MOLECULAR MODELING AND SIMULATIONS OF THE CONFORMATIONAL CHANGES UNDERLYING CHANNEL ACTIVITY IN CFTR

A Dissertation
Presented to
The Academic Faculty

By

Kazi Shefaet Rahman

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Bioengineering in the
Department of Biomedical Engineering

Georgia Institute of Technology

December 2013

MOLECULAR MODELING AND SIMULATIONS OF THE CONFORMATIONAL CHANGES UNDERLYING CHANNEL ACTIVITY IN CFTR

Approved by:

Dr. Stephen C. Harvey, Advisor
School of Biology
Georgia Institute of Technology

Dr. Nael A. McCarty
Department of Pediatrics
Emory University School of Medicine & Children's Healthcare of Atlanta

Dr. King Jordan
School of Biology
Georgia Institute of Technology

Dr. Cheng Zhu
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Peter J. Ludovice
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date Approved: November 5, 2013

Dedicated to my parents.

ACKNOWLEDGEMENTS

I would like to thank my advisors Dr. Harvey and Dr. McCarty for their mentorship. Their insight and wisdom have guided my growth not only as a scientist but also as a person, and I will forever remain indebted to them for taking me under their wing. I am also grateful to my committee members, Drs. King Jordan, Cheng Zhu and Peter Ludovice for their kind advice and feedback. I would like to thank all the members of the Harvey lab — Jared, Bee, Burak, Ethan, Piyush, Anton and Shreyas — and the McCarty lab — Guiying, Brandon, and Danny — with whom I had the honor and pleasure of working. Their support, criticism, and late-night discussions in the office made my time at Tech immensely enjoyable. Finally, I cannot thank enough my parents, my sister and her family, and my fiancée (soon to be wife!) for their unfaltering love and support. I hope I've made them proud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ABC	ATP Binding Cassete
ABS	ATP Binding Site
ADP	Adenosine Diphosphate
AMP-PNP	Adenyl Imidodiphosphate
ATP	Adenosine Triphosphate
C0-CFTR	ATP-free closed channel state of CFTR
C1-CFTR	Partial dimer closed state of CFTR
C2-CFTR	Strained transition state of CFTR
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
CHARMM	Chemistry at Harvard Molecular Mechanics
ECL	Extra-Cellular Loop
GSH	Glutathione
IACUC	Institutional Animal Care and Use Committee
ICL	Intra-Cellular Loop
ISO	Isoproterenol
MD	Molecular Dynamics
MM	Molecular Mechanics
MSA	Multiple Sequence Alignment
MSD	Membrane Spanning Domain
MTS-2-MTS	1,2-ethanediyl bismethanethiosulfonate

NAMD	Not (Just) Another Molecular Dynamics (program)
NBD	Nucleotide Binding Domain
NMR	Nuclear Magnetic Resonance
NPT	Isothermal-Isobaric Ensemble
O-CFTR	Open channel state of CFTR
P-gp	P-glycoprotein
Pi	Inorganic phosphate ion (PO_4^{3-})
POPC	1-Palmitoyl-2-oleoylphosphatidylcholine
R(-domain)	Regulatory (domain)
RD	Relative Divergence
RI	Regulatory Insertion (span of CFTR sequence)
RMSD	Root Mean Square Displacement
TIP3P	Transferable Intermolecular Potential – 3 Point (water model)
TM	Trans-Membrane (Helix)
TMD	Targeted Molecular Dynamics
VMD	Visual Molecular Dynamics (program)

SUMMARY

Mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator protein (CFTR) cause cystic fibrosis (CF), the most common life-shortening genetic disease among Caucasians. Although general features of the structure of CFTR have been predicted from homology models, the conformational changes that result in channel opening and closing have yet to be resolved. We created new closed- and open-state homology models of CFTR, and performed targeted molecular dynamics simulations of the conformational transitions in a channel opening event. The simulations predict a conformational wave that starts at the nucleotide binding domains and ends with the formation of an open conduction pathway. Experimentally confirmed changes in side-chain interactions are observed in all major domains of the protein. We also identified unique-to-CFTR substitutions that may have led to channel activity in CFTR. Molecular modeling and simulations are used to compare the effects of these substitutions against a canonical ABC transporter, and suggest that gain of channel function in CFTR may have risen from loss of ATPase function at its NBDs. The models and simulation add to our understanding of the mechanism of ATP-dependent gating in this disease-relevant ion channel.

CHAPTER 1

INTRODUCTION

ABC Transporters

The ABC transporter superfamily is a large and ancient family of membrane proteins, found in all phyla [1]. Members of the ABC superfamily catalyze the membrane transport, in an ATP-dependent manner, of a wide variety of substrates ranging from small inorganic ions and metabolites to large hydrophobic drugs and polypeptides. At a minimum, all ABC transporters have four domains: two membrane spanning domains (MSDs) consisting of transmembrane alpha helices (TMs) embedded in the lipid bilayer, and two nucleotide binding domains (NBDs) located in the cytoplasm [2]. A large number of studies indicate that the binding and hydrolysis of ATP at the NBDs leads to changes in the conformation of the membrane domains [3]. The binding of ATP leads to dimerization of the NBDs, which is thought to initiate a conformational cascade that is transmitted to the membrane-spanning domains via the intracellular loops (ICLs) that emerge from the cytoplasmic ends of the TMs. This results in the MSDs alternating between outward-facing and inward-facing states. This "ATP switch" concept has been supported by recently published crystal structures of the nucleotide-bound bacterial ABC exporter Sav1866 [4], P-glycoprotein (P-gp) (from M. musculus [5] and C. elegans [6]) and the bacterial exporter TM287/288 [7] in the nucleotide-free inward-facing states, and liganded and unliganded states of the MsbA lipid transporter [8]. Despite their ubiquity, however, the precise mechanisms by which ATP binding and hydrolysis and NBD

dimerization lead to conformation changes in the remaining domains of these ABC transporters are poorly understood.

Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

Cystic fibrosis (CF) — the most common lethal genetic disease among Caucasians — is caused by mutations in the gene that encodes cystic fibrosis transmembrane conductance regulator (CFTR) [9,10], a large transmembrane protein belonging to the ABC transporter superfamily [11,12]. This ATP-gated chloride ion channel plays a central role in ion transport and water movement across epithelia such as those that line the airways and intestinal tracts [13-15]. Loss of CFTR function leads to the impaired mucociliary clearance and consequent chronic airway infection seen in CF patients [16], whereas excessive enterotoxin-mediated CFTR activation can result in secretory diarrhea [17].

Characteristic of its ABC transporter heritage, CFTR has a domain architecture (**Figure 1.1**) comprising two membrane-spanning domains (MSD1 and MSD2), each containing six transmembrane α-helices (enumerated TM1 through TM12); and two cytosolic nucleotide binding domains (NBD1 and NBD2) that have ATPase activity [2]. Unlike any other ABC protein, however, CFTR also has a large regulatory (R) domain, situated between NBD1 and MSD2 in its sequence. The R domain has multiple phosphorylation sites which, along with ATP binding and hydrolysis at the NBDs, play a major role in the regulation of channel activity [18,19].

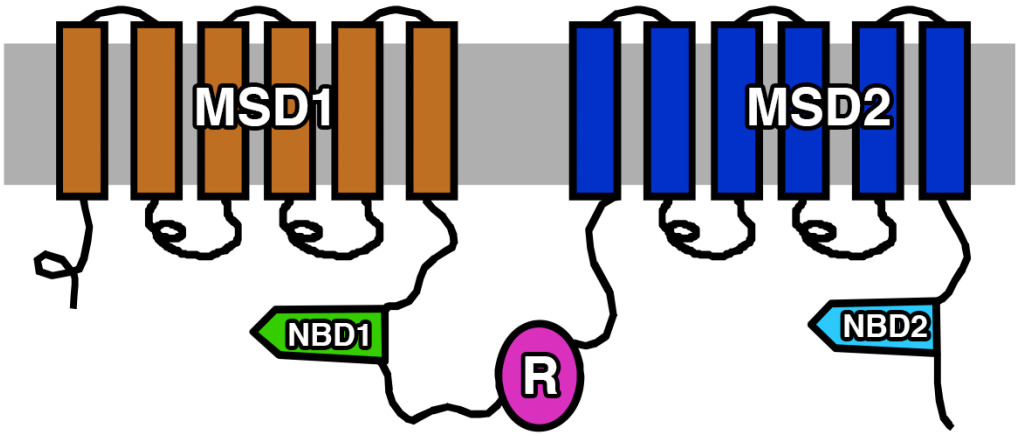

Figure 1.1: Domain topology of CFTR

While over 1,500 disease-causing mutations of the CFTR gene are known (Cystic Fibrosis Mutation Database, http://www.genet.sickkids.on.ca/), over 90% of cystic fibrosis cases are cause by the deletion of a phenylalanine residue at position 508 — ΔF508 [20]. This mutation is thought to hamper CFTR folding and signal transduction by destabilizing the interface between the NBD1 and ICL4 domains, although accurate models of the interactions of this residue in the native state do not exist.

CFTR: Degraded Transporter Turned Ion Channel

Most ABC transporters mediate active membrane transport through an alternating access mechanism, but CFTR is unique in this group as the only member known to function as an ion channel, primarily for chloride ions [14]. The structural basis for this important difference in function, however, is not yet known. Studies have also investigated the role that CFTR plays in transporting glutathione, the most prominent regulator of redox status

in the lung [21]. In the lungs of CF patients, glutathione concentration in the airway surface liquid is drastically reduced, which contributes to oxidative stress. CFTR therefore seems to be a dual-function protein that bears channel activity, but also appears to retain residual transporter activity. In fact, CFTR is thought to have evolved from an alternating access transporter that was somehow degraded to allow the formation of a continuous pore for ion conduction and for NBD-initiated conformation changes to gate this pore [22,23].

In most ABC transporters, ATP binding at the two NBDs results in the formation of a head-to-tail NBD dimer that drives a conformational change in the whole protein which enables membrane transport [3].

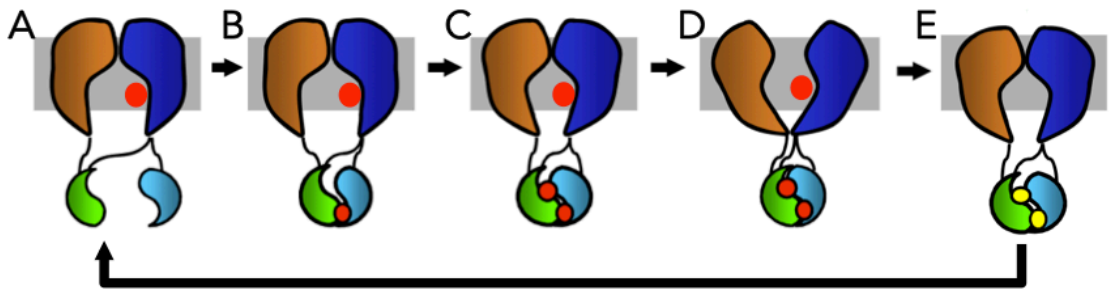

Figure 1.2: ATP-regulated alternating access in atypical ABC transporter

In a typical ABC exporter transport cycle (**Figure 1.2**), for instance, the intracellular substrate to be pumped out first binds to a pocket in the membrane-spanning domain of the protein in its "inward-facing" conformation (A). Then, the binding of cytosolic ATP

to the ATP binding sites (ABSs) results in the formation of a dimer between the NBDs (B, C). This initiates a chain of conformational changes that results in the switching of the protein to an "outward-facing conformation" (D). The substrate is then free to diffuse into the periplasm (this may be aided by a concomitant change in the structure of the binding pocket that leads to a lower affinity interaction between the substrate and the protein in the outward-facing state). In addition to being able to bind ATP molecules, the NBDs of most ABC transporters are also capable of catalyzing the hydrolysis of bound ATP to ADP when in dimer formation [24] (E). The energy released from ATP hydrolysis initiates a second chain of events that eventually leads to dissociation of the two nucleotide binding domains. This conformation change reverses the effect of dimerization, reverting the molecule to the original "inward-facing" state. Release of the hydrolysis by-products (ADP + Pi) allows the cycle to start anew. Unlike other ABC proteins, however, CFTR shows significant disparity in both the sequence and function of the two NBDs. Whereas the site in NBD2 is capable of both ATP binding and hydrolysis, the equivalent site in NBD1 is thought to have poor hydrolytic activity, engaging in stable interactions that retain the nucleotide at the binding site for long periods [25]. Conformational transitions in CFTR are therefore thought to involve at least four distinct states (**Figure 1.3**):

1) **C0**: an apo closed-channel state where both nucleotide-binding sites are empty and the NBDs are completely dissociated;
2) **C1**: a closed-channel state where the NBD1 site is occupied and the NBDs are partially dimerized;

3) **C2**: a strained transition state where the NBDs are fully dimerized but the channel is still closed to conduction; and

4) **O**: an open-channel state with tightly dimerized NBDs in which both of the binding sites contain ATP.

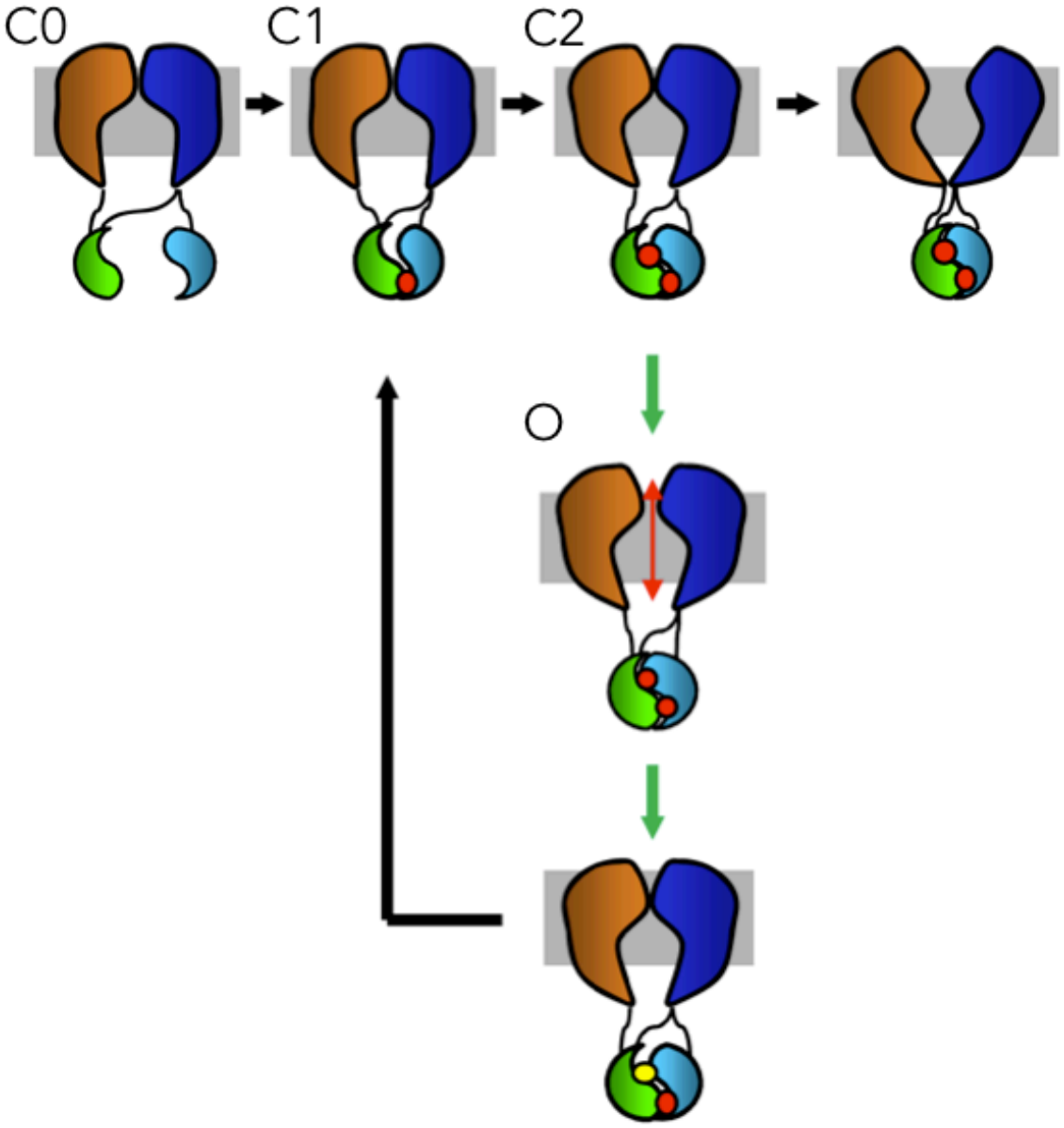

Figure 1.3: States in CFTR's gating cycle

While high-resolution X-ray crystal structures of the nucleotide binding domains exist [26-28], structural information on the MSDs and the unique-to-CFTR R domain has been scarce, with only low-resolution NMR and cryo-EM electron density maps currently available [29,30]. However, since ATP binding and tight NBD dimerization are necessary for channel opening [31], it has been hypothesized that CFTR evolved from a degraded ABC exporter whose ATP-bound outward-facing state corresponds to the open CFTR channel [23]. Due to the low ATP turnover rate at NBD1, the dominant closed channel conformation is thought to be the partial-dimer C1 state, although the fully dissociated C0 state is also likely to be closed to ion conduction.

Translational Impact

There are nearly 50 known ABC transporters in humans and 13 genetic diseases associated with defects in them [32]. Furthermore, several ABC transporters capable of transporting large amphipathic compounds are thought to contribute to the resistance of cancer cells to chemotherapy. Understanding the structural basis for the pump cycle of these transporters can aid in the design of drugs that can correct defects in mutants, or of chemotherapeutic agents that can circumvent or suppress ABC transporter-related multi-drug resistance.

In addition, understanding the dynamic structure of the pore in CFTR in particular is important for several reasons. First, this channel is clinically relevant to the pathogenesis of CF, the most common lethal autosomal recessive disease among Caucasians, affecting more than 30,000 Americans and many more worldwide. The development of

pharmacological modulators for CF requires a better understanding of the function of the normal channel. Second, CFTR is also involved in two other diseases of note: polycystic kidney disease (the fourth leading cause of end-stage renal disease in America) and secretory diarrhea including cholera (a worldwide health problem). Understanding the dynamic structure of the pore may lead to the design of clinically useful inhibitors to treat the rampant secretion associated with these conditions. Understanding the nature of the stable open state will aid in the rational design of drugs that can lock mutant CFTR channels open in the plasma membranes of CF airway epithelial cells, leading to increased chloride secretion and amelioration of disease. residues that may be crucial to the maintenance of the open pore and for ion conduction. Finally, because computational models of CFTR can be validated experimentally with very high resolution using electrophysiology techniques that are unavailable for the study of other ABC transporters, we may be able to extrapolate the observations in CFTR to understanding how the substrate binding pockets of clinically-relevant ABC transporters change structure during the pump cycle.

References

1. Higgins CF (1995) The ABC of channel regulation. Cell 82: 693–696.
2. Biemans-Oldehinkel E, Doeven MK, Poolman B (2006) ABC transporter architecture and regulatory roles of accessory domains. FEBS Lett 580: 1023–1035. doi:10.1016/j.febslet.2005.11.079.
3. Higgins CF, Linton KJ (2004) The ATP switch model for ABC transporters. Nat Struct Mol Biol 11: 918–926. doi:10.1038/nsmb836.
4. Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180–185. doi:10.1038/nature05155.

5. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722. doi:10.1126/science.1168750.

6. Jin MS, Oldham ML, Zhang Q, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. Nature 490: 566–569. doi:10.1038/nature11448.

7. Hohl M, Briand C, Grütter MG, Seeger MA (2012) Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 19: 395–402. doi:10.1038/nsmb.2267.

8. Ward A, Reyes CL, Yu J, Roth CB, Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci U S A 104: 19005–19010. doi:10.1073/pnas.0709388104.

9. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066–1073.

10. Rommens J, Iannuzzi M, Kerem B, Drumm M, Melmer G, et al. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245: 1059–1065. doi:10.1126/science.2772657.

11. Dean M, Hamon Y, Chimini G (2001) The human ATP-binding cassette (ABC) transporter superfamily. J Lipid Res 42: 1007–1017.

12. Gadsby DC, Vergani P, Csanády L (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature 440: 477–483. doi:10.1038/nature04712.

13. Quinton PM (1990) Cystic fibrosis: a disease in electrolyte transport. FASEB J 4: 2709–2717.

14. Anderson M, Gregory R, Thompson S, Souza D, Paul S, et al. (1991) Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. Science 253: 202–205. doi:10.1126/science.1712984.

15. Bear CE, Li C, Kartner N, Bridges RJ, Jensen TJ, et al. (1992) Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). Cell 68: 809–818. doi:10.1016/0092-8674(92)90155-6.

16. Pilewski J, Frizzell R (1999) Role of CFTR in airway disease. Physiological reviews 79: S215–S255.

17. Field M, Semrad C (1993) Toxigenic diarrheas, congenital diarrheas, and cystic fibrosis: disorders of intestinal ion transport. Annu Rev Physiol 55: 631–655.

18. Cheng SH, Rich DP, Marshall J, Gregory RJ, Welsh MJ, et al. (1991) Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. Cell 66: 1027–1036. doi:10.1016/0092-8674(91)90446-6.

19. Welsh MJ, Winter MC (1997) Stimulation of CFTR activity by its phosphorylated R domain. Nature 389: 294–296. doi:10.1038/38514.

20. Cheng SH, Gregory RJ, Marshall J, Paul S, Souza DW, et al. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell 63: 827–834. doi:10.1016/0092-8674(90)90148-8.

21. Gao L, Kim KJ, Yankaskas JR, Forman HJ (1999) Abnormal glutathione transport in cystic fibrosis airway epithelia. Am J Physiol 277: L113–L118.

22. Jordan IK, Kota KC, Cui G, Thompson CH, McCarty NA (2008) Evolutionary and functional divergence between the cystic fibrosis transmembrane conductance regulator and related ATP-binding cassette transporters. Proc Natl Acad Sci U S A 105: 18865–18870. doi:10.1073/pnas.0806306105.

23. Bai Y, Li M, Hwang T-C (2011) Structural basis for the channel function of a degraded ABC transporter, CFTR (ABCC7). J Gen Physiol 138: 495–507. doi:10.1085/jgp.201110705.

24. Moody JE, Millen L, Binns D, Hunt JF, Thomas PJ (2002) Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. J Biol Chem 277: 21111–21114. doi:10.1074/jbc.C200228200.

25. Basso C, Vergani P, Nairn AC, Gadsby DC (2003) Prolonged nonhydrolytic interaction of nucleotide with CFTR's NH2-terminal nucleotide binding domain and its role in channel gating. J Gen Physiol 122: 333–348. doi:10.1085/jgp.200308798.

26. Lewis HA, Buchanan SG, Burley SK, Conners K, Dickey M, et al. (2003) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23: 282–293. doi:10.1038/sj.emboj.7600040.

27. Lewis HA, Zhao X, Wang C, Sauder JM, Rooney I, et al. (2005) Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 280: 1346–1353. doi:10.1074/jbc.M410968200.

28. Lewis HA, Wang C, Zhao X, Hamuro Y, Conners K, et al. (2010) Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. J Mol Biol 396: 406–430. doi:10.1016/j.jmb.2009.11.051.

29. Rosenberg MF, O'Ryan LP, Hughes G, Zhao Z, Aleksandrov LA, et al. (2011) The

cystic fibrosis transmembrane conductance regulator (CFTR): three-dimensional structure and localization of a channel gate. J Biol Chem 286: 42647–42654. doi:10.1074/jbc.M111.292268.

30. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, et al. (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol 14: 738–745. doi:10.1038/nsmb1278.

31. Vergani P, Lockless SW, Nairn AC, Gadsby DC (2005) CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. Nature 433: 876–880. doi:10.1038/nature03313.

32. Gottesman MM, Ambudkar SV (2001) Overview: ABC transporters and human disease. J Bioenerg Biomembr 33: 453–458. doi:10.1023/A:1012866803188.

CHAPTER 2

HOMOLOGY MODELS OF CFTR IN THE OPEN & CLOSED CHANNEL STATES

Abstract

We created a homology model of the open-channel state of CFTR, incorporating a combination of secondary structure predictions, hydropathy analysis, and pore accessibility indicators in the template-target alignment. Soft semi-harmonic restraints were added between particular residues in order to ensure that established salt bridges and distance constraints derived from disulfide crosslinking experiments in CFTR double cysteine mutants were satisfied. The established R352-D993 and R347-D924 salt bridge residue pairs are observed in our model, and a review of the surface accessibility status of predicted pore-lining residues in the critical TM6 and TM12 helices reveals that our model has more of these residues exposed to the aqueous conduction pore than previous open-state models. Close inspection of the pore structure reveals a narrow region, centered upon S341 in TM6 and T1134 in TM12, which separates the transmembrane pore into an inner and outer vestibule, consistent with the functional role of S341 in anion selectivity in CFTR. In addition, we also created a model of the closed state of CFTR using the inward-facing X-ray crystal structure of murine P-glycoprotein (P-gp) as a template, following a similar procedure to that used in generating the open state. We demonstrate the predictive power of our open and closed-state models by examination of the state-dependent accessibility of the arginine at position 334.

Introduction

Understanding structure-function relationships in CFTR will aid in the development of therapeutic treatments for the underlying cause of cystic fibrosis. To aid in this effort, X-ray crystal structures of the soluble first nucleotide binding domain (NBD1) [1], and a low resolution structure of the full protein from 2D crystals [2] have been published. However, despite sustained effort by numerous groups over several years, the structure of the rest of the protein has not yet been determined at adequate resolution to allow strong inference of structure-function relationships. The lack of a reliable molecular structure has hampered efforts to design new therapies that correct the basic defect in CF, and to understand the mechanism by which existing CFTR corrector drugs operate. Furthermore, interpretation of experimental data (e.g. from electrophysiological studies) is often challenging in the absence of structural framework in which to contextualize these results.

Homology modeling — often termed comparative modeling — has been explored in the past to propose candidate structural models of the full CFTR protein. In the following, we will briefly review the central ideas of homology modeling, the state-of-the-art in modeling of CFTR and related ABC transporter proteins, and describe the need for improved structural models of CFTR.

Homology Modeling

Homology modeling is a theoretical method (as opposed to an experimental method, such as X-ray crystallography or NMR) for the prediction of protein structure. However,

unlike purely physical theoretical approaches — which usually rely on modeling the interaction between the atoms of a protein using a variety molecular mechanics methods — homology modeling leverages knowledge of the experimentally determined structures of related *template* proteins to predict the conformation of the *target* protein of interest. The key observation that allows this method to be feasible is that small changes in the sequence of protein usually result in small changes in structure; in fact, structural homology is found to be more strongly preserved than sequence homology [3]. It stands to reason, therefore, that if an experimental structure related to a protein of interest in amino acid sequence can be found, it may serve as a useful starting point for the creation of a model structure of the target protein.

Many homology modeling protocols have been proposed and implemented, but the method proposed by Sali & Blundell [4] has received the most widespread acceptance. This is partly due to its relative ease of use for novices, and partly due to the powerful flexibility it affords to more experienced practitioners. In their method, one begins by supplying a superposition of the amino acid sequences of the template and target proteins. This alignment may be generated from a pairwise comparison of the sequences, optimizing against a suitable set of amino acid substitution penalties (e.g. as is usually done with a BLAST search), or one may employ additional indicators of good fit in more advanced methods, some of which will be discussed in the Methods section of this chapter. The modeling algorithm then generates, based on the sequence alignment, a network of spatial restraints (expressed as probability density functions) for a number of structural features expected in the target, based on the template. This includes, among other features, C-α–C-α distances, main-chain N–O distances, and main-chain and side-

chain dihedral angles. A 3D model that conforms to the input restraints as much as possible is then generated and optimized using a combination of conjugate gradient energy minimization and molecular-dynamics-based simulated annealing. One useful feature of this protocol is that it allows the inclusion of user-derived restraints in addition to the restraints automatically generated from the sequence alignment. This feature will be exploited in our modeling as described in the Methods section below.

ABC Transporter Crystal Structure Templates

The first step in the homology modeling procedure described above is to obtain a suitable 3D crystal structure of a protein closely related to CFTR. Of particular interest are structures from the ABC transporter superfamily of which, as mentioned earlier, CFTR is a member.

