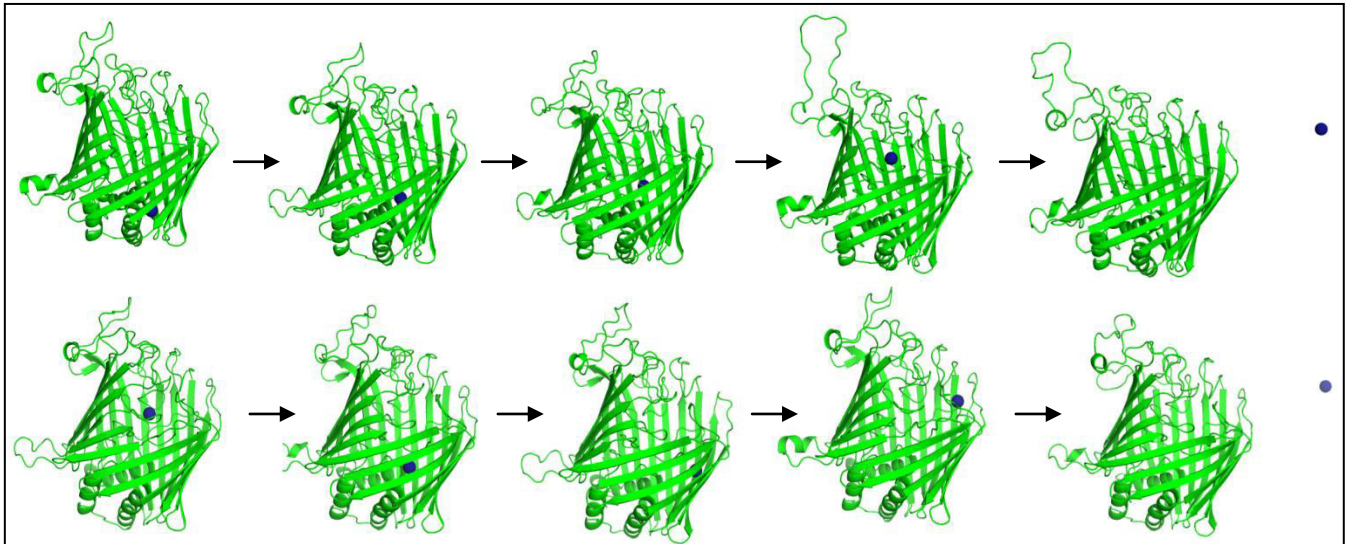


**Figure 4.6** Water conduction in wild type Wzi. Snapshots of Wzi protein (green coloured cartoon) at 10 ns intervals over a 50 ns trajectory. Water (blue coloured surface) is observed to gradually fill up the protein channel.

Detailed investigation shows that the water molecules entering Wzi do not just fill up the channel, but are also conducted across it. Figure 4.7(a) traces the movement of a single water molecule from the periplasmic to the extracellular side. In addition, we also observe that water molecules can enter into the pore from the extracellular side, circulate within the protein and then exit from the same side without crossing the channel. Figure 4.7 (b) depicts the entry of a water molecule from the extracellular side, followed by its circulation within the protein and exit from the same side.

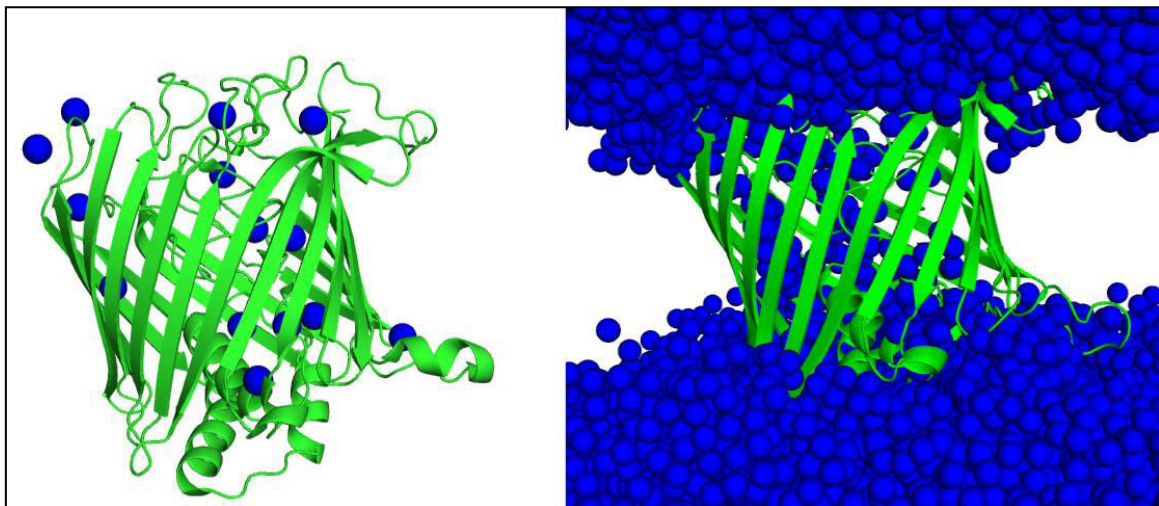
#### 4.2.1 Comparison with crystal structure of Wzi

To further justify the water conducting property observed in the simulations, we compare the results with the crystal structure of Wzi. There are 14 water molecules present in the crystal structure of Wzi [PDBID: 2YNK.pdb] (Figure 4.8 (left)), of which 6 reside completely within the beta barrel.



**Figure 4.7** Transport of a single water molecule (shown as blue spheres) through Wzi protein (coloured green): (a) from the periplasmic side to the extracellular side and (b) Circulating within the channel.

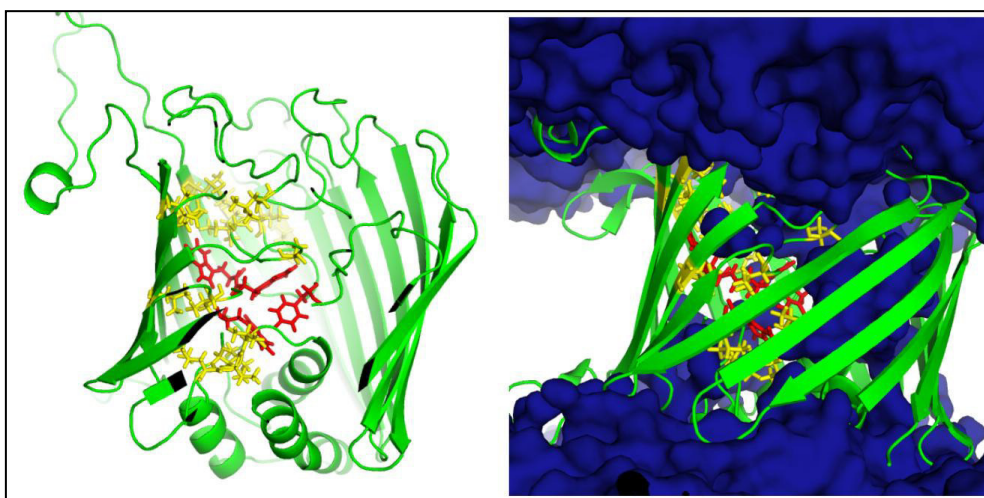
Since the crystal structure just provides a snapshot of Wzi, its water conduction property was not realized earlier. The current MD simulation alone provides evidence for the ability of Wzi to conduct water. Thus, the presence of water molecules in the Wzi crystal validates and strengthens our observation of the water conducting capability of Wzi beta barrel protein.



**Figure 4.8** Comparison of crystal structure of Wzi crystal structure (left) and Wzi wild type protein after 10 ns simulation (right). Protein is shown in green and water is displayed as blue spheres.

#### 4.2.2 Role of hydrophobic plug in regulating water conduction

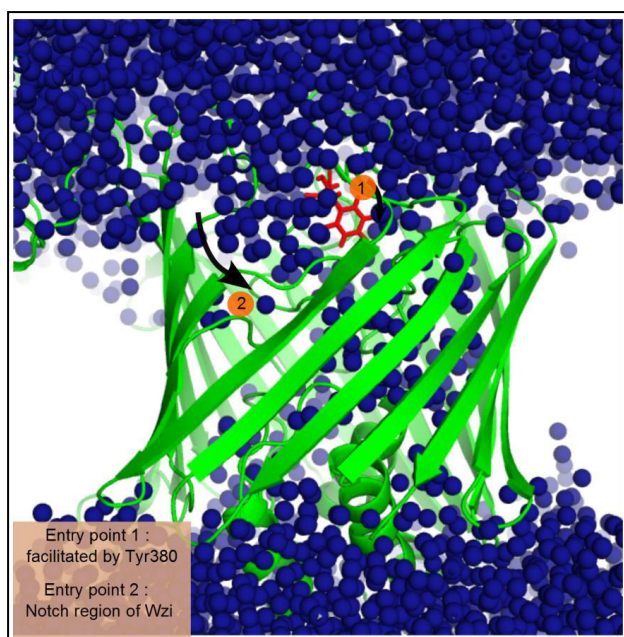
Interestingly, water transport through the protein is restricted to a specific pathway, mediated by the presence of a hydrophobic plug. The plug is observed in the interior of the Wzi protein (Figure 4.9), formed by aromatic residues Trp174 & Trp175 of loop L3, Phe109 of L1 and Trp53 which is part of the connecting loop between periplasmic alpha helices H1 and H2. Multiple non-aromatic hydrophobic residues (Ala, Val, Leu and Ile) are also found in this region, as part of the plug. This hydrophobic plug alters the movement of water (both circulating and crossing) through the channel, confining it to a specific path.



**Figure 4.9** Hydrophobic bundle of Wzi. A hydrophobic bundle consisting of aromatic residues Trp174, Trp175, Phe109 and Trp53 (red sticks) and non-aromatic residues Ala, Val, Leu and Ile (yellow sticks) in the interior of Wzi protein (shown as cartoon) alters the course of water (blue surface).

#### 4.2.3 Entry of water from the extracellular side

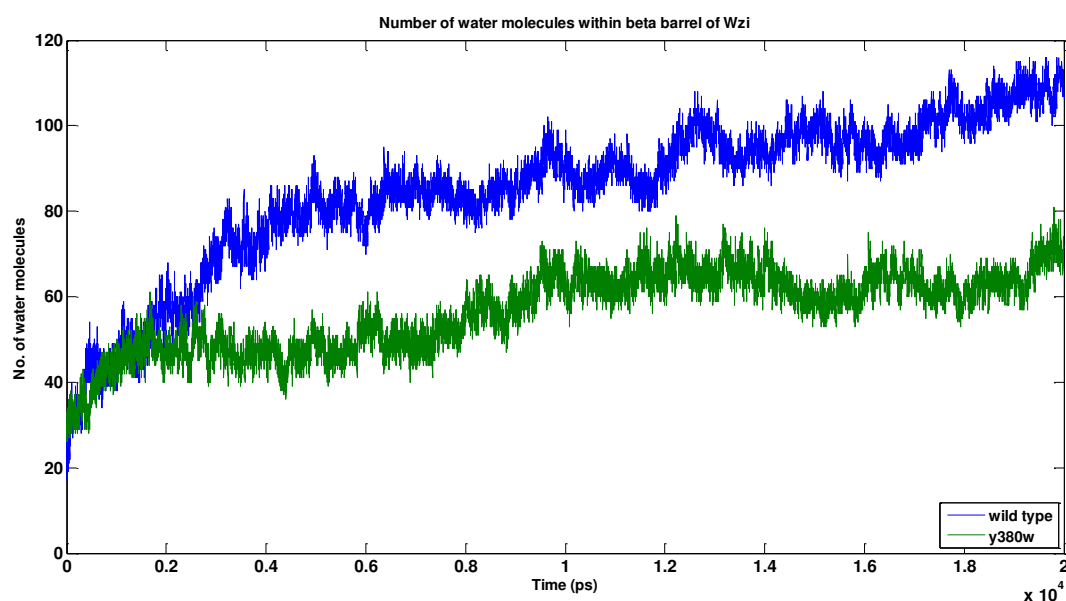
There are primarily two points of entry for water molecules from the extracellular side into the protein as shown in Figure 4.10. Residue Tyr380 plays an important role in facilitating the entry of water molecules into the channel via the first path. To further understand the importance of this residue in regulation of water conduction, MD simulations of two extreme point mutants Y380A and Y380W are carried out. Y380W is considered for the simulation based on the assumption that it can block water entry through entry point 1 (Figure 4.10) & can act as a blocker, whereas, Y380A is designed as it can widen the pore and upregulate water conduction. The second point of entry of water into the channel is via the notch region of Wzi.



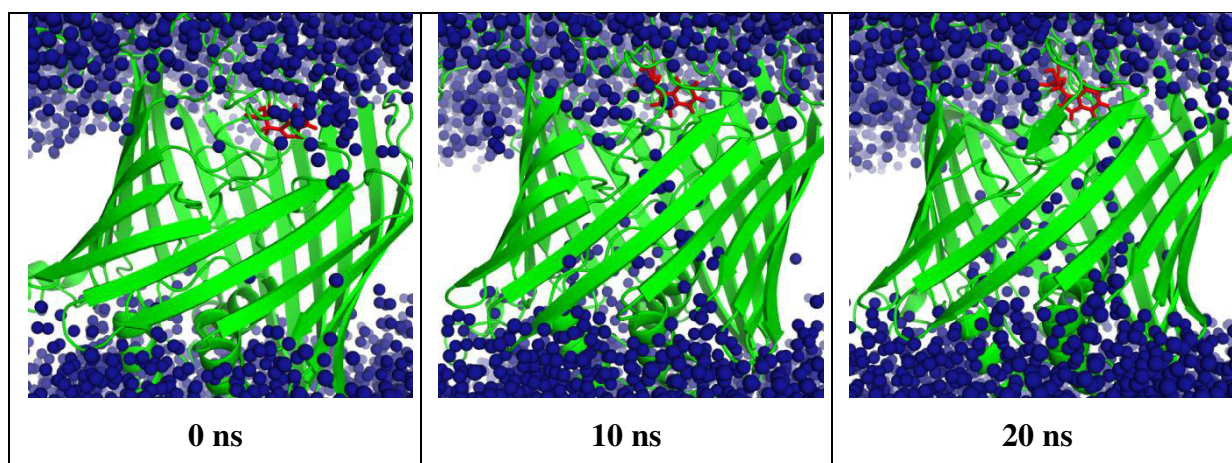
**Figure 4.10** Entry of water (blue spheres) into Wzi protein (green cartoon) from the extracellular region. Tyr380 facilitates entry of water from point 1, while the second point of entry is via the notch region.

