

# **Structural Elucidation of Human Cystine/Glutamate Transporter: xCT**

**Anirudh C.R**

A dissertation submitted for the partial fulfilment of the BS-MS dual  
degree in Science.



**Indian Institute of Science Education and Research Mohali  
April 2017**



## **Certificate of Examination**

This is to certify that the dissertation titled “**Structural Elucidation of Human Cystine/Glutamate Transporter: xCT**” submitted by **Mr. Anirudh C.R** (Reg. No: MS12032) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee find the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Sabyasachi Rakshit

Dr. Shashi B. Pandit

Dr. Monika Sharma

(Supervisor)



## **Declaration**

The work presented in this dissertation has been carried out by me under the guidance of Dr. Monika Sharma at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Anirudh C R

(Candidate)

Dated: April 21, 2017

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr.Monika Sharma

(Supervisor)



## **Acknowledgements**

Firstly, I would like to convey my sincere gratitude to Dr. Monika Sharma for giving me an opportunity to complete my masters, for all the open discussions and suggestions that were very crucial and for providing me freedom to explore and understand various methods over the course of my project. I would like to thank my committee members, Dr. Sabyasachi Rakshit and Dr. Shashi Bhushan Pandit for engaging discussions and suggestions during presentations, which helped me over the span of my thesis work.

I would like to express my deepest thanks to my close friends who made these five long years very much exciting and worthwhile. I would also like to thank my lab (AB-4B) members for helping me with technical problems and creating a relaxed, friendly lab environment that made my short stay very pleasant.

I would like to express my sincere thanks to IISER Mohali for all the facilities and for the opportunity to pursue and complete my undergraduate study.

Finally, I would like to thank my parents and close relatives for their continuing support and motivation, which helped me to reach my goals.



## List of Figures

Fig 1.1.1: A schematic diagram for the classification of membrane transporters based on their mode of transport.....	2
Fig 1.1.2: A depiction of the different classes of Carrier Transporters as explained in the above paragraph.....	3
Fig 1.1.3: Na <sup>+</sup> /K <sup>+</sup> pump shown above is an example of primary active transport.....	3
Fig 1.1.4: Facilitated diffusion in cell membrane depicting ion channels (left) and carrier proteins (three of them) on the right.....	4
Fig 1.2.1: A schematic flowchart of the Sx <sub>c</sub> <sup>-</sup> .....	5
Fig 1.2.2: Glutathione (GSH) Metabolism.....	6
Fig 1.2.3: The overall structure of system xc <sup>-</sup> .....	7
Fig 1.3.1: The Alternate access mechanism cycle.....	9
Fig 2.1.1 : A schematic flowchart of the steps involved in comparative modeling employed in MODELLER.....	14
Fig 2.2.1 : An outline of the steps which are involved in membrane building by the Membrane Builder module in CHARMM-GUI.....	17
Fig 2.4.1: A diagram depicting the intermolecular and intramolecular interactions between protein and ligand.....	24
Fig 3.1.1: The hits from SWISS-MODEL (Table 3.1.1).....	35
Fig 3.1.2:(a); DOPE profile (b); Structural Alignment of xCT (target) and 3gi9C (template) showing xCT colored in blue and 3gi9C colored in red.....	36
Fig 3.1.3: (a) DOPE profile for Ci using template as 4djiA.pdb (b) Structural Alignment of xCT (target) and 4djiA (template) showing xCT colored blue and 4djiA colored red.....	37
Fig 3.1.4: (a) DOPE profile for Co using 3ncyA.pdb as a template (b) Structural alignment of xCT (target) and 3ncyA (template) showing xCT in blue color and 3ncyA in red color.....	37
Fig 3.1.5 : The overall view of Cio modelled structure.....	38
Fig 3.1.6: (a) Intracellular View ; (b) Extracellular View.....	38
Fig 3.1.7 : The overall view of Ci modelled structure.....	39

Fig. 3.1.8: (a) Intracellular View ; (b) Extracellular View.....	39
Fig 3.1.9 : The overall view of Ci modelled structure.....	40
Fig 3.1.10: (a) Intracellular View;(b) Extracellular View.....	40
Fig 3.1.11: This an example of pore lining in Cio structure.....	41
Fig 3.2.1: The assembled xCT (Cio) membrane /protein complex.....	42
Fig 3.2.2: The assembled membrane /protein complex for xCT in the Ci conformation.....	43
Fig 3.2.3, 3.2.4 : The structural alignment of xCT (after membrane building) as shown in blue animation and xCT (after MD run).....	44
Fig 3.3.1: Depicts the clustering analysis done for xCT (in Ci conformation) and L-Cys <sub>2</sub> ....	45
Fig. 3.3.2: Binding Pocket of the Ligand in Xct (Ci conformation).....	46
Fig 3.3.3: The clustering analysis done for xCT (in Co conformation) and L-Cys <sub>2</sub> <sup>-</sup> .....	47
Fig. 3.3.4: Binding Pocket of the Ligand in xCT (Co conformation).....	48
Fig. 3.3.5: L-Cys <sub>2</sub> and xCT interactions for Ci conformation.....	49
(Fig. 3.3.6): L-Cys <sub>2</sub> and xCT interactions for Co conformation.....	50
Fig 3.4.1: The alignment of Ci with Co before TMD run.....	51
Fig. 3.4.2: The RMSD between the initial Ci and final Co structure over 3 ns.....	52
Fig 3.4.3: A structural alignment between Ci and Co conformations after the TMD run .....	53

## List of Tables

TABLE 2.3.1 : This is a table of restraints applied on the proteins, lipid molecules in the bilayer during the simulations to assure a gradual relaxation of the system from its initial uncorrelated high energy state.....	30
TABLE 3.1.1.....	42



## Abbreviations

L-Cys<sub>2</sub>: L-Cystine

L-Glu: L-Glutamate

ATP: Adenosine triphosphate

ADP: Adenosine diphosphate

P<sub>i</sub> : Phosphate

SLC: Solute carrier

DMPC: 1,2-Dimyristoly-*sn*-glycero-3-phosphocholine

SGLT: Sodium-glucose linked transporter

BLAST: Basic local alignment search tool

FASTA: Fast alignment

PDB: Protein data bank

VMD: Visual molecular dynamics

PMV: Python molecule viewer

RMS: Root mean square

RMSD: Root mean square deviation



# Contents

<b>List of Figures</b>	<b>i</b>
<b>List of Tables</b>	<b>ii</b>
<b>List of Photographs</b>	<b>iii</b>
<b>Notation</b>	<b>iv</b>
<b>Abstract</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Methodology</b>	<b>12</b>
<b>3 Results and conclusions</b>	<b>33</b>



## **Abstract**

The human cystine/glutamate antiporter xCT is a membrane protein transporter belonging to the family of Heteromeric Amino acid Transporters (HAT), which regulates the influx of L-Cystine (L-Cys<sub>2</sub>) and efflux of L-glutamate (L-Glu). xCT has been linked to several central nervous system functions and protection of the cell from oxidative stress. Delineating its structure would aid in understanding structure-function relationships, thereby help formulate therapeutic targets for diseases, build our understanding of the residues that aid in binding of the substrates and those that are critical for the permeation of the substrates across the membrane to give us an entire picture of the mechanism of an antiporter.



# CHAPTER 1

## INTRODUCTION

### 1.1 An Introduction to Membrane Transport Proteins

Biological organisms consist of cells built of essential macromolecules like proteins, carbohydrates, lipids and nucleic acids. Proteins are tagged as “work horses” of cell and virtually all the catalytic activities are performed by proteins. About one-third of the proteins present in a cell are embedded in the cellular membrane, and about one-third of these are critical for catalyzing the transport of molecules or ions across the plasma membrane<sup>1</sup>. A handful of molecules enter or leave cells, or cross cellular membranes, without the aid of proteins. Even the transport of molecules like water or urea, which can diffuse across pure phospholipid bilayer are usually accelerated by transport proteins.

Transport proteins are integral membrane proteins, i.e they exist permanently within the span of the membrane within and span the membrane across which they transport substances. Depending on their mode of transport or the way they assist the movement of substrates, these membrane transport proteins can be classified (Fig. 1.1.1) into two main types (i) Active transport and (ii) Facilitated Diffusion (passive transport). The latter are further broadly categorized as (i) Channels and (ii) Carriers.

## Chapter 1: Introduction

A brief outline of the membrane transporters and their classifications are shown below as a schematic flowchart.

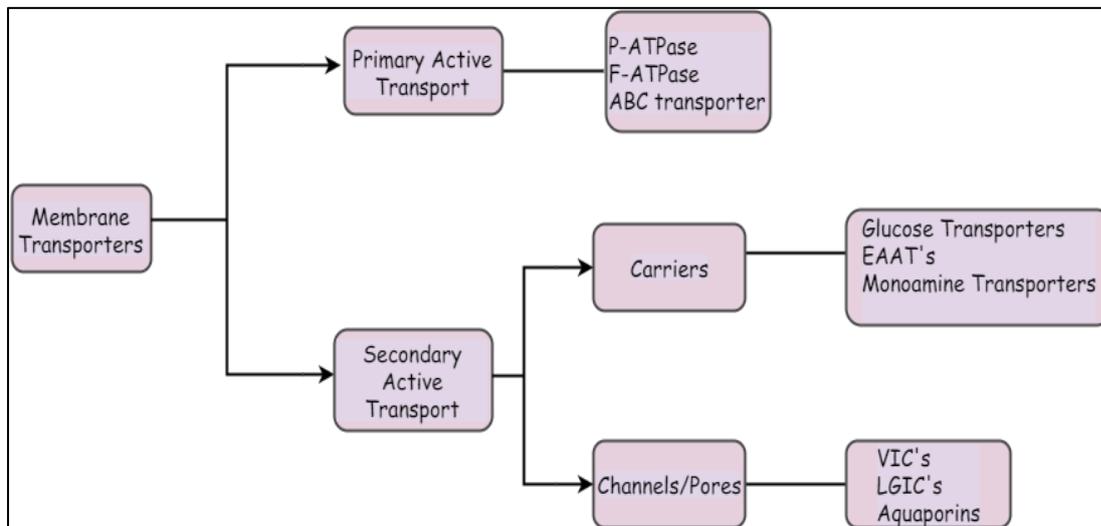


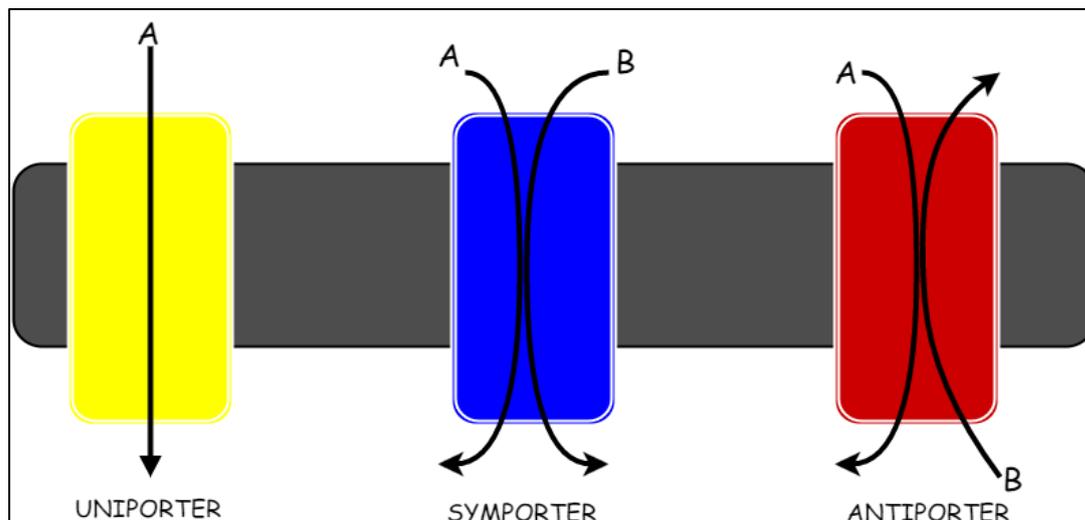
Fig 1.1.1: A schematic diagram for the classification of membrane transporters based on their mode of transport.

The carriers and channels differ in their mode of permeation of solutes. Carriers aren't simultaneously open to the extracellular and intracellular environment, i.e either it is inward open or outward open. In contrast, a channel can be open to both environments simultaneously, permitting solutes to diffuse without intervention. Unlike channel proteins which only transport substances through membrane passively, carrier proteins can transport ions and molecules either passively (facilitated diffusion) or via secondary active transport which involves the use of an electrochemical gradient in lieu of energy produced in the cell like ATP hydrolysis.

There are three main types of carrier transporters that have been identified (Fig. 1.1.1). *Uniporters* transporters which transport one molecule at a time down a concentration gradient. These include **Glucose transporter**, **Monoamine transporters** which move glucose or amino acids across the cellular plasma membrane. In contrast, *antiporters* and *symporters* transporters couple the movement of one type of ion or a small molecule against its concentration gradient to the movement of a different ion or molecule down its concentration gradient, well known studied antiporters are **Na<sup>+</sup>/H<sup>+</sup> antiporters** and the

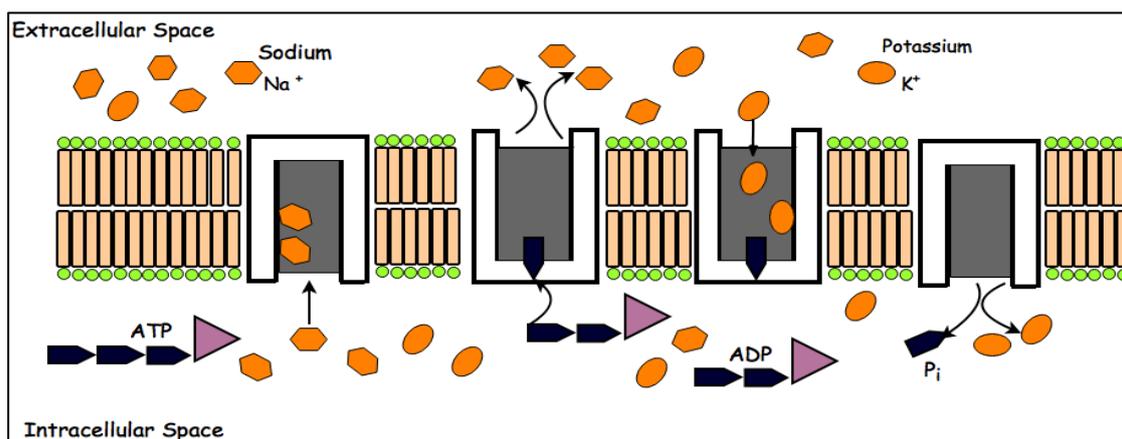
## Chapter 1: Introduction

$\text{Na}^+/\text{Ca}^{2+}$  exchanger and SGLT1 in the intestinal epithelium transports sodium ions and glucose across luminal membrane of epithelial cells.



*Fig 1.1.2: A depiction of the different classes of Carrier Transporters as explained in the above paragraph.*

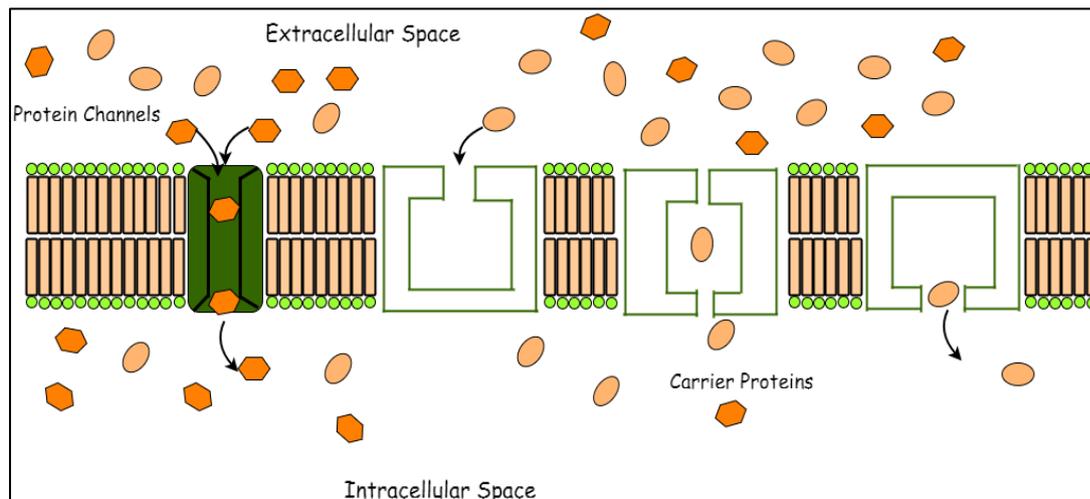
ATP-powered pumps are ATPases that uses the energy of ATP hydrolysis to move ions or small molecules across the membrane against an electrochemical gradient. This process is referred to as active transport, where the energy derived from hydrolysis of ATP to ADP is coupled to the movement of the substances against the electrochemical gradient. The overall reaction --- ATP hydrolysis to ADP and  $\text{P}_i$  and the “uphill” movement of ions and small molecules --- is energetically favorable<sup>2</sup>



*Fig 1.1.3:  $\text{Na}^+/\text{K}^+$  pump shown above is an example of primary active transport.  $\text{Na}^+$  is being exported out against its concentration gradient using the energy which is obtained from ATP hydrolysis to ADP and  $\text{P}_i$  as shown by the two carrier proteins on the left and the proteins on the right employ secondary active transport for transporting potassium. Figure adapted from WIKIPEDIA.*

## Chapter 1: Introduction

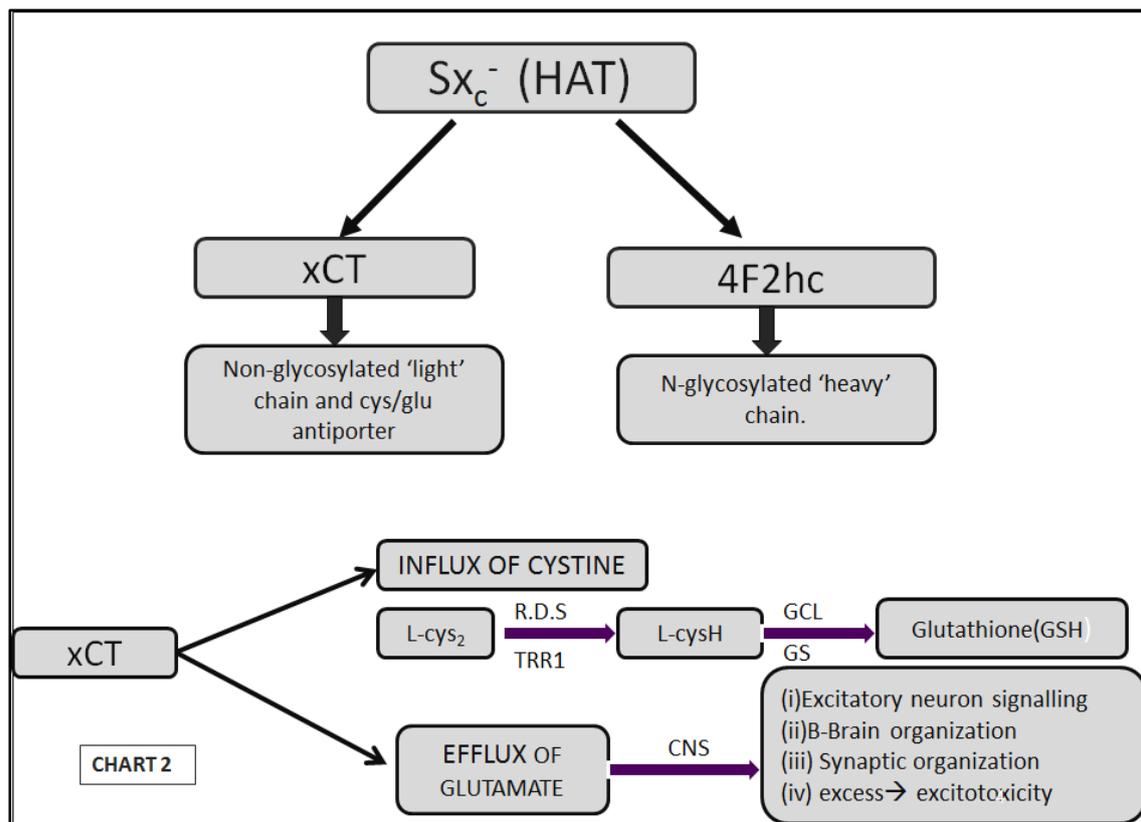
Facilitated diffusion is the movement of molecules or ions across the biological membrane through specific protein transporters and requires no cellular energy input (also known as passive transport). This method is specially used in the case of large polar molecules and charged ions which cannot diffuse freely across the lipid bilayer.