Our current understanding of structure-function relationships in ABC transporters is based on a small number of X-ray structures of varying quality. The structures of the bacterial ABC exporters Sav1866 (**Figure 2.1A**) (determined at a resolution of 3.0 Å) [5] and MsbA (3.7 Å) [6] were solved with ATP or non-hydrolyzable ATP analogs bound at the nucleotide binding sites. The NBDs in these structures form dimers, whereas the transmembrane domains are splayed apart, forming a large cavity accessible from the extracellular side. On the other hand, the structures of P-glycoprotein from mouse (**Figure 2.1B**) (determined at 3.8 Å resolution) [7] and *E. coli* [8] and of two low-resolution structures (>5 Å) of bacterial MsbA [6] have been determined in the absence of nucleotide. The NBDs are disassociated in these structures and far apart spatially,

whereas their membrane-spanning domains display an inward-facing cavity accessible from the cytoplasm.

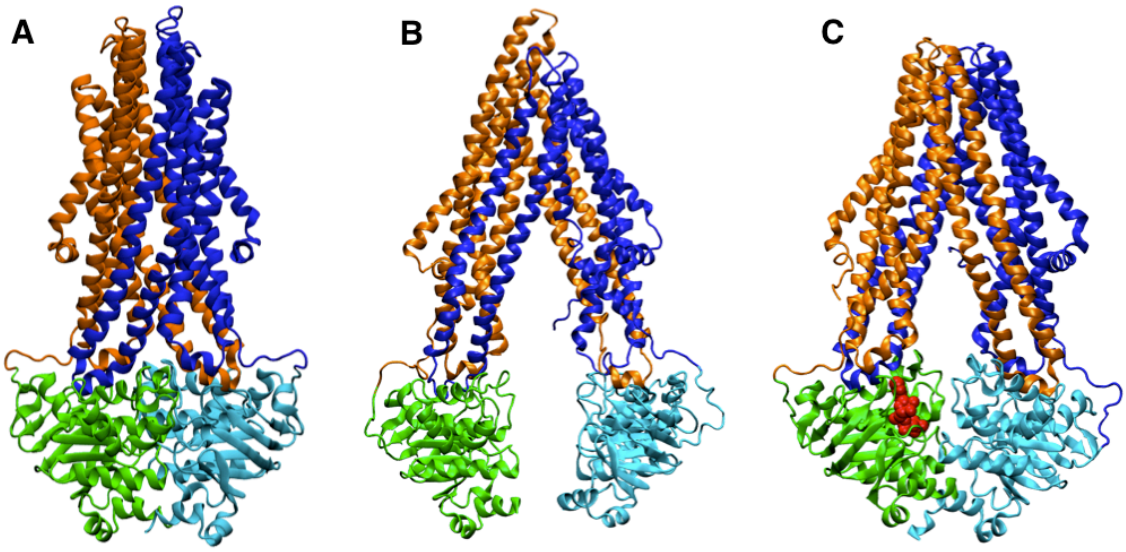

Figure 2.1: Ribbon representations of the crystal structures of A) Sav1866 [9], B) mouse P-gp [7], C) TM287/288 [10]

The crystal structure of TM287/288 (**Figure 2.1C**), an ABC exporter from *T. maritima*, was recently published at a resolution of 2.9 Å. This structure is of particular interest to CFTR researchers due to the presence of a degenerate ATP catalytic site on NBD1 that deviates from the ABC consensus sequence in the same positions as CFTR. The persistence of nucleotide at that site during crystallization led to the formation of a partial dimer structure that is likely to be analogous to the partial dimer C1 state that CFTR primarily occupies as its closed channel conformation.

Existing Homology Models of CFTR

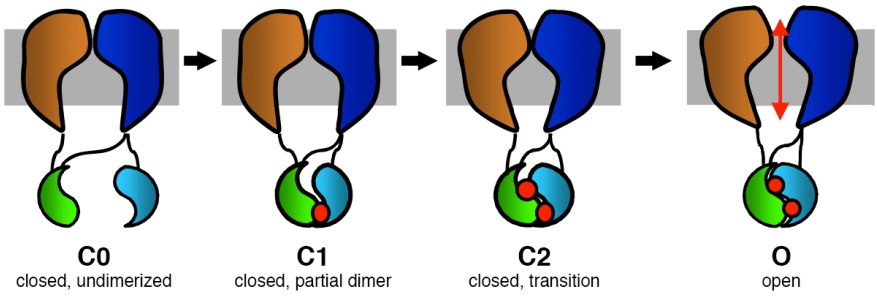

Figure 2.2: Stages in gating cycle of CFTR

Recall, from the proposed gating cycle (shown in **Figure 2.2**) that CFTR is thought to adopt one of several structural states, each of which corresponds to different stages of its transition pathway from the functional closed channel to the open channel. Although the changes in channel conduction are likely due to structural changes in the channel pore and the extra- and intracellular gates (located in the membrane-spanning domains (MSDs)), the transition is thought to be precipitated by changes in the structure and dimerization interactions of the intracellular nucleotide-binding domains (NBDs). Thus, while ATP binding and hydrolysis serve to power the alternating-access pump cycle of other ABC transporters, it appears that, CFTR, despite altering its primary role to that of a passive ion channel, has retained its ATP binding and catalytic domains to drive the gating of its conducting pore. To wit, the binding of ATP at both nucleotide binding sites and the resulting tight dimerization of the nucleotide binding domains is thought to be essential to channel opening in CFTR [11]. In the search for suitable templates for

homology modeling, therefore, it appears likely that an ABC transporter crystallized with bound ATP and tightly dimerized nucleotide binding domains would serve as a good basis for modeling the open channel O state of CFTR.

Following this line of reasoning, the elucidation of the complete structure of the bacterial multidrug transporter Sav1866 in an outward-facing pose [9] led to tremendous excitement in the field and its subsequent use as a template for several homology models of the CFTR open channel state [12-17]. The Sav1866 structure has several features that makes it a good candidate for use in modeling CFTR. First, it is a multidrug exporter (i.e. it was responsible for pumping substrates from the intracellular environment out to the extraceullar side) and the sub-family of ABC transporters with closest homology to CFTR — the ABC-C subfamily — are all also multidrug exporters. Secondly, it was crystallized at sufficiently high resolution to allow determination of important surface-lining residues and specific inter-residue contacts that may be important in maintaining structure (and may thus allow inference of similar information in the modeled CFTR structure). Finally, the assembled molecule had a topological layout — two nucleotide binding domains and two membrane-spanning domains each with six transmembrane α-helices and long intracellular loops — similar to CFTR (sans, of course, the unique-to-CFTR R-domain).

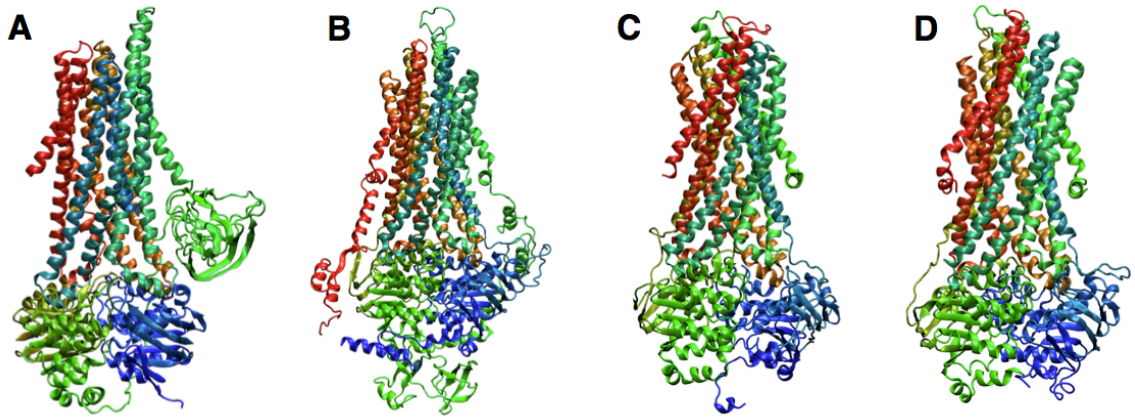

Figure 2.3: Existing homology models of CFTR by A) Serohojis et al. [18], B) Mornon et al. [13], C) Alexander et al. [19], and D) Dalton et al. [15]

However, researchers quickly realized that, while there was a fair degree of conservation between the amino acid sequences of CFTR and Sav1866 in their respective nucleotide binding domains, the sequence homology in the MSDs was very low (<20%), making the transmembrane segments of CFTR the most difficult to align to the Sav1866 template. This is unfortunate, since this region constitutes the conduction pore and is therefore critical to CFTR's function as an ion channel. The implications of this challenge were reflected in the quality of the models generated. While the gross structures of all the models were similar (**Figure 2.3**) a closer analysis reveals significant discrepancies in the register and secondary structure assignment of these models, with no clear indications of any model being clearly better than the others. Prior experimental evidence (particularly from electrophysiology experiments) suggested several experimental features that must exist in any valid model of the CFTR open channel structure. To our disappointment, when we inspected all the models for which coordinates were made available to us for conformance to these features, we found that none of them met all the criteria for a truly

valid model (**Table 2.1**). This signaled to us the need for a new model of the open CFTR channel state, based on the same Sav1866 template, but designed with cognizance of the need to maintain important features known from decades of experimental studies on CFTR.

Table 2.1: Expected structural features of CFTR

	Riordan	Callebaut	Dawson
R334 at outer mouth of pore?	Yes	Yes	Yes
S341 at narrowest point of pore?	No	Yes	Yes
Kink in TM6 at V350?	Yes	No	No
R352 and D993 pointing to each other?	Yes	No	Yes
R347 and D924 pointing to each other?	No	No	No

It also has been suggested that the inward-facing apo state of the typical ABC transporter may correspond to the C0 closed channel in CFTR. Such a choice may be justified for two reasons. First, the fact that ATP binding is found to be necessary for channel opening implies that ATP-free structures of homologous proteins may correspond to the closed CFTR channel. Secondly, all the nucleotide-free crystal structures published so far display a membrane-spanning domain with a wide cellular cavity that narrows to a close in the transmembrane region of the protein. This narrow (in fact, occluded) region may

represent the extracellular gate of the CFTR pore, which prevents conduction in the closed channel state.

Mornon et al. [14] proposed such a structure based on homology modeling using as template a corrected X-ray crystal structure of the bacterial ABC lipid flippase MsbA [6]. However, this template suffered from poor resolution (5.5 Å), compounding the problem of alignment described above, and making it more difficult to ascertain the accuracy of their homology model. With the publication of new high-resolution (3 Å) ATP-free apo crystal structures of the P-glycoprotein [7,8] — which, like CFTR's nearest relatives, is also a multidrug exporter, and also has twelve transmembrane helices — we saw an opportunity to improve the state-of-the-art once more with the creation of a new homology model of the closed channel state of CFTR based on P-glycoprotein.

Methods

Homology Modeling of the Open State

Sequence Alignment

A homology model of the open state of CFTR was created using the crystal structure of Sav1866 [5] as a template. As a general rule of thumb, homology modeling based on amino acid sequence comparison alone is thought to be reliable only if the sequence similarity between the target and the template exceeds 30%. Due to low sequence homology between CFTR and Sav1866 in the membrane-spanning domains (~16%), we had to seek out alternative strategies to increase our confidence in our alignment.

It has been demonstrated that the addition of additional profiles to the alignment can aid in generating better end results. Analogous to the dictum "(3D) structure is more strongly conserved than sequence" — the very basis of comparative modeling — it has been observed that other features such as secondary structure elements, patches of hydrophobicity, and regions of transmembrane embedment are also strongly conserved even in the face of changing sequence over the course of evolution. For example, if a protein contains a span of sequence that forms a transmembrane alpha helix, one may reasonable expect that the overall helicity of this region — a property of its amino acid composition — is maintained in a comparable structural region in its close homologs, even if the exact amino acid sequence is not. Similarly, one may also expect the amphipathic nature of this helix (due to the fact that it is alternately pore-facing and membrane-facing) to also persist in homologous proteins, perhaps to a greater degree than the particulars of its amino acid sequence. There exist algorithms that can predict, based on sequence, the propensity of a particular sequence region to form particular secondary structure elements, its hydrophobicity, and its likelihood of being embedded in a membrane. It has been demonstrated [20] that superimposing and aligning the results from these prediction algorithms for both the template and target sequences can greatly improve accuracy of conventional amino-acid-substitution-based alignments.

In creating an open state model, a sequence alignment between human CFTR and *S. aureus* Sav1866, we generated profiles based on sequence homology (using the BLOSUM62 substitution matrix), a sliding-window average of predicted hydrophobicity (using the KD hydrophobicity scale [21]), PSIPRED predictions of secondary structure [22], and OCTOPUS transmembrane topology predictions [23]. These were used as

weighted inputs to the program AlignMe [20]. In addition, the correct alignment of buried and exposed residues in transmembrane helix 6 (TM6) and 12 (TM12) was enforced by adding a custom measure of similarity to the AlignMe input. First, a profile of percentage solvent accessible surface areas was generated for each residue in the Sav1866 template in TM6 and TM12 using VMD [24]. A counterpart profile on the CFTR sequence was then generated by conducting an extensive literature review on CFTR structure-function studies that reveal information on whether particular residues in these important pore-lining helices are exposed (i.e. pore-facing) or buried (i.e. in the faces of the helix turned away from the pore), including cysteine scanning and open pore blocker experiments. Residues in the CFTR profile were assigned a score of 100 if predicted to be exposed, 0 if buried. The program then used this information to align residues in the CFTR sequence thought (from experiment) to be pore-accessible to sites on the Sav1866 template with high solvent-accessible surface areas (from the crystal structure). The region of the CFTR sequence containing the R-domain (residues 650 through 856) was omitted from all of the modeling in this study, since no analogous structure exists in the ABC templates, and it is in fact thought to be largely unstructured [25].

Restraint-based Homology Modeling

The Sav1866 template structure was crystallized with ADP at both NBD binding sites, but these were removed in our homology models. Homology modeling was carried out in Modeller v. 9.10 which efficiently implements the method of Sali & Blundell described earlier [4]. An advantage of this method of homology modeling is the ability to add user-defined spatial restraints to the final structure. Our group and others have demonstrated

the existence of salt-bride interactions between the residues R347-D924 [26] and R352-D993 [27]. In addition, several residues in the nucleotide binding domains and the intra-cellular loops are thought to be cross-linkable, suggesting they exist in the open state in close spatial proximity. This allowed us to supply a set of distance restraints (**Table 2.2**) in the form of soft semi-harmonic potentials to further refine the model. In the modeler protocol, these restraints were added to those automatically generated based on the provided template structure (Sav1866, PDB: 2HYD) and sequence alignment (from the AlignMe results above), to create initial models that were then subjected to 5000 steps of conjugate gradient energy minization and a standard simulated annealing protocol using molecular dynamics. Ten models were thus generated, and then evaluated based on observed maintenance of the known salt bridges and solvent accessibility of predicted pore-lining residues in TM6 and TM12. The best of these — henceforth referred to as the *O-CFTR* model — was chosen for further study.

The stereochemical quality of the final model was validated using PROCHECK [28]. The radius profile of the ion conduction pore along the axis normal to the lipid bilayer was analyzed using the program HOLE [29].

Table 2.2: Distance restraints used in homology modeling of O-CFTR

Residues	Distance	Reference
R352-D993	≤ 5 Å (salt bridge)	{Cui:2008hh}
R347-D924	≤ 5 Å (salt bridge)	{Cotten:1999vr}
M348-T1142	≤ 13 Å	{Chen:2004jy}
T351-T1142	9 Å ≤ d ≤ 13 Å	{Chen:2004jy}
S605-A1374	≤ 8 Å	{Mense:2006et}
S549-S1248	≤ 8 Å	{Mense:2006et}
S434-D1336	8 Å ≤ d ≤ 16 Å	{Mense:2006et}
S549-S1374	8 Å ≤ d ≤ 16 Å	{Mense:2006et}

Homology Modeling of the Closed State

Sequence Alignment

A similar procedure to that described above was used to generate a sequence alignment between CFTR and the closed state template, murine P-glycoprotein (PDB: 3G5U) [7], using sequence, hydrophobicity and predicted secondary structure information. An interesting observation was made prior to generating the sequence alignment: an initial comparison of the lengths of the extracellular loops revealed similarities between lengths of the extracellular loops (ECLs) of the first membrane-spanning domain (MSD1) of CFTR and the second (MSD2) of P-gp and vice versa (**Figure 2.4**). Furthermore, we noted that the site of N-glycosylation in the ECLs occurs in ECL4 (the first ECL of MSD2) in CFTR, and in ECL1 (the first ECL of MSD1) in P-gp. Therefore, we chose to

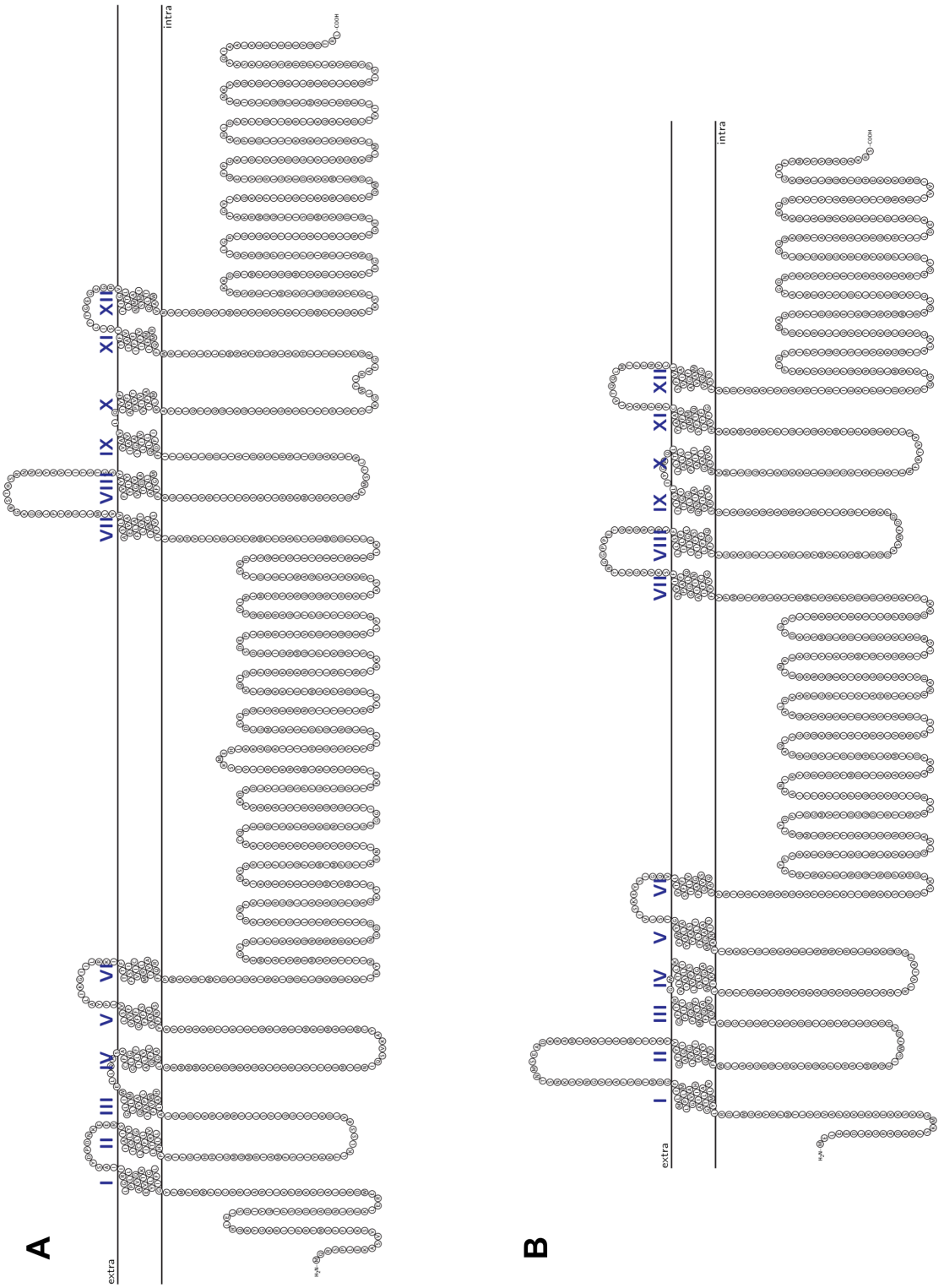

Figure 2.4: Membrane topologies of A) CFTR and B) P-gp.

align MSD1 of CFTR with MSD2 of P-gp, and *vice versa*. Given the low sequence homology between P-gp and CFTR in the MSDs, it is not possible to determine whether this domain switching has biological significance. However, the assertion that full-transporters such as CFTR and P-gp may have been independently assembled from ancestral half-transporter genes [30] does not preclude this possibility.

Homology Modeling

Unlike in the modeling of the open state, we did not have a large body of structural information that could be incorporated in our modeling of the closed state. This may, in part, be due to the fact that it is difficult to design experiments in which one can definitely conclude that the protein was present and in the closed channel conformation. We can reasonably expect, however, that gross reorganization of the transmembrane topology would not occur in the transition from the closed state to the open state. That is to say, residues embedded in the transmembrane region in the creation of the *O-CFTR* already built model were expected to persist in the membrane in the closed state model. We thus first generated predictions of the position of the transmembrane region in the *O-CFTR* model using the Orientation of Proteins in Membranes (OPM) database [31]. This information was added to as a profile to the CFTR sequence, and the corresponding profile in the P-gp sequence was generated by inspection of the template P-gp structure. In this way, residues predicted to be in the transmembrane region in the O-CFTR model were aligned to corresponding transmembrane residues in the P-gp template, ensuring topological consistency between our two generated models. No additional restraints were used during modeling in Modeller 9.10.

PROCHECK was used to check stereochemical quality of the generated C0-CFTR model which revealed 93% of the residues to be in the most-favored regions, 7% in allowed regions, and none in disallowed regions.

Results

New Homology Model of the Open CFTR Channel

We created a homology model of the open-channel state of CFTR based on a Sav1866 template using the comparative modeling method of Sali & Blundell [4]. The full protein was modeled, with the exception of the unique-to-CFTR R-domain, for which a suitable homologous structure could not be found. Due to low sequence homology (particularly in the MSDs), primary structure alone could not be used to generate a reliable alignment. Instead, a combination of secondary structure predictions, hydropathy analysis, and pore accessibility indicators were incorporated into the alignment (**Figure 2.5**). Soft semi-harmonic restraints were manually added between particular residues in order to ensure that established salt bridges and distance constraints derived from disulfide crosslinking experiments in CFTR double cysteine mutants were satisfied. Ten such models were generated and energy-minimized using a simulated annealing protocol, then ranked based on maintenance of the established R347-D924 [26] and R352-D993 [27] salt bridges, and accessibility of residues in TM6 and TM12 predicted to be pore-lining. The model that best fit these criteria (designated O-CFTR) was chosen for further study (**Figure 2.6**). PROCHECK was used to check the validity of our model, and showed 95% of residues to be in the most favorable regions with none in disallowed regions (**Figure 2.7**).

Figure 2.5: Sequence alignment used to generate O-CFTR model

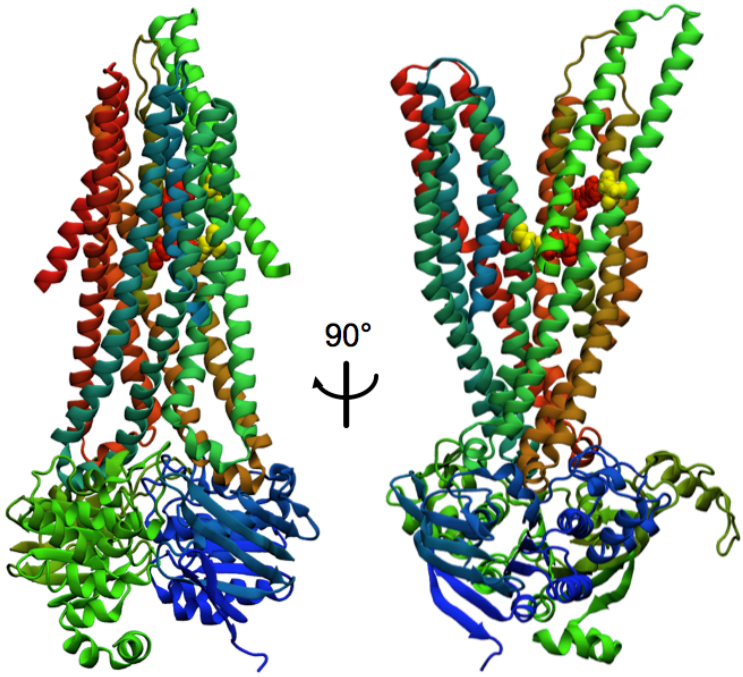

Figure 2.6: Ribbon representation of O-CFTR model

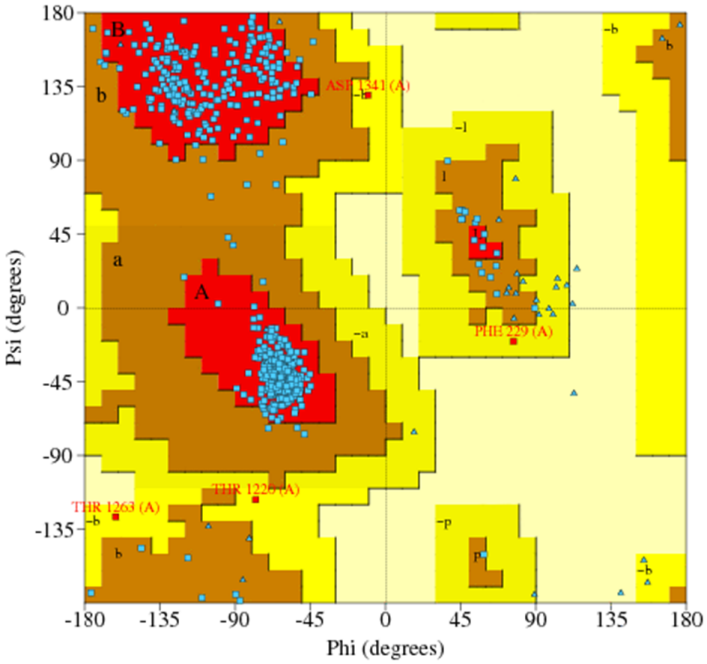

Figure 2.7: Ramachandran plot of O-CFTR model

By design, the O-CFTR model contains several structural features expected to be found in the real CFTR structure. The established R352-D993 and R347-D924 salt bridge residue pairs are observed in our model to have side chains oriented toward each other with distances <5 Å between their interacting guanidinium and carboxyl groups (**Figure 2.8**). This is in contrast to existing homology models, none of which correctly depicts these salt-bridges. Indeed, when we add our *O-CFTR* model to the list of expected features that motivated our modeling project, we are vindicated by the finding that our model does in fact meet all the criteria we had initially set for it. Not that several of these criteria were not explicitly controlled for, but were rather bona fide predictions of our

model. This gives us confidence that our model can, in fact be used to make useful predictions about CFTR structure and function.

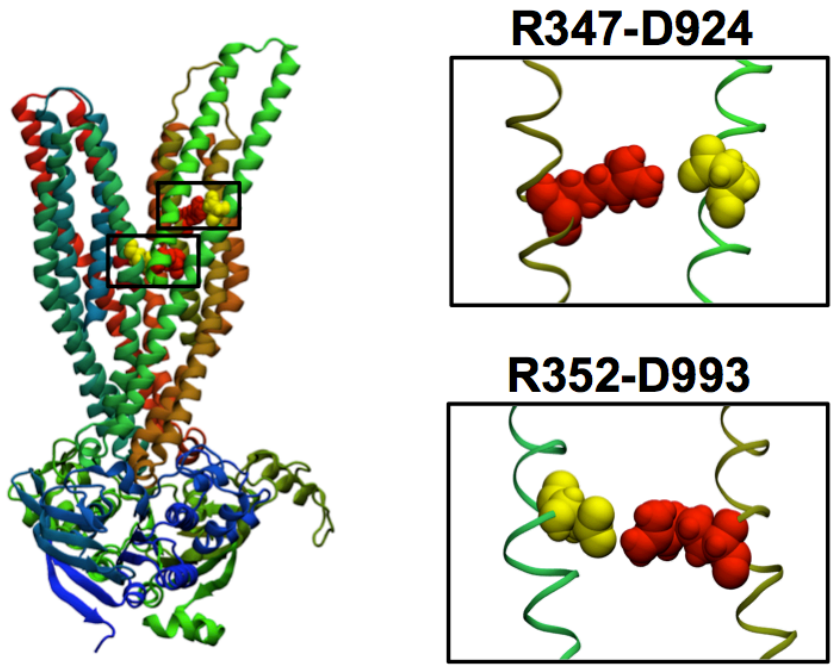

Figure 2.8: Salt bridges in O-CFTR model

We also reviewed the surface accessibility status of predicted pore-lining residues in the critical TM6 [27,32-42] and TM12 [32,43-46] helices. In order to determine accessibility status, we measured the percentage of the total surface area accessible to solvent for each residue of interest in each available model. A commonly employed cutoff of 20% was used as a classifier threshold [47]. Analysis of solvent-accessible surface area in these helices revealed that our *O-CFTR* model has more of these residues exposed to the

aqueous conduction pore than several previous open-state models, including those of Serohijos et al.[18], Mornon et al. [14], Dalton et al. [15], and the 5 ns snapshot of the 30 ns molecular dynamics (MD) simulation obtained from Norimatsu et al. [16] (**Table 2.3**). Of 18 residues expected to be pore-accessible based on cysteine scanning and open pore blocker experiments, 12 were predicted to be accessible (>20% surface area accessible to solvent) in our model.

