STRUCTURAL DYNAMICS OF ABC TRANSPORTERS

by

Akarun Ayça Ersoy

B.S., Chemical Engineering, Boğaziçi University, 2020

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2022

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my thesis supervisor Prof. Türkan Haliloğlu for her guidance, endless support and positivity. It was a great pleasure and honor to work with her in the past three years.

I am also very grateful to my thesis committee members; Prof. Kutlu Ülgen and Assoc. Prof. Nurcan Tunçbağ for their time, attention, and recommendations.

It was a great opportunity to be supported by Tübitak 1001 Project no. 119F392. I would like to specially thank my friends and lab mates in the Polymer Research Center; Bengi Altıntel, Özge Duman, Yiğit Kutlu, Zeynep Erge Akbaş, and my mentor Burçin Acar. I have learned so much from them and wouldn't have been able to go on without their help. I would also like to thank my friends Ayşe Saliha Korkut and Duygu Kaya for giving me motivation, support, and coffee.

I would also like to thank all of my friends in the Boğaziçi University Mountaineering Club, who all make Boğaziçi and the world a better place. A special thanks goes out to Cansu, Berkay, Songül and Ali Budak for their support, friendship, and encouragement.

Lastly, I am grateful for my amazing parents whom I love and respect endlessly. They listen to my countless opinions and problems even if they don't want to, and offer solutions even when I am not willing to listen. This thesis is dedicated to my father who taught me to plan meticulously, and to my mother, who always knows what is best.

ABSTRACT

STRUCTURAL DYNAMICS OF ABC TRANSPORTERS

The ATP-binding cassette transporters (ABC transporters) translocate substrates across membranes by ATP binding and hydrolysis. To understand the complex structural dynamics, transporters were investigated on three case studies: heterodimeric multidrug exporter TmrAB, cystic fibrosis transmembrane conductance regulator (CFTR), and manganese ion importer PsaBC. Anisotropic Network Model-Langevin Dynamics (ANM-LD) simulations were performed between four different conformations of TmrAB. The analysis for each transition revealed the allosteric couplings between the ATP binding in nucleotide binding domains (NBD) and extracellular gates on transmembrane domains (TMD). Transfer Entropy (TE) calculations showed that TMDs and NDBs of TmrAB become drivers at different stages during the translocation cycle. The ANM-LD simulations of dephosphorylated to phosphorylated CFTR transition revealed that the initial movements in the NBDs are followed by the movement of the gating areas in TM. The causal relationships between these movements were supported by TE calculations. CFTR potentiator drug Ivacaftor has been shown to have a global effect, mimicking the effects of ATP binding and gating residues. Corrector drug Lumacaftor has a local effect on TM1, 2, 3, and 6. The dynamics of PsaBC was characterized using TE calculations, and testable predictions were made. Four possible gating residues were identified along with an allosteric pocket on TM8. Finally, TE calculations of many ABC transporters showed the differences and similarities between different types. Type I transporters were shown to be controlled by TMDs in the inward facing (IF) conformation, and NBDs and substrate binding protein (SBP) in the outward facing (OF) conformation. Type II transporters were always driven by NBDs, regardless of conformation. Type IV transporters showed more variability, but for most, TMDs are driving for IF and NBDs are driving for OF conformations. Dynamic characterization methods show potential in classifying ABC transporters.

ÖZET

ABC TAŞIYICILARININ YAPISAL DİNAMİKLERİ

ATP-bağlı kaset (ABC) taşıyıcı proteinleri, ATP bağlanma ve hidrolizinden elde ettikleri enerji ile hücresel zarlar üerinden madde geçisinde görev yaparlar. Karmaşık yapısal dinamikleri anlamak için üç farklı ABC taşıyıcı protein incelenmiştir: heterodimer peptit taşıyıcı TmrAB, kistik fibrosis transmembrane kondüktans regülatorü (CFTR) ve manganez iyon taşıyıcısı PsaBC. TmrAB proteini için kristal yapıların bolluğu, dört farklı konformasyon arasında Anizotropik Ağ Modeli-Langevin Dinamiği (ANM-LD) simülasyonları gerçekleştirilmesine imkan sağlamıştır. ATP bağlanma rezidüleri ve hücredışı kapılar arasındaki allosterik bağlantıları açığa çıkarılmıştır. Transfer Entropi (TE) sonuçları ise transmembran (TMD) ve nükleotit-bağlanma alanlarının translokasyon döngüsü boyunca olayların sürücüsü olduklarını göstermiştir. CFTR için ANM-LD simülasyonları gerçekleştirilmiştir. NBDlerdeki başlangıç hareketlerinin kapı bölgesi tarafından takip edildiği görülmüştür. Bu bölgeler arasındaki sebepsel ilişki, TE hesaplamaları ile de desteklenmiştir. CFTR ilaçları, ivacaftor ve lumacaftor karşılaştırılmıştır. Ivacaftorun yapı üzerinde ATP bağlanmasını ve kapı hareketini taklit eden global bir etkisi olduğu görülürken, lumacaftorun bağlandığı TM1, 2, 3 ve 6ya etki ettiği görülmüştür. Manganez iyon taşıyıcısı da TE hesapları ile karakterize edilmiş ve deneysel olarak test edilebilecek tahminler yapılmıştır. Yeni kapı rezidüleri tahmin edilmiş ve transmembran helix 8 üzerinde bir allosterik cep bulunmuştur. TE hesapları farklı ABC taşıyıcıları arasındaki dinamik davranış farklarını ve benzerliklerini göstermiştir. Tip 1 taşıyıcılar, içe dönükken TMDler tarafından sürülürken, dışa dönükken NBD ve sübstrat bağlayan protein (SBP) tarafından sürülmektedir. Tip II taşıyıcılar hep NBDler tarafından kontrol edilmektedir. Tip IV taşıyıcılar ise daha çok farklılık göstermekle beraber içe dönükken TMDler, dışa dönükken NBDler tarafından sürülmektedir. Dinamik karakterizasyon metodları, ABC taşıyıcılar için yeni bir sınıflandırma için umut vermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

C_{ij}	Cross-correlation between residues i and j
DF	Deformation factor
$\mathbf{d}_i$	Distance between the coordinates of atom i in two structures
$\mathbf{D}_i$	Distance difference vector
$\mathbf{H}$	Hesiian matrix
k	Mode
k_B	Boltzmann constant
m_i	Most overlapping ANM mode for cylce i
$modemax$	Maximum number of modes in the mode pool
N	Number of atoms or residues
r_c	Cut-off radius
$\mathbf{R}_i$	Position vector of residue i
t	Time
T	Absolute temperature
$\mathbf{u}_k$	kth eigenvector
$\mathbf{U}$	Matrix of eigenvectors
X_i	X coordinate of residue i
Y_i	Y coordinate of residue i
Z_i	Z coordinate of residue i
Δ	Change in a parameter
γ	Spring force constant
$\mathbf{\Gamma}$	Kirchoff (or connectivity) matrix
λ	kth eigenvalue
τ	Time delay

LIST OF ACRONYMS/ABBREVIATIONS

1D	One Dimensional
3D	Three Dimensional
A	Alanine
Adk	Adenylate Kinase
ANM	Anisotropic Network Model
Asym	Asymmetrical Conformation
ATP	Adenosine Triphosphate
BtuCD	Vitamin B12 Transporter
C	Cysteine
CC	Cross Correlations
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CH	Coupling Helix
D	Aspartic Acid
E	Glutamic Acid
ECL	Extracellular Loop
F	Phenylalanine
G	Glycine
GNM	Gaussian Network Model
H	Histidine
I	Isoleucine
ICL	Intracellular Loop
IF	Inward Facing Conformation
K	Lysine
L	Leucine
LD	Langevin Dynamics
M	Methionine
MalFGK2	Maltose Transporter Protein
MD	Molecular Dynamics
MntBC	Manganese Transporter Protein

N	Asparagine
NBD	Nucleotide Binding Domain
OF	Outward Facing Conformation
P	Proline
PDB	Protein Data Bank
PsaBC	Bacterial Manganese Transporter
Q	Glutamine
R	Arginine
RD	Regulatory Domain
RMSD	Root Mean Square Difference
S	Serine
SBP	Substrate Binding Protein
T	Threonine
TD-CC	Time Delay Cross Correlations
TE	Transfer Entropy
TECol	Transfer Entropy x Collectivity Score
TMD	Transmembrane Domain
TmrAB	Peptide Transporter Protein
V	Valine
Y	Tyrosine

1. INTRODUCTION

Proteins are macromolecular machines that perform a wide variety of essential cellular tasks. While experimental methods are able to capture the sequence and structure of dominant conformations, their dynamics are still a mystery. Protein dynamics bridge the gap between protein structure and protein function [1]. Each protein goes through many shapes while performing its function. The characterization of these conformations is essential to understanding the mechanism of action. Experimental methods are able to capture snapshots of these events, but cannot describe the behavior along a trajectory. Computational tools offer answers to many topics of interest such as; identification of key functional sites and allosteric pockets, determination of the order-of-events through a trajectory, and the causal relationships between residues.

Even though proteins go through many conformational shapes to perform their function, they spend most of their time in one or two states. Transitional states are very challenging to capture. Only by certain mutations, increased concentrations or added agents, they are able to be captured experimentally. These modifications add bias to results and possibly overstate or understate the importance of certain conformational states [2]. Only recently, it has been possible to capture high energy states in adenylate kinase (Adk), calmodulin and Src kinase in a recent seminal work by Stiller [3]. Protein function is also closely linked with allosteric regulation. Allostery is the phenomenon that energetically links remote sites of a system to create a functional response [4]. Allosteric modulators, such as the binding of a small molecule, leads to a population shift in the proteins conformational ensemble. Understanding the allosteric network of a protein is essential to controlling protein function [5, 6].

Computational simulation methods such as molecular dynamics (MD) are very successful at predicting the conformational dynamics of biological macromolecules; however, for larger biological systems and larger conformational changes between stable conformational states the computational cost significantly increases [4]. To sample the intermediate, high energy states, enhanced sampling MD methods such as metadynam-

ics, simulated annealing, or replica exchange are used [7]. On the other hand, elastic network models significantly simplify the system and make obtaining results much faster [8]. ANM-LD combines the Anisotropic Network Model (ANM) with Langevin dynamics (LD) to search of the conformational space. ANM-LD method uses only the intrinsic dynamic modes of motion accessible for a protein through its functional journey [8, 9]. By choosing dynamic modes in such to enable the transition to the target, this method is able to obtain the intermediate states and identify important dynamic modes for transition. ANM-LD is able to obtain much faster snapshots during the transition period, creating almost a movie of the process [8, 9]. Other computational approaches utilizing MD simulation trajectories [7] or elastic network models [10], such as the transfer entropy analysis reveals causal relationships between events A combination of these approaches makes the ordering of events possible.

ATP binding cassette (ABC) transporters are a superfamily of proteins that transport a wide variety of substrates across cellular membranes. ABC transporters serve important functions within most of the human cells, including the transport of most drugs and metabolites. Their mutations and malfunctions are linked to many diseases, thus understanding these structures and dynamics would be important for cancer research, pharmacokinetics, and pharmacogenetic disorders [11].

The aim of this thesis is to analyze the dynamics ABC transporters using computational approaches such as ANM-LD and transfer entropy, and to characterize the order of events for further research and drug development. In this work, three different members of the ABC transporter superfamily are analyzed with different motives. TmrAB is a heterodimeric bacterial multidrug resistance protein from *Thermus thermophilus*. Hofmann et al., have obtained nine cryo-EM structures that show the protein in different conformations [12]. This allows simulations to be performed between multiple initial and target structure combinations, allowing a full analysis of the energy landscape. The second example is the cystic fibrosis transmembrane conductance regulator (CFTR) that is responsible for cystic fibrosis (CF). The allosteric network and mechanism of action for CFTR is uncharacterized. Identification of allosteric sites is essential for drug development. Finally, the manganese transporter MntBC from *Bacil-*

lus anthracis is directly linked to anthrax [13]. In 2021, Neville et al., have obtained the crystal structure of PsaBC from *Streptococcus pneumoniae* in the inward-facing conformation [14]. The mechanism of translocation is uncharacterized and key sites remain unknown. Other ABC transporters are also analyzed using the transfer entropy method [10, 15]. Observing the similarities and differences between their dynamics makes a novel classification based on their dynamics possible.

1.1. ATP-Binding Cassette (ABC) Transporter Superfamily

ATP binding cassette (ABC) transporters are one of the largest and oldest superfamilies and they are present in all living organisms [2]. ABC transporters are involved in the transport of sugars, amino acids, proteins, ions and metabolites across the cell membrane and other intracellular membranes of endoplasmic reticulum and mitochondria [16]. These substrates are involved in important metabolic functions inside the cell, therefore their transport is essential to many cellular events. There are 48 ABC proteins encoded in the human genome [17]. Mutations in ABC transporter encoding genes are linked to many diseases such as cystic fibrosis, liver disease, and Alzheimer's disease [18, 19].

Substrate transport across cell membranes can be passive or active. Passive transport utilizes the concentration gradient caused by different concentrations inside and outside the cell. Active transport either couples the movement of molecules that are moving in opposite directions or uses the energy released by ATP hydrolysis. ABC transporters use energy from ATP binding and hydrolysis to move substrates against an electrochemical gradient [11]. ABC transporters can transport substrates into or out of the cell membrane. This movement requires large conformational changes [20].

All ABC transporters have an ATP binding cassette, also named the nucleotide binding domain (NBD). This sequence is largely conserved across the members of the ABC transporter superfamily. Common conserved motifs of the NBDs include; Walker A and B, ABC signature motif, the H loop and the Q loop. Each protein has two NBDs that bind ATPs at the interface. The hydrolysis of ATPs in this domain provides the

energy for active transport. The conserved motifs are either involved in ATP binding and hydrolysis or assisting the binding of NBDs.

ABC transporters have two transmembrane domains (TMDs) that are also called membrane spanning domains (MSDs). The TMDs of ABC exporters are made up of twelve transmembrane (TM) helices and TMDs of ABC importers have between 10 to 20 TM helices [11]. Unlike the highly conserved NBDs, there is great variety in the sequences, lengths, folds, and surfaces of the TMDs. In most ABC importers, these domains are four different polypeptide chains that join together to create the transporter. In bacterial exporters, one TMD and one NBD are fused together and create a half-transporter. Some eukaryotic exporters even have all domains on the same polypeptide chain [21].

Initial classifications for this superfamily were made based on the direction of the transport [22]. Some transporters move substrates from the cytoplasmic region out to the periplasmic region (exporters), some move substrates from the periplasmic into the cytoplasmic region (importers) and some transporters are bi-directional. For substrates to be shuttled across the cell membrane, proteins must undergo conformational changes.

The alternating access model was generally accepted for the transport mechanism of ABC transporters. In this model, proteins start from an inward facing conformation and shift to a nucleotide bound outward facing conformation. During this shift, the TM helices are swapped between the two TMDs [23]. There are other proposed mechanisms such as occluded, outward only and substrate capture mechanisms.

The first ABC transporter to be characterized by Locher and Rees in 2002 was the B12 importer BtuCD [24]. In the 20 years since, the structures of many other ABC transporters have been identified. Advanced imaging techniques such as cryo-EM, SAXS and NMR can not only capture proteins at their resting states but also sometimes at their high-energy intermediate conformations [2]. These crystal structures have made it possible to perform computational studies that reveal the transport mechanism and dynamics.

Two major conformational shapes that are common for transporters are inward-facing (IF) and outward facing (OF) conformations [25]. Both of these conformations are essential for substrate binding and release. ABC importers are at their resting state in the OF conformation. On the contrary, ABC exporters are at their resting state in the IF conformation. The characterization of the transient conformations is highly challenging. Only by certain modifications such as mutations or changing transport rates, experimental results on transitional conformations are able to be obtained [12].

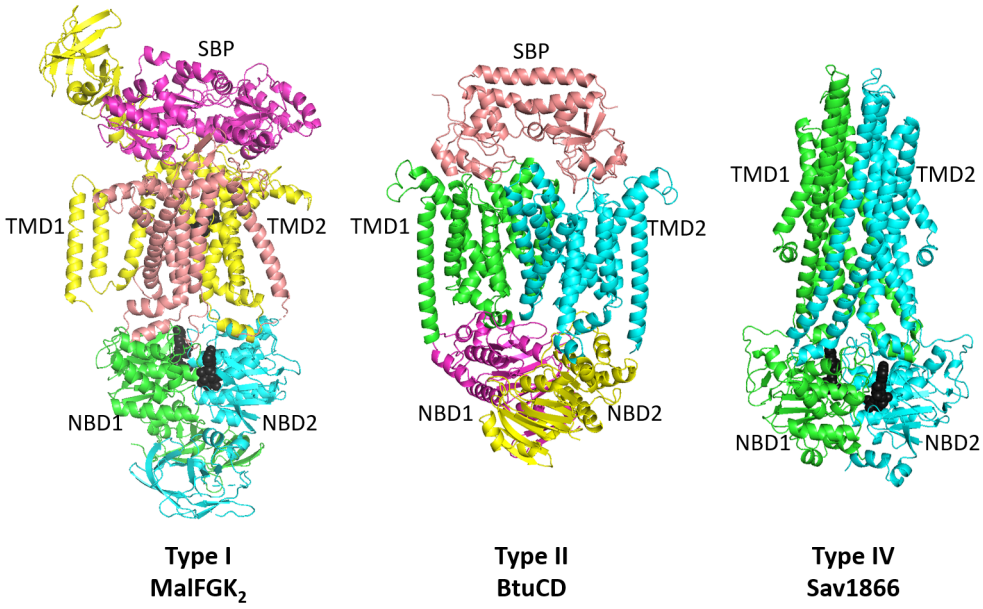

Figure 1.1. Transporters from Types I, II, and IV.

An abundance of experimental information has made a classification based on TMD folds possible. Previously, types of gene subfamilies and chromosomal location had been used to classify ABC transporters into subgroups [16]. Most recently, Thomas et al., have used the differences in the TMD folds to classify ABC transporters into seven groups [22]. This classification takes into account the diversity of the TMD folds and evolutionary networks between bacterial and eukaryotic members of the ABC transporter superfamily. The most prominent types are type I, II and IV. Type I and II ABC transporters are all importers and they have helix organization $(5-6)+(5-6/8)$ and $10+10$ respectively. Type IV transporters can be both exporters or importers

and have helix organization $6 + 6$. Figure 1.1 shows prominent examples from these three types. Maltose importer $MalFGK_2$, B12 importer BtuCD and multidrug exporter Sav1866 are shown for types I, II and IV respectively. Coloring is made based on chains and domains are labelled. As this classification is done based on the fold of the protein, it requires structural knowledge.

Some of the most prominent members of the ABC transporters require accessory domains that facilitate or regulate translocation. These domains are not a core part and their removal does not affect the transport function [26]. ABC importers often require a substrate-binding protein (SBP) to deliver the substrate from the external side to the surface of the cell [21]. ABC importers feature a docking site for the attachment of SBPs. This site relays a signal to the NBDs for the binding and hydrolysis of ATPs [26]. Figure 1.1 shows $MalFGK_2$ and BtuCD with their SBPs. SBPs of type II importers have higher binding affinities to substrates compared to type I importers. This is because type II importers often transport less abundant substrates [25]. ABC exporters do not need an additional unit to capture the substrate from the cytoplasmic side. They instead have a substrate binding site at the interface of the TMDs that is accessible in the IF conformation.

1.2. Multidrug Resistance Protein TmrAB

TmrAB is an ABC exporter from the gram-negative bacteria *Thermus thermophilus*. TmrAB is a heterodimeric protein, with each half containing one transmembrane domain (TMD) and one nucleotide binding domain (NBD). It is a type IV transporter and has the same helix organization as Sav1866 presented in Figure 1.1. TmrAB also serves as a model system for the human transporter associated with antigen processing (TAP) [12]. It's shown that TmrAB can restore antigen transport in TAP-deficient cells [20]. TmrAB is a great case study for dynamic characterization because nine distinct functional states of TmrAB have been characterized by cryo-EM [12].

In Figure 1.2, the organization of TmrAB is presented. The top part lies inside the cell membrane (TMD) and the bottom part is for ATP binding (NBD). While all ABC transporters have two ATP-binding sites formed at the interface of the two NBDs, more than half of the ATP-binding sites of human ABC transporters are asymmetric [27]. In TmrAB, ATP1 can hydrolyze ATP and is named canonical or the catalytic ATP binding site. ATP2 cannot hydrolyze ATP and is named the non-canonical or the degenerate ATP binding site. The presence of one degenerate and one catalytic ATP binding site is common in half of the ABC transporters [25]. There are six transmembrane (TM) helices on each TMD. TM1 is connected to a short, horizontal helix termed the elbow helix. Helices 2 and 3, and helices 4 and 5 are connected by intracellular loops. The intracellular loop between 4 and 5 extends into NBD2 [28]. This geometry is termed 'domain swapped'. Type IV transporters all feature coupling helices that protrude into the opposite domain and create a point of contact. Types without domain swapped coupling helices have a large space at the center (such as Type II importer BtuCD) [21].

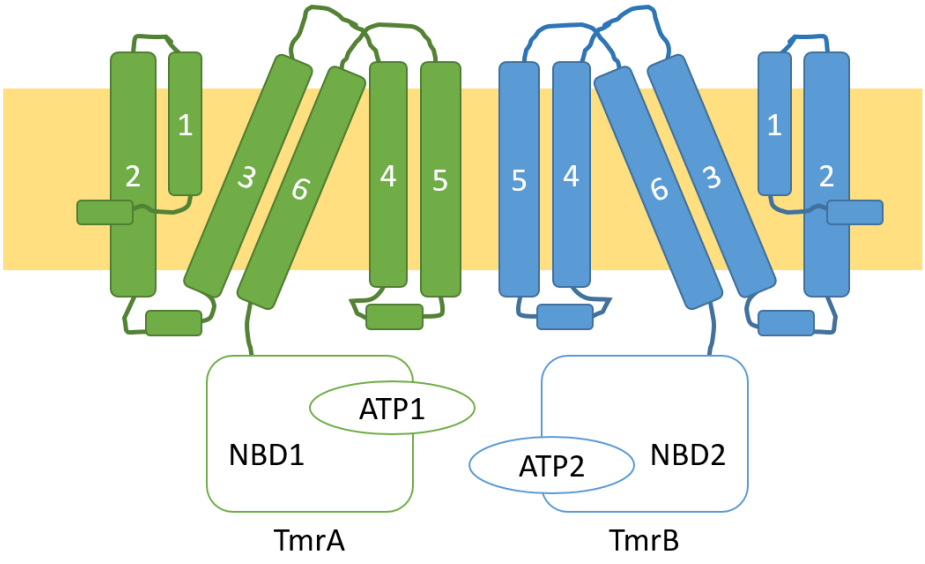

Figure 1.2. TmrAB helix organization and domains.

Inward facing conformations need to be narrow (IF^{narrow}) or wide (IF^{wide}) depending on the size of the substrate, and these two conformations are in equilibrium. Hofmann et al., have found that TM6 helix controls the size and opening of the binding

cavity. In the IF states, the degenerate ATP binding site is occupied by an ATP and the catalytic ATP binding site is occupied by an ADP. When both sites are bound to an ATP, this triggers NBD dimerization and the closure of the cytoplasmic gate. This allows the periplasmic gate to open and facilitates a change to the outward facing conformations. The closing of the TMD happens together with the release of a phosphate from ATP, not ATP hydrolysis. The substrate is released to the extracellular side in the outward facing open (OF^{open}) conformation. Before ATP is hydrolyzed, the system fluctuates between OF open and occluded states. After hydrolysis, the protein takes on an unlocked return asymmetric conformation ($Asym$). In this structure, the NBDs are dimerized and the extracellular gate is closed [12].

Hofmann et al., and Stefan et al., have aimed to delineate the translocation cycle of TmrAB. Single turnover assays have shown that ATP binding is the key to the transition from IF to OF. This transition is responsible for substrate translocation. ATP-occluded states are the pre-hydrolysis states. ATP hydrolysis and phosphate release at the catalytic ATP binding site starts the OF to IF transition. Notably, the degenerate ATP binding site never hydrolyses ATP to ADP. The release of the phosphate triggers the change to the unlocked return (or asymmetric) conformation. This process reactivates the peptide and ATP binding [12, 20].

1.3. Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

The ABC transporter superfamily has one unique member. The cystic fibrosis transmembrane conductance regulator (CFTR) is an ABC transporter but unlike the other members it's an ion channel. The CFTR protein is responsible for regulating the chloride ion concentration in the cell. This transport occurs across the electrochemical gradient of the chloride ions, unlike other ABC transporters that utilize energy to move substrates against their concentration gradients. The chloride concentration also affects the movement of water into and out of the cell. The mutations in the CFTR gene causing a malfunctioning of the protein leads to the disease cystic fibrosis (CF). CF is the most common genetic disorder among people of northern European descent [29]. Chronic lung infections due to increased mucus, constriction of flow in the pancreatic

ducts, restriction of bowel activity, and a characteristic increased salinity in sweat are among the many symptoms of CF. There is currently no known cure for CF and even with improved symptomatic treatments the life expectancy is around 45-50 years [17].

Like other ABC transporters, CFTR has two nucleotide binding domains (NBDs) and two transmembrane domains (TMDs). It also has a 200+ residue regulatory domain (RD). These domains are all part of a single chain. The domain connectivity is TMD1-NBD1-RD-TMD2-NBD2. There are multiple available crystal structures for human and zebrafish CFTR in phosphorylated (ATP-bound) and dephosphorylated (ATP-free) conformations [29–32]. However, none of these crystal structures are able to capture the structure for the RD. While the crystal structure of the RD isn't available, it's known that RD contains 19 sites for protein kinase A (PKA) phosphorylation [29]. In the inward facing, closed conformation RD is wedged between the domains and prevents the NBDs from closing. The phosphorylation of the RD is thought to allow the transportation channel to open. After the phosphorylation of RD and ATP binding to the NBDs, RD is dislocated and moves to the side. In this phosphorylated state the transport channel is open.

The dephosphorylated conformation of CFTR is very similar to the IF conformations of other type IV fold ABC transporters. Similar to the translocation process of other ABC transporters, CFTR requires ATP binding for translocation to occur. However, unlike active transport, channel activity does not require an occluded state where both cytoplasmic and periplasmic gates are closed. In turn, CFTR only features a periplasmic gate but not a cytoplasmic gate. Currently, the conducting state of CFTR is uncharacterized, and it is unknown if this conformation resembles a characteristic OF conformation. Similar to TmrAB, CFTR has two ATP binding sites but only one of these sites has the ability to hydrolyze ATP. ATP site 1 is degenerate and ATP site 2 is catalytic. These sites are formed at the interfaces of the two NBDs. For ATP site 1, the head subdomain lies at NBD1 and the tail subdomain lies at NBD2. The opposite is true for ATP site 2. The core or 'head' subdomain that forms the ATP site features an A-loop, Walker A motif, Q-loop, H-loop, and Walker B motif. The 'tail' subdomain features the D-loop and the conserved signature sequence [33].

There are over 300 disease-causing mutations. The most common CF-causing mutation is the deletion of a phenylalanine amino acid (delF508), which makes CFTR more likely to degrade before it's able to reach the plasma membrane [34]. Other mutations to the CFTR gene can cause defects related to reduced protein expression, function, and stability. There are different drug and therapy options for CF patients. Small molecule CFTR modulators are being used to counteract the effects of CF.