*Fig 1.1.4: Facilitated diffusion in cell membrane depicting ion channels (left) and carrier proteins (three of them) on the right. Figure adapted from WIKIPEDIA.*

## 1.2 An Overview of xCT Membrane Transporter

xCT known as the human cysteine/glutamate antiporter is a part of the amino acid transporter  $Sx_c^-$  that typically mediates the exchange of extracellular L-cystine (L-Cys<sub>2</sub>) and intracellular L-glutamate (L-Glu) across the cellular plasma membrane.  $Sx_c^-$  is a member of the heteromeric amino acid (HAT) family. These transporters are heterodimers (Fig. 1.2.1) composed of two types of II N-glycosylated 'heavy chain' (4F2hc or rBAT, Solute Carrier SLC3 family) linked covalently via a disulphide bond to a non-glycosylated 'light chain' xCT (SLC7 family, SLC7A11). The light chain xCT is credited for the transport activity of the dimer while the heavy chain (4F2hc) acts in the trafficking of the light chain and is essential for cellular surface expression.<sup>3</sup>

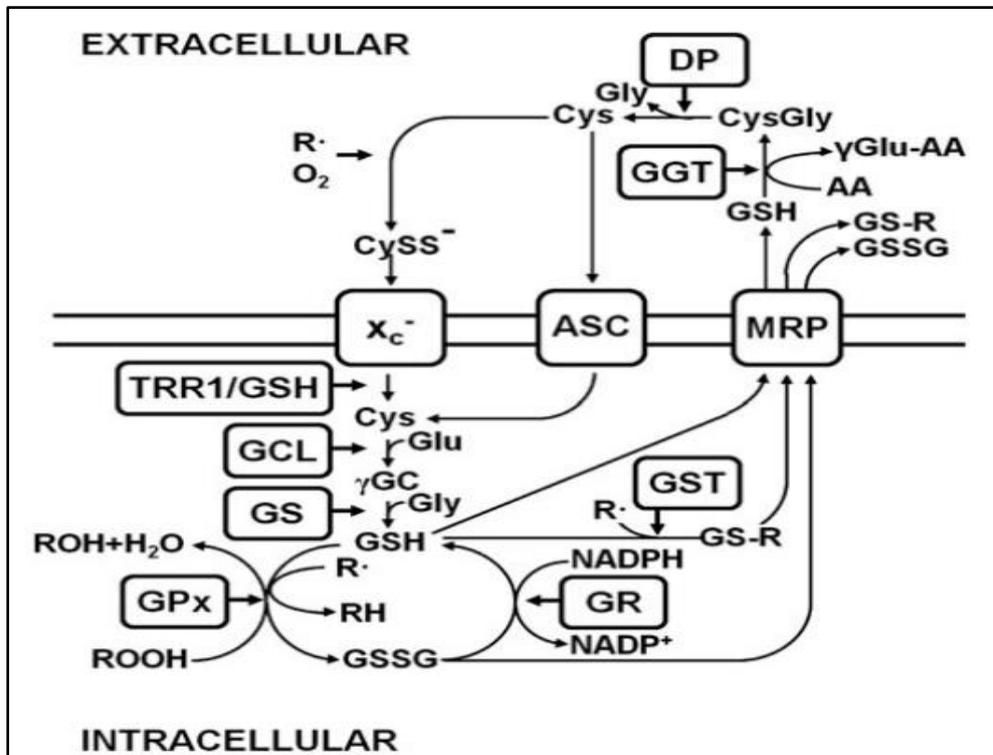


*Fig 1.2.1: A schematic flowchart of the  $Sx_c^-$  and its components along with their functions. Source: Bridges, R. J.; Natale, N. R.; Patel, S. A. 2012.*

(Fig. 5), the influx of L-cystine (L-Cys<sub>2</sub>) mediated by xCT is the rate-limiting step for the formation of the important antioxidant glutathione (GSH), and along with cysteine it forms a key redox couple on its own. Glutathione (GSH) is very much critical for the cell since it is involved in oxidative protection, detoxification of xenobiotic substances and therefore amino acid deprivation, xenobiotic exposure and oxidative stress have all been shown to trigger the upregulation of  $Sx_c^-$ .<sup>3</sup> Glutamate is the most excitatory neurotransmitter in the brain and is implicated for many central nervous system (CNS) functions. When released synaptically, glutamate activates ionotropic glutamate receptors which are present in the postsynaptic part of the synapse. Overactivation of ionotropic receptors induces neuronal death, a pathway known as excitotoxicity.<sup>4</sup> When both sides of the  $Sx_c^-$  are taken into account, the range of CNS processes to which this transport protein has been linked is quite remarkable, it includes as mentioned before oxidative protection, the operation of the blood

## Chapter 1: Introduction

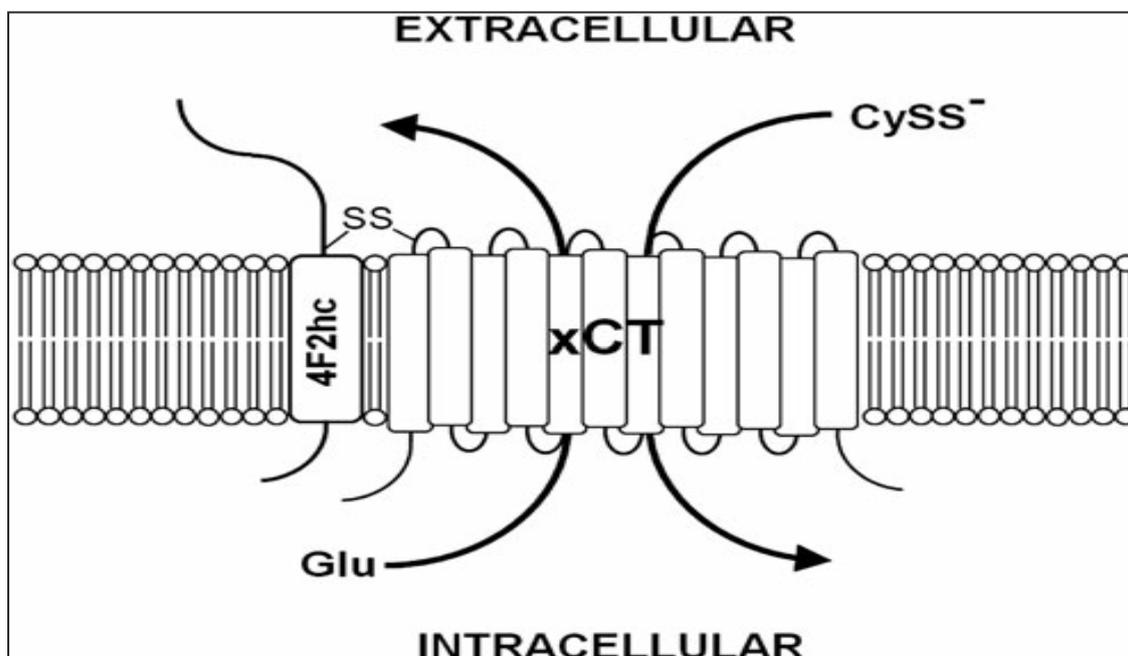
brain barrier, neurotransmitter release, synaptic reorganization, viral pathology, chemosensitivity and chemoresistance.<sup>3</sup>



*Fig 1.2.2: Glutathione (GSH) Metabolism. There is an uptake of Cystine (CySS<sup>-</sup>) by system xc<sup>-</sup> (Sxc<sup>-</sup>). Within the cell, CySS<sup>-</sup> is reduced to cysteine (L-CysH) by thioredoxin reductase 1 (TRR1) or GSH. The synthesis of γ-glutamyl cysteine (γ-GC) from glutamate (Glu) and Cys is catalyzed by Glutamate Cysteine Ligase (GCL), and glutathione synthase (GS) generates GSH by adding glycine (Gly). GSH reduces radicals (R<sup>•</sup>) nonenzymatically and organic hydroperoxides catalyzed by GSH peroxidase (GPx) and is converted to GSH disulfide (GSSG). GSSG is converted back to GSH by GSH reductase (GR), a reaction that uses nicotinamide adenine dinucleotide phosphate in its reduced form (NADPH) as a co-factor. GSH S-transferase (GST) forms GSH adducts (GS-R) from organic molecules (R) and GSH, which is effluxed along with GSH and GSSG from the cell by multi-drug resistance proteins (MRP). The ecto-enzyme γ-glutamyl transferase (GGT) transfers the γ-glutamyl moiety of GSH to an acceptor amino acid (AA), which in turn forms cysteinyl glycine (CysGly), which is cleaved by a dipeptidase (DP) to Cys and Gly. GGT and DP are both membrane-bound enzymes. Cys is either taken up by cysteine transporters, like system alanine-serine-cysteine (ASC), or extracellularly oxidized to CySS<sup>-</sup>, which is again taken up by Sxc<sup>-</sup>. Adapted from **Bridges, R. J.; Natale, N. R.; Patel, S. A. 2012.***

## Chapter 1: Introduction

$Sx_c^-$  has been found in a variety of cells, both outside (for e.g fibroblasts, macrophages, and endothelial cells.) and inside the central nervous system [e.g astrocytes, microglia, immature cortical cells and glioma cell lines. The characterization of  $Sx_c^-$  immunohistochemically i.e by using antibodies to 4F2hc and xCT shows that the transporter part (xCT) is localized to neurons and glia in the CNS. Furthermore,  $Sx_c^-$  was highly expressed at the borders of CNS and the periphery including the vascular endothelial cells, ependymal cells, choroid plexus and the leptomeninges. Similarly, the expression of xCT and 4F2hc was also detected immunohistochemically in the brush border membranes of the kidney and the duodenum.



*Fig 1.2.3: The overall structure of system  $x_c^-$  comprises of the 4F2 heavy chain (4F2hc) and the light chain, xCT, which are linked by a covalent disulphide bond (-S-S-).  $Sx_c^-$  imports cysteine (CysSS<sup>-</sup>) in exchange for glutamate.<sup>4</sup>*

Although  $Sx_c^-$  can transport both L-Cys<sub>2</sub> and L-Glu in both directions, the intracellular concentration of L-Cys<sub>2</sub> is very small because intracellular cysteine is reduced and the intracellular glutamate concentration is generally higher than in the extracellular space, system  $x_c^-$  generally imports cysteine while exporting glutamate.<sup>4</sup>

$Sx_c^-$  has been shown to (i) utilize either L-Cys<sub>2</sub> or L-Glu as substrates, with each one acting as an inhibitor of the other; (ii) function as an antiporter, mediating 1:1 exchange

## Chapter 1: Introduction

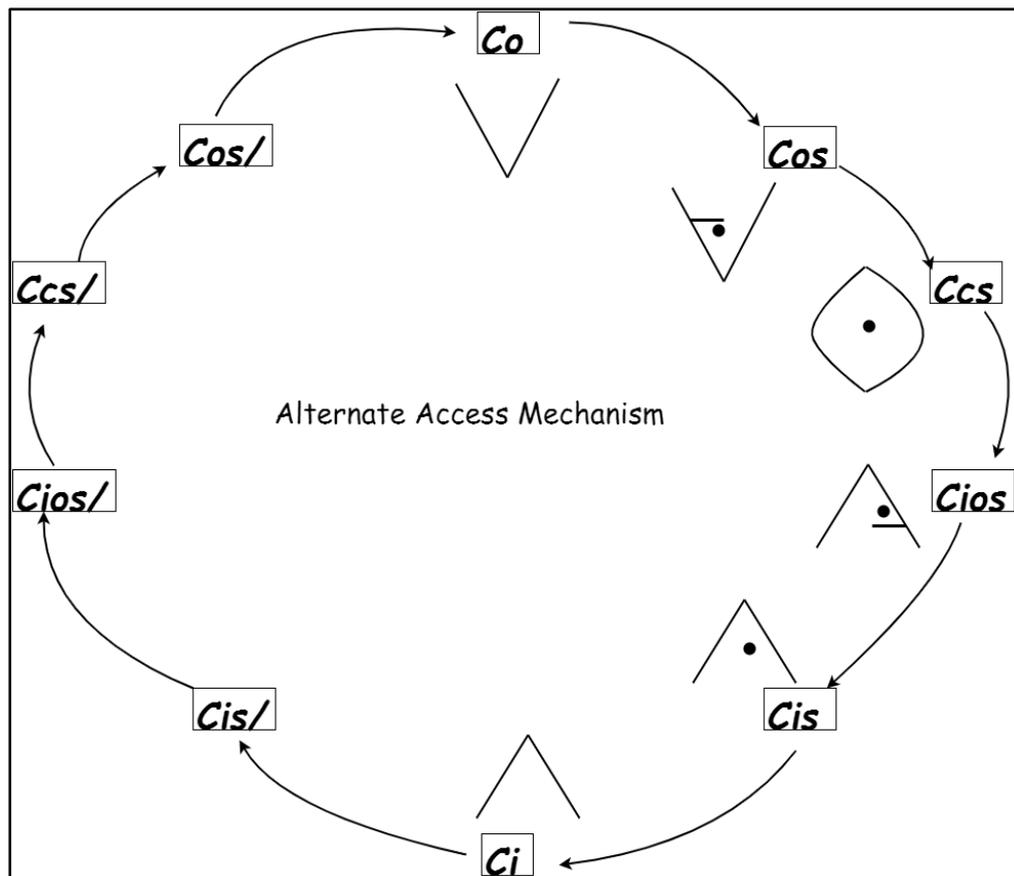
of an intracellular and extracellular amino acid; (iii) act in an electroneutral manner, a chlorine-dependent and sodium-independent fashion (iv) transport L-Gly or L-Cys<sub>2</sub> in an anionic form.

### 1.3 ALTERNATE ACCESS MECHANISM CYCLE

The movement of solutes across the membrane which are facilitated by transporter proteins has been understood to follow an alternating access mechanism where conformational changes expose the binding site located centrally for one or more substrates on either side of the plasma membrane.<sup>5</sup>

In the alternate access mechanism cycle, the substrate binding site which is usually located centrally in the transmembrane protein is linked to the two sides of the membrane by permeation pathways, only one of which is accessible at a time. The major facilitator superfamily (MFS)<sup>6</sup> of membrane transporters has been well studied where they have been able to obtain multiple conformations for each MFS transporter (e.g LacY<sup>7</sup> and Xyle) which can provide a platform or a framework for delineating the mechanism of alternate access. Another well-known membrane transporter is LeuT<sup>5</sup> (a bacterial homologue of the mammalian neurotransmitter transporter) which is a transmembrane protein which consists of two structurally similar motifs (TMHs1-5 and TMHs6-10) that are related by a pseudo C2 axis rotation with respect to the plane of the membrane, and that this particular feature might form the two alternate mediation /permeation pathways that lead to one side of the membrane environment or the other.

A literature survey of homologous proteins obtained from SWISS-MODEL (as discussed in the next chapter) for xCT showed that there was considerable similarity/homology with LeuT and therefore we have speculated that it might also follow the Alternate Access Mechanism cycle (Fig. 1.3.1)



*Fig 1.3.1: The Alternate access mechanism cycle shown with different conformations. For xCT, since it is an antiporter, the first half of the cycle is for the uptake of L-Cystine (L-cys<sub>2</sub>) and the second half (which is depicted by a '/') is for exporting L-Glutamate (L-Glu). Co as outward open conformation; Cos as occluded substrate bound outward open conformation; Ccs as substrate bound closed conformation; Cios as occluded substrate bound inward open conformation; Cis as substrate bound inward open conformation; Ci as inward open conformation.*

#### 1.4 MOTIVATION FOR SELECTING THESIS TOPIC.

The fact that system xc- is critical for many CNS functions, warrants the need to understand the structural aspects of xCT transporter. The importance of knowing its structure can aid us in understanding many different factors which include:

- a) Substrate activity for example Ligand Screening (investigating competitive assays which screen the molecules that inhibit the binding of the amino acids L-Cys<sub>2</sub> or L-Glu with the transporter xCT)

## Chapter 1: Introduction

- b) Therapeutic targets for neurodegenerative disorders linked to system xc-, like Alzheimer's, Parkinson's, gliomas and schizophrenia and treating drug addictions.
- c) Constructing Models that provide a strategy that help us collectively analyze and visualize the Structure-Activity Relationship(SAR's) that govern ligand-binding site interactions.

Hence, probing the structural aspects of human cysteine/glutamate antiporter xCT is necessary and this is the motivation for selecting my thesis topic.

## REFERENCES:

- (1) Milton H. Saier, J. In *Transmembrane Transporters*; Quick, M. W., Ed.; p 1.
- (2) Lodish, H.; Berk, A.; Zipursky, S. In *Molecular Cell Biology.*; W. H. Freeman: New York, 2000.
- (3) Bridges, R. J.; Natale, N. R.; Patel, S. A. **2012**.
- (4) Lewerenz, J.; Hewett, S. J.; Huang, Y.; Lambros, M.; Gout, P. W.; Kalivas, P. W.; Massie, A.; Smolders, I.; Methner, A.; Pergande, M.; Smith, S. B. **2013**, *18* (5), 522–555.
- (5) Forrest, L. R.; Zhang, Y.-W.; Jacobs, M. T.; Gesmonde, J.; Xie, L.; Honig, B. H.; Rudnick, G. .
- (6) Yan, N. **2015**, 257–285.
- (7) Stroud, R. M. **2007**, *104* (2), 1445–1446.