Table 2.3: Pore accessibility of residues in TM6 and TM12

		% Solvent-Accessible Surface Area				
		Current	**Serohijos**	**Mornon**	**Norimatsu**	**Dalton**
TM6	**K335**	35	0	17	40	30
	F337	21	38	37	0	28
	T338	30	13	23	20	19
	S341	25	12	32	12	23
	I344	14	27	41	8	13
	V345	23	2	41	30	13
	M348	29	26	17	14	4
	V350	8	0	0	1	0
	R352	29	26	43	9	10
TM12	**T1134**	23	10	36	28	28
	N1138	28	19	24	25	13
	M1140	7	0	0	8	2
	S1141	24	0	13	14	15
	T1142	21	27	1	0	2
	Q1144	7	12	13	27	19
	W1145	42	17	31	36	30
	V1147	0	0	0	0	1
	N1148	9	0	36	26	21
# In Agreement		**12**	**5**	**10**	**7**	**6**

Interestingly, decreased sensitivity to open pore blockers by mutations at V350 [42] have suggested its important role in CFTR channel structure, but none of the Sav1866-based models, including our current model, show it to be accessible from the pore. V1147 on TM12, also thought to be pore-accessible [42,45] is similarly buried in all the models. This suggests that the effects of mutations at V350 and V1147 are indirect, or that a departure of CFTR structure from the Sav1866 template exists at the cytosolic end of the transmembrane region. Indeed, a drawback of our model is that there is no clear entrance to the transmembrane pore from the cytoplasmic end, a consequence of the alternating-access exporter function of the Sav1866 template shared by most existing homology models of the open CFTR channel.

A closer inspection of the pore structure in the transmembrane region of our O-CFTR model was conducted using the program HOLE [29] (**Figure 2.9**). This program traverses the model structure in the direction parallel to the pore axis, mapping the largest discs that can be circumscribed within the pore at every height and returning the pore radius profile as a function of distance along the axis. In addition, it also can also be used to generate graphical representations of the pore. This analysis reveals the pore radius in our O-CFTR model to be continuously larger than 3.6 Å (the diameter of a dehydrated chloride ion [48]), indicating that the channel shown in our model is indeed open to conduction. We note also the existence of a narrow region, centered upon S341 in TM6 and T1134 in TM12, which separates the transmembrane pore into an inner and outer vestibule. This pore architecture is consistent with recent predicted models [15,16] as well as with experimental evidence suggesting the existence of a wide inner vestibule capable of binding pore-occluding blockers [49], and the narrowing of the pore in the

region of S341 [16]. Interestingly S341 is thought to play a vital role in anion selectivity in CFTR, and its structural position in the narrow region of the pore may be vital to its discriminatory role [33].

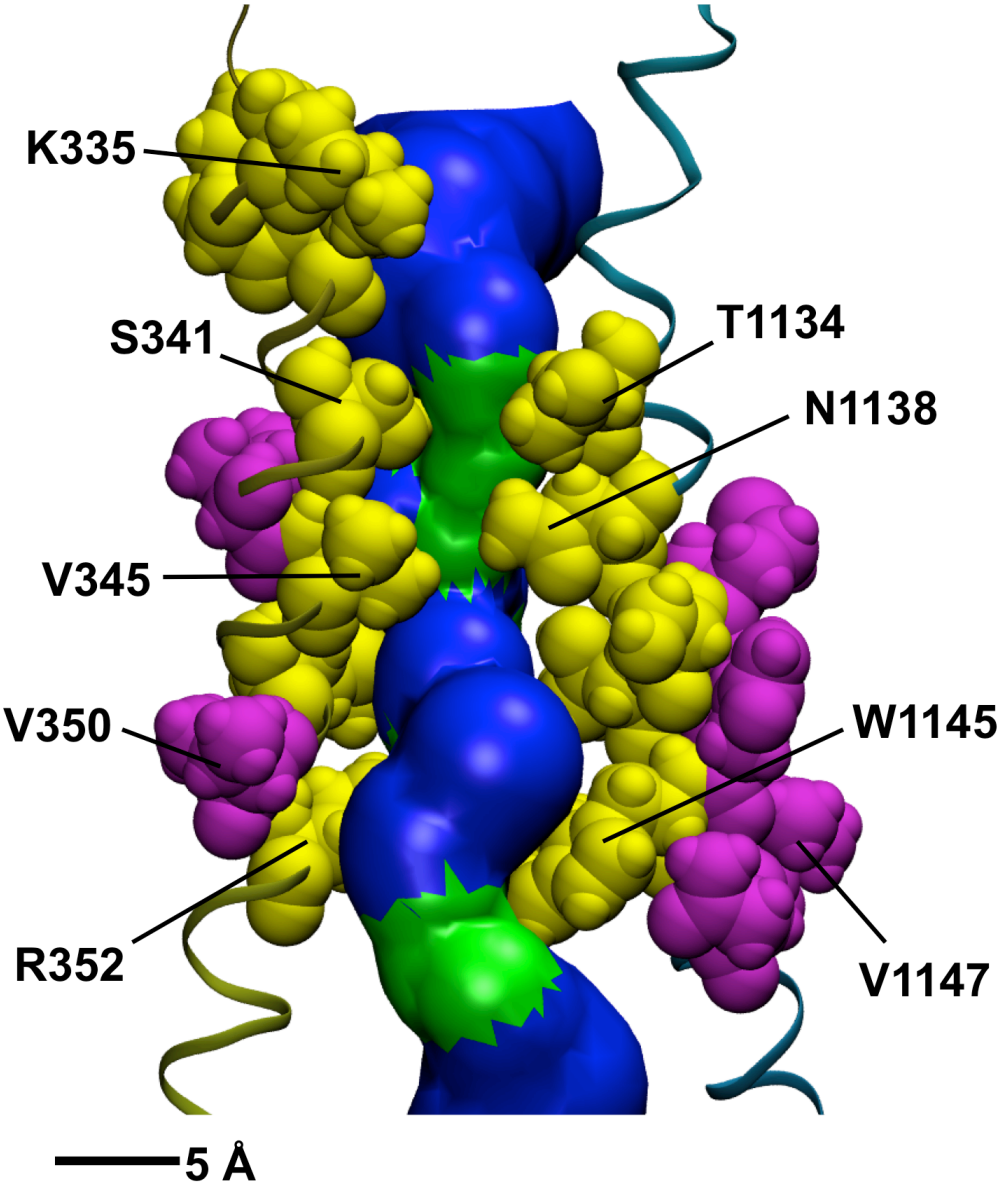

Figure 2.9: Close-up of the transmembrane pore in O-CFTR

New Homology Model of the Closed CFTR Channel

In addition to a Sav1866-based open state model, we also created a model of the proposed nucleotide-free closed state of CFTR using the inward-facing X-ray crystal structure of murine P-glycoprotein (P-gp) [7] as a template.

Prior to generating a sequence alignment between CFTR and P-gp, we noted a discrepancy in the sizes of the extracellular loops (ECLs) in the P-gp structure and our proposed O-CFTR model. In particular, ECL1 between TM1 and TM2 in P-gp (~41 residues long) was much larger than ECL1 in O-CFTR (~12 residues), but similar in length to ECL4 between TM7 and TM8 in O-CFTR (~33 residues) (Figure S2). Therefore, in order to facilitate modeling, we chose to use P-gp's MSD1 as a template for CFTR's MSD2 and vice versa. A similar procedure to that used in generating the open state model was then followed, creating a sequence alignment between CFTR and mouse P-gp and manually adjusting it to ensure a topological match between the O-CFTR model already generated and the closed state model (**Figure 2.10**). The final closed state homology model (**Figure 2.11**) was created in MODELLER, and validated for stereochemical quality using PROCHECK [28].

As may be expected from the structure of the P-gp template, the NBDs in C0-CFTR are completely dissociated, and the MSDs tilt toward each other to form a constriction near the extracellular end of the transmembrane region. This gate, comprised of the extracellular-end residues in TM6, TM12, TM1 and TM7 along with ECL1, ECL3, ECL4, and ECL6, and perhaps with contributions from N-linked glycans (not modeled)

in ECL4, provides an effective barrier to permeation, consistent with the observation that the CFTR channel is closed in the absence of NBD dimerization.

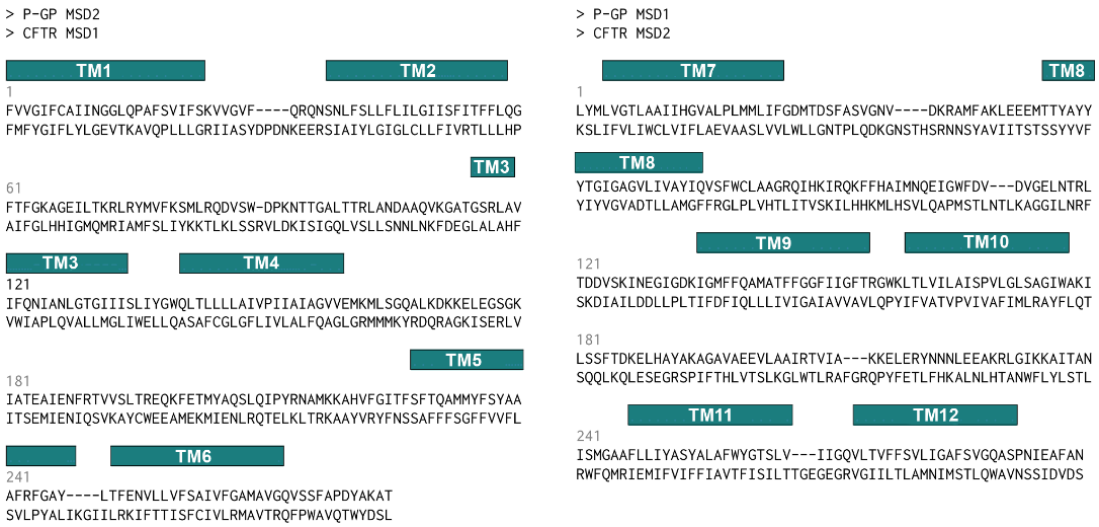

Figure 2.10: Sequence alignment used to generate C0-CFTR model

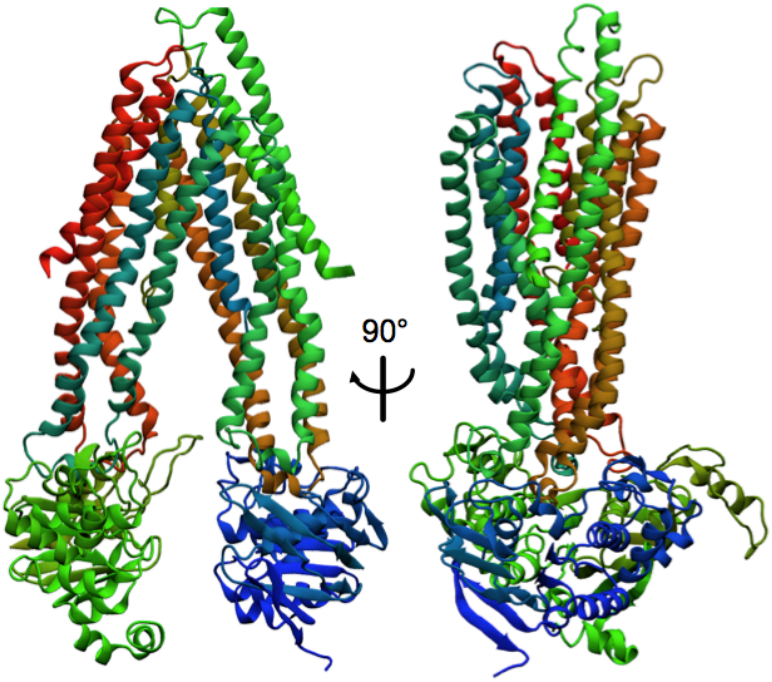

Figure 2.11: Ribbon representation of C0-CFTR model

State-Dependent Accessibility of R334C

The arginine at position 334 on TM6 lies at the outer mouth of the pore, and has been suggested to play a role in attracting anions [50,51]. It has been observed, however, that modification of cysteine mutants at this site (R334C) occurs only in the closed state, with the site apparently being inaccessible when the channel is open [52-54]. Norimatsu et al. [54] have suggested that this state-dependent accessibility may be due to movement of residues neighboring R334, and a comparison of this region in our two models provides support for this hypothesis. In an open state O-CFTR model in which the arginine at position 334 was mutated to a cysteine in silico, R334C is buried by many of its neighboring residues in ECL3 and is found to have a fractional solvent-accessible surface area (SASA) of only 3.8%. By comparison, the C0-CFTR model presents R334C in a relatively exposed conformation, with 32.9% of its surface area accessible to solvent **(Figure 2.12**). This increased accessibility may translate to the higher observed rates of reaction of R334C with extracellular sulfhydryl-modifying reagents in the closed state. A similar SASA analysis of the C0-CFTR and O-CFTR models with the native arginine residue reveals that this position may exhibit similar state-dependent accessibility in wildtype CFTR (**Figure 2.13**).

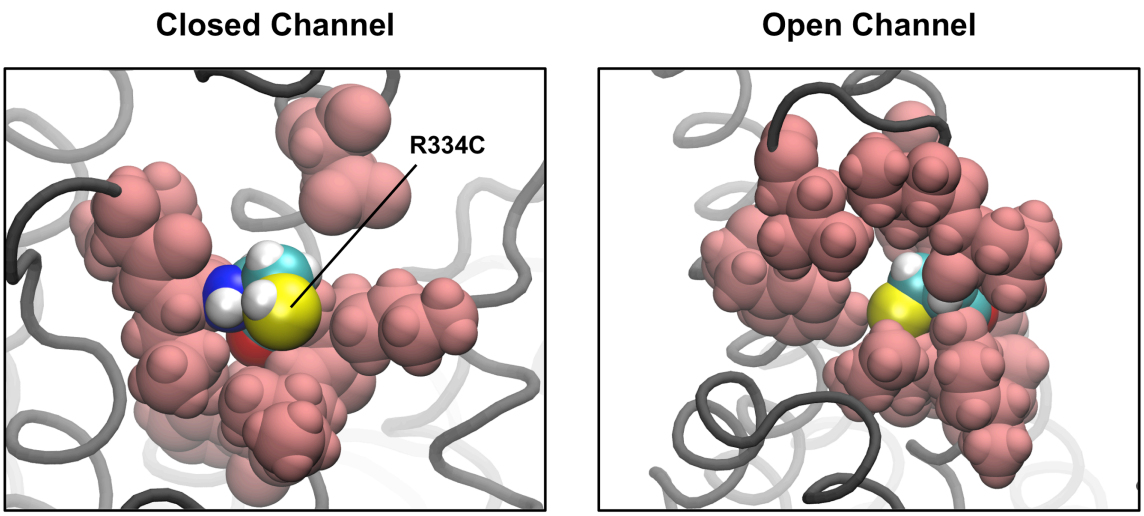

Figure 2.12: State-dependent accessibility of R334C

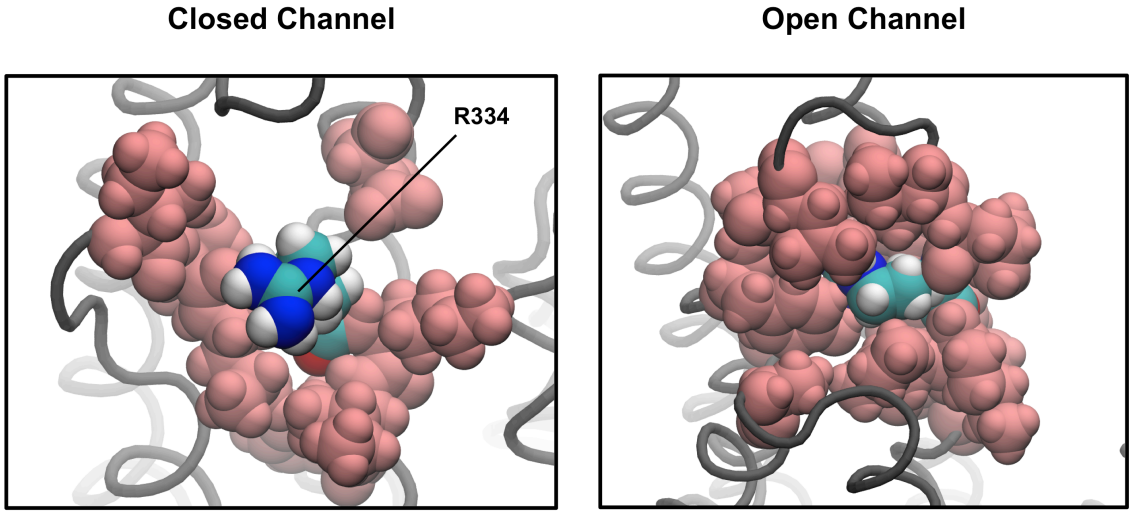

Figure 2.13: State-dependent accessibility of R334

Discussion

New Homology Models

Understanding structure-function relationships in CFTR will aid in the development of therapeutic treatments for the underlying cause of CF. Data from experimental studies, in addition, provide insight about functional states (also reflecting ensembles of conformers) and interactions that are more likely than others. Our aim in this study was to create a single consensus model that best incorporates much of the available experimental information about the CFTR open channel structure. Previous models proposed by Serohijos et al. [18], Alexander et al. [19] and Norimatsu et al. [16] did not include the established salt bridge between residues R347 and D924 [26], and the R352-D993 [27] salt bridge interaction did not exist in the open state model of Mornon et al. [14]. Our Sav1866-based O-CFTR model correctly accommodates all of this structural information, and, like the models from Dalton [15] and Norimatsu [16] and colleagues, contains a transmembrane pore region depicting inner and outer vestibules separated by a narrow region near the selectivity-conferring residues S341 and T338 [33]. Our C0-CFTR closed channel model uses a mouse P-gp template reported at much higher resolution than the MsbA structure previously used by Mornon et al. [14]; our model also included a swap of the N- and C-terminal to improve the alignment (see Methods) that we believe provides a superior alignment than the naïve implementation.

Although CFTR is thought to cycle between a closed state with partially-dimerized NBDs (C1) and a tight-dimer open state (O) during its action as a chloride channel, it is also known to transport large substrates such as the glutathione antioxidant tripeptide (GSH)

[55]. Given the disparity in size between the chloride ion and GSH, transport of GSH may require the much larger inner vestibule provided by the fully NBD-dissociated state represented in our C0-CFTR model.

While some experimental data, such as the salt bridge interactions and accessibility of pore-lining residues, were built into our models as constraints, we also observe that the models accurately predict results that were not explicitly controlled for during model generation. For example, cysteine mutants at residue R334 in the outer mouth of the pore [52-54] have been demonstrated to be more reactive to MTS reagents in the closed channel state than in the open channel. This residue is more exposed in our C0-CFTR model than in our O-CFTR model. These findings give us confidence in the predictive power of these models and in their ability to explain and interpret functional data in terms of specific structural changes in the CFTR molecule. It is our hope that these improved models of CFTR will serve as a useful visualization tool for experimenters in designing and interpreting results from structure-function studies, and as an accurate reference state from which to begin simulation studies of CFTR.

References

1. Lewis HA, Buchanan SG, Burley SK, Conners K, Dickey M, et al. (2003) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23: 282–293. doi:10.1038/sj.emboj.7600040.

2. Rosenberg MF, O'Ryan LP, Hughes G, Zhao Z, Aleksandrov LA, et al. (2011) The cystic fibrosis transmembrane conductance regulator (CFTR): three-dimensional structure and localization of a channel gate. J Biol Chem 286: 42647–42654. doi:10.1074/jbc.M111.292268.

3. Hubbard TJ, Blundell TL (1987) Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Eng 1: 159–

171.

4. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815. doi:10.1006/jmbi.1993.1626.

5. Dawson RJP, Locher KP (2007) Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. FEBS Lett 581: 935–938. doi:10.1016/j.febslet.2007.01.073.

6. Ward A, Reyes CL, Yu J, Roth CB, Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci U S A 104: 19005–19010. doi:10.1073/pnas.0709388104.

7. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722. doi:10.1126/science.1168750.

8. Jin MS, Oldham ML, Zhang Q, Chen J (2012) Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. Nature 490: 566–569. doi:10.1038/nature11448.

9. Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180–185. doi:10.1038/nature05155.

10. Hohl M, Briand C, Grütter MG, Seeger MA (2012) Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 19: 395–402. doi:10.1038/nsmb.2267.

11. Vergani P, Lockless SW, Nairn AC, Gadsby DC (2005) CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. Nature 433: 876–880. doi:10.1038/nature03313.

12. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066–1073.

13. Mornon J-P, Lehn P, Callebaut I (2008) Atomic model of human cystic fibrosis transmembrane conductance regulator: membrane-spanning domains and coupling interfaces. Cell Mol Life Sci 65: 2594–2612. doi:10.1007/s00018-008-8249-1.

14. Mornon J-P, Lehn P, Callebaut I (2009) Molecular models of the open and closed states of the whole human CFTR protein. Cell Mol Life Sci 66: 3469–3486. doi:10.1007/s00018-009-0133-0.

15. Dalton J, Kalid O, Schushan M, Ben-Tal N, Villà-Freixa J (2012) New model of cystic fibrosis transmembrane conductance regulator proposes active channel-like conformation. J Chem Inf Model 52: 1842–1853. doi:10.1021/ci2005884.

16. Norimatsu Y, Ivetac A, Alexander CS, Kirkham J, O'Donnell N, et al. (2012) Cystic fibrosis transmembrane conductance regulator: a molecular model defines the architecture of the anion conduction path and locates a “bottleneck” in the pore. Biochemistry 51: 2199–2212. doi:10.1021/bi201888a.

17. Furukawa-Hagiya T, Furuta T, Chiba S, Sohma Y, Sakurai M (2013) The power stroke driven by ATP binding in CFTR as studied by molecular dynamics simulations. J Phys Chem B 117: 83–93. doi:10.1021/jp308315w.

18. Serohijos AWR, Hegedus T, Aleksandrov AA, He L, Cui L, et al. (2008) Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function. Proc Natl Acad Sci U S A 105: 3256–3261. doi:10.1073/pnas.0800254105.

19. Alexander CS, Ivetac A, Liu X, Norimatsu Y, Serrano JR, et al. (2009) Cystic fibrosis transmembrane conductance regulator: using differential reactivity toward channel-permeant and channel-impermeant thiol-reactive probes to test a molecular model for the pore. Biochemistry 48: 10078–10088. doi:10.1021/bi901314c.

20. Stamm M, Staritzbichler R, Khafizov K, Forrest LR (2013) Alignment of helical membrane protein sequences using AlignMe. PLoS One 8: e57731. doi:10.1371/journal.pone.0057731.

21. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132.

22. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404–405.

23. Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24: 1662–1668.

24. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–38, 27–28.

25. Baker JMR, Hudson RP, Kanelis V, Choy W-Y, Thibodeau PH, et al. (2007) CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol 14: 738–745. doi:10.1038/nsmb1278.

26. Cotten JF, Welsh MJ (1999) Cystic fibrosis-associated mutations at arginine 347 alter the pore architecture of CFTR. Evidence for disruption of a salt bridge. J Biol Chem 274: 5429–5435.

27. Cui G, Zhang Z-R, O’Brien ARW, Song B, McCarty NA (2008) Mutations at arginine 352 alter the pore architecture of CFTR. J Membr Biol 222: 91–106. doi:10.1007/s00232-008-9105-9.

28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26: 283–291. doi:10.1107/S0021889892009944.

29. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MSP (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 14: 354–360. doi:10.1016/S0263-7855(97)00009-X.

30. Jordan IK, Kota KC, Cui G, Thompson CH, McCarty NA (2008) Evolutionary and functional divergence between the cystic fibrosis transmembrane conductance regulator and related ATP-binding cassette transporters. Proc Natl Acad Sci U S A 105: 18865–18870. doi:10.1073/pnas.0806306105.

31. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22: 623–625. doi:10.1093/bioinformatics/btk023.

32. Zhang Z-R, Zeltwanger S, McCarty NA (2000) Direct comparison of NPPB and DPC as probes of CFTR expressed in Xenopus oocytes. J Membr Biol 175: 35–52.

33. McCarty NA, Zhang Z-R (2001) Identification of a region of strong discrimination in the pore of CFTR. Am J Physiol Lung Cell Mol Physiol 281: L852–L867.

34. Gong X, Burbridge SM, Cowley EA, Linsdell P (2002) Molecular determinants of Au $(CN)_2^-$ binding and permeability within the cystic fibrosis transmembrane conductance regulator Cl^- channel pore. J Physiol 540: 39–47.

35. Gupta J, Evagelidis A, Hanrahan JW, Linsdell P (2001) Asymmetric structure of the cystic fibrosis transmembrane conductance regulator chloride channel pore suggested by mutagenesis of the twelfth transmembrane region. Biochemistry 40: 6620–6627.

36. Liu X, Zhang Z-R, Fuller MD, Billingsley J, McCarty NA, et al. (2004) CFTR: a cysteine at position 338 in TM6 senses a positive electrostatic potential in the pore. Biophys J 87: 3826–3841. doi:10.1529/biophysj.104.050534.

37. Aubin CNS, Linsdell P (2006) Positive charges at the intracellular mouth of the pore regulate anion conduction in the CFTR chloride channel. J Gen Physiol 128: 535–545. doi:10.1085/jgp.200609516.

38. Beck EJ, Yang Y, Yaemsiri S, Raghuram V (2008) Conformational changes in a pore-lining helix coupled to cystic fibrosis transmembrane conductance regulator channel gating. J Biol Chem 283: 4957–4966. doi:10.1074/jbc.M702235200.

39. Fatehi M, Linsdell P (2008) State-dependent access of anions to the cystic fibrosis transmembrane conductance regulator chloride channel pore. J Biol Chem 283: 6102–6109. doi:10.1074/jbc.M707736200.

40. Serrano JR, Liu X, Borg ER, Alexander CS, Shaw CF, et al. (2006) CFTR: Ligand exchange between a permeant anion ($[Au(CN)_2]^-$) and an engineered cysteine (T338C) blocks the pore. Biophys J 91: 1737–1748. doi:10.1529/biophysj.105.078899.

41. Bai Y, Li M, Hwang T-C (2010) Dual roles of the sixth transmembrane segment of the CFTR chloride channel in gating and permeation. J Gen Physiol 136: 293–309. doi:10.1085/jgp.201010480.

42. Cui G, Song B, Turki HW, McCarty NA (2012) Differential contribution of TM6 and TM12 to the pore of CFTR identified by three sulfonylurea-based blockers. Pflugers Arch 463: 405–418. doi:10.1007/s00424-011-1035-1.

43. Bai Y, Li M, Hwang T-C (2011) Structural basis for the channel function of a degraded ABC transporter, CFTR (ABCC7). J Gen Physiol 138: 495–507. doi:10.1085/jgp.201110705.

44. Zhou J-J, Li M-S, Qi J, Linsdell P (2010) Regulation of conductance by the number of fixed positive charges in the intracellular vestibule of the CFTR chloride channel pore. J Gen Physiol 135: 229–245. doi:10.1085/jgp.200910327.

45. Qian F, Hiani El Y, Linsdell P (2011) Functional arrangement of the 12th transmembrane region in the CFTR chloride channel pore based on functional investigation of a cysteine-less CFTR variant. Pflugers Arch 462: 559–571. doi:10.1007/s00424-011-0998-2.

