MULTI-SCALE MODELING OF THE STRUCTURE AND DYNAMICS OF MACROMOLECULES

Adrian Wendil R. Serohijos

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics & Astronomy.

Chapel Hill
2009

Approved:

Richard Superfine, Ph.D.	*Chair*
Nikolay V. Dokholyan, Ph.D.	*Advisor*
Timothy C. Elston, Ph.D.	*Reader*
Hugon J. Karwowski, Ph.D.	*Reader*
Jianping Lu, Ph.D.	*Member*

ABSTRACT

ADRIAN W.R. SEROHIJOS: Multi-scale modeling of the structure and dynamics of macromolecules

(Under the mentorship of Nikolay V. Dokholyan)

Biology is defined by phenomena that are inherently complex spanning multiple length and time scales. To understand these processes, there is a need for multi-scale approaches that provide a coherent framework for describing and interrogating these phenomena. Here, we employ multiple approaches to investigate specific biological systems. The first system we studied was the cytoplasmic dynein motor, a protein that walks along the microtubule tracks in cells. The major objective in the dynein motors field is to understand its mechanism. Specifically, what is dynein's structure and how does it transduce chemical energy into mechanical work? We proposed a theoretical structural model of the motor and performed normal mode analysis and molecular dynamics on the motor unit structure. These studies hypothesized new structural features in the dynein motor unit and proposed a potential mechanism for energy transduction [5,6,80]. The second system we studied was the CFTR channel, which regulates ion transport in the apical membrane of epithelial cells. Mutations in the CFTR protein are the basis of the cystic fibosis disease. One of the primary question is how a single residue deletion (Phe508) lead to ~90% of cystic fibrosis cases. We performed molecular dynamics simulation of the first nucleotide-binding domain of CFTR and showed that the wild type and mutant exhibit a difference in their folding kinetics, in agreement with experiments. These simulations also determined the potential structural

origin of this misfolding defect. We also proposed a complete model of the CFTR channel to identify the location of the Phe508 residue in the whole protein. This result is important in understanding another aspect of the ΔF508 defect, which is the misassembly of the whole CFTR protein during its biosynthesis.

ACKNOWLEDGEMENTS

Foremost of all, I would like to thank my mentor Nikolay V. Dokholyan for the countless opportunities to work on good science. The depth of his patience and wisdom in life is unfathomable. He certainly taught me much more than science.

I would also like to thank current and former members of the Dokholyan group, with special mention to Feng Ding and Yiwen Chen. Not only did I earn esteemed colleagues but also good friends.

I would also like to thank my collaborators John Riordan, Tamas Hegedus, and the rest of the Riordan group for placing us right at the forefront of cystic fibrosis research; and to Timothy Elston and his group for introducing me to the exciting world of molecular motors. Scientific output is less a function of personal effort and talent than the people one works with; I could not be more fortunate in this regard.

Thanks to the Board of Governors Fellowship, Program in Molecular and Cellular Biophysics, and the American Heart Association Predoctoral Fellowship for funding. Thanks also to the members of my committee for supporting my efforts.

Thanks to Judith and Nanette for keeping me company and introducing to a whole new community of Filipinos and friends in Chapel Hill and the Philippines. Lastly, I would like to thank my family, friends, and loved ones for their infinite support.

Adrian W.R. Serohijos
Chapel Hill, NC
4 March 2009

Ha akon higugmaon nga Mama

(To my beloved Mama)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

The permissions for the adaption of published figures and texts in the dissertation were granted by the publishers.

Dynein Molecular Motor

1. A.W.R. Serohijos, Y. Chen, F. Ding, T.C. Elston, and N.V. Dokholyan, "A new structural model reveals energy transduction in dynein" (2006) *Proc. Natl. Acad. Sci. USA*, 103: 18540-18545.

2. D. Tsygankov, A.W.R. Serohijos, N.V. Dokholyan, and T.C. Elston, "Kinetic models for the coordinated stepping of cytoplasmic dynein" (2009) *J. Chem. Phys.,* 430: 25101.

3. A.W.R. Serohijos*, D. Tsygankov*, S. Liu, T.C. Elston, and N.V. Dokholyan, "Multiscale approaches for studying energy transduction in dynein" *Submitted*. [*Equal contribution]

Cystic Fibrosis and CFTR

4. A.W.R. Serohijos, T. Hegedus, A.A. Aleksandrov, L. He, L. Cui, N.V. Dokholyan, J.R. Riordan, "Phenylalanine 508 forms interdomain contact in the CFTR structure crucial to folding and function" (2008) *Proc. Natl. Acad. Sci. USA*, 105: 3256-3261.

5. A.W.R. Serohijos*, T. Hegedus*, J. Riordan, and N.V. Dokholyan, "Diminished self-chaperoning activity of the ΔF508 mutant CFTR results in protein misfolding" (2008) *Public Library of Science Computational Biology*, 4: e1000008. [*Equal contribution]

6. T. Hegedus, A.W.R. Serohijos, N.V. Dokholyan, L. He, J.R. Riordan, "Computational studies reveal phosphorylation dependent changes in the unstructured R domain of CFTR" (2008) *J. Mol. Biol.*, 378: 1052-1063.

7. L. He, A.A. Aleksandrov, A.W.R. Serohijos, T. Hegedus, L.A. Aleksandrov, L. Cui, N.V. Dokholyan J.R. Riordan, "Multiple membrane-cytoplasmic domain contacts in CFTR" (2008) *J. Biol. Chem.*, 283: 26383-26390.

Protein Folding, Protein Engineering, and Ion Channels

8. Y. Chen*, F. Ding*, H. Nie*, A.W.R. Serohijos*, S. Sharma*, K.C. Wilcox*, S. Yin*, and N.V. Dokholyan*, "Protein folding: then and now" (2008) *Archives of Biochemistry and Biophysics*, 468: 4-19. [*Equal contribution]

9. J. Hao*, A.W.R. Serohijos*, G. Newton, G. Tassone, D.C. Sgroi, N.V. Dokholyan, and J.P. Basilion, "Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers" (2008) *Public Library of Science Computational Biology*, 4: e1000138. [*Equal contribution]

10. S. Ramachandran, A.W.R. Serohijos, L. Xu, G. Meissner, and N.V. Dokholyan, "Ryanodine receptor pore structure and function" *Public Library of Science Computational Biology*, In press.

Chapter 1

Introduction

In its most elementary sense, all processes that define biology are governed by the dynamics, structure, and interactions between biomolecules such as proteins, nucleic acids such as DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), and lipids. Processes involving these biomolecules are inherently multi-scale in time and length [1]. For example processes involving proteins range from 10^{-15} s (chemical reaction) to 10^{4} s (aggregation) and from 10^{-11} m (chemical bond) to 10^{-6} m (protein complexes) (Fig. 1.1) [2]. Understanding biology then entails understanding the structure, dynamics, and interactions between these molecules.

Thus in recent years, there is a compendium of modeling and simulation methodologies that aim to operate across the various time and length scales. The obvious objective of these approaches is to provide a coherent and consistent framework for describing and understanding complex phenomena.

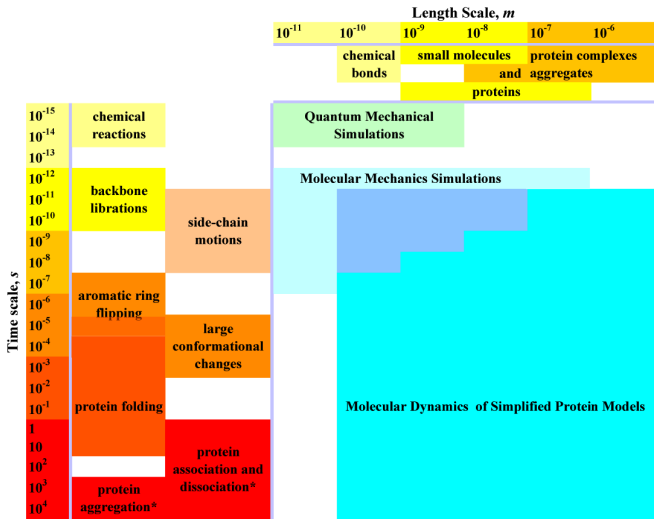

Figure 1.1. Time and length scales of processes involving proteins. The left side of the figure contains examples of processes occurring at various time scales. Protein association, dissociation, and aggregation (*) are concentration dependent and may span longer times than presented here. Examples of molecular sizes are at the top. Three simulation approaches – quantum mechanical, molecular mechanics, and molecular dynamics simulations with simplified protein model – outline the time and length scales accessible to these approaches. The time-length scales area, corresponding to molecular dynamics simulations, signifies a range of simplified protein models used in simulations, i.e. to access all outlined scales one may need to use a number of mutually-consistent simplified protein models. [Diagram is taken from [2]]

This work in particular develops and uses multi-scale modeling of protein dynamics and structure to investigate two outstanding problems in biology and medicine: (1) elucidating the structure and mechanism of dynein, a motor protein that is fundamentally involved in the active transport within cells and (2) understanding the structural basis of the misfolding of the CFTR channel, which eventually leads to cystic fibrosis.

We start in chapter Chapter 2 where we constructed a model of the motor unit of dynein, a motor protein that utilizes energy from hydrolysis to walk along cytoskeletal filaments, in particular the microtubule [3,4]. The dynein motor unit is the domain that hydrolyses ATP then generates force. We employed homology modeling to build the structure of the individual domains comprising the motor unit [5]. Then, to determine the

oligomeric organization of the structure, we fit the individual models of the motor unit to a low-resolution EM density derived from negatively stained electron microscopy [5]. These studies determined that the structure of the motor putatively forms an asymmetric heptamer, which may be important in generating the sequence of conformational changes during the motor's force production.

In Chapter 3, we elucidated the conformational dynamics of the motor unit that may be associated with its force generation. Using a simple analysis of potential protein conformations, we performed normal mode analysis of the structural model constructed from Chapter 2 [5]. This analysis show that the motor contains a mobile "rough" side, which incidentally is the non-catalytic site of the motor unit, and a less mobile "smooth" side of the motor, the site containing the catalytic parts of the motor unit. Molecular dynamics simulations of the motor unit using a simplified protein model corroborate these observations [6].

We move on to the next system under consideration, the CFTR channel, an ATP-binding cassette protein (ABC) that regulates ion transport in the apical membrane of epithelial cells [7,8]. The absence of a functional membrane in epithelial cells in the fundamental cause of cystic fibrosis (CF), the most common genetically inherited disease among people of European ancestry. Of the ~1500 mutations in *cftr* gene that are associated with cystic fibrosis, 90% of CF cases are attributed to the deletion of one single residue (Phe508) in the fisrt nucleotide-binding domain (NBD1) [9]. There are two outstanding hypotheses on how the Phe508 deletion leads to the disease: (1) the loss of the Phe508 backbone may shift a fraction of that mutant NBD1 off the wild type folding pathway, causing misfolding and eventual rapid degradation of the whole protein [10,11]; (2) the absence of the Phe508 side-chain prevents the correct post-translational assembly of all CFTR domains [12]. The detailed structural origin of the perturbed kinetics of NBD1 leading

either to the co-translational arrest or of the protein misassembly leading to post-translational arrest is unknown. In Chapter 4, we explored the structural basis of the folding kinetics defect induced in NBD1. By performing multiple folding simulations of wild type and mutant NBD1s, we identified the metastable folding intermediates and the folding pathways of both the wild type and mutant NBD1s [13]. We also defined a measure of the folding propensity for each of the NBD1 constructs. These analyses showed that indeed this difference in kinetics could be reproduced in simulations [13]. Moreover, from the structures of the intermediate states, we found that this difference in kinetics could be attributed to the conformation of specific loops in the nucleotide-binding domain, especially that of the so-called S6-H6 loop [13]. In fact, when this loop in NBD1-DF508 mutant was forced to that of the wild type conformation, we were able to partially "correct" the folding defect of the protein [13]. Preliminary experimental results likewise validate our model.

In Chapter 5, we addressed the misassembly of the ΔF508 CFTR mutant. We constructed a structural model of the whole CFTR molecule to identify the location of the Phe508 in the whole protein and to determine the specific interdomain interfaces it mediates. This interface is presumably perturbed upon Phe508 deletion [14,15]. From the model, we found that Phe508 in NBD1 interacts with the second membrane-spanning domain through the fourth cytoplasmic loop (CL4). This predicted interface and other interfaces in the model have been verified extensively using experiments [14,16].

Lastly, in Chapter 6, we synthesize the knowledge gained from the development of multi-scale models of protein structure and dynamics and its application to specific biological systems. We also provide an outlook of how this type of modeling may be applicable to other systems and what insights may be derived from studying them.

Chapter 2

Structure of the cytoplasmic dynein motor unit

Life signifies movement. At the cellular level, these movements are mediated by motor proteins that use the energy derived from ATP (adenosine triphosphate) to move the cell and transport materials *within* the cell. The cell is like a city that requires a network of roads and cars to transport people and cargoes to make the city sustainable (Fig. 2.1). Cytoskeletal motors (proteins that walk on actin or microtubule filaments, serve as "highways" within the cells) transport materials towards and away from the center of the cell. Cytoplasmic dynein in particular drives nearly all minus-end (towards the cell's center) directed microtubule-based movement in eukaryotic cells [17] (Fig. 2.1). Biologically, the function of dynein includes spindle formation and chromosome segregation and the transport of numerous cargoes including viruses, RNA, signaling molecules, and organelles [18]. Aberrant dynein function has been associated with various major human diseases such as schizophrenia, lissencephaly, and motor neuron degeneration [19]. Thus, elucidating the structure and mechanism of this fundamental motor protein is essential in understanding fundamental biological processes and the basis of major human diseases.

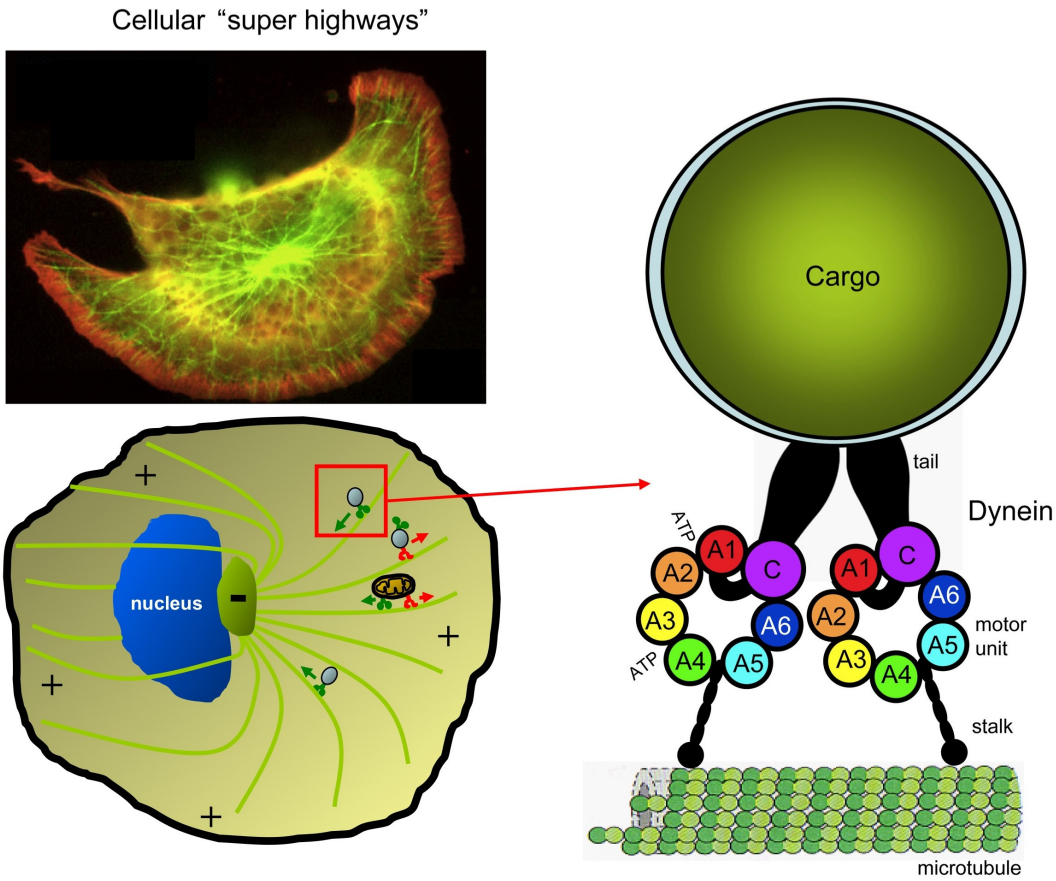

Figure 2.1. Cellular "super highways" and dynein. *(Upper left)* Image of a cell showing the cytoskeletal filaments actin (red) and microtubules (green). *(Lower left)* Simplified schema of a cell showing microtubules running from the cellular periphery towards the nucleus. Motor proteins such as dynein and kinesin walk along the microtubule transporting cellular substructures. *(Right)* Close-up view of a dynein dimer with a load walking towards the nucleus. A dynein monomer consists of a tail that binds to its load, a motor unit that hydrolyzes ATP and generates force, and a stalk that binds to the microtubule track.