# CHAPTER 2

## INTRODUCTION

The focus of this chapter is to illustrate the different methodologies, which are key to delineate different structural aspects of the human cysteine transporter xCT. The chapter will be split into three parts, the first part will emphasize on structural modeling of xCT followed by membrane building and conventional molecular dynamics refinement, which include minimization and equilibration of the assembled membrane/protein complex. The second part will focus on the ‘docking’ of the human cysteine /glutamate antiporter with the substrate (ligand) which can provide us with the possible interactions between the ligand and the receptor (Xct). The final part will include Targeted Molecular Dynamics (TMD), which is an enhanced molecular dynamics simulation technique employed to investigate different conformations in the alternate access cycle (as discussed in the previous chapter) using appropriate forces/ restraints on the membrane/ protein system.

The results for the methodologies for **Part 1**, **Part 2** and **Part3** will be discussed in the next chapter.

## METHODOLOGY (PART 1)

### 2.1 HOMOLOGY/ COMPARATIVE MODELLING

The structure of human cysteine/glutamate antiporter is not available and since elucidating the structural aspects is necessary for understanding how the protein functions, we first need to obtain a model of the structure.

## Chapter 2: Methodology (Part 1)

The objective of this method is to obtain a three-dimensional (3-D) structure of an unknown protein (the target) based on the structure of related or homologous proteins (the templates) which are in the database. The conditions which are necessary for obtaining a useful model are that there must be some considerable similarities between the target and the template sequences and that the correct alignment between them must be constructed via modeling. The applicability of this method and its robustness can be explained as follows; (i) any small change in the protein sequence can lead to a small change in its 3D structure (this is reason why it can be applied) ,(ii) the number of different structural folds that a protein adopts is limited and the since (iii) the number of experimentally determined structures are increasing at an exponential rate comparative modelling can be applied to many protein sequences<sup>1</sup>. Furthermore, since 28% of the known protein sequences have at least a 25% residue identity /similarity with one of the known protein structures, this technique can predict structures of sequences that are an order of magnitude more than those determined by experiments<sup>2</sup>.

There are different software available for homology or comparative modeling and ones which are useful for 'model evaluation' Each of them employs its own algorithms or procedures for predicting the structure of the protein. In the forthcoming pages, we have utilized the software MODELLER<sup>2345</sup> for modeling of protein three-dimensional structures.

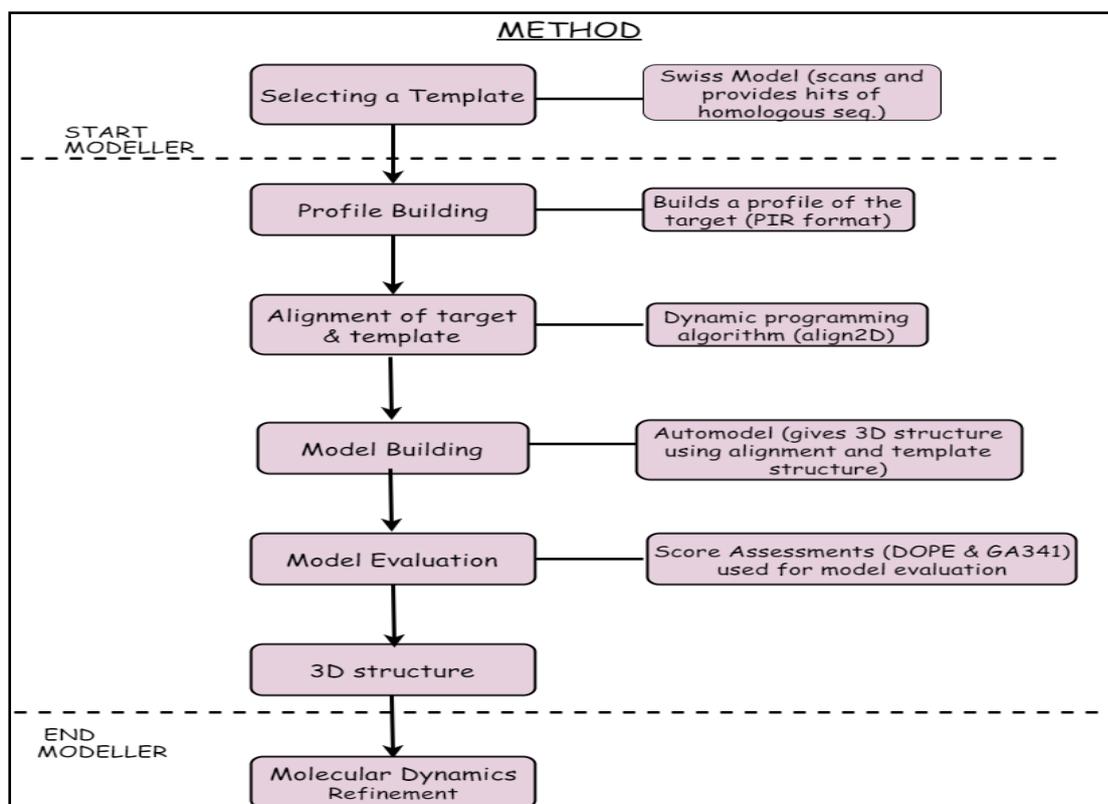
Any comparative modeling program or algorithm involves four main steps.

- a) The first step involves the identification or a search of proteins with known three-dimensional structure that are homologous or related to the target sequence.
- b) The second step includes the alignment with the target sequence and selects those known structures which will be utilized as templates.
- c) The third step is model building, where a model is built for the target sequence given its alignment with the template structures.
- d) The final and the fourth step involves model evaluation based on different criteria.

## Chapter 2: Methodology (Part 1)

The main distinction between different homology modeling techniques is how the model is calculated or generated (third step) and since I had previously mentioned, for each one of the steps mentioned above different programs and softwares are available.

### Comparative Modelling using MODELLER



*Fig 2.1.1 : A schematic flowchart of the steps involved in comparative modeling employed in MODELLER. Source Becker, O. M.; MacKerell, A. D.; Roux, B. (Biologist); Watanabe, M. Computational biochemistry and biophysics; M. Dekker, 2001.*

For the first step in modeling, there are three basic classes of protein comparison methods which can be useful for fold assignment or identification. The first class is where the target sequence is compared with the known sequence database of protein independently using a pairwise sequence comparison (e.g FASTA<sup>6</sup> and BLAST<sup>7</sup>). The second class uses the multiple sequence comparison to improve the sensitivity of the search (e.g PSI-BLAST<sup>8</sup>). The last and third class performs a comparison of a protein sequence

## Chapter 2: Methodology (Part 1)

(target sequence) with a protein structure. MODELLER can also compare automatically of the known 3D structures related to a given target/sequence of interest using a dynamic programming algorithm for pairwise sequence comparison.

However, the method which has been followed in the thesis work is by employing SWISS-MODEL Template Library<sup>910112</sup> (SMTL) which uses BLAST and HHBlits<sup>13</sup> (which employs a Hidden Markov Models HMM-HMM alignment of protein sequences) searches to scan the sequence database and provide a list of hits of homologous proteins.

In MODELLER, (i) the first step (Fig. 2.1.1) involves building a profile which puts the target sequence into a PIR format readable by MODELLER and the output is an alignment file with an extension ‘.pir’.

(ii) The next step in (Fig. 2.1.1) involves aligning the target with the template. The alignment is carried out by a local dynamic programming tool known as *align2d()* which is different from the conventional sequence-sequence alignment because it takes into account the template’s structural information while constructing the alignment. The task is achieved by placing a variable gap penalty function that tends to place gaps in solvent exposed regions and curved regions between two positions close in space. The resulting alignment errors are considered to be reduced to 1/3<sup>rd</sup> relative to those that occur in standard sequence-sequence alignment<sup>2345</sup>.

(iii) The third and the most important part of modeling is model building. MODELLER employs the **automodel** class which calculates a 3D model of the target completely automatically using the target-template alignment from the previous step. This step is forms an integral part of MODELLER software. It uses the alignment information to extract spatial restraints, and a 3D model is obtained by the satisfaction of spatial restraints. The restraints are expressed as a ‘conditional’ probability distribution functions (pdfs) and can be expressed by the following equations:

$$p(x | a, b, \dots, c) \approx W_{x,a,b,\dots,c} \approx f(x, a, b, \dots, c, q) \quad \text{Eqn (2.1.1)}$$

where  $W$  matrix spanned by the elements  $x, a, b, \dots, c$ , the observed relative frequencies of occurrence of  $x$  given  $a, b, \dots, c$  and  $f$  is an analytical function which is fitted to  $W$ .  $f$  is a function with parameters ‘ $q$ ’ which are obtained by applying the least-squares principle, i.e the best  $q$  is that which minimizes the function:

## Chapter 2: Methodology (Part 1)

$$r.m.s = \sqrt{\sum_{x,a,b,\dots,c} [W_{x,a,b,\dots,c} - f(x, a, b \dots, c, q)]^2} \quad Eqn (2.1.2)$$

and  $W$  is calculated from the absolute frequencies  $W'$  using :

$$W_{x,a,b,\dots,c} = \frac{W'_{x,a,b,\dots,c}}{\sum_x W'_{x,a,b,\dots,c}} \quad Eqn (2.1.3)$$

where  $W'$  are the absolute frequencies which are obtained by counting the number of occurrences of each  $(x,a,b,\dots,c)$  values in the database of known proteins structure.

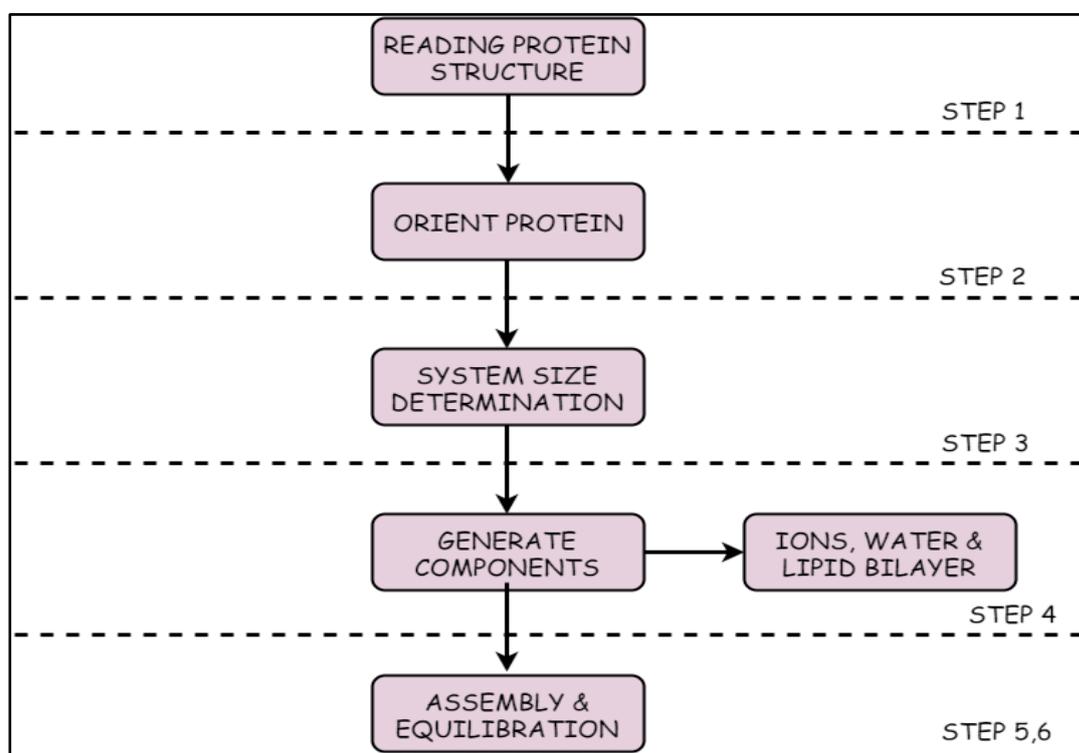
(iv) The final step involves model evaluation; usually, it is assumed that the errors pop up from two main sources, the failure of the conformational search to identify the optimal conformation and failure of the scoring function to determine the optimal conformation<sup>1</sup>. In MODELLER, several models are calculated for the same target and the best model is selected that has the lowest value of the scoring or the objective functions like Discrete Optimized Protein Energy (DOPE)<sup>14</sup>, or the highest value of GA341<sup>1516</sup> assessment scores. Finally, to have a complete picture, a DOPE score profile (plot of DOPE energy vs residue number) of the model profile and the template profile can be obtained to show regions of structurally variable regions (high DOPE energies) like the loops and structurally conserved regions (low DOPE energies). Programs like PROCHECK, MODEVAL, WHATCHECK and many more are also quite useful for model evaluation for evaluating whether the model has good 'stereochemistry.'

Once the best model 3D structure of the protein is obtained, molecular dynamics refinement is performed to equilibrate the structure and perform other dynamics. However, prior to molecular dynamics refinement, it is necessary to build a membrane environment as human cysteine transporter (xCT) is a transmembrane transporter and therefore constructing an environment that mimics the cellular plasma membrane (lipid bilayer) would fit to serve the purpose of further investigations like understanding substrate permeation.

## 2.2 MEMBRANE BUILDING

Molecular dynamics comes in as a handy tool for an in-depth understanding of functions and interactions of membrane proteins with the surrounding environment at the atomistic level. However, as compared to globular protein solvation dynamics, the membrane/protein complex simulations can be computationally quite challenging. Membrane building for one of the models obtained from homology modeling has been carried out efficiently with the CHARMM-GUI<sup>17</sup> server. It provides a web-based graphical user interface to generate various molecular simulation complex systems and generates input files for further use in advanced simulation techniques.

Membrane Builder<sup>18192021</sup>, a module in CHARMM-GUI, aims to help the user to set up simulation of complex protein/membrane systems directly on the web through a generalized and automated building procedure which comprises of orientation of the protein, system size determination, generation of the lipid bilayer using a variety of lipid molecules (for eg. DPPC, DMPC, POPC, CHL1 etc.) , the pore water, the bulk water and ions and then assembling them into a rectangular or hexagonal form.



*Fig 2.2.1 : An outline of the steps which are involved in membrane building by the Membrane Builder module in CHARMM-GUI. Adapted from Jo, S. 2007, PLoS ONE 2(9):e880.*

## Chapter 2: Methodology (Part 1)

- (i) The first step carried out in the web browser GUI is to read the protein structure. Users are given the freedom to upload their protein structure or specify the PDB ID's and the database to download the pdb file. In the case for xCT, the structure obtained from homology modeling was given as the input.
  
- (ii) The second step involves orientation of the protein, where the user is given two options of orientation. Since usually the proteins obtained from PDB aren't oriented properly, the membrane builder can aid the user to place the protein appropriately in the lipid bilayer by either aligning it along the principle axis along Z or a vector connecting two residues (two C $\alpha$  atoms) along the Z-axis. For xCT, since we cannot identify two residues (two C $\alpha$  atoms) which are parallel with the Z-axis, latter method will not be very useful and therefore orienting the protein along the principle axis i.e., along Z would be much more feasible. The assumption followed in Membrane Builder is that the membrane normal is parallel to the Z-axis and the center is at Z=0 Å. Once the alignment is done, the module will calculate the protein cross-sectional area (the profile is along the Z-axis. The cross-sectional profile can be used to figure out the protein areas corresponding to the maximum values at  $10 \text{ \AA} < Z < 20 \text{ \AA}$  and  $-20 \text{ \AA} < Z < -10 \text{ \AA}$  respectively<sup>21</sup>.
  
- (iii) The third step of membrane building involves system size determination. Once the protein areas are obtained from step 2, one can estimate the number of lipid molecules to set up the system of specific dimension or shape (rectangular or hexagonal) in the XY direction<sup>18</sup>. The module provides the option of (i) either choosing the ratio of the lipids in the upper or lower leaflet or (ii) the number of lipid components for the XY size determination according to a guess X and Y length option which has to be specified by the user. The system size along the Z axis can be calculated by the thickness of the bulk water layer from the protein extent along the Z axis. After system size determination, Membrane Builder provides a summary of the system size and the possible lipid packing and the water thickness if at all the user would want to make any changes<sup>21</sup>.

## Chapter 2: Methodology (Part 1)

- (iv) The fourth step is based on the system size determination. From the information obtained from the previous step of size determination, the module will generate components to completely solvate the protein, like the lipid bilayer, bulk water (-xCT is not a pore protein and thus, pore water generation was not done) and counter ions. The ions can be generated either by specifying the concentration or by neutralization method which will give the number of positive and negative ions required to neutralize the system and the placing of these ions are usually carried out by Monte Carlo method.
- (v) The last two steps indulge in the assembly of the generated components and equilibration. The advantage of using the web-based GUI is that it is very user-friendly as the user is provided with the option of starting over again without exiting the server if any problem occurs. Once the generated components are assembled, it is necessary to relax the system from its initial uncorrelated state and for equilibration process the builder prepares input files with different parameters on the system components which will be discussed in detail in the next section.

Another critical aspect of the Membrane Builder is lipid bilayer building. The Builder provides two options for the user based on which the lipid bilayer can be built they are (i) Insertion Method (ii) Replacement Method

- (i) The first one is called as the Insertion method where a hole is created by applying weak radial repulsive forces in the lipid bilayer during the bilayer simulations, and the protein is inserted into the hole. Usually, the insertion method is very efficient for regular and cylindrical proteins since it utilizes the lipid library which has a variety of holes in the pre-equilibrated lipid bilayer, and it takes less than a minute for generating the lipid bilayer<sup>21</sup>.
- (ii) However, the replacement method creates a distribution of pseudo lipid-like atoms around the protein and then substituting them with actual lipid molecules chosen randomly from the lipid library. The distinction between the two methods comes in the fact that the replacement method wraps lipid molecules around the protein and therefore can be virtually be utilized for any membrane protein, including transmembrane proteins, monotopic proteins and interfacial proteins and therefore is quite versatile<sup>21</sup>.

### 2.3 MOLECULAR DYNAMICS REFINEMENT

Molecular dynamics simulations are crucial for understanding the interaction between biological macromolecules at the atomistic timescale. Since it is hard experimentally to capture these motions, MD acts as a tool providing a detailed modeling at the atomistic scale.

In contrast to the conventional Monte Carlo approach, which is restricted to mainly positional information of the system, molecular dynamics can provide information on the magnitude of fluctuations in both momentum and position space and also the time evolution of the system in the phase space. Therefore, it is very useful for monitoring time-dependent processes in complex biological systems to study their structural, dynamic and thermodynamic properties by solving the appropriate equation of motion that govern the motion of the molecules in the system<sup>1</sup>.