46. Chen EY, Bartlett MC, Loo TW, Clarke DM (2004) The DeltaF508 mutation disrupts packing of the transmembrane segments of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 279: 39620–39627. doi:10.1074/jbc.M407887200.

47. Wallin E, Tsukihara T, Yoshikawa S, Heijne GV, Elofsson A (1997) Architecture of helix bundle membrane proteins: An analysis of cytochrome c oxidase from bovine mitochondria. Protein Sci 6: 808–815. doi:10.1002/pro.5560060407.

48. Marcus Y (1997) Ion Properties. New York: Marcel Dekker. pp.

49. Linsdell P (2006) Mechanism of chloride permeation in the cystic fibrosis transmembrane conductance regulator chloride channel. Exp Physiol 91: 123–129. doi:10.1113/expphysiol.2005.031757.

50. Smith SS, Liu X, Zhang Z-R, Sun F, Kriewall TE, et al. (2001) CFTR: covalent and noncovalent modification suggests a role for fixed charges in anion conduction. J Gen Physiol 118: 407–431.

51. Gong X, Linsdell P (2003) Molecular determinants and role of an anion binding site in the external mouth of the CFTR chloride channel pore. J Physiol 549: 387–397. doi:10.1113/jphysiol.2002.038232.

52. Zhang Z-R, Cui G, Liu X, Song B, Dawson DC, et al. (2005) Determination of the functional unit of the cystic fibrosis transmembrane conductance regulator chloride channel. One polypeptide forms one pore. J Biol Chem 280: 458–468. doi:10.1074/jbc.M409626200.

53. Zhang Z-R, Song B, McCarty NA (2005) State-dependent chemical reactivity of an engineered cysteine reveals conformational changes in the outer vestibule of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 280: 41997–42003. doi:10.1074/jbc.M510242200.

54. Norimatsu Y, Ivetac A, Alexander CS, O'Donnell N, Frye L, et al. (2012) Locating a plausible binding site for an open-channel blocker, GlyH-101, in the pore of the cystic fibrosis transmembrane conductance regulator. Mol Pharmacol 82: 1042–1055. doi:10.1124/mol.112.080267.

55. Kogan I, Ramjeesingh M, Li C, Kidd JF, Wang Y, et al. (2003) CFTR directly mediates nucleotide-regulated glutathione flux. EMBO J 22: 1981–1989. doi:10.1093/emboj/cdg194.

CHAPTER 3

TARGETED MOLECULAR DYNAMICS SIMULATIONS OF CHANNEL OPENING IN CFTR

Abstract

We have employed a targeted molecular dynamics simulation using the C0-CFTR model as an initial structure and O-CFTR model as a target to investigate the conformational changes CFTR undergoes during channel opening. In our simulations, the most significant change during the initial stages of the transition involved the translation of the NBDs toward each other, and the concomitant contraction of the cytoplasmic vestibule. Such an NBD-initiated transition is consistent with the long-standing hypothesis that ATP binding at the NBDs initiates channel opening in CFTR. Snapshots of our simulation trajectory yield structures that bear close resemblance to the proposed intermediate states of CFTR during its gating cycle. Through our simulations, we have discovered a novel salt bridge between R334 and E217 predicted to break during channel opening which was confirmed experimentally using functional crosslinking. Notably, many of the amino acids that participate in other interactions predicted by our simulations are also involved in CF disease-associated mutations. These predictions should offer insight into the structural defects that lead to CFTR dysfunction in these mutants, and, ultimately, into mechanisms that could enable their correction.

Introduction

In the previous chapter, we built new homology models of the closed and open channel states of CFTR, which we believe constitute the most accurate representations of CFTR structure in its various functional states to date. Our model structures are highly consistent with the available experimental structure-function data on CFTR, and also demonstrate excellent predictive and explanatory potential by uncovering structural features that are highly plausible (in light of available experimental data) even though they were not explicitly constrained during modeling. Despite these advantages, however, we are bound to admit that these models represent mere "snapshots" of a large ensemble of possible structures that the CFTR protein likely adopts in its native environment. No biological system exists *in vacuo*; they are all subject to myriad interactions with their surrounding environment. A single CFTR molecule resides in a complex milieu of mixed aqueous (at its nucleotide-binding and intracellular and extracellular loop domains) and non-polar (at its transmembrane domains) environments with which it constantly interacts through kinetic, electrostatic, van der Waals, and hydrophobic forces. Thus, while our static models may represent accurate estimates of aggregate "consensus" states, they are, nevertheless, convenient fictions that do not capture the dynamics and ensemble behavior of the real molecule.

In addition, we note that our model for the closed state was modeled on a template (the crystal structure of murine P-glycoprotein [1]) that did not contain bound nucleotide at either of CFTR's two ATP binding sites (ABSs). The ABSs are composed of motifs from both the N-terminal and the C-terminal nucleotide-binding domains (NBD1 and NBD2, respectively), and formation of the ABSs is dependent upon NBD dimerization,

precipitated by ATP binding. ATP hydrolysis and release leads to de-dimerization, and the NBDs in this structure are consequently completely separated (and, therefore, also in our model of *C0-CFTR* based on this structure). However, unlike most other ABC transporters including P-gp, one of CFTR's ATP binding sites (ABS1) has impaired (perhaps abolished) catalytic activity. While this site is capable of binding to an ATP molecule, it cannot subsequently hydrolyze and release the nucleotide. Indeed, gating in CFTR is thought be controlled primarily by ATP binding and hydrolysis at the other binding site (ABS2), while ATP is retained at the degenerate ABS1 over several gating cycles [2]. The NBDs cannot, therefore, de-dimerize at this location, and CFTR is instead to occupy a "partial dimer" structure with ATP retained at ABS1 during the dominant closed channel phase of its gating cycle (C1). The fully de-dimerized state represented by our P-gp based *C0-CFTR* model is thought to be rarely visited under physiological conditions with ATP [3].

In order to investigate the dynamic nature of the protein molecule, as well as to visit intermediate states between the extremes represented by the models we developed in the previous chapter, we turn to the methods of molecular dynamics.

Molecular Dynamics

Molecular interactions and the resulting forces that drive their dynamics are accurately described by quantum mechanics and the solutions to Schrödinger's equation. If we know the positions and quantum states of every atom in a system, we can describe the forces acting on each atom and the evolution of the system over time. Due to the large number of parameters and the incredibly complex computations involved in such a calculation,

however, solving this problem using currently available computational tools is infeasible for large biological macromolecules.

One common approach to solving this problem for *N*-body systems is to use the molecular mechanics (MM) approximation. Several simplifying assumptions must be made in applying MM computation to a system. First, every atom is treated as a single spherical particle, with a volume typically corresponding to the Van der Waals radius for that atom. Secondly, atoms are assigned "partial charges" depending on their type and context, and electrostatic interactions between distant atoms are treated as Coulombic forces. Finally, the functional forms of bonded interactions are truncated to second-degree polynomials, allowing them to be treated as Hookean springs with pre-defined equilibrium bond lengths and spring constants. "Bonded interactions" typically include distance restraints between two bonded atoms, restraints on the angle subtended by three bonded atoms, and restraints on the dihedral angle between four atoms bonded in a chain. Parameters for each of these terms (e.g. partial charge values, spring lengths and force constants) may be empirically derived or fitted from first-principle quantum mechanical calculations on smaller molecular systems. There may exist additional interaction terms, added (often empirically) to improve the accuracy of the resulting "molecular mechanics force field", and the exact functional forms of all the interactions vary by implementation. For instance, one common MM potential that is popular for biomolecular simulations is the CHARMM22 force field [4], defined as:

$$V = \sum_{\text{bonds}} k_b \left(b - b_0\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{dihedrals}} k_\phi \left(\phi - \phi_0\right)^2$$
$$+ \sum_{\text{impropers}} k_\omega \left(\omega - \omega_0\right)^2 + \sum_{\text{Urey-Bradley}} k_u \left(u - u_0\right)^2 + \tag{3.1}$$
$$+ \sum_{\text{non-bonded}} \left\{ E \left[\left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\}$$

Given the position of every atom in a system, information about the atomic parameters (mass, charge, radius), and a graph of the bond connectivity between atoms, one can use the equation above (or its equivalent for some other force field) to calculate the conformation-dependent energy (termed the MM energy) for any given "snapshot" of a system.

Molecular dynamics (MD) takes this idea a step further, by postulating that the dynamic evolution of a system can be calculated by integration of Newton's laws of motion for a system governed by a molecular mechanics potential of the form given above. To wit, the above equation can be used to calculate the forces on each atom and, since we also know the assigned mass of each atom, we can calculate the acceleration of each atom using Newton's Second Law of Motion. Thus, given initial positions, velocities and masses, once can compute (deterministically) the motions of a molecule under a given MM potential.

The dynamic nature of the system necessitates numerical integration as a function of time. Since the potential energy terms in Equation 3.1 depend on the conformation of the system, and the conformation of the system, in turn, evolves in response to these energy terms, one must alternate between computing the forces on the atoms and their evolving positions and velocities. The characteristic "timestep" for this integration is governed by

the highest frequency motions within it. In atomistic simulations of molecules, the short period of the C—H bond vibrational typically limits MD timesteps to 1 or 2 femtoseconds.

Note that the MM potential in equation 3.1 is a linear superposition of several different force types, and that it can be extended to include user-defined forces on a system (e.g. the force of an AFM tip pushing on a molecule).

MD Simulations of CFTR

MD simulations have previously been used to study the dynamics of existing homology models of CFTR. The first such attempt was made by Alexander et al. [5], and they illustrated the utility of MD simulations of CFTR by demonstrating the formation of the R352-D993 salt bridge [6] which did not exist in their initial model over the course of a 5 ns MD simulation. These simulations were subsequently extended by Norimatsu et al [7].

The longest simulations of CFTR to date have been carried out by Furukawa-Hagiya et al. [8]. Unlike in previous attempts that ran MD on open channel CFTR models, they began their simulations with CFTR in the closed channel state. Through MD simulations, they tested the hypothesis that ATP binding at the nucleotide-binding domains causes conformational changes that propagate to the membrane-spanning domains (MSDs), ultimately resulting in channel opening. However, even when they extended their simulations out to 100 ns, they failed to detect significant changes in the conformation of the MSDs. This could have been due to poor modeling of the interactions between MSDs and NBDs responsible for transducing the conformational signal, a failure of the MM

potential to capture the motions of the molecule with sufficient accuracy, or simply because the simulations were not extended long enough. It must be noted, however, that a 100 ns simulation approaches our current limits with regards to feasible MD simulations on systems of this size. Furthermore, the accuracy of the MM potentials when used over longer timescales can also be called to question (particularly due to the increased accumulation of integration errors).

When deciding on MD simulation techniques, therefore, we realized the "free" MD simulations where no forces other than the internal molecular forces were allowed to act on the system would not be particularly informative over the timescales accessible to us. Fortunately, alternative techniques that employ user-defined forces to study particular transition pathways exist, and we have employed one of these approaches — targeted molecular dynamics (TMD) — in the studies that follow. In particular, we employed TMD to study the evolution of our *C0-CFTR* closed channel model towards the structure of the *O-CFTR* open channel structure in a simulation designed to mimic one possible route in a channel opening event.

Methods

Targeted Molecular Dynamics

System Set-up

In order to carry out a realistic simulation of a channel opening event, the closed state-protein model, *C0-CFTR*, was first embedded in an simulation box mimetic of its cellular environment.

First, the *C0-CFTR* model was put in the correct orientation by structurally aligning it with the P-gp template downloaded the Orientation of Proteins in Membranes [9] database. The membrane protein structures stored in this database are positioned such that they intersect an imaginary membrane bilayer centered on the *x*-*y* plane (with the positive *z*-axis pointing in the extracellular direction), using predictions based on the nature of the residues at the surface of the protein (e.g. there will be a high propensity for hydrophobic surface residues to be positioned near the hydrophobic core of the membrane bilayer). Next, a 120 Å×120 Å membrane patch consisting of 343 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) residues was generated in VMD [10]. This patch was generated by duplicating square patches of POPC residues arranged in a bi-layer with density and bilayer thickness set to match physiological values. Disorder was introduced into the patches by randomizing the orientations of the lipid molecules. The correctly-oriented *C0-CFTR* molecule was placed within the generated lipid bilayer, and lipid molecules within 0.6 Å of any protein atom were deleted, in addition to any lipid molecules within the protein cavity. The system was then solvated by the randomized

addition of water molecules. Due to the large number of water molecules in a typical biomolecular simulation, significant computational savings can be had by implementing a simplified model for each molecule. In these simulations, a 3-site water model for which the CHARMM27 potential was optimized called TIP3P [11] was used, and a total of 51,636 water molecules were added to the system. 146 K^+ atoms and 155 Cl^- atoms were also added to neutralize the net charge of the system and bring the ionic strength of the solution up to 150 mM. The full simulation box measured 120 Å × 120 Å × 165 Å and comprised 218,356 atoms (**Figure 3.1**).

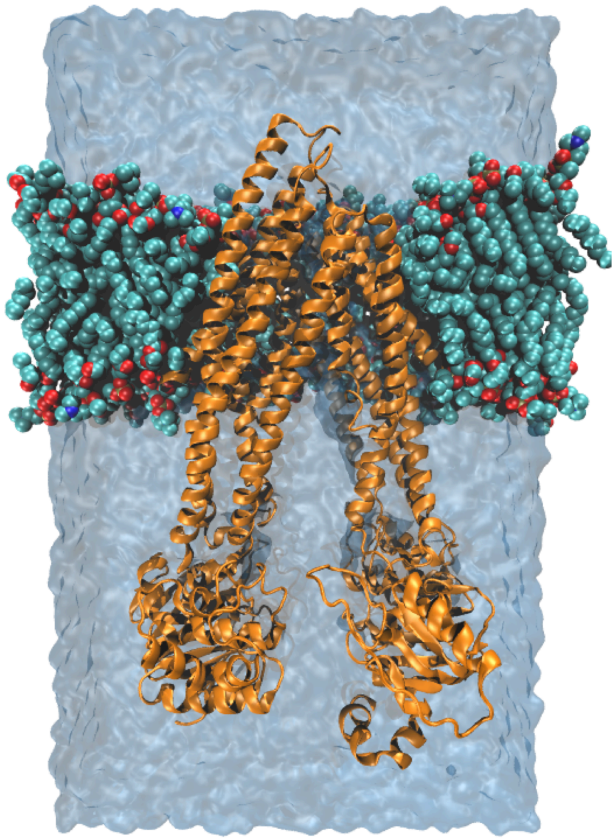

Figure 3.1: Snapshot of a unit simulation box

MD Simulations

NAMD 2.8 [12] was used for all simulations on this system, using the CHARMM27 force field [4] in a Langevin temperature and pressure controlled (NPT @ 300K) ensemble. The boundaries of the simulation box were set to be periodic (i.e. images of the unit cell extend infinitely in every dimension). Long-range electrostatics calculations in such a periodic system are easily computed using particle-mesh Ewald summation method. The SHAKE algorithm [13] was used to fix hydrogen bonds allowing the use of a 2 fs timestep.

First, 5000 steps of conjugate gradient energy minimization to remove steric clashes within the system as set up (i.e. atoms that were positioned within Van der Waals radii of each other). The lipid membrane tails in the original set-up are artificially straight (in order to facilitate protein insertion), and were first allowed to "melt" together to remove gaps between the generated patch and the embedded. In order to do this, all the atoms of the system except for the membrane hydrophobic tails were frozen, and then a short (1 ns) MD simulation was run in order for the membrane tails to equilibrate. This was followed by a 10 ns MD equilibration of the entire system with only the protein held restrained, carried out primarily in order to allow the surrounding solvent to equilibrate and arrange itself around the protein molecule. Finally, the entire system was allowed to equilibrate for a further 10 ns.

Targeted molecular dynamics (TMD) simulations were then carried out in order to transform the *C0-CFTR* state to the open channel *O-CFTR* state. In the method of

targeted molecular dynamics used here, a TMD force is added to MM potential (Equation 3.1) of the form:

$$V_{\mathrm{TMD}} = k_{\mathrm{TMD}}\left(RMSD(t) - RMSD*(t)\right)^2 \tag{3.2}$$

In order to apply this force, we first defined a set of target coordinates. This was provided by the coordinates of our *O-CFTR* model, which were aligned to the OPM structure of Sav1866 as described earlier for our *C0-CFTR* model. *RMSD*(*t*) in the above equation is then calculated to be the instantaneous best fit root-mean-square-displacement of the backbone C-α atoms at time *t* after the start of simulations from the target coordinates (the *O-CFTR* model). The *RMS**(*t*) term evolved linearly from the initial RMSD (*RMSD*(0), i.e. the RMSD between the *C0-CFTR* model and the *O-CFTR* model) to zero over the course of 10 ns. This added term, therefore, has the effect of reducing the RMSD between the protein and the target as the simulation progresses. The TMD force constant, k_{TMD}, was set to 5 kcal/(mol Å^2).

Snapshots of the simulation trajectory taken at 0 ns, 2.5 ns, 5 ns, 7.5 ns, and 10 ns were saved and their pores were analyzed along the axis normal to the lipid bilayer using the program HOLE [14]. Distances between interacting pairs of residues were measured using VMD [10].

Electrophysiology Experiments

Experimental validation of the models was carried out in the lab of Dr. McCarty at Emory, primarily by Dr. Guiying Cui. Experimental approaches were identical to those described in detail previously [6]. To wit, stage V-VI oocytes were isolated from

Xenopus laevis, under a protocol approved by the IACUC at Emory University, and injected with cRNA encoding wildtype human CFTR or mutants bearing cysteines engineered at position 334 and/or 217 on the wildtype background. Macroscopic currents were measured using an Axon Axoclamp 900A amplifier (Molecular Devices, Sunnyvale, CA) at a holding potential of –60 mV, and were activated upon stimulation of the co-expressed human beta2-adrenergic receptor by exposure to 10 μM isoproterenol added to the bath solution (ND96, containing in mM 96 NaCl, 2 KCl, 1 MgCl2, and 5 HEPES, pH 7.5). Cells were exposed to 1 mM 1,2-ethanediyl bismethanethiosulfonate (MTS-2-MTS; Toronto Research Chemicals, Canada) in the presence or absence of isoproterenol.

Results

Targeted Molecular Dynamics Simulations

We employed TMD simulations to generate a possible molecular trajectory of a channel opening event. Our initial simulation system consisted of the *C0-CFTR* homology model embedded in a POPC lipid bilayer membrane patch and solvated in a 150 mM KCl solution with explicit TIP3P water molecules. A biasing force was applied to the backbone C-α atoms to reach the *O-CFTR* target structure over a 10 ns simulation, arriving at a final RMSD between the simulation structure and *O-CFTR* of ~0.3 Å. Snapshots of the progressing simulation were taken at 0 ns, 2.5 ns, 5 ns, 7.5 ns, and 10 ns, and are further analyzed below (**Figure 3.2**). The PDB coordinate files for each of these snapshots can be found at: http://harvey.gatech.edu/cftr/PLoSONE_2013/.

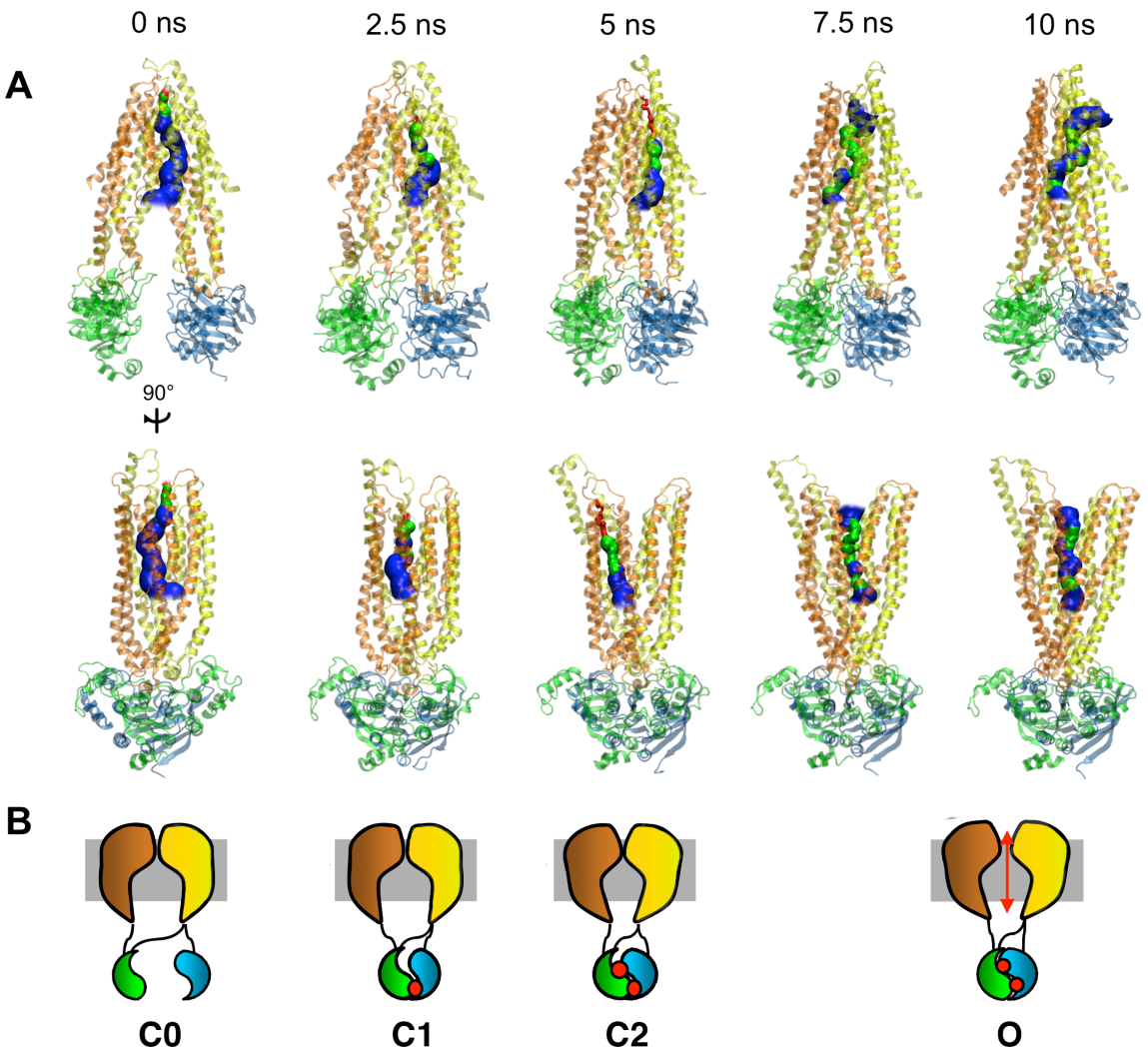

Figure 3.2: A) Snapshots along the TMD simulation, B) Corresponding states of CFTR's gating cycle

Visual inspection of the TMD trajectory reveals a progressing "conformational wave" during the opening transition (Figure 4), similar to that seen in nicotinic acetylcholine receptors during their transition between closed and open states [15]. This comprises three main stages:

1) pushing together of NBDs and contraction of the inner vestibule (C0 → C1),

2) NBDs forming a tight dimer and bending at the intracellular extensions of the TM helices that form the intracellular loops (ICLs) (C1 → C2), and
3) propagation of conformational change through TM helix rotation plus translation of the two transmembrane "wings" of the molecule — comprised of TM1 + TM2 + TM9 + TM10 + TM11 + TM12 and TM7 + TM8 + TM3 + TM4 + TM5 + TM6, respectively — away from each other in the direction perpendicular to the NBD dimer interface (C2 → O).

This sequence of events is further corroborated by analysis of the pore radius profiles for each of the snapshots taken during the trajectory. The dominant pore in each snapshot was analyzed using the program HOLE [14], and is depicted in **Figure 3.2** as a tube within the protein cavity, colored blue for regions with radius > 3 Å, green where the pore radius is between 3 Å and 1.8 Å, and red where the radius is < 1.8 Å, indicating that it is too narrow to allow chloride conduction. As noted, we see a contraction of the cellular cavity over the first half of the simulation due to NBD dimerization. During this period, the pore is never larger than a chloride ion indicating that all the structures generated until this stage correspond to closed CFTR channels. At 7.5 ns NBD dimerization is more or less complete, and we finally see unbending of the intra-cellular loops that leads to expansion of the transmembrane pore. At this stage we finally see a pore large enough to allow chloride conduction, indicating that this structure likely corresponds to the open CFTR channel. The conformation transition completes over the last quarter of the simulation with further expansion of the pore leading to the full conductance state at 10 ns.

This NBD-initiated transition pathway is consistent with the notion that ATP binding regulates channel gating by driving conformational changes in the transmembrane pore [16].

Note that we have chosen the labels for these snapshots (*C0*, *C1*, *C2* and *O*) to coincide with the nomenclature used to describe the various functional states of CFTR on purpose, since we believe that each of these structures best corresponds to these functional states. For instance, we believe that the *C1-CFTR* structure at 2.5 ns represents the C1 partial dimer state that CFTR primarily occupies as a closed channel. Due to the similarity of its asymmetric nucleotide binding domains to those proposed in CFTR, a recently published crystal structure of the bacterial ABC transporter TM287/288 [17] that has nucleotide bound in the degenerate catalytic site may serve as a useful template for the C1 partial dimer state of CFTR. A comparison of the NBDs from this structure to those in our snapshots confirms that the 2.5 ns snapshot of our TMD trajectory most closely resembles the purported partial dimer closed state of TM287/288 (**Figure 3.3**).

A

B

Snapshot	NBD RMSD from TM287/288 structure (Å)
0 ns	5.55
2.5 ns	4.17
5.0 ns	4.88
7.5 ns	5.04
10 ns	5.23

Figure 3.3: Comparison of NBDs from simulation and TM287/288 structure. A) Overlay of TM287/288 (red) and 2.5 ns snapshot (green). B) RMSD distance betwee TM287/288 and various snapshots.

Tracking Conformational Changes

In order to track the progression of the conformational wave as it travels from the NBDs through the ICLs to the transmembrane region, we charted the formation and breaking of the residue pair interactions shown in (**Figure 3.4A**) over the course of the simulation. In the NBDs, we tracked the distance between the side chains of residues R555 (NBD1) and T1246 (NBD2), which have been suggested to participate in a hydrogen bond interaction upon NBD dimerization [16]. In the TM helices, the formation of the R352-D993 [6,18] salt bridge was followed. In the ICLs, we discovered a novel salt bridge interaction between residues R258 and E282 in our simulations that exists in the closed state but is broken in the transition to the open channel as the passing conformational wave causes these residues to move apart.

Tracking these interactions over the course of the 10 ns simulation confirms the overall direction in which the transition progresses across the different domains of the protein (**Figure 3.4B**). The distance between R555 and T1246 decreases monotonically over the first 4 ns of the simulation as the NBDs are pushed together, and is then locked with the formation of a stable hydrogen bond at ~5 ns. The breaking of the closed-state R258-E282 salt bridge follows at ~6 ns, as the intracellular extensions of TM4 and TM5 slide past each other with the contraction of the cytoplasmic vestibule. The conformational wave then leads to rotations in TM6 and TM9, bringing residues R352 and D993 closer together until a stable salt-bridge is formed between them after ~7 ns.