Dyneins are multi-component complexes that are constructed around one to three heavy chains that contain the ATPase and motor activities. Dynein is a member of the ancient AAA+ (ATPases associated with diverse cellular activities) family of ATPases that includes a wide variety of proteins [20]. The heavy chain is composed of a *tail, which* binds to various cargos and other intermediate proteins, a microtubule binding *stalk*, and a *motor unit* that

binds and hydrolyzes ATP and putatively is the site for force production (Fig. 2, right panel). Sequence analysis of dynein's motor unit indicates that it consists of six concatenated AAA subunits (denoted as A1 to A6 in the schema), an extended stalk that contains a microtubule binding domain, and a C-terminal domain that is twice the size of an AAA subunit (Fig. 2.1, right panel) [20,21,22]. Although experimental studies using EM revealed significant insights into the structure of dynein, modeling the motor at the atomic level is essential in investigating its mechanism.

The stepping mechanism of single dynein has been explored by many groups using single molecule assays (Fig. 2.3). Studies of bead movement driven by cytoplasmic dynein in vitro suggest that single molecule dynein molecules are *processive*, that is, single molecules of dynein are capable of taking multiple steps along the microtubule track without detaching [23,24,25].

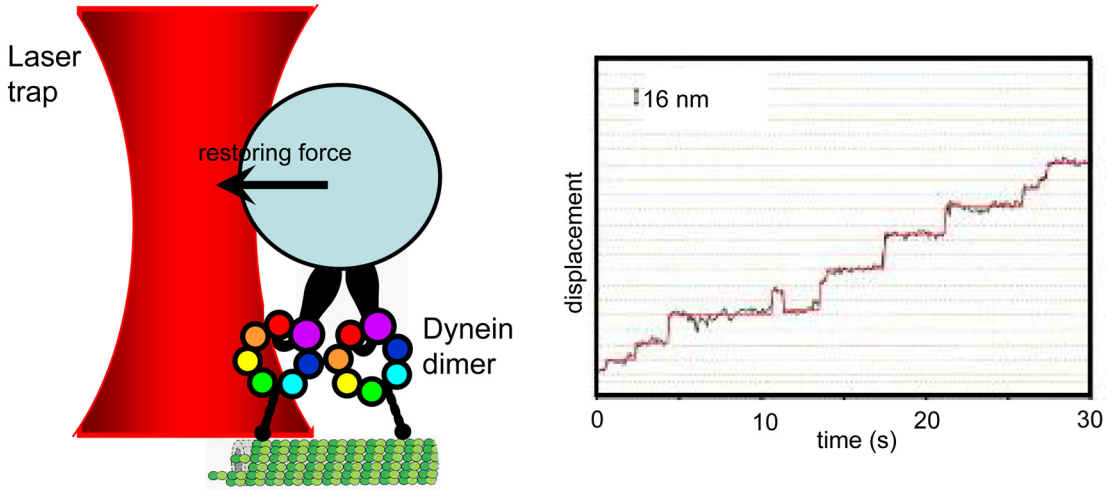

Figure 2.2. Dynein stepping behavior. *(Right)* Schema of single molecule optical trapping experiment, which is a standard tool employed in the mechanical manipulation of single molecules. *(Left)* Typical position *vs* time plots of single dynein moving along the microtubule. [Trajectory adapted from [25]].

The mechanism of other cytoskeletal motors such as myosin and kinesin, are better studied and more understood. It has been shown that nucleotide-driven conformational changes of their mechanism elements power the hand-over-hand stepping of their two identical motor domains [26,27]. In contrast, the mechanism of processivity in dynein is much less understood, and dynein's distinct evolutionary origin and structural features of this motor suggest that its mechanism differs considerably from other molecular motors. First of all, their stepping behaviors are already different. Using dynein molecules labeled with quantum-dots, it has been shown that dynein takes both small (~8 nm) and large (12-24 nm) step sizes with occasional backward stepping, which are rarely observed in Kinesin 1 or Myosin V.

In this chapter, we first address the structure of the dynein motor unit. The dynamics of the motor unit and the stepping of the motor unit is address in the next chapter.

2.1 Methods

To build a structure of the dynein motor unit, we first performed homology modeling of the individual domains that comprise the motor unit region. This individual domains need to be assembled together to finally determine the quaternary structure of the protein. We used an experimentally derived low-resolution electronic map of motor unit in determining organization of these various domains.

2.1.1 Homology modeling

Homology modeling (or comparative modeling) is a method for deriving all-atom theoretical models of protein tertiary structure by copying the topology of a related protein with a known structure, usually derived from x-ray crystallography or NMR (nuclear magnetic resonance) [28]. The fundamental assumption of the method is that proteins that

exhibit high sequence similarity (~30 % or greater) would most likely exhibit the same structure. This assumption is based on the observation that protein tertiary structure is better conserved than amino acid sequence [28]. While the structure derived from homology modeling cannot be as definitive as those derived from X-ray crystallography or NMR, the models are useful for generating hypothesis and directing experimental work.

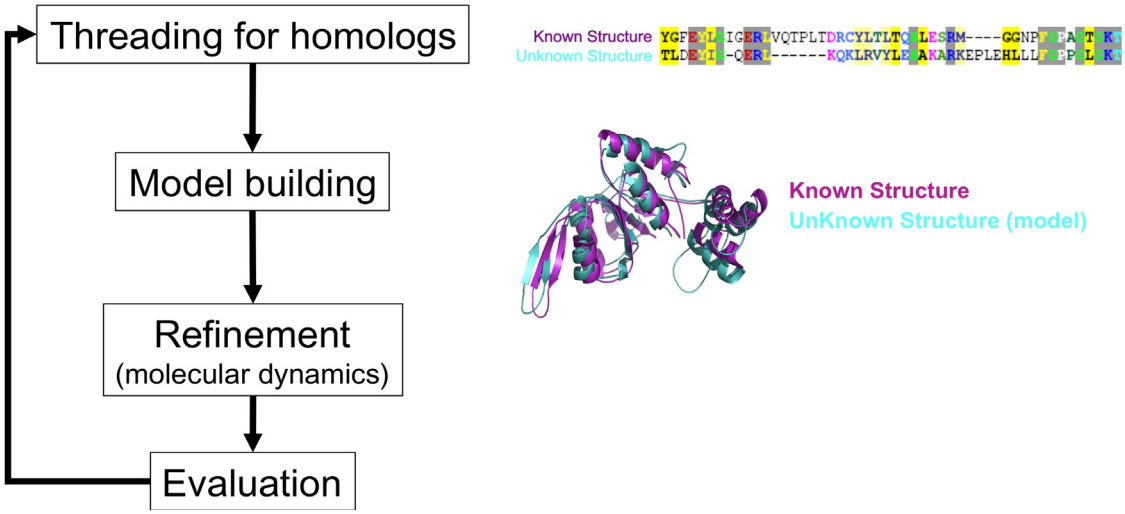

Figure 2.3. Homology modeling. To construct a model, the sequence of the protein of unknown structure is mapped to the sequence of the protein with known structure. The mapping optimizes the alignment of conserved residue regions *(left panel)*. An all-atom model is then constructed by copying the backbone topology (and when possible, the rotameric states of the side-chains) of the known structure *(left panel)*. The model structure is refined using molecular dynamic simulations. The quality of the model structure may be evaluated for correctness of backbone and side-chain geometry, exposure of charge residues, burriedness of hydrophobic residues, among other things. This set of procedures is performed iteratively until we arrived at a model of reasonable quality.

The sequence of the motor domain of cytosolic dynein heavy chain of slime mold *D. discodeum* (GenBank accession no. P34036) was submitted for threading to 3DJury (http://bioinfo.pl/meta)[29,30]. The templates used in building the AAA1, AAA2, and AAA4 were the Holliday junction migration motor protein RuvB from *Thermus thermophilus* HB8 (PDB ID code 1HQC; [31]), clamp loader gamma complex of *Escherichia coli* DNA

polymerase III (PDB ID code 1JR3; [32]), and eukaryotic clamp loader (PDB ID code 1SXJ; [33]), respectively. AAA3 and AAA5 were modeled from the same model used by AAA1. Similar to AAA2, AAA6 was built from the clamp loader. We constructed the atomic models using the Homology suite of INSIGHTII (Accelrys, San Diego, CA).

To construct a structural model for the C domain, we followed a protocol similar to the one used to construct models of the AAA subunits. We found that the C domain's first 290 residues consist entirely of α-helices, whereas the remaining 128-residue stretch includes five β-strands and terminates with a helix (Fig. 2.5). We then determined a family of candidate proteins that represent good structural templates for the two stretches of the C domain. Interestingly, the candidate templates for the first 290 residues were structures of the complement component C3d [Protein Data Bank (PDB) ID code 1GHQ], which attaches to foreign antigens during immune response [34]. Using the C3d fragment as template, the first 290 residues acquired a dome-shaped α–α toroidal fold [34]. The remaining five β-strands and last helix were built from the plecktrin homology (PH) domain of the Leukemia-associated RhoGEF (PDB ID code 1TXD) [35], which folds into a flattened seven-stranded β-barrel capped with a C-terminal helix. To obtain the complete structure of the C domain, the two subdomains were docked together using rigid body docking. The α-helical stretch shows higher homology with its template structure than the remaining β-strands suggesting that the function of the C domain is performed by the more conserved α-helical stretch.

2.1.2 Fitting all-atom models to EM-maps

To preserve functionally relevant interactions between domains AAA1–AAA4 and to construct a regular tetramer for this portion of the motor, we superimposed the models of these subunits onto the σ^{54} RNA polymerase activator NtrC1 (PDB ID code 1NY6) [36]. This

protein has a known homogenous heptamer structure consisting of AAA subunits with active catalytic sites. Next, we used the vector-quantization method implemented in SITUS [37,38] to fit the AAA1–AAA4 tetramer to the EM density (Fig. 2.4). To obtain a preliminary orientation of the remaining domains, the atomistic models of subunits AAA5, AAA6, and the C domain were fit separately to their corresponding electron density lobes. We also imposed the constraint that AAA5, IDR4, and AAA6 form a continuous peptide. Thus, AAA5 was oriented such that its C terminus faced the coiled coil. Similarly, AAA6 was oriented such that its N terminus faced IDR4 (Fig. 2.4).

Finally, to arrive at the complete model, we docked IDR4 and the rest of the inter-domain regions to the seven domains using a rigid-body docking protocol and shape complimentarity as criteria. When the complete atomic model was refit to the EM density of the entire motor unit, SITUS [38] ab initio identified the correct orientation of the domains with a correlation of 0.74 ($P < 10^{-316}$).

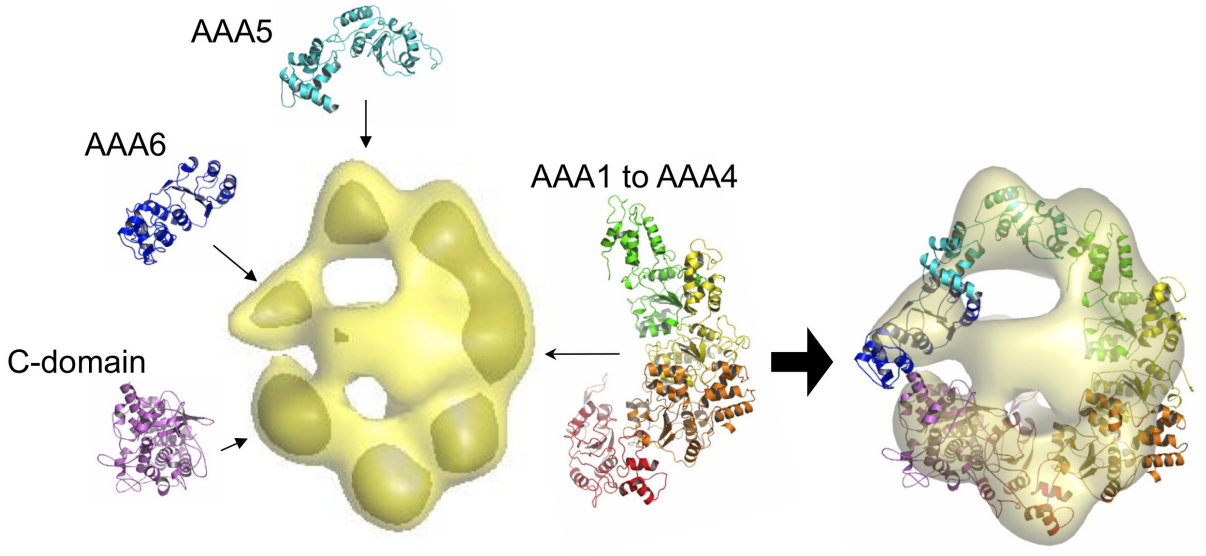

Figure 2.4. Fitting of all-atom models to low-resolution electron miscroscopy maps. All-atom models of the domains within the motor unit are fitted into their corresponding densities in the EM map to determine the quaternary organization of the domains.

2.2 Results

2.2.1 Models of the individual domains

We systematically constructed a complete structural model of the motor unit of cytoplasmic dynein from *D. discoideum*. The model includes the six AAA subunits, the linker regions that connect the subunits, and the C domain (Fig. 2.5). All AAA enzymes consist of two structurally conserved units: an α/β Rossman fold subdomain and an α-helical globular subdomain. Despite a <20% sequence identity between proteins in the AAA superfamily, up to 50% of equivalent C_α positions are within 2 Å rmsd [39]. To produce improved folds for the six AAA subunits of the dynein motor unit, we used 3DJury (Section 2.2.1) to search for candidate homologs of these subunits and the linker regions that connect them. 3DJury produces a consensus structure template based on the results of multiple independent structure prediction algorithms. The structural templates obtained in this way have consensus scores well above the confidence threshold of 50, which offers a prediction accuracy of 90%. Using these initial alignments, we then built atomic models of the AAA domains and their adjacent linkers using the homology modeling suite in Insight II (See Section 2.2.1). To evaluate the accuracy of the models for each subunit, we compared the local environments of the residues in the predicted structures to the population-averaged residue environments determined from known structures. The profiles score show the current model has better fold than earlier proposed models [21].

To determine the residues that form dynein's primary catalytic core, which is located between the first and second AAA subunits, we docked an ATP molecule to the glycine-rich P-loop (1969-GPAGTGKT-1976), which is the putative binding site for the nucleotide phosphate tail (Fig. 2.6). Within 5 Å of the docked nucleotide, we found conserved residues

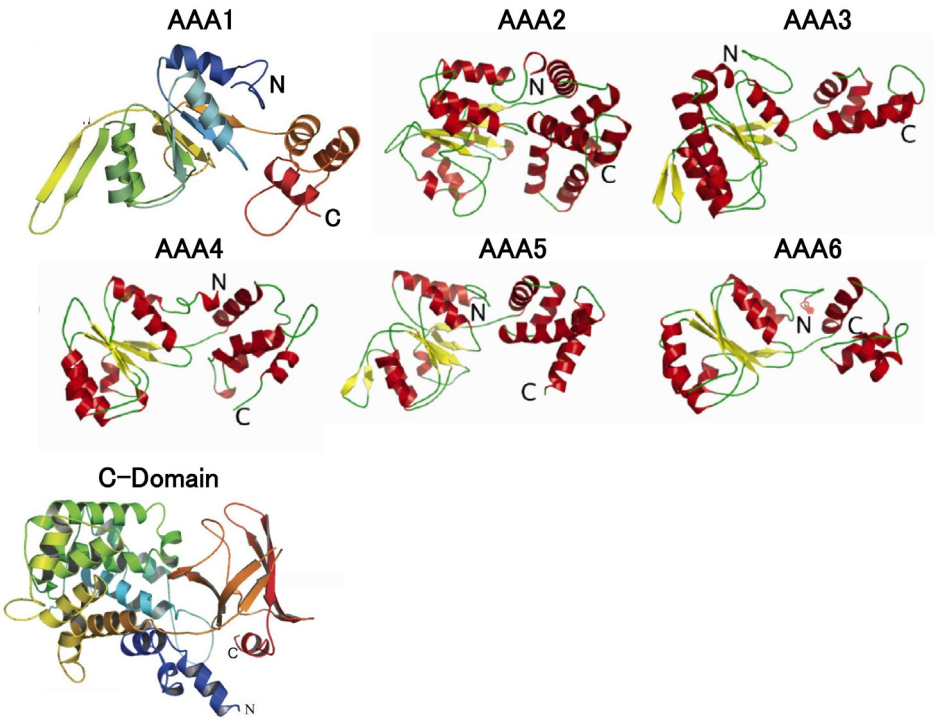

Figure 2.5. All-atom models of domains within the motor unit. Six AAA domains (ATPases associated with various cellular activities) and a C-domain comprise the motor unit. [Figure adapted from [5]]

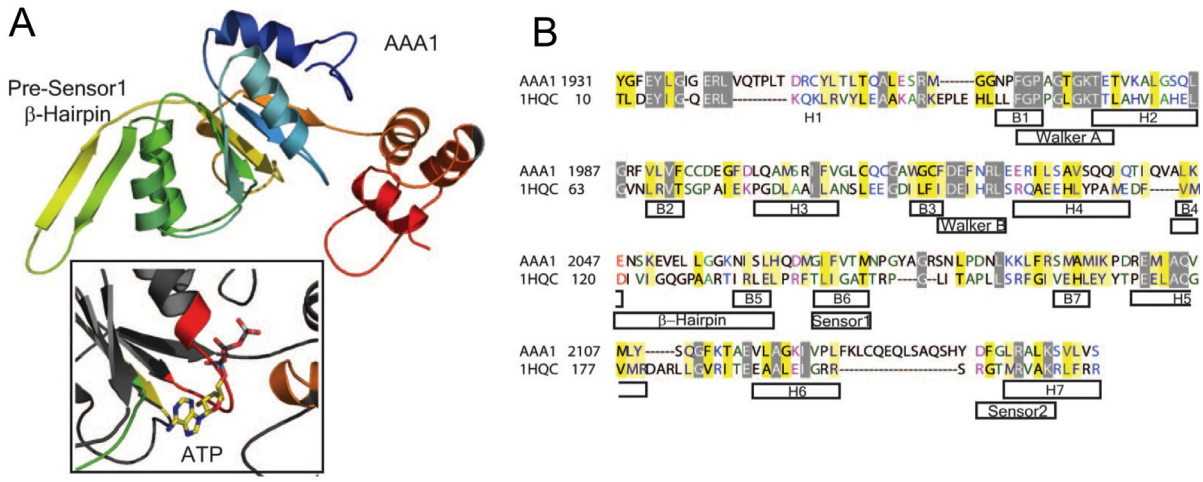

Figure 2.6. Dynein's putative catalytic site. (A) Canonical AAA fold (α/β-subdomain and all-α subdomain) with a synapomorphic pre-Sensor-1 β-hairpin insertion. (*Inset*) Putative site of primary hydrolysis with ATP docked. Walker A (i.e., P-loop) is red, Walker B is blue, Sensor 1 is yellow, and Sensor 2 is orange. **(B)** Sequence alignment between the *Dictyostelium discoideum* AAA1 and template structure RuvB domain (PDB ID code 1HQC)

in the Walker A and Walker B motifs that bind the β and γ NTP phosphates in all P-loop NTPases. These conserved residues found in dynein include K1975 in Walker A; D2021, E2022, and R2025 in Walker B; and R2145 in Sensor 2 (Fig. 2.6). These results are consistent with recent biochemical studies showing the dynein mutant K1975T trapped in a strong-binding state and devoid of motile activity[40].