The main ‘simulation protocol’ followed in any dynamic simulation are the follows:

- (i) **Data Preparation:** This step involves obtaining the initial coordinates and assignment of initial velocity. Since the equation which governs the motion of molecules are second order differential equations, like Newton's equation of motion, Langevin (stochastic dynamics), etc., they require two initial conditions for each degree of freedom to perform the integration. In this case the two initial conditions are  $\{r(0)\}$  and initial velocities  $\{v(0)\}$ . The initial coordinates can be obtained from experimentally determined structures, mainly from NMR or X-Ray crystallographic experiments. The initial velocities aren't however obtained experimentally but are usually randomly assigned from the Maxwell velocity distribution by specifying the system temperature T.

$$P(v) dv = \left( \frac{m}{2\pi k_b T} \right)^{1/2} \exp \left[ \frac{-mv^2}{2k_b T} \right] dv \quad \text{Eqn (2.3.1)}$$

Another important aspect in this step is the minimization of the experimentally determined structures. The key aspect of minimization is to alleviate any local stress due to nonbonded interaction, as well as to relax the experimentally determined bond distances and angle distortions.

## Chapter 2: Methodology (Part 1)

- (ii) **Heating or coupling the system to a heat bath:** This step is necessary to get the system to the desired temperature and minimize any fluctuations in its energy which comes as function of temperature, The simplest method that keeps the temperature in an MD process to its desired value is the velocity rescaling of the atoms at each time step by a factor of  $(T_0/T)^{1/2}$ , where T is the current instantaneous temperature, and  $T_0$  is the required temperature. Another efficient method proposed by Berendsen<sup>22</sup>, that used the scaling of velocities of all the atoms of the system to couple it to a constant heat bath operating at a temperature  $T_0$ . The heat bath also has a relaxation time which is characteristic to it and therefore, each velocity is scaled by a factor  $\lambda$  which is given by:

$$\lambda = \left[ 1 + \frac{\Delta t}{2\tau_T} \left( \frac{T_0}{T} - 1 \right) \right]^{\frac{1}{2}} \quad \text{Eqn (2.3.2)}$$

where  $\Delta t$  is the integration time step size and  $\tau_T$  is the relaxation time step, and T is the instantaneous temperature. This method is commonly referred to as ‘constant temperature molecular dynamics. This method is usually accompanied by shorter equilibration dynamics.

- (iii) **Equilibration:** This step involves relaxing the system over a long time to ensure that the whole system has some stability and it is devoid of erratic fluctuations. In cases with large membrane/ protein systems, algorithms (RATTLE<sup>23</sup> or SHAKE<sup>24</sup>) which can freeze high frequency vibrations like bond stretching and bending involving H atoms are commonly used in order to remove the limitations on the integration time step and therefore they can increase the value of  $\Delta t$  which enables us to probe longer time scale dynamics. Equilibration simulations can span from tens of picoseconds to several hundred picoseconds.
- (iv) **Production :** Once the system is equilibrated, the dynamics of the system is considered to be significant for calculating the observables of interest, and the trajectory obtained after solving the equations of motions is stored for further analysis. Usually, a typical production run takes from several picoseconds (ps) to sub-nanoseconds (ns) or these days, microseconds ( $\mu$ s) depending on the system size and computational power.

## Chapter 2: Methodology (Part 1)

- (v) **Analysis:** The last part of any simulation is the analysis of the generated trajectories. Different analysis can be performed, such as structural deviations, fluctuations, solvent accessibilities, transport coefficients, conformational sampling and clustering using enhanced MD techniques, essential mode analysis etc.

In the case of xCT, the final output obtained after building the membrane using the CHARMM-GUI server are a set of input files which are fed into the molecular dynamics software, I used NAMD<sup>25</sup> (scalable molecular dynamics) to minimization and equilibration of the system of membrane/protein complex.

Step	Ensemble	Time-Step	Simulation time	Force Constants for harmonic restraints			
				Protein backbone <sup>1</sup>	Protein sidechain <sup>1</sup>	Lipid Head <sup>2</sup>	Lipid Tail <sup>3</sup>
1.	NPT	1 fs	25 ps	10.0	5.0	5.0	500
2	NPT	1 fs	25 ps	5.0	2.5	5.0	200
3	NPT	1 fs	25 ps	2.5	1.0	2.0	100
4	NPT	2 fs	100 ps	1.0	0.5	1.0	100
5	NPT	2 fs	100 ps	0.5	0.1	0.2	50
6	NPT	2 fs	100 ps	0.1	0.0	0.0	0

1. Positional harmonic restraints in kcal/(mol.A<sup>o2</sup>).

2. Planar harmonic restraints [in kcal/(mol.A<sup>o2</sup>) ] to keep the lipid head groups in the initial assigned values.

3. Dihedral restraints [in kcal/(mol.rad<sup>2</sup>) ] to keep the cis-double bonds in the acyl chains (lipid tails) and to keep the C2 chirality in the glycerol backbone

**TABLE 2.3.1** : This is a table of restraints applied on the proteins, lipid molecules in the bilayer during the simulations to assure a gradual relaxation of the system from its initial uncorrelated high energy state<sup>21</sup>. Adapted from Jo, S. 2007, PLoS ONE 2(9):e880.

## **METHODOLOGY (PART 2):**

### **2.4 AUTODOCK FOR DOCKING STUDIES.**

Studying the docking of ligand to its receptor ( macromolecule in general) is indeed a challenge because the standard binding free energy computation is non-trivial owing to not only the translational and rotational variations but also the conformational entropies which need to be fully captured in a finite-length statistical simulation<sup>26</sup>. In order to address the issue, different docking methods have been developed computationally over the last few years which use different search methods to get the conformation of the ligands in the bound state to the macromolecular targets and also estimate the free energy change upon binding.

AutoDock<sup>27282930</sup> is one of the most robust methods developed for solving the abovementioned task. AutoDock has been refined over the years and the newer versions have implemented three new search methods: (i) genetic algorithm (GA) (ii) a local search algorithm (LA) and (iii) an efficient adaptive global-local search method which is based on Lamarckian genetics, called as the Lamarckian Genetic Algorithm (LGA). Another advancement is the development of the empirical binding free energy force fields which aid in the binding free energy predictions, and the docked ligand binding constants. The basic idea behind the use of genetic algorithm will be briefly presented in the coming paragraphs, but the details of implementation of the search methods will be omitted.

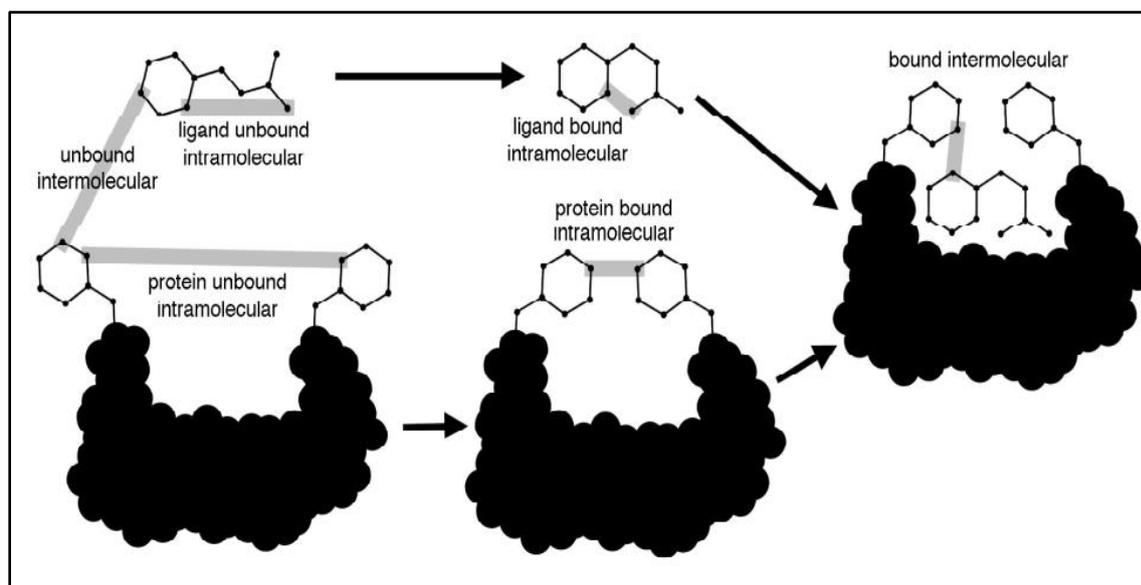
In the case of molecular docking, the genetic algorithms operate on the premise which states that the particular interactions and arrangements of the protein and the ligand can be defined by a set of values which can describe the orientation, translation and the conformation of the ligand with respect to the protein. These are called as the ligand's state variable, and each state variable is associated with a gene in the genetic algorithm. The *fitness* corresponds to the total energy of ligand-protein interaction and is evaluated with an energy function. The algorithm implements the following steps: (i) random pairing of individuals for mating using the crossover process in which the progeny inherit genes from either parent (ii) random mutation in offspring which might change the gene by a random amount and (iii) selection of the progeny in the current generation based on the fitness of

## Chapter 2: Methodology (Part 2)

the individuals. The creation of random population of individuals is looped over generations, an iterative process which occurs until the maximum number of energy evaluations or the maximum number of generations is reached<sup>29</sup>.

### • Theory behind the Free Energy Scoring Function

As I mentioned previously, AutoDock uses an empirical free energy force field for conformational evaluation during docking. A large number of protein-inhibitor complexes have been used for parametrizing the force fields for which both the structure and inhibition constants  $K_i$  are well known<sup>30</sup>.



*Fig 2.4.1: A diagram depicting the intermolecular and intramolecular interactions between protein and ligand. Adapted from Morris, G. M.; Goodsell, D. S.; Pique, M. E.; Lindstrom, W. L.; Huey, R.; Hart, W. E.; Halliday, S.; Belew, R.; Olson, A. J. User Guid. 2009, 1–49.*

In (Fig2.4.1), the force field is evaluated in two steps overall,

- (i) First, the intramolecular energies are calculated for the transition from the free or unbound state to the bound state for both the protein and the ligand conformations.
- (ii) Second, the intermolecular energies are computed for combining the ligand in their bound conformation.

## Chapter 2: Methodology (Part 2)

The complete force field includes six pair-wise potentials (V) and an estimate of the loss in the conformational entropy upon binding ( $\Delta S_{conf}$ ) and is expressed as follows:

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad Eqn (2.4.1)$$

where P refers to the “protein” and L refers to the “ligand” in the docking evaluation.

Each of the pairwise energetic terms consists of terms necessary for evaluating the dispersion/repulsion Lenard-Jones (LJ) kind-of potential, hydrogen bonding potential, electrostatics and desolvation potentials. It is expressed as follows:

$$V = W_{vdw} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \left( \frac{q_i q_j}{e(r_{ij})r_{ij}} \right) + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad Eqn (2.4.2)$$

**W**'s are optimized for the free energy calibration, which is based on experimentally determined set of binding constants. In the abovementioned equation, the first term is a typical LJ (6/12) potential for dispersion/repulsion interactions. The second term is a directional H-bond based on the 10/12 potential. **E(t)** function provides the directionality based on the angle t from the equilibrium H-bonding geometry. The third term in the equation is a screened electrostatic potential (Coulombic). The last term is a desolvation potential that consists of the volume of the number of atoms (**V<sub>j</sub>**) that surround a given atom ‘i’ and shelter it from the solvent molecules, and a weighted solvation parameter (S) and an exponential term with  $\sigma$  as the distance-weighting factor<sup>30</sup>.

### • Steps involved in AutoDock Calculations

Four major steps go into the AutoDock calculations: (1) Coordinate file preparation using **AutoDockTools**<sup>27</sup> (2) pre-calculation of the atomic affinity potentials mentioned in the equations using **AutoGrid** (iii) ligand docking using **AutoDock** (iv) analysis of the docking using **AutoDockTools**.

## Chapter 2: Methodology (Part 2)

- (i) Preparation of the coordinate file: The file format as read by AutoDockTools is an extension to the PDB format, known as PDBQT. The ligand and receptor coordinate files required for the consecutive steps are created in this format which includes polar hydrogen, where the ADT adds polar hydrogens and merges the nonpolar hydrogens (i.e., which is connected to carbon) to the heavy atom itself, addition of partial atomic charges and different atom types ; ADT has several different atom types which are separate for aromatic and aliphatic carbon atoms, and also types that distinguish the polar atoms that form H-bonding to those that do not. It also includes information on the flexibility of the receptor side-chains and creates a separate file for it and decodes the torsional degrees of freedom for the ligand of choice.
- (ii) AutoGrid Calculations: This step involves in the pre-calculation of the affinity potential terms for each atom type in the ligand molecule that helps to accelerate the docking calculations. This method places the protein in a 3-D lattice of equally spaced points (a grid), and a 'probe' atom is placed at each point. The interaction energy of each of the probe atoms with the entire protein is assigned to the grid points. The ligand conformation energies are later evaluated based on the values from the grids during the docking calculations.
- (iii) Docking using AutoDock: As I have mentioned before the docking simulations are carried out using search methods like genetic algorithm (GA) which performs a global search , or the local search (LS) which is adaptive in nature since it can alter the step size depending on the success or failures (i.e., increases or decreases in energy) in the consecutive steps, or a hybrid of the above methods called as Lamarckian Genetic Algorithm (LGA). However, the most efficient method is the LGA method which is run for several times to obtain several docking conformations and the best solution ('selection') is identified after detailed analysis and after checking the consistency of the results.
- (iv) Analysis using AutoDockTools: Different tools are employed for analysis of the docking simulations by ADT, which include clubbing or clustering of similar conformations, conformation visualization ranked by the binding energy, visualizing ligand-protein interactions in 3-D.

## **METHODOLOGY (PART 3)**

### **2.5 ENHANCED MOLECULAR DYNAMICS:**

Molecular dynamics has provided an important tool for deciphering the atomistic details of biomolecular interactions and has been viewed as a general sampling method for studying the changes in the conformations of biomolecules. However, sampling over all of the conformational space is a cumbersome task due to time constraint and computational complexity. The inaccuracy of the force fields to describe all the necessary interactions as seen in the experiments coupled with the high computational cost render the conventional MD techniques inefficient to cover longer timescales, which are linked to functional aspects of the biomolecules<sup>3132</sup>. The main objective in the pages that follow would be to employ enhanced sampling techniques known as Targeted Molecular Dynamics (TMD) to go from one relevant conformational sub-state to the other in a considerable timescale.

In the previous chapter, the alternating access mechanism was discussed briefly and also a speculation that xCT, operate via this mechanism. TMD allows one to ‘steer’ one conformation to another; and thus, investigate the transition pathways associated between two conformations. TMD has been extensively used for studying conformational changes and binding interactions in biologically important receptors, (i) TMD simulations performed to examine the substrate binding process and to understand the structural changes in *E.coli* **MurD**, a three –domian ATP hydrolysis driven muramyl ligase<sup>33</sup> and (ii) TMD study of the coupling of the C-loop closure and channel gating in Nicotinic receptors<sup>34</sup> were a few research articles which mention the versatility of this method.

The basic principle of Targeted Molecular Dynamics (TMD) is that a subset of atoms in a particular conformation in a simulation is ushered towards a final ‘target’ structure by applying steering forces. The root-mean-square distance between the target structure and the current coordinates is computed at each time-step (i.e., after first aligning

## Chapter 2: Methodology (Part 3)

the target structure and the current coordinates). The forces that act upon each atom in the system is given by the gradient of the potential function which is expressed as follows:

$$U_{TMD} = \frac{1}{2} \frac{k}{N} [RMS(t) - RMS^*(t)]^2 \quad Eqn (2.5.1)$$

where  $RMS(t)$  is the instantaneous best-fit root mean square distance of the current coordinates from the target coordinates, and  $RMS^*(t)$  linearly evolves from the initial RMSD between the current and the target coordinates at the first TMD step to the final RMSD at the last TMD step. The spring constant ' $k$ ' is scaled down by  $N$  the number atoms in the system (which must be the same for the current and the targeted structure). TMD module of NAMD has been used for our studies where transition of xCT conformation from inward open, substrate free Ci (as initial structure) to outward open, substrate free Co (as target structure) has been studied.

In NAMD, both initial and target structure files are defined as input. The beta column of the TMD structure file can be used to assign integer values so that the atoms can be separated into non-overlapping constraint domains. The forces on the atoms belonging to one domain will be calculated independently of the other domains. If the occupancy column in the TMD file is non-zero then the atoms are biased towards the target structure and if the altloc field is non-zero then the atoms are fitted to the target structure and within each domain, the atoms which are fitted to the target structure can be different from that which are biased. Langevin dynamics and harmonic restraints are required if the forces acting on the system neither conserves energy or momentum, i.e., if the spring constant is not the same for all the target atoms within a domain.

The parameters necessary for a TMD run in NAMD are the follows<sup>35</sup>:

- (i) **TMDk**: the elastic constant for the TMD forces on the atom, if this is not mentioned then the occupancy value for the atoms in the TMD file will be used as a constant for that atom.

## Chapter 2: Methodology (Part 3)

- (ii) **TMDOutputFreq:** This parameter sets the frequency of the output or how often the TMD outputs, which are the *targetRMS*, *currentRMS*, *ts* (timestep) are printed in the output file.
  
- (iii) **TMDFile:** The user can specify the target structure PDB file in this option. The file must contain the same number of atoms as present in the initial structure file and also the atoms present must have the same index as the structure file. The target structure coordinates are obtained from the targeted atoms in the file and the non-targeted atoms are ignored.
  
- (iv) **TMDFirstStep:** The parameter which commands when the TMD run should start, i.e the user can specify when the TMD run should start in a production run.
  
- (v) **TMDLastsStep:** It supervises the program when the TMD run should terminate, in other words, TMD forces are only applied between the *TMDFirstStep* to *TMDLastStep*.
  
- (vi) **TMDInitialRMSD:** It is necessary to have a reference RMSD value (in Å) which is obtained by aligning the initial structure with the target structure. The target RMSD linearly evolves from this value to the final RMSD value during the TMD run.
  
- (vii) **TMDFinalRMSD:** Depending on whether the system is be steered or away from the target structure, the *TMDFinalRMSD* can be lesser or greater than the *TMDInitialRMSD*. The forces are applied if and only if the  $RMS(t)$  is between the *TMDInitialRMSD* and  $RMS^*(t)$ .

## REFERENCES:

- (1) Becker, O. M.; MacKerell, A. D.; Roux, B. (Biologist); Watanabe, M. *Computational biochemistry and biophysics*; M. Dekker, 2001.
- (2) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234* (3), 779–815.
- (3) Webb, B.; Sali, A. In *Current Protocols in Bioinformatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016; p 5.6.1-5.6.37.
- (4) Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29* (1), 291–325.
- (5) Fiser, A.; King, R.; Do, G.; Andrej, S. **2000**, 1753–1773.
- (6) Pearson, W. R.; Lipman, D. J. **1988**, *85* (April), 2444–2448.
- (7) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
- (8) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. **1997**, *25* (17), 3389–3402.
- (9) Waterhouse, A.; Arnold, K.; Bienert, S.; Bertoni, M. .
- (10) Kiefer, F.; Arnold, K.; Kunzli, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Res.* **2009**, *37* (Database), D387–D392.
- (11) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22* (2), 195–201.
- (12) Guex, N.; Peitsch, M. C.; Schwede, T. *Electrophoresis* **2009**, *30* (S1), S162–S173.
- (13) Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. *Nat. Methods* **2011**, *9* (2), 173–175.
- (14) Shen, M.; Sali, A. **2006**, 2507–2524.