There are five classes of small molecules: read through agents, correctors, potentiators, stabilizers, and amplifiers [35]. Read through agents are used to combat premature stop codon mutations, they stimulate the continuation of translation. Potentiators and correctors are the most common modulators and they are often used in tandem. Potentiators are used to increase channel activity, and correctors increase the amount of CFTR in the cell membrane [36]. Ivacaftor, previously named VX-770 is the most commonly used potentiator drug. Lumacaftor (VX-809), tezacaftor (VX-661), and elexacaftor (VX-445) are all commercially available corrector drugs [37]. These drugs bind to different allosteric cavities on CFTR. Ivacaftor binds to the TMDs at the protein-cell membrane interface. It binds to a cavity created by TM helices 4, 5, and 8. This cavity is also a hinge region in TM8 [34]. S308, F312, and F931 are the most important residues in ivacaftor binding [38]. On the other hand, corrector drugs are known to bind to another cavity in the TMDs. Fiedorczuk and Chen have identified that lumacaftor binds to a binding pocket formed by TM1, 2, 3, and 6 [36]. The other two class of drugs, stabilizers and amplifiers, anchor the CFTR at the plasma membrane and increase the effectiveness of the other drugs respectively [35].

1.4. Manganese Transporter

There are many members of the ABC transporter superfamily that are responsible for the uptake of metal ions. Trace transition metals are vital in many biological processes. However, they are also toxic in high concentrations. It's essential for cells to have a complex metal uptake system in place to mine this scarce resource and to avoid toxic overload. MntBC-A is ABC transporter for manganese transport from the bacterial human pathogen *Bacillus anthracis*. The manganese importer MntBC-A is

also of interest due to its role in the progression of anthrax [13,39]. MntBC-A is a type II importer and previously has been modeled based on the available crystal structures of other type II importers. These importers are, B12 vitamin importer BtuCD and heme importer BhuUV [24,40]. While these proteins are structurally similar to MntBC, the differences in substrate size and charge caused challenges.

Most recently, the manganese transporter PsaBC from gram-positive bacterial pathogen *Streptococcus pneumoniae* in open-inward conformation has been characterized [14]. The sequence of PsaB and PsaC are respectively 41% and 51% identical to MntB and MntC. Thus PsaBC serves as a good template for a structural model for MntBC. There are three available models for MntBC based on PsaBC; the homology model, the Hidden Markov model with HHPred [41] and the AlphaFold model based on neural network algorithms [42].

Understanding the mechanism of action and dynamics of manganese transport is key to counteracting diseases such as anthrax [13]. Type II transporters often feature an extra substrate binding protein (SBP). SBPs aid in the capture of the metal ions through their higher affinities [14]. The SBP-bound conformation for the manganese transporter is structurally uncharacterized. The interactions between TMDs and SBP are of great interest. Computational approaches also serve as a great tool to identify potential allosteric sites. Neville et al. have identified certain residues have been identified as serving key functions; L43 is a part of the cytoplasmic gate, D46 and H50 are coordination sites, and L104 and I199 control pore closure. Also, residues 117-135 are known to form the extracellular gate [14]. Identification of other key sites or allosteric pockets is of utmost importance to control or manipulate transport.

2. MATERIALS AND METHODS

2.1. ABC Transporter Dataset

ATP binding cassette (ABC) transporters are a superfamily of proteins that transport different substrates across the cellular membranes [43]. There are 48 ABC transporter proteins encoded in the human genome [17]. ABC transporters serve many functions in the cell and are linked to many life-threatening diseases. Starting in the early 2000s, the crystal structures of ABC transporters became available. The first full ABC transporters to be imaged were the vitamin B12 transporter BtuCD from *E. coli*, the metal chelate transporter from *H. influenzae* and molybdate transporter ModBC from *A. fulgidus* [21, 24, 44].

Thomas et al., have classified ABC transporters based on their TMD folds into seven subtypes. In this work types I, II, and IV are considered. Type I and II transporters are all importers [22]. For type I, the calculations are performed on maltose importer MalFGK, alginate importer AlgM1M2SS, trehalose importer SugABC, molybdate importer ModBC and Methionine importer. From type II, B12 importer BtuCD, heme importer BhuUV, manganese importer PsaBC is considered. Type IV transporters show more variability in terms of their substrate direction and substrate size. The type IV transporters in this work are: daunorubicin exporter Tm287/238, peptide exporter TmrAB, iron siderophore importer IrtAB, peptide exporter McjD, siderophore importer YbtPQ, and multidrug importer Sav1866. The conformations of the crystal structures, bound ligands and respective references are presented in Table 2.1.

Table 2.1. ABC Transporter Dataset.

Type	Transporter	PDB ID	Ref.
Type I	**MalFGK** (maltose importer)	3PUY	[45]
		3FH6	[46]
	AlgM1M2SS (Alginate importer)	4TQU	[47]
	SugABC (trehalose importer)	7CAG	[48]
		7CAD	[48]
	ModBC (molybdate importer)	2ONK	[44]
		3D31	[49]
	MetNI (methionine importer)	6CVL	[50]
		3TUI	[51]
Type II	**BtuCD** (B12 importer)	1L7V, 4R9U	[24]
		2QI9	[52]
		4FI3	[53]
	BhuUV (heme importer)	5B57	[40]
		5B58	
	PsaBC (manganese)	7KYP	[14]
Type IV	**TM287/288**(Daunorubicin export)	4Q4A	[54]
		6QUZ	[55]
	TmrAB (peptide exporter)	6RAF	[12]
		6RAN	
		6RAJ	
	IrtAB (iron sid. importer)	6TEJ	[56]
	McjD (peptide exporter)	4PL0	[57]
	YbtPQ (siderophore importer)	6P6J	[58]
	Sav1866	2ONJ	[59]
	CFTR	5UAK	[29]

2.2. Computational Methodology

2.2.1. Gaussian Network Model

Proteins at their resting state aren't stationary but rather they are in a constant state of fluctuation [60]. These fluctuations can be modeled by elastic network models. The first example of an elastic network model is the one-dimensional Gaussian Network Model (GNM). Elastic network models describe proteins as a series of beads and springs. Each alpha carbon atom is a node/bead that is connected to another alpha carbon within the cutoff distance through a series of springs. These nodes are assumed to fluctuate around their equilibrium locations following a Gaussian principle. The network of interactions of each node pair are shown on a Kirchhoff matrix, Γ. The equilibrium correlation between each pair (i and j) is written as,

$$\langle \Delta \boldsymbol{R}_i \cdot \Delta \boldsymbol{R}_j \rangle = \left(\frac{3k_b T}{\gamma} \right) \left[\boldsymbol{\Gamma}^{-1} \right]_{ij} \tag{2.1}$$

where k_b is Boltzmann constant, T is absolute temperature (K), and γ is the spring force constant. The Kirchhoff matrix is written as,

$$\boldsymbol{\Gamma}_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } \mathbf{R}_{ij} \leq r_{\mathrm{c}} \\ 0 & \text{if } i \neq j \text{ and } \mathbf{R}_{ij} > r_{\mathrm{c}} \\ -\sum_{i,i \neq j} \boldsymbol{\Gamma}_{ij} & \text{if } i = j \end{cases} \tag{2.2}$$

where each item is described. The inverse of the Kirchhoff matrix can be decomposed into the sum of contributions from orthogonal individual motions, or GNM modes. This is expressed as,

$$\left[\boldsymbol{\Gamma}^{-1} \right]_{ij} = \left[\boldsymbol{U} \left(\boldsymbol{\Lambda}^{-1} \right) \boldsymbol{U}^T \right]_{ij} = \left[\lambda_k^{-1} \boldsymbol{u}_k \boldsymbol{u}_k^T \right]_{ij} \tag{2.3}$$

where U is the eigenvectors matrix, u_k is a single eigenvector, k refers to the kth eigenvector that contains information on the displacements of nodes. The kth eigenvalue and kth GNM mode are proportional [60, 61].

2.2.2. Anisotropic Network Model

The Anisotropic Network Model (ANM) is a three dimensional elastic network model [62, 63]. ANM is an extension of the ideas described for GNM. This model takes into account the orientational differences in the interactions between alpha carbons. The main assumption for ANM is the assumption that equilibrium residue fluctuations are anisotropic. This incorporates the X, Y, and Z directions of the fluctuations into account independently. Each interaction is represented by a spring that is associated with a force constant. Instead of the Kirchhoff matrix, these force constants are shown on a Hessian matrix. Each element of the hessian matrix is actually a 3x3 matrix that has orientational anisotropic information. The Hessian matrix is written as,

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{2.4}$$

where each element of the Hessian is written as a group of second partial derivatives. For each residue pair i and j, the correlation is calculated from their equilibrium positions. This correlation is decomposed into a series of $3N - 6$ orthogonal motions or modes. This relationship is deduced either from the sums of the fluctuations,

$$\langle \Delta \mathrm{R}_\mathbf{i} \cdot \Delta \mathrm{R}_\mathbf{j} \rangle = \langle \Delta X_i \Delta X_j \rangle + \langle \Delta Y_i \Delta Y_j \rangle + \langle \Delta Z_i \Delta Z_j \rangle \tag{2.5}$$

or the inverse Hessian,

$$\langle \Delta \mathrm{R}_\mathbf{i} \cdot \Delta \mathrm{R}_\mathbf{j} \rangle = \frac{k_B T}{\gamma} \left[\widetilde{\mathrm{H}}^{-1}_{3i-2,3j-2} + \widetilde{\mathrm{H}}^{-1}_{3i-1,3j-1} + \widetilde{\mathrm{H}}^{-1}_{3i,3j} \right]. \tag{2.6}$$

2.2.3. ANM-LD

Proteins are complex systems with often tens of thousands of atoms per molecule. This limitation makes it very challenging to simulate the trajectories of proteins by classic simulation approaches like molecular dynamics (MD) and langevin dynamics (LD). ANM-LD is a hybrid computational method. Simplification by modeling makes it possible to use Langevin dynamics, which is a computationally costly, high-resolution method, in an efficient computational scheme [8, 9]. ANM-LD is used to explore the transition pathways between a given initial and target structure. This process aims to cover the energy landscape and to generate intermediate conformations along this transition path. The convergence between the initial and target state conformations are measured by the root mean square differences (RMSD), which are calculated as,

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N}}, \tag{2.7}$$

where N is the total number of atom that are incorporated into the RMSD calculation, d_i is the distance between the coordinates of each corresponding atom on the two structures. The ANM-LD simulation process is as follows:

(i) Initial and target state conformations are selected. Energy minimization for both of these conformations are performed.

(ii) The ANM modes of the minimized initial structure are calculated using its coordinates.

(iii) Difference vector D_i between initial and target structures are calculated.

(iv) The most overlapping ANM mode with the difference vector is identified.

(v) An intermediate structure is generated by perturbing the initial conformation along the most overlapping ANM mode using the deformation factor (DF).

(vi) The generated intermediate is energy minimized and Langevin Dynamics simulation is performed to generate a new conformation.

(vii) RMSD is calculated between the intermediate and target conformation.

(viii) Steps (ii-vii) are repeated until RMSD sufficiently converges or until a predefined

number of cycles have been performed.

(ix) A set of intermediate conformations and the transition pathway trajectory is obtained. The selected ANM modes, RMSD differences, collectivity measures have been recorded for each cycle.

The parameters for this simulation are the number of available modes to the initial conformation ($mode_max$), deformation factor (DF) to perturb the intermediate conformations. Additionally, in the LD part temperature (T), number of energy minimization steps ($stepmin$) and number of LD simulation steps ($stepsim$) can be changed.

2.2.4. GNM-Based Transfer Entropy

GNM-based transfer entropy method provides an estimate of the direction of information flow between two residues i and j for a certain time delay τ [10]. GNM produces an expression to describe the correlation between zero-time fluctuations of residue i and the future-time fluctuations of j [61]. This is written as,

$$\langle \Delta \boldsymbol{R}_i(0) \cdot \Delta \boldsymbol{R}_j(\tau) \rangle = (3k_B T/\gamma) \sum_k \left[\lambda_k^{-1} \boldsymbol{U}_k \boldsymbol{U}_k^T \right]_{ij} e^{-\lambda_k \tau/\tau_0} \tag{2.8}$$

where τ_0 is the characteristic vibrational dynamics time for all proteins in folded state. A combination of equal time and time delay correlations are used to calculate the Shannon entropies. This property describes the amount of transfer entropy or information transfer and is written as,

$$T_{i\to j}(\tau) = S\left(\Delta R_j(t+\tau) \mid \Delta R_i(t)\right) - S\left(\Delta R_j(t+\tau) \mid \Delta R_i(t), \Delta R_j(t)\right). \tag{2.9}$$

From this relation, the net entropy transfer from residue i to j at tau is summarized as,

$$\Delta T_{i\to j}(\tau) = T_{i\to j}(\tau) - T_{j\to i}(\tau), \tag{2.10}$$

where $\Delta Ti \rightarrow j(\tau)$ is a measure of the direction of information flow from residues i to j at a given time delay τ. Transfer entropy is a measure of the reduction in the uncertainty of the future fluctuations of atom j given the fluctuations in atoms i and j at time t. This reveals some residues as entropy sources, meaning that they send information to other residues. Others are revealed to be entropy sinks, meaning that they receive information. This method also allows more global conclusions to be drawn such as how certain areas of a protein interact with each other. Along with the transfer entropy calculations, the degree of collectivity is also calculated. This is a measure of how many other residues or atoms are influenced by a motion from a given residue i [64]. This is an additional measure to determine which residues are the most collective or global sources of information. The degree of collectivity between the movements of residues is written as,

$$\kappa_{i,s} = \frac{1}{N} \exp\left(-\sum_{j=1}^{N} \alpha \left(\Delta T_{ij,k}(\tau)\right)^2 \log\left(\alpha \left(\Delta T_{ij,k}(\tau)\right)^2\right) \right) \tag{2.11}$$

where i is a given residue and s is the selected subset of slow modes. N is the number of total residues and $\Delta T_{ij,k}(\tau)$ is the positive net transfer entropy from residue i to j in a given GNM mode k at future time τ [15].

3. RESULTS AND DISCUSSION

3.1. Determination of the Order-of-Events for TmrAB

TmrAB is an ABC exporter from the gram-negative bacteria *Thermus thermophilus*. Hofmann et al., have reported nine crystal structures with high resolution obtained by cryo-EM. These structures show TmrAB during its complete translocation cycle [12]. The available conformations are unlocked-return asymmetrical (PBD ID: 6RAL, 6RAM), inward-facing narrow (PDB ID: 6RAF), inward-facing wide (PDB ID: 6RAG, 6RAN), outward-facing open (PDB ID: 6RAH, 6RAJ), and outward-facing occluded (PDB ID: 6RAK, 6RAI). Table 3.1 shows the RMSD differences between each conformation. The names have been abbreviated to show only the last letter of the PDB ID.

Table 3.1. RMSD between the crystal structures of TmrAB

	L	**M**	**F**	**G**	**N**	**H**	**J**	**K**	**I**
6RAL	0								
6RAM	0.91	0							
6RAF	6.99	6.64	0						
6RAG	6.27	5.85	2.90	0					
6RAN	6.39	5.96	3.22	1.30	0				
6RAH	3.70	3.91	8.05	7.56	7.63	0			
6RAJ	3.77	3.98	8.12	7.61	7.67	0.70	0		
6RAK	0.94	1.43	7.24	6.62	6.74	3.57	3.63	0	
6RAI	0.84	1.32	7.21	6.59	6.71	3.64	3.72	0.65	0

3.1.1. ANM-LD Simulations for TmrAB

ANM-LD simulations are performed between the four distinct conformations of heterodimeric ABC exporter TmrAB. These four conformations are asymmetric (Asym,

PDB ID: 6RAL), inward facing narrow (IF_{narrow}, PDB ID: 6RAF), inward facing (IF_{wide}, PDB ID: 6RAG), and outward facing (OF, PDB ID: 6RAJ). Parallel runs are performed for different mode restrictions of 100 and 30 modes. Deformation factor is 0.4 for all runs. Figure 3.1 shows the four conformations that the ANM-LD simulations are performed between. The six double ended arrows between them indicate the 12 different paths and energy landscapes that have been taken by these simulations.

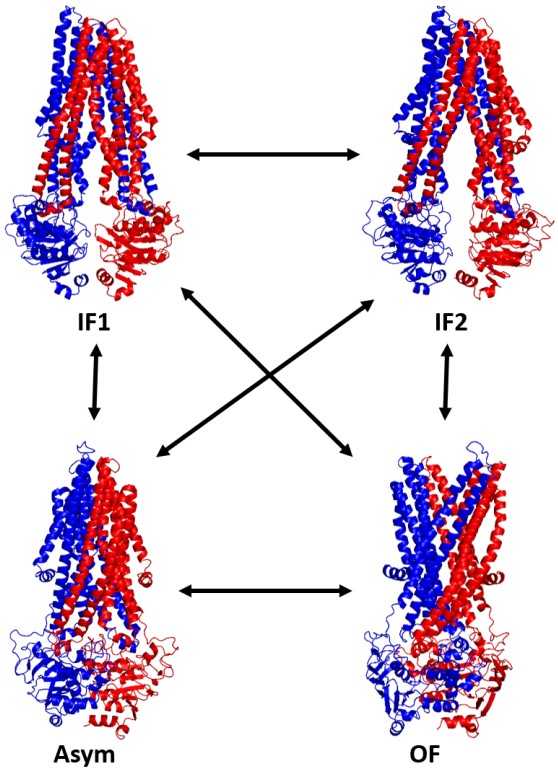

Figure 3.1. Summary of the conformations and paths for ANM-LD simulations.

3.1.1.1. Identification of Key Modes for Transition.

The RMSD changes for each cycle of the ANM-LD simulations is given in Figure 3.2. The plots are organized based on the initial conformation. The first panel shows the RMSD changes when the asymmetrical (Asym) conformation is selected as the initial conformation. Initially, the greatest RMSD difference is with IF1 and the smallest is with OF conformation. The RMSD stops decreasing for the Asym to OF run around the 50th cycle and fluctuates around 1.30. Similarly, Asym to IF1 and Asym to IF2

runs stops decreasing around the 70th and 100th cycles respectively. The second panel shows the RMSD changes for runs starting from IF1, third panel shows runs starting from IF2, and the final panel shows runs starting from OF conformation. In all cases, the RMSD drop occurs the fastest for the smallest initial RMSD difference.

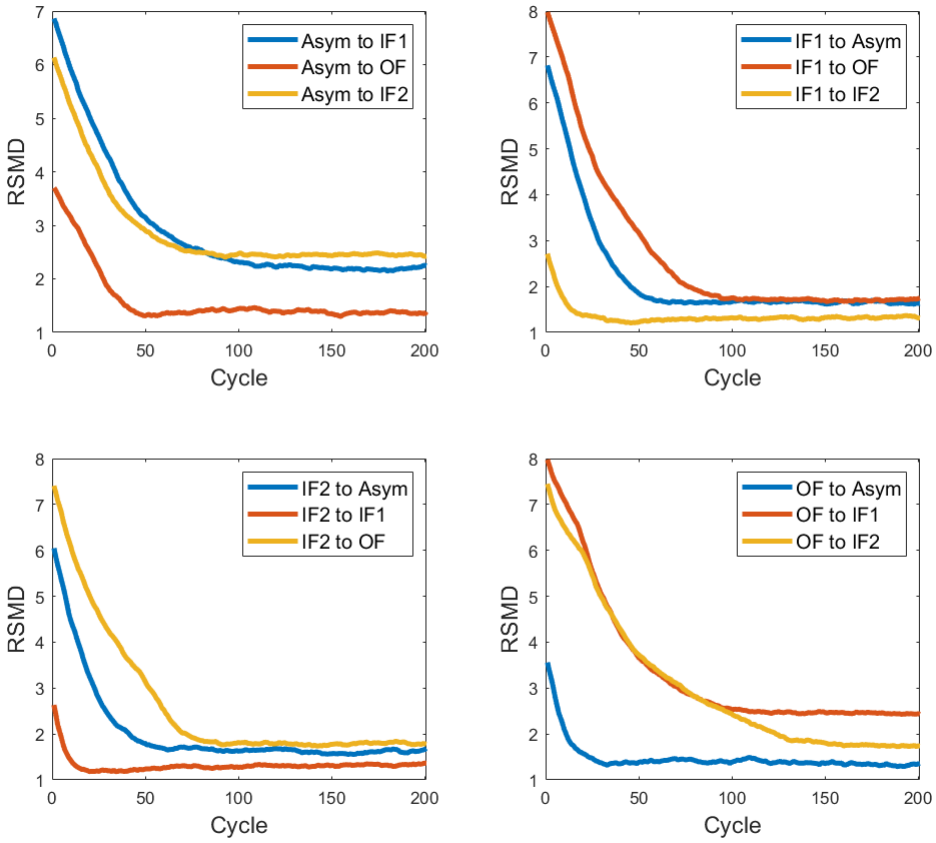

Figure 3.2. RMSD changes for the ANM-LD simulations regarding TmrAB.

A summary of each runs containing the initial and final RMSD and the most selected ANM modes during the ANM-LD simulations is presented below. Additional parallel runs for these initial and target structure combinations have confirmed these mode selections. Identification of key modes for transitions is essential in the determination of the order-of-events and the characterization of dynamics.

Table 3.2. Summary of runs starting from 6RAL.

Initial	**Target**	**Initial RMSD**	**Final RMSD**	**Most Selected Modes**
6RAL	6RAF	6.98	2.14	2, 7, 9, 11, 17, 19, 20
	6RAG	6.26	2.40	2, 5, 7, 11, 19, 23, 29
	6RAJ	3.77	1.30	1, 2, 4, 5, 8, 17, 23, 29

Table 3.3. Summary of runs starting from 6RAF.

Initial	**Target**	**Initial RMSD**	**Final RMSD**	**Most Selected Modes**
6RAF	6RAL	6.98	1.61	1, 5, 7, 9, 11
	6RAG	2.98	1.20	1, 5, 6, 7, 9, 14, 18, 23, 30
	6RAJ	8.12	1.60	1, 5, 9, 10, 11, 30

Table 3.4. Summary of runs starting from 6RAG.

Initial	**Target**	**Initial RMSD**	**Final RMSD**	**Most Selected Modes**
6RAG	6RAL	6.26	1.55	2, 4, 5, 8, 11, 12, 13, 16, 28
	6RAF	2.89	1.18	5, 6, 10, 13, 15, 26, 29
	6RAJ	7.61	1.73	1, 2, 3, 4, 5, 11, 12, 13

Table 3.5. Summary of runs starting from 6RAJ.

Initial	**Target**	**Initial RMSD**	**Final RMSD**	**Most Selected Modes**
6RAJ	6RAL	3.77	1.29	1, 3, 4, 5, 22, 23, 25
	6RAF	8.12	2.42	2, 9, 27, 29, 30
	6RAG	7.61	1.72	2, 9, 13, 16, 28

3.1.1.2. Cross Correlations.

For each of the ANM-LD simulations, cross calculations have been calculated. Cross correlation maps reveal information about the coupling between the movements of each residue during the transition. Along the diagonal is the interaction of each residue with itself, this interaction shows perfect correlation. If there is strong correlation between residues that are far apart in space, this indicates an allosteric connection [9]. Eight key residues are selected from key sites to observe the coupling between different areas. These residues are G396(A) from ATP site 1, G376(B) from ATP site 2, substrate binding sites F261(A) and W297(B), and Y125(A), I226(A), H111(B), and V211(B) from ICLs 1, 2, 3, and 4 respectively [12]. For each cross correlation map presented in this section, the structures are colored based on these key residues.

Figure 3.3 shows the cross correlation map for the transition from asymmetrical (6RAL) to IF-narrow (6RAF) conformation. There are large red and blue segments, this indicates rigid global body movements. As it can be seen from Figure 3.1, the main difference between these conformations is the separated NBDs and opened cytoplasmic gate. In the asymmetrical conformation, TMDs are close together and both cytoplasmic and periplasmic gates are closed. The transition from Asym to IF1 is triggered by the release of ATP from the degenerate site and this is the rate-limiting step of the translocation cycle [12].

There is clear coupling between TMD1 and NBD2, and TMD2 and NBD1. Panel A, shows the coupling between G396 from ATP site 1 and the other areas. NBDs are negatively correlated with each other, meaning that they move at the same time but in opposite directions. This is expected, as the transition to $IF1$ requires an opening of the NBDs. The substrate binding residue F261 from TmrA shows coupling with NBD2 and substrate binding residue W297 from TmrB is coupled with ICL1 and ICL4. ICLs are coupled with the NBDs they are on top of, ICLs 1 and 4 are coupled with NBD1, and ICL2 and ICL3 are coupled with NBD2. Both substrate binding residues are coupled with the lower TMDs, close to the cytoplasmic gates. However, the correlation of the NBDs extends upwards towards the periplasmic side of the TMDs. During the transition from Asym to IF1, only the cytoplasmic gate is opening.

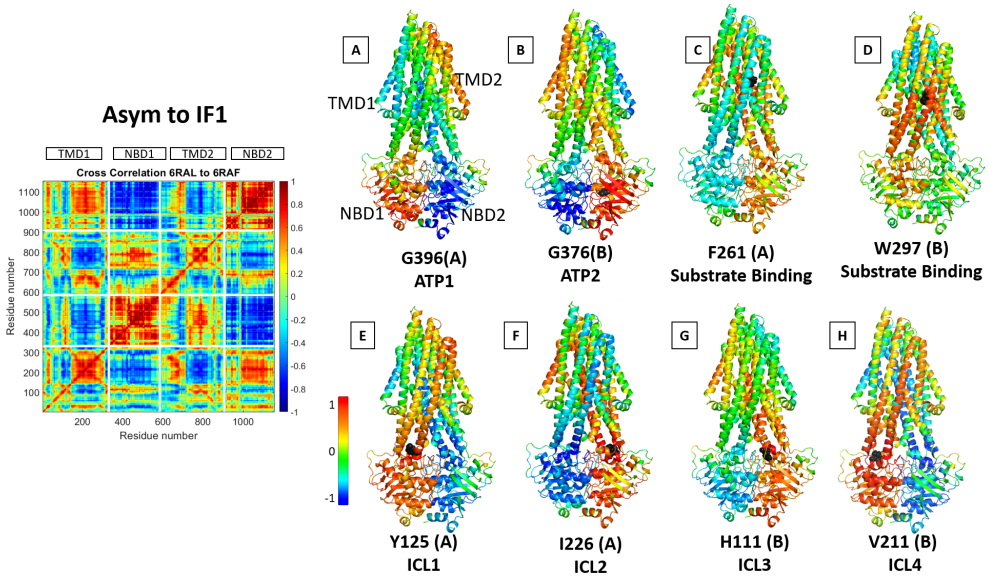

Figure 3.3. Cross correlation results of the transition from 6RAL to 6RAF.

Figure 3.4 shows the cross correlation map for the transition from asymmetrical (6RAL) to IF2 (6RAG) conformation. While this map is very similar to the previous one, there are slight differences. In this case, NBDs are interacting with both TMDs more. There is also more fragmented interaction between the two TMDs. The coloring for this transition produces very similar results as well. Only significant difference is from the substrate binding residues. W297(B) is correlated with NBD2, whereas for Asym to IF1 it was correlated with NBD1. F261(A) was correlated with NBD2, now it isn't correlated with either NBD.

The third run that has the Asym conformation as the initial conformation is presented in Figure 3.5. The target conformation is the outward facing conformation, the initial RMSD between these two conformations is 3.77 Å. For this transition, there is very little movement necessary by the NBDs. The main change that needs to occur is the closure of the periplasmic gates in the TMDs. This is reflected in the figure, as both substrate binding residues are correlated with this area. NBD1 is also slightly correlated with the extracellular side of the TMDs, this is the catalytic NBD. The main correlation occurs between the extracellular side of the TMDs and the ICLs. ICL3 especially is highly correlated with the periplasmic gating area. Both substrate binding residues

are coupled with the extracellular side as well. If this change were to happen without visiting the inward facing states, it would result in a futile cycle as the translocation pathway never opens enough for the transport a peptide [12].