#### 4.2.3.1 Simulations of Y380 mutants

MD simulation of Y380W shows the water conduction through entry point 1 to be blocked. The number of water molecules within the beta barrel of the Y380W mutant is reduced to nearly half of the wild type simulations (Figure 4.11). This significant reduction emphasizes the importance of Y380 in influxing water from the extracellular region. It is noteworthy that water still enters into the protein via the notch region. Snapshots of the system at 10 ns intervals showing water conduction blocked by Y380W is depicted in Figure 4.12.

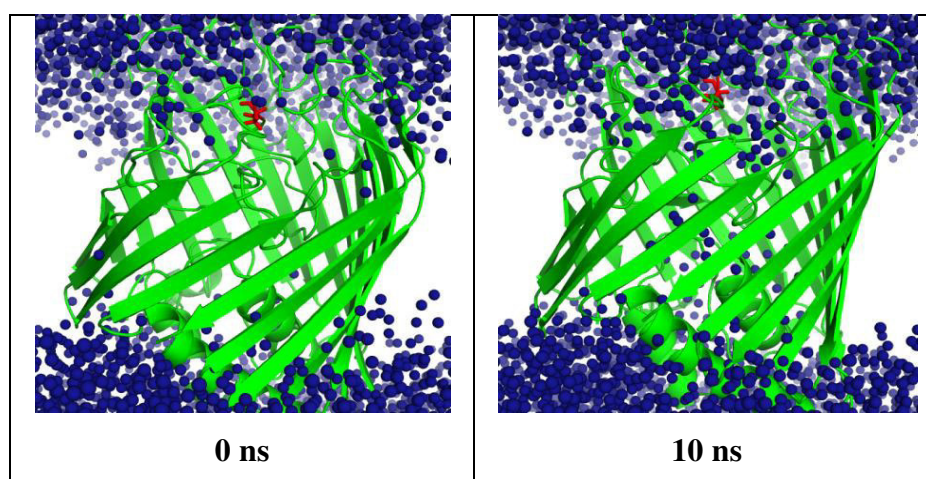


**Figure 4.11** Comparison of number of water molecules within beta barrel of Wzi wild type protein (blue) and Y380W point mutant (green). Calculations are the same as in Figure 4.5.



**Figure 4.12** Water conduction of Y380W mutant. Snapshots at 0 ns, 10 ns and 20 ns of Y380W mutant of Wzi protein (coloured green) blocking conduction of water influx (blue spheres) by Trp380 (red sticks).

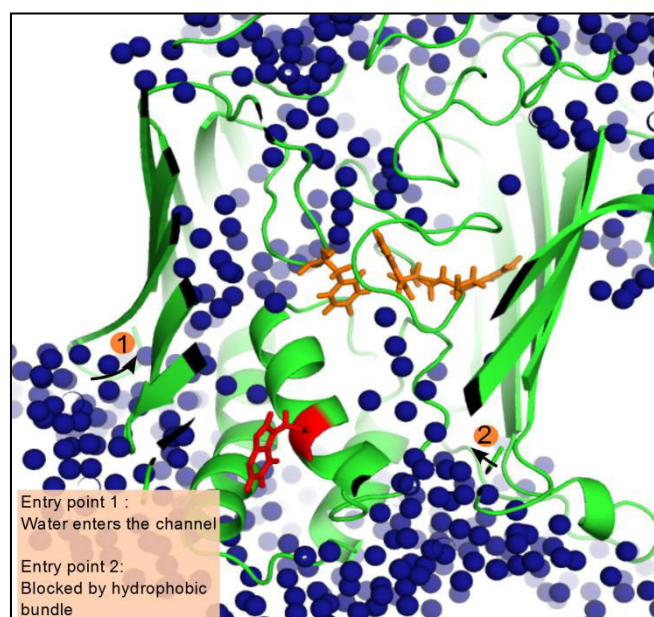
We also carry out simulation of a second point mutant, Y380A. In this case, water enters from the notch region as well as from the top of the protein. Ala380, being a smaller residue, does not hinder the passage of water. Snapshots of the system showing water conduction at 10 ns is depicted in Figure 4.13.



**Figure 4.13** Water conduction through Y380A mutant. Snapshots at 0 ns and 10 ns of the y380a mutant Wzi protein (coloured green) showing conduction of water (blue spheres). Ala380 is shown in red.

#### 4.2.4 Entry of water from the periplasmic side

There are two points of entry of water into the protein from the periplasmic side, as shown in Figure 4.14. Water passing through point 1 enters the protein channel and gets conducted across it. However, water entering from point 2 gets blocked by the hydrophobic plug of Wzi and is unable to cross the channel.



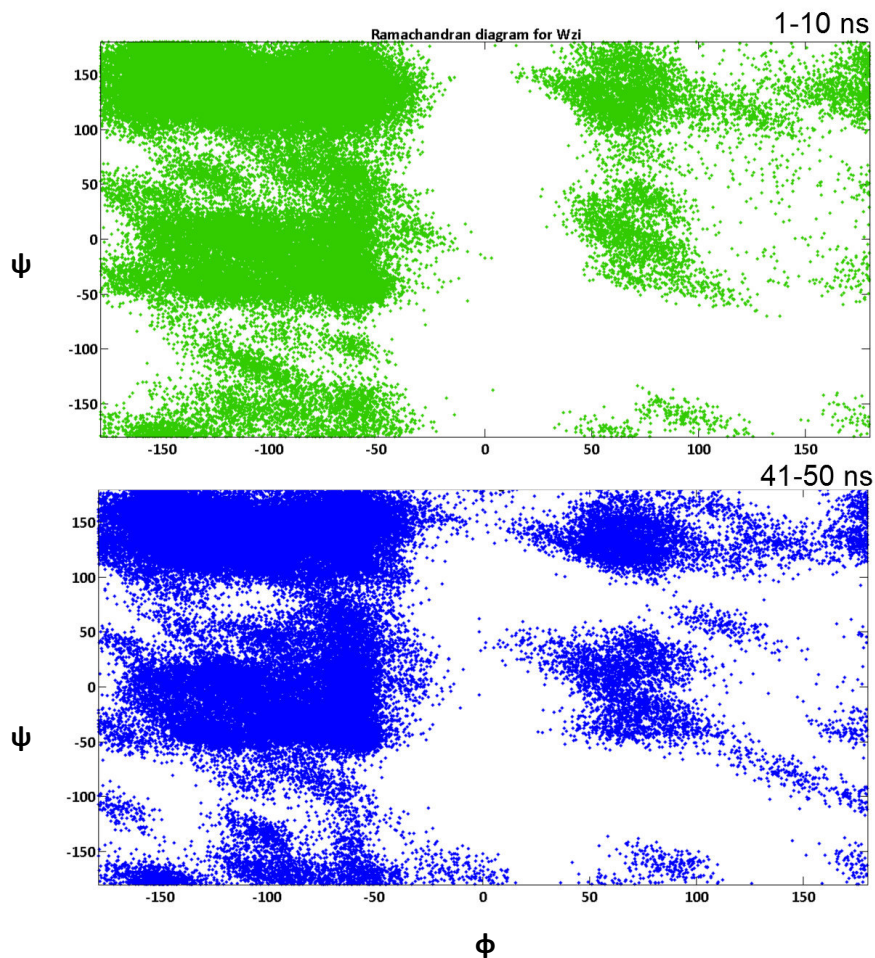
**Figure 4.14** Entry of water (shown as blue spheres) from the periplasmic side into Wzi (coloured green). Water entering from point 1 gets conducted across the channel while that from point 2 is blocked by the hydrophobic bundle (coloured orange). Trp39 of helix H1 (shown in red) blocks the entry of water.

Hydrophobic residue Trp39 which is part of periplasmic alpha helix H1, lies close to the periplasmic end of the protein and blocks the water molecules and directs them along a specific path. Mutants in which this residue is absent show a change in the water conduction pattern as discussed in section 4.6.

#### 4.2.5 Secondary structural changes

Ramachandran plots of the whole protein over the first 10 ns and last 10 ns of simulations (Figure 4.15) indicate that the values of  $\phi$  and  $\psi$  remain within the allowed regions.

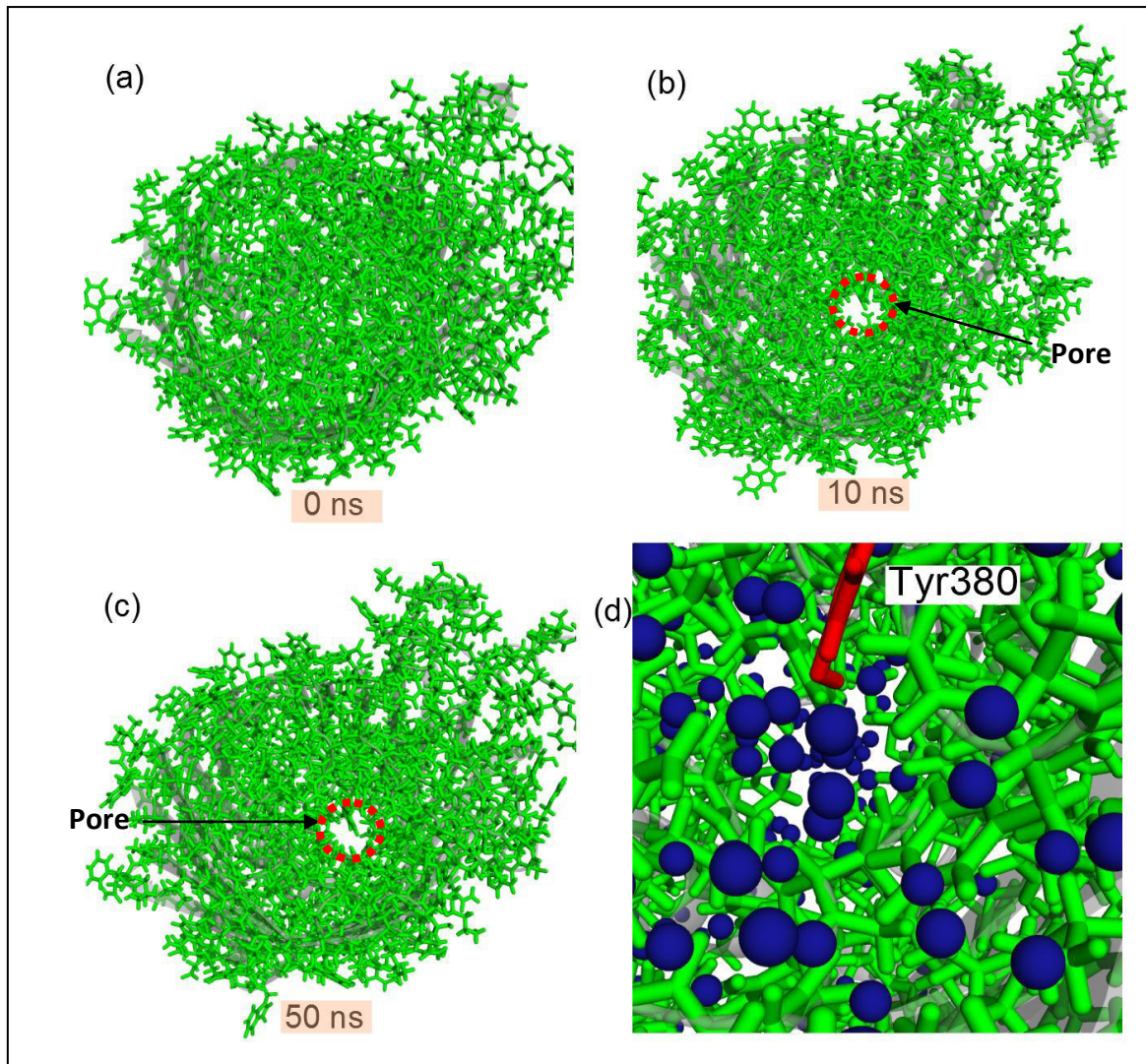




**Figure 4.15** Ramachandran diagram for wild type Wzi protein. Phi and Psi angles over the first 10 ns (shown in green) and last 10 ns (shown in blue) of the simulations fall within the same range.