- (15) Melo, F.; Sánchez, R.; Sali, A. *Protein Sci.* **2002**, *11* (2), 430–448.
- (16) John, B.; Sali, A. *Nucleic Acids Res.* **2003**, *31* (14), 3982–3992.
- (17) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
- (18) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. *Biophys. J.* **2009**, *97* (1), 50–58.
- (19) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. *J. Comput. Chem.* **2014**, *35* (27), 1997–2004.
- (20) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. .
- (21) Jo, S.; Kim, T.; Im, W. **2007**, No. 9.
- (22) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
- (23) Andersen, H. C. *J. Comput. Phys.* **1983**, *52* (1), 24–34.
- (24) ———; Forester, T. R.; Smith, W. *J Comput Chem* **1998**, *19*, 102–111.
- (25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781–1802.
- (26) Gumbart, J.; Chipot, C. **2013**, 1–29.
- (27) Harris, R. M.; Goodsell, D. S. **2012**, No. July, 1–8.
- (28) Huey, R.; Morris, G. M.; Forli, S. **2012**.
- (29) Docking, A. J. A.; Autodock, A. J. **2009**, 1639–1662.
- (30) Morris, G. M.; Goodsell, D. S.; Pique, M. E.; Lindstrom, W. L.; Huey, R.; Hart, W. E.; Halliday, S.; Belew, R.; Olson, A. J. *User Guid.* **2009**, 1–49.
- (31) Spiwok, V.; Sucur, Z.; Hosek, P. *Biotechnol. Adv.* **2015**, *33* (6), 1130–1140.
- (32) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *BBA - Gen. Subj.* **2014**, *1850* (5), 872–877.

- (33) Perdih, A.; Kotnik, M.; Hodoscek, M.; Solmajer, T. *Proteins Struct. Funct. Bioinforma.* **2007**, *68* (1), 243–254.
- (34) Cheng, X.; Wang, H.; Grant, B.; Sine, S. M.; McCammon, J. A. *PLoS Comput. Biol.* **2006**, *2* (9), e134.
- (35) Targeted Molecular Dynamics (TMD)  
<http://www.ks.uiuc.edu/Research/namd/2.10b1/ug/node47.html> (accessed Apr 15, 2017).

## CHAPTER 3

### RESULTS

The results will be discussed in two parts. The first part will discuss the results gathered for structural modeling of xCT, membrane building and conventional molecular dynamics simulations that were performed as discussed in the previous chapter. The second part will have the results obtained for the docking simulations and targeted molecular dynamics (TMD) run.

#### PART 1

##### 3.1 COMPARATIVE MODELING RESULTS

This section will comprise of the results from SWISS-MODEL Template Library (SMTL) and MODELLER (homology modeling).

Top 15 hits of homologous proteins obtained using SWISS-MODEL are shown in (Table 3.1.1).

## CHAPTER 3: RESULTS (PART 1)

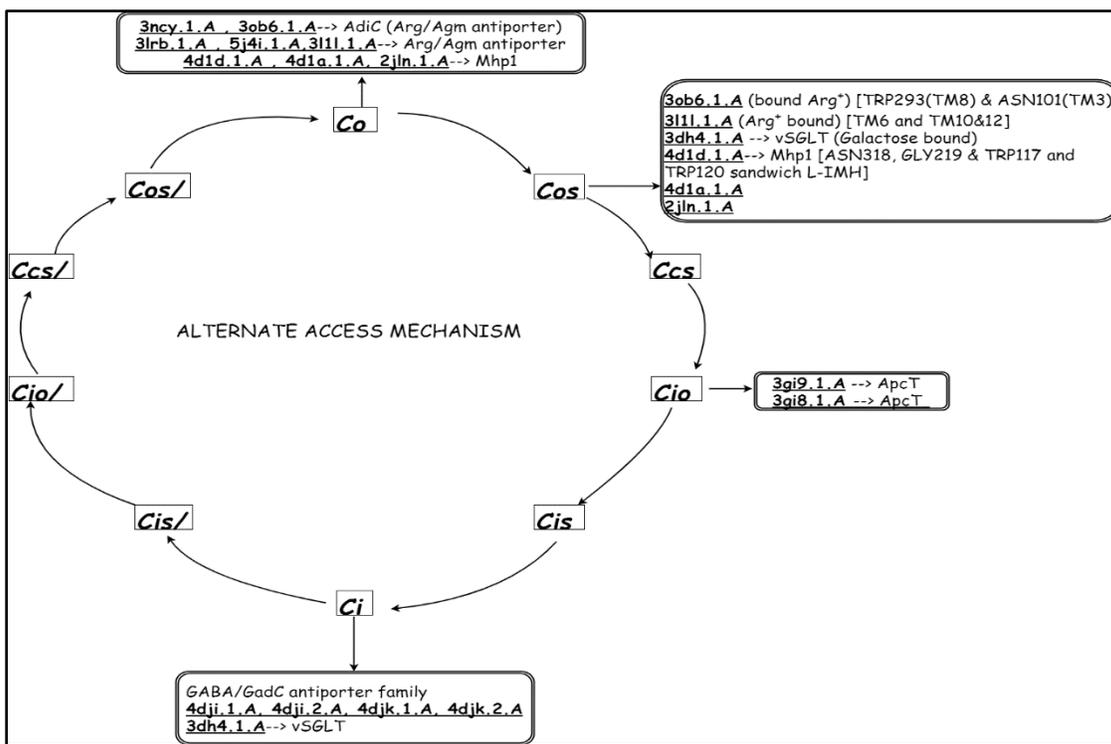
TABLE 3.1.1

Template	Seq. Identity	Oligo-state	Found by	Method	Resolution	Seq. Similarity	Coverage	Description
3ncy.1.A	20.19	hetero-oligomer	HHblits	X-Ray	3.20 Å	0.30	0.84	AdiC
3ob6.1.A	20.43	homodimer	HHblits	X-ray	3.00 Å	0.30	0.83	Adic arginine:agmatine antiporter
4dji.1.A	17.41	monomer	HHblits	X-ray	3.19Å	0.28	0.85	Probable glutamate/gamma-aminobutyrate antiporter
4dji.2.A	17.41	monomer	HHblits	X-ray	3.19Å	0.28	0.85	Probable glutamate/gamma-aminobutyrate antiporter
4djk.1.A	17.41	monomer	HHblits	X-ray	3.10Å	0.28	0.85	Probable glutamate/gamma-aminobutyrate antiporter
4djk.2.A	17.41	monomer	HHblits	X-ray	3.10Å	0.28	0.85	Probable glutamate/gamma-aminobutyrate antiporter
3lrb.1.A	20.48	homo-dimer	HHblits	X-ray	3.61Å	0.30	0.83	Arginine/agmatine antiporter
5j4i.1.A	20.53	homo-dimer	HHblits	X-ray	2.21Å	0.30	0.83	Arginine/agmatine antiporter
3l1l.1.A	20.34	homo-dimer	HHblits	X-ray	3.00Å	0.30	0.82	Arginine/agmatine antiporter
3gi9.1.C	18.29	hetero-oligomer	HHblits	X-ray	2.48Å	0.29	0.82	Uncharacterized protein MJ0609
3gi8.1.A	18.05	hetero-oligomer	HHblits	X-ray	2.59Å	0.29	0.82	Uncharacterized protein MJ0609
3dh4.1.A	10.76	homo-tetramer	HHblits	X-ray	2.70Å	0.25	0.57	Sodium/glucose cotransporter
4d1a.1.A	10.23	monomer	HHblits	X-ray	3.40Å	0.25	0.43	HYDANTOIN TRANSPORT PROTEIN
4d1d.1.A	10.23	Monomer	HHblits	X-ray	3.70Å	0.25	0.43	HYDANTOIN TRANSPORT PROTEIN
2x79.1.A	9.30	Monomer	HHblits	X-ray	3.80Å	0.24	0.43	HYDANTOIN TRANSPORT PROTEIN

## CHAPTER 3: RESULTS (PART 1)

There were, however, a few PDB ID's repeated in (Table. 3.1.1) which were found using BLAST search method (as mentioned in homology modeling part in chapter 2) and also crystallized at different resolutions.

Among these, initially, PDB ID: 3GI9.1.C<sup>1</sup> which is Na<sup>+</sup> independent amino acid secondary transporter belonging to the Amino acid polyamine organo-cation transporter ApcT, was chosen as the template since the literature review of ApcT transporters suggested its amino acids alignment and secondary structures had conserved features in LeuT and human xCT. Later, 4DJI.1.A<sup>2</sup> (Ci) and 3NCY.1.A<sup>3</sup> (Co) (Fig. 3.1.1) were chosen as templates based on the sequence identity, target coverage and sequence similarity (Table. 3.1.1) values.

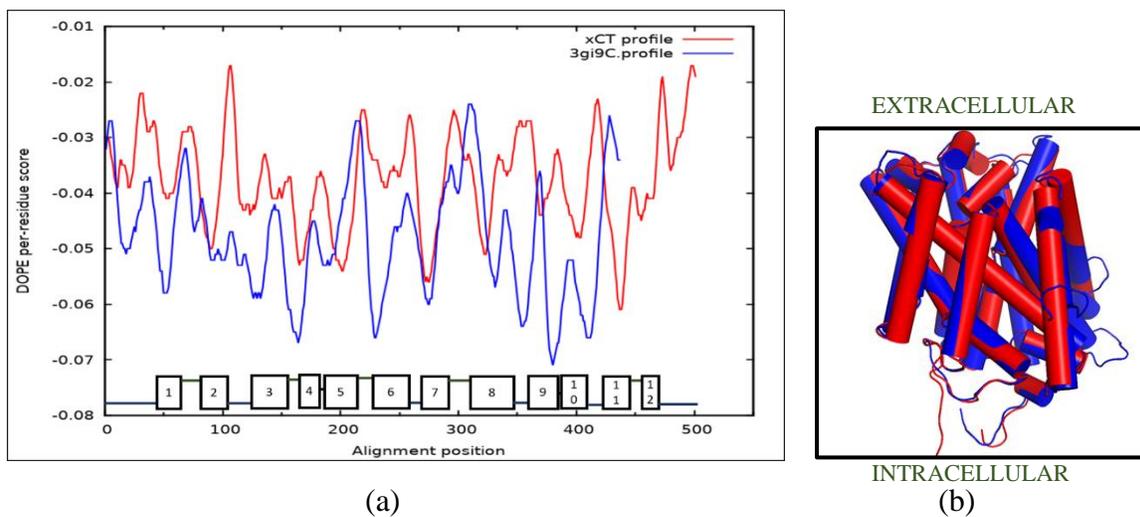


*Fig 3.1.1: The hits from SWISS-MODEL (Table 3.1.1) grouped into different conformations present in the alternate access mechanism (Fig.1.3.1) cycle based on literature survey of each of them, which mentioned the conformation in which they were crystallized.*

Results from MODELLER (homology modeling)

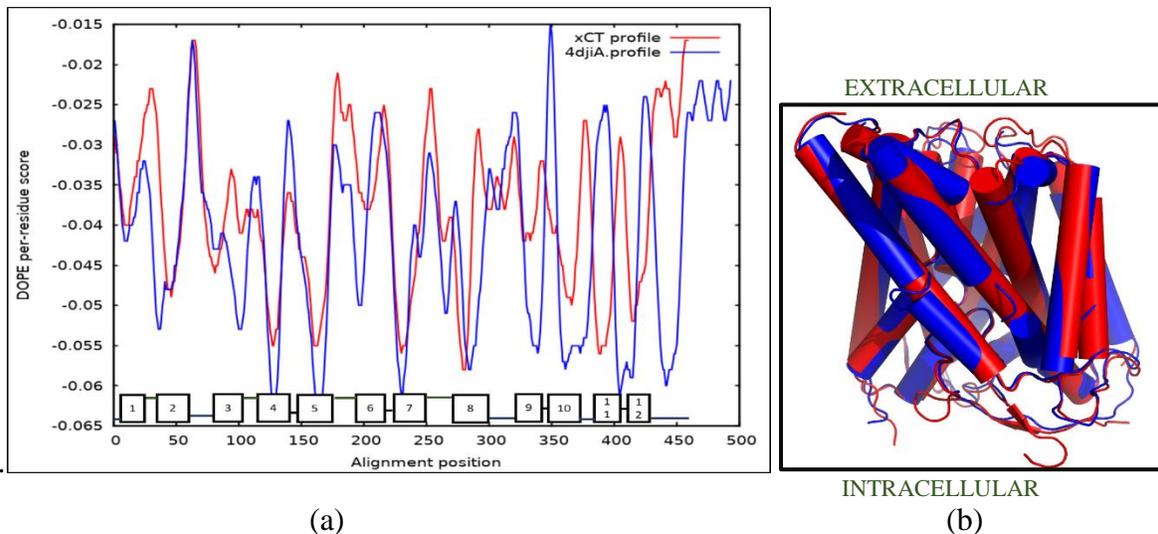
Homology modeling was performed using one PDB ID (Fig 3.1.1) each corresponding to Co, Cio and Co conformations as templates and xCT as the target. Initially, for Cio ,3gi9C was used as a template and 10 models were generated. The model with the lowest DOPE (section 2.1, chapter 2) was chosen as the best model for further analysis.

For Ci and Co conformations, 1000 models were generated, and similarly, the DOPE score was calculated for each one of the models and the model with the lowest DOPE score was selected for refinement, docking and enhanced MD simulations.

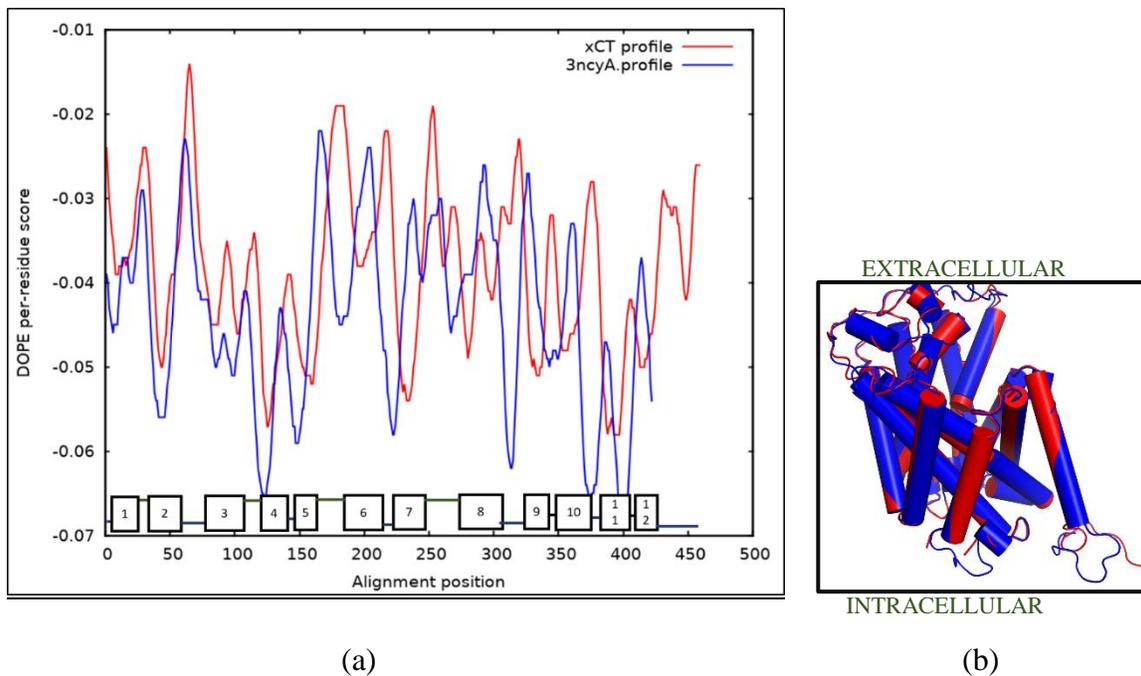


*Fig 3.1.2:(a); DOPE profile of both target and template showing regions of crests (high energy/DOPE score) near the “structurally variable” loop regions (brown and blue dashed lines, depicting inner and outer membrane regions respectively) and the long C-terminal ‘tail’ of the target, and troughs (low DOPE score) near the “structurally conserved” regions (transmembrane helices). Using **GNUPLOT**<sup>4</sup> (b); Structural Alignment of xCT (target) and 3gi9C (template) showing xCT colored in blue and 3gi9C colored in red; **RMSD: 0.276 Å**. Using **PyMOL**.<sup>5</sup>*

CHAPTER 3: RESULTS (PART 1)

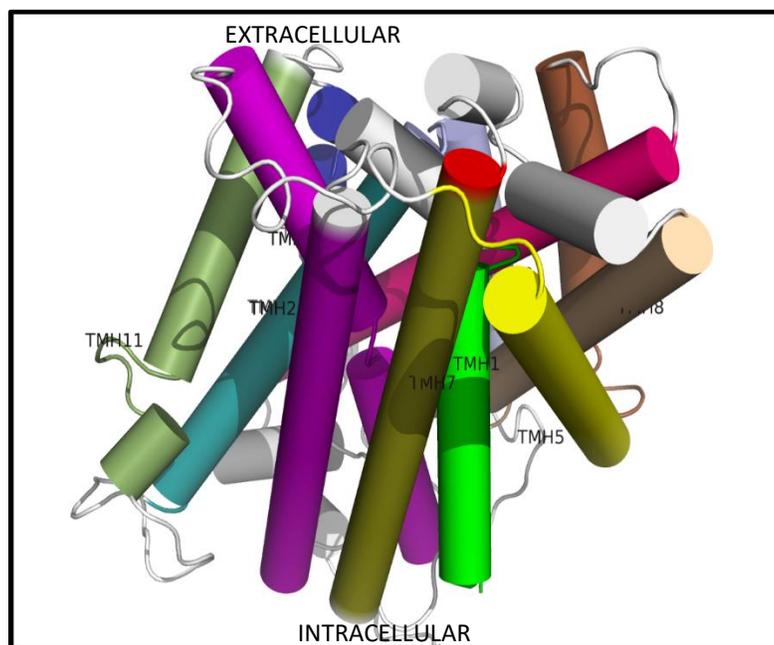


*Fig 3.1.3: (a) DOPE profile for Ci using template as 4djiA.pdb. 1000 models were generated, hence more peaks and troughs are seen. **GNUPLOT.** (a) Structural Alignment of xCT (target) and 4djiA (template) showing xCT colored blue and 4djiA colored red ; **RMSD : 1.441 Å.** Using **PyMOL.***