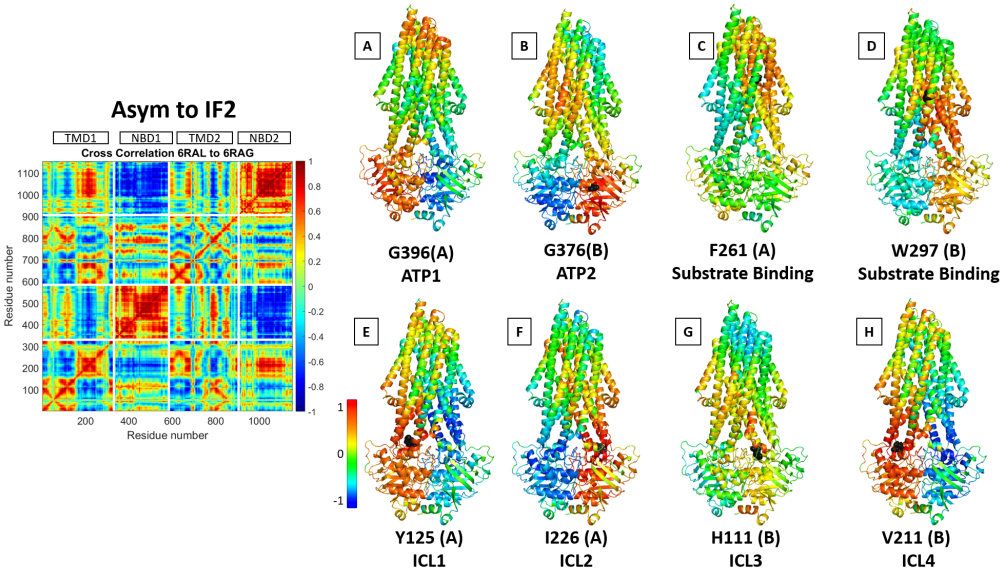

Figure 3.4. Cross correlation results of the transition from 6RAL to 6RAG.

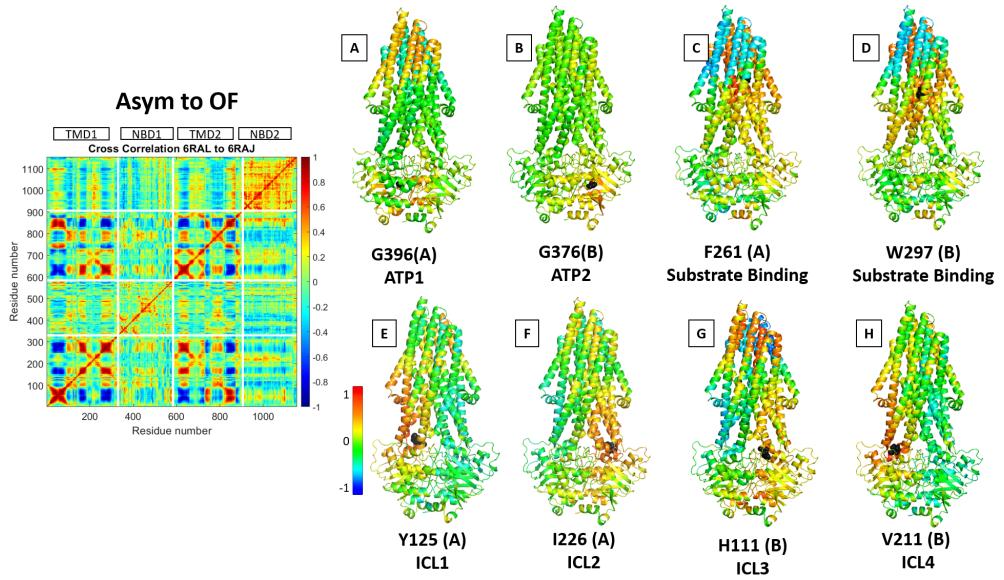

Figure 3.5. Cross correlation results of the transition from 6RAL to 6RAJ.

The second group of cross correlation maps are showing the transitions starting from the IF1 conformations. Figure 3.6 shows the cross correlation map for which transition

and structure colored based on the couplings of the key residues. This is the reverse of the transition which transition presented in Figure 3.3. While these maps aren't identical, they are very similar. TMD1 is clearly coupled with NBD2 and TMD2 is coupled with NBD1.

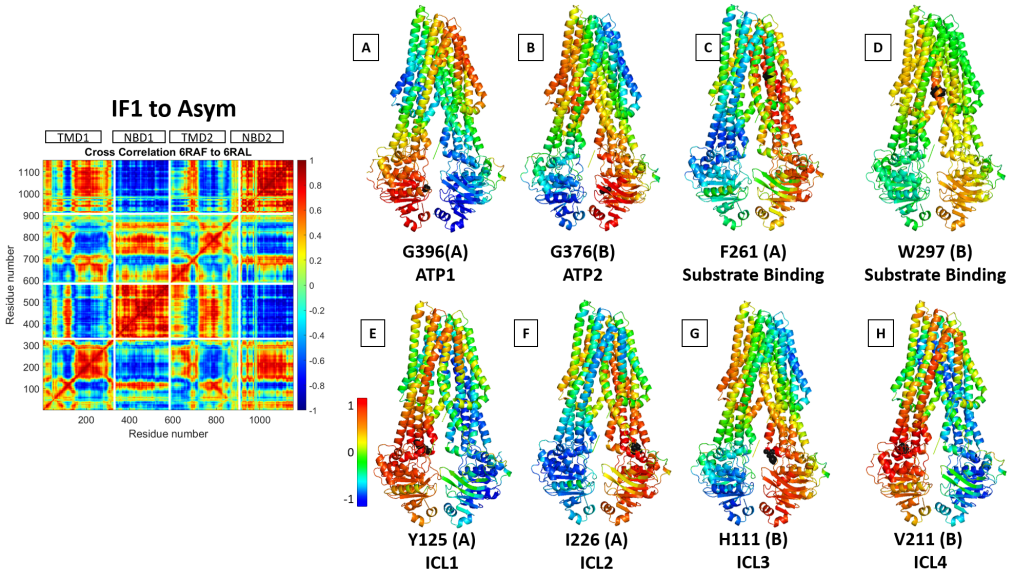

Figure 3.6. Cross correlation results of the transition from 6RAF to 6RAL.

Figure 3.7 shows the cross correlations and the coupling between residues for the transition between inward facing narrow and wide states. The RMSD between these conformations is 2.89 Å. The correlations are very minor for this transition, as it had been for the Asym to OF transition. The major movements need to happen in the NBDs and TM6 of TMD1. TmrAB undergoes this transition in order to accommodate for larger substrates. There is slight coupling between TMD1 and NBD1, and TMD2 and NBD2. This is the opposite that was observed for Asym to IF2 transition. The substrate binding sites aren't correlated with either of the NBDs.

Figure 3.8 contains the correlations for the transition from IF1 to OF. This is the translocation transition. In order for this transition to occur, substrate binding needs to happen at the TMDs and ATP binding needs to occur at the NBD interface. Both ATP binding sites are highly correlated with both TMDs, specifically the extracellular sides and the translocation cavity. ICLs show strong coupling with the NBDs they are

on top of. ATP binding is essential for this transition to occur, as it triggers NBD dimerization that closes the intracellular gate [12].

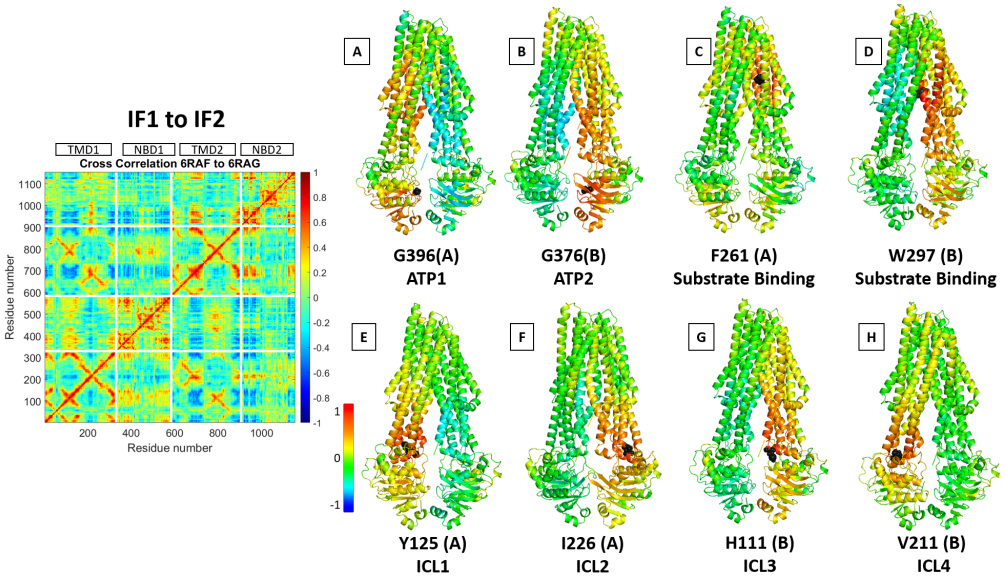

Figure 3.7. Cross correlation results of the transition from 6RAF to 6RAG.

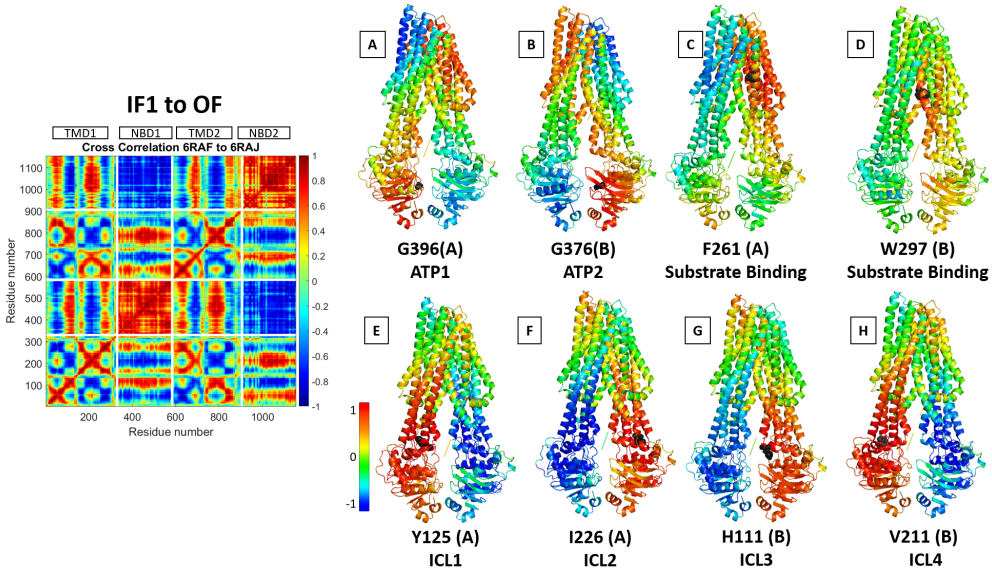

Figure 3.8. Cross correlation results of the transition from 6RAF to 6RAJ.

The third group of cross correlation maps start from the inward facing wide (IF2) conformation. In this conformation, the NBDs are the farthest apart, in order to accommodate larger substrates. Figure 3.9 shows the transition from IF2 to Asym.

One key difference between this figure and Figure 3.6 is TM6 of both TMDs. For IF1, TM6 helix of TmrA is correlated with NBD2 and TM6 of TmrB is correlated with NBD1. For IF2 the opposite is true. TM6 helix is also known to control the plasticity of the binding cavity and it causes the main difference between these two conformation [12]. The other interaction patterns are very similar but IF1 to Asym has stronger couplings between residues.

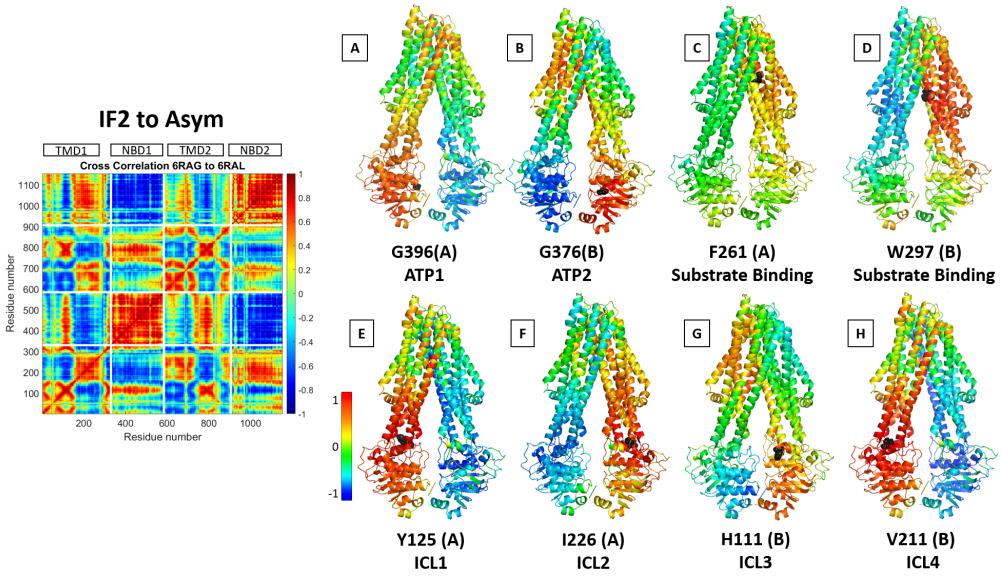

Figure 3.9. Cross correlation results of the transition from 6RAG to 6RAL.

The two inward facing states narrow and wide are in equilibrium. The protein fluctuates between these conformations until a substrate is bound to the substrate binding residues and the catalytic ATP binding site is phosphorylated. The enlargement of the binding cavity is controlled by TM6. Figure 3.10 is the opposite transition to Figure 3.7. The correlations are stronger for IF1 to IF2 transition, especially the correlations between TMD1-NBD2 and TMD2-NBD1. The colored structures reveal the coupling between residues to stay relatively local.

Final figure regarding IF2 is the transition from IF2 to OF. Similar to Figure 3.8 this change is triggered by ATP binding to both sites. The results are almost identical, with the exception of a stronger coupling of substrate binding residue W297(B) in Figure 3.11 (D) to both NBDs.

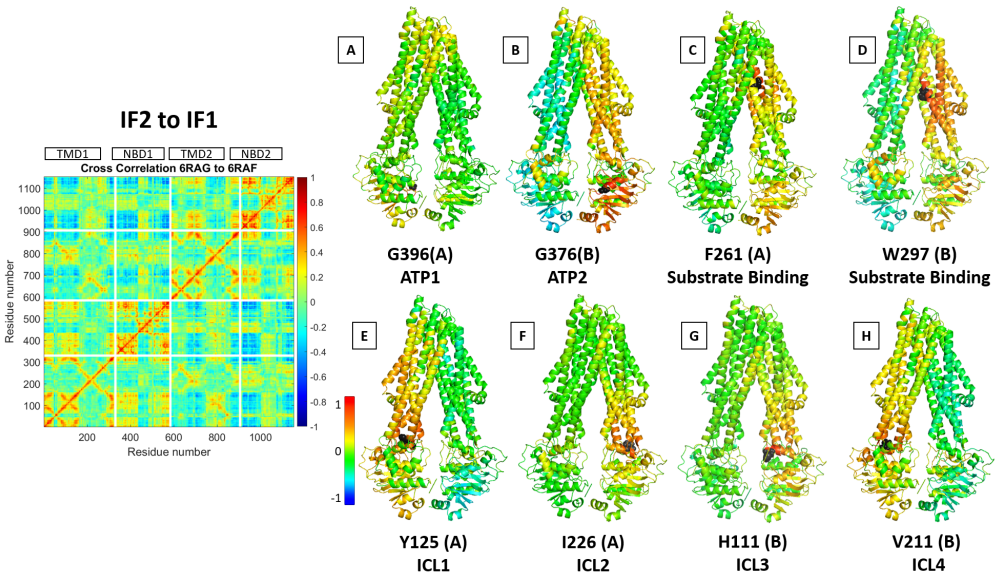

Figure 3.10. Cross correlation results of the transition from 6RAG to 6RAF.

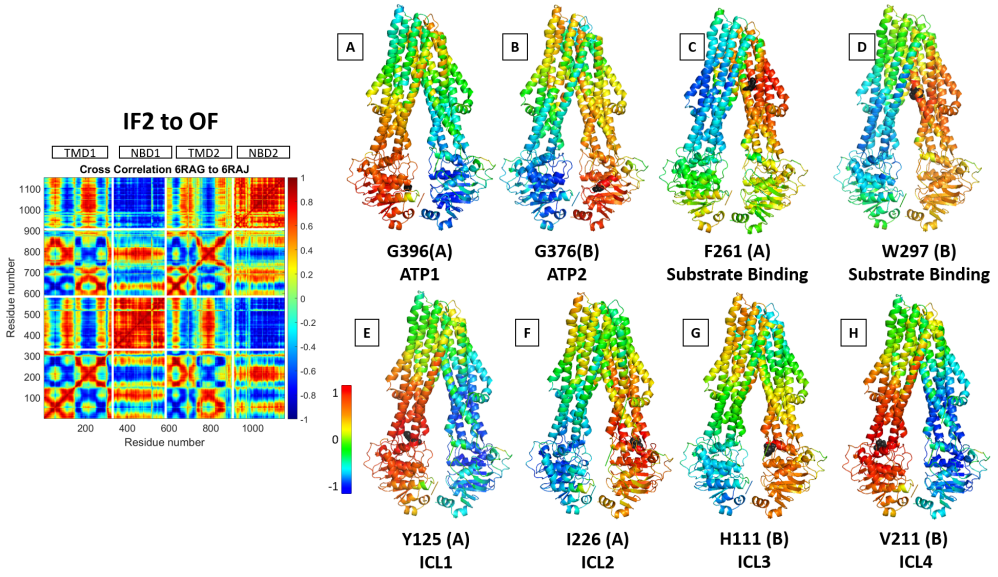

Figure 3.11. Cross correlation results of the transition from 6RAG to 6RAJ.

The final group of cross correlation maps start from the outward facing (OF) conformation. This conformation is significantly different form the IF conformations, the NBDs are bound together and the TMDs are separated. This is also the only conformation where two ATPs are bound to both ATP binding sites.

Figure 3.12 shows the cross correlation map and the coupling between key residues for the transition from the OF to Asym conformations. This transition requires ATP hydrolysis and the release of a phosphate from the catalytic ATP site [12]. The major change in the structure happens in the TMDs, the periplasmic gate need to close and TMDs should lock into each other. The RMSD for this transition is 3.77 Åand as seen in Figure 3.5 the coupling between residues is not very strong. The main correlations are observed between the two TMDs. Compared to the opposite direction, the coupling is even less in this direction. In Asym to OF there were slight interactions between the TMDs and the NBDs, for OF to Asym there is virtually no correlation with the NBDs. OF state is highly transient and in the ANM-LD transition the RMSD is able to drop in just 15 cycles. The colored structures in Figure 3.12 show that there is little coupling between domains. The only significant coupling occurs for panel (c/d), substrate binding residues F261(A) and W297 (B) are coupled with both TMDs.

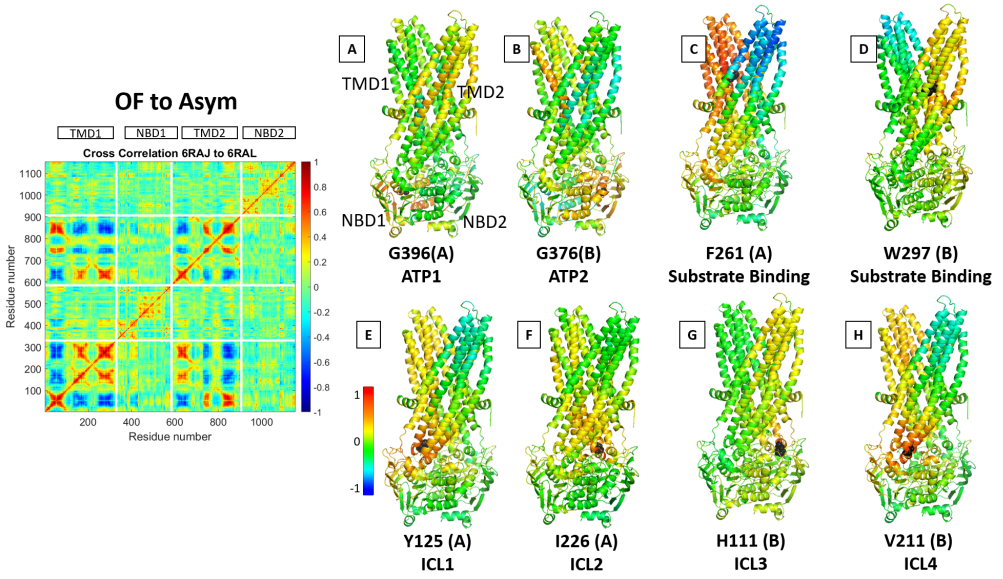

Figure 3.12. Cross correlation results of the transition from 6RAJ to 6RAL.

Figure 3.13 and Figure 3.14 show the transitions from Asym to IF1 and IF2, respectively. The respective initial RMSDs for these transitions are 8.12 Åand 7.61 Å. The protein should also visit the occluded and asymmetrical conformations during these transitions. There are more correlated segments in Figure 3.13, especially around residues forming ICL3 on TmrB (plot index: 690-700). This area is highly correlated

with NBD1. TM6 on TmrB is connected to ICL3 it blocks the entry to the translocation cavity.

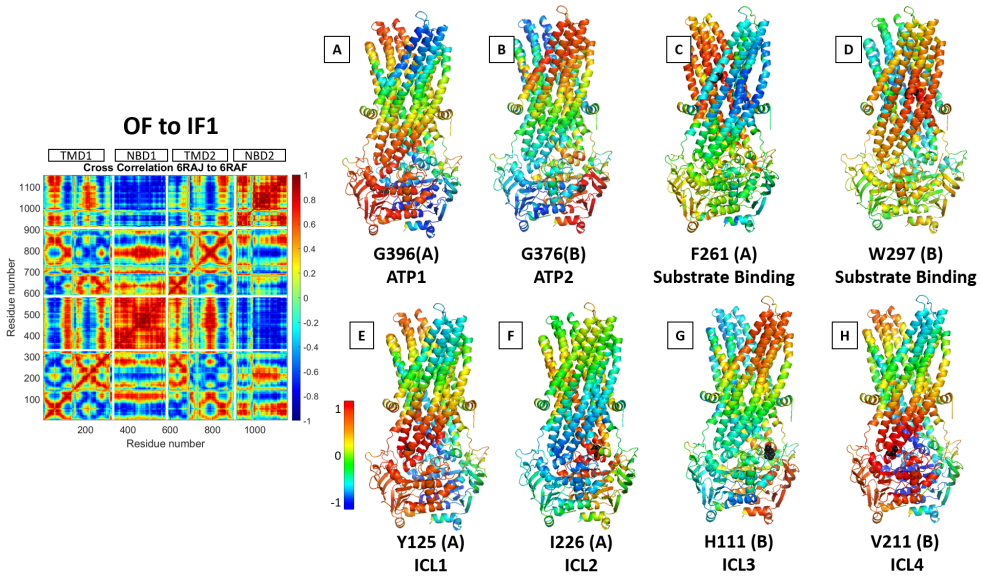

Figure 3.13. Cross correlation results of the transition from 6RAJ to 6RAF.

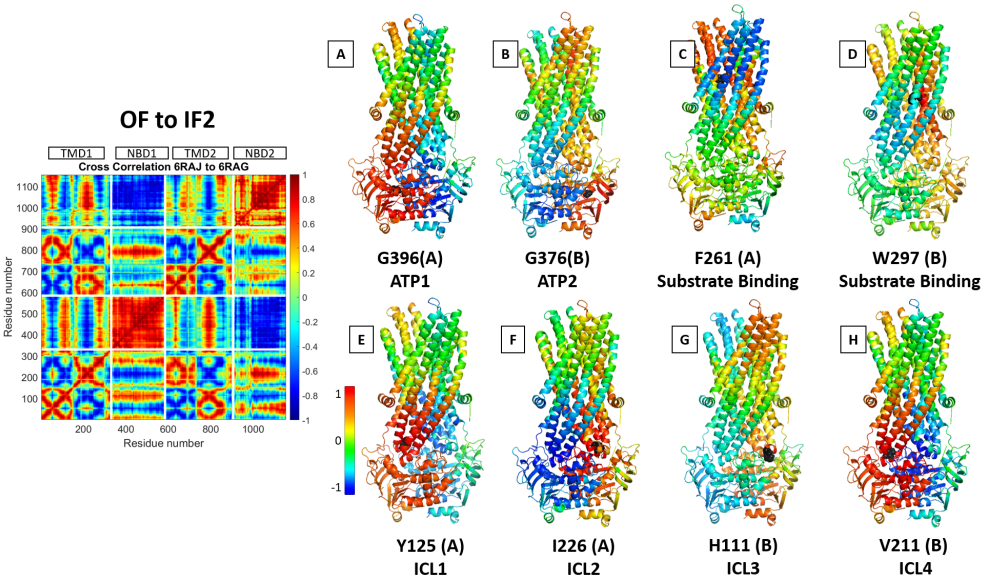

Figure 3.14. Cross correlation results of the transition from 6RAJ to 6RAG.

3.1.2. Transfer Entropy Results

Transfer entropy analysis is performed for all nine of the crystal structures. Each produces a different dynamic information flow pattern. In this section, the detailed results for four distinct conformations (Asym, IFnarrow, IFwide, OF) are presented. The RMSD and structural differences between Asym and OFoccluded conformations are small, so their TE profiles are similar. For each conformation the transfer entropy map is presented along with the degree of collectivity, TExCollectivity score and cumulative TE for each residue. On the TE maps, the domains are separated by white lines. Each square describes how domains are interacting with each other. Residues that are transferring the most amount of entropy are named entropy sources or information sources. Residues that are receiving information are called entropy sinks.

Starting from the asymmetrical (Asym) conformation (PDB ID: 6RAL), Figure 3.15 shows that TMDs are the driving domains. The time delay value (tau or τ) is selected to produce the greatest amount of information transfer. Largest entropy source peaks and residues with highest collectivity are clustered around the extracellular gate and the ATP binding sites. All of the domains are interacting with each other. The ADP and ATP binding sites as shown in Figure 3.15 with vertical red and green lines, correspond to the minor peaks on cumulative TE plot.