#### 4.2.6 Pore opening

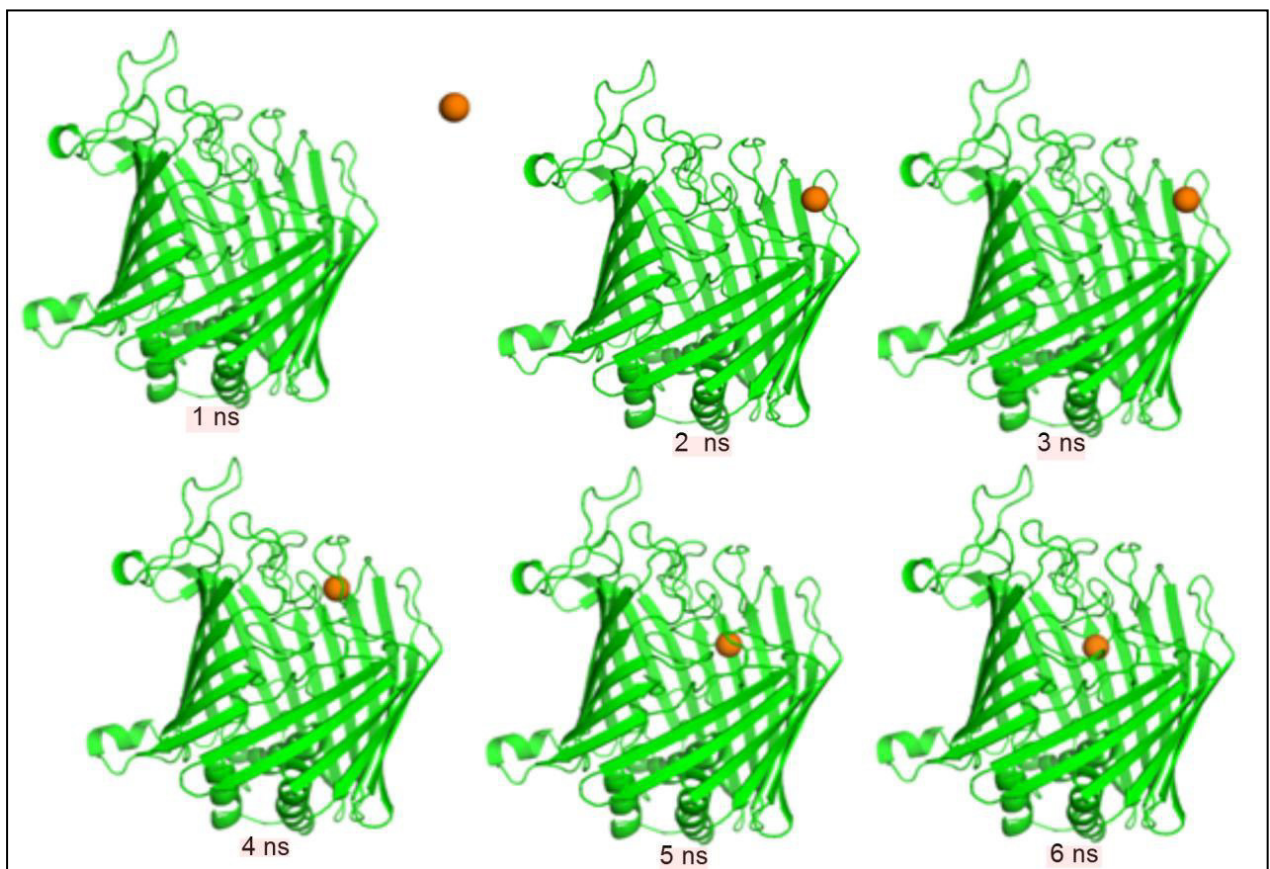
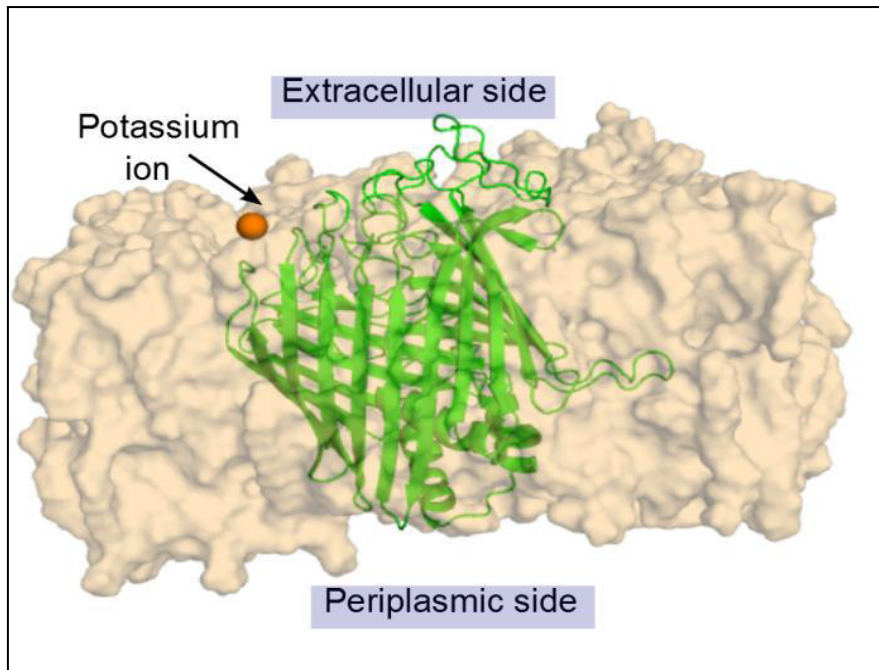
Interestingly, during the course of simulations, we observe the formation of a pore at the extracellular side of Wzi protein. Figure 4.16 (a) depicts the structure of Wzi viewed from the extracellular region at 0 ns, where the pore is absent. The snapshots at 10 ns and 50 ns clearly depict pore formation. Figure 4.16 (d) shows the passage of water molecules into the pore. The ‘OH’ group of Tyr380, projects into the pore and facilitates the entry of water into it.



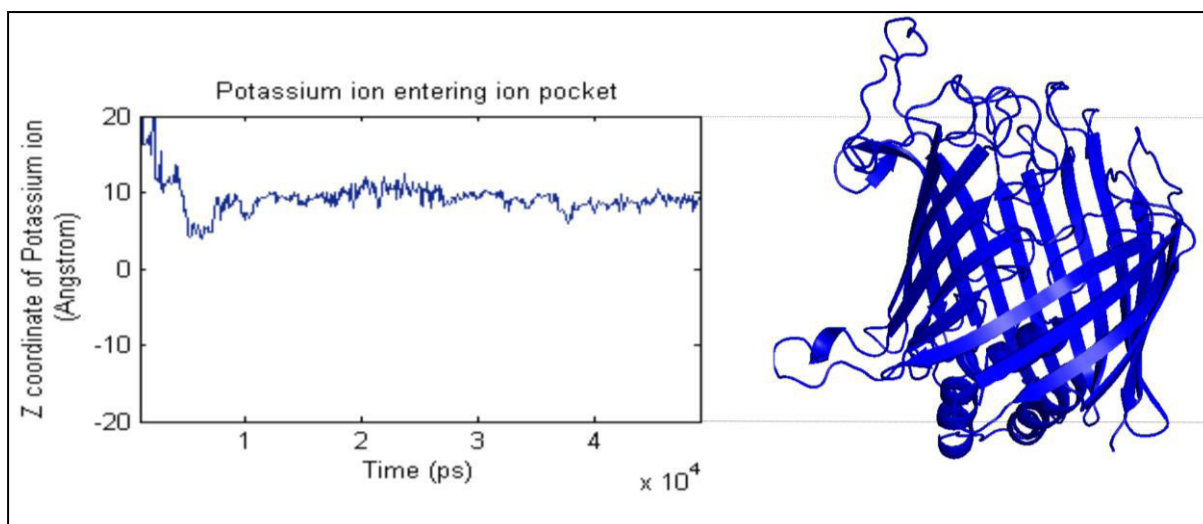
**Figure 4.16** Pore opening at the extracellular region of Wzi protein. Snapshots of Wzi (coloured green) as viewed from the extracellular region at (a) 0 ns, (b) 10 ns and (c) 50 ns. Pore formation is demarcated by red circle in (b) and (c). (d) depicts water molecules passing into the pore at 50 ns. Tyr380, which facilitates water entry, is shown in red in stick representation.

### 4.3 Ion binding pocket of Wzi

A second major observation is the presence of an ion binding pocket on the extracellular face of the protein. A single potassium ion enters into the pocket from the extracellular region and remains stably associated with the protein throughout the simulation timescale (Figure 4.17). Ion entry starts ~2-3 ns and the ion enters inside the pore completely by ~6ns. It then remains inside the channel upto 50 ns. Figure 4.18 depicts the Z coordinate of the potassium ion as it enters into the ion pocket.



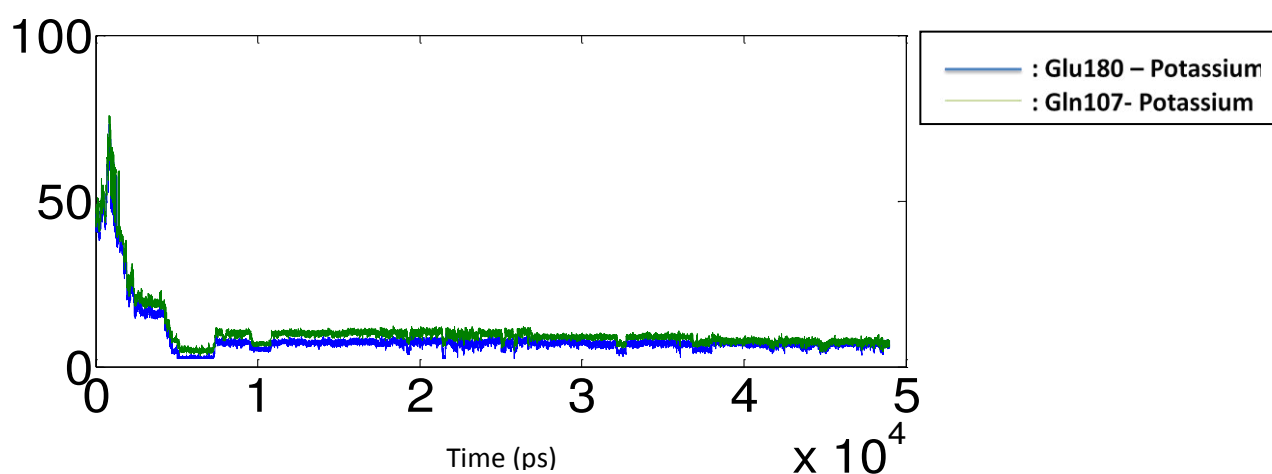
**Figure 4.17** Entry of Potassium ion into Wzi. Top, Snapshot at 2ns of Wzi protein (coloured green) in a lipid bilayer (shown as surface) and potassium ion (orange sphere) on the extracellular side of the membrane, prior to ion entry into the channel. Bottom, snapshots depicting the entry of potassium ion (orange) into the protein (green) from 1 to 6 ns.



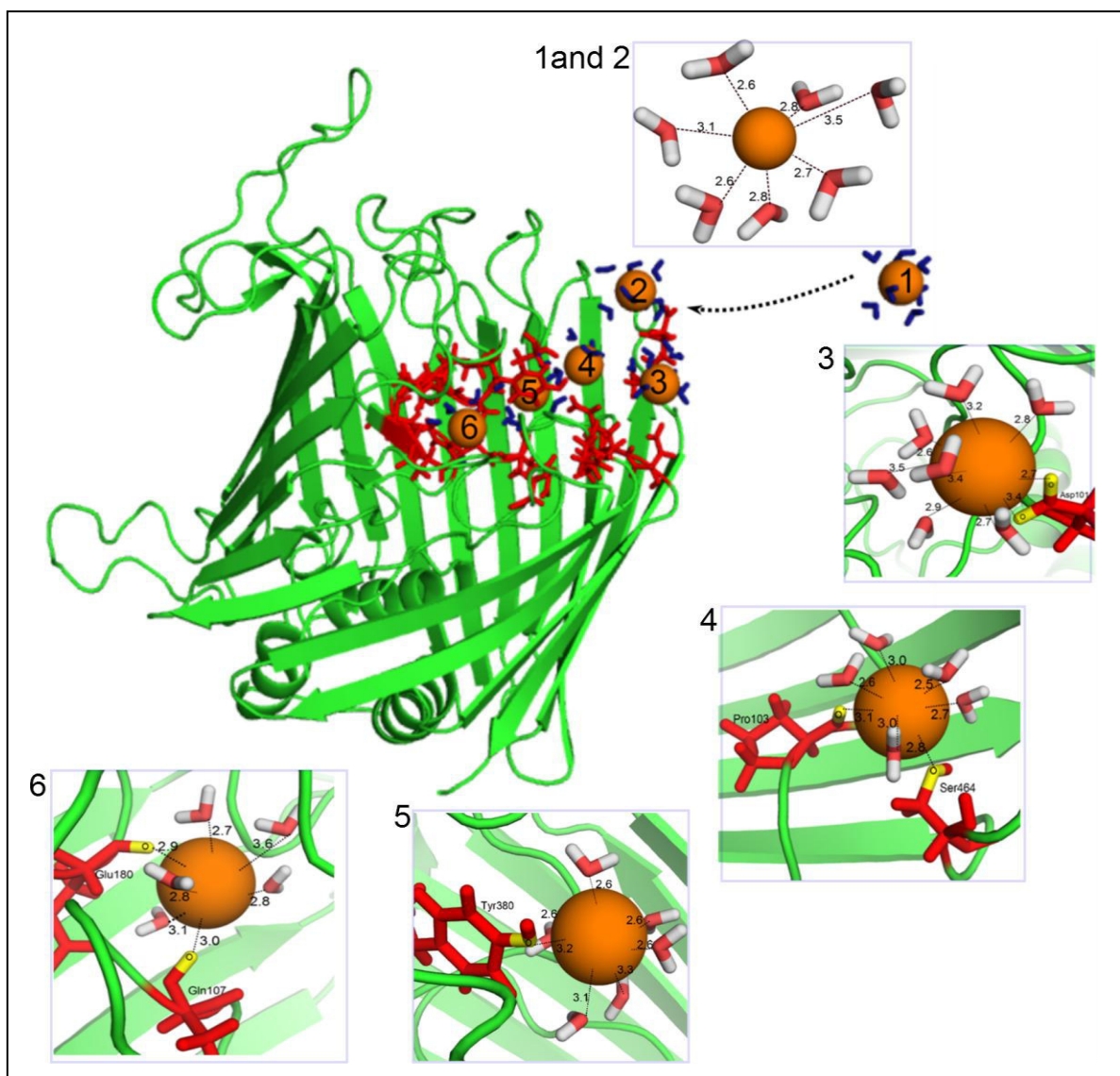
**Figure 4.18** Z coordinate of potassium ion entering into the Wzi channel. The potassium ion enters the channel at ~2-3 ns and remains within the pocket upto 50 ns.

The potassium ion initially interacts alternately with two negatively charged residues on the extracellular side, namely Asp101 and Asp463. These negatively charged residues attract the positively charged ion towards the protein. Note that the two residues are found to be conserved across several *E.coli* and *Klebsiella* species.

Once the potassium ion crosses the Asp residues, it moves further inwards, coordinating with the carbonyl oxygen of Tyr380. It then interacts with the carbonyl oxygens of residues Gln107, Tyr179, Glu180, Ala385 and Asp187 in an alternating fashion. Once it is completely within the pocket, the potassium ion interacts with backbone carbonyl group of mainly 2 residues, Glu180 and Gln107 (Figure 4.20). The distance between the potassium ion and carbonyl oxygen of these residues over 50 ns clearly indicates the coordination of the ion with these residues (Figure 4.19).

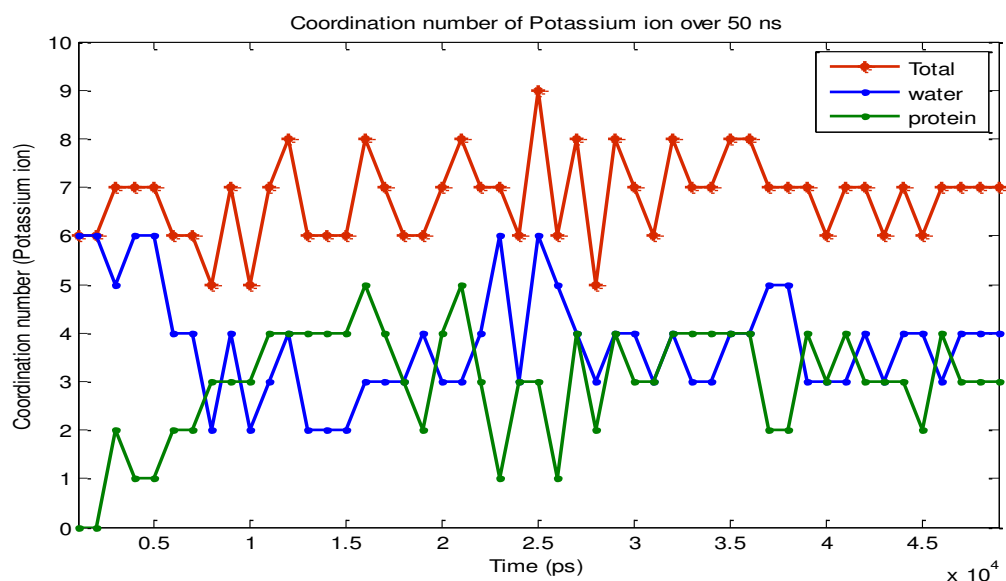


**Figure 4.19** Time vs Distance profile of potassium ion & backbone carbonyl oxygen of Glu180 (green) and Gln107 (blue).