Interestingly, the interdomain region between subunits AAA5 and AAA6 (denoted as IDR4) is 231 residues long, comparable with the size of an AAA unit (whereas the length of the other interdomain regions IDR1, IDR2, and IDR3 are 79, 103, and 92, respectively) (Fig. 2.7). If IDR4 possesses a globular fold, then it would manifest as an additional lobe in the reconstructed EM densities (Fig. 2.7) [41], and the motor would appear as an octamer. On the other hand, one of the densities on the face of the motor unit forms a long arch that spans the ring formed by the AAA subunits (Fig. 2.7), and is suggestive of a coiled coil. The IDR4 is sufficiently long to span the ≈8-nm facial density of the motor unit. Moreover, coil prediction algorithms assign a coiled-coil structure in the AAA5-IDR4 sequence, although the length of the predicted coil varies for dyneins from different species (Fig. 2.7). The search for structural homologs also resulted in several coiled-coil structures. On the basis of these results, we built IDR4 as a coiled coil using the cytoplasmic domain of serine chemotaxis receptor (PDB ID code 1QU7)[42] as a template.

The smaller lump on the face opposite the arching density could be the remnant of the dimerizing tail used in the EM studies. Another EM study [43] where the dynein tail has been labeled with antibody-Fab tag showed that the tag is not rigid and can be found at various positions around the planar ring. The study suggests that the tail domain docks into the center of the ring and that the tail sequence immediately adjacent to the docking point is flexible.

Thus, only the point of attachment near of the tail will exhibit a density because the flexible part will be averaged out, making the smaller facial density the more viable candidate for docking of the N-terminal tail.

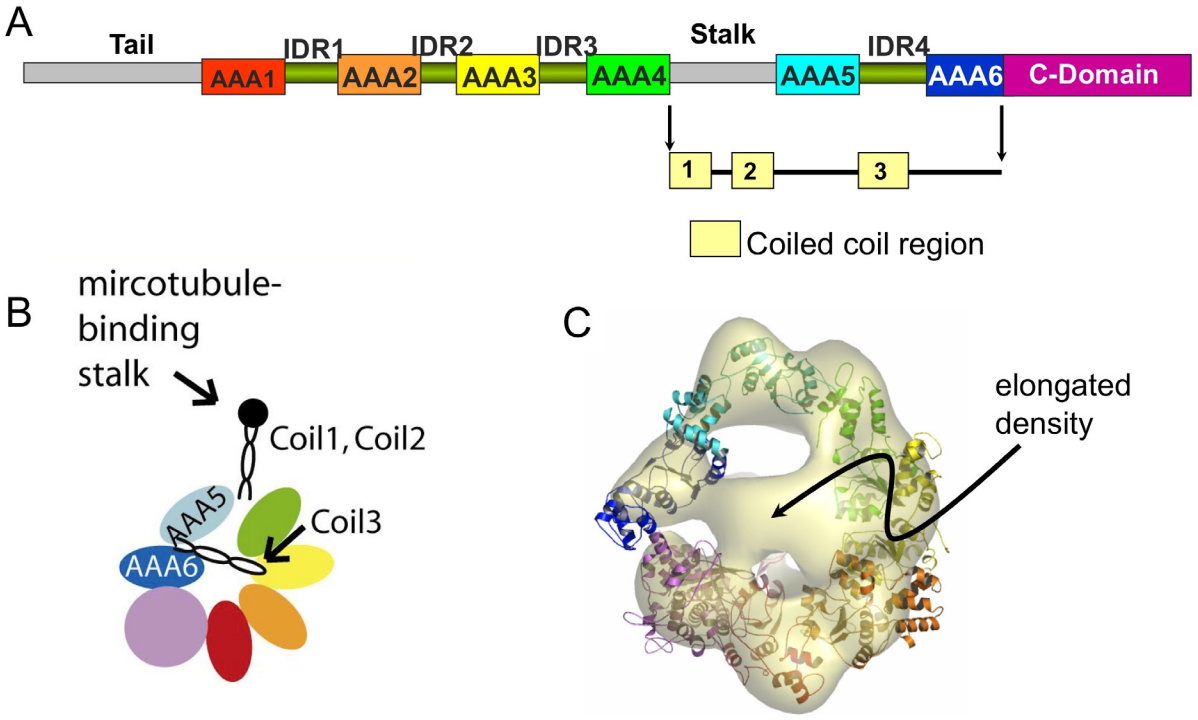

Figure 2.7. Potential coiled-coil conformation of IDR4. (A) Sequence map of dynein showing the various domains. Apart from the AAA units and the C-domains, there are also interdomain regions (IDRs) predicted to be primarily helical. There are three regions predicted to be coiled coils; the first two are already known to form the microtubule-binding stalk (panel B). We postulated that the third coiled-coil region correspond to the elongated density in the EM map (panel C).

2.2.2 Motor domain organization

The predicted structural model of the cytoplasmic dynein motor unit consists of six AAA domains and a C-terminal domain arranged in an asymmetric heptameric ring. The conserved AAA1-AAA4 domains form a tetramer that is organized similarly to other AAA homomer complexes. The less well-conserved AAA5, AAA6, and C-terminal domains constitute the rough side of the motor. This asymmetric organization of the motor complex is

consistent with the postulated evolutionary origin of the molecule in which the primordial homodimer pairs AAA1–AAA2 and AAA3–AAA4 combined to form a tetramer, with the subunits AAA5 and AAA6 representing later additions to the motor [44].

Another intriguing feature of the homology model is the IDR4 linker that connects subunits AAA5 and AAA6. The model predicts that this structure accounts for the observed density that spans the motor ring. In addition to contributing to the overall rigidity of the motor, IDR4 provides a route for force propagation from the rigid smooth edge of the motor where the nucleotide-binding sites are located to the flexible rough edge. Specifically, IDR4 extends from AAA5, which is at the base of the microtubule-binding stalk, to AAA3, whose nucleotide-binding pocket regulates the motor's processivity [40,44]. The IDR4 structure provides a clue to the important question of how distant functional sites communicate with each other to generate a coordinated mechano-chemical cycle.

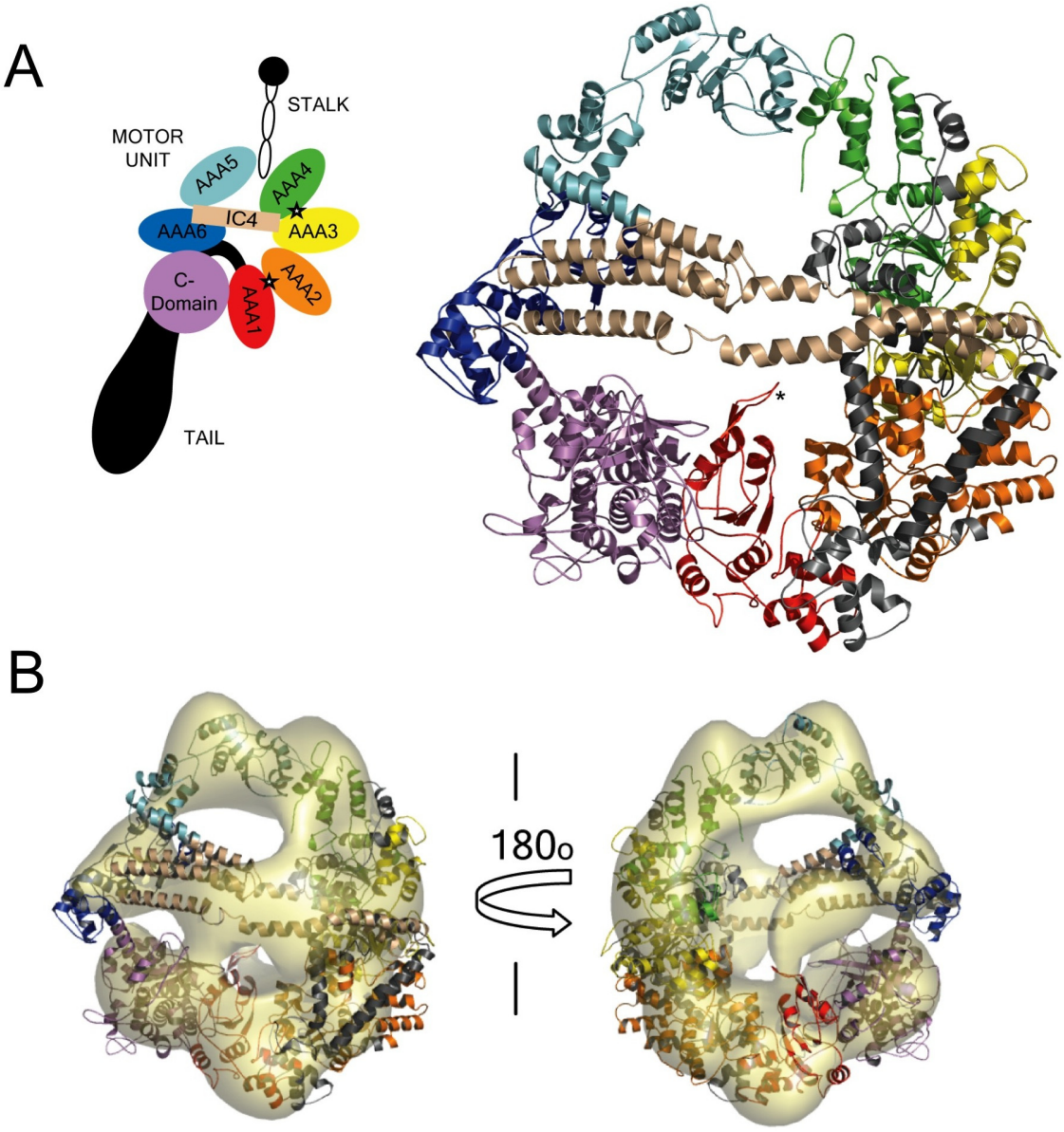

Figure 2.8. Complete model of the motor unit. (A) All-atom model of the motor unit. Domains are colored according to schema. **(B)** The model fitted to a low-resolution EM map [41]. [Figure adapted from [5].]

Chapter 3

Conformational dynamics of the dynein motor unit

To understand the mechanism of the dynein motor, we investigated the conformational changes that may be relevant with its force production. First, we used a simple normal mode analysis of the protein to determine the most likely dynamic conformations of the protein. Second, we performed molecular dynamics simulation using simplified models of proteins and a non-traditional molecular dynamics approach called molecular dynamics simulations. The results in this chapter have been described in two articles [5,6].

3.1 Methods and Models

3.1.1 Preliminary investigation of the motor dynamics using normal mode analysis

We performed normal mode analysis to establish the motor unit's dominant modes of motion. Normal mode analysis has been shown to accurately identify structural sites that function as pivots and, therefore, can be used to infer global motions of large molecular complexes [45]. Normal mode analysis also can be used to explore the intrinsic flexibility of molecular structures. In this analysis, the interactions between heavy atoms (Cα only) within 8 Å were approximated using a harmonic potential. Equations of motion were then computed by diagonalization of the Hessian matrix (mass-weighted second derivatives of the potential energy matrix) (Fig. 3.1). The eigenvalues of the matrix correspond to the mode frequencies and the associated vectors are the normal modes.

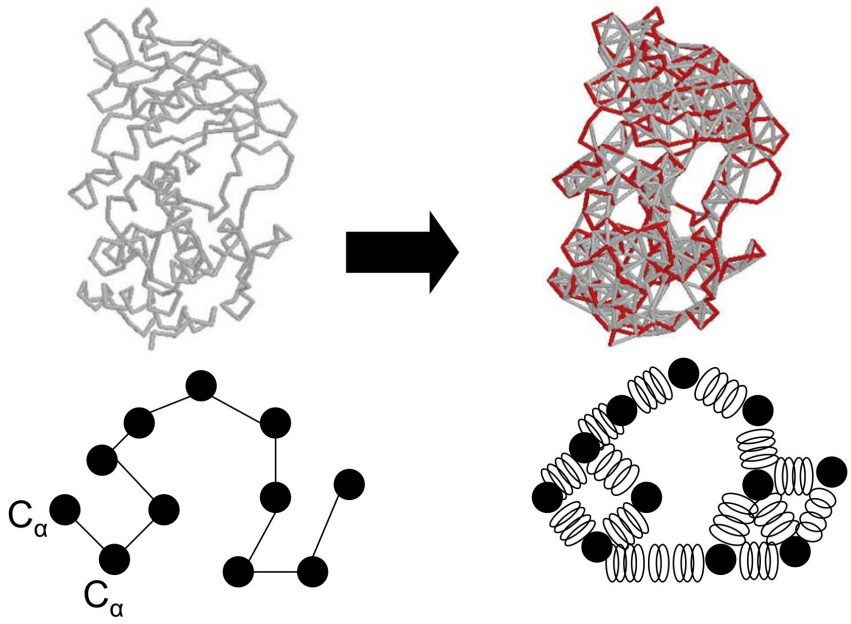

Figure 3.1. Normal mode analysis and elastic network model. In normal mode analysis, we find a set of basis vectors (normal modes) describing the molecule's concerted atomic motion and spanning the set of all 3*N* - 6 degrees of freedom. By modeling the interatomic bonds as springs and analyzing the protein as a large set of coupled harmonic oscillators, one can calculate a frequency of periodic motion associated with each normal mode, and then attempt to find normal modes with low frequencies. The low-frequency modes are indicative of the long time scale movement of the macromolecule.

3.1.2 Molecular dynamics simulation of the motor unit

To make the dynamics simulation of this large molecule tractable, we used a simplified protein model, a simplified interaction potential between atoms in the model, and a fast sampling algorithm called discrete molecular dynamics (DMD)[2,46,47].

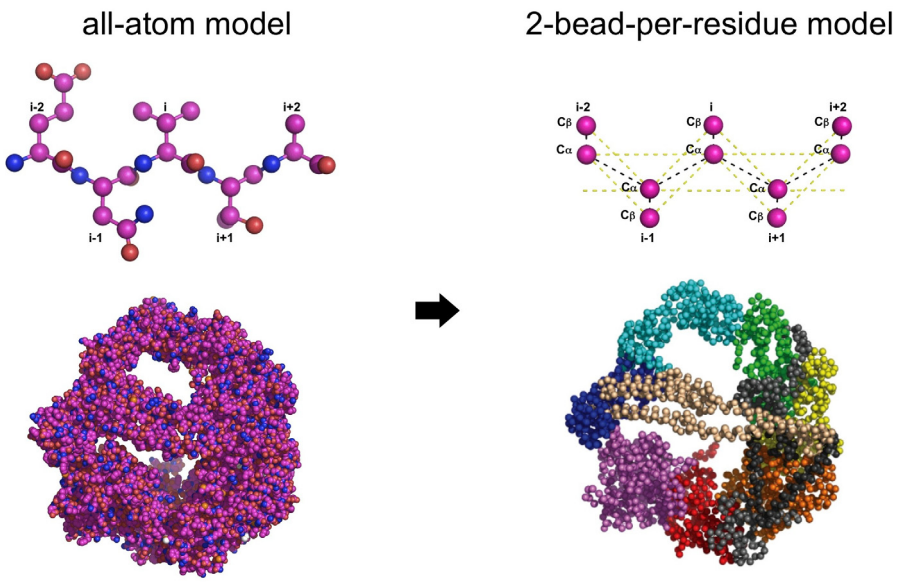

Figure 3.2. Simplified protein model.. The all-atom model of the dynein motor (Left) can be represented by by its Cα and Cβ atoms. Simulation of the simplified model enables the investigation of the motor dynamics over long time scales.