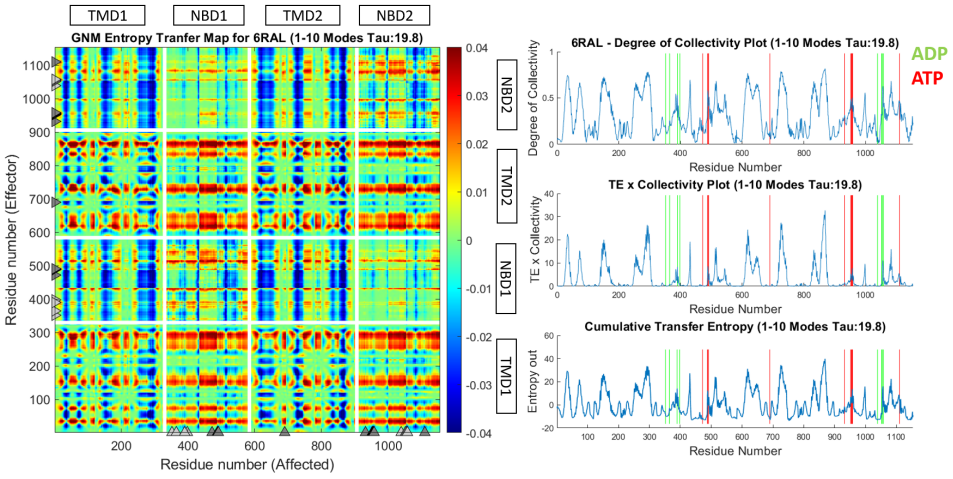

Figure 3.15. Transfer entropy and collectivity results for 6RAL, 1-10 modes.

When the slowest mode is removed, alternative or hidden information transfer patterns are revealed. For 2-10 modes, Figure 3.16 shows that the domains are interacting with their own types (TMD1-TMD2 and NBD1-NBD2). The only point of interaction between TMD and NBDs are the intracellular loops. All 4 loops behave as entropy sources, and both NBDs are entropy sinks to these loops. Finally, the next slow mode is removed and calculations are performed for 3-10 modes. Bottom panels of Figure 3.16 show the information flow from TMDs. There is no information flow from the NBDs. The entropy sources on the TMDs are located around the translocation cavity especially around the periplasmic and cytoplasmic openings.

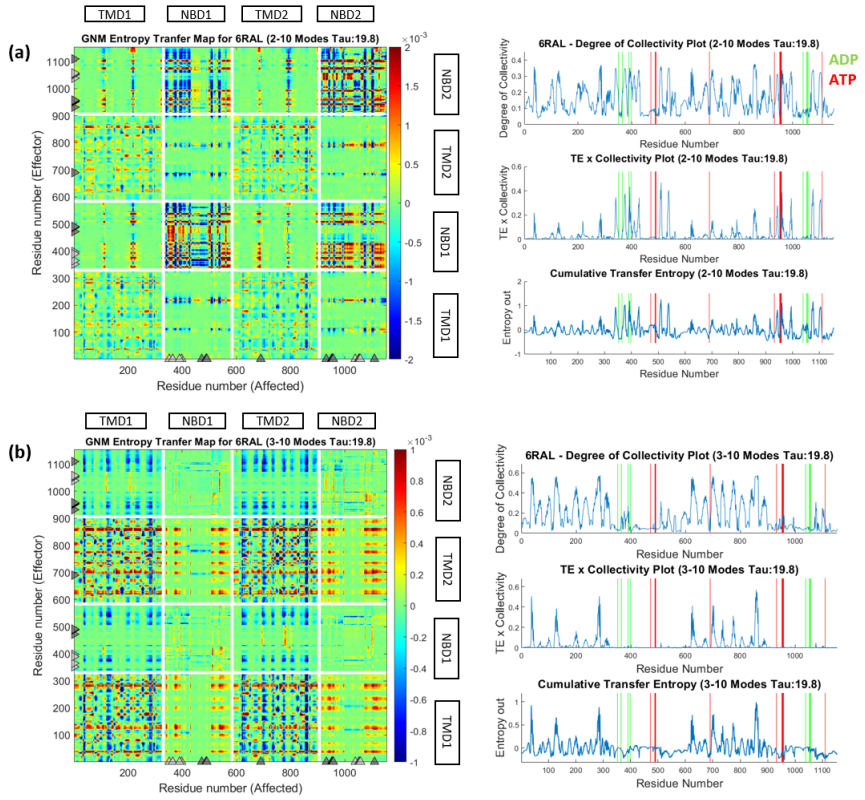

Figure 3.16. Transfer entropy and collectivity results for 6RAL, a) 2-10 b) 3-10 modes.

Figure 3.17 shows the asymmetrical structure of TmrAB colored for the cumulative TE values. The residues that are sending the most information are red and the ones receiving the most information are shown in blue. The residues that are located at the

peaks of the cumulative transfer entropy plot are shown with spheres. These are the most important TE sources.

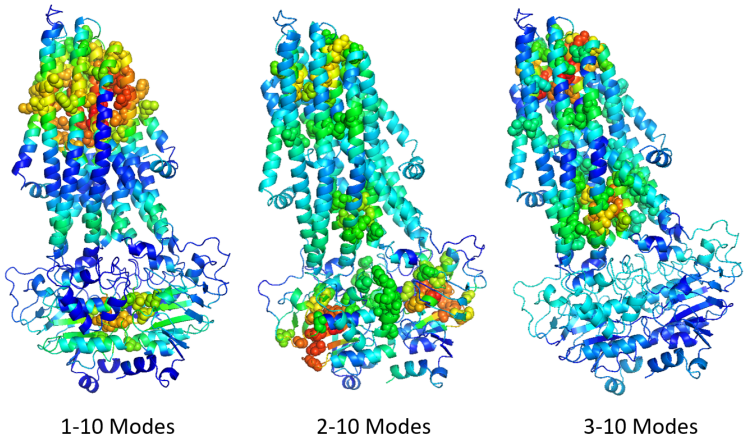

Figure 3.17. Cumulative transfer entropy values shown on 6RAL.

The TE calculations are repeated for the IFnarrow (IF1) conformation (PDB ID: 6RAF). The NBDs are separated and the ATPs are only bound to one side of the NBDs. In this case, NBDs appear to be the driver of events. Entropy sources on the TMDs are clustered around the periplasmic gate. TM6 (residues 286-332) is an entropy source, this area is known to shift during the transition from IFnarrow to IFwide [12]. Figure 3.18 shows the TE plot and collectivity analysis results.

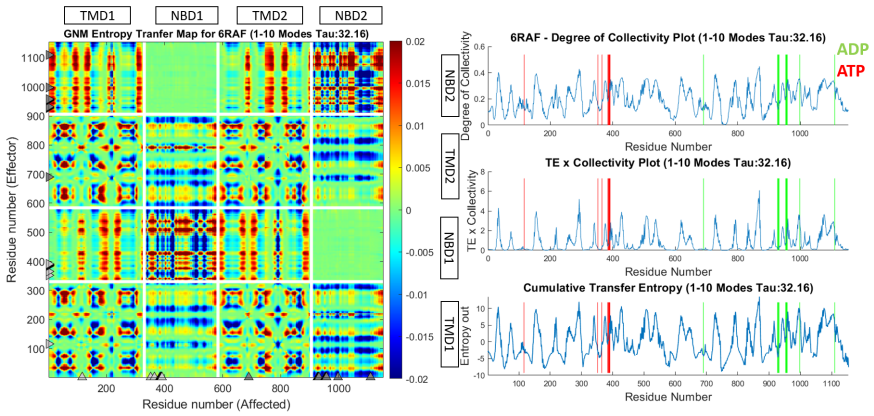

Figure 3.18. Transfer entropy and collectivity results for 6RAF, 1-10 modes.

The movement of TM6 takes place in order to accommodate larger substrates. There are entropy sinks on the TMDs (red vertical lines) that are receiving information from both of the NBDs. There is no interaction between the NBDs. The peaks on the NBDs partially match the ATP binding sites. NBD2 appears to have a more prominent effect on the TMDs, the ATP binding sites on it also match the peaks better. The behavior of the NBDs is asymmetrical, as is common with type IV ABC transporters [22].

The slowest mode and next slow mode are removed and the results for 2-10 and 3-10 modes are presented in Figure 3.19. For 2-10 modes, the NBDs are clearly the driving domains. The only entropy sources on the TMDs are the intracellular loops (ICLs). Two of these are coupling helices. ICLs also act as entropy sinks to the NBDs, meaning that they receive information from the whole domains.

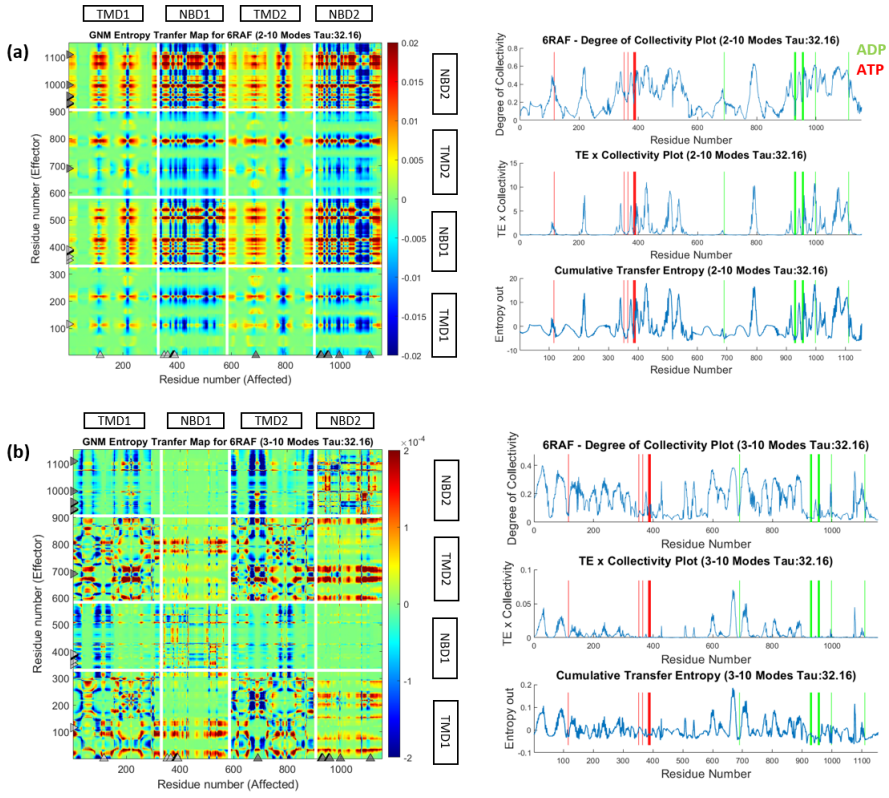

Figure 3.19. Transfer entropy and collectivity results for 6RAF, a) 2-10 b) 3-10 modes.

Even though NBDs are driving, the ATP binding sites partially match the peaks. Similar to IF1, NBD2 has a better match to the entropy source peaks. For 3-10 modes the driving domains are the TMDs. This profile is very similar to Asym conformation 3-10 modes presented in Figure 3.16. There is low collectivity for the peak residues (0.4>). The entropy sources are not around the translocation cavity but towards to cell membrane. The elbow motif (11-24) is an important entropy source. This cavity formed behind the elbow motif could be an allosteric site and could be good candidate for drug studies.

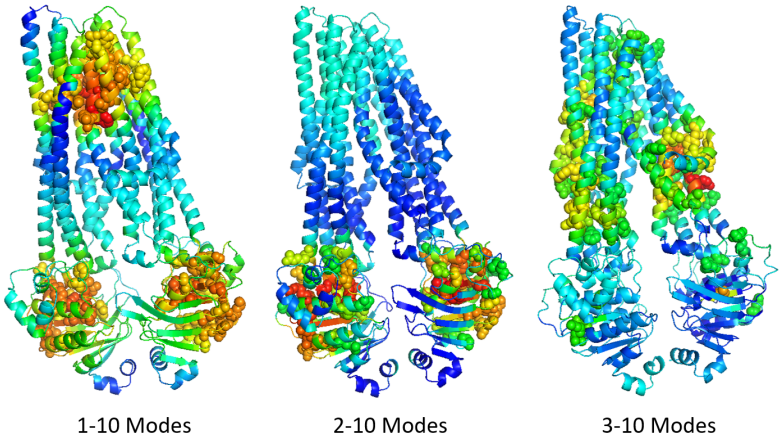

Figure 3.20. Cumulative transfer entropy values shown on 6RAF.

Figure 3.21 shows the TE results for IFwide (IF2) conformation (PDB ID: 6RAG). Also in this case the NBDs are physically separated and ATPs are only bound to one side of the NBDs. TE map shows that the driver of events are the NBDs. The entropy source peaks perfectly match the ATP binding sites. Noticeably there is interaction between the NBDs, which was not present in IFnarrow conformation. The importance of the TMDs has decreased, there is less information flow from TMDs to NBDs. However, the opposite information flow (NBDs to TMDs) has increased. As this is the pre-translocation conformation, this could suggest that ATPs are start driving once the substrate is bound to the TMDs.

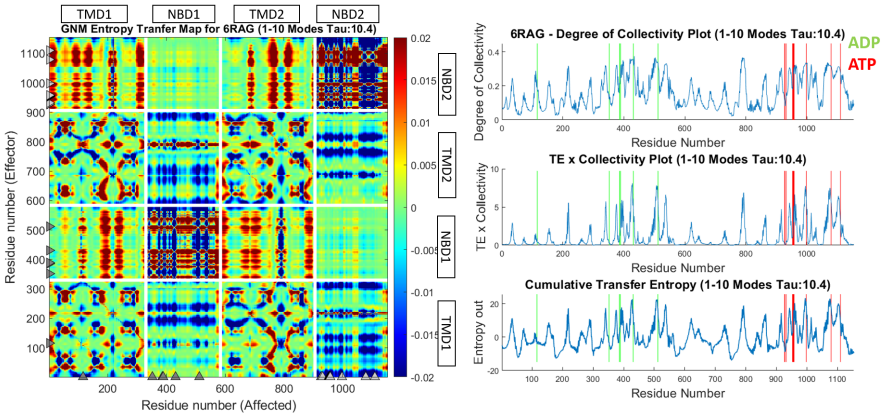

Figure 3.21. Transfer entropy and collectivity results for 6RAG, 1-10 modes.

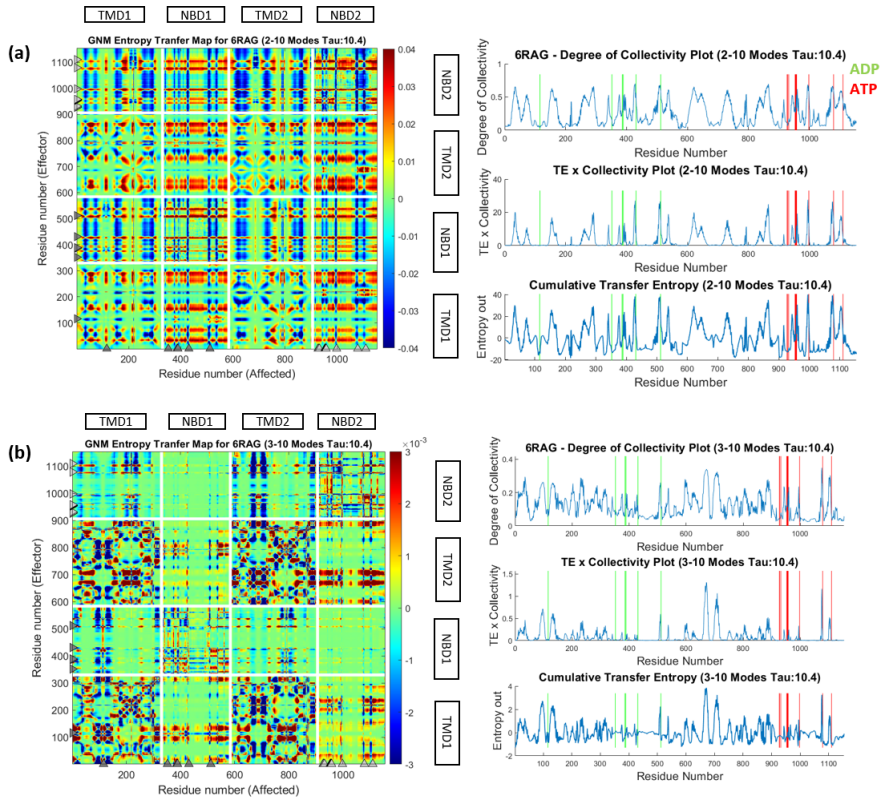

Figure 3.22. Transfer entropy and collectivity results for 6RAG, a) 2-10 b) 3-10 modes.

For 2-10 modes, the TE map Figure 3.22 is very similar to the 1-10 modes of asymmetrical conformation shown in Figure 3.15. This indicates that entropy transfer patterns never fully disappear but shift in importance as conformation changes. ATP binding

residues are again entropy sources. The extracellular gate is also an important site. For 3-10 modes, peaks are located towards the cell membrane. While there were differences for 1-10 modes between IFnarrow and IFwide conformations, the lower mode combinations are quite similar.

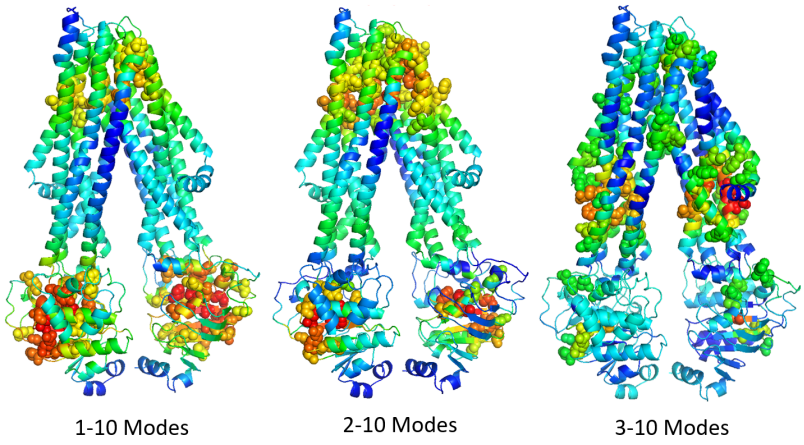

Figure 3.23. Cumulative transfer entropy values shown on 6RAG.

Finally, the results are presented for the outward facing conformation (PDB ID: 6RAJ). This is the conformation after transport has occurred and it is the most unstable conformation [12]. Figure 3.24 shows the TE map, and the collectivity results.

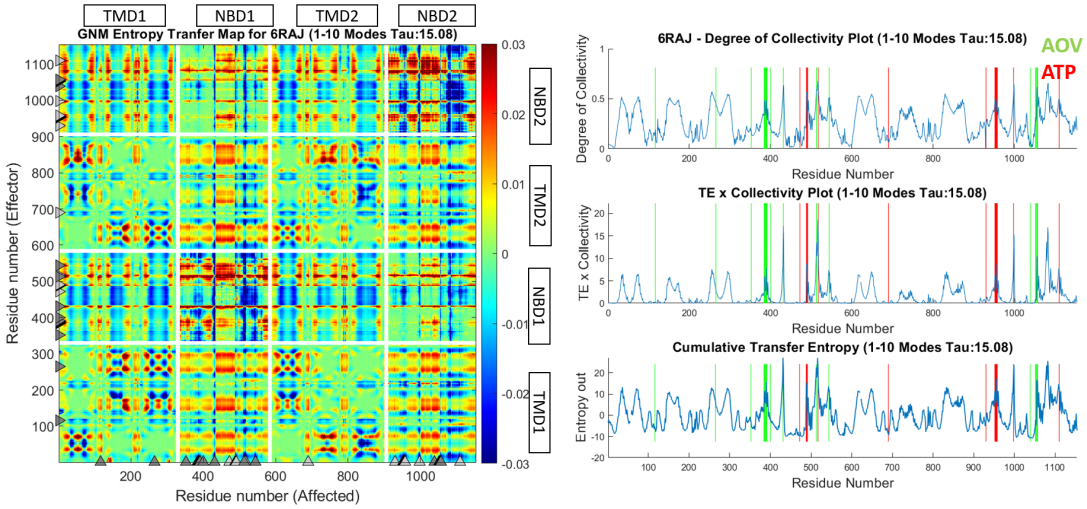

Figure 3.24. Transfer entropy and collectivity results for 6RAJ, 1-10 modes.

Both TMDs and NBDs contain entropy source residues and all of them are interacting with all other domains. All of the ATP binding residues match the entropy source peaks on the right panel. ATP binding sites were also entropy sources for IFwide conformation, this means that they are driving the conformational changes throughout the whole translocation. Figure 3.26 shows that the entropy source peaks for 1-10 modes have shifted downwards in the translocation channel. This area includes the known substrate binding residues F261(A), H246(A) and W297(B) [28]. Removal of the slowest mode yields a similar profile to asymmetrical conformation shown in Figure 3.16. It also resembles the previously dominant profile from IFwide shown in Figure 3.21. As previously observed, the information transfer patterns change in order depending on the conformation. For both 2-10 and 3-10 modes, in the OF state TMDs are only interacting with each other and NBDs are only interacting with each other. Intracellular loops and coupling helices are the only points of interaction across domains of different types.

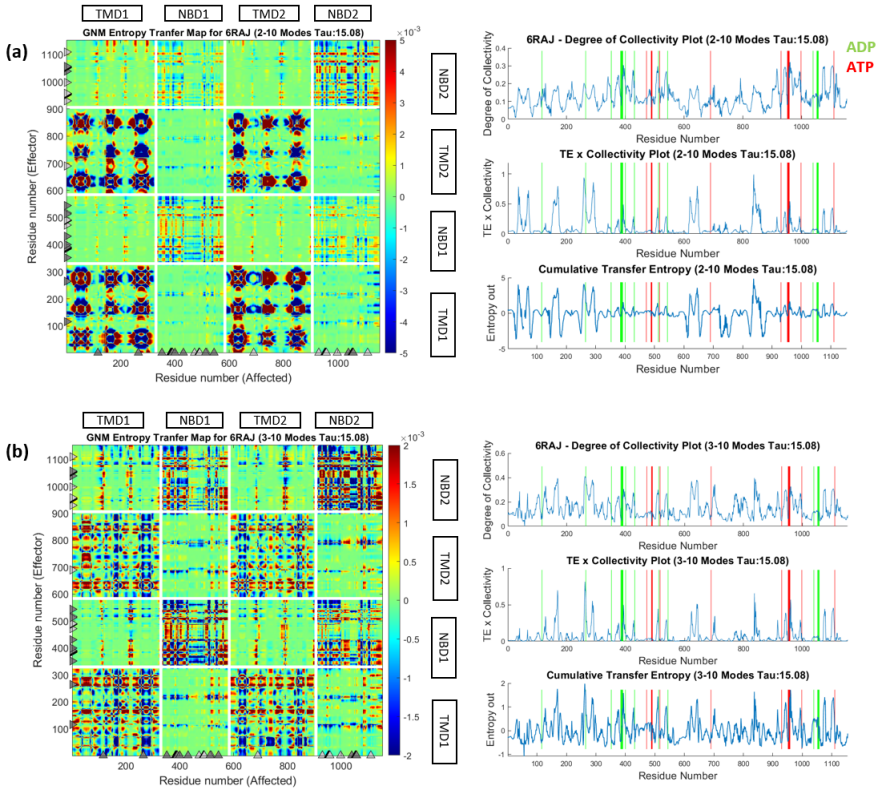

Figure 3.25. Transfer entropy and collectivity results for 6RAJ, a) 2-10 b) 3-10 modes.

To summarize the transfer entropy results for TmrAB, there are clear information transfer patterns for transport. The importance of distinct information patterns changes depending on the conformation. Starting from asymmetrical conformation, where all domains are in close contact, there is information flow both ways and TMDs are driver of events. NBDs are interacting with each other and their behavior is symmetrical. Then the structure moves onto the inward facing narrow conformation, and the NBDs slightly separate. The drivers become the NBDs but there is still information flow from the TMDs as well. There is no interaction between the NBDs and NBD2 appears to be more prominent. There is partial overlap between the peaks and ATP sites. TM6 shifts slightly and the translocation cavity becomes wider. In the IFwide conformation, NBD is clearly the driving domain and the ATP binding sites match the peaks. There is more interaction and possible cooperation between the NBDs. As translocation occurs and TmrAB moves to the outward facing state, all four domains become the drivers. Now both ATP binding sites are bound to an ATP and they match the entropy source peaks exactly.

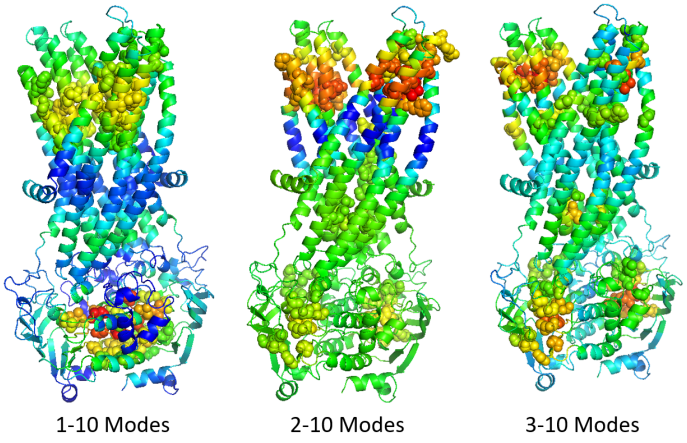

Figure 3.26. Cumulative transfer entropy values shown on 6RAJ.

3.2. Characterization of the Cystic Fibrosis Transmembrane Regulator

The cystic fibrosis transmembrane conductance regulator (CFTR) is a unique member of the ABC transporter superfamily. Unlike the other members which are active transporters, CFTR is an ion channel and it transports substrates using their concen-

tration gradient. In 2017, Liu et al., characterized the first structure for the human CFTR [30]. Currently, there are multiple available crystal structures for phosphorylated and dephosphorylated zebrafish and human CFTRs. There are also drug-bound structures for commercially available drugs ivacaftor, lumacaftor, and tezacaftor. In the dephosphorylated, ATP-free conformation (PDB ID: 5UAK), the NBDs are separated and the transporter is inward-facing. In the phosphorylated, ATP-bound conformation (PDB ID: 6MSM), the NBDs are bound together through the ATPs. This is not the conducting conformation as the gating residues aren't sufficiently separated. The conducting conformation is still uncharacterized. In neither of these PDB structures, the regulatory domain (RD) isn't characterized.

Residues 1-69 form the lasso motif at the beginning of TMD1. Residues 70-390 form TMD1 and 845-1173 form TMD2. Residues 390-637 form NBD1, and residues 1202-1451 form NBD2. The missing portion between 637-845 is the RD. Figure 3.27 shows both human conformations colored based on domains.

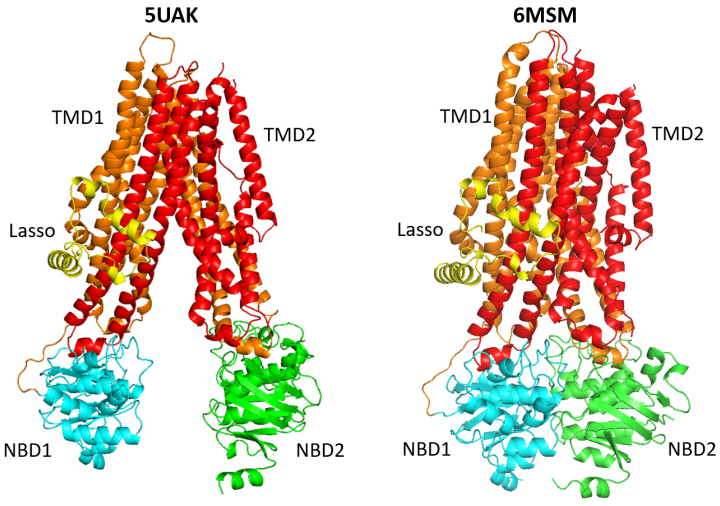

Figure 3.27. Dephosphorylated and phosphorylated conformations of human CFTR protein.