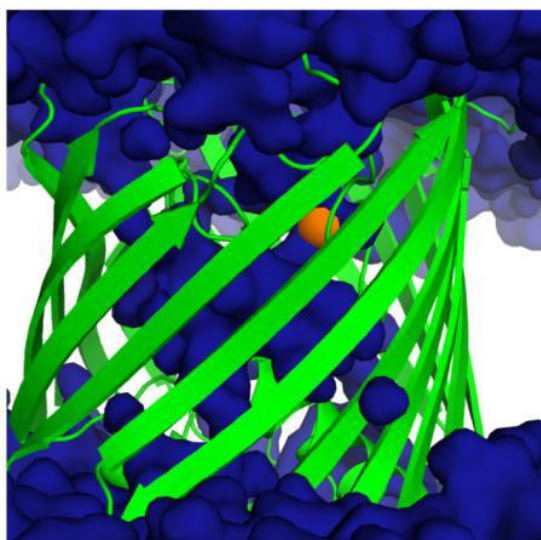
**Figure 4.20** Potassium ion (orange sphere) entering into Wzi protein (green coloured cartoon) is depicted. Movement of the ion from ~3ns (when it starts entering into the protein) to 6 ns (when it is completely within the pocket) is traced. Coordinating waters are shown in blue in stick representation. Details of the interactions are shown in boxes on the right. Note that water molecules within 3.6 Å of the ion alone are shown.

The coordination number of the potassium ion entering the channel is also quantified with respect to time (Figure 4.21), using a cutoff distance of 3.6 Å (Michael Thomas et al., 2007). Prior to entry into the pocket, the ion is surrounded by a shell of water molecules (Figure 4.20 (1)). In order to enter the narrow protein pocket, the ion is stripped of a few of the surrounding water molecules and it compensates for this dehydration energy by interaction with carbonyl oxygens of the protein. Once the ion enters into the pocket, a coordination number of at least 1 is observed with the protein backbone carbonyl groups.



**Figure 4.21** Number of coordination bonds formed by the potassium ion is plotted as a function of time over the entire simulation timescale. Coordination number with water (blue) decreases and that with the protein (green) increases as the ion enters into the protein. The total number of coordination bonds remains  $\sim 7-8$ . Note that a cutoff distance of  $3.6 \text{ \AA}$  is used.

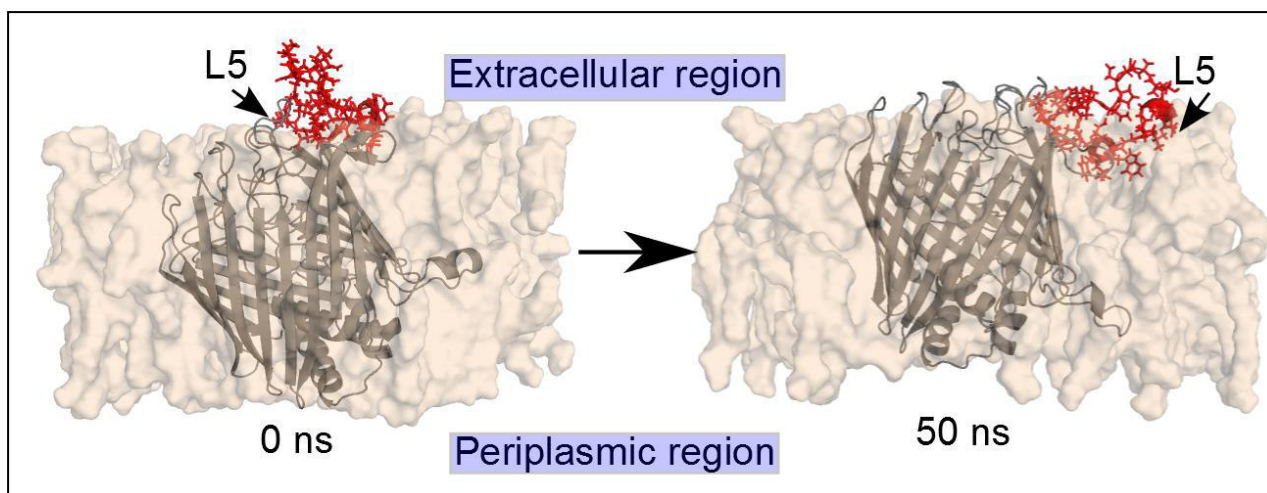
We observe that the entry of ion into the pocket is along the same path as the water molecules entering from the extracellular region (Figure 4.22) via entry point 1 (Figure 4.10), and the ion is surrounded closely by water molecules even inside the channel, which temporarily coordinate bond with it (see figure 4.20). Further movement of the ion towards the interior of the protein is blocked by the presence of an aromatic residue. The path of water is also similarly blocked and it follows a different path from that point. The ion, being larger in size, remains fixed within the pocket.



**Figure 4.22** The potassium ion (shown as orange sphere) lies in the path of water conduction (shown as blue surface) of Wzi (green coloured cartoon).

## 4.4 Importance of extracellular loop 5

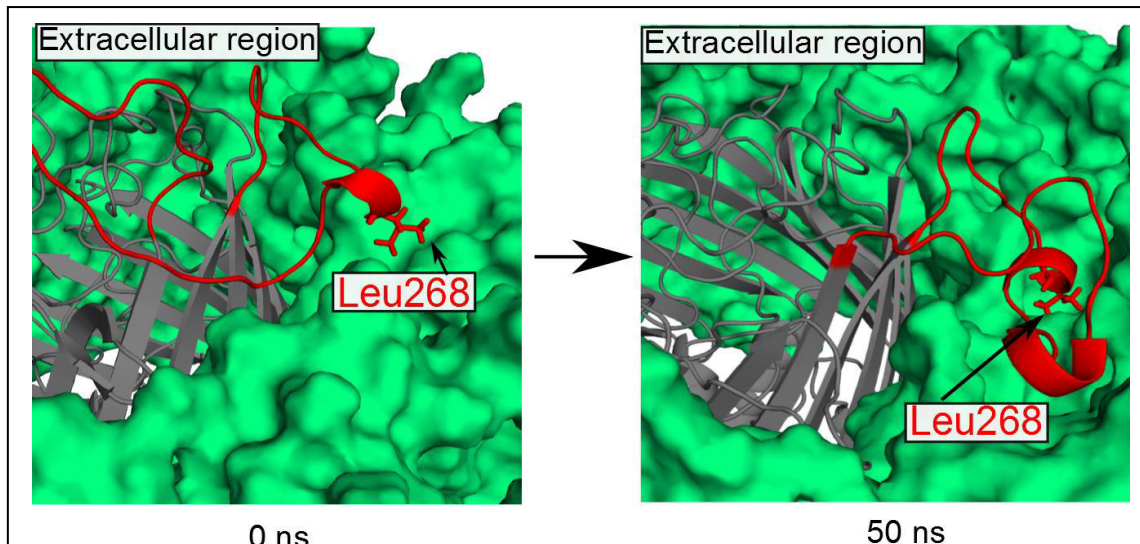
### 4.4.1 Membrane insertion of L5



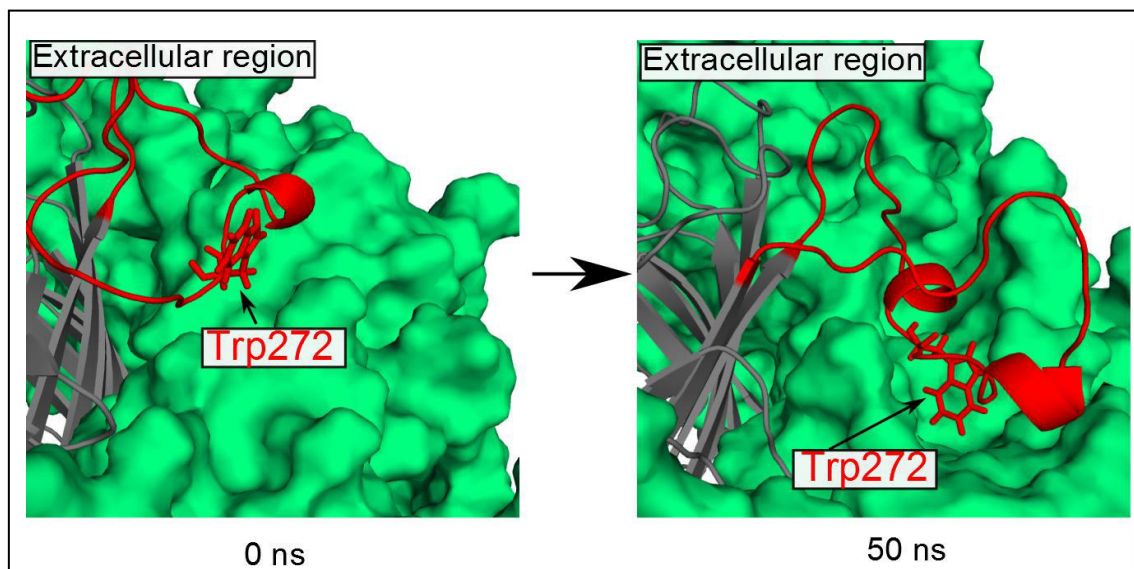
**Figure 4.23** Membrane insertion of L5. Loop L5 (shown as red coloured sticks) of Wzi protein (coloured grey) splays outwards from the barrel and becomes highly extended at the end of the 50 ns simulation trajectory.

As shown in Figures 4.23 to 4.27, extracellular loop L5 undergoes conformational changes during the course of simulation. It adopts an extended conformation and is involved in interactions with the lipid bilayer. In Figure 4.23, the splaying out of the loop L5 can be clearly observed. Figure 4.27 depicts the changes in loop conformation over the simulation trajectory, as viewed from the extracellular region. Membrane interactions of L5 occur at three separate points. Two of these interactions are hydrophobic, facilitated by residues Leu268 and Trp273. In both these cases, we observe that the hydrophobic residue which was initially above the membrane is getting inserted into it, as can be visualized from the snapshots at 0 ns and 50 ns (Figures 4.24 and 4.25).

A third important interaction of L5 with the membrane is observed, wherein two charged residues of L5, namely positively charged Lys278 and negatively charged Asp279 interact with the carbonyl and amino groups of POPE residues respectively via hydrogen bonding interactions (Figure 4.26). While the former is involved in the N-H ... O hydrogen bond, the latter is involved in strong electrostatic interaction with the positively charged amino head group of POPE.

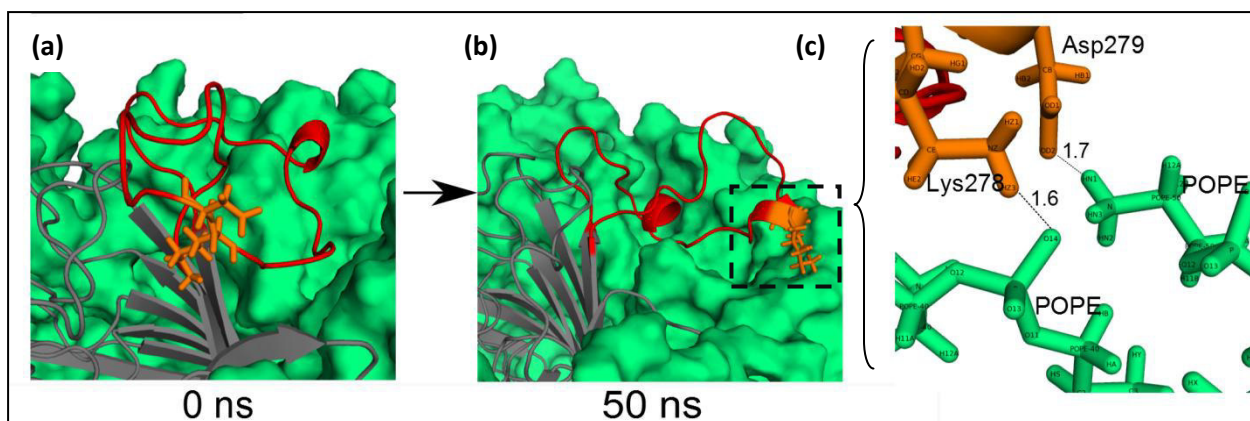


**Figure 4.24** Insertion of Leu268 of extracellular loop L5 (shown as red sticks) of Wzi protein (shown in grey as cartoon) into the lipid membrane (shown as green surface). Snapshot at 0ns is shown on the left and at 50 ns on the right.