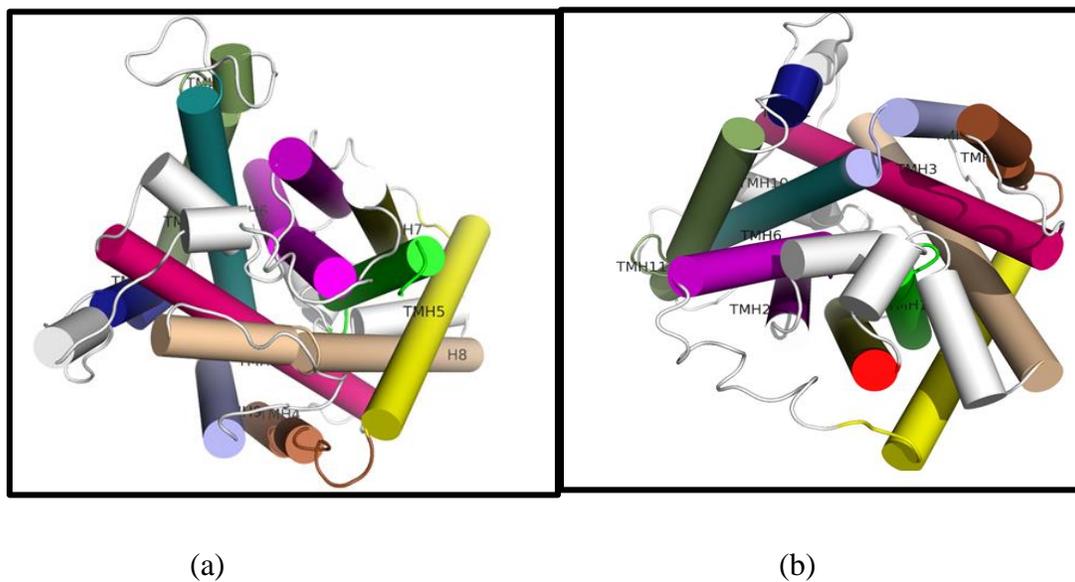


*Fig 3.1.4: (a) DOPE profile for Co using 3ncyA.pdb as a template. Similar to (Fig. 3.132). 1000 models were generated. **GNUPLOT.** (b) Structural alignment of xCT (target) and 3ncyA (template) showing xCT in blue color and 3ncyA in red color; **RMSD : 0.515 Å.** Using **PyMOL.***

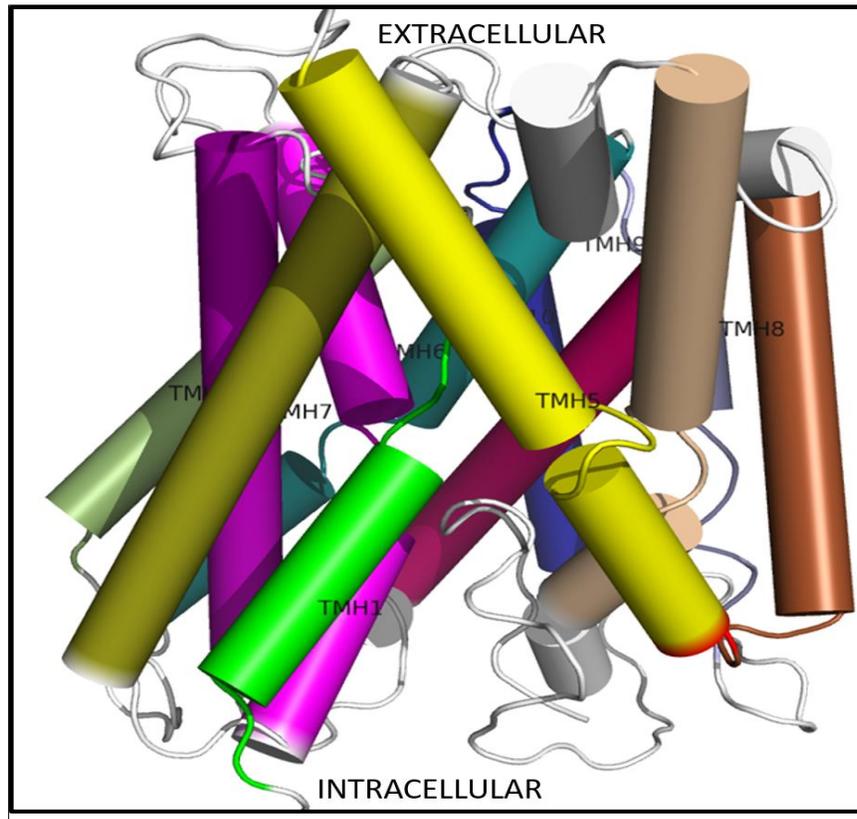
Structures of the Ci, Cio and Co Models:



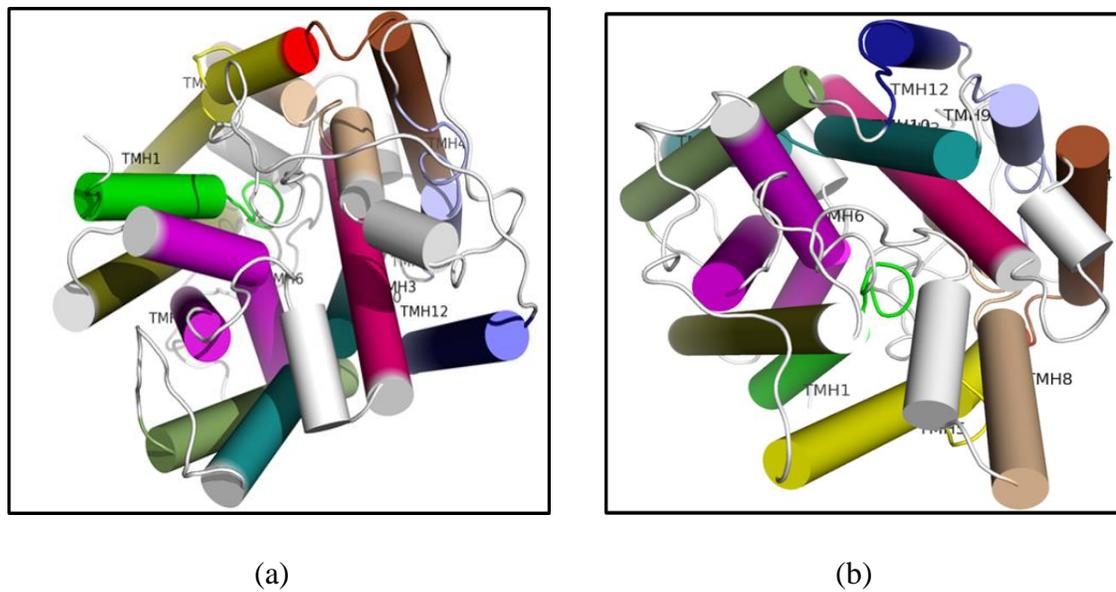
*Fig 3.1.5 : The overall view of Cio modelled structure, animated using PyMOL*



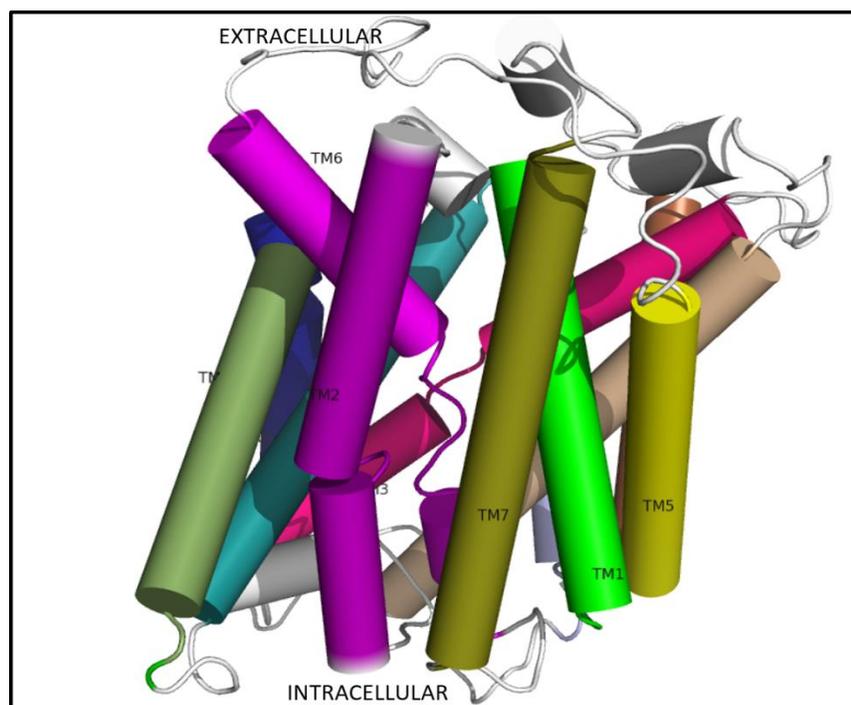
*Fig 3.1.6: (a) Intracellular View ; (b) Extracellular View*



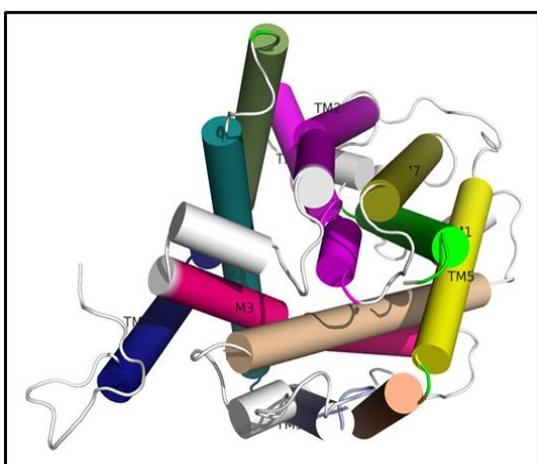
*Fig 3.1.7 : The overall view of Ci modelled structure. Using **PyMOL***



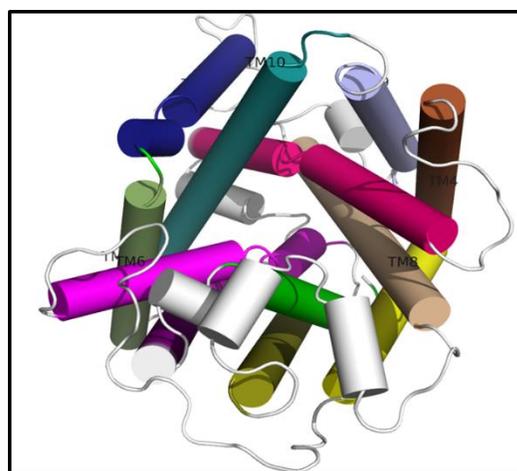
*Fig. 3.1.8: (a) Intracellular View ; (b) Extracellular View*



*Fig 3.1.9 : The overall view of Ci modelled structure. Using **PyMOL***



(a)

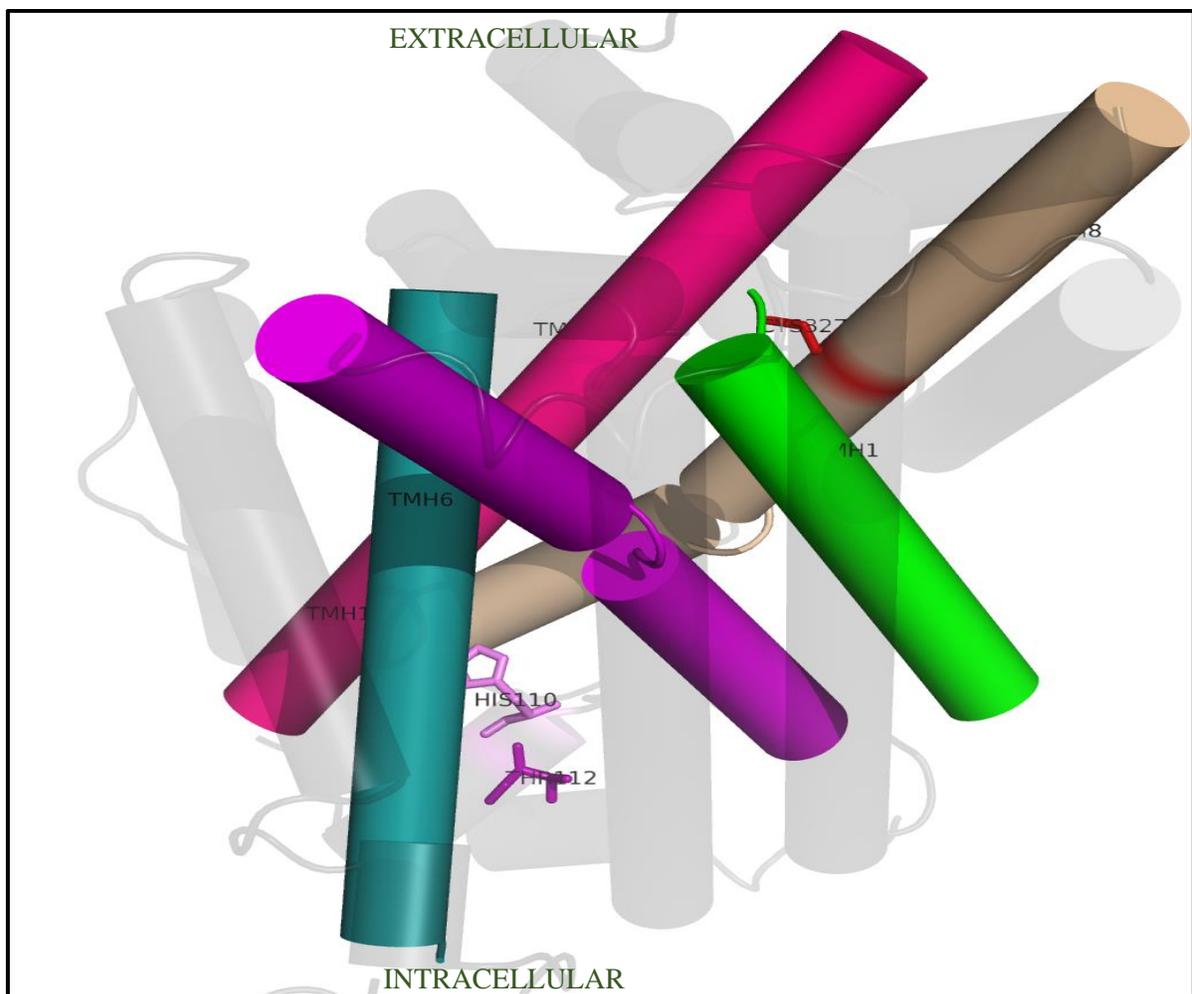


(b)

*Fig 3.1.10: (a) Intracellular View;(b) Extracellular View*

### Pore Lining Helices and Critical Residues

According to the literature survey, experiments have shown that transmembrane helices (TMH 1, TMH3, TMH6, TMH8 and TMH10) form the pore for substrate permeation and have identified the residues responsible for uptake of L-Cys<sub>2</sub>.



*Fig 3.1.11: This an example of pore lining in Cio structure. Pore helices are TMH1 , 3 , 6 , 8 & 10 and the residues which have been identified are HIS 110, THR 112 and CYS 327. Using PyMOL*

## 3.2 MEMBRANE BUILDING & MD REFINEMENT

### Assembly of the Membrane/Protein System.

For the xCT (Cio conformation) and membrane system, the assembled components are;

Waterbox generation : 20022 water molecules

Lipids : Upper Leaflet : Total = 80 ; Lower Leaflet : Total = 80

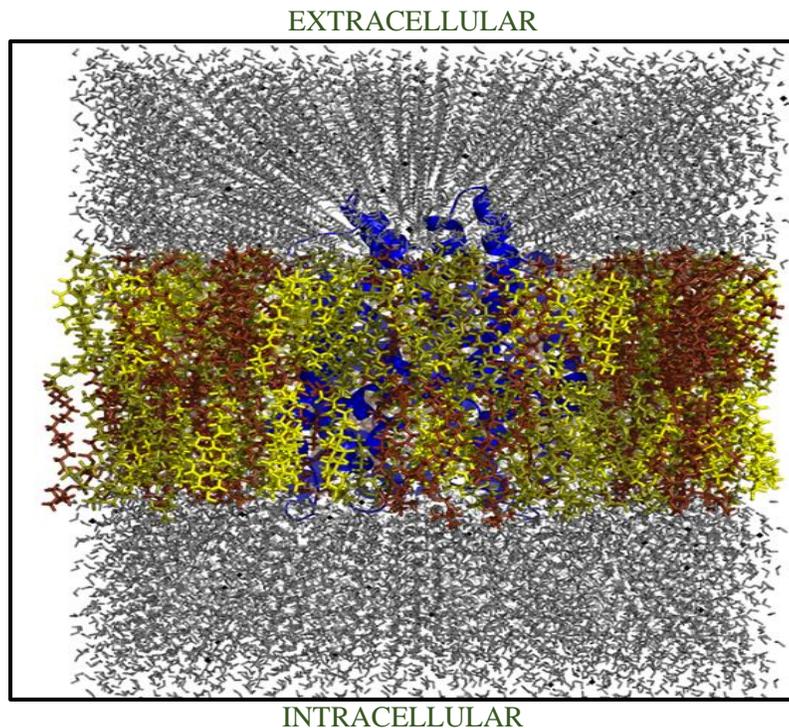
Cholesterol (CHL1) :5 ; Ratio (UL:LL) = 1:1

1,2-Dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC): 50 ; Ratio(UL:LL) = 1:1

1-Palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC): 25 ; Ratio (UL:LL) = 1:1

Carried out Neutralization to generate ions ( $K^+$  and  $Cl^-$ ); Total Ions: 120

$N(K^+) = 54$  and  $N(Cl^-) = 66$



*Fig 3.2.1: The assembled xCT (Cio) membrane /protein complex. Using **PyMOL***

## CHAPTER 3: RESULTS (PART 1)

For the xCT (Ci conformation) and membrane system, the assembled components are;