3.2.1. ANM-LD Simulations for CFTR

ANM-LD simulations are performed between the dephosphorylated, ATP-free human CFTR and the phosphorylated, ATP-bound human CFTR. The initial RMSD difference between these two structures is 7.52 Å. During the transition from dephosphorylated to phosphorylated conformation, the RMSD drops to 2.09 Åwhen 100 modes are available and 2.51 Åwhen 30 modes are available.

When the dephosphorylated conformation is selected as the initial conformation and the phosphorylation is selected as the target, in five parallel runs the slowest ANM mode (ANM 1) is selected for the first 5-10 cycles. This is because most of the RMSD difference comes from the positioning of the NBDs. The first ANM mode is about the movement of the NBDs and its hinges are located around the extracellular gate. The most selected ANM modes during this transition are 1, 2, 3, 6, 7, 18 and 23. ANM modes are compared with the GNM modes and most similar GNM modes are determined for further analysis. ANM 1 is most similar to GNM 1. ANM 3 is most similar to a combination of GNM 5 and 6. ANM 7 is again most similar to GNM 1. These three modes are the most important dynamic modes for the transition from dephosphorylated to phosphorylated conformations. The other most selected modes ANM 2 and 6 are most similar to GNM 1 and 2 respectively. Figure 3.28 shows the ANM mode shapes of the most selected modes and their corresponding GNM modes.

The mode shaped of the most selected ANM modes are presented in Figure 3.29. Both ANM1 and ANM3 are only moving the NBDs. While ANM1 is about a movement of the complete NBDs, ANM3 causes a partial movement. ANM7 moves the NBDs, lasso motif and the intracellular loops. This area is also close to where RD is located. Especially in the initial ANM-LD cycles, only these modes are selected. This reveals that during the transition from dephosphorylated to phosphorylated the first movement is the closure of the NBDs.

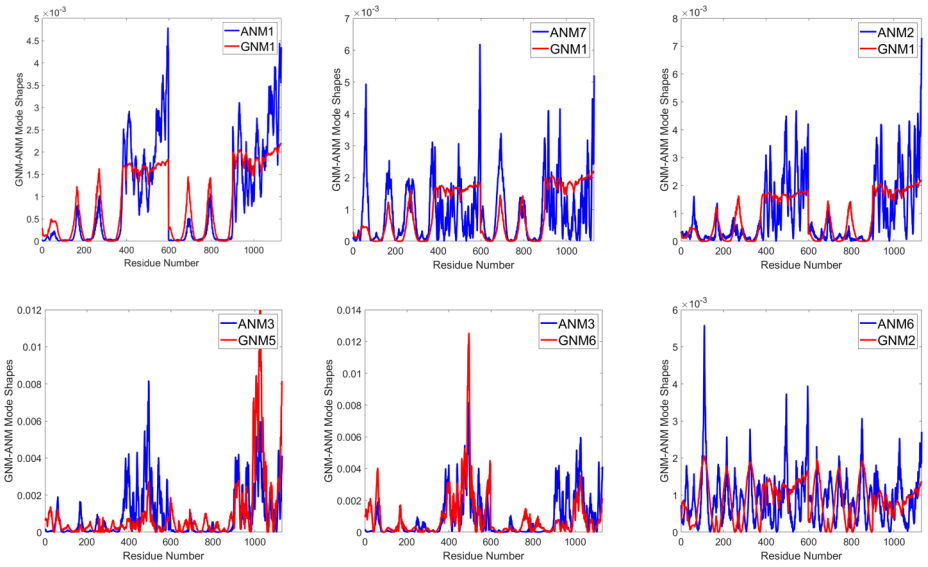

Figure 3.28. Comparison of ANM and GNM modes for 5UAK.

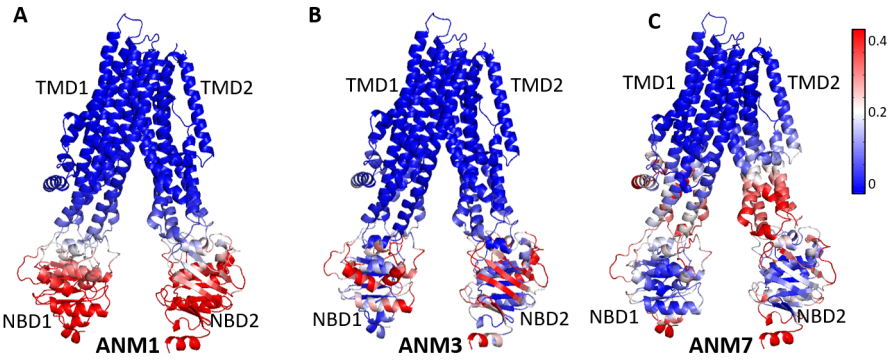

Figure 3.29. Mobilities of the most selected ANM modes, shown on 5UAK.

3.2.1.1. Cross Correlations.

From the ANM-LD cycles, cross correlation maps are obtained. These maps reveal information about the coupling between the movements of different areas of a structure. Along the diagonal is the interaction of each residue with itself, this interaction shows perfect correlation. It's expected that two residues that are next to each other sequentially will also show strong coupling [9]. However, if there is strong correlation between residues that are far apart in space, this indicates an allosteric connection. Figure 3.30 shows the equal time cross correlation map of dephosphorylated CFTR

during its transition to the phosphorylated conformation. Red colors indicate a movement in the same direction and blue colors indicate movement in the opposing direction. The map has continuous sections of red or blue, showing global rigid body movements. The black lines on the plot show the domain terminals. The rigid body movements do not exactly match the domains. ICLs of TMDs also behave as a part of the NBDs. The RMSD converges in the first 48 cycles of the ANM-LD simulations, and fluctuates around 2.09 for the next cycles. For the cross correlation calculations only cycles 1-48 is considered.

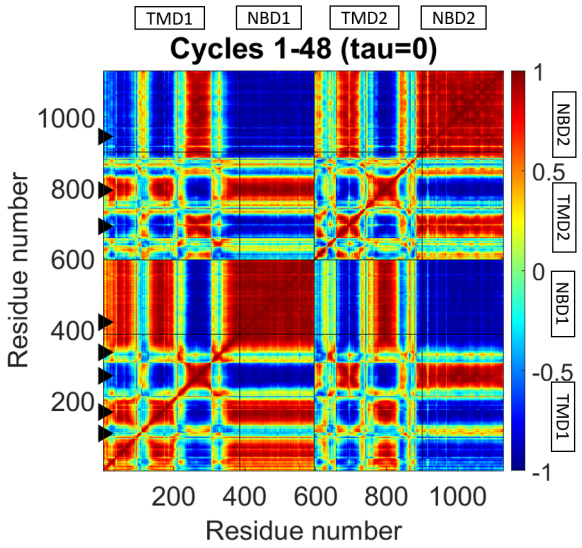

Figure 3.30. Equal time cross correlation map for 5UAK to 6MSM transition.

Eight key residues are colored based on their equal time cross correlations. Figure 3.31 shows the areas each residue is coupled with. The reference residue is shown with a black sphere on the structure. These residues are also shown with black triangles on Figure 3.30. T460 is an ATP binding site from NBD1, it is part of the degenerate ATP binding site. NBD1 is correlated with the lasso motif, and intracellular loops (ICLs) 1 and 4. S1251 is another ATP binding site but it is from the catalytic NBD2. NBD2 is coupled with intracellular loops 2 and 3. T338 is a gating residue, and is slightly correlated with NBD2 (yellow) and highly correlated with the surrounding TM residues. K114 lies on the extracellular loop 1, that is known to essential to gating mechanism and stabilize the open state of the conducting channel [65]. ECL1 is coupled with

intracellular loop 2 (coupling helix 1) and NBD2. R170 lies on ICL1 and is coupled with L1065 on ICL4 up to the permeation pathway. V272 is part of ICL2 and T963 is part of ICL3. ICL2 and 3 are coupled with NBD2. The ICLs that are coupled with the NBDs are topographically on top of their respective NBDs.

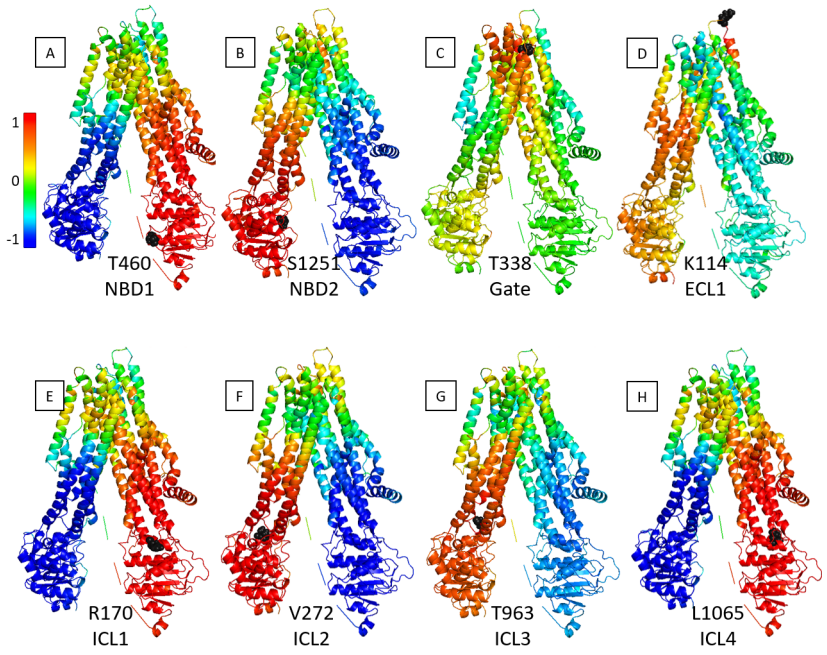

Figure 3.31. Equal time cross correlations colored on 5UAK.

The equal time cross correlation map for this transition is similar to the equal time cross correlation maps of TmrAB for IF-OF transition shown on Figure 3.11. This shows that even though CFTR is an ion channel unlike an active transporter like the rest of ABC transporters, it behaves similarly dynamically. Both TmrAB and CFTR share a similar TMD fold structure and are both Type IV ABC transporters. The most significant difference is in the first 100 residues that make up the lasso motif on the CFTR, this area is coupled with NBD1.

3.2.1.2. Determination of the Leading and Lagging Events.

Time delays are introduced to the cross correlation calculations. This additional time delay reveals information about leading and lagging events. Similar to the equal time cross correlation maps, red coloring on the map shows movement in the same direction

between residues and blue coloring shows movement in the opposite direction. However, time delay means that the movements of the other residues occurs *tau* cycles later. Figure 3.32 shows four different time delay cross correlation maps. The time delay amount is 6, 12, 18, and 24 cycles respectively. The time delays represent temporal coarse graining at different levels. The first one ($TD = 6$) is very similar to the equal time cross correlation maps. The relationship between the correlated segments persists even with the added lagging amount. When the delay is increased to 12 cycles, some of the correlations are reduced and some disappear. Around a delay of 18 cycles, some of the correlations are close to zero, especially auto-correlations for NBDs are close to zero, and some become more apparent or alter the sense. Finally, around a delay of 24 cycles, the sense of the correlations are reversed.

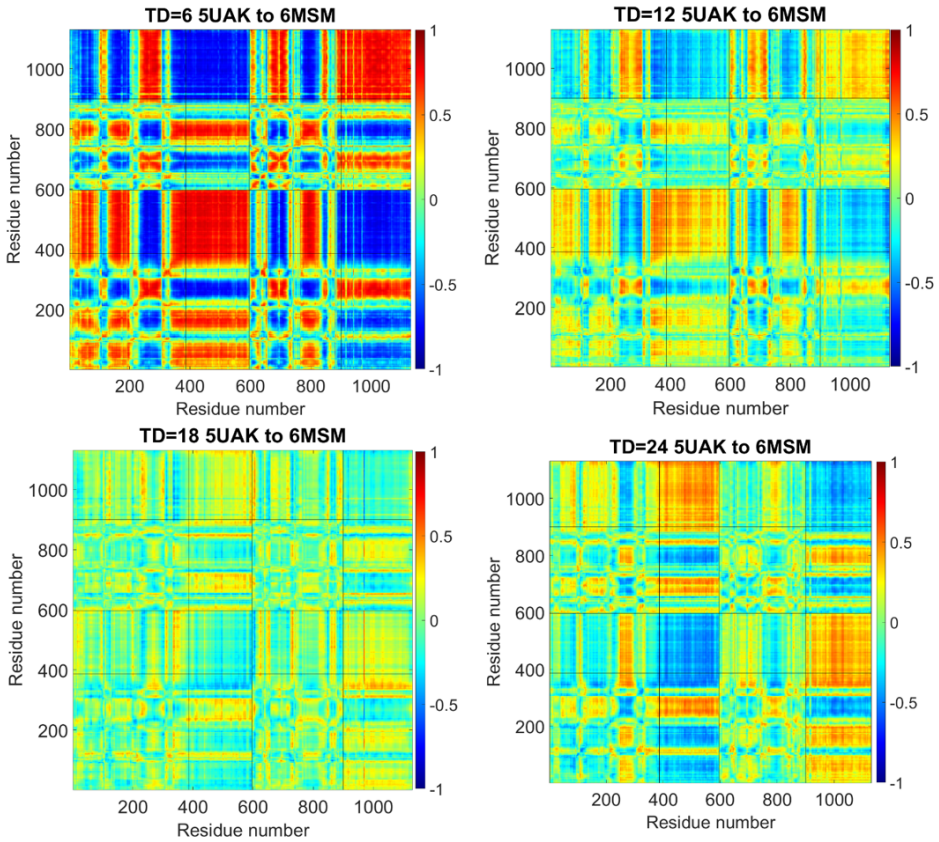

Figure 3.32. Time delay correlation maps for differing time delays for 5UAK to 6MSM transition.

The time delay that causes most correlations to disappear is of interest because the remaining correlations are very persistent. A detailed analysis of the correlated segments is performed for a time delay of 16 cycles. Figure 3.33 shows the time delay cross correlations for a time delay of 16 cycles, for the first 48 cycles of ANM-LD simulations. The colorbar is manipulated to highlight the remaining interactions between residues. In order to observe the behaviors, different key residues are selected as the leading events. These are the same residues from Figure 3.31. Where that figure showed the correlated movements at the same time, Figure 3.34 shows the correlations in time.

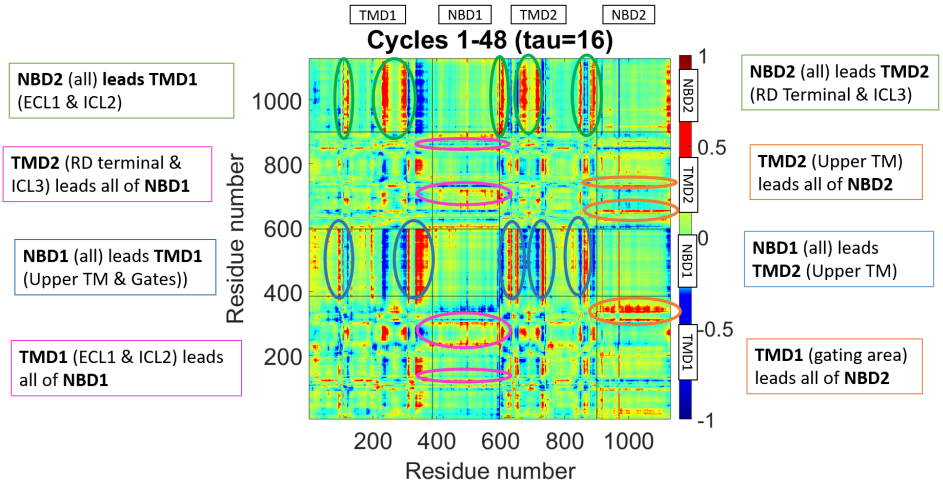

Figure 3.33. Time delay ($\tau = 16$) cross correlation map for 5UAK to 6MSM transition.

When T460 of NBD1 is leading, the movement of upper regions of TMDs appear as the lagging events. This region includes the gating residues such as T338. When the gating area (T338) moves, the lagging movement occurs in the complete NBD2 and ICL2 and ICL3. T338 also slightly leads NBD1, which is colored light blue. When S1251 (an ATP binding residue from NBD2) moves, the helices connected to ICL2 and ECL1 follow. S1251 is also negatively correlated with the gating areas. A movement from V272 (ICL2) and D110 (ECL1) are followed by a movement from all of NBD1. These leading and lagging/following regions imply a plausible dynamic cycle as a signature of this specific transition pathway. The anticipated cycle goes from NBD1, to the gating area on TMDs, to NBD2, to ICL2 and ECL1, and finally back to NBD1. This cycle is supported by the analysis below.

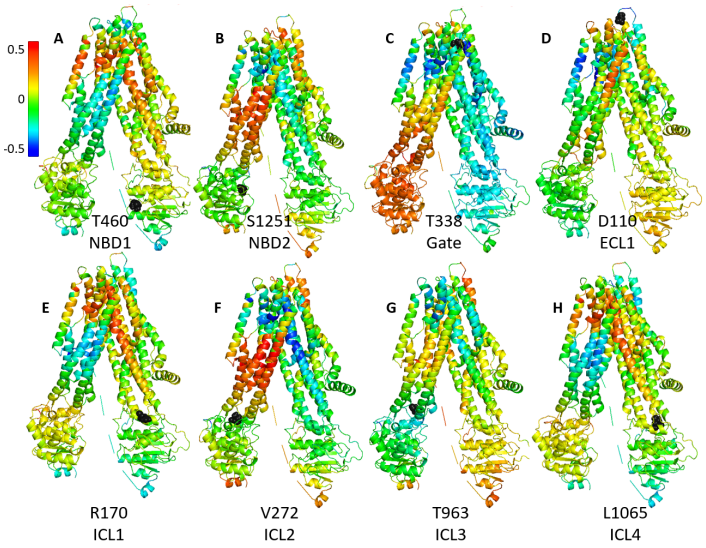

Figure 3.34. Time delay ($\tau = 16$) cross correlations colored on 5UAK.

The time delay cross correlations map for 16 cycles is further analyzed by dividing the time window of 1-48 cycles into three time segments to elaborate the order-of-events during the transition. For this analysis the mean cross correlations are calculated for the complete 1-48 cycles. However, only the given time window is considered. In the first time segment, only the events from cycles 1-10 are considered as the leading events. Due to the time delay of 16, lagging events are from cycles 17-27. Figure 3.35, shows the map and the responses of the movements of 8 key residues. T460 from NBD1 is positively correlated with NBD2 and the gating area. S1251 from NBD2 is positively correlated with NBD1 and ICL2 and ICL3. NBD1 and NBD2 have the opposite correlation with each other. The gating area is not highly correlated with any areas, its relationship to both NBDs appears symmetrical. ICL2, ICL3 and ECL1 are again leading the same events in NBD1. ICL1 and ICL4 are leading the movements in NBD2 and gating area.

The second window is considering the leading events from cycles 11-21 and the lagging events from cycles 27-37. In the first 10 cycles, a significant RMSD drop occurs by the closure of the NBDs. In this window more local movements become apparent. Figure 3.36 shows the results for this time window. The correlations have become much less prominent. Especially the leading events from both NBDs causes very lit-

tle response from the rest of the structure. The most response is generated by the movement of the ECL1 and the helices connected to ICL2. These areas are positively correlated with NBD1 and negatively correlated with NBD2. Gating residue T338 is more correlated with NBD2.

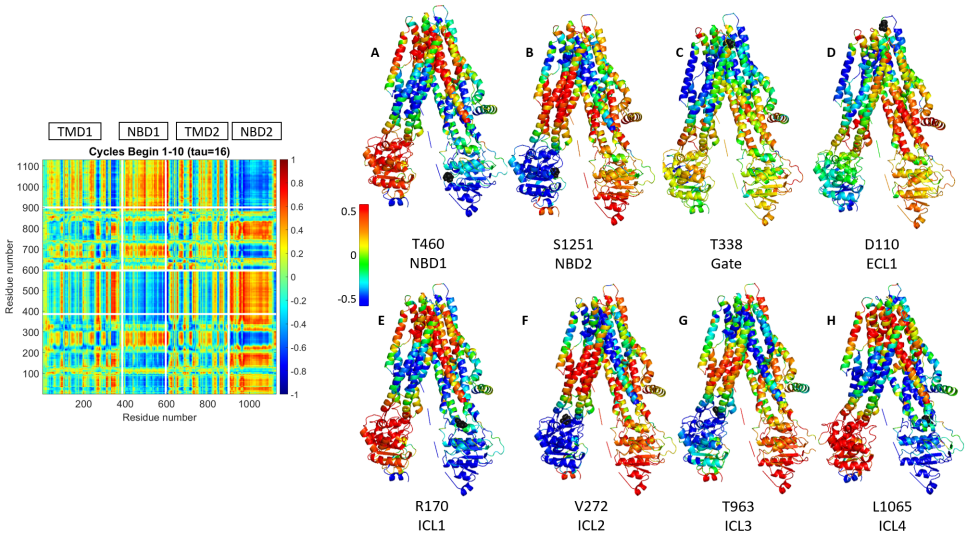

Figure 3.35. Time delay ($\tau = 16$) cross correlations with custom window 1-10.

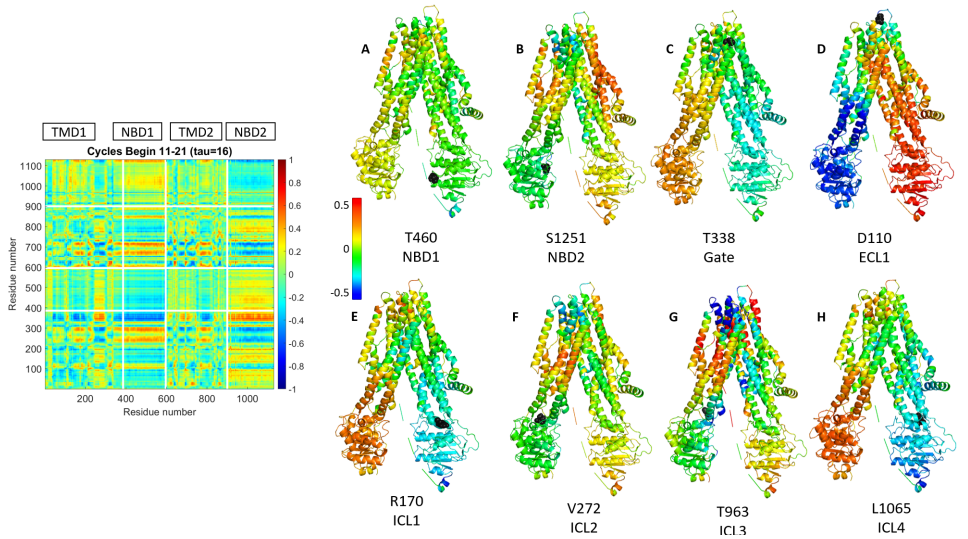

Figure 3.36. Time delay ($\tau = 16$) cross correlations with custom window 11-21.

The final and third window is considering the leading events from cycles 22-32 and the lagging events from cycles 38-48. The pattern in Figure 3.37 is similar to the pattern in Figure 3.36 but in the opposite direction. The most important segments are the

NBDs. They are negatively correlated with each other, leading each other in opposite directions. NBD1 is leading the gating area, and the gating area (T338) is leading NBD2. NBD2 is leading ICL2, ICL3 and ECL1.

Time delay cross correlation analysis gives a detailed insight into the order of events during the transition from dephosphorylated, ATP-free conformation to the phosphorylated, ATP-bound conformation. The selected ANM modes had already shown that the first movements during the transition were by the NBDs. Time delay analysis shows that this movement is followed by the movement of the gating areas. NBDs initially cause a positive movement to the other NBD, then no effect, finally a response in the opposite direction. There are multiple concurrent information flow cycles in the structure. One cycle could be the movement from NBD1 is leading the gating area, gating area is leading NBD2, NBD2 is leading ICL2/3 and ECL1 and ICL2/3 is leading NBD1. Another cycle can be written by starting from NBD2 and going in the opposite direction. This analysis only reveals the order of motions, but not the causal relationships that underlie these movements. In order to determine causality, it is necessary to combine this information with the results from the transfer entropy analysis.

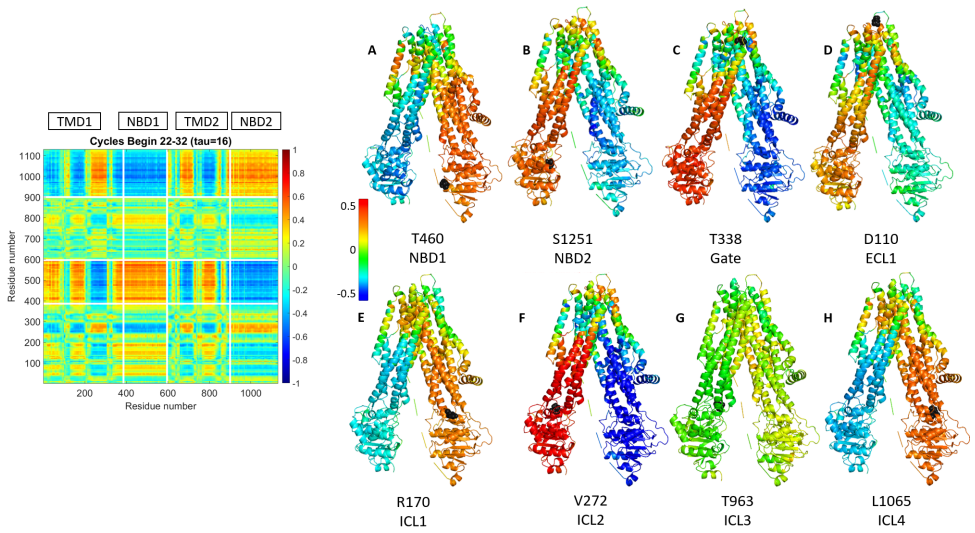

Figure 3.37. Time delay ($\tau = 16$) cross correlations with custom window 22-32.

3.2.2. Transfer Entropy Analysis for CFTR

3.2.2.1. Locations of Entropy Sources and Driving Domains.

Transfer entropy (TE) analysis is first performed for the dephosphorylated human CFTR (PDB ID: 5UAK) for the most collective 10 GNM modes. The map on Figure 3.38. shows how each residue affects the others. On the y-axis are the effector residues (entropy sources) and on the x-axis are the affected residues (entropy sinks or receivers). An unbroken horizontal red line means that that residue (on the y-axis) is greatly affecting all of the other residues. The domain terminals are indicated on the structure by the white lines and tags. ATP binding residues are indicated on the collectivity and cumulative TE plots.

When 1-10 modes are considered for dephosphorylated, ATP-free CFTR (PDB ID: 5UAK) the entropy sources are accumulated mostly on the NBDs. The only entropy sources on the TMDs are on the intracellular loops. The peak residues (entropy source residues) do not perfectly match the ATP binding sites for the dephosphorylated conformation and the peaks broad rather than sharp.

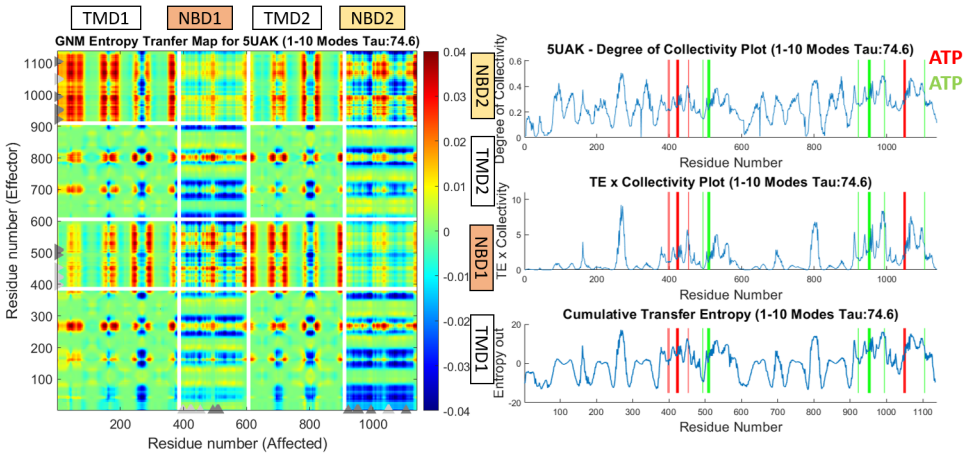

Figure 3.38. Transfer entropy map and collectivity results for 5UAK, 1-10 modes.