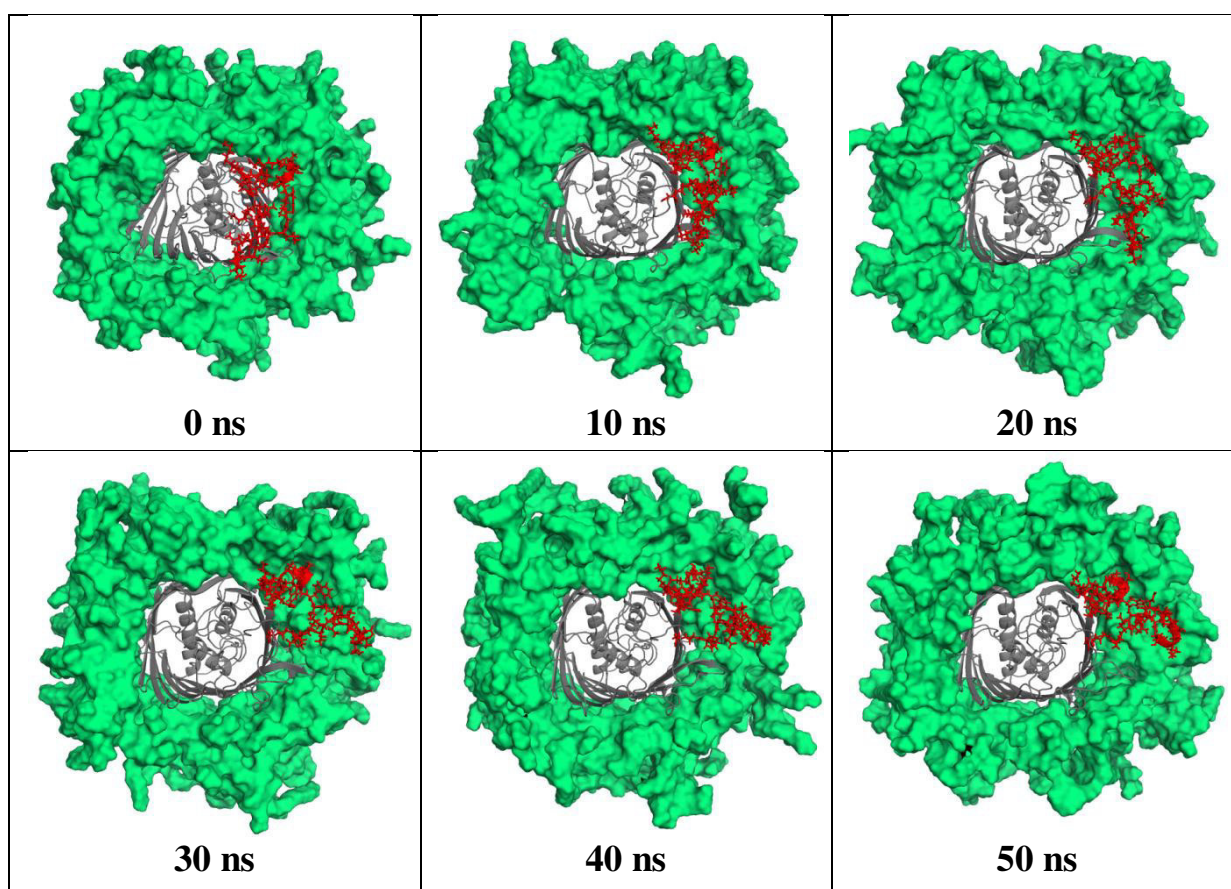
**Figure 4.25** Insertion of Trp272 of extracellular loop L5 (shown as red sticks) of Wzi protein (shown in grey as cartoon) into the lipid membrane (shown as green surface). Snapshot at 0ns is shown on the left and at 50 ns on the right.





**Figure 4.26** Interaction of charged residues, Lys278 and Asp279 (shown as orange sticks) of extracellular loop 5 (red) of Wzi protein (coloured grey) with the lipid bilayer (green surface) at (a) 0 ns, (b) 50ns. (c) Enlarged view of interaction of residues Lys278, Asp279 (orange sticks) with POPE residues (green sticks).

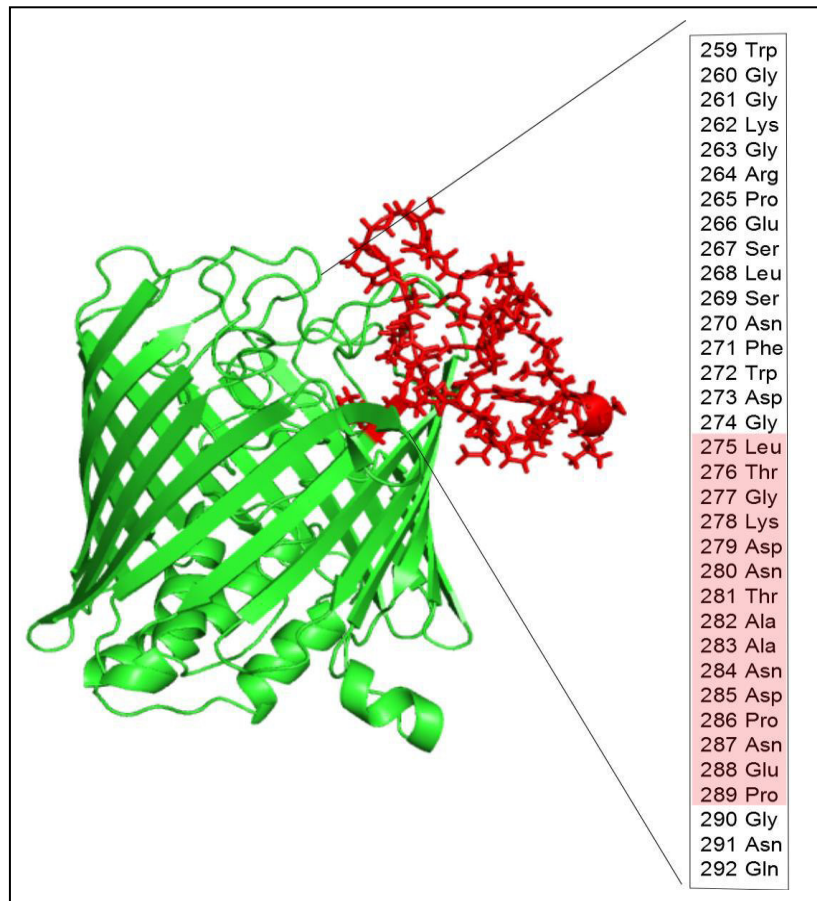
Snapshots at 10 ns intervals, depicting the interaction of L5 with the membrane, are shown in Figure 4.27. The loop splays away from the barrel and adopts an extended conformation at the end of 50 ns of simulations.



**Figure 4.27** Interaction of loop L5 with the lipid bilayer. Loop L5 is shown as red sticks, Wzi protein as grey cartoon and the lipid bilayer is depicted as green surface.

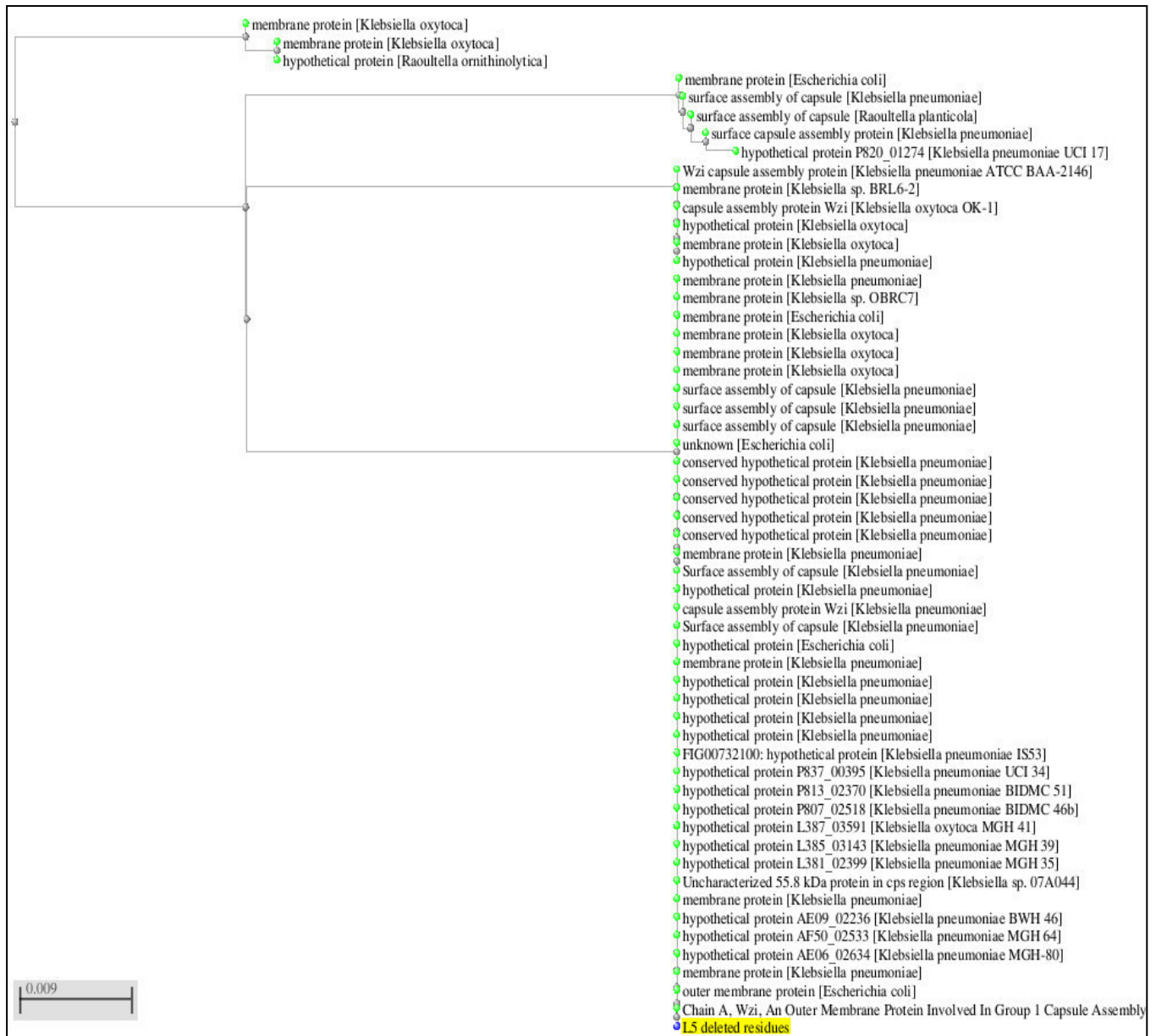
#### 4.4.2 Bioinformatics analysis

15 residues of the extracellular loop L5 of Wzi are absent from the crystal structure, namely, L-T-G-K-D-N-T-A-A-N-D-P-N-E-P (shown in Figure 4.28). To understand the significance of these residues, bioinformatics studies are carried out as detailed below.



**Figure 4.28** Wzi protein (cartoon) depicting the extracellular loop L5 (shown in red in stick representation). Sequence of L5 is in box (right), with the 15 deleted residues highlighted in pink.

Protein BLAST of the 15 residues is carried out with maximum sequence difference set at 10% and the results obtained are shown in tree format in Figure 4.29. “Fast minimum evolution” algorithm is used to generate the tree (Desper and Gascuel, 2004). The Grishin distance model is used, where evolutionary distance between two sequences is modelled as expected fraction of amino acid substitution per site given the fraction of mismatched amino acids in the aligned region (Grishin, 1995).

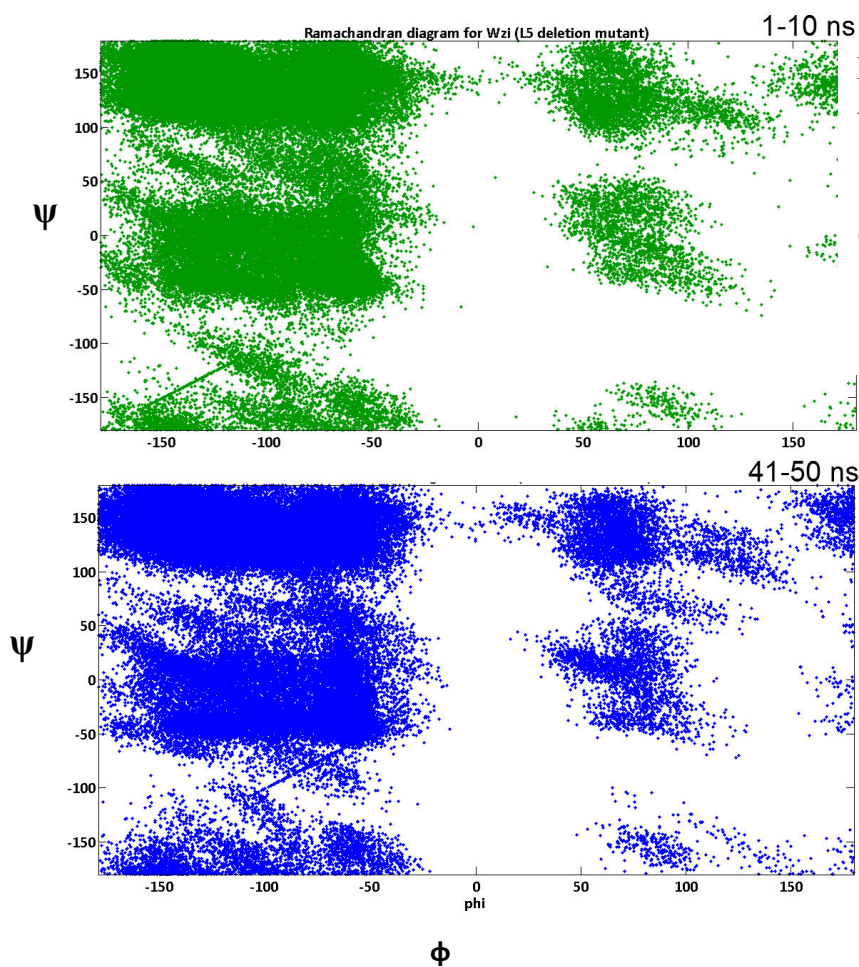


**Figure 4.29** Distance tree derived from the protein BLAST of deleted L5 residues. Maximum sequence difference is set to 10% for the BLAST to fetch the sequences that are 90% identical. Grishin distance model and Fast minimum evolution are used to generate the tree.

The 15 residues of L5 that are absent in the crystal structure possess sequence similarity mainly with bacterial membrane proteins. Most of the closely related hits belong to *E. coli* *Klebsiella pneumoniae* and *Klebsiella oxytoca* species. Hits are also observed from *Raoultella planticola* and *Raoultella ornithinolytica*. In order to understand the significance of these loop 5 residues, we carry out simulations, in which the 15 residues are deleted.

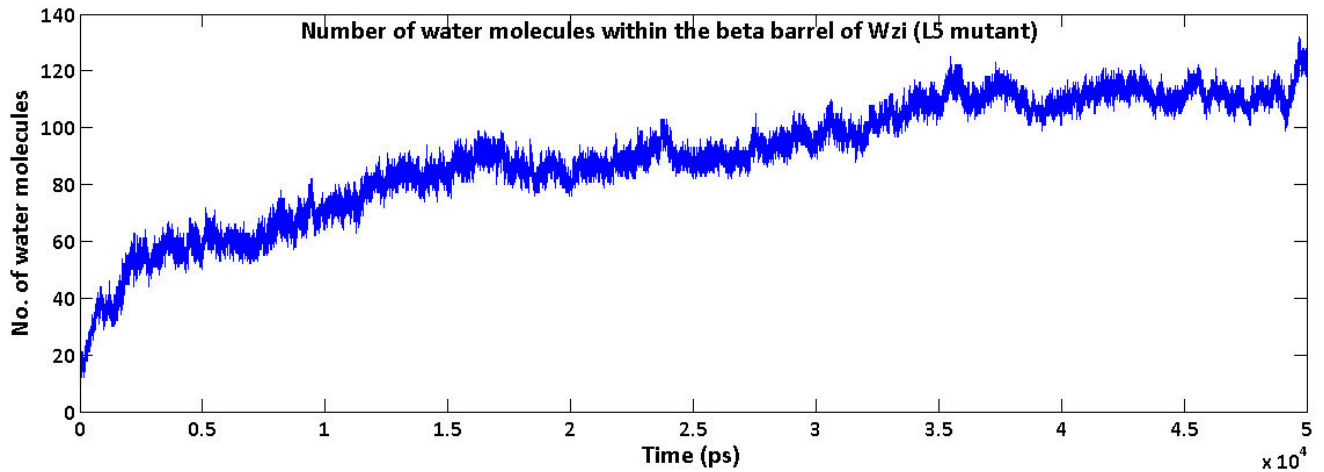
### 4.4.3 L5 deletion simulations

Ramachandran plots of the whole protein over the first 10 ns and last 10 ns of simulations (Figure 4.30) indicate that the values of  $\phi$  and  $\psi$  remain within the allowed regions. Water conduction is also not affected as reflected in the plot of number of water molecules within the barrel with respect to rime (Figure 4.31). The number of water molecules increases from 20 to nearly 120, similar to the wild type.