Waterbox generation : 17244 water molecules

Lipids: Upper Leaflet : Total = 114 ; Lower Leaflet : Total = 114

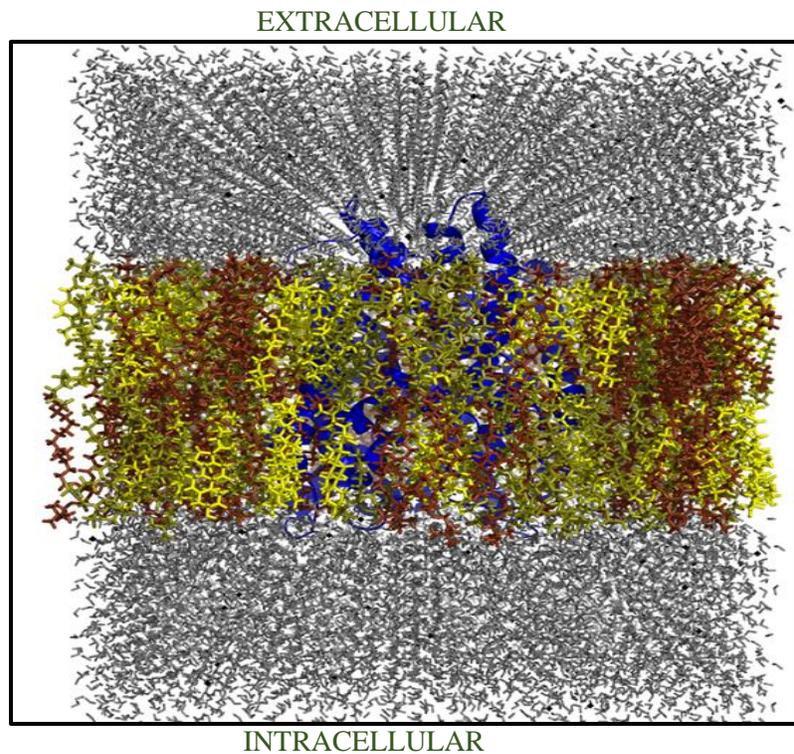
Cholesterol (CHL1) :38 ; Ratio (UL:LL) = 1:1

Dipalmitoylphosphatidylcholine (DPPC): 38 ; Ratio(UL:LL) = 1:1

1-Palmitoyl 2-oleoyl phosphatidylcholine (POPC): 38 ; Ratio (UL:LL) = 1:1

Carried out Neutralization to generate ions ( $K^+$  and  $Cl^-$ ); Total Ions: 104

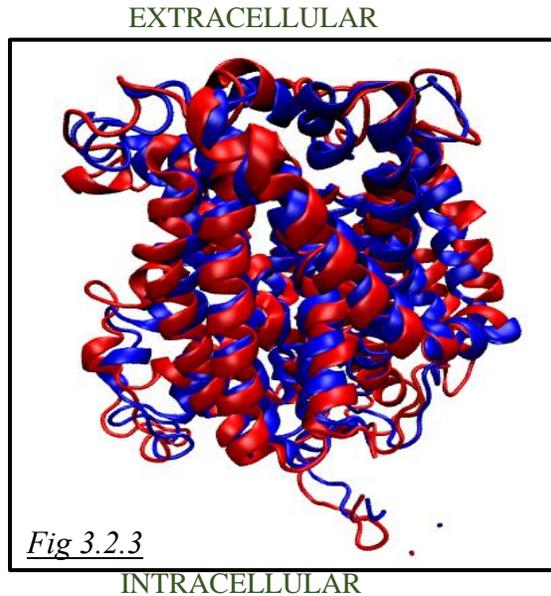
$N(K^+) = 49$  and  $N(Cl^-) = 55$



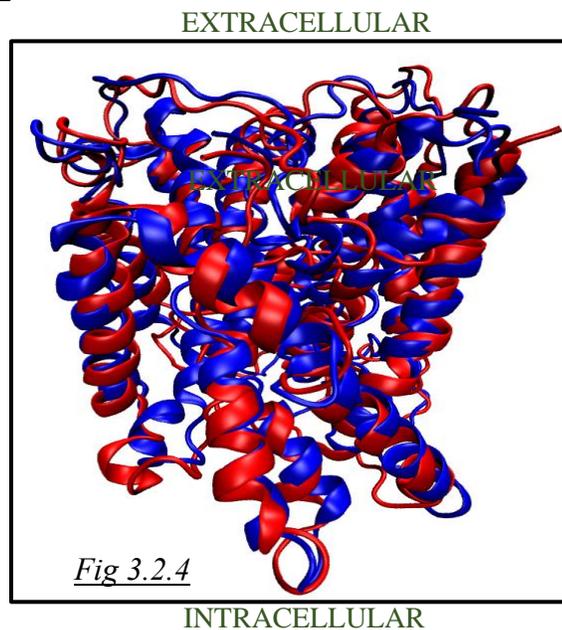
*Fig 3.2.2: The assembled membrane /protein complex for xCT in the Ci conformation. Using PyMOL.*

## CHAPTER 3: RESULTS (PART 1)

The MD equilibration / relaxation over time causes the transmembrane helices (TMH) to shift, hence we see differences in the structural alignment of the TMH's in (Fig 3.2.3, Fig 3.2.4) and can be measured using the RMSD from the pre-equilibrated and post equilibrated structures shown below.



xCT(pre-equilibration) → Blue  
xCT(post-equilibration) → Red

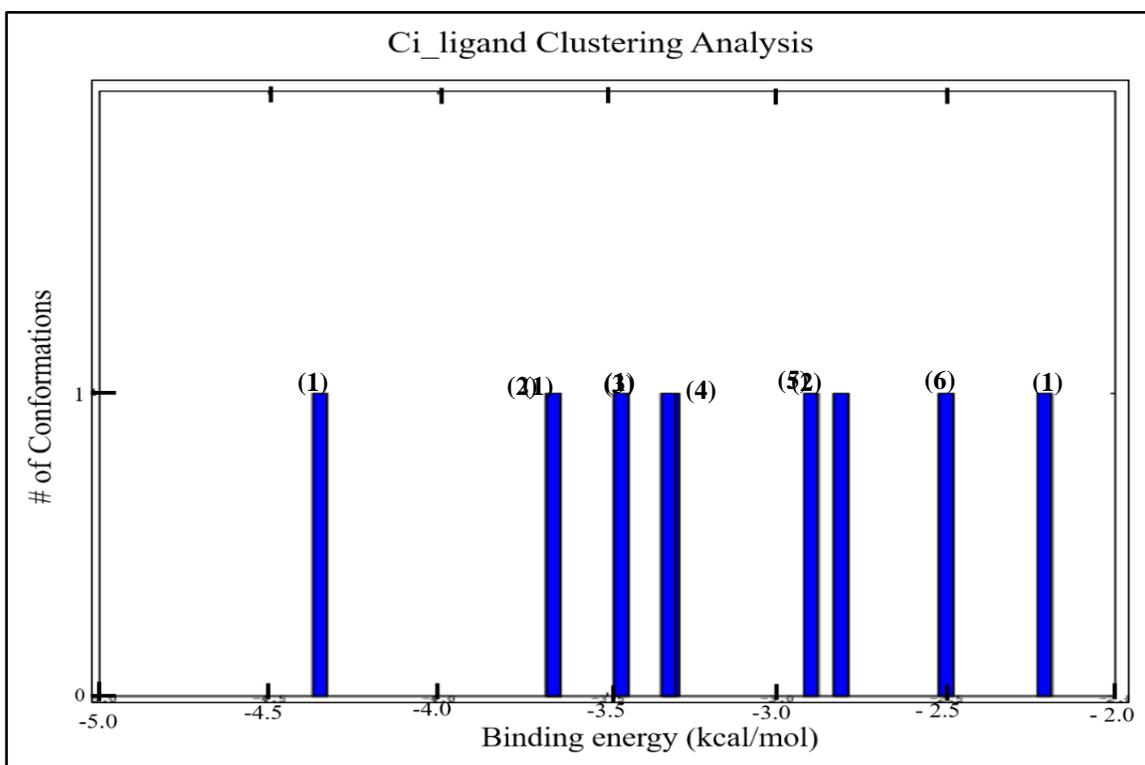


(Fig 3.2.3, 3.2.4) : The structural alignment of xCT (after membrane building) as shown in blue animation and xCT (after MD run) as shown in red animation for Cio and Co conformations respectively. The RMSD's are  $2.969 \text{ \AA}$  and  $2.663 \text{ \AA}$  respectively. Using **VMD**<sup>6</sup>.

PART 2:3.3 DOCKING RESULTS USING AUTODOCK:

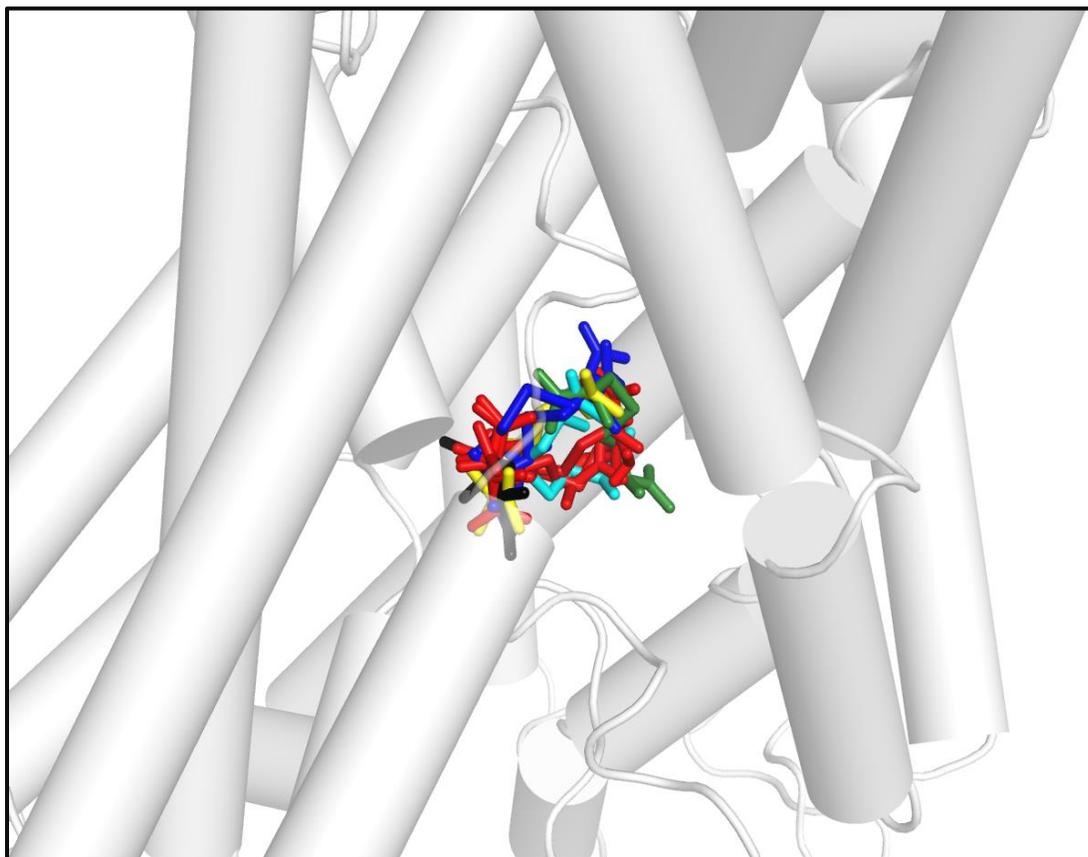
For the docking simulations, the ligand, L-cystine (L-Cys<sub>2</sub>) was chosen from PubChem \* and then converted to its anionic form (L-Cys<sub>2</sub><sup>-</sup>). The torsional degrees of freedom for L-cystine was 10 (calculated as the total number of possible torsions in ligand minus the torsions that only rotate hydrogen bonds), this is a very critical value which is used for predicting the different bound conformations of the ligand in the receptor.

After docking was completed using AutoDock, clustering analysis was carried out to group conformations into clusters based on the mean binding energy. (Fig 3.3.1 and Fig 3.3.2) are the Cluster Histograms which are the output of this calculation done using AutoDockTools.

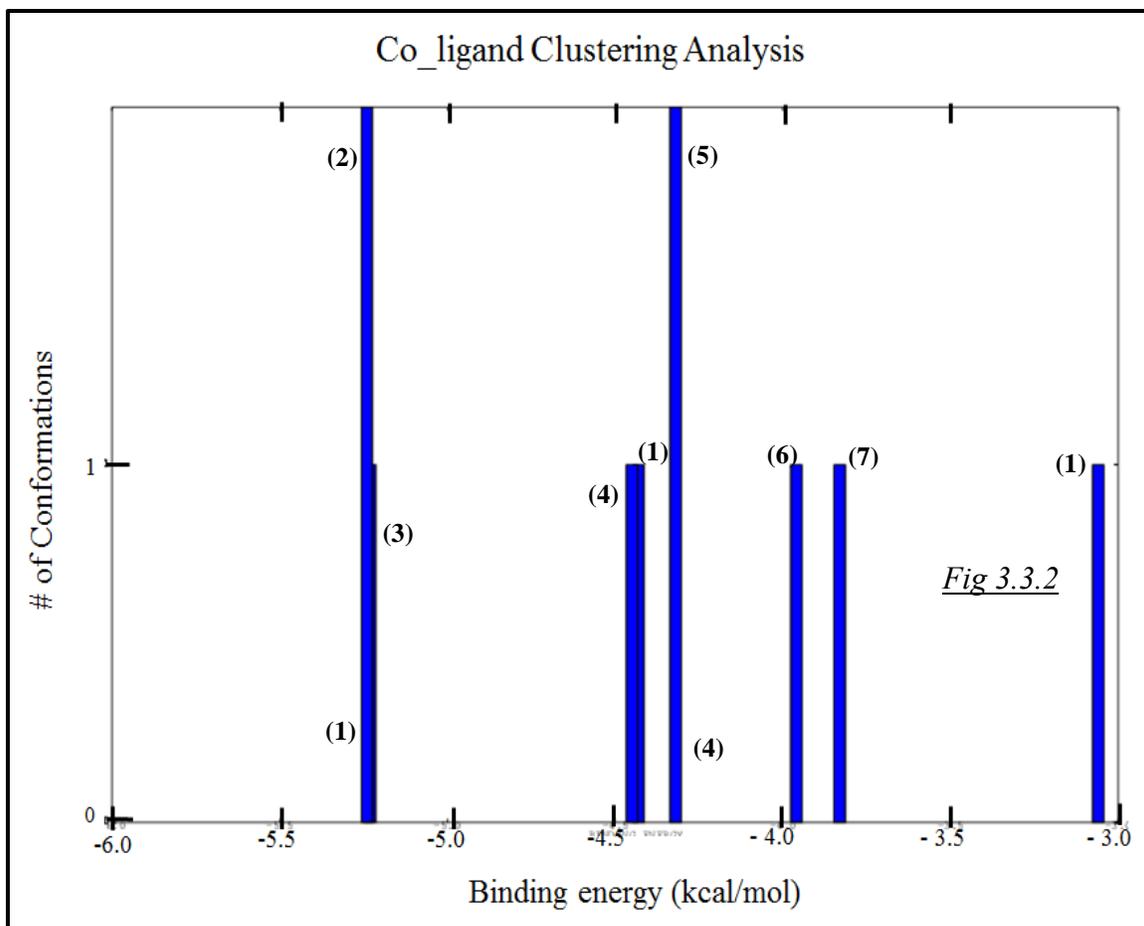


*Fig 3.3.1: Depicts the clustering analysis done for xCT (in Ci conformation) and L-Cys<sub>2</sub><sup>-</sup>, the different conformations have been further grouped into six groups, (1), (2), (3), (4), (5) and (6) based on their binding poses in the binding pocket region in the receptor. Using PMV<sup>7</sup>.*

Binding Pocket for the ligand in xCT (Ci conformation)

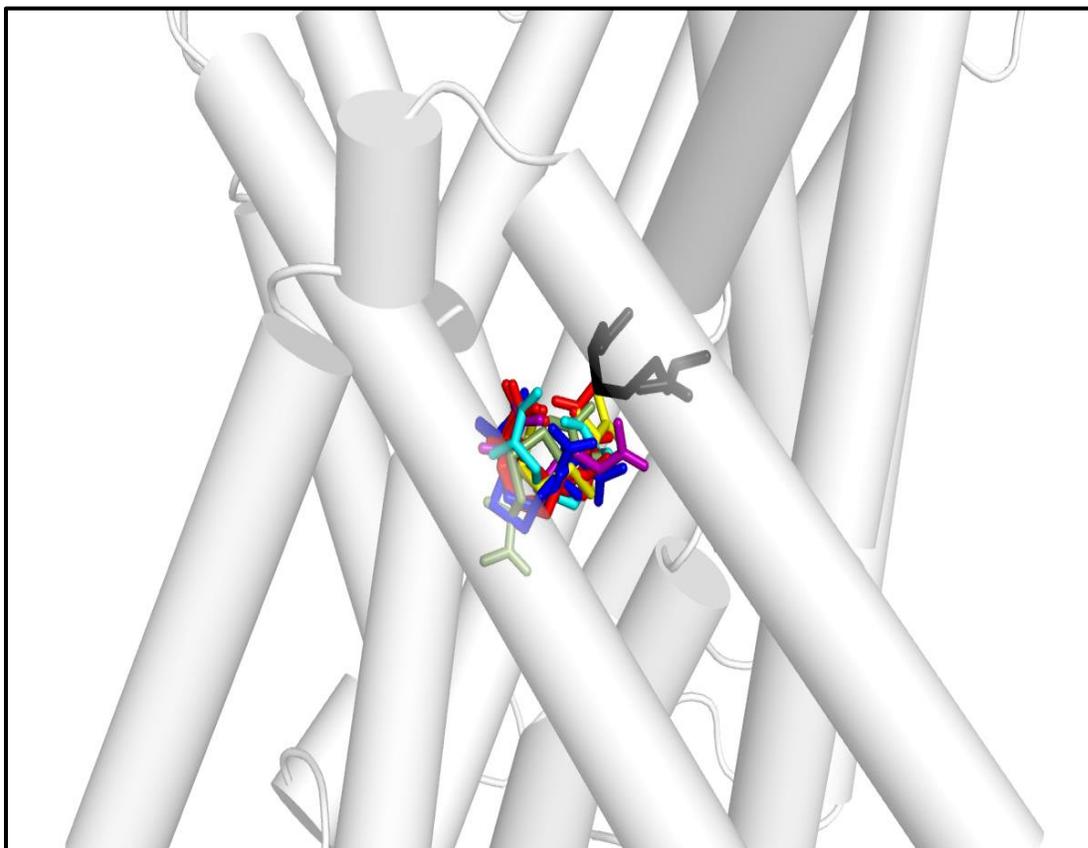


*Fig. 3.3.2: Different groups shown in (Fig. 3.3.1) have been animated with different colors, Group (1) includes the ligand conformations {0,1,3,9} is colored as red; Group (2) includes conformation {2,7} is colored blue; Group (3) includes conformation {4} which is colored black; Group (4) includes conformation {5} colored yellow; Group (5) includes conformation {6} colored cyan and finally Group (6) which includes conformation {8}. Using **PyMOL***