The protein model consisted only of Cα and Cβ atoms [48](Fig. 3.2). The interaction potential used in the simulation is the structure-based Go-interaction [49,50], where the residues that were proximate in the native state were assigned an attractive interaction, but those that were not were assigned a repulsive interaction. The total potential energy of a model protein was then $U = \sum_{i \neq j=1}^{N} U_{ij}$, where i and j denoted residues i and j, U_{ij} was the matrix of interactions

$$U_{ij} = \begin{cases} +\infty, & \left|\vec{r}_i - \vec{r}_j\right| \le a_0 \\ \varepsilon_{ij}\Delta_{ij}, & a_0 < \left|\vec{r}_i - \vec{r}_j\right| \le a_1 \\ 0, & a_1 < \left|\vec{r}_i - \vec{r}_j\right| \end{cases}$$

Here, a_0 was the hard core diameter, a_i was the maximum interaction distance between residues and ε_{ij} was the interaction strength between residue *i* and residue *j*, which set the energy scale. $||\Delta_{ij}||$ was a matrix of contacts with elements $\Delta_{ij} = 1$, if $\left|\vec{r}_i^{NS} - \vec{r}_j^{NS}\right| \le a_1$ and $\Delta_{ij} = -1$, if $\left|\vec{r}_i^{NS} - \vec{r}_j^{NS}\right| > a_1$, where $\vec{r}_i^{NS}$ was the position of the *i*th residue in the native conformation. We penalized the non-native contacts by imposing $\varepsilon_{ij} < 0$. Temperature units were taken in terms of the typical value of interaction strength ε_{ij} divided by the Boltzmann constant k_B, i.e., in units of ε_{ij}/k_B.

The strength of the interaction between residues in contact ε_{ij} defines the energy units. Physically, $\varepsilon_{ij} \approx$ 1–2 kcal/mol, which is approximately the contribution to protein stability from a hydrogen bond. The time unit (tu) is estimated to be the shortest time between particle collisions in the system (~0.1 ns).

The evolution of this simplified protein model with simplified atomic interactions was calculated using DMD. In contrast to traditional molecular dynamics which employs continuous potentials, the DMD algorithm uses discretized square well potentials [2,46,47], thus all particles move at constant velocity until the before the soonest collision. That the state of the system is necessarily updated only in the event of a collision enables DMD to access the long time scale dynamics of large proteins.

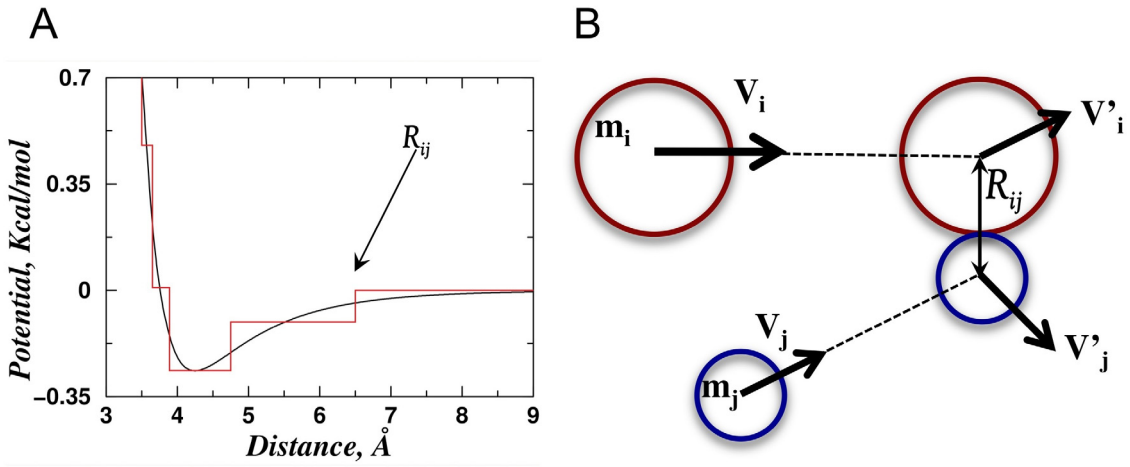

Figure 3.3. Discrete molecular dynamics. **(A)** Interaction potential between atoms are discretized (red) as opposed to being continuous in traditional molecular dynamics simulations (black). **(B)** Because of the discretized potential, the evolution of the system is now driven the collision between particles and entails the calculation of momentum and energy conservation equations.

3.2 Results

In the final model of the motor unit, the interdomain regions and AAA units form a compact backbone (Fig. 2.8). The most closely packed part of the motor consists of its smooth side where the nucleotide-binding P loops are found. We hypothesized that this compact structure is essential for efficiently transducing forces generated at the ATP hydrolysis site to the extended stalk that contains the microtubule binding domain and is located between subunits AAA4 and AAA5. To investigate this possibility, we performed normal mode analysis to establish the motor unit's dominant modes of motion (Fig. 3.1 and 3.4). Fig. 3.1 illustrates atomic displacements associated with the three lowest frequency vibrational modes. The frequencies of modes 2 and 3 are 1.28 and 1.56 times larger than that of mode 1. From Fig. 3.1, it is evident that the most mobile domains are AAA5, AAA6, and the C domain, whereas AAA1–AAA4 form a more compact structure. These observations are

made quantitative in Fig. 3.4(B), which lists the RMSD of the Cα atoms of each subunit for the first three normal modes.

In mode 1, the AAA5 subdomain exhibits an upward motion, whereas AAA6 partially rotates about the IDR4 linker (Fig. 3.4(C)). On the other hand, AAA1 to AAA4 and their linkers exhibit minimal displacement. Interestingly, AAA5 is positioned at the base of the stalk that interacts with the microtubule. The fact that the dominant motion of the lowest frequency normal mode occurs at the base of the stalk suggests that the stalk tilts during the motor's power stroke. Mode 2 is characterized by a "squeeze" applied to subunit AAA5 and the C domain coupled with an outward motion by AAA6 (Fig. 3.4(C)). Similar to mode 1, in mode 2, AAA1–AAA4 and their linkers exhibit minimal movement. EM 3D reconstructions of the motor unit with stalks positioned at 0°, 25°, and 45° relative to vertical show greatest variation in electron densities corresponding to subunits AAA5 and AAA6 [41] (Fig. 3.4(C)). The direction and magnitude of the domain displacements determined for modes 1 and 2 are consistent with these observations (Fig. 3.4(C)). For example, the motion predicted to occur in modes 1 and 2 is consistent with the reorientation of subunit AAA6's density observed for different stalk positions (Fig. 3.4(C)).

Using the simplified model described above, we extensively characterized the dynamics of dynein. First, we performed discrete molecular dynamics (DMD) simulations for 10^6 time units (approximately a few milliseconds) with initial temperatures from T=0.1 e/k_BT to T=2.0 e/k_BT (see section 2.1.2). These equilibrium simulations allowed us to determine the thermal denaturation curve of the dynein head and the melting temperature. Using the observation that the native state of the protein is slightly below the melting

temperature, we then performed molecular dynamics simulations near the identified melting point.

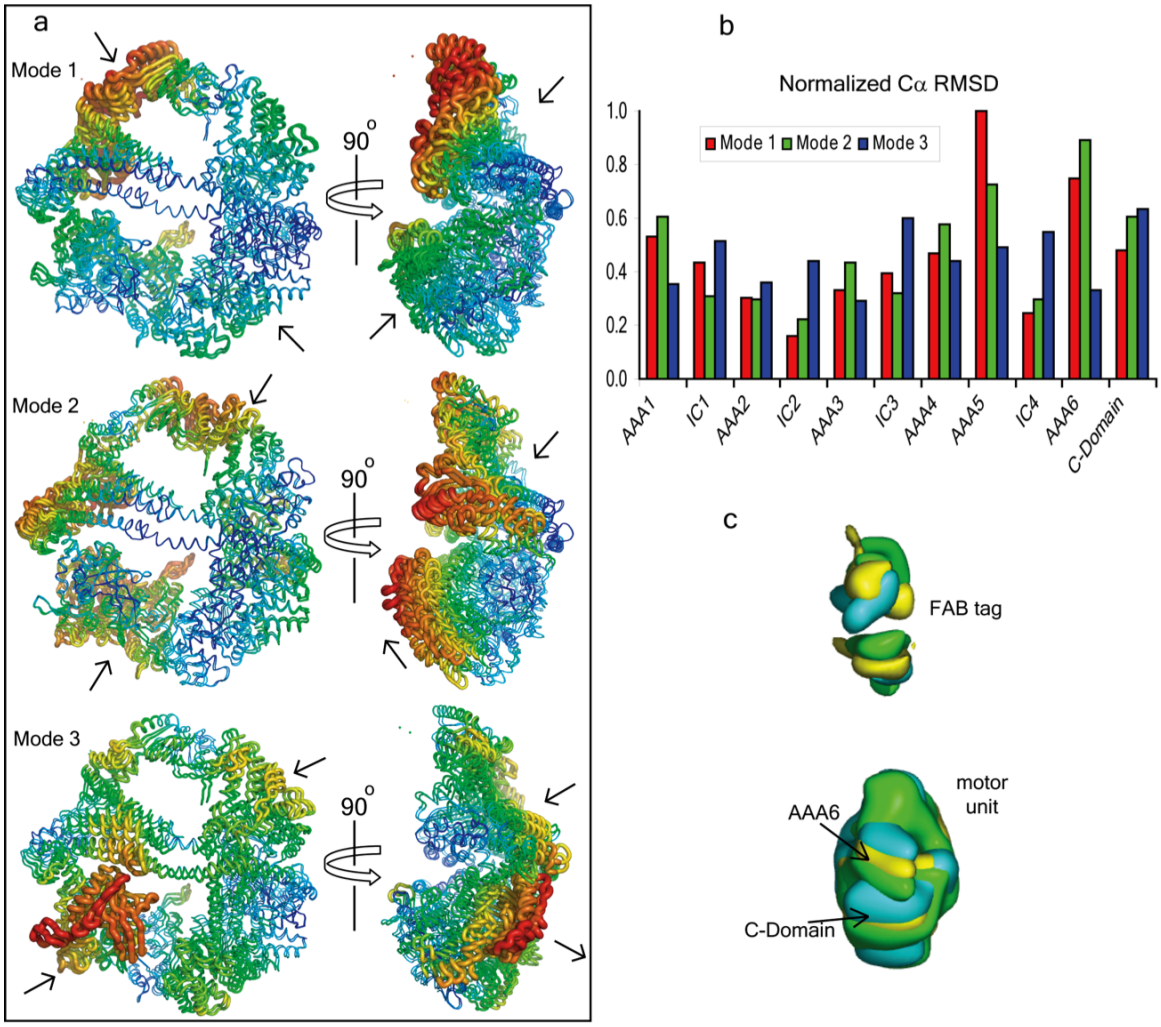

Figure 3.4. Lowest frequency normal modes of dynein motor unit. **(A)** Superposition of two structures displaced in opposite directions along the normal mode. The size of the backbone is proportional to fluctuations of the Ca atoms. Arrows indicate the directions of the dominant vibrations. AAA5, AAA6, and the C-domain exhibit the most prominent variation in domain architecture in the three normal modes. **(B)**, Average rmsd of the C_a in a domain, normalized by the largest displacement and weighted by inverse frequency. **(C)** Superposition of reconstructed 3D structures of the motor unit in three distinct stalk conformations fromEM studies by Samso and Koonce [41]. In the three stalk positions, the side formed by AAA5, AAA6, and the C-domain exhibit the largest variation, in agreement with normal mode calculations. [Image adapted from [5]]

To quantify the fluctuations of all the domains, we calculated the per residue root mean square deviation (RMSD) with respect to the initial structure. The average fluctuations of residues within a particular domain are shown in Fig. 3.5. In agreement with the normal mode analysis, we found the "rough" side of the motor composed of AAA5, IDR4, AAA6, and C-domain exhibits the largest fluctuations, whereas the "smooth" side, which is composed of the AAA1 to AAA4 is a more compact structure. Interestingly, the ATP-binding domains are located on the smooth side, suggesting that only minor conformational changes in the catalytic binding pocket are induced upon hydrolysis or product release, however, these conformations are then propagated to and amplified by the rough side of the motor.

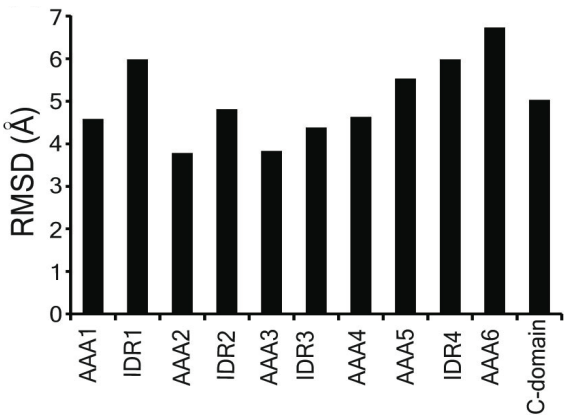

Figure 3.5. Averaged domains fluctuations from equilibrium molecular dynamics. The per residue root-mean-square deviation with respect to the starting conformation was calculated over the equilibrium simulation run.

3.3 Summary

Our analysis of the three lowest frequency normal modes indicates that large-scale motions of the motor primarily involve movements in subunits AAA5, AAA6, and the C domain, whereas subunits AAA1–AAA4 function as a rigid structure. This finding is

consistent with recent observations from EM reconstructed structures [41,43,44]. We speculate that the subunits AAA1–AAA4 provide the motor with a stationary backbone against which forces generated in the primary catalytic site can act. This generates conformational changes that propagate sequentially through the C domain, AAA6 and AAA5 and terminate with a movement of the microtubule-binding stalk (Fig. 3.6).

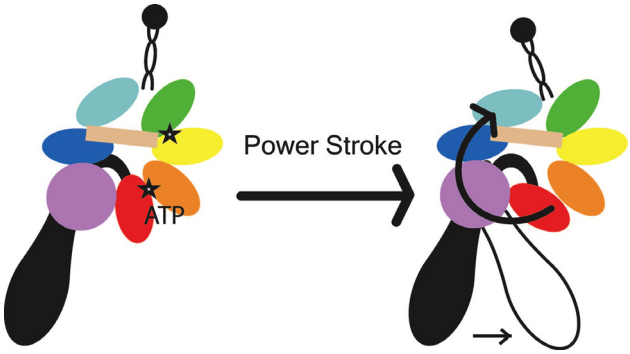

Figure 3.6. Model of power stroke. Binding ATP or release of ADP_Pi in the hydrolytic sites (indicated by stars) induces conformational change that is primarily propagated through the C domain, AAA6, and AAA5. These domain reorientations cause the stalk or tail to flex about the junction that connects them to the motor unit, thus generating the power stroke.

There are three current models for dynein's power stroke. In the first model, ATP causes a rotation of both the stalk and the tail about the junctions that connect them [21,27,41]. The second model assumes that a conformational change of the tail swings the motor unit and the stalk together [44]. Lastly, the third model assumes that a flexible structural linker between the motor unit and tail bends upon coordinated conformational rearrangements of the AAA domains [51]. From our structural model and normal mode analysis, model 2 is unlikely because of the large motions in AAA5 to which the stalk is docked. We propose the possible conformational rearrangements of the domains movements within the motor unit that is the basis of either model 1 or model 2: Binding or release of

ATP or ADP induces conformational change in the catalytic domain between AAA1 and AAA2. Because of the rigid structure formed by subunits AAA1-AAA4, the disturbance is propagated in a clockwise direction through the C domain, AAA6, and AAA5, causing the microtubule-binding stalk to flex. The change in the angular position of the stalk possibly alters the microtubule binding affinity of the stalk's globular tip. These conformational changes may also play regulatory role, consistent with the findings in enzymatic studies of dynein domain fragments suggesting that the stalk autoinhibits ATP or ADP release in AAA1 and AAA3, and that the C domain also affects the ATPase activity [51].

In a recent cover article in the journal *Cell* [52], a new EM study of tagged and truncated dynein constructs showed that the ring-like architecture of the motor unit only consists of six domains. Contrary to our model and earlier experimental results, the C-domain is not an integral part of the ring. With this revised architecture of the motor unit, the model of energy transduction proposed in this chapter needs to be revised accordingly. The revision of the model is an endeavor in the immediate future.

The issue of whether the proposed structure of the IDR4 is indeed coiled coil or not, and whether it spans the motor ring or not, is still a point of contention in the field. This issue may be resolved by higher resolution structural studies of the motor unit.