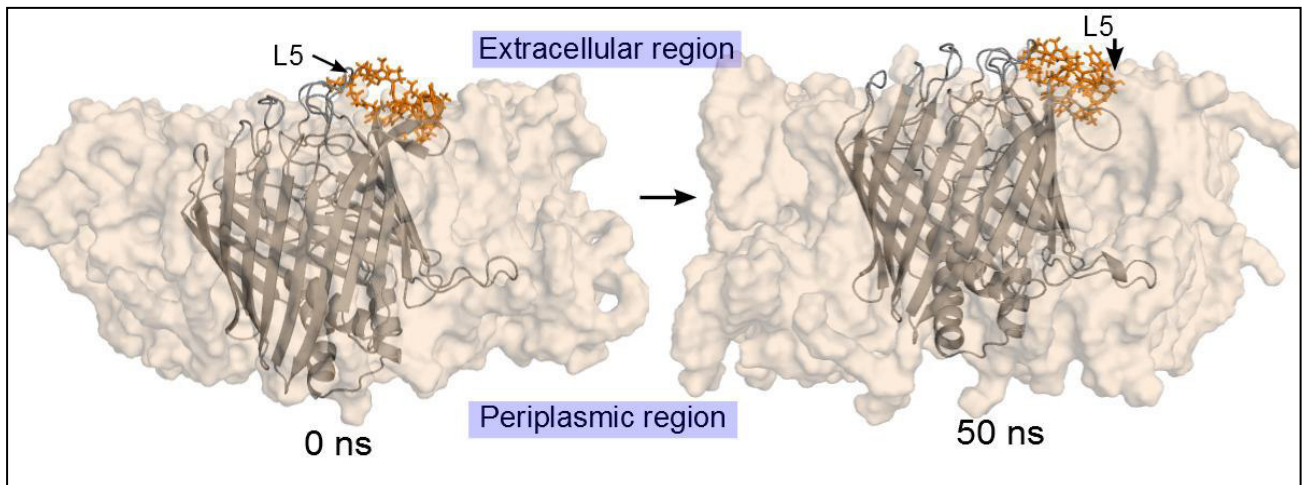


**Figure 4.30** Ramachandran diagram of Wzi loop deletion mutant over the first 10 ns (shown in green) and last 10 ns (shown in blue) of the simulations.

The major difference observed between the wild type noted and the L5 deletion mutant is the loss of interaction of L5 with the lipid membrane. As the loop is 15 residues shorter in this case, no significant membrane insertion is observed (Figure 4.32). Leu268 tends to move towards the membrane, but is limited by the loop size as depicted in Figure 4.33. Trp272, which inserts into the membrane in case of wild type Wzi does not exhibit any membrane interaction in this mutant. In addition, Lys278 and Asp279 are deleted in this mutant, resulting in the loss of interactions with the lipid head groups seen in case of wild type Wzi.

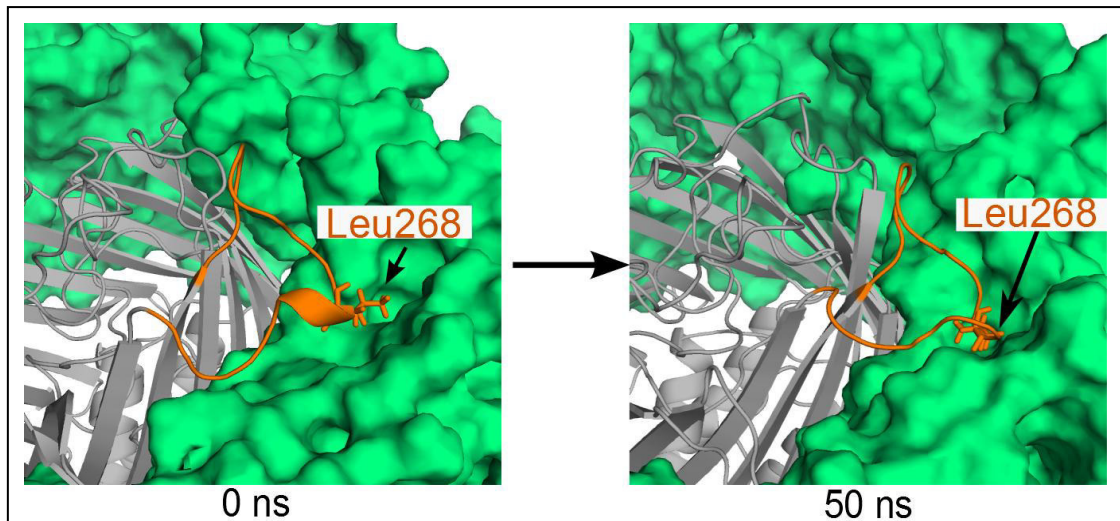


**Figure 4.31** Number of water molecules within the beta barrel plotted as a function of time. Calculations are the same as in Figure 4.5.



**Figure 4.32** Loop L5 (orange sticks) of Wzi protein (grey cartoon) does not undergo any major structural change over 50 ns simulations. Lipid bilayer is shown in surface representation.

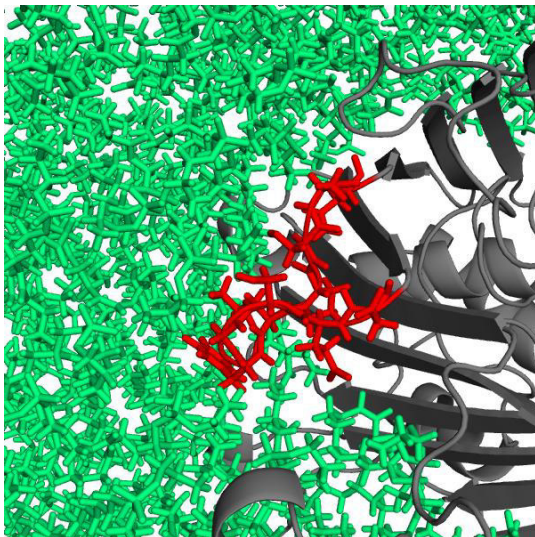
Thus, the major conclusion from our study of extracellular loop L5 is that it does not appear to have any role in water conduction, but plays an important role in membrane interactions and in anchoring the protein into the lipid bilayer.



**Figure 4.33** Lack of Leu 268-POPE interaction in Wzi L5 deletion mutant. Leu268 is shown as orange sticks and the lipid bilayer as green surface. Note that there is no loop splaying as observed in case of wild type L5.

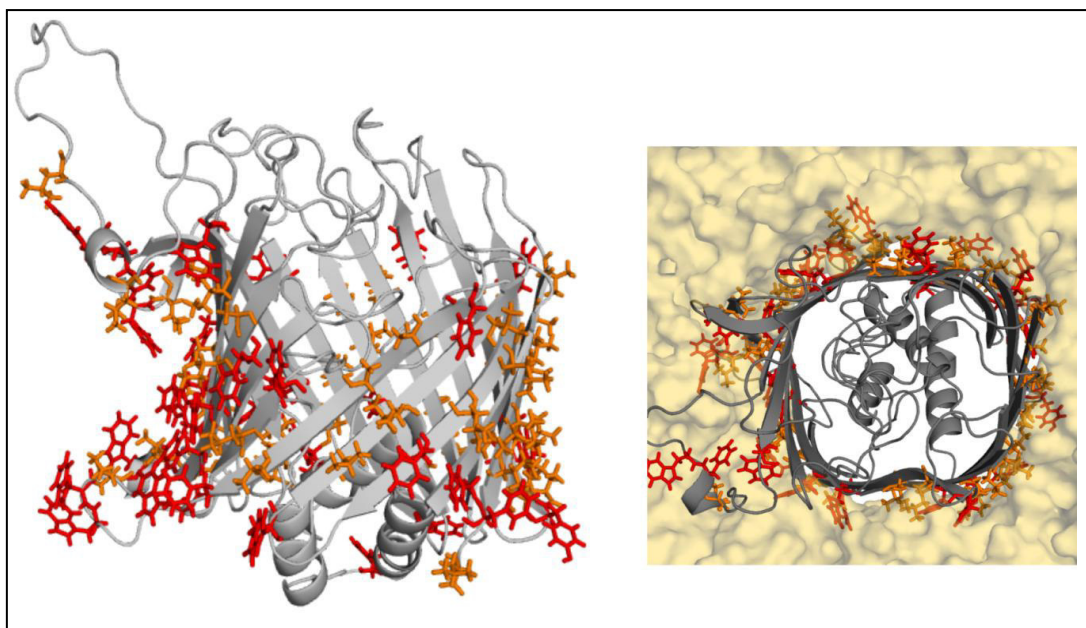
#### 4.5 Other membrane protein interactions of Wzi

In addition to L5, other extracellular loops are also found to interact with the beta barrel. Loop L6 is a short loop consisting of 9 residues of which Phe324 is involved in hydrophobic interactions with the lipid bilayer as shown in Figure 4.34.



**Figure 4.34** Hydrophobic interaction of extracellular loop 6 (red sticks) of Wzi protein (coloured grey) with the lipid bilayer (green sticks).

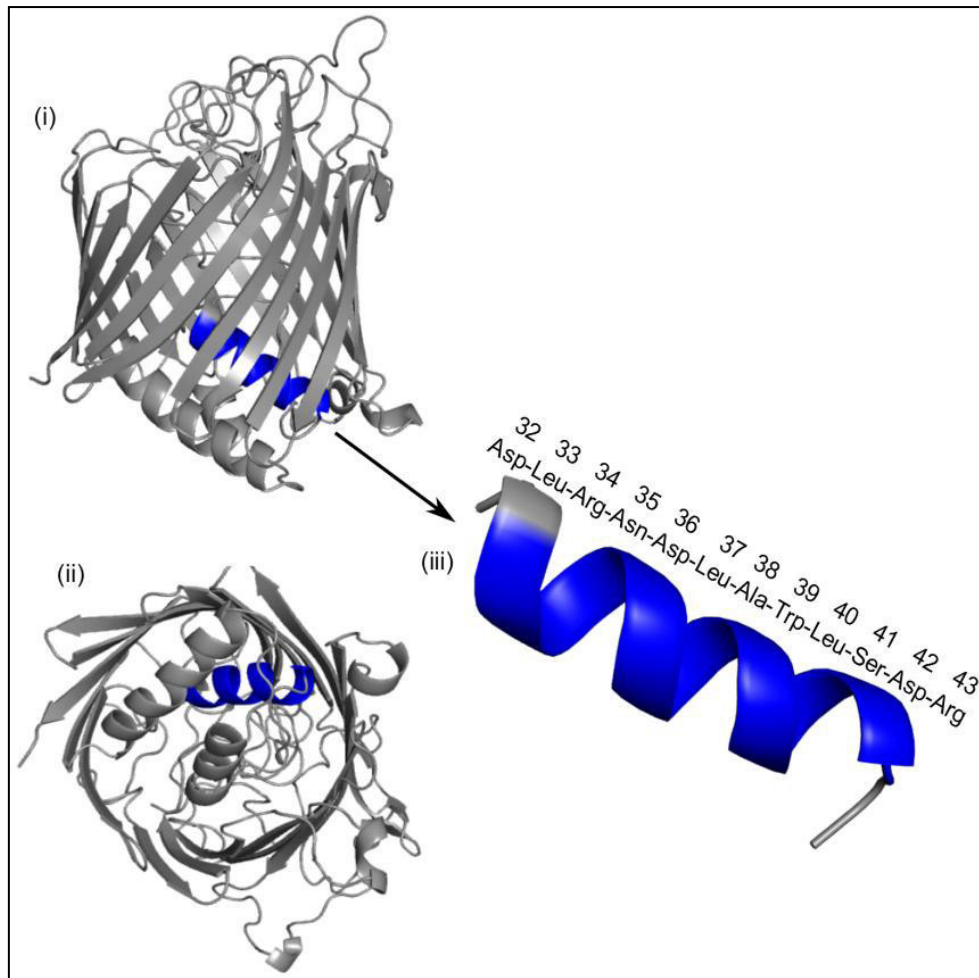
Apart from the lipid bilayer interactions of loops L5 and L6, several hydrophobic residues, both aromatic (Phe, Tyr and Trp) and non-aromatic (Ala, Val, Leu and Ile) project outwards from the protein and interact hydrophobically with the lipid bilayer as shown in Figure 4.35.



**Figure 4.35** (Left) Hydrophobic residues of Wzi (Aromatic residues, Phe, Tyr and Trp shown in red and non-aromatic residues, Ala, Val, Leu, Ile shown in orange) that project outwards from the barrel can be seen. (Right) View of Wzi from extracellular side with hydrophobic residues projecting towards the membrane. Wzi is shown as grey coloured cartoon and the lipid membrane is shown as yellow coloured surface.

#### 4.6 Helix mutants of Wzi

Previous studies have shown that the substitution mutations of the middle five residues (35 to 39) and last 5 residues (39 to 43) residues of helix H1 to glycine have a negative effect on its binding activity to K30, though it is not directly involved in the anchorage of K30 CPS. However, substitution mutation of the first five residues (32 to 36) of H1 to glycine does not affect the K30 binding activity. The reason for the negative effect of the mutations was unknown (Simon R Bushell et al., 2013). In this context, we carried out MD simulations of the 3 helix mutants (with residues 32-36 mutated to Gly, residues 35-39 mutated to Gly and residues 39-43 mutated to Gly) of Wzi. Figure 4.36 depicts the Wzi protein, with helix H1 highlighted in blue. The sequence of H1 is also provided.