Figure 3.38 also shows that NBD1 is mostly affecting TMD2 and NBD2 is mostly affecting TMD1, in contrast to its domain connectivity (TMD1-NBD1 and TMD2-NBD2). This is in agreement with the domain swapping previously observed for Type

IV ABC transporters [66, 67]. Another key observation from this map is that NBD2, which is the catalytic NBD, has a more prominent effect on the TMDs.

When entropy transfer calculations are performed for the phosphorylated, ATP bound CFTR (PDB ID: 6MSM) for 1-10 modes, the entropy source peaks become much sharper. The entropy sources in this case are on the NBDs and the periplasmic region of the TMDs. In this conformation, the ATP binding residues perfectly align with the TE peaks. This points to the conclusion that ATP binding focuses the information flow from the ATP sites. Both degenerate and catalytic ATP binding sites are entropy sources. In this conformation, there are also entropy source peaks on the TMDs, specifically around the extracellular opening. Known ion coordinating residues F337, T338, S341 and N1138 are all entropy sources for 1-10 modes. Additionally, a cleft formed by TM4, 5, and 8 is an entropy source area. In this conformation, all domains contain entropy sources. This indicates that in the conducting state all domains are working cooperatively.

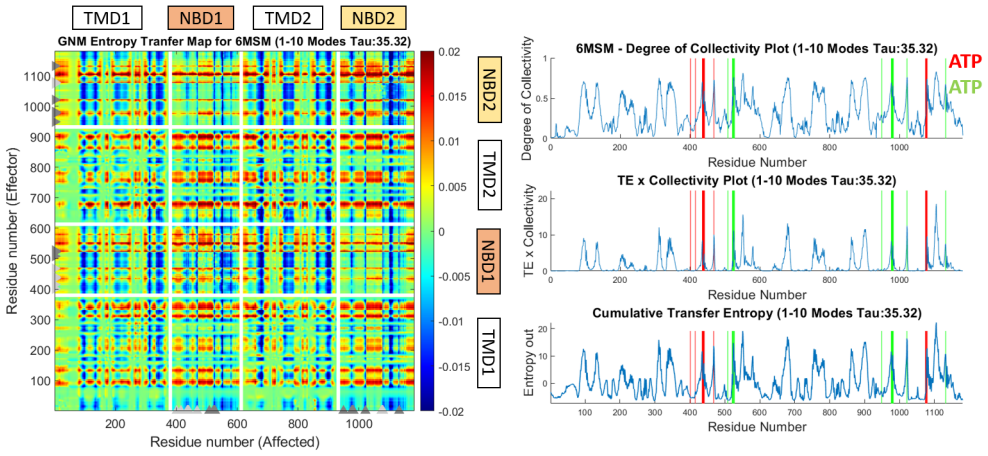

Figure 3.39. Transfer entropy map and collectivity results for 6MSM, 1-10 modes.

When the most collective mode is removed, latent entropy sources are revealed. When 2-10 GNM modes are considered for the dephosphorylated CFTR (PDB ID: 5UAK), the entropy sources are migrated to the ATP binding sites and the periplasmic side. This TE map is similar to the map produced by the phosphorylated CFTR for 1-10 modes. This shows that this information transfer pattern is present in other conformations

but ATP binding makes it the dominant one. The entropy sources on the TMDs are surrounding the gates and the ion permeation pathway.

One more slow mode is removed, and calculations are performed using 3-10 modes for 5UAK. The profile becomes distinctly asymmetrical, TMD1 and NBD1 contain entropy sources. Besides the degenerate/catalytic difference, another possible source of asymmetry is the regulatory domain. While the regulatory domain is uncharacterized in these crystal structures, the areas where RD would bind appear as entropy sources. The entropy source cluster in TMD1 is close to the known helix from RD and the rest are all on the cavity behind the lasso motif. Figure 3.41 shows the entropy source peaks from 1-10, 2-10 and 3-10 modes for the dephosphorylated conformation.

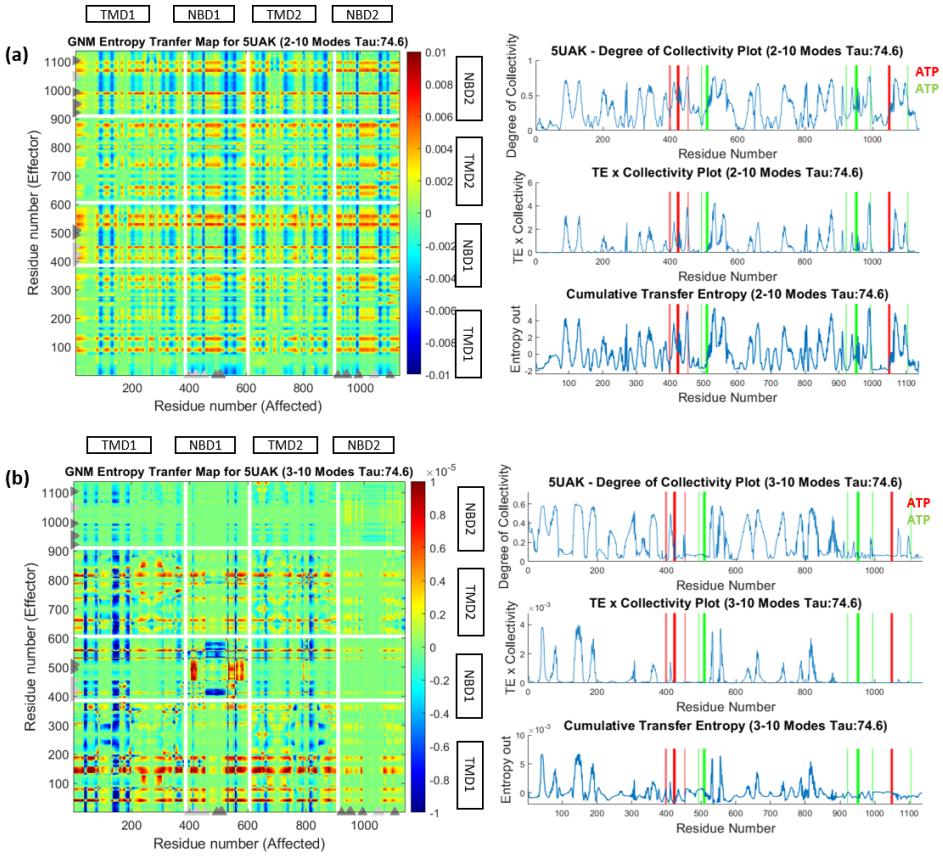

Figure 3.40. Transfer entropy map and collectivity results for 5UAK a) 2-10 b) 3-10 modes.

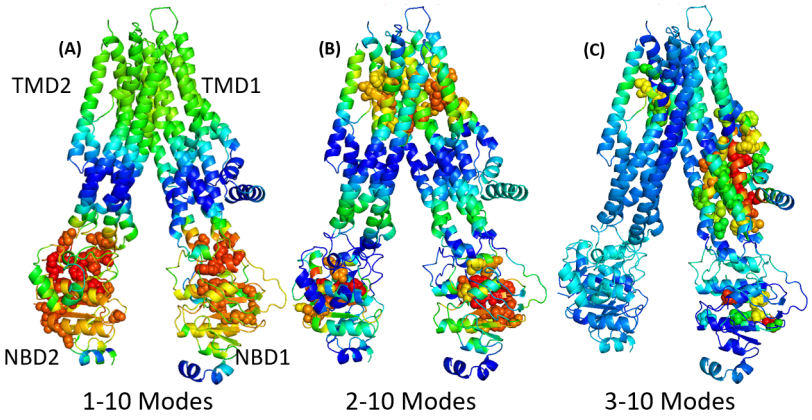

Figure 3.41. Cumulative transfer entropy colored on 5UAK.

For the ATP-bound, phosphorylated conformation, removing the slowest mode (2-10 modes), reveals an asymmetry surrounding the same area where RD would be placed, such as residues 1076-1080 in TMD2. Both for 2-10 and 3-10 modes, there is a clear asymmetry. Zhang et al., have observed an electron density on the side of NBD1 and hypothesized that this is where the other half of RD would be placed [29]. This area appears as an entropy source cluster for both 2-10 and 3-10 modes. Figure 3.42 shows locations of the entropy source peaks for the phosphorylated, ATP-bound conformation for 1-10, 2-10 and 3-10 modes.

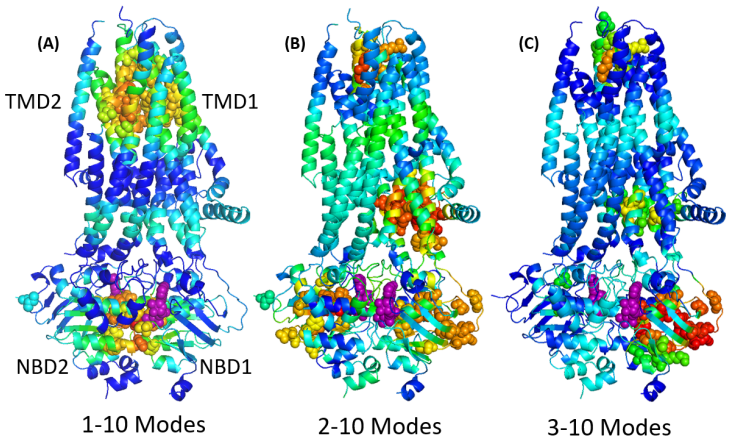

Figure 3.42. Cumulative transfer entropy colored on 6MSM.

Similar to the cross-correlation maps, the transfer entropy maps of CFTR for both conformations are similar to the transfer entropy maps of TmrAB, shown in Figure 3.21

and Figure 3.24. ATP-free CFTR (5UAK) has a similar profile to IF-wide state of TmrAB (6RAN). ATP-bound CFTR (6MSM) has a similar profile to OF state TmrAB (6RAJ). As the ATP-bound structure is not the conducting conformation, it could be presumed that the map for CFTR would resemble OF state TmrAB even more in its conducting conformation. There another key difference between other ABC transporters and CFTR. While other ABC transporters have a periplasmic and cytoplasmic gate, CFTRs only have a periplasmic gate. For 1-10 modes both TmrAB and CFTR show a cluster of entropy source peaks around the periplasmic gate. Removal of the slowest mode reveals other entropy sources surrounding the translocation cavity and especially towards the cytoplasmic end. However, for CFTR even in other mode combinations there is a lack of entropy sources on other parts of the translocation cavity. Transfer entropy method is able to capture this key difference.

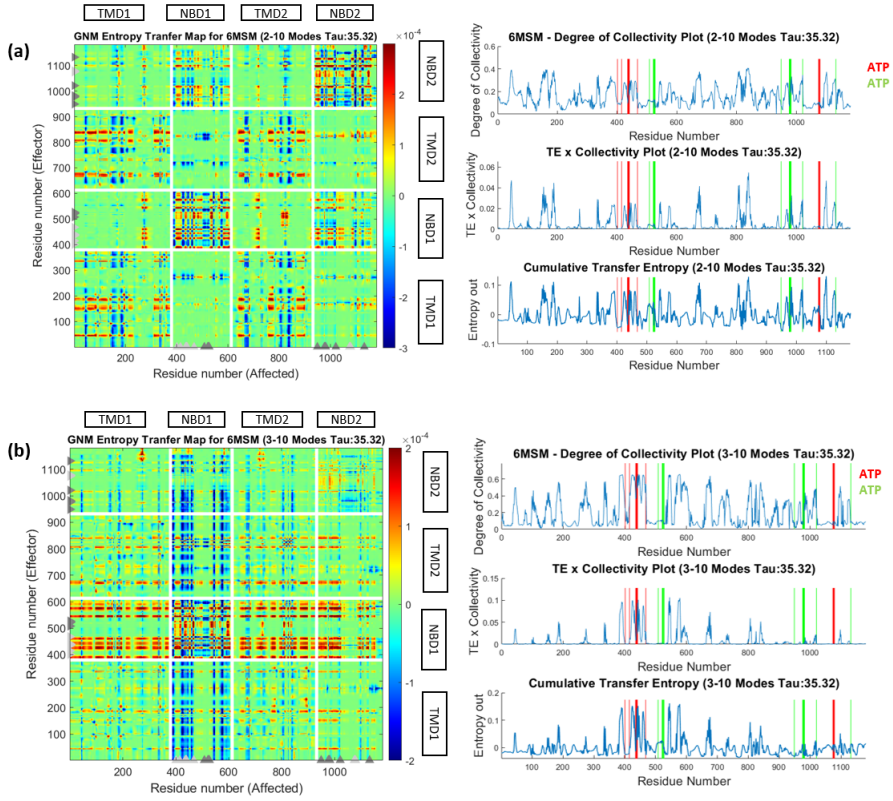

Figure 3.43. Transfer entropy map and collectivity results for 6MSM a) 2-10 b) 3-10 modes.

3.2.2.2. Information Flow from Key Residues.

Besides revealing the locations of the entropy source residues, TE analysis reveals the information flow directions and patterns between specific residues. Each row on the TE maps is how that residue affects each of the other residues. The same key residues are selected for this analysis as the ones for the cross correlation calculations. Figure 3.44 shows the information flow each residue to the rest of the structure. The areas that receive the most information are shown in red. There is information flow from the two NBDs to opposite areas. NBD1 is affecting helices connected to ICL2 and 3. NBD2 is affecting the lasso motif and helices connected ICL1 and 4. There is little information flow from the gating area or ECL1. Coupling helices behave similar to the NBDs they are on top of.

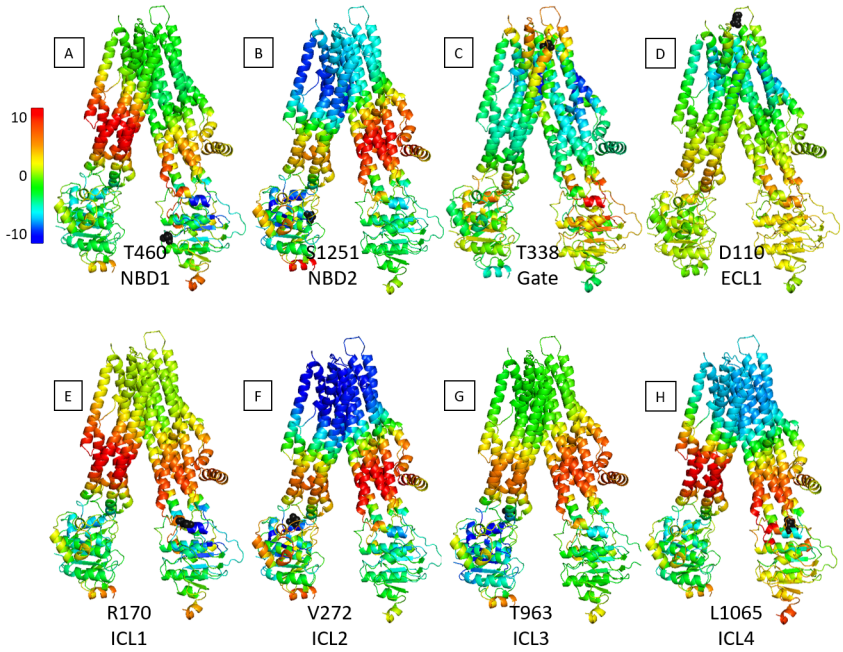

Figure 3.44. Transfer entropy 1-10 colored on 5UAK.

In all available structures only a 17-residue helix from the regulatory domain (RD) is characterized. When these residues are incorporated into the calculations, they appear as entropy source residues. The structure is colored for a single residue on this helix (unknown amino acid), the effects are observed on the periplasmic side of TMDs, surrounding the gating residues. Additionally, the lasso motif and the terminal of NBD1 is affected. This area on NBD1 is where the RD would be connected. RD is known to

physically block the ion transport channel in the dephosphorylated conformation [29]. Figure 3.45 shows the affected areas from the regulatory domain residue. The effects on the periplasmic gate points to an allosteric control by the RD on transport.

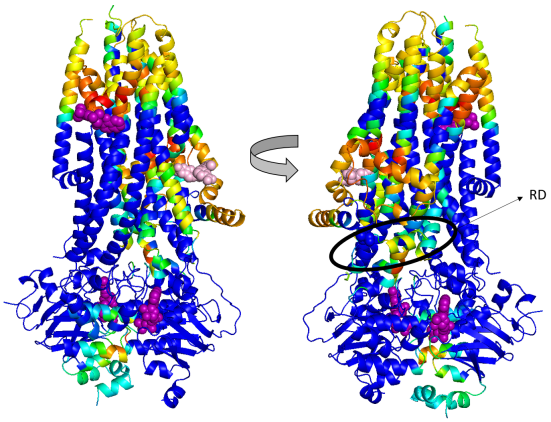

Figure 3.45. Affected residues from RD.

3.2.2.3. Transfer Entropy Using Most Selected Modes.

The most selected ANM modes during the transition from dephosphorylated, ATP-free to phosphorylated, ATP-bound conformations are 1st, 3rd, and 7th modes. When GNM modes were compared to these three, ANM1 and 7 were similar to GNM1 and ANM3 was similar to GNM 5 and 6. Transfer entropy calculations are repeated again with these modes that are essential for conformational change. Figure 3.46 shows the transfer entropy maps for 1, 5, and 6 modes. For this mode combinations, most of the entropy source residues are located on the NBDs and the ICLs. This is similar to the distribution observed for 1-10 modes in Figure 3.41.

Figure 3.47 shows the selected key residues colored for their effects based on the calculations of 1, 5, and 6 modes. T460 from NBD1 and S1251 from NBD2 are affecting the upper regions of the TMDs, close to the gating area. Gating residue T338 is slightly affecting NBD2. ICL1 and 2 have a similar effect profile to the NBDs for this mode combination.

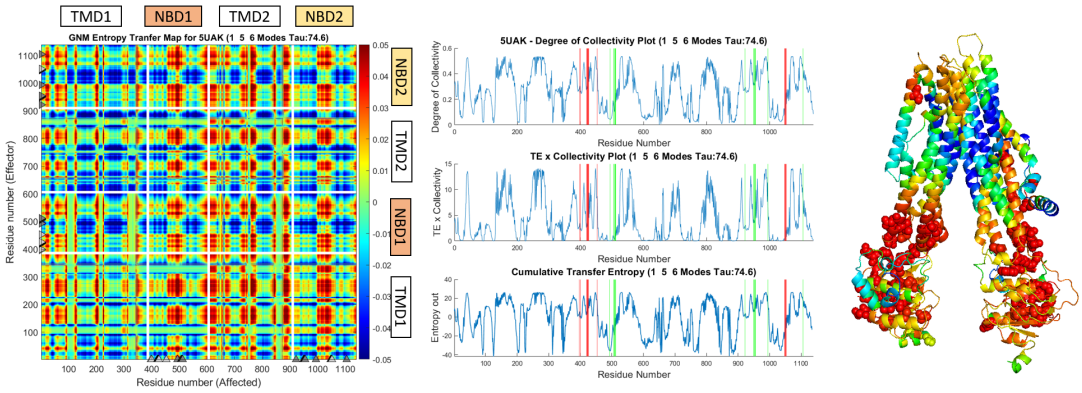

Figure 3.46. Transfer entropy results for 5UAK, 1-5-6 modes.

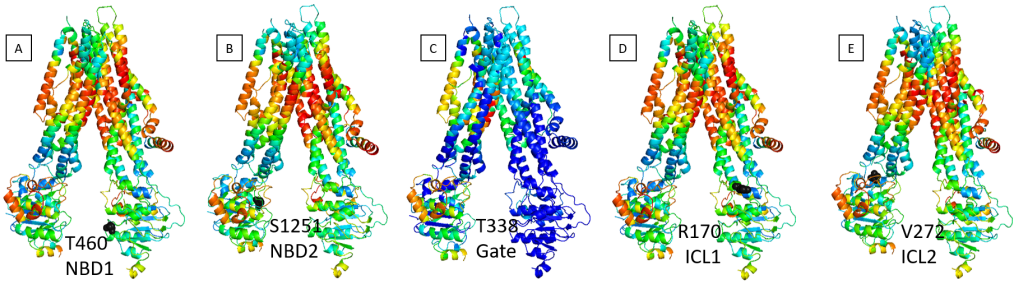

Figure 3.47. Transfer entropy colored 5UAK, 1-5-6 modes.

3.2.3. Connecting ANM-LD and TE Results

ANM-LD results and the subsequent time delay cross correlations analysis revealed the order-of-events during the dephosphorylated to phosphorylated conformational transition. The NBDs started the movements and the gating areas followed. The ICLs and coupling helices served as a connection point between the two NBDs. This analysis had revealed the leading and lagging events but not the causal relationships. ANM-LD results also identified the key modes for this conformational change, GNM modes 1, 5, and 6. The transfer entropy calculations for this mode combination has the same information flow pattern as the time delay cross correlation relationships.

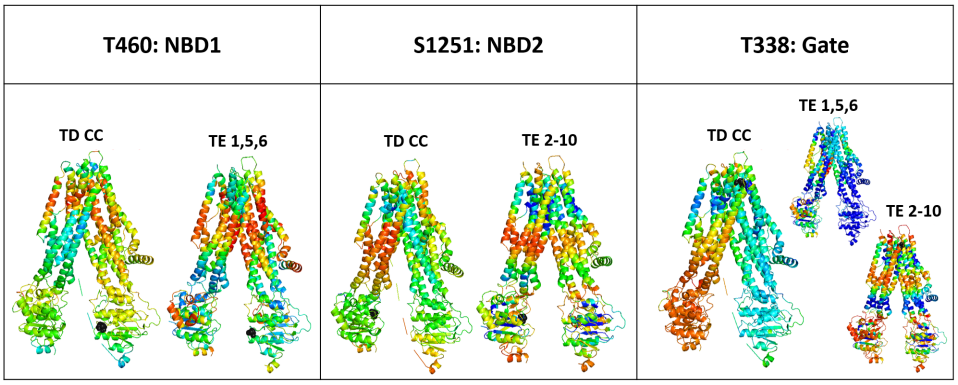

Figure 3.48. Similar TD-CC and TE effects.

There is clear information flow from the NBDs to the gating areas. As the 1st mode is the most dominant, its removal reveals latent information pathways. TE results for 2-10 modes shows the information flowing to ICLs, which are the other key areas during transition. Also for both 1, 5, 6 modes and 2-10 modes, there is more information flow from the gating area to NBD2. Figure 3.48 shows the similarities of the TD-CC and TE results.

3.2.4. Mechanism and Dynamics of CFTR Drugs

There are multiple drugs known to be effective in cystic fibrosis treatment. Potentiators that increase open-channel probability and correctors that increase 3D structure stability are two types of drugs that are commercially available. Ivacaftor is a potentiator drug that is especially used for the common G551D mutant [34]. This drug is known to bind to a cavity formed by TM4, TM5 and TM8. Liu et al., have obtained the ivacaftor and ATP-bound crystal structure in 2019 (PDB ID: 6O2P). Calculations on this structure and comparison with the ivacaftor-unbound structure reveal important features about the mechanism of action. Figure 3.39 had shown an entropy source cluster surrounding the ivacaftor binding cavity formed by TM4, 5, and 8. All of the ivacaftor binding residues are entropy sources.

Lumacaftor and tezacaftor are two commercially available corrector drugs. Fiedorczuk et al., have reported the lumacaftor bound crystal structure in both dephosphorylated

(PDB ID: 7SVR) and phosphorylated (PDB ID: 7SVD) conformations [36]. Both of these drugs bind to the same cavity formed by TM1, 2, 3, and 6 behind the lasso motif. This drug is reported to work by linking these TM helices that are thermodynamically unstable.

For the dephosphorylated conformation, Figure 3.40 shows that this exact cavity is an entropy source cluster for 3-10 modes. For the phosphorylated conformation, this area is revealed to be an entropy source when even more modes are removed and calculations are performed for 5-10 modes. In fact, all of the lumacaftor binding residues perfectly align with the peaks on TMD1 for 5-10 modes. Figure 3.49 shows the drug bound structures colored for the mode combinations that cause their binding cavities to become entropy source clusters.

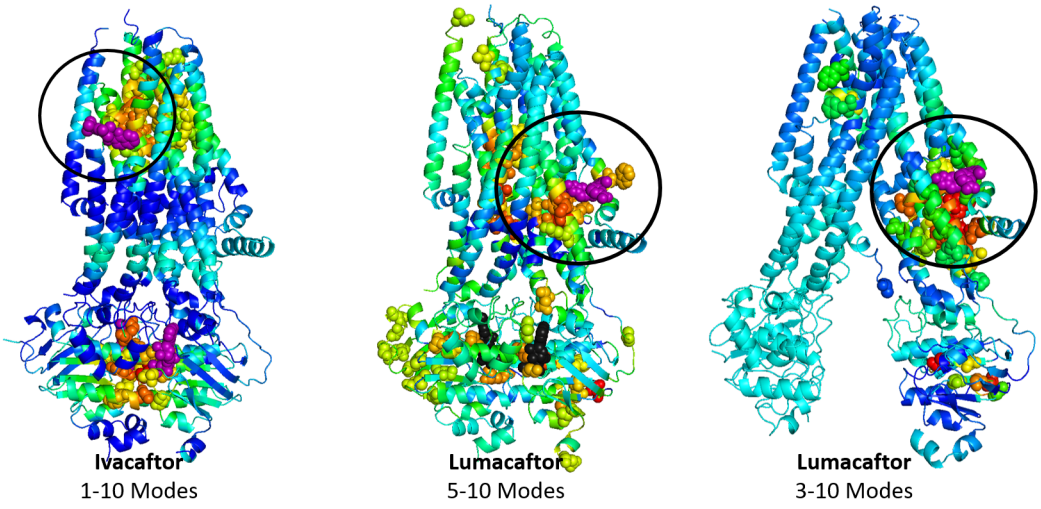

Figure 3.49. Transfer entropy results that show drug binding cavities as entropy sources.

Next, the affected residues from TE calculations are analyzed. Understanding how drug binding residues affect the rest of the structure is a useful tool in understanding the mechanisms of action. Commercially available potentiator drug ivacaftor is known to interact with F312 from TMD1. It's also one of the entropy source residues for the phosphorylated conformation calculated using 1-10 modes and it is highly collective residue. Figure 3.50(b) shows that F312 has a global effect on the structure, both ATP

binding sites and multiple TM helices are affected. Lumacaftor binding regions are also affected, this may point to an inherent relationship between these two drug binding regions. This is notable because these two drugs are often administered together [35]. Ivacaftor binding residues are also entropy source residues for 4–10 modes. The effects of another ivacaftor binding residue F931 is observed. It affects ion permeation pathway residues L102 and Y914. For both 1-10 and 4-10 modes, the effects are global.