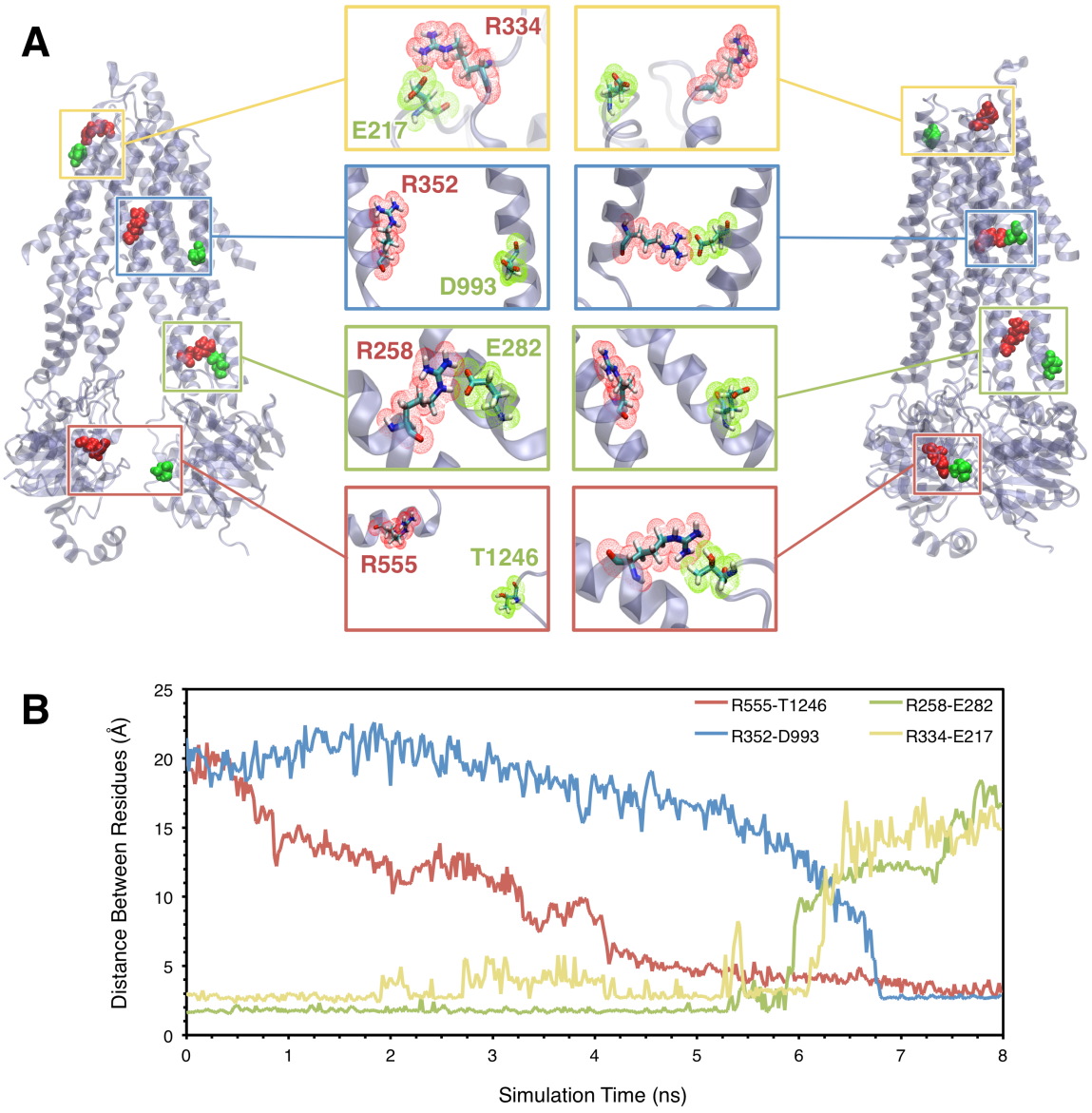

Figure 3.4: Tracking the conformational changes in CFTR channel opening by following particular side-chain interactions. A) Close-ups of tracked interactions. B) Tracking the formation and breaking of these interactions over the course of the simulation.

The pore radius profile at different points of the simulation provides another means of tracking the transition (**Figure 3.5**). Snapshots of the trajectory at 0, 2.5, 5, 7.5, and 10 ns were saved and the pore radius along the pore axis in each was determined using the program HOLE [14]. Over the first 5 ns, we noted a contraction of the inner vestibule, presumably as a result of the cytosolic domains being pushed together. During this period, the conduction pore radius is never larger than a chloride ion throughout its length, demonstrating that the channel is indeed closed when the NBDs are dissociated or in the early stages of dimer formation in our simulation. Subsequently, the formation of a tight NBD dimer results in conformational changes in the TM region leading to its expansion. A through-pore that is continuously larger than a chloride ion develops at 7.5 ns and is maintained through the end of the simulation, and is characterized by an inner vestibule and a large outer vestibule separated by a "narrow region" centered around S341 in TM6. Such a pore profile is in agreement with experimental data suggesting the existence of a narrow region and anion selectivity filter around S341 [18,19].

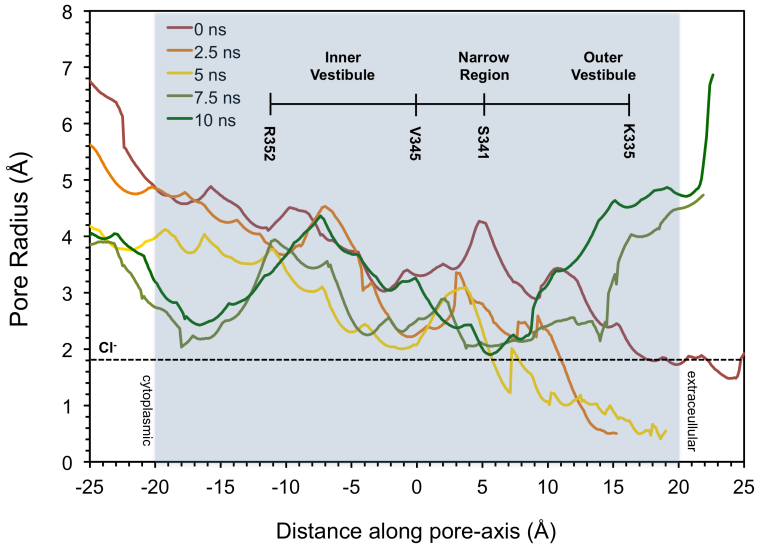

Figure 3.5: Pore radius profiles over the course of the simulation

Side Chain Interactions

As the channel opening transition sweeps through the structure, the local microenvironment around each residue could change as a result of backbone translation and rotation, exposing them to interactions with different neighboring residues over the course of the transition. In order to assess a subset of these interactions — those formed as a result of electrostatic attraction with basic residues — we analyzed all 104 arginine and lysine residues in our models for close contacts with neighboring negatively charged (aspartate or glutamate) and hydroxyl-containing (serine, threonine, tyrosine) amino acids for salt bridge and hydrogen-bonding interactions (**Table 3.1**). "Close contacts" were defined as being interactions in which atoms of each residue were within 5 Å of each other, and each interaction was classified as being closed-state (existing primarily over the first half of the channel opening TMD simulation), open-state (existing primarily in the latter half), or persistent interactions. Beside the previously established R347-D924 [20], R352-D993 [6], and R555-T1246 [16] interactions, our simulations also predict the existence of 89 additional interactions between residues in different domains that have not previously been reported. Interestingly, 86 of the residues involved in these interactions have been linked to CF-causing missense and deletion mutations [21]. We hypothesize that these interactions may be vital either to stabilizing a particular CFTR state or to the overall structural integrity of the protein.

Table 3.1: Side-chain interactions over the course of the simulation

SIDE-CHAIN INTERACTIONS OBSERVED DURING GATING IN TMD SIMULATIONS			
Closed State (Undimerized)	**Open State (Dimerized)**	**Comments**	**Clinically Relevant Mutations**
K95 **E1104**		Closed-state inter-TM salt bridge interaction	E1104K
R104 **E116**		Closed-state salt bridge interaction; may persist in open state	E116Q, E116K
K114 **D110** D112 **E115** **E116**		Interactions between cluster of neighboring acidic residues	D110H, D110Y, D110N, D110E, delE115, E116Q, E116K
	R117 D112	Open-state salt bridge interaction	R117C, R117G, R117H, R117P, R117L
(**R170** D173)		Weak interaction between neighboring residues	R170C, R170G, R170H
K190 **E193**		Closed-state salt bridge interaction between neighboring residues	E193K
K246 D249		Persistent salt bridge interaction between neighboring residues	
R248 **D985**	(**R248** D363)	Closed-state salt bridge broken as TM4 moves away from TM9 and towards TM6 during opening	R248T, D985Y, D985H
K254 **E292**		Closed-state inter-TM helix salt bridge interaction	E292K
R251 D984		Closed-state salt bridge broken as TM4 moves away from TM9 and towards TM6 during opening	
R258 **E282**		Closed-state salt bridge broken due to relative motion of TM4/TM5 extensions	R258G, E282D
K283 **E279**		Persistent salt bridge interaction between neighboring residues	E279D
R289 **E282** E286		Closed-state triangular salt bridge interaction may stabilize ICL2	E282D
	(K329 **D891**)	Weak open-state interaction may help hold ECLs together	D891G
R334 **E217**	(**R334** Y325)	Closed-state salt bridge interaction may help hold ECLs together; possible open state hydrogen bonding interaction	R334W, R334L, R334Q, E217G
K335 **D891**		Closed-state salt bridge interaction may help hold ECLs together	D891G
	R347 **D924**	Open-state inter-TM salt bridge interaction may stabilize pore structure	R347C, R347H, R347L, R347P, D924Y, D924N
(**R352** S1141)	**R352** **D993**	Weak closed-state inter-TM hydrogen bond interaction; Open-state inter-TM salt bridge interaction may stabilize pore structure	R352W, R352G, R352Q, D993Y, D993G
K411 **E407**		Persistent salt bridge interaction may stabilize NBD1	E407V
K413 E479		Persistent salt bridge interaction may stabilize NBD1	
(R419 **D443**)		Weak interaction may stabilize NBD1	D443Y
	K447 E395	Open-state salt bridge interaction may stabilize NBD1	
	K464 **T604** S605	Open-state hydrogen bond interactions may stabilize NBD1	K464N, T604I, T604S
R450 E391 D567		Persistent triangular salt bridge interaction may stabilize NBD1	
R487 D567 **D565**	R487 D567	Triangular salt bridge interaction may stabilize NBD1	D565G
R516 **D565**		Persistent salt bridge interaction may stabilize NBD1	R516G, D565G
R518 E514		Persistent salt bridge interaction may stabilize NBD1	
K522 **E527**		Persistent salt bridge interaction may stabilize NBD1	E527Q, E527G
(**R553** S495)		Weak interaction may stabilize NBD1	R553G, R553Q
R555 **D529**	**R555** **T1246** **D529**	Persistent salt bridge interaction may stabilize NBD1; Open-state inter-NBD hydrogen bond interaction	R555G, D529H, D529G, T1246I
K584 **E608** D648		Open state triangular salt bridge interaction may stabilize NBD1	E608G
K593 **E588**		Persistent salt bridge interaction may stabilize NBD1	E588V
(K598 **D565**)		Weak interaction may stabilize NBD1	D565G
(**R600** **D614**)		Weak interaction may stabilize NBD1	R600G, D614G, D614Y
K606 E583		Persistent salt bridge interaction may stabilize NBD1	
K611 **E608**		Persistent salt bridge interaction may stabilize NBD1	E608G
K612 **E588** **E608**		Persistent triangular salt bridge interaction may stabilize NBD1	E588V, E608G
K892 D110 E116		Closed-state triangular salt bridge interaction may help hold ECLs together	
	(R933 D993)	Weak open-state interaction	
K951 **E282** S262		Closed-state inter-TM helix interactions	E282D
	R975 E1046	Open-state inter-TM salt bridge interaction	
	K978 S1045 E1046	Open-state inter-TM interactions	
	K1041 **D1152**	Open-state inter-TM salt bridge interaction	D1152H
	R1048 E1044	Open-state salt brige interaction between neighboring residues	R1048G
R1066 E474		Persistent salt bridge interaction may stabilize ICL4-NBD1 interaction	R1066C, R1066S, R1066H, R1066L
R1070 E1075		Closed-state salt bridge interaction may stabilize ICL4	R1070W, R1070Q, R1070P
R1128 E1126		Persistent salt bridge interaction may stabilize NBD2	
(K1213 E1266)		Weak interaction may stabilize NBD2	
R1239 **E1401** **E1409**		Persistent triangular salt bridge interaction may stabilize NBD2	R1239S, E1401K, E1401A, E1401G, E1409K, E1409V
R1245 E1417 E1418	R1245 **E528** **D529**	Closed-state triangular salt bridge interaction may stabilize NBD2; Open-state triangular salt bridge interaction stabilize NBD1-NBD2 dimer	E528K, E428D, D529H, D529G
	(K1250 **D579**)	Weak open-state inter-NBD interaction	D579Y, D579A, D579G
(R1259 **E278**)		Weak closed-state interaction may stabilize ICL2-NBD2 interface	delE278
R1301 **D1305** **D1312**		Persistent triangular salt bridge interaction may stabilize NBD2	D1305E, D1312G
K1317 E1314		Persistent salt bridge interaction may stabilize NBD2	
R1325 D1320 E1329		Persistent triangular salt bridge interaction may stabilize NBD2	
K1334 E1329 D1336		Closed-state salt bridge interaction between neighboring residues in NBD2	
	R1358 S271	Open-state hydrogen bond interaction may stabilize ICL2-NBD2 interface	R1358S
	K1363 E1314	Open-state salt bridge interaction may stabilize NBD2	
(K1365 D1394)	(K1365 **D1270**)	Weak interactions may stabilize NBD2 in open and closed states	D1270N, D1270Y
R1385 **E1409** E1405		Persistent triangular salt bridge interaction may stabilize NBD2	E1409K, E1409V
(R1386 **E1321**)		Weak closed-state interaction may stabilize NBD1	E1321Q
K1389 **E1409**		Persistent salt bridge interaction may stabilize NBD2	E1409K, E1409V
R1403 E1405	R1403 E1405 S605	Persistent salt bridge interaction may stabilize NBD2; Open state inter-NBD hydrogen bond interaction	
K1420 E1418 **E1228**		Persistent triangular salt bridge interaction may stabilize NBD2	E1228K, E1228G
R1422 E1417		Persistent salt bridge interaction may stabilize NBD2	R1422W
K1429 D1425		Persistent salt bridge interaction may stabilize NBD2	

Experimental Support for the Models

One of the intraprotein interactions suggested by our modeling effort — which has not been indicated in previous models — is a salt bridge between residues R334 and E217 at the outer mouth of the pore (**Figure 3.6**). This salt bridge is seen in our simulations to be stable throughout the proposed C0, C1, and C2 closed states, finally breaking after approximately 6.3 ns of the simulation have elapsed — following the formation of the tight NBD dimer and changes in interactions in the ICLs — during the final transition to the open channel O state (**Figure 3.4B**).

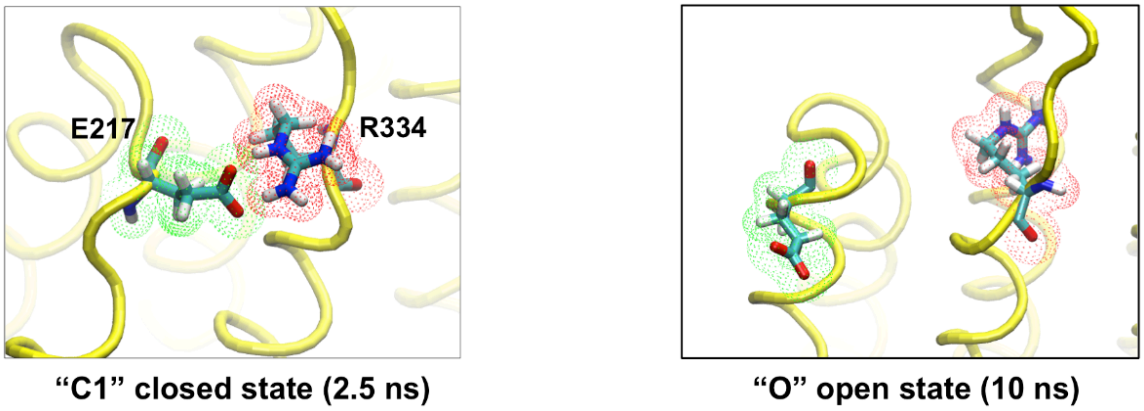

Figure 3.6: Closed-state E217-R334 salt bridge

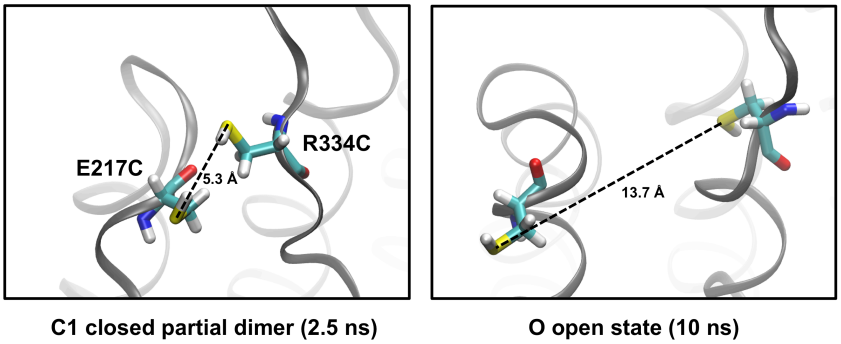

Figure 3.7: R334C-E217C double mutant

To provide experimental confirmation of the closed and open structures, and the transitions between the two, we asked whether cysteines engineered at these two positions in the double-mutant R334C/E217C-CFTR could be functionally crosslinked. Indeed, *in silico* mutations of positions 334 and 217 to cysteines retain the side chains of their amino acids in close proximity (**Figure 3.7**) We expressed the double mutant channels in *Xenopus* oocytes, and measured macroscopic currents.

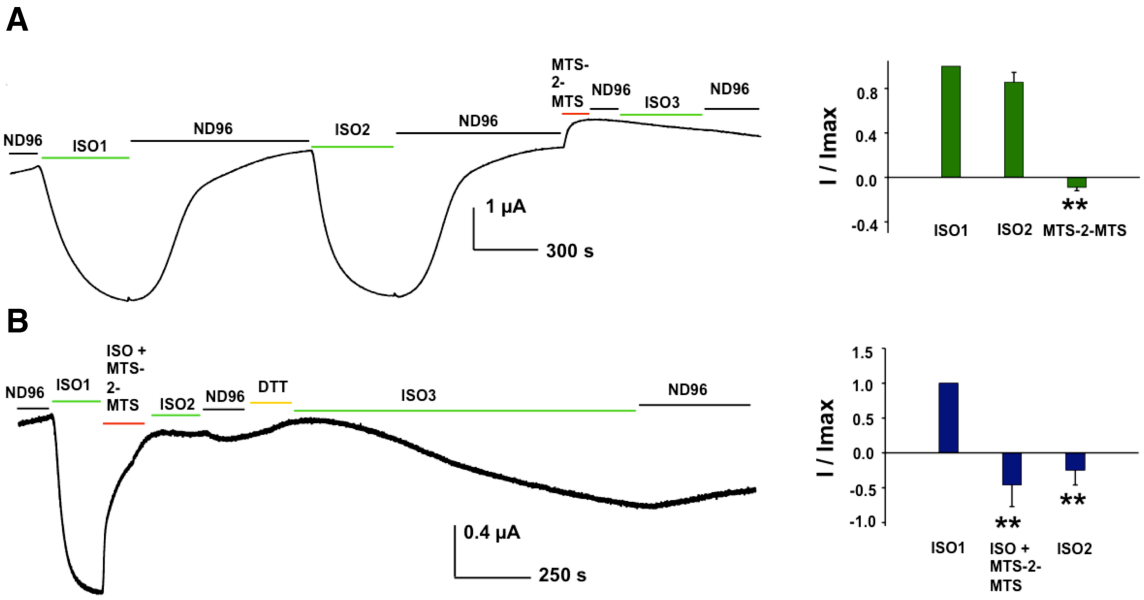

Figure 3.8: Crosslinking R334C to E217C locks CFTR channels into the closed state. (Data collected by Dr. Guiying Cui.)

The traces in (**Figure 3.8**) show that R334C/E217C-CFTR can repeatedly be activated by stimulation of the co-expressed beta2-adrenergic receptor using isoproterenol, without substantial decrement in peak current prior to exposure to the crosslinker MTS-2-MTS. When the same cell was exposed to MTS-2-MTS in the absence of isoproterenol (**Figure**

3.8A), when most of the channels should be closed, subsequent exposure to isoproterenol failed to activate CFTR channels to the same degree as prior to MTS-2-MTS; these results are consistent with the notion that R334C and E217C are positioned very near each other in the channel closed state. In the lower trace (**Figure 3.8B**), the cell was exposed to MTS-2-MTS in the continuing presence of isoproterenol, which resulted in a rapid decrease in current; after washout of isoproterenol and crosslinker, channels could not be reactivated by a second exposure to isoproterenol alone. When fully activated under physiological conditions, CFTR channels are only open ~40% of the time [22]. The loss of macroscopic current in this experiment, then, represents trapping of channels in the closed state by the MTS-2-MTS crosslinker. Subsequent exposure to DTT to break the disulfide bond(s) enabled activation upon re-application of isoproterenol. In control experiments, exposure to MTS-2-MTS did not have similar effects on the single mutants R334C-CFTR and E217C-CFTR, with respect to the ability to re-open channels after MTS-2-MTS exposure (**Figure 3.9**). Similarly, exposure of the double mutant to monofunctional MTS reagents did not impact the ability to activate CFTR channels upon subsequent reapplication of isoproterenol (**Figure 3.10**). These are consistent with the notion that R334 and E217 approach each other closely in the closed state of CFTR and that this salt bridge breaks as the channel opens, as predicted by our TMD simulations.

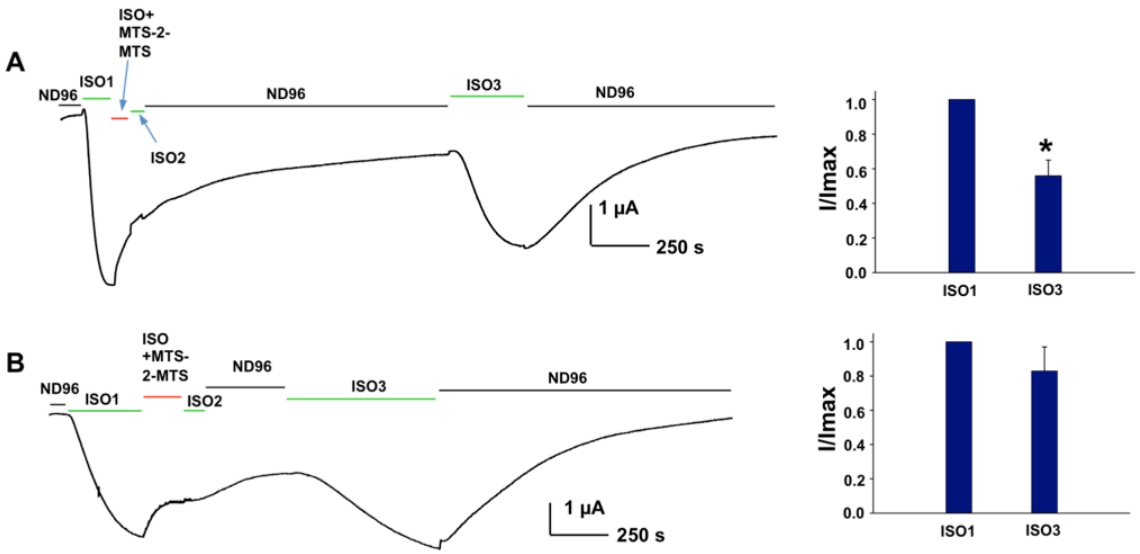

Figure 3.9: Effects of 1 mM MTS2-2MTS on R334C-CFTR and E217C-CFTR channels. Representative traces (left) and summary data (right) for macroscopic currents measured from R334C- (A) and E217C-CFTR (B) by two-electrode voltage clamp. Channels were activated by exposure of the oocyte to isoproterenol (ISO). ND96 = control bath solution. The bifunctional MTS reagent was not capable of covalently locking closed CFTR channels bearing a single cysteine. (Data collected by Dr. Guiying Cui.)

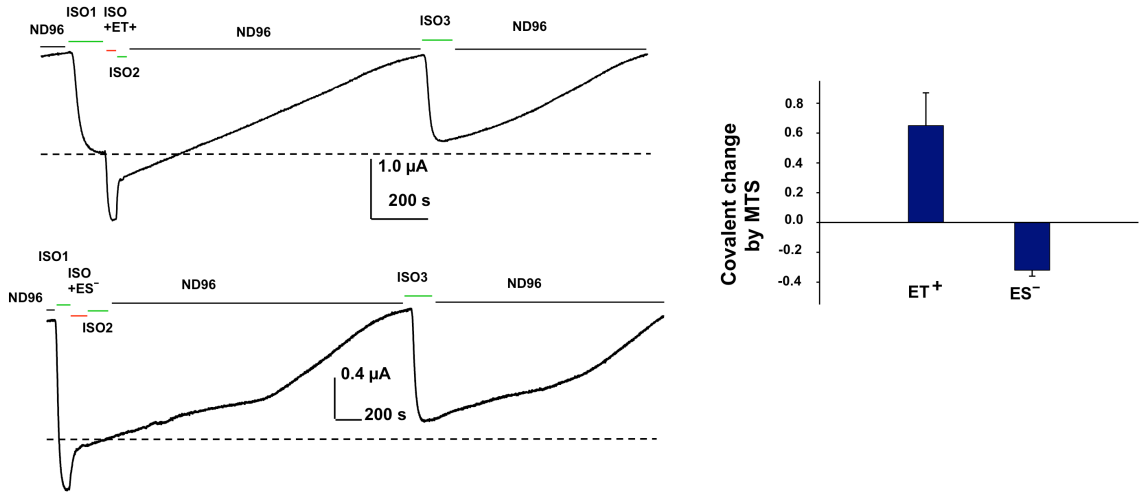

Figure 3.10: Effect of monofunctional MTS reagents $MTSET^+$ (ET^+) or $MTSES^-$ (ES^-) in the presence of isoproterenol (ISO). ND96 = control bath solution. Neither $MTSET^+$ nor $MTSES^-$ inhibited the subsequent activation by ISO. (Data collected by Dr. Guiying Cui.)

Discussion

Targeted Molecular Dynamics Simulations

Recent studies have employed MD simulations to sample the conformational space around particular CFTR homology models [5,7,8]. Simulating large-scale conformational transitions using conventional MD, however, is unfeasible in typical timescales. In a recent study by Furukawa-Hagiya et al. [8], a free MD simulation starting with a homology model of the CFTR closed state with MgATP bound in the NBDs did not reveal any significant widening of the transmembrane pore over 100 ns. In the present study, we used targeted molecular dynamics to generate a trajectory of one possible transition pathway between *C0-CFTR* and *O-CFTR*. Although the use of a biasing force in this technique prevents definite prediction of the true transition pathway, TMD has been successfully employed to study several other ABC proteins [23,24] and it can provide valuable information about events that may take place during the transition, including large-scale motions of backbone structures, and the changing interaction partners and surface accessibility of particular residues as the transition progresses.

It should be noted that, while the pattern of changes in interactions exhibit a general trend from the NBDs upward, the vertical position along the pore axis of a particular interacting pair in our simulation is not an absolute determinant of the order in which it changes. Thus, while the closed-state salt bridge pair R334-E217 is located in the extracellular loops — further away from the NBDs than any of the other pairs shown in **Figure 3.4** — it appears to break at 6.3 ns, before transmembrane R352-D993 salt bridge

is formed, indicating that at least partial freeing of the ECLs may be required for the helical rotations that lead to the transmembrane pore opening.

In our simulations the most significant change during the initial stages of the transition from C0 (0 ns) involves the translation of the NBDs toward each other and the concomitant contraction of the cytoplasmic vestibule below R352, as suggested by Figure 6. Such an NBD-initiated transition is consistent with the long-standing hypothesis that ATP binding at the NBDs initiates channel opening in CFTR [16]. After 2.5 ns, the NBDs come into close approach, but the R555-T1246 inter-NBD hydrogen bond has not yet formed, and the channel is closed to ion conduction (Figures 4, 5). Of the trajectory snapshots analyzed, the NBDs of the 2.5 ns snapshot have the lowest RMSD when superimposed onto the NBDs of the recently crystallized TM287/288 bacterial ABC transporter [17] which, like CFTR, has asymmetric NBDs containing one degenerate nucleotide binding site. This suggests that the 2.5 ns snapshot may approximate the C1 partial dimer state that CFTR primarily occupies when closed to ion conduction. Halfway through the trajectory, we observe a state where the NBDs are close enough for a hydrogen bond between R555 and T1246 to form, but the pore radius in the transmembrane region still remains too constricted to allow the passage of a chloride ion. This configuration may represent the ATP-bound full dimer closed channel transition state that CFTR is thought to adopt during its gating cycle (C2) [16]. A thermodynamic analysis by Csanády et al. [25] suggested that this transition state contains a high degree of conformational strain, and analysis of our simulation trajectory reveals kinks in the intracellular loop regions in this state that may contribute to this. Lending further support to the 5 ns structure as a model for the C2 closed transition state, we note that known

open-state interactions such as the R352-D993 salt bridge have not yet occurred at this stage (**Figure 3.4B**). In the latter half of the simulation, there is an expansion of the transmembrane pore region (**Figure 3.5**), culminating in a through-pore with radius continuously greater than that of a chloride ion (**Figure 3.2**). As the *O-CFTR* target state is approached, the structure contains inner and outer vestibules separated by a (possibly selectivity-conferring) narrow region. We note that this narrow, open region persists during the entire transition between the C1 and O states in our simulations, despite a widening of the pore in the outer vestibule. The stability of this region was demonstrated by Liu et al. [26] who showed that the constriction in this region to extracellularly applied reagents remains stable over temperatures from 22°C to 37°C, despite major conformational changes in the outer vestibule.