Chapter 4

Misfolding of mutant CFTR NBD1 domains

CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) is an ATP-binding cassette (ABC) protein found in apical membranes of epithelial cells (Fig. 4.1). It is a chloride channel involved in the regulation of salt secretion and reabsorption. CFTR is a multidomain, integral membrane protein containing two transmembrane domains, two nucleotide-binding domains (NBD1 and NBD2), and a regulatory region (R domain) (Fig. 4.1(B)). The absence of a functional CFTR channel in the plasma membrane is the fundamental cause of the cystic fibrosis (CF), the most common genetically inherited disease among populations of European descent. CF patients have altered epithelial ion transport that leads to decreased hydration of epithelia in the gut, kidney, pancreas, and airways (Fig. 4.1(B)) [7,8,53]. Decreased surface liquid volume in the airways impairs mucociliary clearance which in turn leads to respiratory bacterial infection [54,55]. Chronic pulmonary damage caused by bacterial infection dramatically decreases patients' life expectancies.

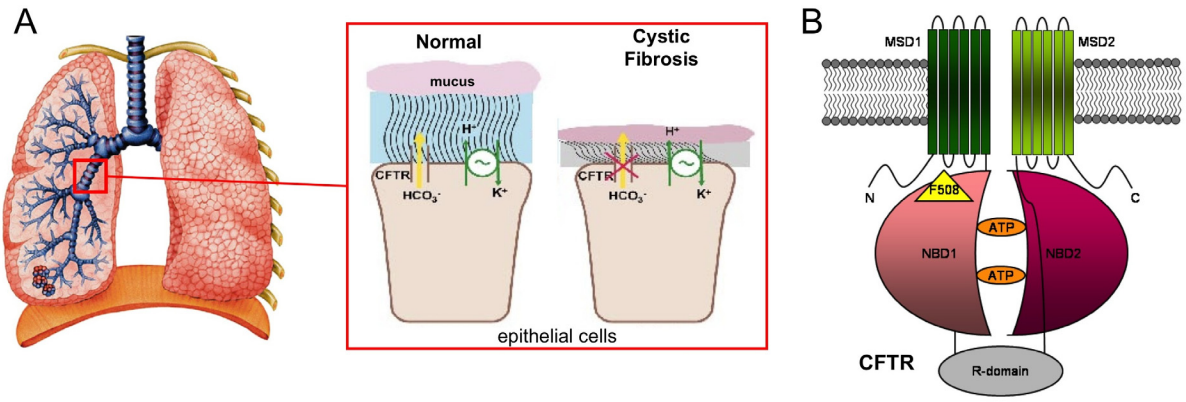

Figure 4.1. CFTR and cystic fibrosis. **(A)** CFTR is a chloride channel found in apical membranes of epithelial cells. The absence of a functional CFTR channel in the epithelial cell membranes leads to hydration of the airway surface layer, eventually leading to the cystic fibrosis. **(B)** CFTR is an ATP-binding cassette protein consisting of membrane-spanning domains (MSD), nucleotide-binding domains (NBD), and a regulatory region (R domain). The deletion of a single residue Phe508 in NBD1 is associated with ~90% of CF cases.

Although more than 1,500 mutations are known in CFTR [56], approximately 90 % of CF patients carry the allele with the deletion of the codon for phenylalanine at position 508 (Phe508)[57], which is located in the first nucleotide-binding domain (NBD1) of CFTR (Fig. 4.1(B)). Experimental studies suggest that the CFTR-ΔF508 (CFTR protein with deleted Phe508) may be arrested at two stages during its biogenesis. First, the loss of the Phe508 backbone may shift a fraction of that mutant NBD1 off the wild type folding pathway, causing misfolding of and eventual rapid degradation of the whole protein (Fig. 4.2), 1st quality control checkpoint) [10,11]. Second, that the absence of the Phe508 side-chain prevents the correct post-translational assembly of all CFTR domains (Fig. 4.2, 2nd quality control checkpoint) [12]. The detailed structural origin of the perturbed kinetics of NBD1 leading either to the co-translational arrest or post-translational arrest is unknown. Understanding the molecular basis of CFTR's arrest at these two quality control checkpoints is essential in the development of therapeutic treatment.

Despite the extensive research was done in recent decades to find a cure, only symptomatic treatments are currently available. Since the detailed molecular mechanism of the CFTR function and the effect of the mutations are not known, moreover the structure of the full length CFTR channel remains to be solved, drug discovery has been limited to high throughput screening assays. The biggest outstanding question in cystic fibrosis is the molecular basis of the fast degradation of the CFTR protein with the most prevalent ΔF508 mutation. The answer could accelerate rational CF drug development.

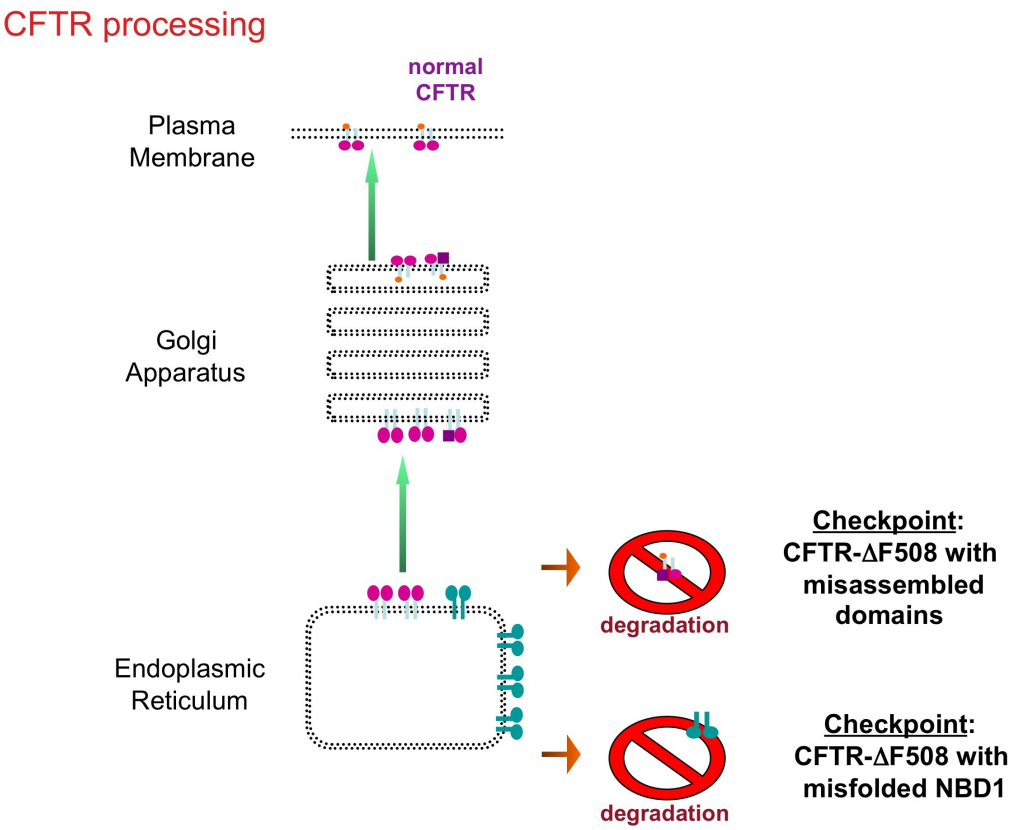

Figure 4.2. Arrest in the processing of the CFTR-ΔF508 mutant. Synthesis of proteins within the cells is tightly regulated by cellular quality control systems, a process that is conceptually akin to a manufacturing production line. CFTR is synthesized in the endoplasmic reticulum and transported to the plasma membrane at the surface of the epithelial cell. The defective CFTR-DF508 may be targeted for degradation at two stages: (1) the deleted Phe508 residue shifts NBD1 off the wild type folding pathway and (2) the missing Phe508 prevents the packing of NBD1 with MSD2 resulting in a misassembled protein.

In this chapter, we explore the structural basis of the misfolding of NBD1 mutants. In general, when a protein proceeds from the unfolded state to the native state in the multidimensional free energy landscape (conceptual cartoon shown in Fig. 4.3), it accesses metastable folding intermediates along the way. The sequence of intermediate states accessed by the protein defines its folding pathway. This folding pathway may be perturbed in the case of mutants, which we hypothesize to be the case for CFTR NBD1.

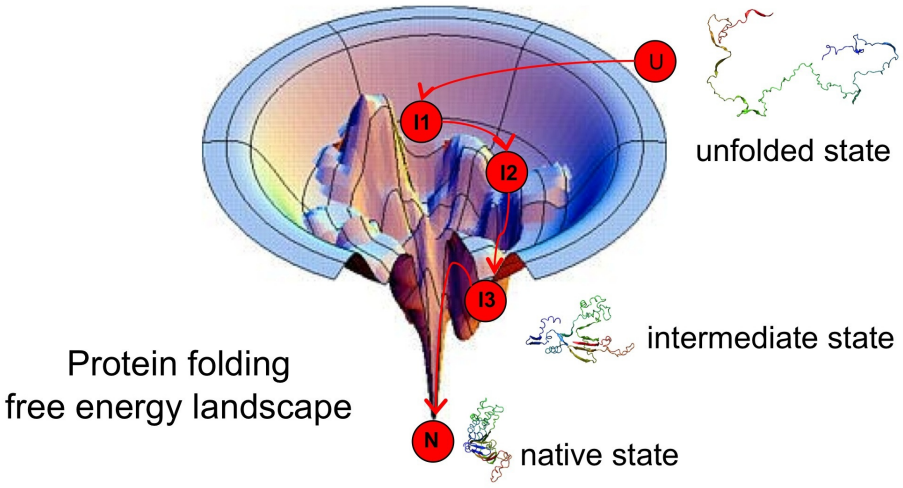

Figure 4.3. Protein folding energy landscape. As the protein proceeds from the unfolded state to the native (folded) state, it accesses metastable folding intermediates along the way. This native folding pathway may be perturbed when the protein is mutated.

Using molecular dynamics simulation of a simplified protein model of a single NBD1, we recapitulated the observation that there is no significant difference in the thermodynamic stabilities of the wild type and mutant [11]. This recapitulation of experimental observation points to the validity of the protein model. Next, by performing multiple folding simulations, we constructed the folding pathways of the wild type and

mutant NBD1s, and showed that indeed these pathways are different. We also showed that this difference could be attributed to the conformation of some loop regions in the NBD1.

4.1 Methods and Models

4.1.1 Simplified model of a protein

To access time scales of NBD1 folding, we used a simplified protein model that still maintained important features of the protein such as side-chain packing. Amino acid residues were modeled as follows: (1) glycines are represented by three beads (-N, C_α, C′); (2) phenylalanine, tyrosine, tryptophan, and histidine by five beads (-N, C_α, C′, C_β, C_γ), and (3) all other residues by four beads (-N, C_α, C′, C_β) [58]. This protein model has been successfully employed in studying protein aggregation [58]. In the simulations, we used the available crystal structures of wild type and mutant NBD1: wild type (PDB ID 2BBO), ΔF508 mutant (PDB ID 1XMI) and F508A mutant (PDB ID 1XMJ)[59,60]. The missing loop between E403 and L436 in both wild type and mutant NBD1 is reconstructed using a loop-search algorithm in SYBYL (Tripos Assoc. Inc, St. Louis, MO).

4.1.2 Simplified interaction using the Go-model and discrete molecular dynamics

To determine the long-range interaction between the particles in the simplified protein model, we used the Go-model described in Section 3.1.2. In this particular Go-model, two residues are said to be in contact if their heavy atoms are within a distance of 4.5 Å. To calculate the evolution of the system, we also used discrete molecular dynamics as described in Section 3.1.2.

4.1.3 Equilibrium simulations protocol

Using discrete molecular dynamics (Section 3.1.2), long equilibrium simulations at various temperatures were performed to investigate the equilibrium dynamics of the CFTR NBD1. The primary objective was to compute the thermal denaturation plot and the heat capacity of the protein. From these plots, we can compare the thermodynamic stabilities of wild type and mutant NBD1 domains.

From long equilibrium simulations of 10^6 time units (tu), we were able to access the long time-scale dynamics of the CFTR NBD1 in the order of 0.5 ms. Each equilibrium simulation consumed approximately 300 CPU hours.

4.1.4 Folding simulations protocol

We performed 300 folding simulations for each NBD1-WT, NBD1-F508A, and NBD1-ΔF508. Starting from fully unfolded chains, the temperature of the system was progressively reduced to allow NBD1 to fold to its native structure. Folding simulations proceeded until $\tau_{max} \sim 60{,}000$ tu, which was chosen to be longer than the typical folding time of the studied sequences [61]. A similar criterion was employed in the studies calculating the folding probability of other proteins [62]. The NBD1 structure in a particular folding run was considered folded it satisfied the following criteria: (1) its energy was less than or equal to -620ε (the energy of the native state), (2) its structure was within 2.5 Å RMSD relative to the native conformation, and (3) the structure possessed correct topological wiring of the secondary structure elements.

The folding probability of either wild type or mutant NBD1 was calculated as

$$\text{Folding Probability} = \frac{\text{Number of successful folding trials}}{\text{Total number of folding trials}}$$

To estimate the error in folding probabilities, each folding trajectory was considered a Bernoulli trial with a binary outcome, folded or unfolded. The variance σ of this Bernoulli process was then $\sigma^2 = p(1-p)/n$, where p was probability and n was the total number of trials.

4.1.5 Analysis of folding simulations

Identification of metastable folding intermediate states

For each folding trajectory that successfully folded the NBD1, we calculated energy probability distribution, which is simply the normalized histogram of the energy over the folding simulation time. The peaks of this normalized energy probability distribution is indicative of metastable folding intermediate states. To identity the dominant intermediate states for the wild type and mutants, a sum of multiple Gaussian curves $\sum_i a_i \exp\left[\left(x-b_i\right)^2 / c_i^2\right]$ was fitted to the average energy probability distribution of successfully folded runs. The parameters a_i, b_i, and c_i were the center, standard deviation and height of the *i*th Gaussian curve, respectively.

Structural characterization of intermediate states

Because of the reduction in dimensionality of the folding process when energy was used as a reaction coordinate, each intermediate state, as defined above, represented an ensemble of NBD1 structures. To identify the primary structural characteristics of each intermediate state, we clustered the structures in the corresponding state and calculated the frequency of contacts formed between pairs of residues. For a particular structure, an $n \times n$ contact matrix was constructed from the n residue NBD1. The value of the cell *(i,j)* was 1

when residue *i* and *j* were in contact (within 4.5 Å) or 0 otherwise. Dominant contacts between residue pairs were then determined from the average contact matrix of all the structures within an intermediate state.

Kinetic accessibilities of the intermediate states and most likely paths

We estimated the probability of transition between states by counting the trajectories that underwent such a transition. The sum of probabilities of the paths emanating from a given state was normalized to 1, which physically meant that the system always exited from its then current intermediate state.

The transition probabilities represented independent conditional probabilities, thus the probability of a sequence of paths to be taken from the unfolded state to the native state was estimated by multiplying the probabilities of the traced edges. The sequence of edges connecting the unfolded and folded state with highest probability was considered the most likely folding pathway.

4.2 Results

4.2.1 Equilibrium dynamics

To determine the equilibrium dynamics and stabilities of the wild type and mutant NBD1, we performed equilibrium simulations (10^6 time units ~ 0.5 ms) of wild type and mutant NBD1 using discrete molecular dynamics (see Methods, Section 4.1.3). From the equilibrium simulations, we calculated the thermal denaturation curve of both NBD1-WT and NBD1-ΔF508 (Fig. 4.4) and observed two stable thermodynamic states, folded and unfolded. In agreement with previous experimental studies by denaturation experiments [10,11], the stabilities of wild type and ΔF508 NBD1 were not significantly different. The slope at the transition temperature of the wild type ($T_m \sim 0.68\ \varepsilon/k_B$) was $9.8 \times 10^3\ k_b$ and the

slope at the transition temperature of the mutant ($T_m \sim 0.70\ \varepsilon/k_B$) was $1.6 \times 10^3\ k_b$ ($\varepsilon \sim 1$–2 kcal/mol and k_B is the Boltzman factor; see Section 3.1.2 for further discussion on units). This shift in slope at the transition temperature indicated a difference in folding cooperativity of NBD1-WT and NBD1-ΔF508 and therefore a difference in folding kinetics.