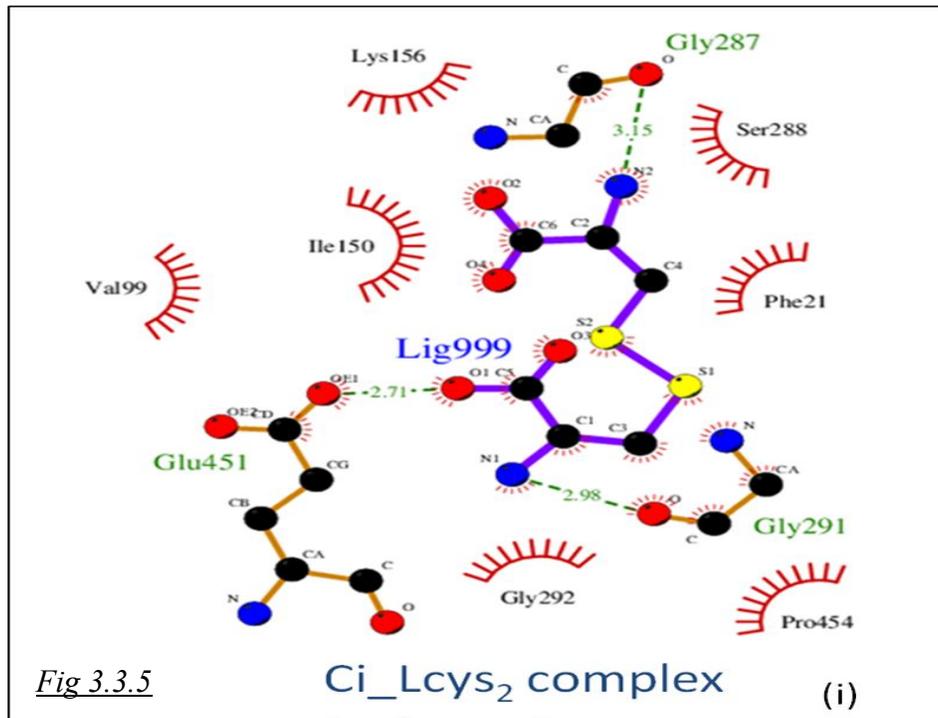


*Fig 3.3.3: The clustering analysis done for xCT (in Co conformation) and L-Cys<sup>2-</sup>, similar to (Fig. 3.3.1) conformations have been grouped into seven groups, (1), (2), (3), (4), (5),(6). Using PMV*

Binding Pocket for the ligand in xCT (Co conformation)



*Fig. 3.3.4: Groups in (Fig. 3.3.3) have been animated with different colors, Group (1) includes the ligand conformations {0,4, 9} colored red; Group (2) includes conformation {1} colored smudge; Group (3) includes conformation {2} colored purple; Group (4) includes conformation {3,5} colored blue; Group (5) includes conformation {6} colored yellow; Group (6) which includes conformation {7} colored cyan and finally Group (7) includes conformation {8} colored black (quite isolated from the pocket). Using **PyMOL***



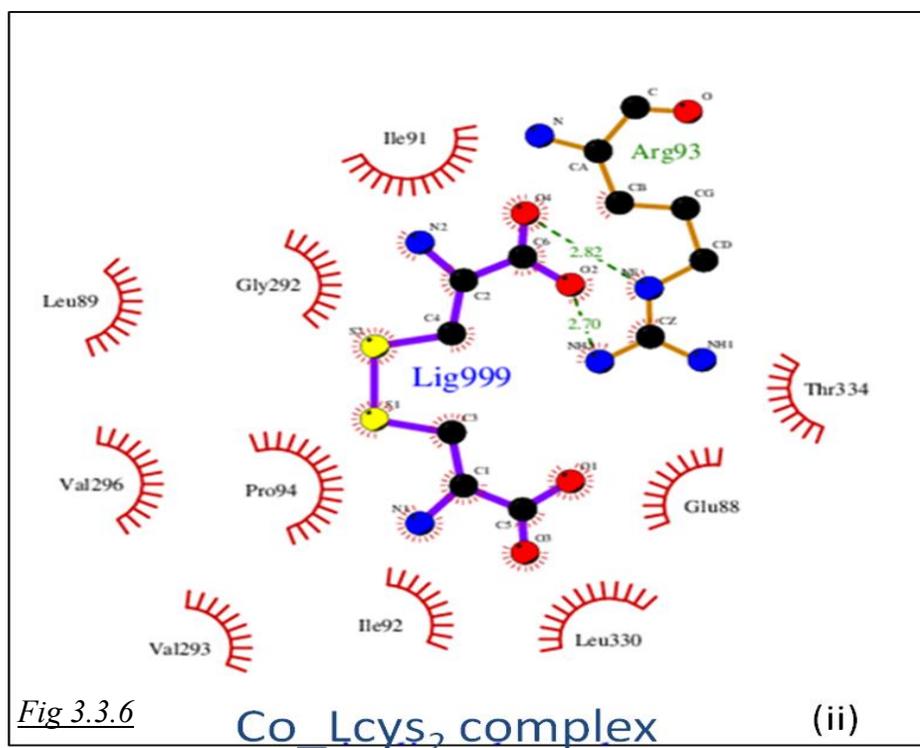
(Fig. 3.3.5): *L-Cys<sub>2</sub>* and *xCT* interactions (*H*-bonded and hydrophobic) for *Ci* conformations. The list of the non-bonded (hydrophobic) interactions and Hydrogen bonded interactions are listed along with the transmembrane helix to which they belong. Using LigPlot<sup>+</sup><sup>8</sup>.

Close Contacts (Hydrophobic Interactions)

1. PHE21 → TMH1
2. VAL99 → TMH3
3. ILE150 → TMH5
4. LYS156 → TMH5
5. GLY287 → TMH8
6. SER288 → TMH8
7. GLY291 → THM8
8. GLY292 → TMH8
9. GLU451 → IL7
10. PRO454 → IL7

Hydrogen Bonding

5. GLY287
7. GLY291
9. GLU451



(Fig. 3.3.6): L-Cys<sub>2</sub> and xCT interactions (H-bonded and hydrophobic) for Ci conformation. The interacting residues have been listed below. Using LigPlot<sup>+</sup>

Close Contacts ( Hydrophobic Interactions)

1. GLU88 → TMH3
2. LEU89 → TMH3
3. ILE91 → TMH3
4. ILE92 → TMH3
5. **ARG93** → TMH3
6. PRO94 → TMH3
7. GLY292 → TMH8
8. VAL293 → TMH8
9. VAL296 → TMH8
10. LEU330 → TMH9
11. THR334 → TMH9

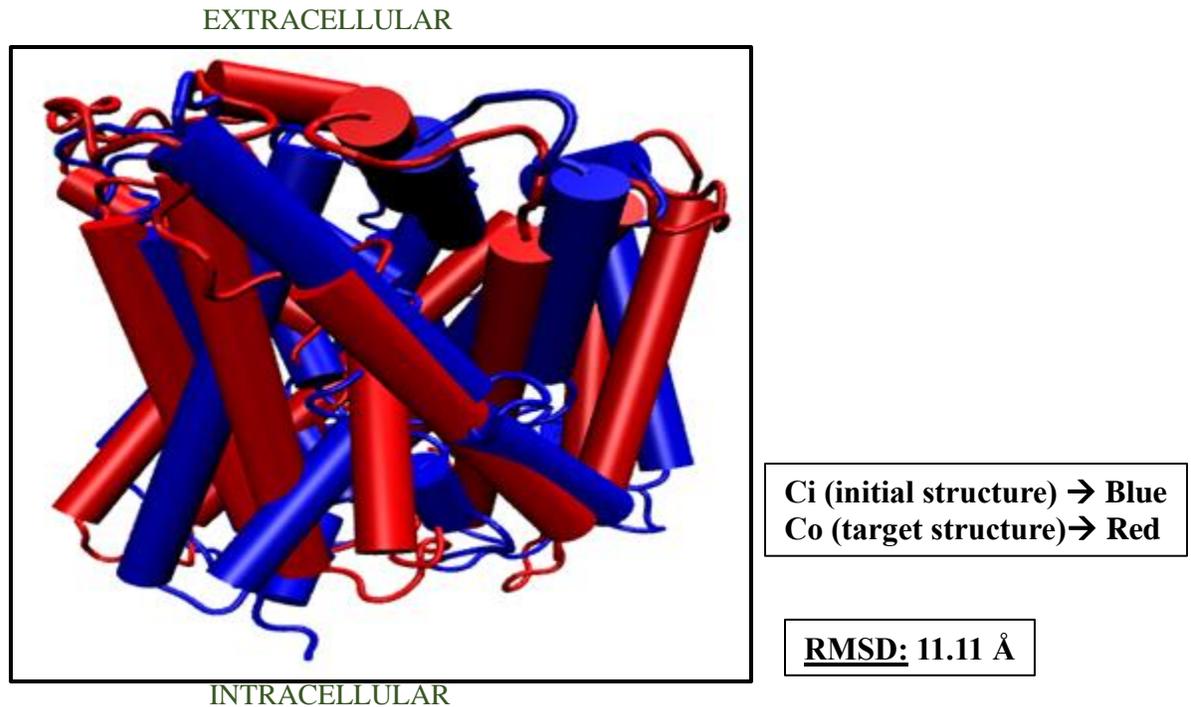
Hydrogen Bonding

5. **ARG93**

### 3.4 TARGETED MOLECULAR DYNAMICS RESULTS:

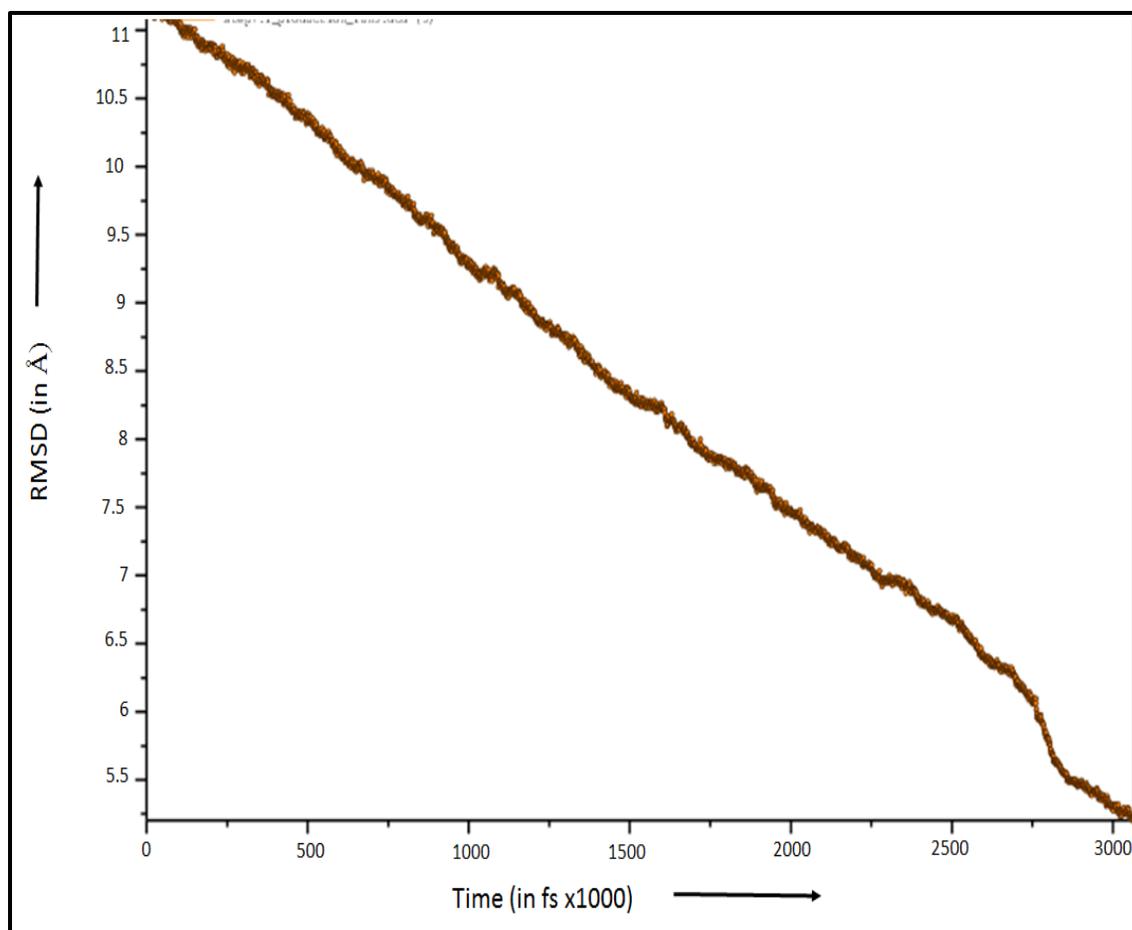
TMD was used to investigate the transition pathways from Ci (initial conformation) to Co (the target structure). Since the two conformations are very different in a sense, the former is an inward (open to the cytoplasm) open conformation and the latter is an outward (open to the extracellular membrane) conformation, the TMH's are also arranged differently.

The structural alignment of the two structures before the TMD run in (Fig. 3.4.1) shows how different the two conformations are relative to each other along with the mentioned RMSD.



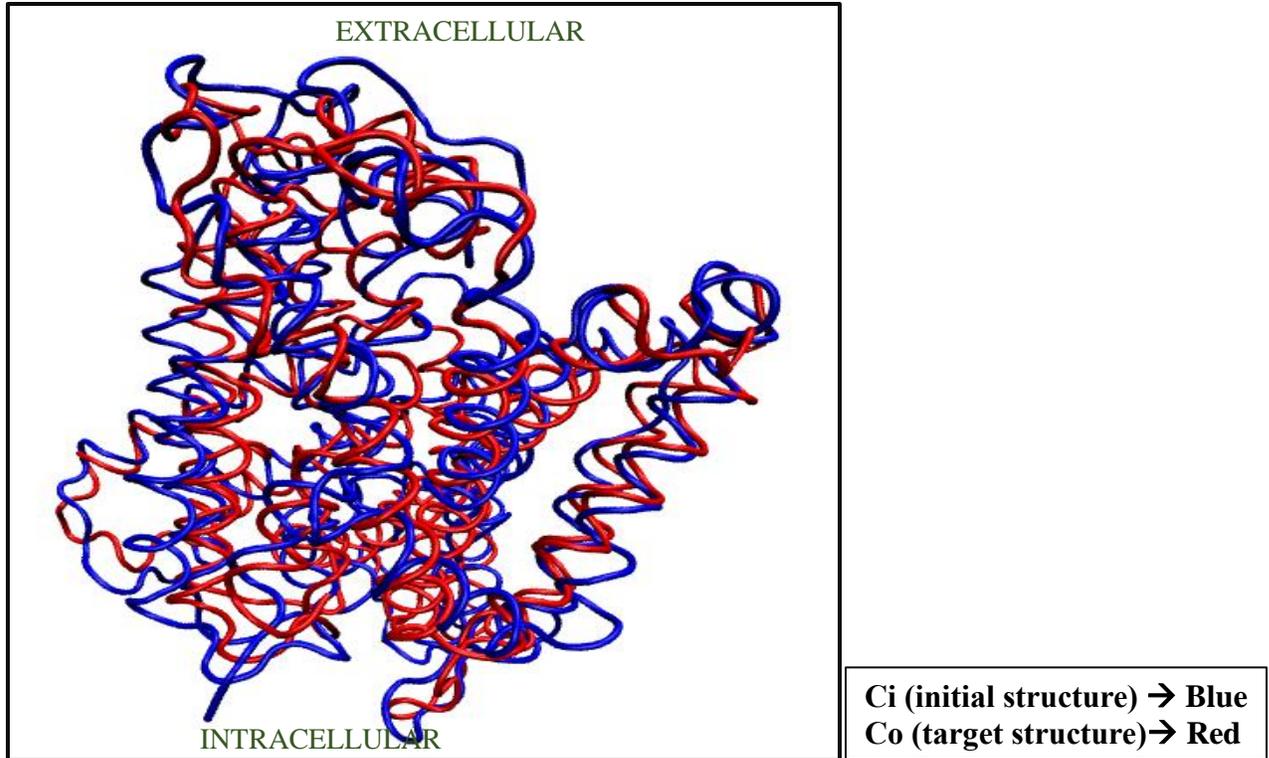
*Fig 3.4.1: The alignment of Ci (initial structure) with Co (final structure) before TMD run. Using VMD.*

## CHAPTER 3: RESULTS (PART 2)



*Fig. 3.4.2: The RMSD between the initial (Ci) structure and target (Co) structure over 3 ns. Avg. = 8.33 Å; Min: 5.2 Å and Max: 11.23 Å. Using VMD*

The coordinates are not being calculated after about 3ns of simulation time and the energies are rapidly increasing after 3ns. The target RMSD, which is 0 Å (as mentioned in the TMD parameters), is not reached and the lowest RMSD as shown in (Fig. 3.4.2) is around 5.2 Å with respect to the target (Co) structure.



*Fig 3.4.3: A structural alignment between Ci and Co conformations after the TMD run. Using VMD.*

### FUTURE WORK:

- Since the RMSD is not going below 5 Å to the desired target value, we need to employ the Umbrella Sampling technique to successively split the RMSD profile into smaller fragments and apply TMD to each of these fragments to capture the entire transition pathway between Ci and Co.
- Once we have the pathway well defined, we can employ methods like Adaptive Biasing Forces (ABF) to pull the substrate along the pathway and capture the interactions.

## CONCLUSIONS:

- (i) Modeling plays a key role in predicting the three-dimensional structure of an unknown protein and obtaining the structure are proper evaluation forms the first step towards understanding the implications to its function.
- (ii) Docking simulations give us an idea of the binding pocket for the ligand, different conformations of the ligand in the pocket and also a theoretical binding energy of the ligand-protein interactions which is usually in good agreement to the experiments.
- (iii) Finally, molecular dynamics acts as a tool to capture these interactions at the atomistic level which can be directly linked to the critical functions of biomolecules. Enhanced MD techniques come in handy at times when the conventional molecular dynamics fails to capture some functional conformations of the biomolecule which may appear at timescales inaccessible to conventional MD due to computational complexity or inaccuracy of the force fields to describe the necessary interactions that correspond to functional aspects of the biomolecule.

## REFERENCES:

- (1) Shaffer, P. L.; Goehring, A.; Shankaranarayanan, A.; Gouaux, E. *Science* (80-. ). **2009**, 325 (5943), 1010–1014.
- (2) Ma, D.; Lu, P.; Yan, C.; Fan, C.; Yin, P.; Wang, J.; Shi, Y. *Nature* **2012**, 483 (7391), 632–636.
- (3) Fang, Y.; Jayaram, H.; Shane, T.; Kolmakova-Partensky, L.; Wu, F.; Williams, C.; Xiong, Y.; Miller, C. *Nature* **2009**, 460 (7258), 1040–1043.
- (4) Janert, P. *Gnuplot in Action*; 2008.
- (5) DeLano, W. L. *Schrödinger LLC wwwpymolorg* **2002**, Version 1., <http://www.pymol.org>.
- (6) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, 14 (1), 33–38.
- (7) Sanner, M. F. *J. Mol. Graph. Model.* **1999**, 17 (1), 57–61.
- (8) Laskowski, R. A.; Swindells, M. B. *J. Chem. Inf. Model.* **2011**, 51 (10), 2778–2786.