As the only ABC protein known to have channel-like properties, insights into CFTR structure are difficult to extract based solely on studies of the crystal structures of one state or another in other ABC transporters. Our simulations provide insights into the nature of the C1 partial dimer and C2 closed transition state and the large-scale motions involved in the transition between these various states.

Novel Predicted Side-Chain Interactions

One of the most promising aspects of our simulations is the ability to track the formation and dissociation of specific residue pair interactions as the backbone structures are transformed between the various channel states, and we report here possible contacts for all positively charged (arginine and lysine) residues over the course of the transition (**Table 3.1**). These predictions will guide experiments to further define the structure of

this unique and unusual protein, and, in this study, a novel salt bridge between R334 and E217 predicted to break during channel opening was confirmed experimentally using functional crosslinking (**Figure 3.8**); this result also helps explain the heretofore unclear role of R334 in CFTR function [27-29]. Notably, many of the amino acids that participate in these predicted interactions are also involved in CF disease-associated mutations. These predictions should offer insight into the structural defects that lead to CFTR dysfunction in these mutants, and, ultimately, into mechanisms that could enable their correction.

References

1. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722. doi:10.1126/science.1168750.

2. Basso C, Vergani P, Nairn AC, Gadsby DC (2003) Prolonged nonhydrolytic interaction of nucleotide with CFTR's NH2-terminal nucleotide binding domain and its role in channel gating. J Gen Physiol 122: 333–348. doi:10.1085/jgp.200308798.

3. Szollosi A, Muallem DR, Csanády L, Vergani P (2011) Mutant cycles at CFTR's non-canonical ATP-binding site support little interface separation during gating. J Gen Physiol 137: 549–562. doi:10.1085/jgp.201110608.

4. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4: 187–217. doi:10.1002/jcc.540040211.

5. Alexander CS, Ivetac A, Liu X, Norimatsu Y, Serrano JR, et al. (2009) Cystic fibrosis transmembrane conductance regulator: using differential reactivity toward channel-permeant and channel-impermeant thiol-reactive probes to test a molecular model for the pore. Biochemistry 48: 10078–10088. doi:10.1021/bi901314c.

6. Cui G, Zhang Z-R, O'Brien ARW, Song B, McCarty NA (2008) Mutations at arginine 352 alter the pore architecture of CFTR. J Membr Biol 222: 91–106. doi:10.1007/s00232-008-9105-9.

7. Norimatsu Y, Ivetac A, Alexander CS, Kirkham J, O'Donnell N, et al. (2012) Cystic fibrosis transmembrane conductance regulator: a molecular model defines the architecture of the anion conduction path and locates a “bottleneck” in the pore. Biochemistry 51: 2199–2212. doi:10.1021/bi201888a.

8. Furukawa-Hagiya T, Furuta T, Chiba S, Sohma Y, Sakurai M (2013) The power stroke driven by ATP binding in CFTR as studied by molecular dynamics simulations. J Phys Chem B 117: 83–93. doi:10.1021/jp308315w.

9. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22: 623–625. doi:10.1093/bioinformatics/btk023.

10. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–38, 27–28.

11. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79: 926–935. doi:10.1063/1.445869.

12. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781–1802. doi:10.1002/jcc.20289.

13. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 23: 327–341. doi:10.1016/0021-9991(77)90098-5.

14. Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MSP (1996) HOLE: A program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 14: 354–360. doi:10.1016/S0263-7855(97)00009-X.

15. Grosman C, Zhou M, Auerbach A (2000) Mapping the conformational wave of acetylcholine receptor channel gating. Nature 403: 773–776.

16. Vergani P, Lockless SW, Nairn AC, Gadsby DC (2005) CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. Nature 433: 876–880. doi:10.1038/nature03313.

17. Hohl M, Briand C, Grütter MG, Seeger MA (2012) Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 19: 395–402. doi:10.1038/nsmb.2267.

18. Cui G, Song B, Turki HW, McCarty NA (2012) Differential contribution of TM6 and TM12 to the pore of CFTR identified by three sulfonylurea-based blockers. Pflugers Arch 463: 405–418. doi:10.1007/s00424-011-1035-1.

19. McCarty NA, Zhang Z-R (2001) Identification of a region of strong discrimination in the pore of CFTR. Am J Physiol Lung Cell Mol Physiol 281: L852–L867.

20. Cotten JF, Welsh MJ (1999) Cystic fibrosis-associated mutations at arginine 347 alter the pore architecture of CFTR. Evidence for disruption of a salt bridge. J Biol Chem 274: 5429–5435.

21. Cystic Fibrosis Mutation Database (n.d.) Cystic Fibrosis Mutation Database. Available:http://www.genet.sickkids.on.ca/cftr/app. Accessed 7 December 2012.

22. Fuller MD, Zhang Z-R, Cui G, McCarty NA (2005) The block of CFTR by scorpion venom is state-dependent. Biophys J 89: 3960–3975. doi:10.1529/biophysj.105.060731.

23. Weng J-W, Fan K-N, Wang W-N (2010) The conformational transition pathway of ATP binding cassette transporter MsbA revealed by atomistic simulations. J Biol Chem 285: 3053–3063. doi:10.1074/jbc.M109.056432.

24. Weng J, Fan KN, Wang W (2012) The conformational transition pathways of ATP-binding cassette transporter BtuCD revealed by targeted molecular dynamics simulation. PLoS One 7: e30465. doi:10.1371/journal.pone.0030465.

25. Csanády L, Nairn AC, Gadsby DC (2006) Thermodynamics of CFTR channel gating: a spreading conformational change initiates an irreversible gating cycle. J Gen Physiol 128: 523–533. doi:10.1085/jgp.200609558.

26. Liu X, Zhang Z-R, Fuller MD, Billingsley J, McCarty NA, et al. (2004) CFTR: a cysteine at position 338 in TM6 senses a positive electrostatic potential in the pore. Biophys J 87: 3826–3841. doi:10.1529/biophysj.104.050534.

27. Gong X, Linsdell P (2003) Molecular determinants and role of an anion binding site in the external mouth of the CFTR chloride channel pore. J Physiol 549: 387–397. doi:10.1113/jphysiol.2002.038232.

28. Zhang Z-R, Song B, McCarty NA (2005) State-dependent chemical reactivity of an engineered cysteine reveals conformational changes in the outer vestibule of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 280: 41997–42003. doi:10.1074/jbc.M510242200.

29. Norimatsu Y, Ivetac A, Alexander CS, O'Donnell N, Frye L, et al. (2012) Locating a plausible binding site for an open-channel blocker, GlyH-101, in the pore of the cystic fibrosis transmembrane conductance regulator. Mol Pharmacol 82: 1042–1055. doi:10.1124/mol.112.080267.

CHAPTER 4

CFTR-SPECIFIC SUBSTITUTIONS IN THE NBDS AND THEIR EFFECTS ON STRUCTURE AND FUNCTION

ABSTRACT

CFTR is the only ABC protein known to function primarily as a channel protein, rather than as an alternating access transporter. The question of *how* it altered its primary function, however, remains unsolved. In these studies, we focused our attention on the nucleotide-binding domains, searching for substitutions unique to CFTR that may have led to gain of channel function. Leveraging the high sequence homology between CFTR and its ABC-C homologs in the NBDs, we employed bioinformatics techniques to identify particular sites of interest that appear to have been positively selected in the CFTR lineage alone. Comparative molecular modeling and simulations of CFTR and Sav1866 NBDs based on the partial dimer template of the TM287/288 crystal structure revealed that these sites participate in important interactions involved in binding to and hydrolyzing ATP, as wells as in transmitting the conformational changes from the NBDs to the other domains of the protein. These interactions are significantly altered in CFTR's NBDs, leading to the hypothesis that loss of ATPase function may have contributed to gain of channel function in CFTR. These studies provide an important advance in our understanding of CFTR's unique role as an ion channel, and how it came to acquire it.

INTRODUCTION

Our studies of CFTR structure through homology modeling and targeted molecular dynamics in the previous chapters have provided insight into the structure and function of this gated ion channel. We have based these investigations on the homology of CFTR with related proteins in the ATP Binding Cassete (ABC) transporter superfamily. As we noted earlier, however, CFTR differs in several important ways from the rest of the ABC proteins.

The Structural Basis of Altered Function in CFTR

Perhaps the most significant difference between the CFTR and the other ABCs lies in its primary function as a chloride ion channel. Most other ABC transporters use the energy from ATP binding and hydrolysis at their nucleotide binding domains (NBDs) to pump substrates into (in the case of importers) or out of (exporters) the cell. How might CFTR differ in order to accommodate its altered primary function?

Membrane-Spanning Domains

First of all, we may speculate that the transmembrane regions of CFTR and other ABC transporters are likely to be very different. Even the most closely related homologs of CFTR — the members of the ABC-C subfamily [1] — are involved in the transport of large hydrophobic molecules. CFTR, on the other hand, serves as a conduit for the passive transit of small anions (chlorides in particular) across the cell membrane. One may therefore expect the structure and biochemical nature of the CFTR pore to be substantially different from its ABC-C cousins in order to accommodate its very different

substrate. Even more importantly, the fact that CFTR functions as a *channel* rather than an alternating-access *transporter* necessitates the existence of an "open state" that is open at both ends of the transmembrane segment of the protein. This feature — antithetical to the function of other ABC proteins — is vital for any viable model of CFTR. However, since all the homology models of CFTR created so far — including our own O-CFTR model described earlier — are based on templates of ABC proteins with distinct inward-facing and outward-facing states (and no "open" state) no modeling strategy has yet successfully yielded a CFTR structure with a truly open pore. This problem is further exacerbated by the poor sequence similarity between CFTR and other ABC proteins (particularly the ones that have so far been crystallized) in the membrane-spanning domains (MSDs), which prevents strong inference of CFTR structure and function in these domains based on these templates. While we were successful in our modeling and simulation efforts described in the previous chapters — due primarily to the large body of experimental data incorporated into our modeling in addition to the controlled nature of the simulations conducted — extending our methodology in order to gain insight into the differences between CFTR and its related ABC transporters would be highly speculative.

R-domain

The R-domain is another anomaly that is unique to CFTR. An analysis by Sebastian et al. [2] revealed that the R-domain sequence is not found in any other protein, and most likely originated through *de novo* expression of previously non-coding sequence beginning ~550–650 million years ago. The R-domain is thought to play an important role in gating the CFTR channel through the phosphorylation of conserved serine residues [3], but it remains to be seen whether the introduction of this novel domain was a priming mutation

for CFTR to evolve channel activity. Studies on the R-domain, however, suffer from the fact that it is largely unstructured, leaving researchers without a framework within which to propose and investigate mechanistic hypotheses about its function. We did not previously include the R-domain in our modeling for this reason, a choice that may be further justified by the fact that CFTR variants lacking the R-domain remain largely functional [4].

Nucleotide-Binding Domains

This leaves the nucleotide binding domains (NBDs) as potential sites for the investigation of differences between CFTR and transporter sequence and structure that may potentially have led to the acquisition of novel channel function in the former.

Unlike the other domains, the NBDs of CFTR show relatively high homology with their counterparts in related ABC proteins. This is consistent with the role of the NBDs: whereas the MSDs must evolve to accommodate the wide range of substrates associated with the various members of the ABC family, all of their NBDs must retain the ability to 1) bind ATP, 2) hydrolyze ATP, and 3) couple the free energy change of ATP binding / hydrolysis to the protein's function as a transporter (or, in the case of CFTR, a channel). The invariant nature of these functions may logically be assumed to give rise to similar structures and sequence of these domains across all the ABCs, and publication of high-resolution crystal structures of the NBDs of several ABC transporters [5-13] and of NBD1 of CFTR [14-16] have confirmed this expectation.

Given its very different functional demands, however, one may still expect to see some divergence in CFTR sequence and structure, even in the highly conserved NBDs. Firstly,

an NBD dimer contains two potential ATP-binding sites, and a full transporter cycle is thought to involve binding and hydrolysis at both these sites. While this may indeed be required for the energetically demanding task of transporting large molecules against their concentration gradients, it seems unnecessarily expensive for the comparatively simple task of opening and closing a channel gate in CFTR. Secondly, if CFTR has indeed evolved from a defunct exporter [1], it must have incorporated changes that allowed it to slow the rate of ATP hydrolysis in order to allow the open channel state to persist for a significant span of time. In fact, to this end, CFTR is one of a small subset of ABC transporters for which one of the ATP-binding sites (ABS1) is thought to be incapable of ATP hydrolysis [17]. Finally, given that the NBDs interact with the intracellular loops (ICLs) of CFTR to transmit the conformational wave initiated by NBD dimerization / disassociation, it is not unreasonable to ask whether changes at the interface between these domains could also have contributed to altered function in CFTR. Indeed, we note with emphasis that the most common CF-causing CFTR mutation — ΔF508 — is the deletion of a residue in NBD1 located at this very transmission interface.

In other words, if CFTR did indeed evolve from a degraded ABC transporter, then the gain of channel activity function could very well be due to loss-of-function mutations in the NBDs that led to impaired ATP binding / hydrolysis and altered transmission of the ATP-driven power stroke from the NBDs to the ICLs.

Conserved Motifs in the Nucleotide Binding Domains of ABC Transporters

While widely varied in the structures and functions of their membrane integral domains, the NBDs of all ABC transporters share several common features. In order to understand

how changes in these conserved regions may have led to channel function in CFTR, we must first familiarize ourselves with some of the more important of these motifs. The location of these motifs in CFTR and Sav1866 sequence are indicated in **Table 4.1**.

Table 4.1: Conserved motifs in NBDs of CFTR and Sav1866

Motif	CFTR ABS1	CFTR ABS2	Sav1866
Walker A	G458 to S466	G1244 to T1252	G374 to T382
Signature	L548 to Q552	L1346 to H1350	L478 to Q482
Q-Loop	C491 to S495	I1289 to V1293	V420 to N424
X-Loop	I539 to G545	F1337 to G1343	T469 to G475
Walker B	L568 to S573	I1366 to E1371	I498 to E5503
H-Loop	V603 to M607	C1400 to I1404	V532 to L536

Walker A

The Walker A motif (also known as the P-loop or phosphate-binding loop) has the sequence pattern GXXXXGK(T/S), where X is any residue and the other letters represent the usual 1-letter code for amino acids. This motif — first identified in the F1-ATPase

[18] — forms a loop that wraps around the triphosphate of an ATP of GTP molecule, binding the nucleotide to the protein. Particularly important are the terminal lysine and hydroxyl-containing (serine or threonine) residues, which coordinate the γ- and β-phosphate groups of the nucleotide.

LSGGQ / ABC Signature Motif

The highly conserved and unique-to-ABC signature motif is the defining characteristic of ABC transporters. Also called the 'C-loop', it contains the important LSGGQ sequence responsible for dimer formation in ABC NBDs [19]: the signature motif of NBD2 binds to the ATP molecule bound to the Walker A motif of NBD1, and vice-versa. The resulting "composite ATP binding site" (ABS) is usually numbered after the NBD contributing the Walker motifs (in the case of the example just mentioned, it would be designated ABS1). Dimer formation mediated by the signature motif is essential for the subsequent dephosphorylation reaction [20].

X-loop & Q-loop

The newly-identified X-loop — located just upstream of the LSGGQ signature motif — has the consensus sequence TEVGERG and is found only in ABC exporters (as well as in CFTR) [21-23]. The Q-loop is a variable span of 4-5 residues following a highly conserved glutamine residue which coordinates the ATP γ-phosphate and magnesium ion at the ABS. Large temperature factors associated with this region in various NBD crystal structures suggest that this loop is highly flexible.

The Q-loop the X-loop are together thought to comprise the primary coupling interface between the NBDs and the rest of the protein through interactions with the intracellular loops.

In one model of the transmission mechanism, ATP binding and hydrolysis in the ABS is sensed by the Q-loop glutamine which then triggers concerted movements in the Q-loop and X-loop [24,25]. These loops, in turn, are coupled to ICLs through which the conformational wave is propagated to the rest of the domains of the protein [21,22,26].

Walker B

There is considerable variability in the Walker B motif among ABC proteins, but the general consensus is that it comprises a β-strand containing four hydrophobic amino acids followed by an aspartate and a glutamate (commonly represented as hhhhDE) [18,27]. Mg^{2+} is often found to be an essential co-factor for ATP binding and hydrolysis and, in ABC proteins, and the Walker B aspartate is thought to co-ordinate these ions. The glutamate residue at the end of the Walker B sequence acts as a catalytic base for the ATP hydrolysis reaction [28-31]. In a subset of ABC transporters — including the ABC-C subfamily of which CFTR is a member — substitution of the catalytic glutamate to an aspartate is thought to impair ATP hydrolysis at one of the ABSs.

H-loop / Switch Region

The H-loop contains an important histidine residue that coordinates the γ-phosphate of the bound ATP and, in tandem with the Walker B glutamate, forms part of the "catalytic dyad" responsible for ATP hydrolysis [8]. The central role of this histidine in organizing

the complex network of hydrogen bonding contacts in the active site has led to it being termed the "linchpin" of ATP hydrolysis in some ABC transporters [30].

D-loop

Closely following the Walker B region, the D-loop has a consensus sequence of SALD. Its exact role remains controversial, although it appears to be involved in both in communicating between the subunits of an NBD dimer [32] and for coordinating attacking water molecules at the ATPase active site [33].

In the following, we will investigate structure/function relationships in the NBDs, employing bioinformatics, structural modeling and simulation techniques in order to understand the differences between CFTR and its related ABC proteins in these important domains.

METHODS

Relative Divergence Calculations in the NBDs

In order to identify sites at which CFTR diverges significantly from its ABC homologs, we first generated a multiple sequence alignment of human CFTR protein sequence against the 11 other human ABC-C proteins. The sequences were retrieved through a BLAST search against the RefSeq database (**Table 4.2**), and a multiple sequence alignment was generated using MUSCLE [34]. In order to filter out apparently divergent sites that were not conserved within the CFTRs of other related organisms, a multiple

sequence alignment of all available CFTR sequences from the vertebrate subphylum was also similarly generated.

Table 4.2: Human ABC-C sequences used to calculate relative divergence

NAME	Accession No.
ABCC7 (CFTR)	NP_000483 / NM_000492
ABCC1 (MRP1)	NP_004987 / NM_004996
ABCC2 (MRP2)	NP_000383 / NM_000392
ABCC3 (MRP3)	NP_003777 / NM_003786
ABCC4 (MRP4)	NP_005836 / NM_005845
ABCC5 (MRP5)	NP_005679 / NM_005688
ABCC6 (MRP6)	NP_001162 / NM_001171
ABCC8 (SUR1)	NP_000343 / NM_000352
ABCC9A (SUR2A)	NP_005682 / NM_005691
ABCC10 (MRP7)	NP_001185863 / NM_001198934
ABCC11 (MRP8)	NP_115972 / NM_032583
ABCC12 (MRP9)	NP_150229 / NM_033226

Residues in CFTR may diverge from the ABC consensus sequence simply due to generally high rates of evolution at that site across all ABCs (if, for instance, the identity of a residue at a particular site has little functional consequence). In order to discriminate

between sites with generally high mutation rates from those which showed divergence particularly in CFTR, we defined the following metric — which we have termed "Relative Divergence" — and calculated it for every CFTR residue (numbered k):

$$RD_k = \frac{\text{Mean pairwise distance between CFTR and other ABC-Cs at site } k}{\text{Mean pairwise distance between all ABC-Cs except CFTR at site } k}$$

$$= \frac{\frac{1}{n_k}\sum_{i=1}^{n_k} S\left(r_{k,\text{CFTR}}, r_{k,i}\right) + \mu}{\frac{2}{n_k\left(n_k - 1\right)}\sum_{i=1}^{n_k}\sum_{j=i+1}^{n_k} S\left(r_{k,i}, r_{k,j}\right) + \mu}$$

where:

- n_k is the number of ABC-C sequences with residues aligned to CFTR residue k,
- $r_{k,i}$ is the identity of the residue in sequence i ($i \in [1, n_k]$) aligned to CFTR site k,
- $S(x, y)$ is a symmetric amino acid substitution function that yields a measure of distance (i.e. dissimilarity) between residues x and y, and
- μ is a pseudocount introduced to normalize the results.

In our analysis, we used — for S — the Grantham substitution matrix [35], since it provides a measure of biochemical dissimilarity and we wanted to highlight residues that have evolved away in CFTR to acquire significantly different biochemical properties from other ABC transporters. We used a pseudocount μ =10 as this empirically proved to normalize our data most effectively.

Homology Modeling of CFTR NBDs

In order to model the NBDs of CFTR with its degenerate ATP binding site, we used the recently published crystal structure of the bacterial exporter TM287/288 (PDB: 3QF4) [32] as a template. The sequence of TM287/288 was extracted from the PDB file and added to the sequences used in the analysis above to generate another multiple sequence alignment in MUSCLE [34]. The pairwise alignment of TM287/288 and CFTR was extracted from the MSA and used as input for homology modeling.

NBD1 of CFTR contains a unique 32-reside segment (amino acids 404-435) not found in other ABCs, often termed the "regulatory insertion". This segment was modeled using a loop search in SYBYL-X (Tripos, Inc.) against its sequence. The remainder of NBD1 and NBD2 were separately modeled in Modeller v10 [36] using the alignments generated above. 10 models for each NBD were generated and the ones with the lowest molpdf scores (a measure of energetic stability and similarity to the template structure [37]) were retained. NBD1 (and the regulatory insertion structure) and NBD2 were then assembled into a single structure and any clashes were eliminated using an MD simulated annealing protocol in Modeller.

Two separate starting structures were then generated based on the NBD dimer generated above: one with MgATP bound only at the degenerate ATP-binding site (CFTR-ATP/apo), and one with MgATP at both ABSs (CFTR-ATP/ATP). The TM287/288 template contained only a magnesium ion and adenyl imidodiphosphate (AMP-PNP, a non-hydrolyzable ATP analogue) at the degenerate ABS1. The template was first aligned to the CFTR model by fitting TM287/288 sequence to CFTR, followed by structural

alignment carried out by minimizing RMSD in PyMol (Schrödinger, Inc.). Following this alignment, the Mg^{2+} ion was placed at the same coordinates in the CFTR model as in the template, and an ATP molecule was constructed at ABS1 based on the position of the AMP-PNP molecule in the template (by replacing the N between the β- and γ-phosphates with an O). The second starting structure — with both ABSs occupied — was created using the same procedure for ABS1, but then also aligning the crystal structure of CFTR NBD1 with bound ATP [14] (PDB: 1R0X) to NBD2 of our model and using the ligand coordinates from this structure to construct MgATP at ABS2 of the model.

Modeling of Sav1866 NBDs

To contrast with the CFTR structures generated above, we also constructed models of the NBDs of the bacterial ABC transporter Sav1866 in a "partial dimer" structure with both ABSs occupied (Sav-ATP/ATP). Coordinates of the nucleotide binding domains from the crystal structure of Sav1866 in a tight dimer configuration with bound Na^+ AMP-PNP (a non-hydrolyzable ATP analogue) [38] (PDB: 2ONJ) were first extracted from the rest of the protein. The Sav1866 NBDs were then aligned to those in the structure of TM287/288 [32] (PDB:3QF4) using PyMol (Schrödinger, Inc.). Magnesium ions were substituted for the Na^+ ions in the structures and the AMP-PNP coordinates were used to construct ATP molecules at both ABSs.

MD Simulations

In order to guard against the sampling problems inherent in simulations, nine separate molecular dynamics simulations of the three starting structures (CFTR-ATP/apo, CFTR-

ATP/ATP, Sav-ATP/ATP) were carried out, totaling nearly 0.5 μs of total MD simulations.

In every case, the starting system was generated by first converting the protein models to all-hydrogen structures (guessing the coordinate of the hydrogen atoms which do not appear in the crystal structure based on pre-defined topologies), then solvating in TIP3P water boxes with at least 15 Å between the boundary of the box and the proteins. K^+ and Cl^- ions were added to each simulation system in order to neutralize the net charge and bring the solution potential to 150 mM. System construction was carried out in VMD [39].

All MD simulations were carried out in NAMD v2.8 [40] using the CHARMM27 force-field topology definitions and parameters [41]. Steric clashes were first eliminated through 50,000 steps of conjugate gradient minimization while holding the protein and MgATP restrained through harmonic restraints. A 5 ns MD simulation was then carried out — again with protein and MgATP held restrained — to equilibrate the solvent. This was followed by nine production runs: two of length 25 ns and one 100 ns-long for each of the three systems. All simulations were carried out in NPT (constant temperature / constant pressure) conditions using Langevin temperature and pressure controls set at 310K and 1atm. The SHAKE algorithm [42] was used to fix hydrogen bond lengths, allowing the use of a 2 fs timestep.

RESULTS

Identification of Divergent Sites in CFTR NBD1

We investigated the differences in the primary structure of CFTR NBDs that may have led to the altered function of the protein by analyzing the sequences of human CFTR and the other 11 members of its closest relatives in the human ABC-C subfamily [1].

In order to tease apart the effects of background mutation (*versus* positive selection for particular traits in CFTR specifically), we computed a metric that we have termed *Relative Divergence* (*RD*) for each site in CFTR NBD1 (see Methods for more details). The rate of evolution at a particular site can be influenced by a variety of factors. If, for instance, a particular residue is located in a region of the protein that is not important for its function, mutations at that site will be more highly tolerated than in function-critical sites across all transporters. On the other hand, sites that are conserved everywhere including CFTR may be vitally important for functions common to all ABCs, but would not yield any information about the lineage-specific evolution of channel activity. *RD* provides a relatively simple measure — a single number — that contains both these overcomes both these pitfalls. Roughly, RD is a measure of the biochemical dissimilarity between CFTR and other transporters at a particular residue normalized to the general degree of variance across all ABC-Cs at that residue. A high value of RD (> 3) signifies that CFTR is significantly different from other ABCCs at a particular site, and that the difference is probably not just due to general variability at that position (i.e., it is conserved in other ABCCs, but divergent in CFTR). Analysis of sequence divergence in the NBDs in this way suggests sites at which positive or directional selection has been at

work, possibly due to the acquisition of novel function (channel vs. transporter activity) in CFTR.

The results of RD analysis for NBD1 and NBD2 are shown in **Figure 4.1** and **Figure 4.2**. Spikes in the plots represent sites of initial interest, which are then filtered based on conservation within CFTRs of different organisms — if a site has high RD when comparing human CFTR to other ABC-Cs but is not conserved within CFTRs, it is not likely to contribute significantly to CFTR-specific function.