Potentiator drugs are known to increase the open probability of the ion transport channel by decoupling ATPase from gate opening [34]. An interesting observation is that three key sites affect the same areas on the structure. Ion permeation pathway residue F337, ivacaftor binding residue F312 and ATP binding residue G551 are all entropy sources for 1-10 modes for the phosphorylated, ivacaftor bound conformation. Their profiles, how much entropy is transferred to other residues, are highly correlated. The correlation coefficient between the profiles of affected residues of Ivacaftor BR and ATP BR is 0.953. The correlation coefficient between gate and ivacaftor is 0.979. G551D is a common mutation treated by ivacaftor. The similarity between their effects could point to the compensation of the allosteric effects of ATP binding by ivacaftor.

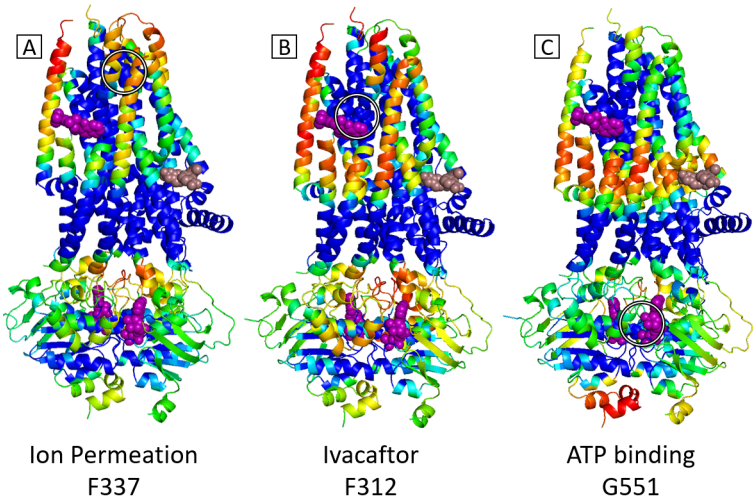

Figure 3.50. Affected residues from F337, F312, and G551.

Another commercially available drug is lumacaftor and it is a corrector drug. Lumacaftor is known to structurally link TM1, 2, 3 and 6, and stabilize TMD1 [36]. It binds to a

cavity behind the lasso motif. On phosphorylated CFTR with lumacaftor (PDB ID: 7SVD), a lumacaftor binding residue S364 is an entropy source for 5-10 modes. The effects of S364 are seen specifically on TMD1 on helices TM1, 2, and 6, very local compared to ivacaftor. For the dephosphorylated CFTR (PDB ID: 7SVR), lumacaftor binding residue F77 has a more spread out effect but mostly around ATP binding site 1 and the ivacaftor binding site. Figure 3.51 shows the both of these structures and the effects of the selected lumacaftor binding residues.

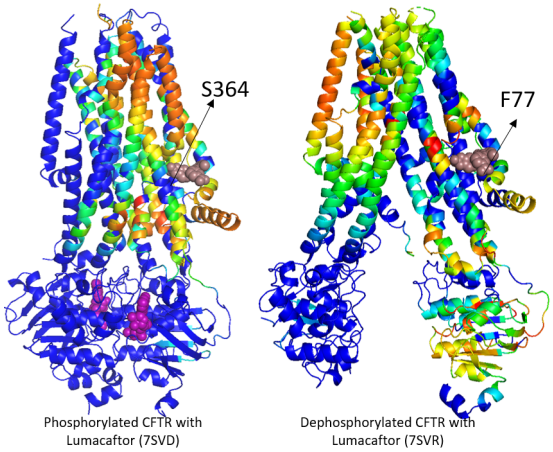

Figure 3.51. Affected residues from S364 and F77.

3.3. Understanding the Mechanism of Action of the Manganese Transporter

3.3.1. Analysis of Crystal Structure of PsaBC

Manganese transporter MntBC from *Bacillus anthracis* and PsaBC from *Streptococcus pneumoniae* are part of the ABC transporter superfamily. Neville et al., have obtained the crystal structure of PsaBC in its open-inward facing (IF) conformational state [14]. First, GNM-based transfer entropy calculations are performed on the available crystal structure (PDB ID: 7KYP). Then, these results are compared to the available models for MntBC.

The crystal structure lacks the ATPs and there is no ligand bound to it. As there are no ATPs bound to the NBDs, they are separated. The exact ligand binding residues are unknown and their prediction is of interest. Additionally, the identification of any allosteric sites offers great potential in the manipulation of manganese transporter function, and the control of anthrax [13].

Transfer entropy analysis is performed on PsaBC for simple GNM mode combinations. When the first 10 slow modes are taken into account, the driver of the information flow appears to be the NBDs. Most of the information flow occurs from the NBDs to the TMDs. However, there is no interaction between the two NBDs. TMDs are also interacting with each other, this might indicate the requirement for cooperative behavior for the passing of the substrate. Additionally, collectivity of each residue is calculated. The largest entropy source residues surround the transport channel and the possible ATP binding sites.

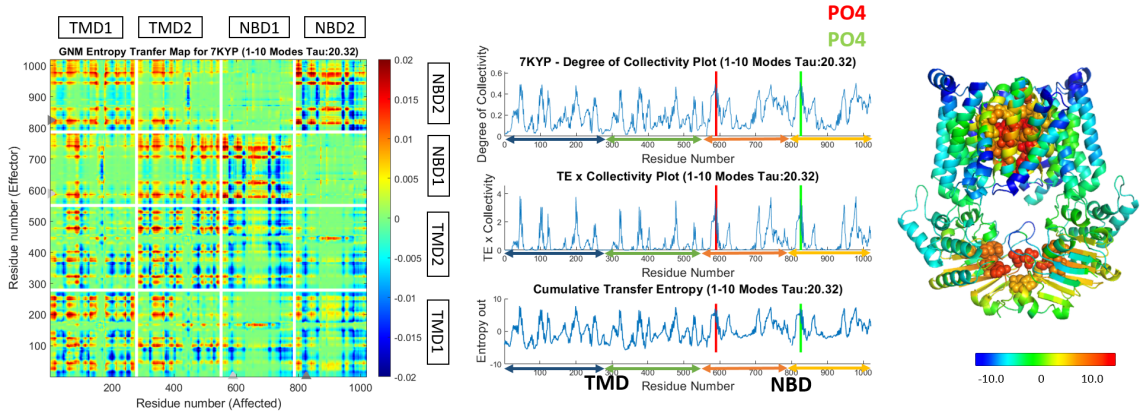

Figure 3.52. Transfer entropy maps and collectivity plots for 7KYP, 1-10 modes.

Multiple known key residues are on the peaks of Figure 3.52, the largest peaks are shown on the structure with spheres. Cytoplasmic gate (L43), coordination sites (D46, H50), pore closure residue (L104, I199), extracellular gate (F121, F125) are examples (Neville et al., 2021). Four other residues are on the peaks but have not been identified as key residues. G97, V202, S250 and V253 have been suggested as key residues involved in substrate transport activity. The peaks for the cumulative transfer entropy plot are labeled and shown on structure in Figure 3.53. Unknown blind predictions are

shown on the structure. The entropy source peaks on TMDs are shown in Table 3.6, known sites are labelled and unknown sites are suggested.

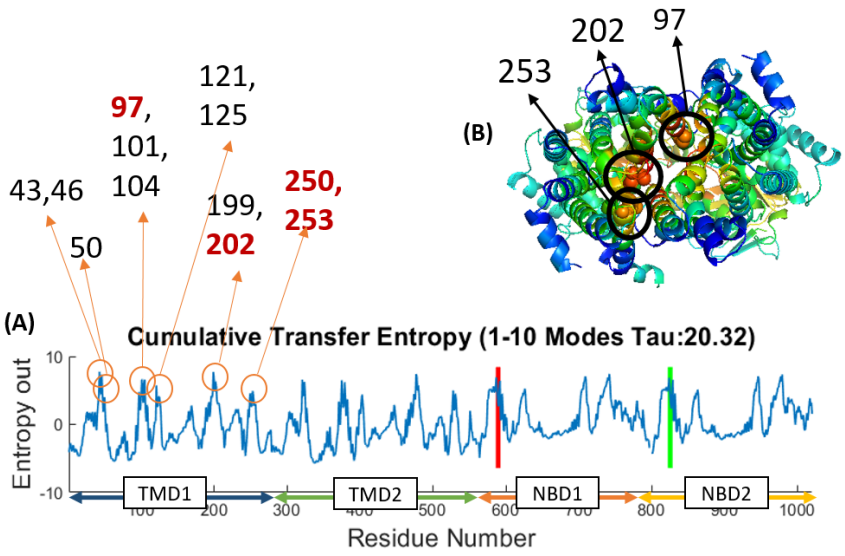

Figure 3.53. Cumulative transfer entropy peaks for 7KYP, 1-10 modes.(a) Peaks and corresponding residues are labeled on the plot. (b) Predicted key residues are shown on structure from top view.

Table 3.6. PsaBC transfer entropy peak residues and known or predicted key sites.

Residues with peak TE	Known or Predicted Sites
L43	Cytoplasmic gate
D46, H50	Coordination site
G97	Unknown blind prediction
S101	7KYP is the F100SPsaBC variant
L104	Control pore closure
F121, F125	Extracellular gate
I199	Control pore closure
V202	Unknown blind prediction
S250, V253	Unknown blind prediction

When the slowest mode is removed (2-10 modes), this reveals a latent allosteric communication path. The driving domains are the NBDs. However, the largest sources of entropy and the most collective residues are not at the NBD interface but more towards the sides. The only entropy source residues on the TMDs are the coupling helices. When another mode is removed (3-10 modes), the driving domain becomes the TMDs. All of the entropy source residues are clustered on the cytoplasmic side of helix TM8, specifically between residues 226-232. It's possible that this helix is controlling the behavior through its interaction with the membrane and it is an allosteric key site.

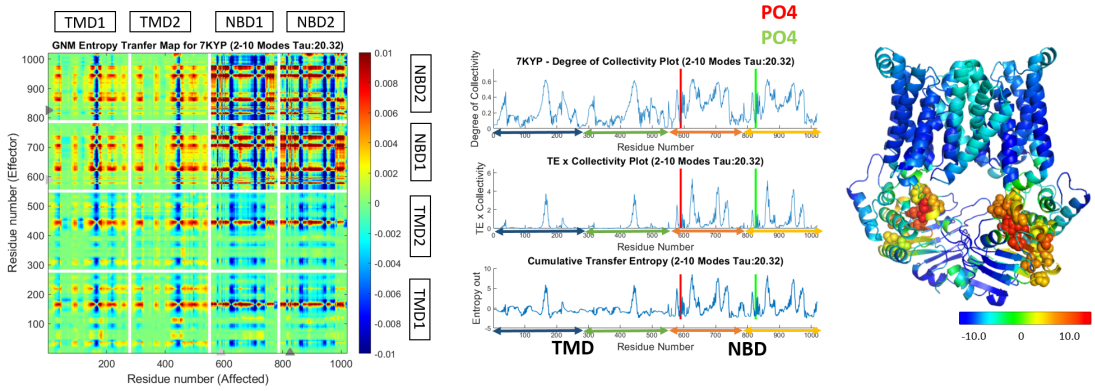

Figure 3.54. Transfer entropy maps and collectivity plots for 7KYP, 2-10 modes.

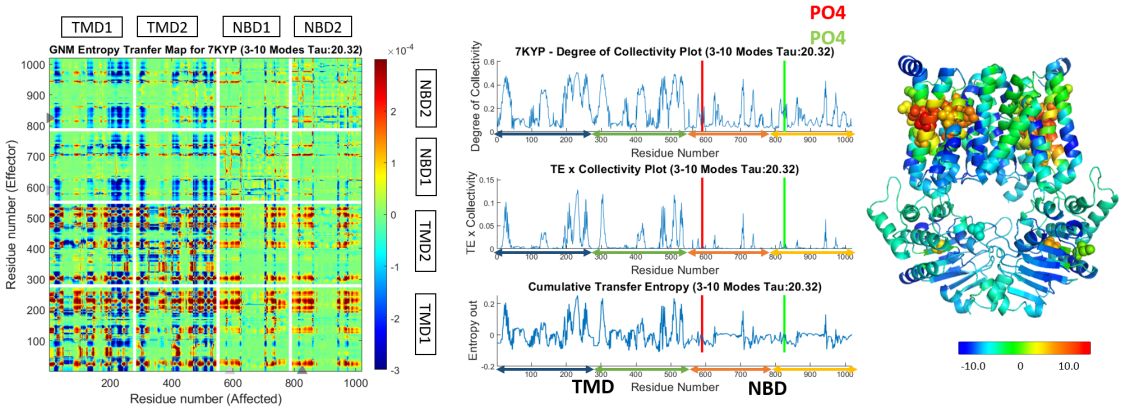

Figure 3.55. Transfer entropy maps and collectivity plots for 7KYP, 3-10 modes.

Another important point of comparison for the manganese transporter is comparing it with other well characterized ABC transporters. Both structurally and from the dynamic characterization of PsaBC, it is most similar to other type II importers.

Specifically, it is dynamically similar to heme importer BhuUV and cobalamin importer BtuCD in IF conformation [24, 40]. However, both cobalamin and heme are metal chelate complexes and manganese is an ionic metal. Therefore, manganese is significantly smaller than both in size and has a positive charge.

This similarity allows the prediction of how manganese transporter will behave when its substrate binding protein PsaA binds to the extracellular side of the transporter. Both BhuUV and BtuCD have crystal structures available in SBP bound conformations. Figure 3.56 shows the transfer entropy map for SBP-unbound (PDB ID: 5B57) and SBP-bound (PDB ID: 5B58) conformations of heme importer BhuUV. In the SBP-unbound conformation, the information transfer occurs between cross domains (NBD1-TMD2, NBD2-TMD1). However, unlike PsaBC the TMDs aren't interacting. For both PsaBC and BhuUV, the driving domains are the NBDs. The similarity between the SBP-unbound TE maps, supports the idea that these two transporters behave similarly dynamically. In the SBP-bound conformation BhuUV-T, the entropy source residues are on the NBDs and the periplasmic gates on the TMDs. SBP is also an entropy source.

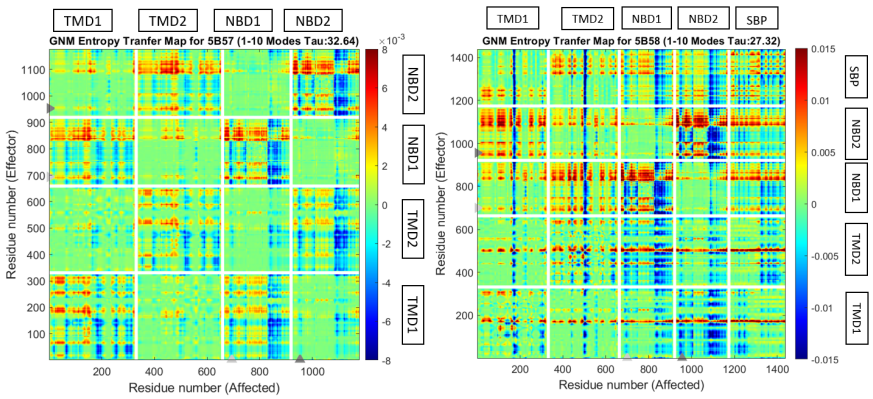

Figure 3.56. Transfer entropy maps and collectivity plots for 5B57 and 5B58 for 1-10 modes.

It's reasonable to assume that for the manganese transporter ATP binding occurs before the binding of SBP and manganese transport. Acar et al., have shown evidence of this claim for another type II ABC transporter BtuCD [9]. Neville et al., have

hypothesized that nucleotide binding drives the rearrangement of TMDs, rather than the docking of SBPs [14]. Transfer entropy results also point to NBD control in the inward facing state. It's likely that ATP binding to the NBDs trigger a change in the TMDs. The periplasmic residues on the TMDs then interact with the SBP. After the SBP is bound, SBP and NBDs coordinate to facilitate substrate transport.

3.3.2. Analysis of MntBC Models

While calculations are performed for PsaBC from *Streptococcus pneumonia*, the actual protein of interest is MntBC from *Bacillus anthracis*. MntBC is the manganese transporter protein that is directly related to the disease-causing mechanism of anthrax. There are multiple models for MntBC, a homology model based on PsaBC, Hidden Markov model with HHPred [41] and the AlphaFold model based on neural network algorithms [42]. The dynamic characteristics of these models are analyzed using the transfer entropy model and compared against the results from PsaBC. While MntBC and PsaBC are not from the same species, they are significantly similar in terms of sequence and structure. This analysis also provides feedback on the reliability of the tested models.

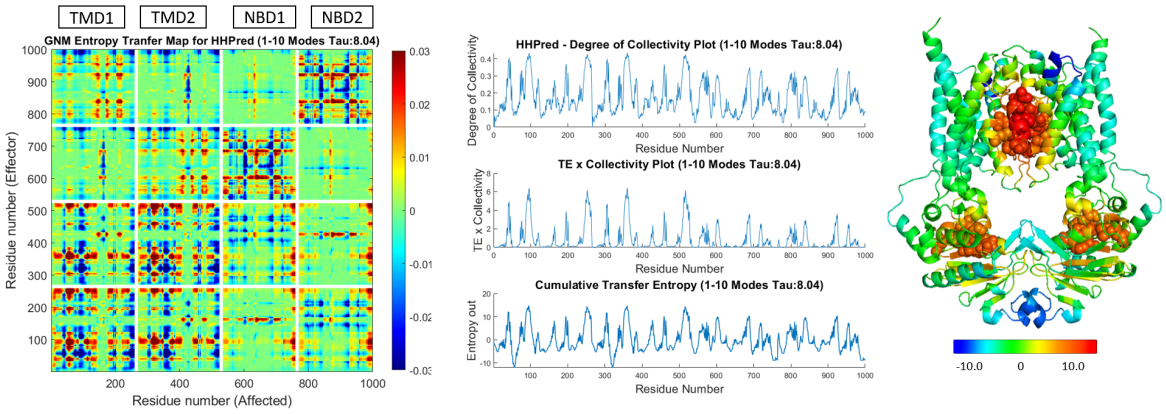

Figure 3.57. Transfer entropy maps and collectivity plots for HHPred model of MntBC for 1-10 modes.

Out of the four models, HHPred model appears to deviate the most from the PsaBC results. The RMSD between HHPred model and PsaBC is 6.39. For 1-10 modes,

the entropy source results are no longer at the NBD interface. The entropy sources surrounding the translocation cavity have shifted downwards towards the cytoplasmic side. Notably, there is also low collectivity for 2-10 and 3-10 modes. Figure 3.57 shows the TE map and collectivity plots for the HHPred model for 1-10 modes.

The RMSD between the AlphaFold model and PsaBC is 2.53. It's less than the HHPred model but notably there is an extended helix at both TMD terminals. This helix appears to distort the results. For 1-10 modes, entropy source residues are spread out towards the sides of the NBDs, rather than at their interface near the ATP binding sites. For lower mode combinations the effect of the extended helix is more pronounced. Figure 3.58 shows the TE map and collectivity plots for the AlphaFold model for 1-10 modes. The translocation cavity is also not covered with entropy sources as it was for the crystal structure of PsaBC.

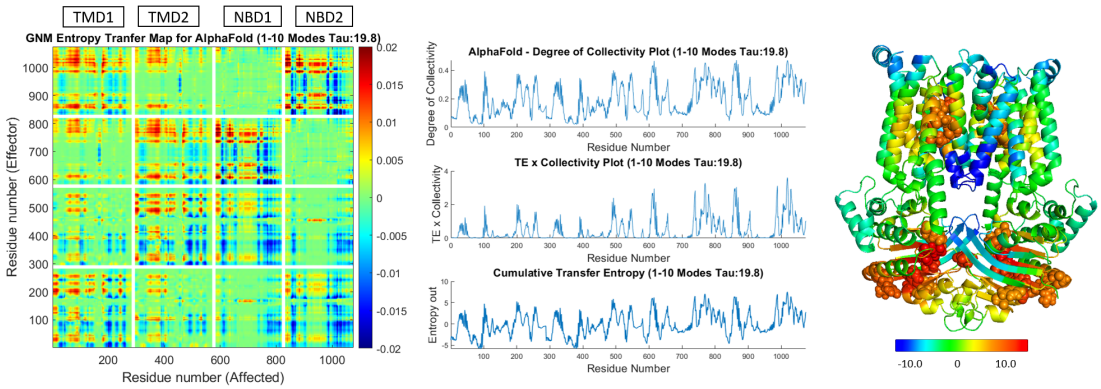

Figure 3.58. Transfer entropy maps and collectivity plots for AlphaFold model of MntBC for 1-10 modes.

The calculations are repeated by manually removing this helix from the model. When the helix is removed, the profile becomes similar to the PsaBC TE map. Most notable change is at the previously described allosteric cavity on the cytoplasmic side of the TMDs. The entropy source cluster has shifted to the neighboring helix onto residues 129-141. Figure 3.59 shows the differing sites for PsaBC (7KYP) and AlphaFold model. It's possible that these two helices form a cavity, and both of them are significant. However, AlphaFold is not as reliable for non-monomer structures or structures with multiple stable, known conformations.

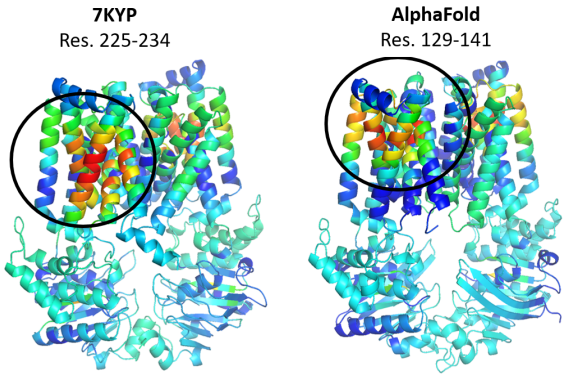

Figure 3.59. Transfer entropy peak cluster from 3-10 modes for 7KYP and AlphaFold model of MntBC.

The homology model results are almost identical to the PsaBC results. The RMSD between these structures is 1.21. Figure 3.60 shows the TE map and collectivity plots for the homology model. All of the entropy source peaks are at the same locations as Figure 3.52. These residues include residues that control pore closure, ion coordination sites, periplasmic gates, and cytoplasmic gates. Along with likely ATP binding sites. Again for 3-10 modes, the entropy source cluster on TM8 is present. Residues 228-236 appear to form an allosteric cavity, especially G229 and L233 are good candidates for mutational analysis. This similarity supports the usage of this model for further calculations.

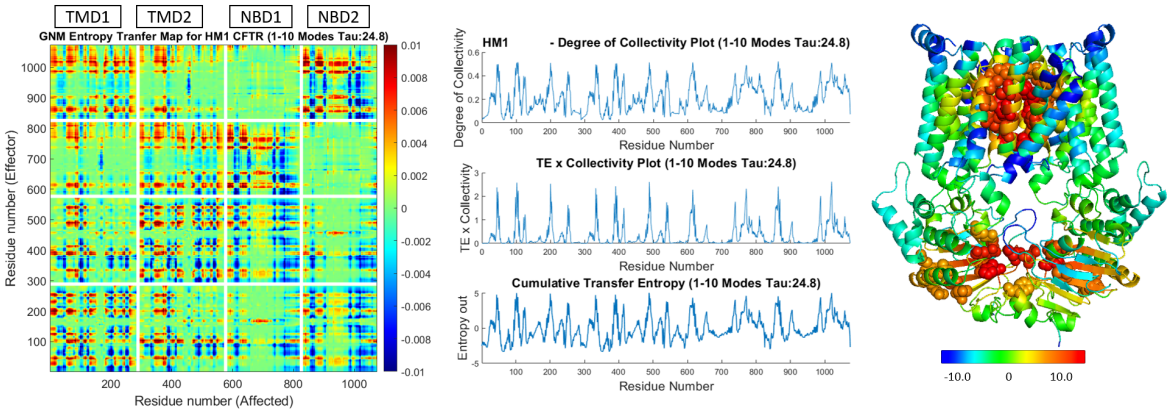

Figure 3.60. Transfer entropy maps and collectivity plots for Homology Model of MntBC for 1-10 modes.

3.4. Dynamic Characterization and Classification of ABC Transporters

There have been multiple characterization and classifications of the ABC transporter superfamily. Initial classifications were made based on the transport direction. Others have used sequential information and evolutionary relationships for classification [16]. Recent developments in x-ray crystallography and cryo-EM imaging have made it possible to determine the 3D structures of many proteins including ABC transporters. These developments have made a classification based on TMD folds and structure possible [22]. There are seven types based on TMD folds, for this analysis on types I, II, and IV are considered.

Another key aspect to transporters is their dynamics. There are many key questions to uncover the dynamic characteristics. How do the four core domains of ABC transporters and the other accessory domains interact with each other? Which domains are the drivers of key events? Is it possible to obtain the same behavior without the presence of a substrate binding protein (SBP)? How can symmetrical or pseudo-symmetrical structures produce asymmetrical behaviors? How does the information flow change between conformations? Which key sites, such as cytoplasmic and periplasmic gates, ATP binding sites, coupling helices etc., are information sources?

Previous results sections have highlighted similarities between different members of the ABC transporter superfamily. Even though CFTR is not an active transporter like the other type IV ABC transporters its correlated movements was similar to the global rigid body movements of TmrAB as shown on Figure 3.11and Figure 3.30. Information flow patterns obtained from transfer entropy method highlighted that CFTR lacks the cytoplasmic gate found in other ABC transporters, through the lack of entropy sources around that area. The comparison of manganese transporter MntBC, with cobalamin transporter BtuCD and heme transporter BhuUV revealed the interaction between TMDs and SBPs. These conclusions were only possible through the comparison of the dynamics of ABC transporters. This section aims to characterize many ABC transporters in a similar manner and to utilize the similarities between them to make new conclusions.

In order to answer these questions and more, multiple computational approaches have been utilized. Mainly, the transfer entropy method was used to understand the information flow patterns between domains [10]. Collectivity of each residue is also calculated as described in the Methods section [15]. For each of the transporters the conformation and bound ligands are reported in the summary tables along with the driving domain and if the ligand binding sites are entropy sources. Some type I and II transporters feature substrate binding proteins (SBP). Results are repeated with and without the SBPs. One feature to consider is that importers tend to have a resting state in the outward facing conformation (OF), whereas exporters tend to rest in the inward facing conformation (IF) [21]. The results presented in this section are only considering 1-10 modes for the transfer entropy calculations. This characterization of ABC transporters has shown that transporters with similar TMD folds tend to have similar dynamic profiles [22].

Table 3.7. Dynamic characterization results for Type I transporters.

Transporter	**PDB**	**Conformation**	**Ligands**	**Driving Domain**	**Peak Match**
MalFGK (maltose)	3PUY	OF with SBP	Maltose, ATP(2)	NBD + SBP	Yes
		OF		TMD	Yes
	3FH6	IF	-	TMD	Yes
AlgM1M2SS (Alginate)	4TQU	IF	Alginate	TMD	Yes
SugABC (trehalose)	7CAG	OF with SBP	Tre., ATP(2)	NBD + SBP	Yes
		OF		TMD	Yes
	7CAD	IF	-	TMD	Yes
ModBC (molybdate)	2ONK	OF with SBP	Molybdate	NBD + SBP	Yes
		OF	-	TMD	-
	3D31	IF	-	TMD	-
MetNI (methionine)	6CVL	OF with SBP	ATP(2)	All	Yes
	3TUI	IF	ADP(2)	TMD	No

Type I proteins are all importers; their substrate is written in parentheses in the transporter column. Type I transporters like MalFGK are driven by TMDs in IF conformation and by the NBDs and the substrate binding protein in OF conformation. When the SBP is removed (but the rest of the protein at OF conformation) TMDs become the driving domain. For the tested type I proteins, regardless of conformation, functional sites appeared to be highly collective and entropy source residues. The results for type I transporters are summarized in Table 3.7.