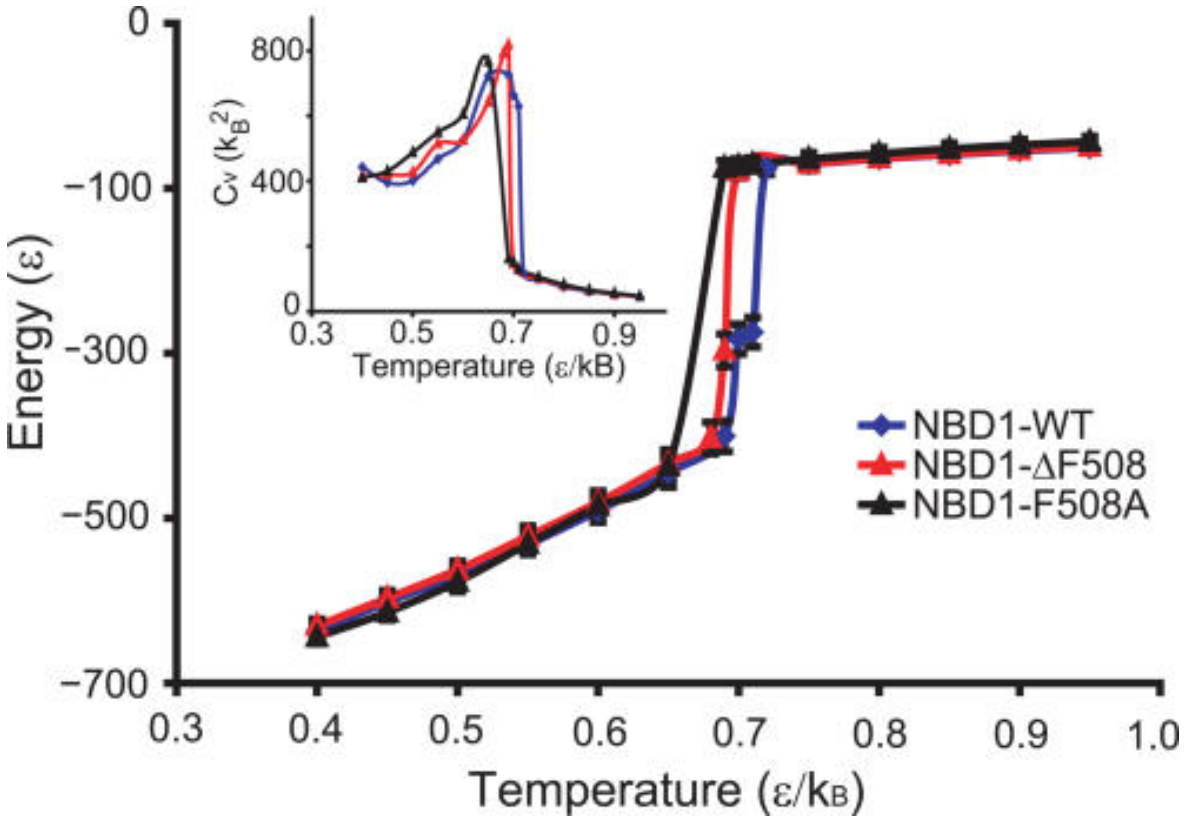

Figure 4.4. Thermodynamics of NBD1-WT, NBD1-F508A, and NBD1-ΔF508. Average equilibrium energy was calculated from long equilibrium simulations (10^6 time units) of NBD1-WT, NBD1-F508A, and NBD1-ΔF508 crystal structures. Error bars represented ± standard deviation. (Inset) The specific heat is calculated as $C_v = (\langle E^2\rangle - \langle E\rangle^2)/T^2$. Wild type and mutant NBD1s exhibit similar thermodynamic states but different dynamics near the folding transition.

4.2.2 Difference in wild type and mutant NBD1 folding propensities

Folding is a stochastic process, thus to investigate in detail the difference in folding kinetics and dynamics of NBD1-WT and NBD1-ΔF508, we performed 300 folding simulations on each of the structures (Section 4.1.4). Starting from fully unfolded chains of NBD1-WTand NBD1-ΔF508, we progressively reduced the temperature of the system to simulate thermal folding. We found that the folding probability [61] of wild type to be 33 ± 3% while that of the mutant was 13 ± 2%. The ratio of NBD1-WT and NBD1-ΔF508 correlated with the ratio of their folding yields derived from folding experiments. Folding yields of NBD1-WT was approximately twice that of NBD1-ΔF508 in the temperature range 10°C to 22°C [11]. Folding simulations of our control structure NBD1-F508A yield a folding probability of 26 ± 4% which was intermediate to that NBD1-WTand NBD1-ΔF508. This folding probability value was in agreement with experimental studies showing intermediate folding efficiencies and maturation levels of NBD1-F508A relative to NBD1-WT [10,11].

4.2.3 Folding pathways

To investigate the molecular origin of the difference in folding yields and probabilities, we mapped the folding pathways of NBD1-WT, NBD1-F508A, and NBD1-ΔF508 by identifying their metastable folding intermediate states. The folding intermediate states of a folding trajectory were exhibited as peaks in the energy probability distributions (Fig. 4.5). Thus, dominant intermediate states in the folding pathways were peaks in the average energy probability distributions (Fig. 4.6). The average energy probability distributions of wild type and the mutant were significantly different (Kolmogorov-Smirnov test; P-value$<1.4\times10^{-292}$), which suggested a significant difference in the folding kinetics of wild type and mutant NBD1. The average fraction of native contacts of NBD1 structures in

an intermediate state follows a distinct distribution (Fig. 4.7), thus, an intermediate state identified using energy as the folding reaction coordinate, forms a distinct collection of NBD1 conformations.

To determine the difference between the sequence of folding events of the wild type, F508, and the F508A control, we estimated the probability of transitions between intermediate states (Fig. 4.8). The difference in transition probabilities of NBD1-WT, NBD1-F508, and NBD1-F508A is shown in Fig. 4.9. The transition probabilities showed some states accessible only to either wild type or mutant NBD1. The difference in state accessibilities between the two suggested a difference in contact pattern formation (nucleation events), which could cause the observed difference in folding yields.

We also calculated the most dominant folding pathways in wild type and mutant NBD1. The most dominant path in wild type follows a sequence of transition Unfolded→S10→S8→S7→S5→S4→S1, while the dominant path in the mutant follows the sequence of transitions Unfolded→S9→S8→S7→S6→S4→S1. Thus, NBD1-WT and NBD1-F508 undergo different sequences of folding events.

4.2.4 Structural modulators of folding kinetics

Because of the reduction in dimensionality of the folding process when energy was used as a reaction coordinate, each intermediate state represented an ensemble of NBD1 structures. To identify the primary structural characteristics of each intermediate state, we clustered structures in the corresponding state and calculated the frequency of contacts formed between pairs of residues (Fig. 4.11). In all intermediate states, we found the most notable structural difference between NBD1-WT and NBD1-F508 occured in the S7-H6 loop. For example, P574 interacted with Q493 in wild type but not in the mutant. Also, F575

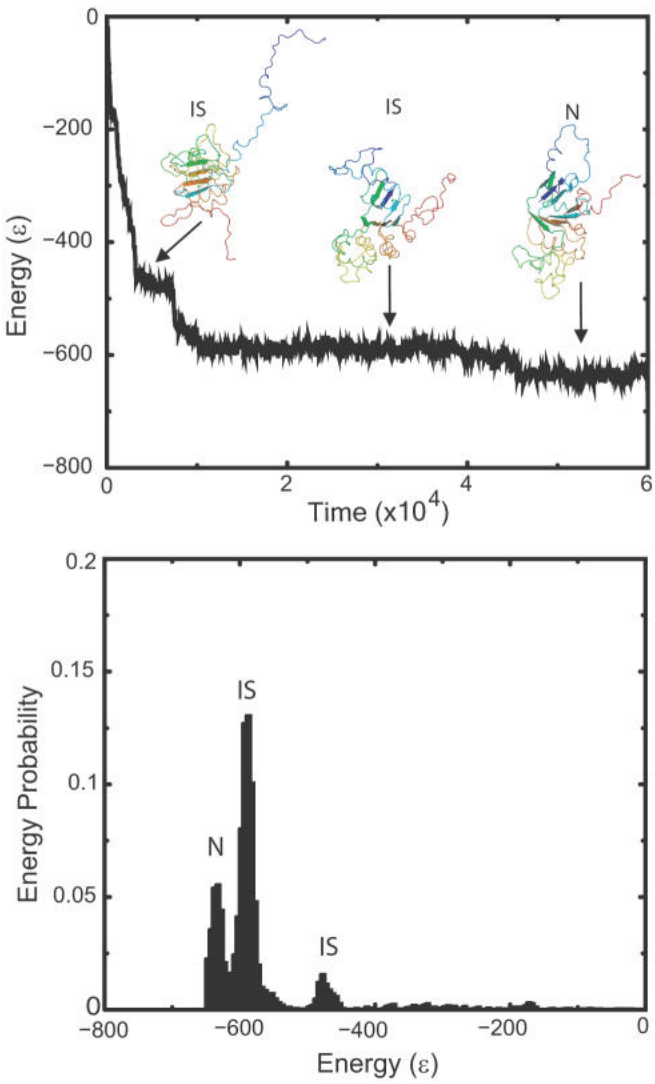

Figure 4.5. NBD1 folding. (Upper panel) We performed annealing simulations in which the temperature was decreased to facilitate the folding of NBD1. Shown is a time evolution of energy starting from unfolded state to the native (N) state. As the protein proceeds towards it native state, it accessed metastable folding intermediate states (IS). (Lower panel) In a normalized energy probability distribution, these intermediate states are observed as peaks.

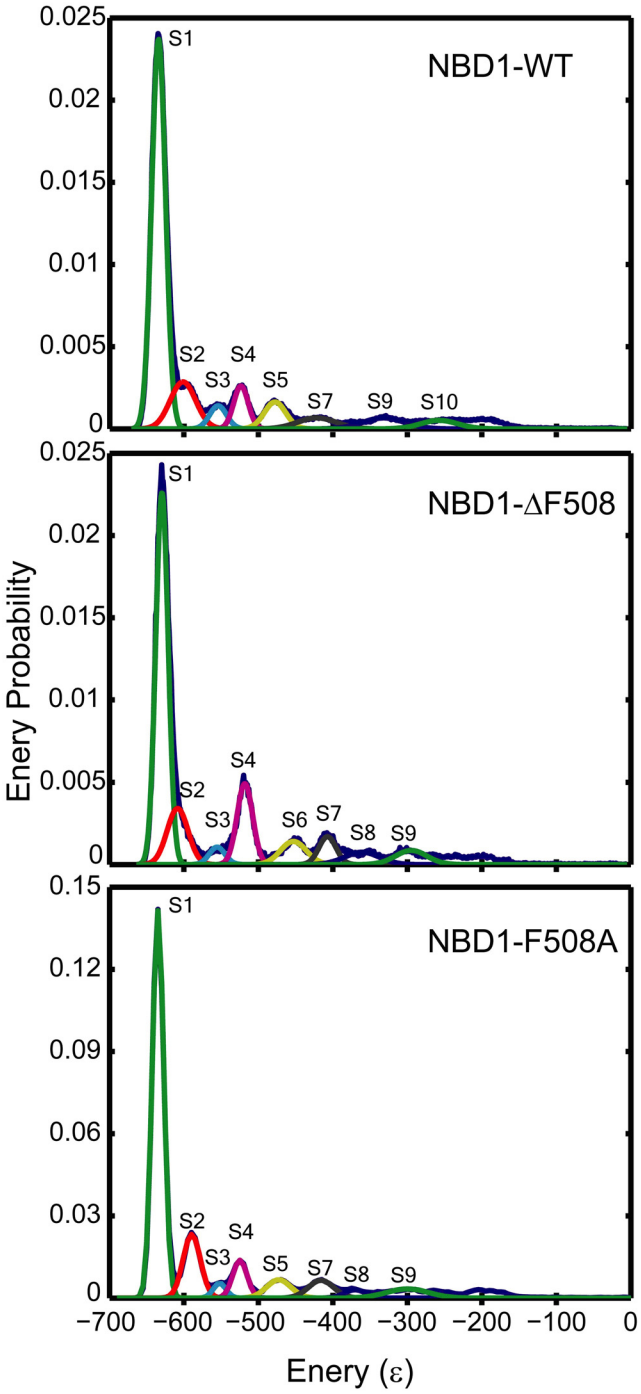

Figure 4.6. Energy probability distributions averaged over all successful folding trajectories. Positions of metastable intermediate states were identified by fitting a sum of Gaussian distributions. Each Gaussian curve corresponded to a putative folding intermediate state.

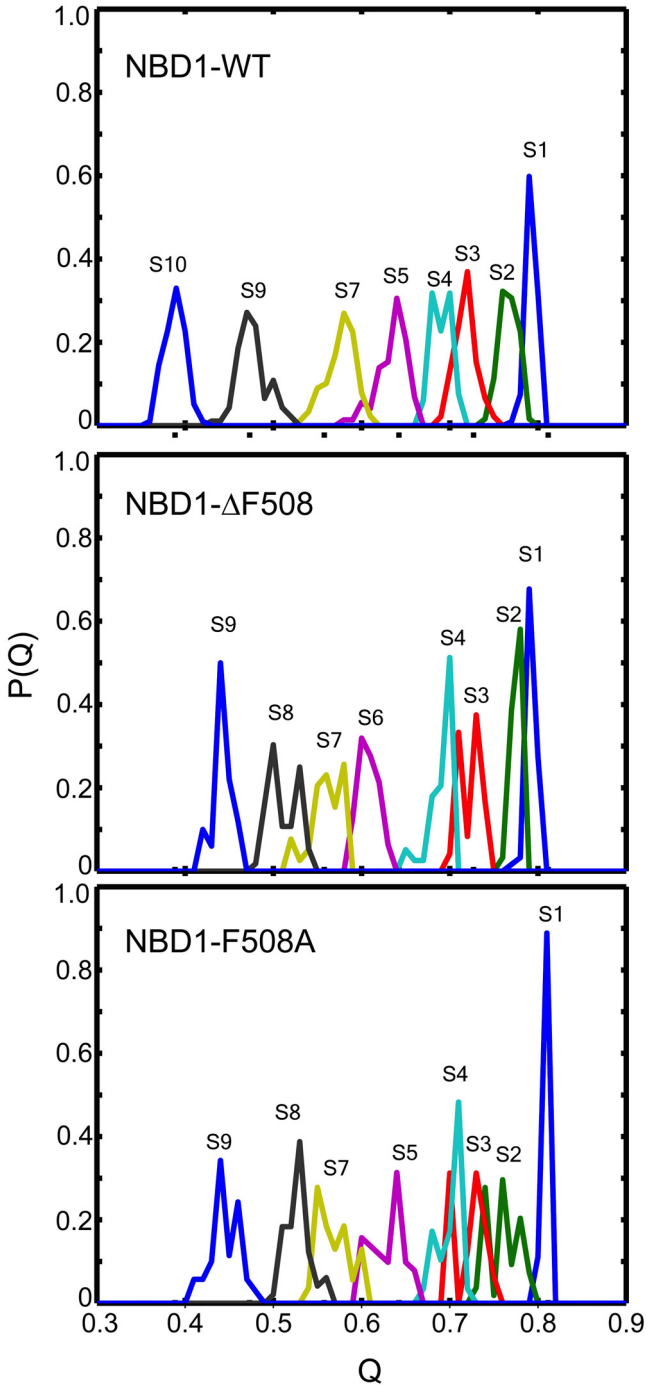

Figure 4.7. Distribution of fraction of native contacts. For a given intermediate state, we calculated the average fraction of native contacts (Q) coming from a particular folding trajectory. The normalized distribution of Q shows that the states defined using energy are structurally distinct.

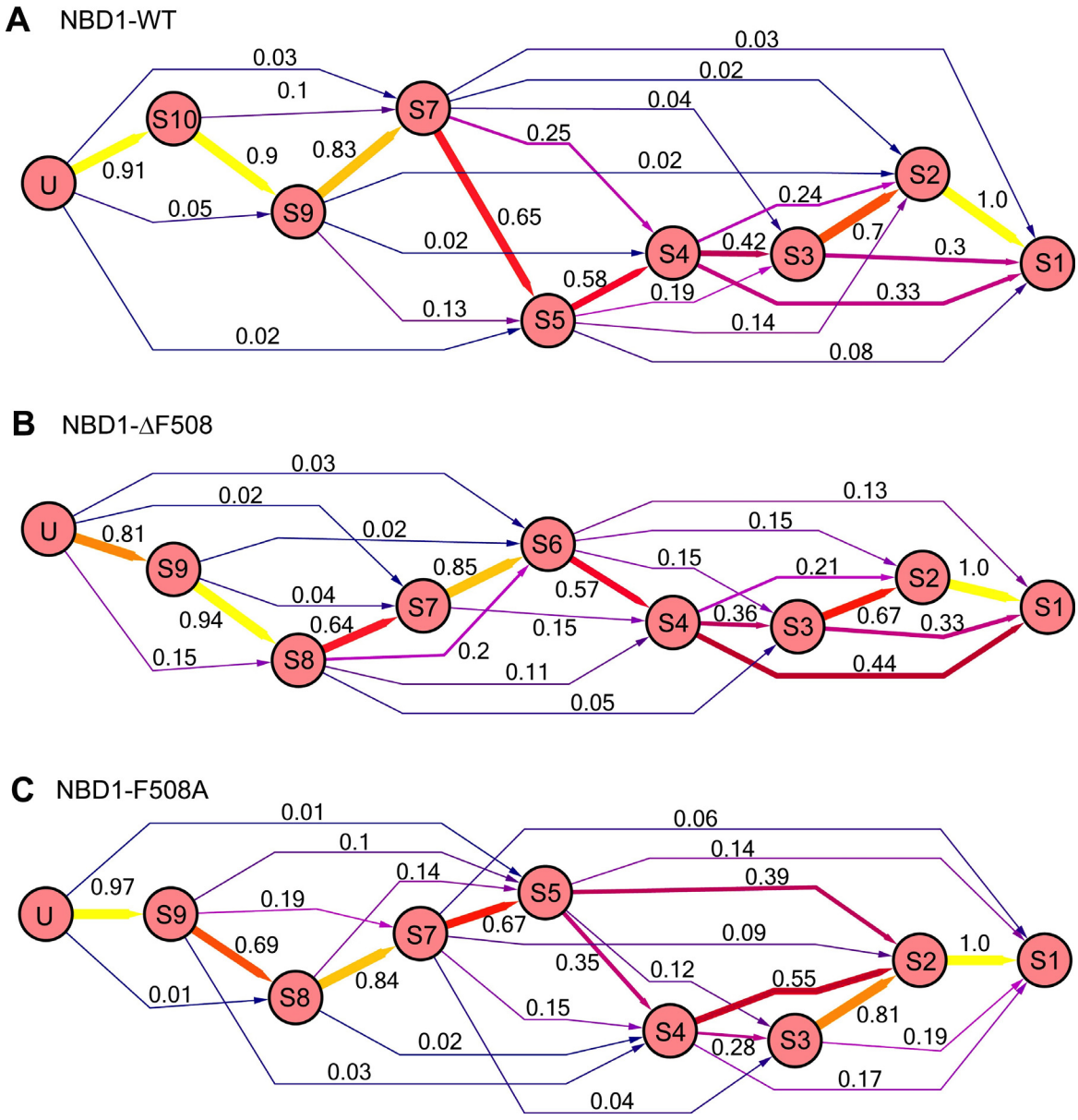

Figure 4.8. NBD1 folding pathways. Probability of kinetic transitions between intermediate states of NBD1-WT, NBD1-ΔF508, and NBD1-F508A. The probability of exiting a state is normalized to 1. The thickness and warmth of the transition edges are rendered proportional to the probability value.