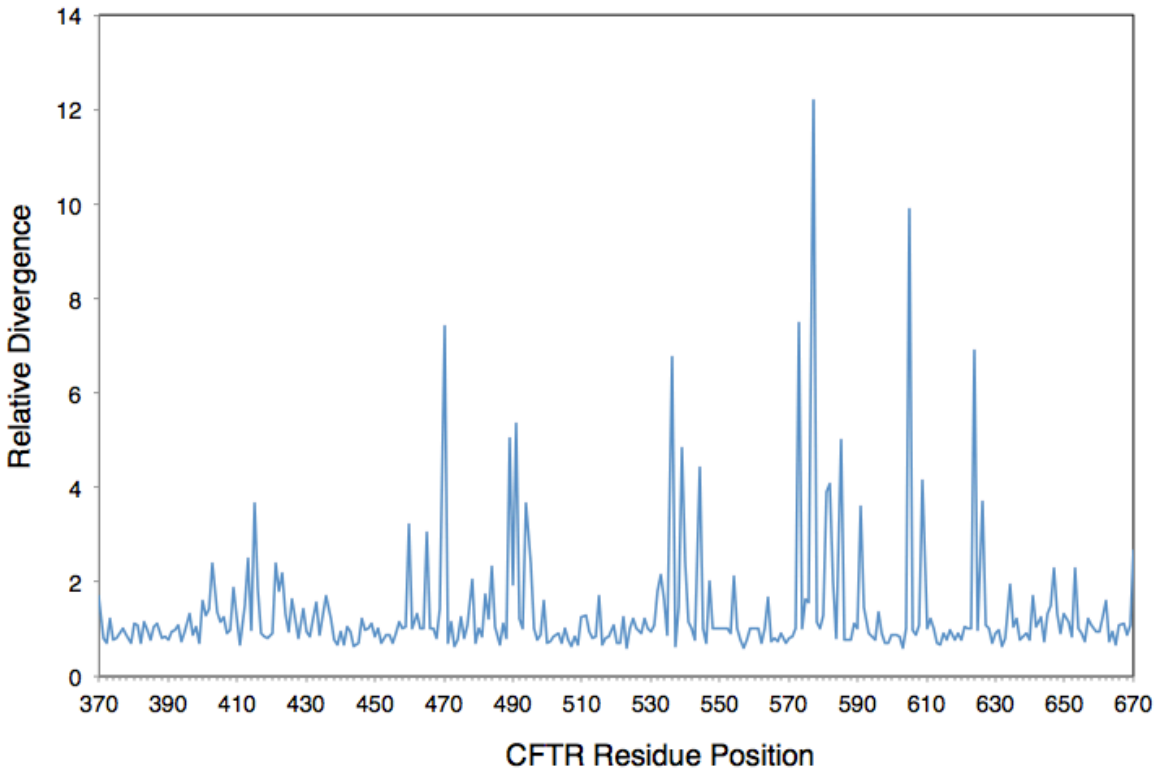

Figure 4.1: Relative divergence of CFTR sites in NBD1

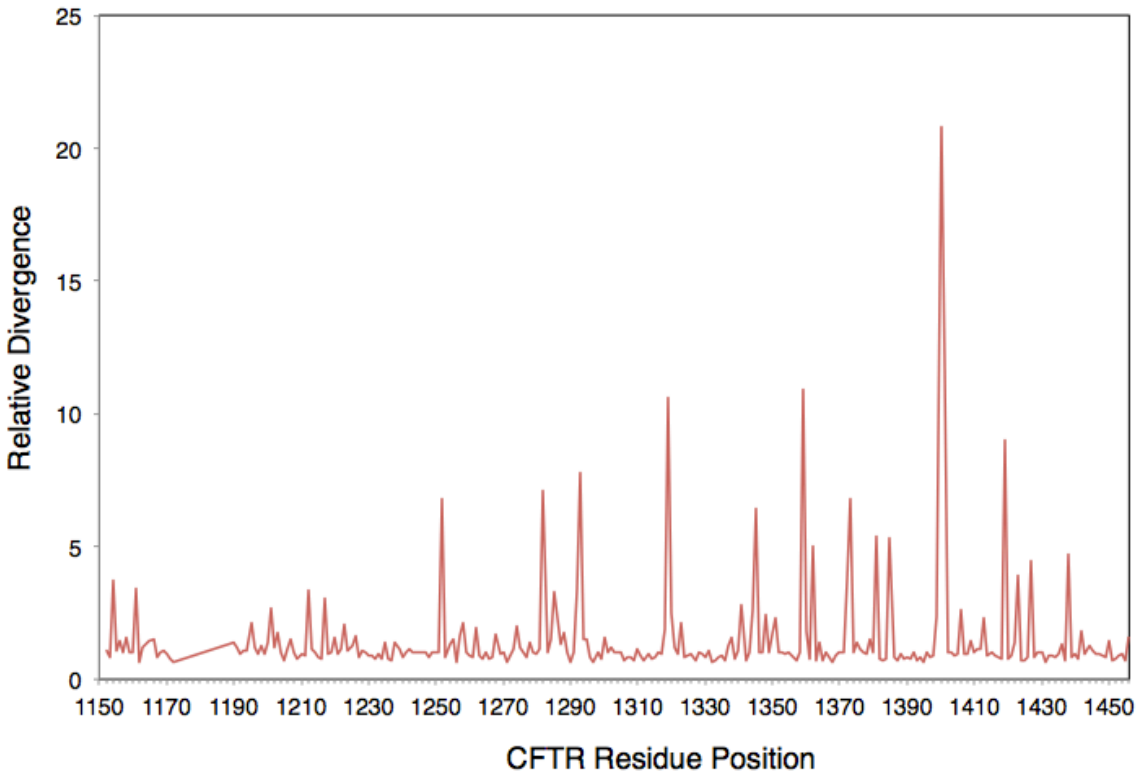

Figure 4.2: Relative divergence of CFTR sites in NBD2

Focusing on the NBD1 domain that contains the catalytically dead binding site, we note the existence of 22 residues with RD > 3. Filtering by conservation within CFTRs leaves 9 residues of interest — S605, S573, Y577, G544, C491, S489, K536, F626, and T582 — that all have the following characteristics:

- within vertebrate CFTR sequences, these sites are highly conserved
- equivalent sites in other (non-CFTR) ABC-C transporters are highly conserved
- equivalent sites in other ABC-C transporters are biochemically different from those in CFTR

In order to gain a better understanding of the possible mechanisms by which these important (and apparently directed) unique-to-CFTR NBD1 substitutions may have led to channel activity in CFTR, we conducted a review of existing experimental studies that allude to each of these twelve residues. We also noted CF-causing mutations occurring at these sites. **Table 4.3** summarizes our findings.

Table 4.3: CFTR-specific substitutions in NBD1

CFTR Residue	ABC-C Consensus	Clinical Mutation	Comments
S605	**H**		The switch region histidine is essential for ATP hydrolysis in other ABC proteins, missing in CFTR.
S573	**D**	**S573C** **S573F**	The acidic residue at the end of the Walker B sequence (D in other ABC-C, E in all other ABCs) is essential for ATP hydrolysis in other ABC, missing in CFTR.
Y577	**A**	**Y577E**	Tyrosine in place of D-loop consensus alanine (SALD)
G544	**K/R**	**G544S** **G544V**	Glycine substitutes a basic residue in the X-loop
C491	**V/A**	**C491R** **C491S** **C491F**	This hydrophilic mutation at the NBD1/ICL4 interface may weaken/alter NBD/MSD coupling.
S489	**A**		This hydrophilic mutation at the NBD1/ICL4 interface may weaken/alter NBD/MSD coupling.
K536	**G**	**K536E**	Mutation to positively charged residue, possibly near intracellular mouth of pore.
F626	**E/R**		May engage in hydrophobic interactions with the regulatory extension(RE) region, and possible binding site for corrector/potentiator CFFT-001 [43]
T582	**V/L**	**T582S** **T582I** **T582R**	This threonine is essential for PKA activation [44]

Modeling of the CFTR Partial Dimer State

Having identified sites in the primary structure that appear to have been under positive selective pressure in NBD1 of CFTR, we investigated the functional role of these sites through computational modeling and simulations of the CFTR's NBDs.

As mentioned earlier, crystal structures of human CFTR NBD1 exist [14,15]. However they were crystallized in the absence of the membrane spanning domains. Furthermore, they were crystallized with bound ATP or non-hydrolyzable ATP analogues, and are thus likely to be different from the structure of CFTR NBDs in the closed channel in their native environments where they interact extensively with the intracellular loops of the MSDs. In previous chapters, we used the structure of murine P-glycoprotein to model the structure of the closed channel state of CFTR. This has proven adequate as a starting point and yielded a number of fruitful hypotheses about the structure of the closed state, but — as we addressed in our prior discussions — it has been shown that CFTR is unlikely to adopt this completely dissociated structure for most of its gating cycle [17]. In particular, it is thought that one of CFTR's two ATP binding sites (ABS1) is not functionally competent; it is capable of binding ATP but not hydrolyzing it. As a result, the NBD dimer is thought to remain in close contact at this site and the majority of the conformational movement that results in channel gating is thought to occur due to binding, dimerization, hydrolysis and dissociation at the functional ABS2 site.

The recent publication of the crystal structure of a TM287/288 [32] — a bacterial transporter which, like CFTR, has a degenerate ABS in NBD1 — opens up the possibility of a more accurate analysis of the "partial dimer" state of CFTR. This protein was

crystallized with bound nucleotide at the degenerate binding site but without any ATP in the competent ABS2. Homology modeling of CFTR NBDs based on the NBDs of this template therefore yielded what we believe to be the most accurate picture yet of the partial dimer structure of CFTR's NBDs (henceforth CFTR-ATP/apo) (**Figure 4.3**, left). An additional structure with ATP at both binding sites (CFTR-ATP/ATP) (**Figure 4.3**, center) was also created in order to study the dynamics of the NBDs upon ATP binding at the functional site.

Modeling of Sav1866 Partial Dimer State

In order to investigate the differences between the NBDs of CFTR and those of an ABC transporter with two canonical binding sites, we also created a partial dimer structure of the bacterial ABC exporter Sav1866. The NBDs from the crystal structure of Sav1866 with the non-hydrolyzable ATP analogue AMP-PNP bound at both ABSs [38] (PDB: 2ONJ), were aligned to the NBDs from the TM287/288 partial dimer crystal structure [32] to generate a model of the partial-dimer Sav1866 state (Sav1866-ATP/ATP) (**Figure 4.3**, right). Both the CFTR models and the Sav1866 model were used as starting structures for subsequence molecular dynamics simulations.

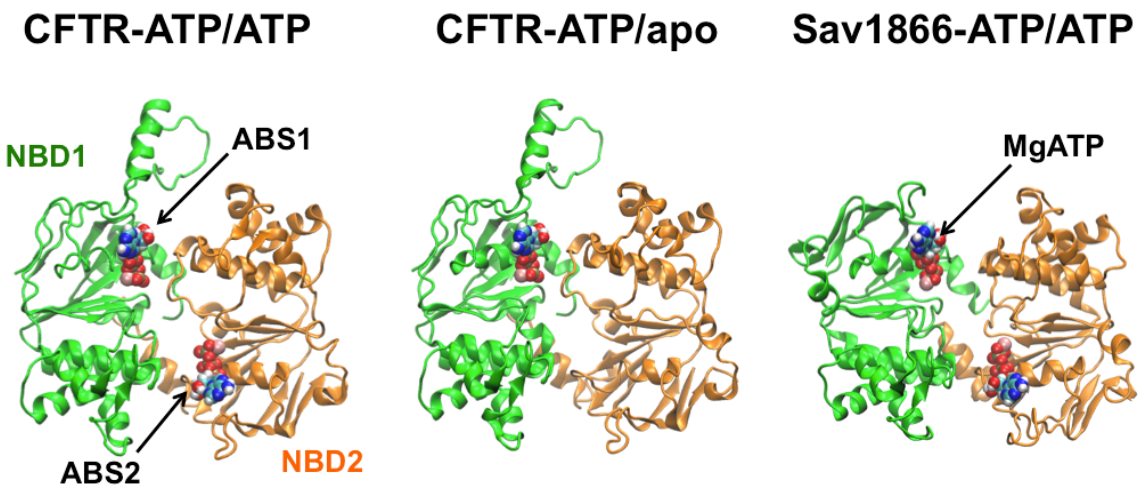

Figure 4.3: Modeled NBD systems in ribbon representation.

Locations of CFTR-specific Subsititions

Mapping the nine residues of interest identified from our earlier bioinformatics analysis onto our molecular model of the CFTR NBD1 (**Figure 4.4**) reveals that these unique-to-CFTR substitutions may be classified into four clusters depending on their location in the NBD structure (and, therefore, the most likely mode by which they influence CFTR function).

1. Substitutions at the ATP-binding sites: S573, and S605. These likely alter ATP binding and hydrolysis and thus affect channel gating. (Orange in **Figure 4.4**.)
2. Substitutions at the interface between the two NBDs and between motifs within the NBDs: G544, Y577. These may alter dimer formation and thus affect gating. (Red.)

3. Substitutions at the NBD-ICL interface: S489, C491. These may impact the transduction of the conformational wave from the NBDs to the MSDs through the ICLs. (Green.)
4. Substitutions at the peripheral surfaces of the NBDs: K536, T582, F626. These sites may be involved in interactions with the R-domain of CFTR, regulate CFTR activity directly as sites of phosphorylation, or otherwise be involved in maintaining the structural integrity of the NBDs. (Blue.)

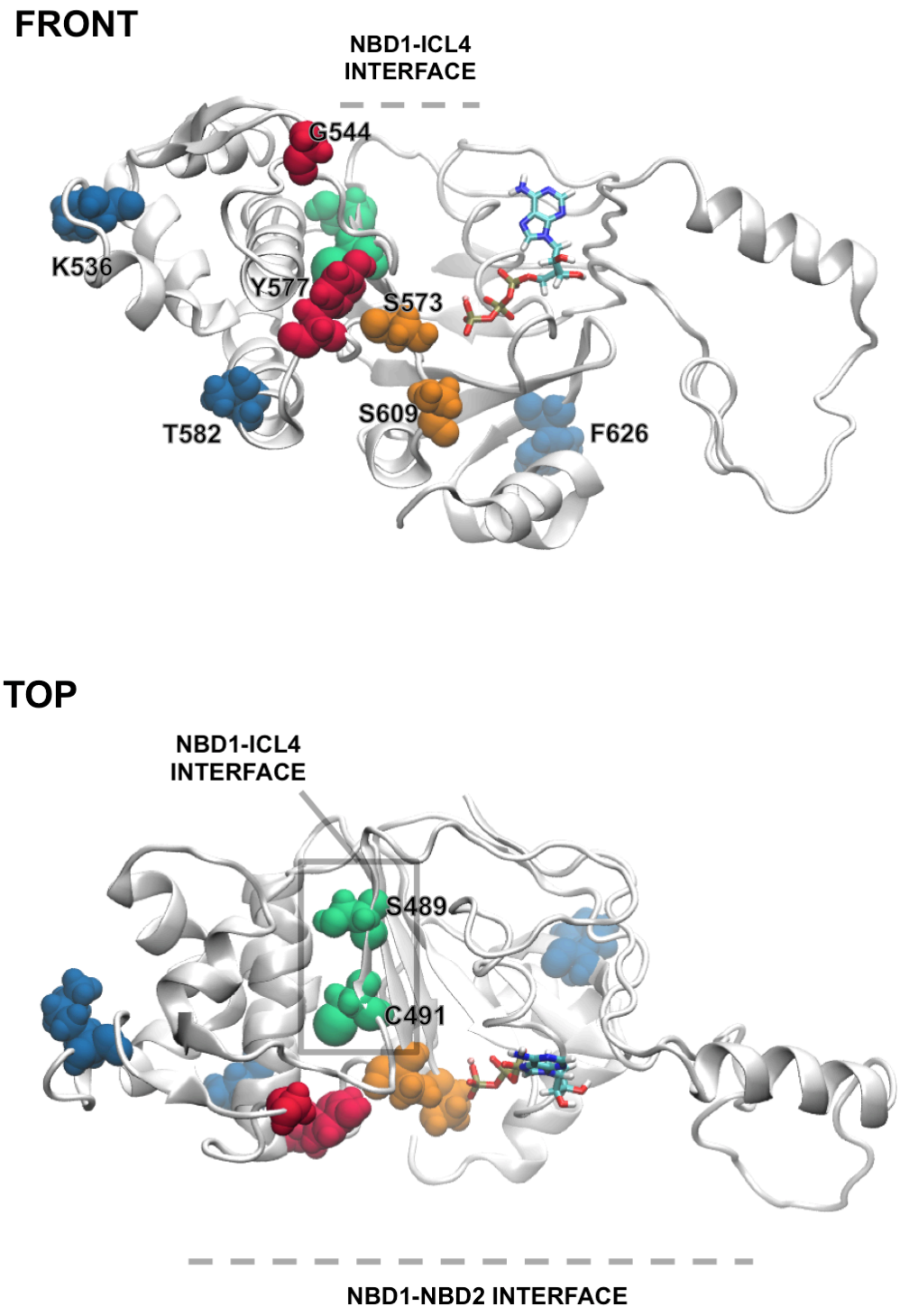

Figure 4.4: Front and top (rotated 90 degrees about the horizontal axis) views of CFTR NBD1 showing locations of CFTR-specific substitutions. ATP molecule shown in stick representation.

MD Simulations of Modeled NBDs

All-atom, unbiased MD simulations were carried out on all three starting systems: CFTR-ATP, CFTR-ATP/ATP and Sav1866-ATP. For each of these, three simulations were carried out: two lasting 25 ns each, and one that was extended to 100 ns, resulting in a total of nine simulations and a total simulated time of 0.45 μs.

Analysis of total RMSD (**Figure 4.5**) for all the simulations revealed that, in all cases, the majority of the conformational change occurred in the first 20 ns of the every simulation. Extending the simulation out to 100 ns showed very little change in any of the structures by either visual inspection of the trajectory or by analysis of the RMSD plots. The final structures from each of the simulations were also remarkably consistent, particularly at the ABSs and the NBD interfaces where we were most interested. Differences in the final RMSDs values between different runs of the same system can be attributed primarily to slightly different organizations of motifs in the periphery of the NBDs. Note that the RI loop — which was modeled *de novo* (since no equivalent segment existed in the available crystal structures) — was omitted from the RMSD plots.

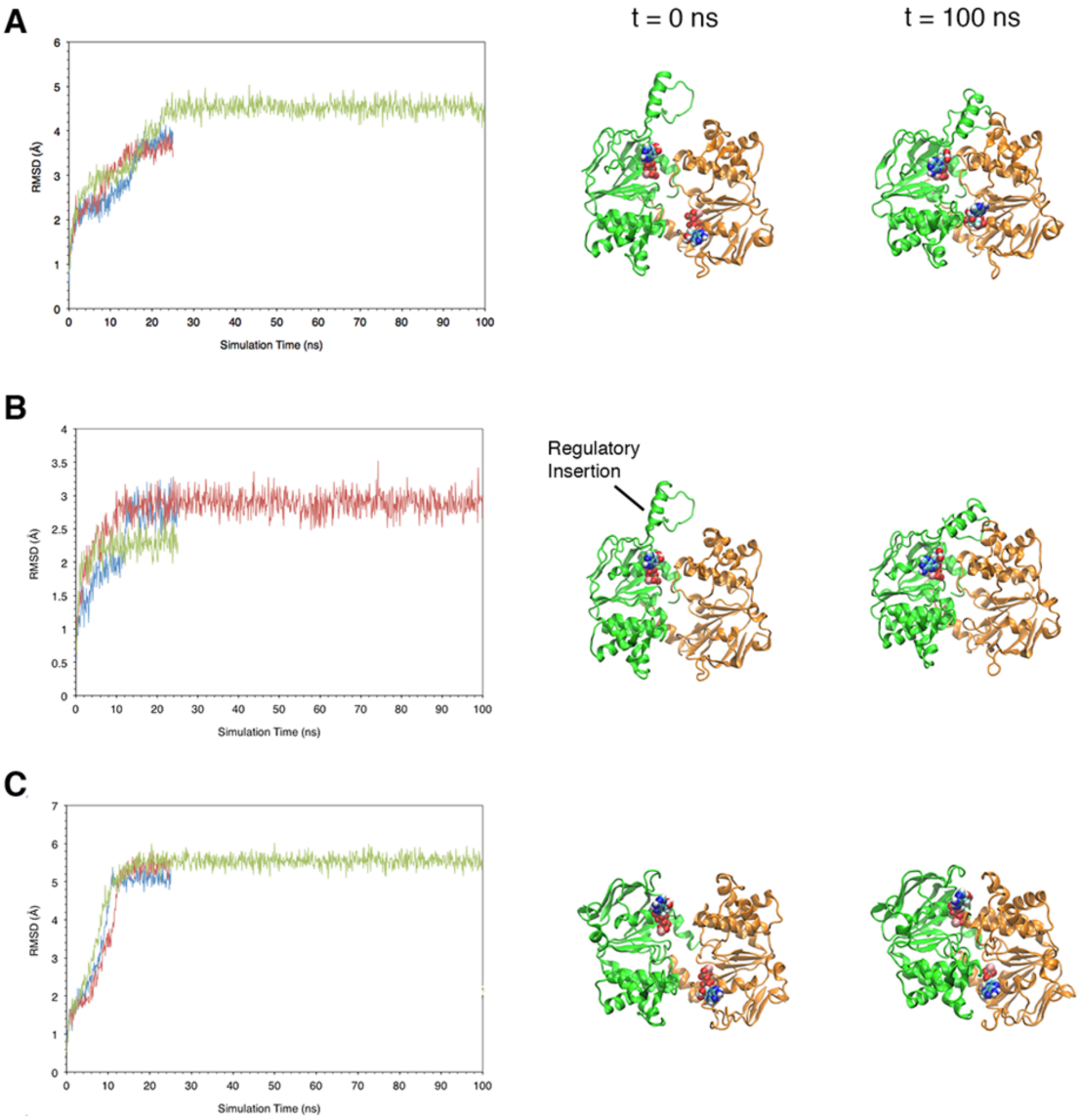

Figure 4.5: RMSD evolution and snapshots of A) CFTR-ATP/ATP, B) CFTR-ATP/apo, and C) Sav1866-ATP/ATP systems. RMSD was measured for protein heavy atoms except those in the regulatory insertion loop of CFTR models.

ATP and the Canonical Signature Motif are Required for Tight NBD Dimerization

In order to investigate whether our CFTR models were capable of forming tight dimers despite having only one functional ABS, and whether ATP binding was necessary for

dimer formation, we observed the trajectories of both the CFTR-ATP/apo and CFTR-ATP/ATP systems over 100 ns. In both cases, we note a nearly immediate collapse of the NBD dimers into more compact configurations, suggesting that our CFTR homology models based on the TM287/288 crystal structure were not minimum energy structures. Following this initial compaction, however, the two systems behaved differently. Whereas the ATP/ATP structure continued to close the inter-dimer gap present in the initial "partial dimer configuration", the ATP/apo structure remained partially open at the apo binding site. As a measure of inter-NBD separation, we plotted the distance between the C-α atoms of the terminal hydroxyl-containing residue in the Walker A motif (T465 in ABS1, S1251 in ABS2) and the second glycine in the signature motif (G1349 in ABS1, G551 in ABS2) (**Figure 4.6**, top). The inter-NBD distance at both these points at 25 ns is maintained throughout the remainder of the 100 ns simulations (**Figure 4.6**, bottom-left). In both cases, the separation is greater for the CFTR-ATP/apo system, although this difference is more pronounced in the ABS2 site, which contains ATP in one of the simulations but is empty in the other. Notably, ATP binding at the competent binding site results in tighter binding than at the degenerate ABS1. Since dimer formation is generally considered a prerequisite for enzymatic hydrolysis in ATP NBDs, this observation may partly explain why ABS1 is catalytically dead.

We note that the inter-NBD distance dips quickly in the ATP/apo simulation and then rebounds slightly at around the 15 ns mark. From visual inspection of the trajectory, we note that this is approximately the point at which the regulatory insertion loop on NBD1 docks onto the periphery of NBD2 in our simulation. However, due to the speculative

nature of the RI structure in our model, it is not possible to say whether this feature of our simulations has biological relevance.

For comparison, we also inspected the inter-NBD distance for our Sav1866-ATP/ATP model as over 100 ns. As in our CFTR simulations, the Sav-based model the initial partial dimer structure of the Sav1866 also underwent a transition to a tight dimer. Plotting the distance between the equivalent sites to those mapped for the CFTR models (S381 and G481 of the opposing NBD) (**Figure 4.6**, bottom-right), we note that the transition was completed within the first 20 ns. The final separation of the ABS2 site (functional in both CFTR and Sav1866) was nearly identical in our Sav1866 and CFTR-ATP/ATP simulations, indicating that this site in CFTR is in fact capable of forming a canonical tight dimer in the presence of ATP. At the ABS1 site, on the other hand, we observe a marked difference between the final separation in CFTR (where this site is degenerate) and Sav1866 (where it is canonical).

Taken together, these results confirm the hypothesis that ATP acts as a molecular "glue" that binds the two NBD dimers together. Upon closer inspection of the final structures produced in our simulations, we note that each ATP molecule interacts through a network of non-covalent interactions with the Walker A motif of one NBD monomer and the LSGGQ signature residues of the opposing monomer (**Figure 4.7**, top). Mutation of the of signature motif to LSHGH at CFTR's non-canonical binding site weakens this inter-NBD interaction, leading to incomplete closure at this site (**Figure 4.7**, bottom).

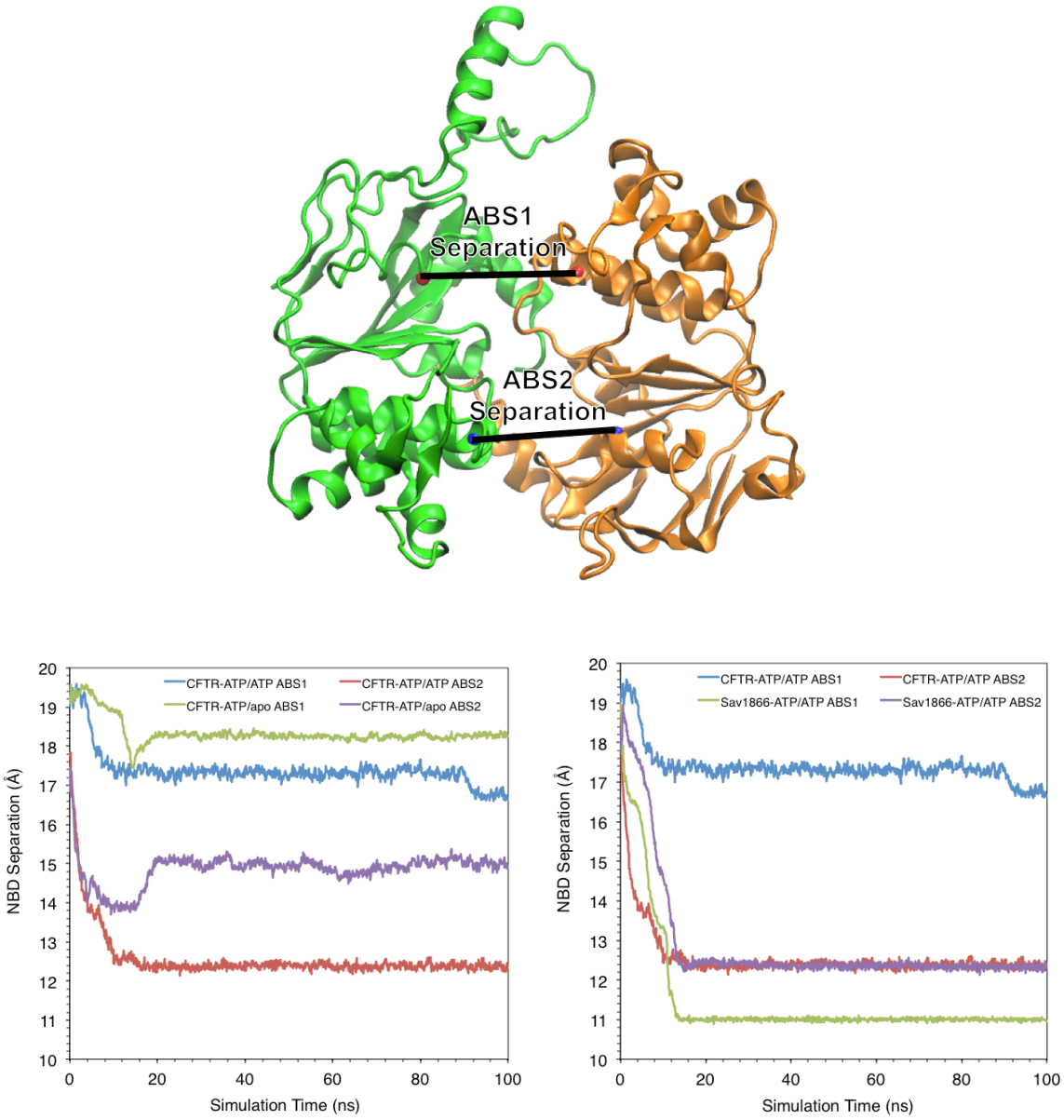

Figure 4.6: Changes in NBD separation at ABS1 and ABS2 for the simulated systems.