For Type II transporters like BtuCD, NBDs are driving for both IF and OF conformations, independent of SBP presence. None of the crystal structures of Type II proteins did not feature the substrates, instead some of the proteins were presented in ATP-bound conformations. The known ATP binding sites matched the entropy source peaks, regardless of conformation. However, it should be noted that only BtuCD (PDB ID: 4R9U) is in OF conformation. The results for type II transporters are summarized in Table 3.8.

Table 3.8. Dynamic characterization results for Type II transporters.

Transporter	**PDB**	**Conformation**	**Ligands**	**Driving Domain**	**Peak Match**
BtuCD (B12)	4R9U	OF	ANP(2)	NBD	Yes
	2QI9	IF with SBP	-	NBD	Yes
		IF (apo)		All	Yes
	4FI3	IF with SBP	ANP(2)	NBD	Yes
		IF (apo)		NBD	Yes
BhuUV (heme)	5B57	IF	-	NBD	-
	5B58	IF with SBP	-	NBD	-
PsaBC (manganese)	7KYP	IF (apo)	-	NBD	-

For Type IV transporters, for most of the structures in this dataset, TMDs are driving for IF conformation and NBDs are driving for OF conformations. However, there are

exceptions to this pattern. Ligand binding residues appeared to be entropy sources in the outward facing conformation and close to the entropy source residues in the inward facing conformation. This was a common pattern observed for all type IV proteins. The variety in the results of type IV transporters could be due to the differences in substrate variety or difference in transport direction. The results for type IV transporters are summarized in Table 3.9.

Table 3.9. Dynamic characterization results for Type IV transporters.

Transporter	**PDB**	**Conf.**	**Ligands**	**Driving Domain**	**Peak Match**
TM237/288 (export)	4Q4A	IF	ANP(1)	TMD	Partial
	6QUZ	OF	AGS(2)	NBD	Yes
TmrAB (export)	6RAF	IF (narrow)	ATP, ADP	TMD	Partial
	6RAN	IF (wide)	ATP, ADP	NBD	Partial
	6RAJ	OF	AOV, ATP	TMD	Yes
IrtAB (import)	6TEJ	IF (apo)	-	TMD	-
McjD (export)	4PL0	OF	Peptide, ANP(2)	All	Yes
YbtPQ (import)	6P6J	IF	Substrate	TMD	No
Sav1866 (export)	2ONJ	OF	ANP(2)	NBD	Yes
CFTR (ion channel)	5UAK	IF	-	TMD	No
	6MSM	OF	ATP(2)	All	Yes

Different mode combinations reveal different interactions within the same structure. For this preliminary work, the results for 1-10 GNM modes are reported. However detailed results for TmrAB and CFTR had shown that latent information flow patterns

are revealed with conformation changes and substrate binding. This also holds true with the binding of SBPs. The information flow patterns when SBPs are bound can be observed in SBP-unbound conformations in other mode combinations (2-10 or 3-10). Other mode combinations can also reveal differences between proteins of the same type. While a definitive classification is not presented in this work, this preliminary work shows promise.

4. CONCLUSION AND RECOMMENDATIONS

4.1. Conclusions

ABC transporters are responsible for the translocation of diverse substrates that are essential to key cellular functions [25]. Understanding the function of ABC transporters requires the understanding of the dynamics, and the allosteric network. In this work, three case studies were investigated and a novel dynamic classification is proposed.

Heterodimeric multidrug exporter TmrAB is one of the best characterized ABC transporters. Hofmann et al., obtained nine crystal structures in different conformations during the translocation cycle [12]. The transitional pathways of TmrAB explored by many successful ANM-LD simulations between four different conformations (Asym, IF1, IF2, OF) were able to relieve the key modes enabling the transitions. The correlated fluctuations provided the basis of dynamic connectivity among the functional sites such as the ATP binding sites, substrate binding residues, and ICLs. TM6 of both TMDs appeared as a key helix controlling the cytoplasmic entrance of the translocation cavity. Furthermore, the transfer entropy calculations revealed that information transfer patterns in correlated sites are conserved for all conformations, but their order differs, i.e. the conformations prime the distinct patterns of dynamic correlations while having the information for the others in store as latent. TMDs are driving the movements in Asym conformation, NBDs are driving in the IF conformations, and all domains are information sources in OF conformation.

Cystic Fibrosis transmembrane conductance regulator (CFTR) is directly responsible for the life threatening disease cystic fibrosis [29,30]. ANM-LD simulations between the dephosphorylated ATP-free conformation (PDB ID: 5UAK) and the phosphorylated, ATP-bound conformation (PDB ID: 6MSM) revealed the key modes for transition. The movement of the NBDs essential for transition, which was mainly encoded in ANM modes 1, 3, and 7. The leading and lagging/following regions by the time delay

cross correlations suggested a dynamic cycle, plausible signature dynamics of this specific transition. The movement of NBD1 is followed by the movement of the gating area, which is followed by NBD2, followed by the movement of ICL2 and ECL1. The causal relationship between residues by the TE calculations disclosed the information flow patterns for different mode combinations. NBDs are mainly the driving domains for the dephosphorylated conformation, which also displayed information flow from the periplasmic gating area and the regulatory domain binding area as latent causal interrelation. ATP binding sites and periplasmic gating residues are driving for the phosphorylated conformation. When transfer entropy was calculated using the key GNM modes determined from ANM-LD simulations, the information flow pattern was remarkably similar to the leading/lagging segments calculated from the time delay cross correlations. This shows that not only are NBDs leading the gating residues, but they are causing the movements.

The mechanism of two CFTR drugs; ivacaftor (potentiator drug), and lumacaftor (corrector drug) are investigated. Ivacaftor and lumacaftor drug binding cavities can be identified by the TE calculations as entropy sources for 1-10 and 5-10 modes, respectively. The ivacaftor has a similar global effect on the structure as the combination of ATP binding and ion permeation residues. Whereas, lumacaftor only affects TM helices it is known to stabilize; TM 1, 2, 3, and 6 [36].

The transfer entropy analysis of PsaBC from the gram-positive bacterial pathogen *Streptococcus pneumoniae* revealed the information pathways in the inward facing conformation [14]. Testable predictions were made regarding residues responsible substrate translocation and an allosteric pocket on the cytoplasmic side of TMDs. Four possible gating residues G97, V202, S250 and V253 are identified along with a possible allosteric pocket on transmembrane helix 8. The comparison of MntBC-A models revealed the homology model to have the closest dynamic profile to the crystal structure and to be the most dependable model. Finally, manganese transporter was compared with other Type II ABC transporters and the relationship between TMDs and substrate binding protein was deciphered.

A novel classification based on the dynamics of ABC transporters is presented. This classification works as an addition to the one presented by Thomas et al. based on the TMD folds [22]. For this classification most common types of ABC transporters (I, II and IV) were characterized using the transfer entropy method. The effects of ligand binding and SBP binding were observed on the dynamics. It was observed that Type I transporters, such as MalFGK are driven by TMDs in IF conformation and by the NBDs and the substrate binding protein in OF conformation. When the SBP is removed (but the rest of the protein at OF conformation) TMDs become the driving domain. Type II transporters like BtuCD, NBDs are driving for both IF and OF conformations, independent of SBP presence. For Type IV transporters, TMDs are driving for IF conformation and NBDs are driving for OF conformations. However, the variety in substrate size and charge and the transport direction makes type IV more variable.

4.2. Recommendations

The ANM-LD results for heterodimeric multidrug exporter TmrAB have been presented in this thesis. The equal time cross correlation relations have been calculated and analyzed. In further research, the time delay cross correlations can be calculated using custom window sizes, and time delay amounts. This would reveal specific events during the translocation cycle of TmrAB. By performing many ANM-LD simulations between different TmrAB conformations, the key modes for transition have been identified. As it was done for CFTR, transfer entropy analysis using the most selected modes can be performed to reveal causal relationships essential for transition. Comparing these causal relationships and the order-of-events from the time delay calculations can illuminate the translocation cycle.

In this work, a preliminary classification based on the dynamic characteristics of ABC transporters have been presented. For this task, transfer entropy and collectivity analysis has been used. The primary classification metrics were the identification of the driving domain between conformations and if the known key sites were the most collective. Other classification parameters can be included regarding the locations of the

entropy source clusters. This would allow easier identification of key sites and allosteric pockets. For this work, only 1-10 mode subset was used. Using other mode subsets would reveal latent connections between residues. Finally, the classification will be expanded to include all ABC transporters with available structural information.

REFERENCES

1. Haliloglu, T. and I. Bahar, "Adaptability of Protein Structures to Enable Functional Interactions and Evolutionary Implications", *Current Opinion in Structural Biology*, Vol. 35, pp. 17–23, 2015.

2. Lewinson, O., C. Orelle and M. Seeger, "Structures of ABC Transporters", *FEBS Letters*, Vol. 594, pp. 3799–3814, 2020.

3. Stiller, J., R. Otten, D. Häussinger, P. Rieder, D. Theobald and D. Kern, "Structure Determination of High-Energy States in a Dynamic Protein Ensemble", *Nature*, Vol. 603, pp. 1–8, 2022.

4. Jiao, W. and E. J. Parker, "Using a Combination of Computational and Experimental Techniques to Understand the Molecular Basis for Protein Allostery", *Advances in Protein Chemistry and Structural Biology*, Vol. 87, pp. 391–413, 2012.

5. Sumbul, F., S. Acuner-Ozbabacan and T. Haliloglu, "Allosteric Dynamic Control of Binding", *Biophysical Journal*, Vol. 109, No. 6, pp. 1190–1201, 2015.

6. Ma, B., C.-J. Tsai, T. Haliloğlu and R. Nussinov, "Dynamic Allostery: Linkers Are Not Merely Flexible", *Structure*, Vol. 19, No. 7, pp. 907–917, 2011.

7. Bernardi, R. C., M. C. Melo and K. Schulten, "Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems", *Biochimica et Biophysica Acta*, Vol. 1850, No. 5, pp. 872–877, 2015.

8. Yang, M., N. Livnat Levanon, B. Acar, B. Aykaç Fas, G. Masrati, J. Rose, N. Ben-Tal, T. Haliloglu, Y. Zhao and O. Lewinson, "Single-Molecule Probing of the Conformational Homogeneity of the ABC Transporter BtuCD", *Nature Chemical Biology*, Vol. 14, p. 715–722, 2018.

9. Acar, B., J. Rose, B. Aykac Fas, N. Ben-Tal, O. Lewinson and T. Haliloglu, "Distinct Allosteric Networks Underlie Mechanistic Speciation of ABC Transporters", *Structure*, Vol. 28, No. 6, pp. 651–663.e5, 2020.

10. Hacisuleyman, A. and B. Erman, "Causality, Transfer Entropy and Allosteric Communication Landscapes in Proteins with Harmonic Interactions", *Proteins: Structure, Function, and Bioinformatics*, Vol. 85, pp. 1056–1064, 2017.

11. Vasiliou, V., K. Vasiliou and D. Nebert, "Human ATP-Binding Cassette (ABC) Transporter Family", *Human Genomics*, Vol. 3, pp. 281–90, 2009.

12. Hofmann, S., D. Januliene, A. R. Mehdipour, C. Thomas, E. Stefan, S. Bruchert, B. T. Kuhn, E. R. Geertsma, G. Hummer, R. Tampe and A. Moeller, "Conformation Space of a Heterodimeric ABC Exporter Under Turnover Conditions", *Nature*, Vol. 571, pp. 580–583, 2019.

13. Gat, O., I. Mendelson, T. Chitlaru, N. Ariel, Z. Altboum, H. Levy, S. Weiss, H. Grosfeld, S. Cohen and A. Shafferman, "The Solute-Binding Component of a Putative Mn(II) ABC Transporter (MntA) is a Novel Bacillus Anthracis Virulence Determinant", *Molecular Microbiology*, Vol. 58, No. 2, pp. 533–551, 2005.

14. Neville, S. L., J. Sjöhamn, J. A. Watts, H. MacDermott-Opeskin, S. J. Fairweather, K. Ganio, A. C. Hulyer, A. P. McGrath, A. J. Hayes, T. R. Malcolm, M. R. Davies, N. Nomura, S. Iwata, M. L. O'Mara, M. J. Maher and C. A. McDevitt, "The Structural Basis of Bacterial Manganese Import", *Science Advances*, Vol. 7, No. 32, p. eabg3980, 2021.

15. Altintel, B., B. Acar, B. Erman and T. Haliloglu, "Subsets of Slow Dynamic Modes Reveal Global Information Sources as Allosteric Sites", *Journal of Molecular Biology*, p. 167644, 2022.

16. Dean, M., Y. Hamon and G. Chimini, "The Human ATP-Binding Cassette (ABC) Transporter Superfamily", *Journal of Lipid Research*, Vol. 42, No. 7, pp. 1007–

1017, 2001.

17. Csanády, L., P. Vergani and D. C. Gadsby, "Structure, Gating, and Regulation of the CFTR Anion Channel", *Physiological Reviews*, Vol. 99, No. 1, pp. 707–738, 2019.

18. Göddeke, H. and L. V. Schäfer, "Capturing Substrate Translocation in an ABC Exporter at the Atomic Level", *Journal of the American Chemical Society*, Vol. 142, No. 29, pp. 12791–12801, 2020.

19. Locher, K. P., "Mechanistic Diversity in ATP-Binding Cassette (ABC) Transporters", *Nature Structural and Molecular Biology*, Vol. 23, pp. 487–493, 2016.

20. Stefan, E., S. Hofmann and R. Tampe, "A Single Power Stroke by ATP Binding Drives Substrate Translocation in a Heterodimeric ABC Transporter", *eLife*, Vol. 9, p. e55943, 2020.

21. Hollenstein, K., R. Dawson and K. Locher, "Structure and Mechanism of ABC Transporter Proteins", *Current Opinion in Structural Biology*, Vol. 17, pp. 412–418, 2007.

22. Thomas, C., S. G. Aller, K. Beis, E. P. Carpenter, G. Chang, L. Chen, E. Dassa, M. Dean, F. Duong Van Hoa, D. Ekiert, R. Ford, R. Gaudet, X. Gong, I. B. Holland, Y. Huang, D. K. Kahne, H. Kato, V. Koronakis, C. M. Koth, Y. Lee, O. Lewinson, R. Lill, E. Martinoia, S. Murakami, H. W. Pinkett, B. Poolman, D. Rosenbaum, B. Sarkadi, L. Schmitt, E. Schneider, Y. Shi, S.-L. Shyng, D. J. Slotboom, E. Tajkhorshid, D. P. Tieleman, K. Ueda, A. Váradi, P.-C. Wen, N. Yan, P. Zhang, H. Zheng, J. Zimmer and R. Tampé, "Structural and Functional Diversity Calls for a New Classification of ABC Transporters", *FEBS Letters*, Vol. 594, No. 23, pp. 3767–3775, 2020.

23. Ford, R. and K. Beis, "Learning the ABCs One at a Time: Structure and Mechanism of ABC Transporters", *Biochemical Society Transactions*, Vol. 47, pp. 23–36,

2019.

24. Locher, K. P., A. T. Lee and D. C. Rees, "The ¡i¿E. coli¡/i¿ BtuCD Structure: A Framework for ABC Transporter Architecture and Mechanism", *Science*, Vol. 296, No. 5570, pp. 1091–1098, 2002.

25. Thomas, C. and R. Tampe, "Structural and Mechanistic Principles of ABC Transporters", *Annual Review of Biochemistry*, Vol. 89, pp. 605–636, 2020.

26. Biemans-Oldehinkel, E., M. Doeven and B. Poolman, "ABC Transporter Architecture and Regulatory Roles of Accessory Domains", *FEBS Letters*, Vol. 580, pp. 1023–35, 2006.

27. Stockner, T., R. Gradisch and L. Schmitt, "The Role of the Degenerate Nucleotide Binding Site in Type I ABC Exporters", *FEBS Letters*, Vol. 594, No. 23, pp. 3815–3838, 2020.

28. Noll, A., C. Thomas, V. Herbring, T. Zollmann, K. Barth, A. R. Mehdipour, T. M. Tomasiak, S. Brüchert, B. Joseph, R. Abele, V. Olieric, M. Wang, K. Diederichs, G. Hummer, R. M. Stroud, K. M. Pos and R. Tampé, "Crystal Structure and Mechanistic Basis of a Functional Homolog of the Antigen Transporter TAP", *Proceedings of the National Academy of Sciences*, Vol. 114, No. 4, pp. E438–E447, 2017.

29. Zhang, Z., F. Liu and J. Chen, "Molecular Structure of the ATP-Bound, Phosphorylated Human CFTR", *Proceedings of the National Academy of Sciences*, Vol. 115, No. 50, pp. 12757–12762, 2018.

30. Liu, F., Z. Zhang, L. Csanády, D. C. Gadsby and J. Chen, "Molecular Structure of the Human CFTR Ion Channel", *Cell*, Vol. 169, No. 1, pp. 85–95.e8, 2017.

31. Zhang, Z. and J. Chen, "Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator", *Cell*, Vol. 167, pp. 483–491, 2016.

32. Zhang, Z., F. Liu and J. Chen, "Conformational Changes of CFTR upon Phosphorylation and ATP Binding", *Cell*, Vol. 170, pp. 483–491, 2017.

33. Hwang, T.-C., J.-T. Yeh, J. Zhang, Y.-C. Yu, H.-I. Yeh and S. Destefano, "Structural Mechanisms of CFTR Function and Dysfunction", *Journal of General Physiology*, Vol. 150, No. 4, pp. 539–570, 2018.

34. Liu, F., Z. Zhang, A. Levit, J. Levring, K. K. Touhara, B. K. Shoichet and J. Chen, "Structural Identification of a Hotspot on CFTR for Potentiation", *Science*, Vol. 364, No. 6446, pp. 1184–1188, 2019.

35. Lopes-Pacheco, M., "CFTR Modulators: Shedding Light on Precision Medicine for Cystic Fibrosis", *Frontiers in Pharmacology*, Vol. 7, p. 275, 2016.

36. Fiedorczuk, K. and J. Chen, "Mechanism of CFTR Correction by Type I Folding Correctors", *Cell*, Vol. 185, No. 1, pp. 158–168.e11, 2022.

37. Baatallah, N., A. Elbahnsi, J.-P. Mornon, B. Chevalier, P. Iwona, N. Servel, R. Zelli, J.-L. Décout, A. Edelman, I. Sermet-Gaudelus, I. Callebaut and A. Hinzpeter, "Pharmacological Chaperones Improve Intra-Domain Stability and Inter-Domain Assembly via Distinct Binding Sites to Rescue Misfolded CFTR", *Cellular and Molecular Life Sciences*, Vol. 78, pp. 7813–7829, 2021.

38. Kleizen, B., J. F. Hunt, I. Callebaut, T.-C. Hwang, I. Sermet-Gaudelus, S. Hafkemeyer and D. N. Sheppard, "CFTR: New Insights Into Structure and Function and Implications for Modulation by Small Molecules", *Journal of Cystic Fibrosis*, Vol. 19, pp. S19–S24, 2020.

39. Kuznetsova, A., G. Masrati, E. Vigonsky, N. Livnat-Levanon, J. Rose, M. Grupper, A. Baloum, J. G. Yang, D. C. Rees, N. Ben-Tal and O. Lewinson, "Titratable Transmembrane Residues and a Hydrophobic Plug Are Essential for Manganese Import via the Bacillus Anthracis ABC Transporter MntBC-A", *Journal of Biological Chemistry*, Vol. 297, No. 4, p. 101087, 2021.

40. Naoe, Y., N. Nakamura, A. Doi, M. Sawabe, H. Nakamura, Y. Shiro and H. Sugimoto, "Crystal Structure of Bacterial Haem Importer Complex in the Inward-Facing Conformation", *Nature Communications*, Vol. 7, 2016.

41. Söding, J., A. Biegert and A. N. Lupas, "The HHPred Interactive Server for Protein Homology Detection and Structure Prediction", *Nucleic Acids Research*, Vol. 33, pp. 244–248, 2005.

42. Jumper, J. M., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. A. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, "Highly Accurate Protein Structure Prediction with AlphaFold", *Nature*, Vol. 596, pp. 583 – 589, 2021.

43. Bock, C., T. Zollmann, K.-A. Lindt, R. Tampé and R. Abele, "Peptide Translocation by the Lysosomal ABC Transporter TAPL Is Regulated by Coupling Efficiency and Activation Energy", *Scientific Reports*, Vol. 9, p. 11884, 2019.

44. Hollenstein, K., D. C. Frei and K. P. Locher, "Structure of an ABC Transporter in Complex with Its Binding Protein", *Nature*, Vol. 446, pp. 213–216, 2007.

45. Oldham, M. L. and J. Chen, "Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State", *Science*, Vol. 332, No. 6034, pp. 1202–1205, 2011.

46. Khare, D., M. L. Oldham, C. Orelle, A. L. Davidson and J. Chen, "Alternating Access in Maltose Transporter Mediated by Rigid-Body Rotations", *Molecular Cell*, Vol. 33-4, pp. 528–36, 2009.

47. Maruyama, Y., T. Itoh, A. Kaneko, Y. Nishitani, B. Mikami, W. Hashimoto and K. Murata, "Structure of a Bacterial ABC Transporter Involved in the Import of

an Acidic Polysaccharide Alginate", *Structure*, Vol. 23, No. 9, pp. 1643–1654, 2015.

48. Liu, F., J. Liang, B. Zhang, Y. Gao, X. Yang, T. Hu, H. Yang, W. Xu, L. W. Guddat and Z. Rao, "Structural Basis of Trehalose Recycling by the ABC Transporter LpqY-SugABC", *Science Advances*, Vol. 6, No. 44, p. eabb9833, 2020.

49. Gerber, S., M. Comellas-Bigler, B. A. Goetz and K. P. Locher, "Structural Basis of Trans-Inhibition in a Molybdate/Tungstate ABC Transporter", *Science*, Vol. 321, No. 5886, pp. 246–250, 2008.

50. Nguyen, P. T., J. Y. Lai, A. T. Lee, J. T. Kaiser and D. C. Rees, "Noncanonical Role for the Binding Protein in Substrate Uptake by the MetNI Methionine ATP Binding Cassette (ABC) Transporter", *Proceedings of the National Academy of Sciences*, Vol. 115, No. 45, pp. E10596–E10604, 2018.

51. Johnson, E., P. T. Nguyen, T. O. Yeates and D. C. Rees, "Inward Facing Conformations of the MetNI Methionine ABC Transporter: Implications for the Mechanism of Transinhibition", *Protein Science*, Vol. 21, No. 1, pp. 84–96, 2012.

52. Hvorup, R. N., B. A. Goetz, M. Niederer, K. Hollenstein, E. Perozo and K. P. Locher, "Asymmetry in the Structure of the ABC Transporter-Binding Protein Complex BtuCD-BtuF", *Science*, Vol. 317, No. 5843, pp. 1387–1390, 2007.

53. Korkhov, V. M., S. A. Mireku and K. P. Locher, "Structure of AMP-PNP-Bound Vitamin B12 Transporter BtuCD–F", *Nature*, Vol. 490, pp. 367–372, 2012.

54. Hohl, M., L. M. Hürlimann, S. Böhm, J. Schöppe, M. G. Grütter, E. Bordignon and M. A. Seeger, "Structural Basis for Allosteric Cross-Talk Between the Asymmetric Nucleotide Binding Sites of a Heterodimeric ABC Exporter", *Proceedings of the National Academy of Sciences*, Vol. 111, No. 30, pp. 11025–11030, 2014.

55. Hutter, C., M. Timachi, L. Hürlimann, I. Zimmermann, P. Egloff, H. Göddeke, S. Kucher, S. Stefanic, M. Karttunen, L. Schäfer, E. Bordignon and M. Seeger,

"The Extracellular Gate Shapes the Energy Profile of an ABC Exporter", *Nature Communications*, Vol. 10, p. 2260, 2019.

56. Arnold, F., M. Weber, I. Gonda, M. Gallenito, S. Adenau, P. Egloff, I. Zimmermann, C. Hutter, L. Hürlimann, E. Peters, J. Piel, G. Meloni, O. Medalia and M. Seeger, "The ABC Exporter IrtAB Imports and Reduces Mycobacterial Siderophores", *Nature*, Vol. 580, pp. 413–417, 2020.

57. Choudhury, H. G., Z. Tong, I. Mathavan, Y. Li, S. Iwata, S. Zirah, S. Rebuffat, H. W. van Veen and K. Beis, "Structure of an Antibacterial Peptide ATP-Binding Cassette Transporter in a Novel Outward Occluded State", *Proceedings of the National Academy of Sciences*, Vol. 111, No. 25, pp. 9145–9150, 2014.

58. Wang, Z., W. Hu and H. Zheng, "Pathogenic Siderophore ABC Importer YbtPQ Adopts a Surprising Fold of Exporter", *Science Advances*, Vol. 6, No. 6, p. 7997, 2020.

59. Dawson, R. J. and K. P. Locher, "Structure of the Multidrug ABC Transporter Sav1866 from Staphylococcus Aureus in Complex with AMP-PNP", *FEBS Letters*, Vol. 581, No. 5, pp. 935–938, 2007.

60. Bahar, I., A. R. Atilgan and B. Erman, "Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential", *Folding and Design*, Vol. 2, No. 3, pp. 173–181, 1997.

61. Haliloglu, T., I. Bahar and B. Erman, "Gaussian Dynamics of Folded Proteins", *Physical Review Letters*, Vol. 79, pp. 3090–3093, 1997.

62. Doruker, P., A. Atilgan and I. Bahar, "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to Amylase Inhibitor", *Proteins: Structure, Function, and Bioinformatics*, Vol. 40, pp. 512 – 524, 2000.

63. Atilgan, A., S. Durell, R. Jernigan, M. Demirel, O. Keskin and I. Bahar, "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model", *Biophysical Journal*, Vol. 80, No. 1, pp. 505–515, 2001.

64. Brüschweiler, R., "Collective Protein Dynamics and Nuclear Spin Relaxation", *The Journal of Chemical Physics*, Vol. 102, No. 8, pp. 3396–3403, 1995.

65. Infield, D., G. Cui, C. Kuang and N. McCarty, "The Positioning of Extracellular Loop 1 Affects Pore Gating of the Cystic Fibrosis Transmembrane Conductance Regulator", *American Journal of Physiology - Lung Cellular and Molecular Physiology*, Vol. 310, p. L403–L414, 2015.

66. He, L., A. A. Aleksandrov, A. W. Serohijos, T. Hegedüs, L. A. Aleksandrov, L. Cui, N. V. Dokholyan and J. R. Riordan, "Multiple Membrane-Cytoplasmic Domain Contacts in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Mediate Regulation of Channel Gating", *Journal of Biological Chemistry*, Vol. 283, No. 39, pp. 26383–26390, 2008.

67. Patrick, A. E. and P. J. Thomas, "Development of CFTR Structure", *Frontiers in Pharmacology*, Vol. 3, 2012.

68. Millan, C. R., M. Francis, N. K. Khandelwal, V. F. Thompson, T. M. Thaker and T. M. Tomasiak, "A Conserved Motif in Intracellular Loop 1 Stabilizes the Outward-Facing Conformation of TmrAB", *Journal of Molecular Biology*, Vol. 433, No. 16, p. 166834, 2021.

69. Wang, W., J. Wu, K. Bernard, G. Li, G. Wang, M. O. Bevensee and K. L. Kirk, "ATP-Independent CFTR Channel Gating and Allosteric Modulation by Phosphorylation", *Proceedings of the National Academy of Sciences*, Vol. 107, No. 8, pp. 3888–3893, 2010.