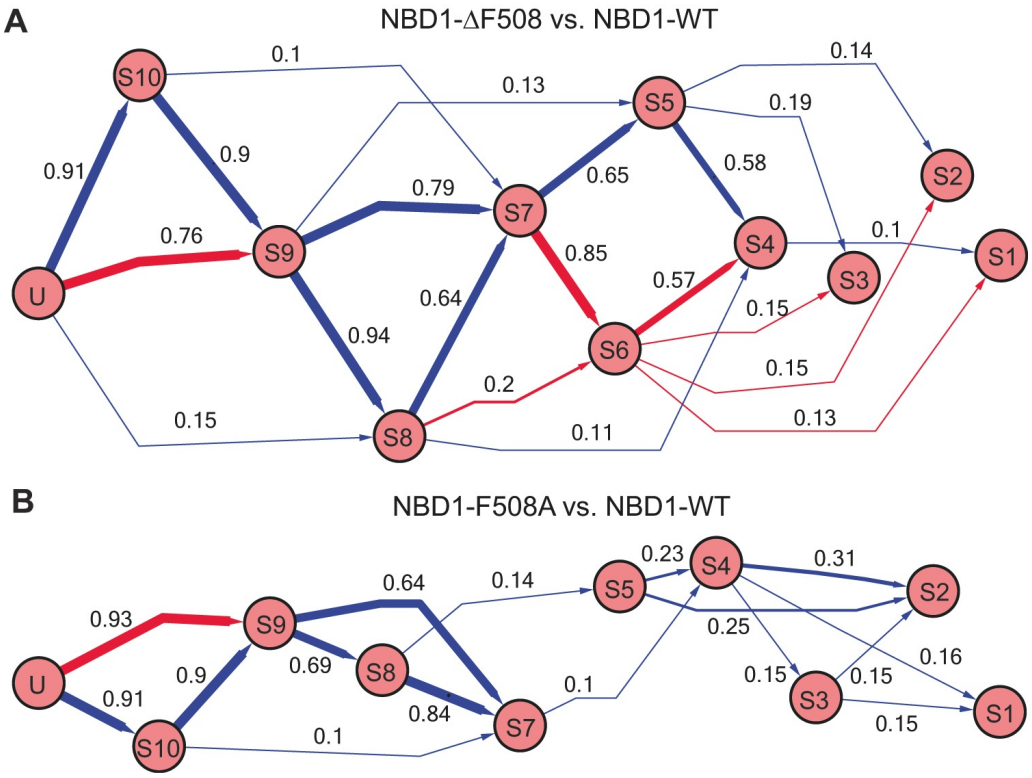

Figure 4.9. Comparison of the folding pathways of wild type NBD1 and its mutants. Shown are the difference transition probabilities between **(A)** NBD1-ΔF508 and NBD1-WT and between **(B)** NBD1-ΔF508 and NBD1-WT. Blue edges denote transitions dominant in the mutant folding pathway, while red edges denote transitions in the wild type folding pathway.

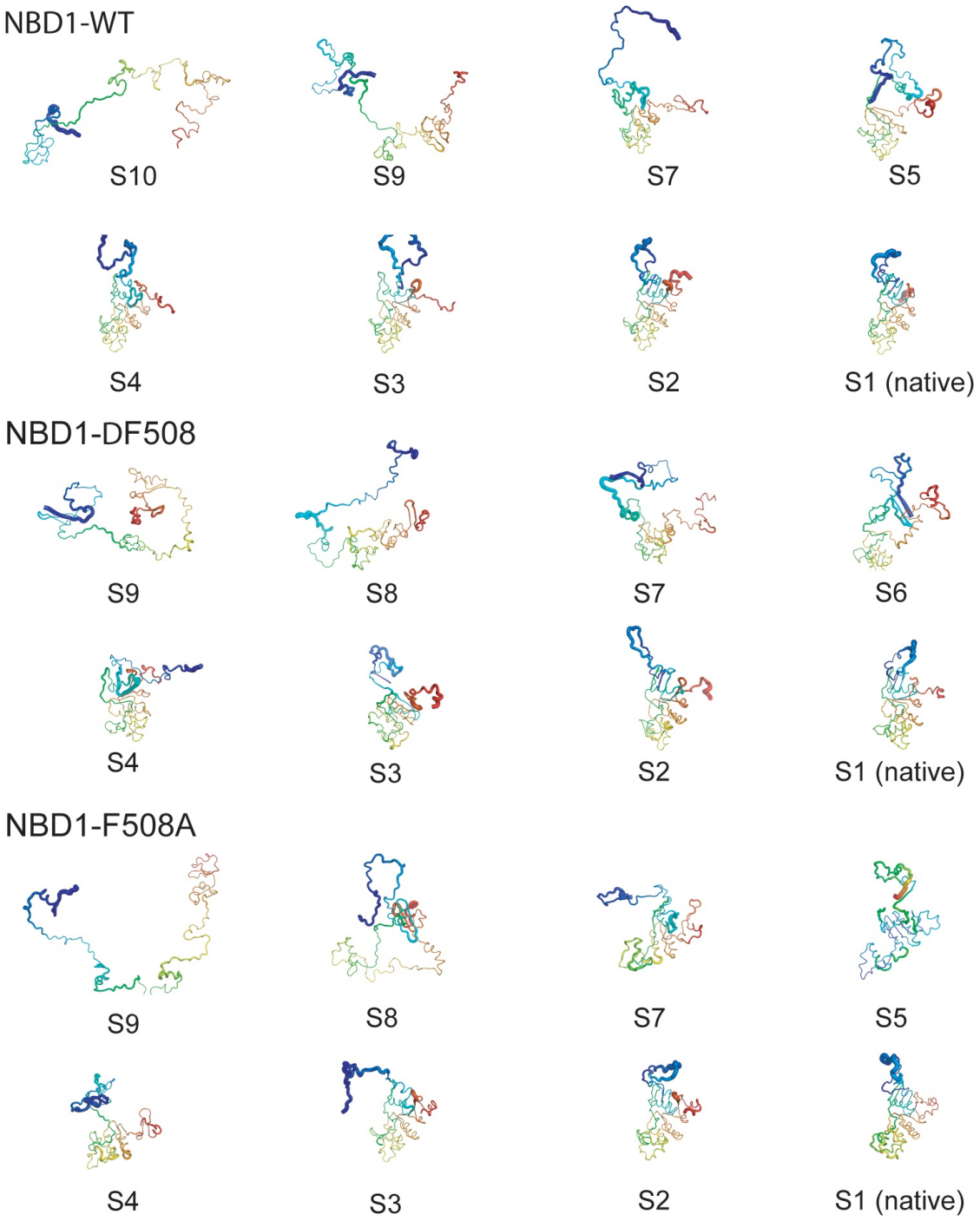

Figure 4.10. Structures of folding intermediates. To identify the structure most representative of an intermediate state, we clustered the structures within the folding intermediate. Shown above are the centroids of the dominant clusters. Diameter of the backbone cartoons is proportional to the average per residue root-mean-square deviation (RMSD) of the structures within the intermediate state. Blue and red represent the N- and C-termini, respectively.

interacted with F587 in the mutant but not in wild type (Fig. 4.11). This pattern of contact formation reflected the difference in NBD1-WT and NBD1-F508 crystal structures that are embedded in the interactions defined according to structure. Additionally, residue pairs that had similar interactions (i.e., attractive or repulsive) in the wild type and mutant crystal structures still exhibited different contacts in the folding intermediate states. These results showed that the pattern of transient contact formation in the wild type was also perturbed by Phe508 deletion. This class of residue pairs included Q525/E585 and C524/I586.

4.2.5 Computational rescue of NBD1-ΔF508

To verify that the identified contact pairs (Q493/P574 and F575/F587) found in the S7-H6 loop were indeed critical in the kinetics of NBD1, we reverted their interactions in NBD1-F508 to their interactions in NBD1-WT and performed folding simulations. In the case of the Q493/P574 pair, the residues were in close proximity in NBD1-WT but not in NBD1-ΔF508, thus we changed the interaction between Q493 and P574 in NBD1-ΔF508 from repulsive to attractive to mimic a possible rescuing mutation. Folding simulations of "rescued" NBD1-ΔF508 yielded a folding probability of 19 ± 2% (Table 4.1). On the other hand, residues F575 and F508 were in close contact in NBD1-ΔF508 but not in NBD1-WT, thus we reverted their interaction in NBD1-F508 from attractive to repulsive. Folding simulations of the second "rescued" NBD1-F508 yielded a folding probability of 20 ± 2%. These folding probabilities of the two "rescued" NBD1-ΔF508s were higher than the 13 ± 2% folding probability of the original NBD1-ΔF508, supporting our findings that the contacts between Q493 and P574 and between F575 and F587 were indeed critical to NBD1 folding.

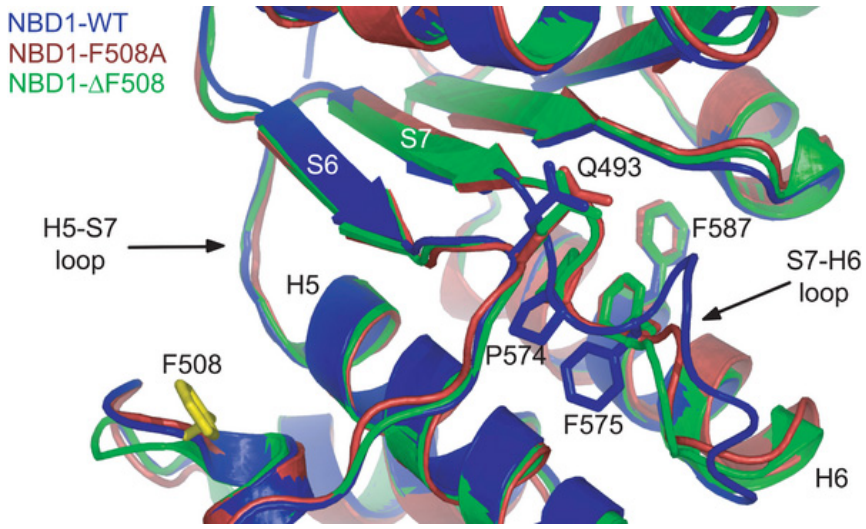

Figure 4.11. Contacts in NBD1-WT that perturbed in the F508A and ΔF508 mutants. Difference between average contact frequencies of structures within intermediate states showed malformed contacts in NBD1-ΔF508 (green) compared to NBD1-WT (blue). These identified malformed contacts in the mutants were critical determinants of NBD1 folding kinetics. In particular, P574 interacted with Q493 in wild type but not in the mutant. Also, F575 interacted with F587 in mutant but not in wild type. Redesigning these contacts to their wild type interactions in the ΔF508 background can potentially rescue NBD1-ΔF508.

Table 4.1. Computational rescue of NBD1-ΔF508. To computationally rescue the NBD1-ΔF508, we forced the loop S7-H6 of the mutant to its wild type conformation. These constructs (shown in gray) exhibit higher folding probability than the original NBD1-ΔF508.

Protein	Computational Folding Probability
wild type	33 ± 3 %
ΔF508	13 ± 2 %
ΔF508 + (Q493/P574)	19 ± 2%
ΔF508 + (F575/F587)	20± 2%

4.3 Summary

Deletion of a single residue, Phe508 in CFTR is present in approximately 90% of cystic fibrosis (CF) patients. Experiments showed that this mutant protein exhibited inefficient biosynthetic maturation and susceptibility to degradation probably due to misfolding of NBD1 and the resultant incorrect interactions of other domains. Using

molecular dynamics simulations of NBD1-WT, NBD1-F508A, and NBD1-ΔF508, we showed that the deletion of Phe508 indeed altered the kinetics of NBD1 folding. We also found that the intermediate states appearing on wild type and mutant folding pathways were populated differently and that their kinetic accessibilities were distinct [13].

We also identified critical interactions not necessarily localized near position 508, such as Q493/P574 and F575/F587, to be significant structural elements influencing the kinetic difference between wild type and mutant NBD1. Forcing these locations to adopt wild type conformations, at least from simulations, rescues the aberrant folding kinetics of the ΔF508 mutant [13].

Chapter 5

Structure of the complete CFTR channel

In this chapter, we investigate the second aspect of the defect associated to the Phe508 deletion, which is that of the misassembly of the whole protein. To determine the origin of the misassembly, it is essential to know where the residue Phe508 is located in the context of the whole protein and identify the specific set of domain-domain interactions that it mediates. We constructed a complete model of the CFTR protein, partly from homology and partly from *ab initio* protein folding. The model predicted, and verified with extensive biochemical experiments in our collaborating laboratory, that Phe508 mediates a crucial interaction between the cytoplasmic and membrane-spanning domains of the CFTR channel. Identification of this crucial interface is important in the targeted rational design of drugs that can rescue the protein.

5.1 Methods and Models

5.1.1 Modeling the CFTR structure from Sav1866

CFTR consists of several domains: nucleotide-binding domains NBD1 and NBD2, membrane-spanning domains MSD1 and MSD2, and a regulatory R domain (Fig. 4.1B).

There exist crystal structures of NBD1 but none for the other domains. The NBD1-NBD2 dimer was constructed by superimposing the NBD1 crystal structure and the homology model of NBD2 [63] on the Sav1866 (PDB ID 2HYD) structure [64]. The conformations of the NBD1-NBD2 dimer agrees with the inter-NBD cross-links observed by Mense et al. [65].

We modeled the membrane spanning domains of CFTR using homology modeling (Section 2.2.1). Because both CFTR and Sav1866, an ABC bacterial multidrug transporter, contain 12 transmembrane helices that are of similar length, we opted to model the CFTR membrane-spanning domains from that of Sav1866. The alignment of Sav1866 and CFTR was dictated by the position of their corresponding membrane-embedded regions and the conserved coupling helices in the intracellular loops (Fig. 4.1B). The membrane-embedded regions of the Sav1866 helices were identified from the PDB_TM database [66], whereas the approximate locations of CFTR TM helices were defined by using the results from earlier glycosylation site insertions [67] and the HMMTOP transmembrane prediction server (www.enzim.hu/hmmtop) [68]. Using the CFTR-Sav1866 alignment, the atomic model of CFTR MSDs was constructed in the Homology suite of INSIGHTII (Accelrys, Inc.). To eliminate clashes and refine the model, the side-chain rotamer states were optimized, and minor backbone fluctuations were introduced by using Medusa [69].