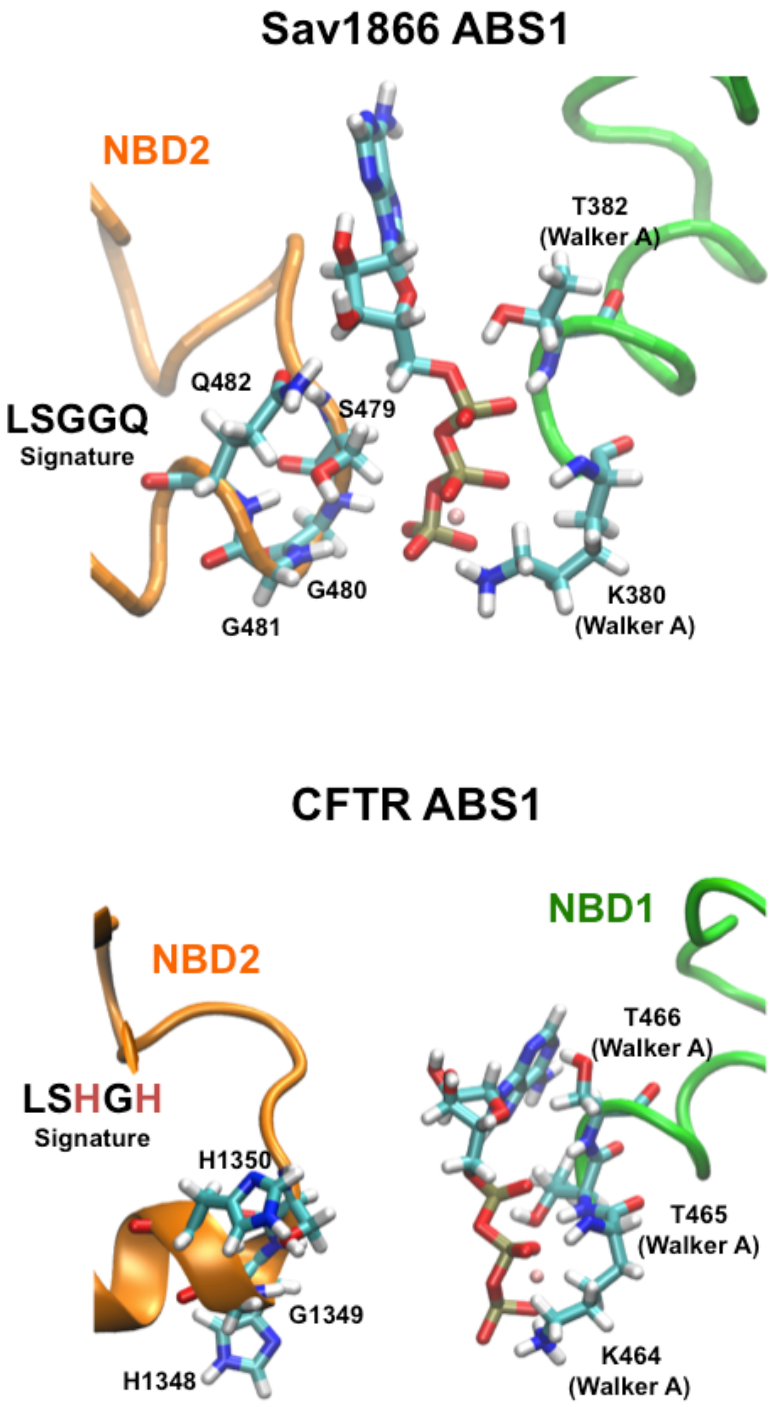

Figure 4.7: Close-up of ATP-binding site 1 in Sav1866 (top) and CFTR (bottom).

CFTR D-loop Substitutions Stabilize the NBD1 Q-Loop Glutamine Q493

The tyrosine residue at site 577 in CFTR is located in the so-called D-loop of NBD1 and is equivalent to an alanine residue in all other ABC-C's as well as in Sav1866 (the second residue of the canonical SALD motif). In our simulations, this tyrosine forms strong hydrogen bonds with the highly conserved catalytic glutamine Q493 in the so-called "Q-loop" of NBD1 (**Figure 4.8**). This glutamine, in turn, coordinates with both the Mg^{2+} and the γ-phosphate in the non-canonical ABS1. Interestingly, the counterpart residue of Y577 in NBD2 — H1375 — is also divergent from the NBD consensus (alanine) and participates in another Q-loop-stabilizing hydrogen bond with Q493. The Q-loop is thought to be a primary mediator of conformational signaling between the NBDs and the ICLs in CFTR [24] and other ABC transporters [25,45].

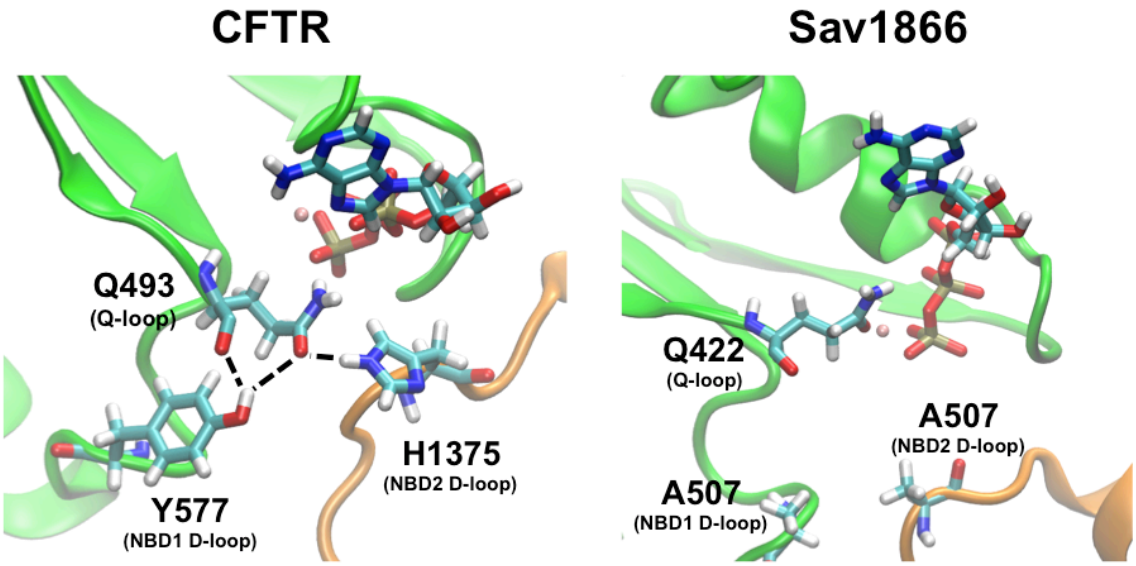

Figure 4.8: D-loop residues stabilize Q-loop glutamine in CFTR but not in Sav1866. Hydrogen bonds represented as dashed lines. ATP and Mg^{2+} bound to ABS1 are also shown in these figures. (Snapshots at t = 100 ns.)

G544 Replaces a Salt Bridge Between the X-loop and Q-loop in CFTR

G544 was also identified in our analysis to be a unique-to-CFTR substitution; the equivalent site in most other ABC-Cs and Sav1866 is a basic arginine or lysine. In our simulations of Sav1866, the equivalent X-loop site (R474) forms a salt bridge with a Q-loop aspartate (D423) that couples these two motifs. The Q-loop and the X-loop together form a major part of the NBD-ICL coupling interface, and their concerted motion may be necessary transmitting the ATP-driven conformation wave from the NBDs to the rest of the protein [25,26,45]. However, since the associated ABS in CFTR is catalytically inactive, the G544 X-loop substitutions may serve to decouple this interface from events at the degenerate binding site.

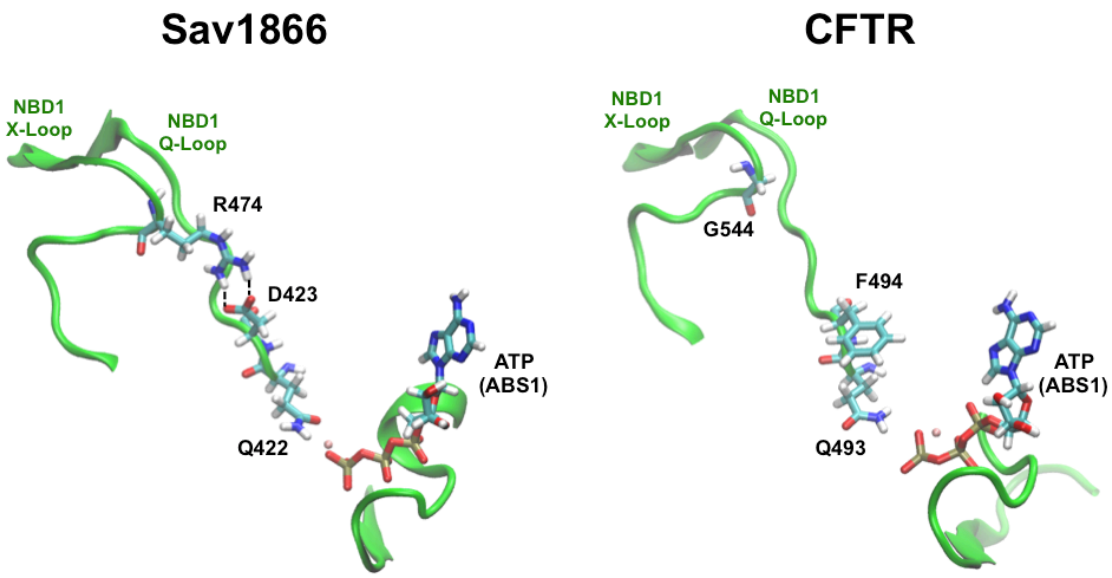

Figure 4.9: The Q-loop and X-loop of Sav1866 are coupled by a salt bridge in Sav1866, but not in CFTR. (Snapshots at t = 100 ns.)

DISCUSSION

CFTR is the only known member of the ABC superfamily known to have channel activity. In this study, we identified sites within the nucleotide binding domains of CFTR that appear to have undergone lineage-specific positive evolutionary selection, and investigated their possible functional roles — particularly with regard to the acquisition of channel function — through molecular modeling and simulations.

By defining a measure of *Relative Divergence* we were able to identify sites that are highly conserved in other ABC transporters, but divergent in CFTR — and therefore likely candidates for understanding the emergence of channel behavior. Many such sites were identified (**Figure 4.1** and **Figure 4.2**), but in the present analysis we focused primarily on the NBD1 sites listed in **Table 4.3**.

CFTR-Specific Substitutions Render ABS1 Catalytically Incompetent

Two substitutions at the ATP-binding site 1 —S573 and S605 — likely affect the efficacy of ATP binding and hydrolysis and thus alter the timing of channel closure. The serine at site 573 in CFTR occurs at the end of the Walker B sequence, replacing the "catalytic carboxylate"-containing aspartate or glutamate amino acid in its ABC homologs. This residue was found in several other ABC transporters to be essential for the conversion of ATP to ADP, serving as the catalytic base for the hydrolysis reaction [28-31]. Similarly, the serine at position 605 is a universally conserved histidine in other ABC transporters, also thought to be crucial for coordinating ATP hydrolysis [30]. We

surmise that CFTR substitutions away from the ABC consensus at these sites result in slowed (or abolished) ATP hydrolysis at ATP-binding site 1.

Our MD simulations of Sav1866 and CFTR also indicate that dimer formation is never complete at the degenerate site, due to substitutions in the LSGGQ signature motif of the opposing NBD (LSHGH in CFTR) that significantly reduce the strength of the non-covalent interactions between the bound ATP molecule and the opposing NBD monomer (**Figure 4.7**). Since dimer formation is known to be essential for ATP hydrolysis in ABC NBDs, these substitutions are likely to further impair nucleotide dephosphorylation in the degenerate site.

Interestingly, this loss of ATPase function may have led to gain of channel function in CFTR. Impaired dephosphorylation of bound ATP at the degenerate site allows the NBD dimer to remain in close contact over prolonged periods relative to other ABC transporters, thus increasing channel open time. In addition, if the NBDs are never completely dissociated, ATP turnover at the functional ABS2 may be sensed more rapidly, resulting in faster tight dimer formation and increasing the efficiency of channel gating.

CFTR-Specific Substitutions Alter Inter-Domain Communication Pathways

Previous studies on ATP binding and hydrolysis in the NBDs of Sav1866 and other ABC transporters reveal the central role of the Q-loop [22,45,46]. This loop has a single highly conserved glutamine that coordinates the MgATP at the ABS. Amino acids immediately downstream from this glutamine, however, are highly variable, and the structure of the Q-

loop has been shown to be alternately extended and retracted in several NBD crystal structures [10,12,47,48]. It has been suggested that the Q-loop glutamine binds to the Mg^{2+} and / or γ-phosphate at the active site. Upon hydrolysis, this interaction is broken, and the Q-loop undergoes a "switching" transition that induces changes in the secondary structure upstream [45]. The Q-loop is also thought to interact strongly with the X-loop motif, participating in collective motions upon the release of MgATP. In turn, both the Q-loop and the X-loop participate in persistent interactions with the intracellular loops of the membrane-spanning domains, and ATP-driven conformational changes in the NBDs are thought to be transmitted to the rest of the protein through this interface [24,25,46].

This transduction pathway — from the ABS to the Q-loop/X-loop to the ICLs — appears to be broken in the case of the degenerate binding site of CFTR in three ways. First, two residues — Y577 and H1375 — that deviate from the consensus D-loop SALD sequence are seen to interact with the ABS1 Q-loop glutamine (Q493) in our simulations (**Figure 4.8**). These interactions may have a stabilizing effect, locking the glutamine in place and preventing the Q-loop "switching" observed in other ABC transporters. Secondly, in our Sav1866 simulations, we observed close coupling of the Q-loop and the X-loop through a salt bridge between D423 and R474 (**Figure 4.9**), which may be important for coordinating the collective motion of these two domains necessary for signal transduction. In CFTR, the residues at the equivalent sites are F494 and G544, respectively, and no coupling interactions are observed between the Q-loop and the X-loop in our CFTR simulations. Finally, residues C491 and S489 — located where the ICLs are thought to dock into NBD1 — are unique-to-CFTR substitutions that deviate significantly from the ABC consensus at equivalent sites (**Table 4.3**); in both cases, the

CFTR variant is significantly more hydrophilic than the consensus residue. We speculate that these substitutions will also alter the NBD:ICL interaction in CFTR, although — without modeling the rest of the protein — we are not yet able to determine how.

The question of how CFTR alone —of all the ABC transporters — came to gain channel activity is a fascinating one that remains to be answered definitely. In the preceding studies, we focused our efforts only on the nucleotide-binding domains of CFTR, asking what changes in the many conserved motifs of these subunits could have led lineage-specific gain-of-function. Due to high homology between CFTR and its ABC-C homologs in the NBDs we were able to employ bioinformatics techniques to identify particular sites of interest which appear to have been positively selected in the CFTR lineage alone. Comparative molecular modeling and simulations of CFTR and Sav1866 NBDs revealed that these sites participate in important interactions involved in binding to and hydrolyzing ATP, as wells as in transmitting the conformational changes from the NBDs to the other domains of the protein; interactions which are significantly saltered in CFTR's NBDs. These studies provide an important advance in our understanding of CFTR's role as an ion channel.

References

1. Jordan IK, Kota KC, Cui G, Thompson CH, McCarty NA (2008) Evolutionary and functional divergence between the cystic fibrosis transmembrane conductance regulator and related ATP-binding cassette transporters. Proc Natl Acad Sci U S A 105: 18865–18870. doi:10.1073/pnas.0806306105.

2. Sebastian A, Rishishwar L, Wang J, Bernard KF, Conley AB, et al. (2013) Origin and evolution of the cystic fibrosis transmembrane regulator protein R domain. Gene 523: 137–146. doi:10.1016/j.gene.2013.02.050.

3. Rich DP, Berger HA, Cheng SH, Travis SM, Saxena M, et al. (1993) Regulation of the cystic fibrosis transmembrane conductance regulator Cl- channel by negative charge in the R domain. J Biol Chem 268: 20259–20267.

4. Bompadre SG, Ai T, Cho JH, Wang X, Sohma Y, et al. (2005) CFTR gating I: Characterization of the ATP-dependent gating of a phosphorylation-independent CFTR channel (DeltaR-CFTR). J Gen Physiol 125: 361–375. doi:10.1085/jgp.200409227.

5. Locher KP, Lee AT, Rees DC (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 296: 1091–1098. doi:10.1126/science.1071142.

6. Chang G (2003) Structure of MsbA from Vibrio cholera: a multidrug resistance ABC transporter homolog in a closed conformation. J Mol Biol 330: 419–430.

7. Chang G, Roth CB (2001) Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. Science 293: 1793–1800. doi:10.1126/science.293.5536.1793.

8. Hung LW, Wang IX, Nikaido K, Liu PQ, Ames GF, et al. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. Nature 396: 703–707. doi:10.1038/25393.

9. Reyes CL, Chang G (2005) Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. Science 308: 1028–1031. doi:10.1126/science.1107733.

10. Diederichs K, Diez J, Greller G, Müller C, Breed J, et al. (2000) Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. EMBO J 19: 5951–5961. doi:10.1093/emboj/19.22.5951.

11. Verdon G, Albers SV, Dijkstra BW, Driessen AJM, Thunnissen AMWH (2003) Crystal structures of the ATPase subunit of the glucose ABC transporter from Sulfolobus solfataricus: nucleotide-free and nucleotide-bound conformations. J Mol Biol 330: 343–358.

12. Yuan YR, Blecker S, Martsinkevich O, Millen L, Thomas PJ, et al. (2001) The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. J Biol Chem 276: 32313–32321. doi:10.1074/jbc.M100758200.

13. Schmitt L, Benabdelhak H, Blight MA, Holland IB, Stubbs MT (2003) Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. J Mol Biol 330: 333–342.

14. Lewis HA, Buchanan SG, Burley SK, Conners K, Dickey M, et al. (2003) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23: 282–293. doi:10.1038/sj.emboj.7600040.

15. Lewis HA, Zhao X, Wang C, Sauder JM, Rooney I, et al. (2005) Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 280: 1346–1353. doi:10.1074/jbc.M410968200.

16. Atwell S, Brouillette CG, Conners K, Emtage S, Gheyi T, et al. (2010) Structures of a minimal human CFTR first nucleotide-binding domain as a monomer, head-to-tail homodimer, and pathogenic mutant. Protein Eng Des Sel 23: 375–384. doi:10.1093/protein/gzq004.

17. Szollosi A, Muallem DR, Csanády L, Vergani P (2011) Mutant cycles at CFTR's non-canonical ATP-binding site support little interface separation during gating. J Gen Physiol 137: 549–562. doi:10.1085/jgp.201110608.

18. J E Walker MSMJRNJG (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1: 945.

19. Ye J, Osborne AR, Groll M, Rapoport TA (2004) RecA-like motor ATPases--lessons from structures. Biochim Biophys Acta 1659: 1–18. doi:10.1016/j.bbabio.2004.06.003.

20. Moody JE, Millen L, Binns D, Hunt JF, Thomas PJ (2002) Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. J Biol Chem 277: 21111–21114. doi:10.1074/jbc.C200228200.

21. Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180–185. doi:10.1038/nature05155.

22. Oancea G, O'Mara ML, Bennett WFD, Tieleman DP, Abele R, et al. (2009) Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. Proc Natl Acad Sci U S A 106: 5551–5556. doi:10.1073/pnas.0811260106.

23. Ward A, Reyes CL, Yu J, Roth CB, Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci U S A 104: 19005–19010. doi:10.1073/pnas.0709388104.

24. He L, Aleksandrov AA, Serohijos AWR, Hegedus T, Aleksandrov LA, et al. (2008) Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating. J Biol Chem 283: 26383–26390. doi:10.1074/jbc.M803894200.

25. Dalmas O, Orelle C, Foucher A-E, Geourjon C, Crouzy S, et al. (2005) The Q-loop disengages from the first intracellular loop during the catalytic cycle of the multidrug ABC transporter BmrA. J Biol Chem 280: 36857–36864. doi:10.1074/jbc.M503266200.

26. Aittoniemi J, de Wet H, Ashcroft FM, Sansom MSP (2010) Asymmetric switching in a homodimeric ABC transporter: a simulation study. PLoS Comput Biol 6: e1000762. doi:10.1371/journal.pcbi.1000762.

27. Hanson PI, Whiteheart SW (2005) AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol 6: 519–529. doi:10.1038/nrm1684.

28. Orelle C, Dalmas O, Gros P, Di Pietro A, Jault J-M (2003) The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. J Biol Chem 278: 47002–47008. doi:10.1074/jbc.M308268200.

29. Tombline G, Bartholomew LA, Tyndall GA, Gimi K, Urbatsch IL, et al. (2004) Properties of P-glycoprotein with mutations in the "catalytic carboxylate" glutamate residues. J Biol Chem 279: 46518–46526. doi:10.1074/jbc.M408052200.

30. Zaitseva J, Jenewein S, Jumpertz T, Holland IB, Schmitt L (2005) H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. EMBO J 24: 1901–1910. doi:10.1038/sj.emboj.7600657.

31. Geourjon C, Orelle C, Steinfels E, Blanchet C, Deléage G, et al. (2001) A common mechanism for ATP hydrolysis in ABC transporter and helicase superfamilies. Trends Biochem Sci 26: 539–544.

32. Hohl M, Briand C, Grütter MG, Seeger MA (2012) Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nat Struct Mol Biol 19: 395–402. doi:10.1038/nsmb.2267.

33. Jones PM, George AM (2012) Role of the D-loops in allosteric control of ATP hydrolysis in an ABC transporter. J Phys Chem A 116: 3004–3013. doi:10.1021/jp211139s.

34. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113. doi:10.1186/1471-2105-5-113.

35. Grantham R (1974) Amino acid difference formula to help explain protein evolution. Science 185: 862–864.

36. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779–815. doi:10.1006/jmbi.1993.1626.

37. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291–325. doi:10.1146/annurev.biophys.29.1.291.

38. Dawson RJP, Locher KP (2007) Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. FEBS Lett 581: 935–938. doi:10.1016/j.febslet.2007.01.073.

39. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33–38, 27–28.

40. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781–1802. doi:10.1002/jcc.20289.

41. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4: 187–217. doi:10.1002/jcc.540040211.

42. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 23: 327–341. doi:10.1016/0021-9991(77)90098-5.

43. Hudson RP, Chong PA, Protasevich II, Vernon R, Noy E, et al. (2012) Conformational changes relevant to channel activity and folding within the first nucleotide binding domain of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 287: 28480–28494. doi:10.1074/jbc.M112.371138.

44. Chappe V, Hinkson DA, Howell LD, Evagelidis A, Liao J, et al. (2004) Stimulatory and inhibitory protein kinase C consensus sequences regulate the cystic fibrosis transmembrane conductance regulator. Proc Natl Acad Sci U S A 101: 390–395. doi:10.1073/pnas.0303411101.

45. Jones PM, George AM (2002) Mechanism of ABC transporters: a molecular dynamics simulation of a well characterized nucleotide-binding subunit. Proc Natl Acad Sci U S A 99: 12639–12644. doi:10.1073/pnas.152439599.

46. Becker J-P, Van Bambeke F, Tulkens PM, Prévost M (2010) Dynamics and structural changes induced by ATP binding in SAV1866, a bacterial ABC exporter. J Phys Chem B 114: 15948–15957. doi:10.1021/jp1038392.

47. Karpowich N, Martsinkevich O, Millen L, Yuan YR, Dai PL, et al. (2001) Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. Structure 9: 571–586.

48. Gaudet R, Wiley DC (2001) Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. EMBO J 20: 4964–

4972. doi:10.1093/emboj/20.17.4964.

CHAPTER 5

CONCLUSIONS & FUTURE WORK

The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel found in the apical membranes of many types of epithelial cells, where it plays a crucial role in ion and fluid homeostasis. Defects in CFTR due to mutation underlie cystic fibrosis (CF), the most common life-shortening autosomal recessive genetic disease among Caucasians. CFTR malfunction is also involved polycystic kidney disease and secretory diarrhea due to cholera. There is a tremendous impetus, therefore, to understand CFTR at a molecular level.

Modeling & Simulations of CFTR Gating

CFTR is a member of the ABC superfamily of proteins, but anomalous in being the only member of this ancient family to exhibit channel activity. Nonetheless, it share several features in common with its ABC transporter cousins, including the formation of dimers between its two intracellular nucleotide binding domains upon ATP binding, which leads to a chain of conformation changing events that culminates — in the case of CFTR — in opening of the transmembrane channel.

In order to visualize the conformational changes in the transition from the closed to open states and to form hypotheses about the underlying mechanisms, we generated new homology models of CFTR in the open and closed states. Our open-state model preserved the expected R347–D924 and R352–D993 salt bridges, and presented a greater number of

the pore-lining residues on transmembrane helices TM6 and TM12 in their experimentally predicted positions compared to previous CFTR models. Our model structures also revealed a possible basis for the observed state-dependent accessibility of residue R334 at the extracellular mouth of the pore.

Furthermore, we performed targeted molecular dynamics simulations between our closed- and open-state models in order to generate a trajectory of the conformational transitions that may occur in a channel opening event. By tracking the formation and breakage of particular interactions in the nucleotide-binding domains, the intra-cellular loops, and the membrane-spanning domains over the course of the trajectory, we found that our simulations demonstrated a conformational wave that was initiated at the nucleotide-binding domains (NBDs) and ended with the formation of an open conduction pathway in the transmembrane pore, consistent with the notion that ATP-binding at the NBDs drives channel opening in CFTR. Analysis of our simulation trajectory also revealed intermediate structures that likely represent the molecular conformation of CFTR at several stages of its gating cycle, including a closed state that bears close resemblance to the partial-dimer crystal structure of the related ABC transporter TM287/288, and a strained transition state with fully dimerized NBDs but with a channel pore that is closed to chloride conduction. Changes in side-chain interactions that were unconstrained in our modeling were also observed in all major domains of the protein, and the predictive value of our models was confirmed through experimental validation of one such novel intra-protein salt bridge (R334–E217) that breaks near the end of the channel opening transition.

Future Work

Even though our models have been verified to some extent, they remain only models. With the expected advent of new crystal structures with greater homology to CFTR — and perhaps a crystal structure of CFTR itself — we believe that there remains room for more the creation of more accurate models. Our use of restraint-based modeling also leaves the door open for the introduction of additional empirically derived constraints as more structure/function data becomes available. We chose to use a biased simulation protocol — targeted molecular dynamics — for computational expediency. The use of more efficient conformational sampling algorithms, and the advent of faster, and cheaper computer hardware could soon allow simulations of the entire CFTR protein in its native environment that capture biologically relevant motions without the need for applying any external forces.

Identification of CFTR-Specific Substitutions and Their Effects

CFTR is the only ABC protein known to function primarily as a channel protein, rather than as an alternating access transporter. In order to begin understanding the molecular basis for this altered function, we focused our attention on the nucleotide-binding domains, searching for substitutions which are unique to CFTR. Leveraging high homology between CFTR and its ABC-C homologs in the NBDs, we employed bioinformatics techniques to identify particular sites of interest that appear to have been positively selected in the CFTR lineage, but are largely conserved elsewhere.

Subsequently, we applied molecular modeling techniques to create models of CFTR and Sav1866 NBDs based on the partial dimer template of the TM287/288 crystal structure, and subjected them to unbiased molecular dynamics simulations in the presence and absence of ATP in their binding sites. Mapping the CFTR-specific substitutions sites onto these models revealed that several of them participate in important interactions involved in binding to and hydrolyzing ATP. Whereas the Walker A motif for binding ATP was conserved at both ATP-binding sites, there are several CFTR-specific substitutions in the catalytic residues in the switch region and Walker B sequence, as well as the signature motif required for dimer formation, of ABS1. Substitutions are also seen which appear to abolish coupling of the Q-loop associated with ABS1 to its ICL interface, potentially altering the transmission of the conformational changes from the NBDs to the other domains of the protein. The fact that these substitutions are unique to CFTR and so is channel activity leads to the hypothesis that loss of ATPase function may have contributed to gain of channel function in CFTR.

Future Work

Having identified sites in the NBDs that are substantially divergent in CFTR, we may begin mutating them back to their ABC-C consensus residues to ask if CFTR channel activity can be abolished (or, perhaps, if transporter activity can be regained) by reverting changes in the NBDs unique to the CFTR lineage. Conversely, we may ask if introduction of CFTR-specific substitutions to a closely related transporter, e.g. ABC-C4, can diminish its transporter activity (and perhaps even lead to channel formation). Ancestral sequence reconstruction methods may also be employed to determine the

nature of the earliest common ancestors of CFTR and its closest ABC homologs, in order to ask which specific changes may have led to the emergence of channel function. In our modeling and simulation, we focused only on the NBDs due to their high homology and, therefore, reliable sequence alignments with related ABC-transporters. It may be interesting to extend these efforts to include the ICLs and even the membrane spanning domains in order to gain a better mechanistic understanding of how CFTR-specific substitutions led to altered function.