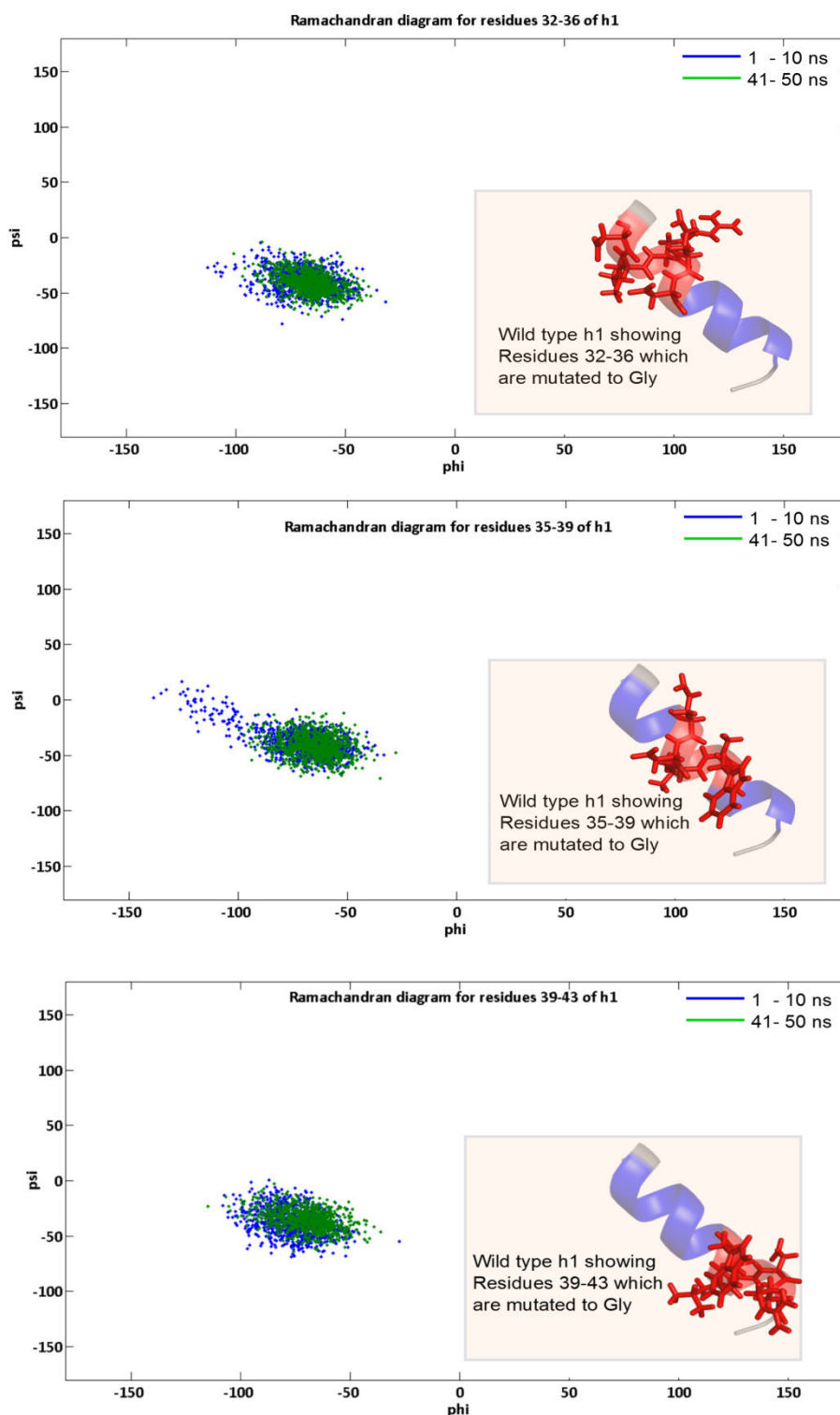


**Figure 4.36** (i) Cartoon representation of Wzi (grey) with helix H1 shown in blue (i) side view, (ii) view from periplasmic side. (iii) Enlarged view of H1 with the sequence of residues 32-43 marked.

#### 4.6.1 Secondary structure

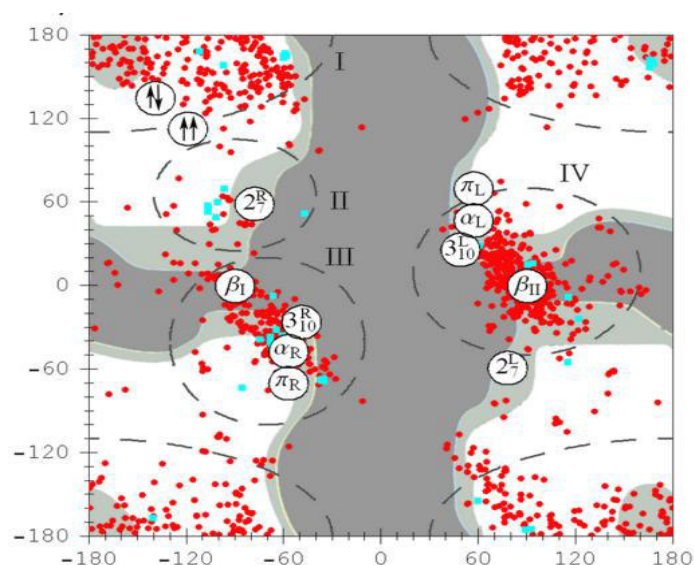
Dihedral angles ( $\psi$ ,  $\phi$ ) of the mutated residues of H1 are calculated for the first and last 10 ns simulations of mutant 1 (Figure 4.37 (top)), mutant 2 (Figure 4.37 (middle)) and mutant 3 (Figure 4.37 (bottom)). It is observed from the Ramachandran plot that no major secondary structural deviations occur. The conformational preference ( $\phi, \psi$ ) of glycine residues in all three mutants stays similar to  $\alpha$  helical conformation ( $\phi = -57^\circ$ ,  $\psi = -47^\circ$ ).





**Figure 4.37 Ramachandran diagram for residues mutated to glycine.**  $\psi$  and  $\phi$  of residues 32 to 36 (top), 35 to 39 (middle) and 39 to 43 (bottom) of helix H1 in mutants 1, 2 and 3 respectively over the first 10 ns (blue colour) and last 10 ns (green colour) of the simulations. Inset: H1  $\alpha$ -helix of wild type Wzi with the residues mutated to glycine shown in red.

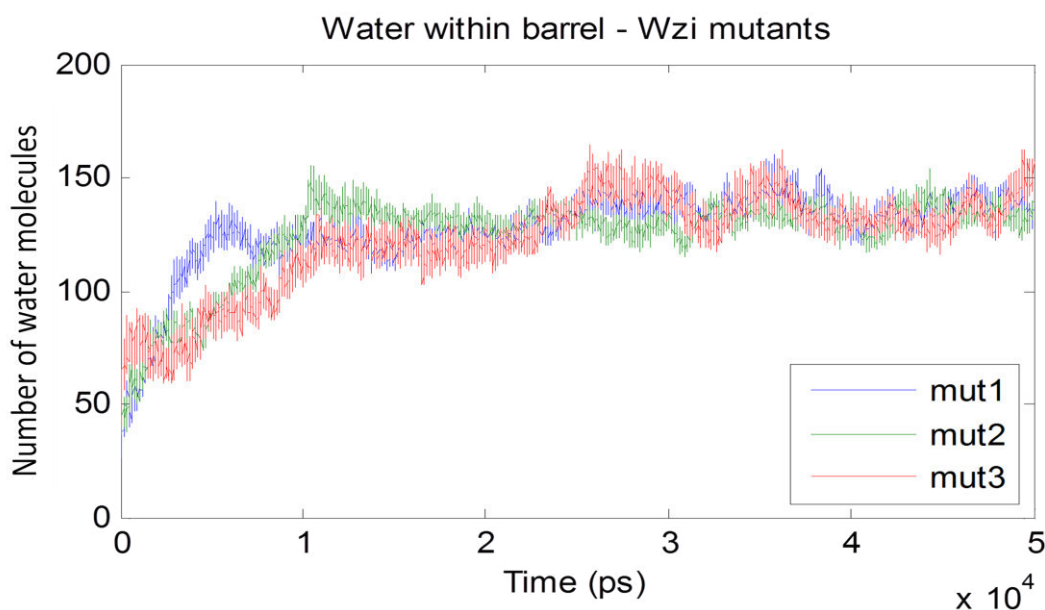
This observation is supported by previous studies where the different stable conformations of Gly tripeptides and hexapeptides were investigated and found to exist in alpha helical conformation along with other secondary structural conformations (Yakubovich, A.V et al., 2006). Figure 4.38 shows a comparison of the dihedral angles of glycine residues from PDB with the steric diagram for polyglycine which was determined experimentally. As can be seen in the plot (Figure 4.38), some population of glycine residues fall in the range of dihedral angles corresponding to alpha-helical conformation.



**Figure 4.38** Comparison of  $\psi$  and  $\phi$  angles of polyglycine residues in protein structures from Brookhaven Protein Data Bank (red) superimposed with the theoretically calculated allowed conformations of a glycine hexamer (encircled). Four minima in the steric diagram have been encircled and labelled I, II, II and IV. Adapted from Yakubovich, A.V et al., 2006.

#### 4.6.2 Water conduction

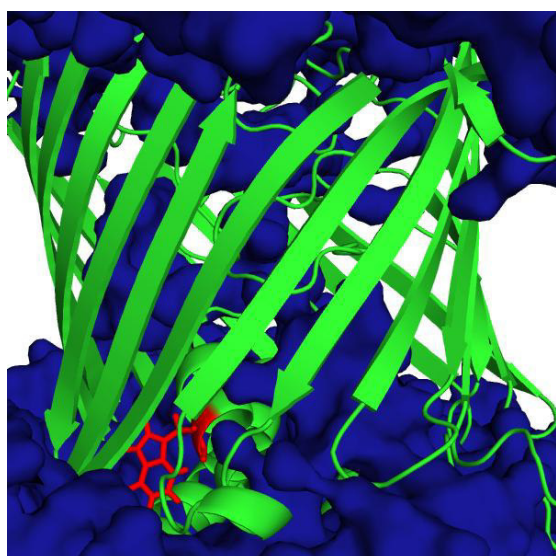
Water conduction through Wzi is observed in all the three helix mutants. The number of water molecules within the beta barrel is plotted as a function of time for each mutant (Figure 4.39). The total number of water molecules does not vary significantly from the wild type protein. This observation is not surprising, as there is no major structural change in the beta barrel region of the protein and consequently, the number of water molecules accommodated within the barrel cannot vary significantly.



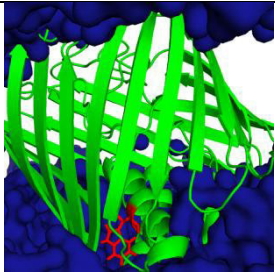
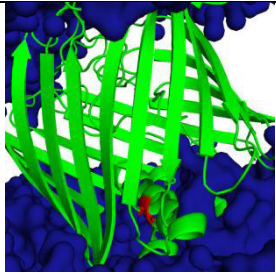
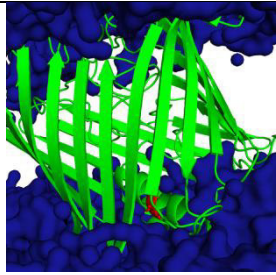
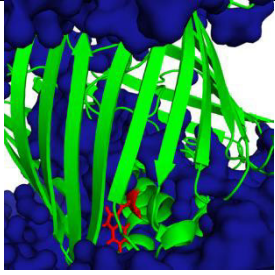
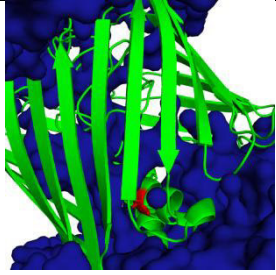
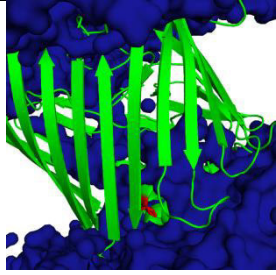
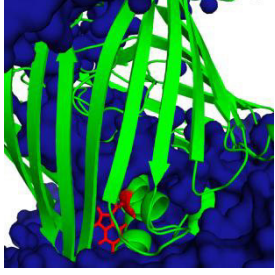
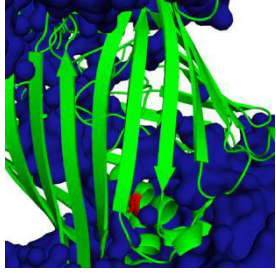
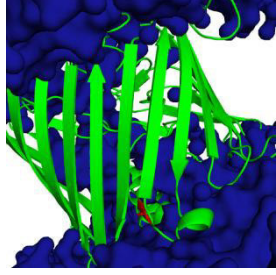
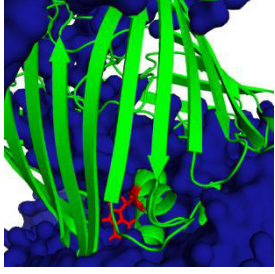
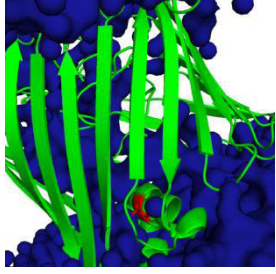
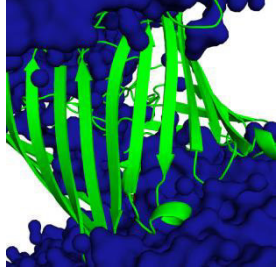
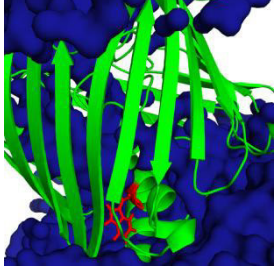
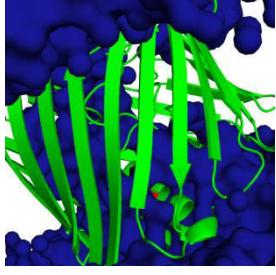
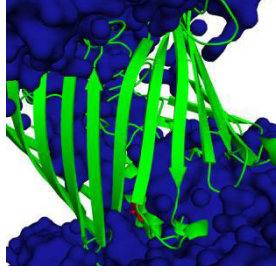
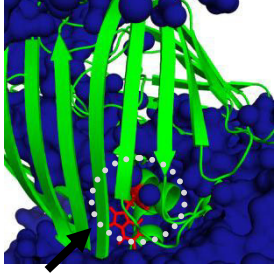
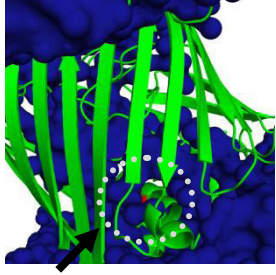
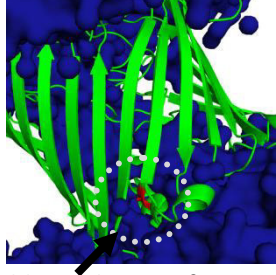
**Figure 4.39** Number of water molecules within the Wzi beta-barrel in case of mutant 1 (blue), mutant 2 (green) and mutant 3 (red). The total number of water molecules is nearly the same in all the three cases at the end of the 50 ns trajectory.

#### 4.6.3 Role of Trp39 in water regulation at the periplasmic end

Yet another intriguing revelation from the simulation of Wzi is the role of Trp39 in regulating entry of water into the channel at the periplasmic end. As mentioned in section 4.2.4, Trp39, which is part of helix H1, blocks entry of water into the channel (Figure 4.40). Notably, in helix mutants 2 and 3, residue Trp39 is mutated to glycine. Glycine, being the smallest amino acid, does not block water molecules like Trp. Consequently, in these cases water conduction occurs through an additional pathway as illustrated by the snapshots of all three mutants at 10ns intervals in Figure 4.41.



**Figure 4.40 Role of Trp-39 (snapshot of wild type Wzi at 10 ns).** Tryptophan-39 (shown in red in stick representation) of helix h1 of Wzi protein (shown in green in cartoon representation) blocks the path of water entering the channel.

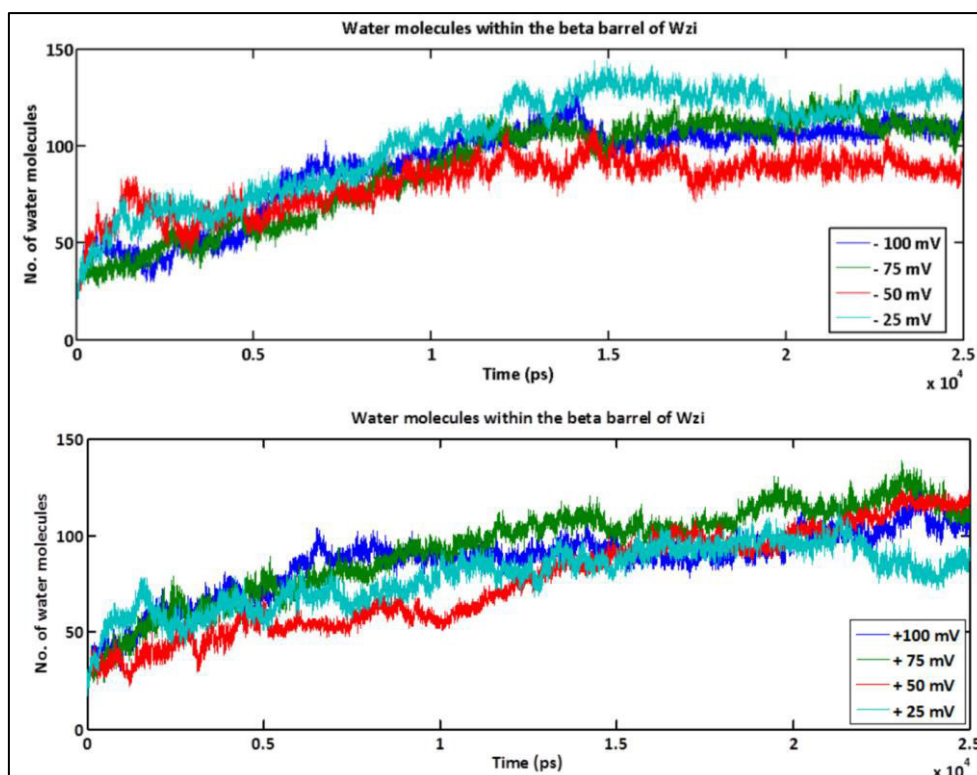
Time	Mutant 1	Mutant 2	Mutant 3
0ns			
10ns			
20ns			
30ns			
40ns			
50ns			
	Water blocked by Trp39	Additional point of entry	Additional point of entry

**Figure 4.41** Snapshots of water (shown in blue in surface representation) entering the protein (shown in green as cartoon) at 10 ns intervals for mutants 1, 2 and 3. Water enters into Wzi from the periplasmic side through an additional route in mutants 2 and 3, while the route is blocked by Tarp39 in mutant 1.

Thus, we hypothesize that the change in the water conduction pattern in case of mutants 2 and 3, due to the absence of Trp39 residue may be the cause for loss of protein function. This argument further gains support from the experimental data which demonstrates the loss of function of glycine substitution mutants 2 and 3 of helix H1, even though they are not directly interacting with the CPS. As the current simulation indicates that Wzi can act as an osmotic pressure regulator, the increased influx of water from the periplasmic side might dilute CPS concentration on the surface and thus reduce its binding to Wzi protein.

#### 4.7 Voltage independent water conduction of Wzi

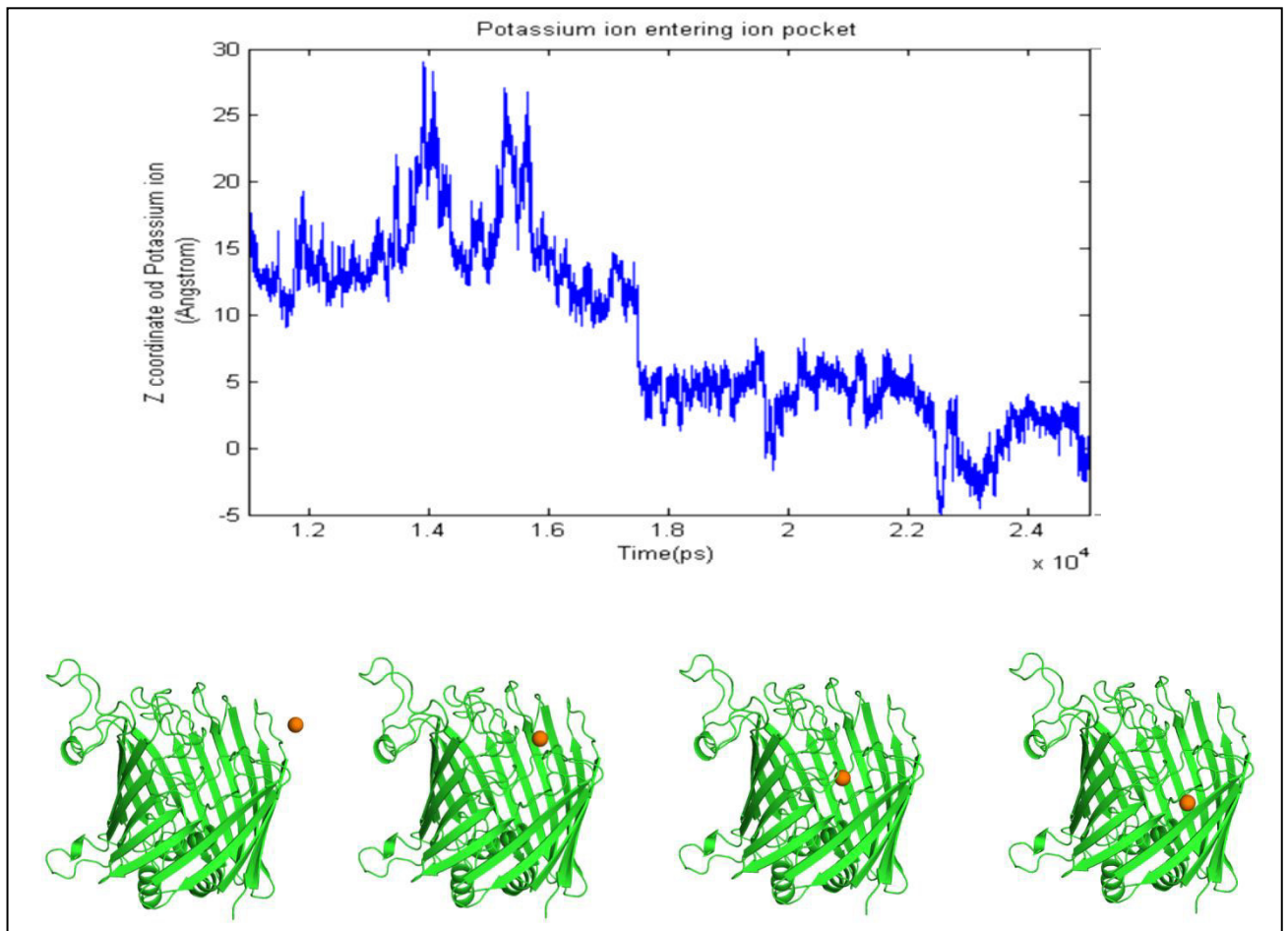
As detailed in Chapter 2, we apply eight different transmembrane voltages, namely -100 mV, -75 mV, -50 mV, -25 mV, +25 mV, +50 mV, +75 mV, +100 mV to the Wzi wild type system, to investigate voltage dependent properties of the protein, if any. Note that the wild type system is simulated at 0 mV for 100 ps prior to application of voltage. The eight systems are each simulated over a 25 ns trajectory.



**Figure 4.42** Number of water molecules present within the beta barrel of Wzi with applied negative voltages (top), -100 mV (blue), -75 mV (green), -50 mV (red), -25 mV (cyan) and applied positive voltages (bottom), +100 mV (blue), +75 mV (green), +50 mV (red), +25 mV (cyan), each simulated over a 25 ns trajectory. Calculations same as in Figure 4.5.

No major structural changes are observed on application of transmembrane voltage. Water conduction is observed in all the systems irrespective of the voltage applied. Figure 4.41 depicts the gradual increase in the number of water molecules present within the beta barrel of Wzi for the applied negative voltages (Figure 4.41 top) and positive voltages (Figure 4.41 bottom). In all eight systems simulated upto 25 ns, the number of waters remains nearly the same.

An ion pocket is observed in two of the simulated systems, namely, the systems with an applied voltage of -25 mV and +75 mV. A single potassium ion is observed to enter Wzi from the extracellular side and coordinate with the carbonyl oxygen of the protein (Figure 4.43) as in the wild type. Note that the ion enters deeper into the pocket in this case than in wild type.



**Figure 4.43 Potassium ion binding to Wzi with the applied transmembrane voltage of -25mV.** (Top), Z coordinate of the potassium entering into the ion binding pocket of Wzi. (Bottom), cartoon representation of Wzi (green) depicting the entry of the potassium ion (orange spheres) over 11 to 25 ns of the simulation timescale.

As the ion entry is observed irrespective of the applied transmembrane voltage (0Mv, -25mV and +75mV), similar binding cannot be ruled out in all the other applied voltage systems. Longer simulation timescales and steered MD studies are required to understand this property better. From the transmembrane voltage simulations, we conclude that Wzi does not exhibit any voltage dependency.

## 4.8 Conclusions

The first major observation in our simulations of Wzi protein is its water conduction property. Conduction of water molecules through Wzi protein occurs irrespective of the loop 5 deletion, helix 1 substitution mutations and application of transmembrane voltage. It thus appears to be a fundamental property of the Wzi beta barrel protein. Although the role of Wzi in the surface anchorage of K30 CPS is well characterized, this is the first study that offers evidence for its role in water conduction. Our investigation also identifies the role of several amino acids that are inside the beta-barrel in regulating water conduction through a specific path. The first among them is a hydrophobic plug within the protein channel that restrains the water molecules to a narrow conduction path. Further, Tyr380 is identified as an important residue aiding in water conduction, with its mutant, Y380W, showing a marked decrease in water conduction. Another key residue is Trp39 of the periplasmic helix, H1. When H1 is mutated to glycine, as in H1 substitution mutants 2 and 3, the water conductance path gets altered with the creation of an additional water entry point at the periplasmic face. We cite this change in water conductance as the possible reason for loss of protein function in mutants 2 and 3 as observed from previous studies, i.e., the increased influx of water from the periplasmic side may dilute the K30 CPS concentration and reduce its binding efficiency to Wzi on the extracellular side. We hypothesize that the highly hydrated Wzi channel provides a suitable environment for the insertion and anchorage of the K30 CPS polymer, conferring high biological relevance to this observation. We also suggest that the concentration gradient created by accumulation of CPS at the cell exterior may drive the transport of water from the periplasm to the extracellular space.

Our second observation is the presence of an ion binding pocket on the extracellular face of Wzi protein. A single potassium ion enters into the pocket and remains stably associated with it throughout the simulation timescale. We hypothesize that this positive ion may act as a crosslinker for the negatively charge CPS molecule and promote its non-covalent attachment to Wzi. We also carry out simulations of wild type Wzi at a range of transmembrane voltages.

However, no major structural deviations or changes in water conduction patterns are observed. Thus, the protein acts as an osmotic pressure regulator, irrespective of the applied transmembrane voltages.

The importance of extracellular loop L5 of Wzi in insertion into the lipid bilayer is also investigated in detail. 15 residues of L5, which are absent from the crystal structure, are found to be conserved across several *Klebsiella* species as well. Wild type Wzi, with the 15 missing residues modelled, shows L5 interacting with the protein at multiple point in an extended conformation and inserting into the membrane through hydrophobic, electrostatic and hydrogen bonding interactions. Though L5 plays an important role in anchoring Wzi into the membrane, it is not involved in water conduction as the Wzi mutant with 15 residues deleted from L5, does not exhibit any changes in water conduction. From the results obtained, we conclude that the major role of L5 is in facilitating protein-bilayer interactions.

Thus, the current investigation successfully characterized the osmotic pressure regulating mechanism of Wzi in detail.

#### 4.9 References

- Bushell, S.R., Mainprize, I.L., Wear, M.A., Lou, H., Whitfield, C., Naismith, J.H., 2013. Wzi is an outer membrane lectin that underpins group 1 capsule assembly in *Escherichia coli*. *Structure* 21, 844–853. doi:10.1016/j.str.2013.03.010
- Desper, R., Gascuel, O., 2004. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Mol. Biol. Evol.* 21, 587–598. doi:10.1093/molbev/msh049
- Grishin, N.V., 1995. Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Mol. Evol.* 41, 675–679.
- Michael Thomas, Dylan Jayatilaka, Ben Corry, 2007. The Predominant Role of Coordination Number in Potassium Channel Selectivity. *Biophys. J.* 93, 2635–2643.
- Rahn, A., Beis, K., Naismith, J.H., Whitfield, C., 2003. A novel outer membrane protein, Wzi, is involved in surface assembly of the *Escherichia coli* K30 group 1 capsule. *J. Bacteriol.* 185, 5882–5890.
- Yakubovich, A.V, Greiner, W., I. A. Solov'yov, A. V. Solov'yov, 2006. Conformational changes in glycine tri- and hexapeptide. *Eur. Phys. J. - At. Mol. Opt. Plasma Phys.* 39, 23–34.