The structural model is consistent with available experimental data on the orientation and packing of transmembrane helices. Pairs of residues such as M348C/T1142C, T351C/T1142C, and W356C/W1145C, which come from transmembrane helices TM6 and TM12, could be cross-linked by molecules of different lengths [70]. Cross-linking between I340C and S877C also exists [71]. In our model, these residue pairs were closer than 23 Å (Fig. 5.2). In another study, R347 (TM5) was found to form a salt bridge with D924

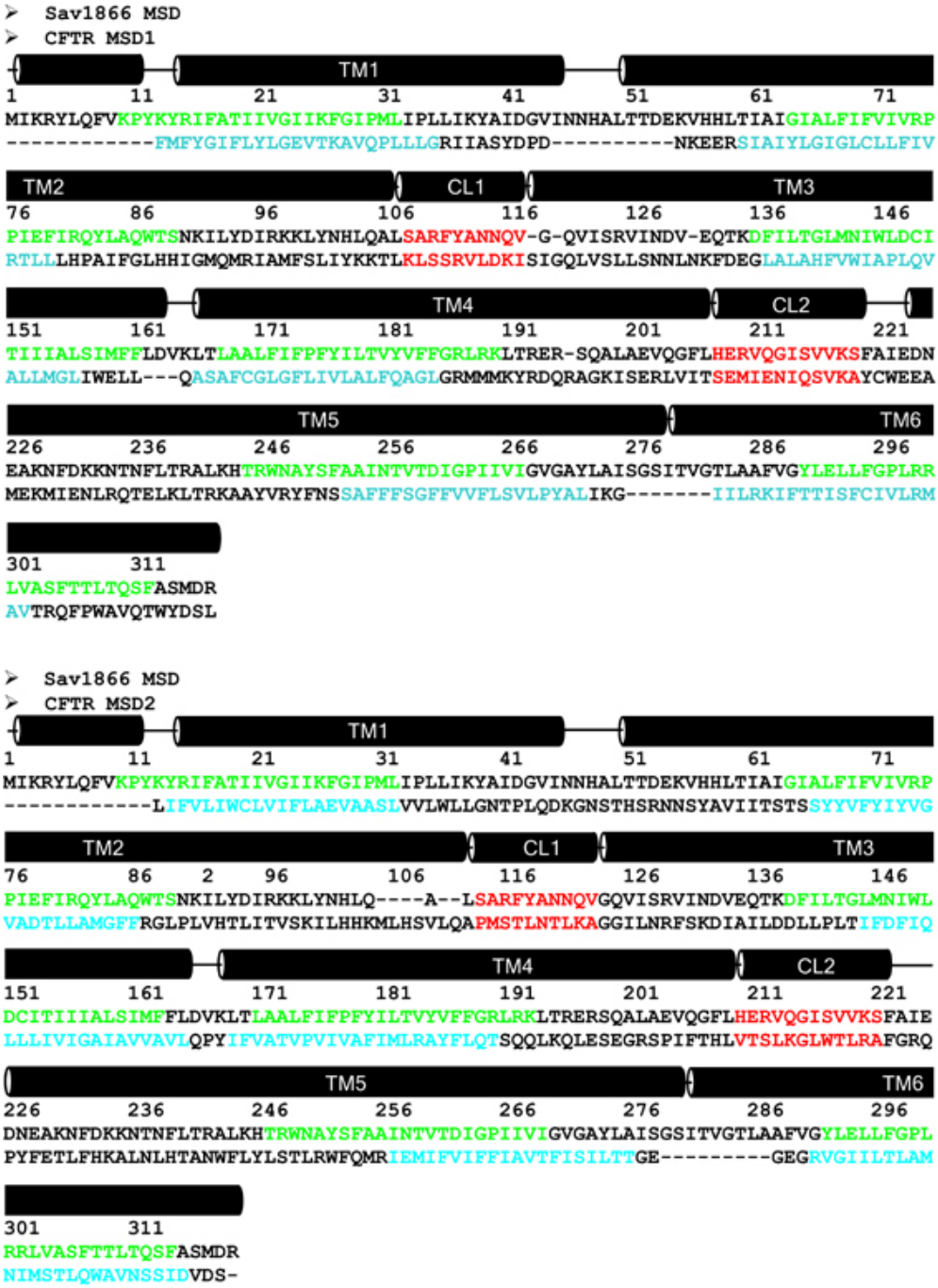

Figure 5.1. Sequence alignment of the membrane-spanning domains of human CFTR and the Sav1866 exporter [64]. Predicted membrane-embedded regions of Sav1866 are colored green, and those of CFTR are blue. Coupling helices are red.

(TM8) [72], which in the model face each other directly and their Cα atoms are separated by 9 Å (Fig. 5.2). Likewise, Therien et al. [73]found that TM3 and TM4 form anti-parallel helices and that Q207 (TM3) and V232 (TM4) form a hydrogen bond between them, suggesting that this pair of residues is structurally close. In the model, Q207 and V232 side chains directly face each other (Fig. 5.2(A)).

Aside from the experimental constraints described in the main text that were satisfied by the structural model, the organization of the membrane-spanning helices also agreed with studies identifying water-accessible residues along the channel pore [74,75]. Akabas *et al*. [74] found that residues G91, K95, and Q98, which are located in TM1, are accessible to water-soluble MTS reagents, which implies that these residues line the CFTR pore. Fig. 5.2(B) shows that indeed these residues face the pore lining in the current model. Another study by the same group [75] found that residues I331, L333, R334, K335, F337, S341, R347, T351, R352, and Q353 (all positioned in TM6) are on the water-accessible surface of the protein. These residues in TM6 are shown to face the CFTR pore lining, which makes them accessible to water (Fig. 5.2(B)).

To identify the ensemble of conformations dynamically accessible to the R domain, Dr. Tamas Hegedus (UNC-CH Department of Biochemistry and Biophysics) and I performed *ab initio* folding of the R-domain and generated decoys of low-energy structures by using discrete molecular dynamics (Section 3.1.2) with an all-atom force field called Medusa [69]. We clustered the decoy set to determine putatively dominant conformations of the R domain. The centroid of the largest decoy set is docked to the CFTR homology model by using ZDOCK [76][77], a rigid-body docking protocol that employs shape complementarity,

desolvation energy, and electrostatics. In docking the R domain model, we imposed the constraint that the C terminus of the R domain is close to the N terminus of MSD2.

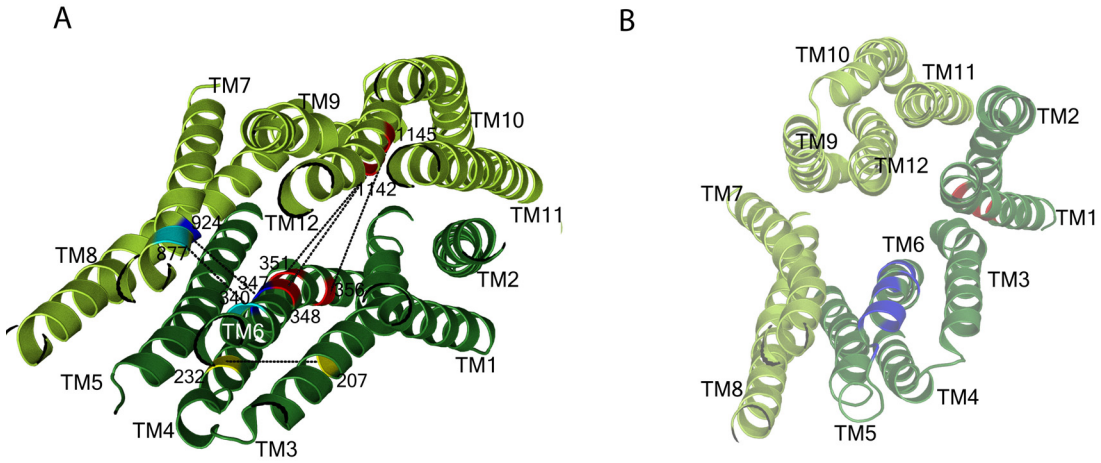

Figure 5.2. Experimental constraints satisfied in the membrane-spanning domains of the homology model. (A) Cross-links can be formed between M348-T1142, T351-T1142, and W356-T1145 (red), which are pairs of residues found between TM6 and TM12 [70]. R347 (TM6) forms a salt bridge with D924 (TM8) (blue) (11). A H-bond can be formed between Q207 (TM3) and V232 (TM4) (yellow) [73]. A recent constraint from cross-linking of I340C-S877C (cyan) is also satisfied [70]. **(B)** Residues G91, K95, and Q98 (colored red) in TM1 are water-accessible, suggesting that they face the channel pore [74]. I331, L333, R334, K335, F337, S341, R347, T351, R352, and Q353 (colored blue), which are all found in TM6, are also water-accessible [75].

5.2. Results

We constructed a 3D structure of CFTR by molecular modeling (see above). Full-length ABC proteins (the protein family to which CFTR belongs) can be grouped into two classes according to the number and conformation of their transmembrane helices. Bacterial importers have variable numbers of helices that are short, positioning their NBDs close to the membrane plane. The exporters such as Sav1866 possess 12 transmembrane helices that are longer than those of the importers, thus their NBDs are farther from the membrane plane. CFTR contains 12 transmembrane helices, and its intracellular loops are of a length similar to

those of Sav1866 [64,78,79], which suggested that CFTR MSDs can be modeled from those of Sav1866. To organize the different domains of CFTR, we followed the tertiary organization of the Sav1866 domains. The structural model is consistent with available experimental data on the orientation and packing of CFTR transmembrane helices (Fig. 5.2).

The complete structural model of CFTR is shown in Fig. 5.3. It exhibits the characteristic domain swapped architecture of Sav1866 whereby one MSD sits on both NBDs. This characteristic topology also predicts that the Phe508 residue is in contact with the cytoplasmic loop 4 (CL4) of the MSD2. The preponderance of other disease-associated mutations in CL4 that are sensitive to CFTR misassembly suggest that indeed this interface is crucial the assembly of the whole protein.

To verify that indeed the predicted interactions are correct, our experimental collaborators, Dr. John R. Riordan and company (UNC-CH Department of Biochemistry and Biophysics) performed chemical cross-linking (Fig. 5.5). Cross-linking can verify whether two residues are spatially close as predicted by the model. This method involves mutating the two residues in question to cysteines in a Cys-less CFTR background (see Fig 5.5). The two cysteines are then induced to form a disulfide bond using bifunctional methane-thiosulfonate (MTS) reagents. If the residues successfully cross-link, the cross-linked species can be detected by a shift of a band in a western blot (see Fig 5.5). Using this methodology, we indeed found that Phe508 plays a central role in this interface because it can be cross-linked to cysteines introduced at many positions in CL4. These positions include Leu-1065, Phe-1068, Gly-1069, and Phe-1074 (Fig. 5.6).

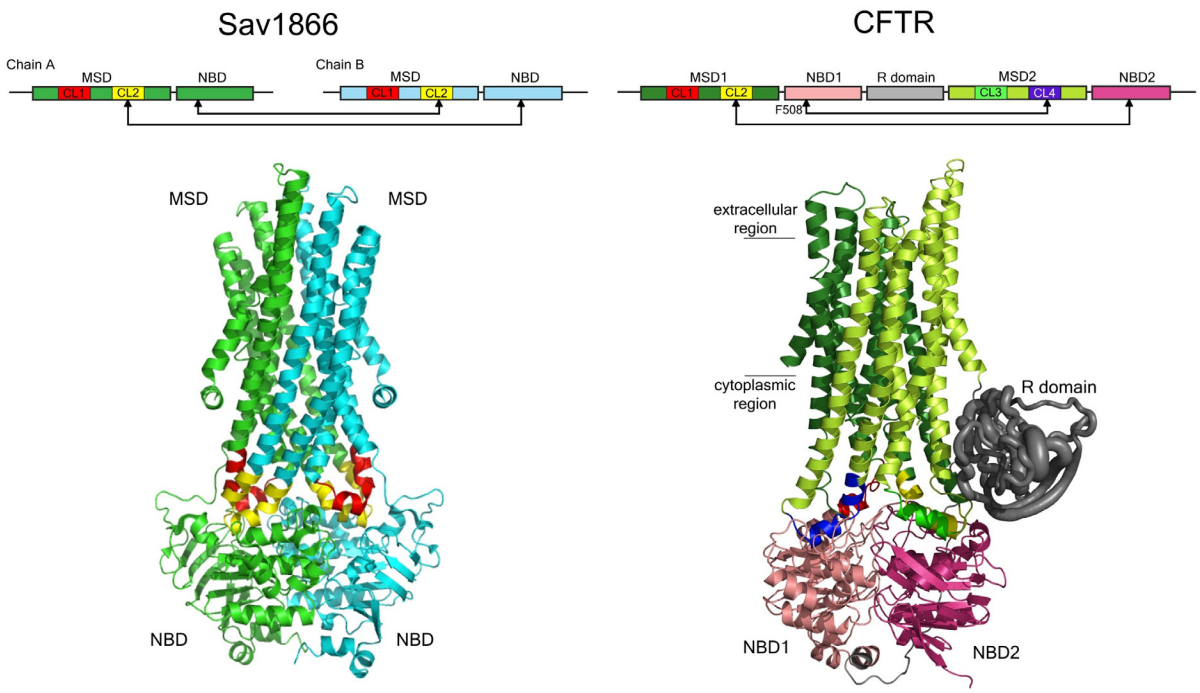

Figure 5.3. Theoretical model of CFTR structure. (Left) Structure of the bacterial multidrug transporter Sav1866 [64], which exhibits a characteristic domain swapping between the two chains. Because of this domain-swapping, each MSD forms in terface with both NBDs. Domains are colored according to schema. MSD: membrane spanning domains; NBD: nucleotide-binding domains; and CL: cytoplasmic loops. (Right) Homology model of CFTR constructed from Sav1866 exporter [14]. The CFTR R domain was approximated by constructing an ensemble of dynamically accessible conformations derived from *ab initio* folding [15].

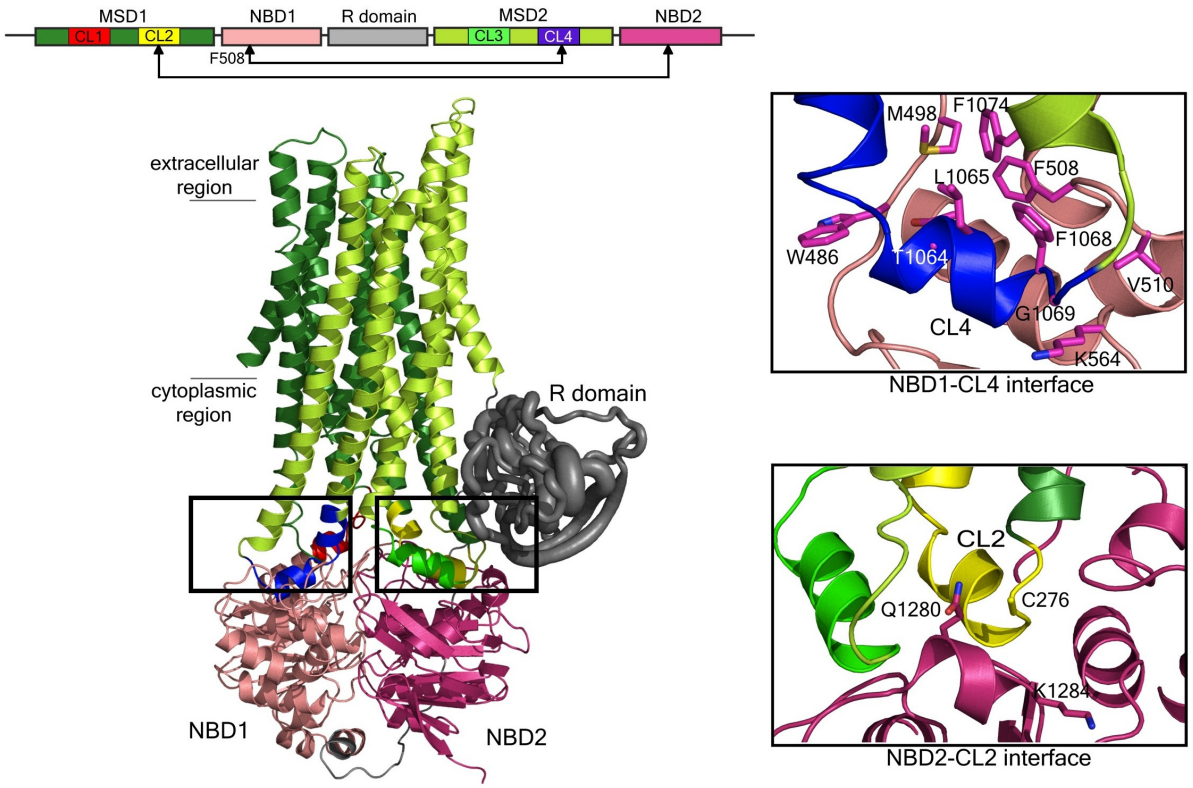

Figure 5.4. Predicted cytoplasmic and nucleotide-binding domain interfaces. The domain swapped architecture of the CFTR model (Left) predicts that NBD1 interacts with MSD2 through the cytoplasmic loop 4 (CL4), and NBD2 with the cytoplasmic loop 2 (CL2) of NBD1. (Right) Specifically, the Phe508 residue is predicted to form an aromatic cluster together with aromtic residues from CL4 [14].

We also showed that the interface mediated by the Phe508 with the CL4 of MSD2 is crucial for channel function. Single-channel gating, which persists after the introduction of a Cys pair at each interface, was completely inhibited by cross-linking (Fig. 5.7). In both cases, this inhibition was completely reversed on reduction with DTT. Hence, these points of contact are integral elements of the structure, and covalent coupling between residues on either side restricts channel activity. This restriction is unlikely to be caused by prevention of signal transmission per se but probably reflects the restriction of dynamics at the interfaces.

5.3. Summary

We constructed a model of the whole CFTR molecule to determine the overall topology of the protein. More importantly, the model identified the location of the Phe508 residue, whose deletion has been known to induce misassembly of the whole CFTR complex. We found Phe508 to mediate a crucial interaction between the NBD1 and the cytoplasmic loop 4 (CL4) of MSD2. This architecture explains the sensitivity of Phe508 and CL4 to mutations that also affect the maturation (presumably due to domain misassembly) of the whole CFTR. These interactions between NBD1-CL4 and NBD2-CL2, and other MSD-NBD interfaces, have been validated experimentally. We likewise showed that these interfaces are crucial to the channel function since cross-linking of cysteines on either side of the interface arrests channel gating, indicating a dynamic interface. The precise identification of the interface perturbed upon the deletion of the Phe508 provides a focused target for drugs that either restore or mimic the role of the lost residue.

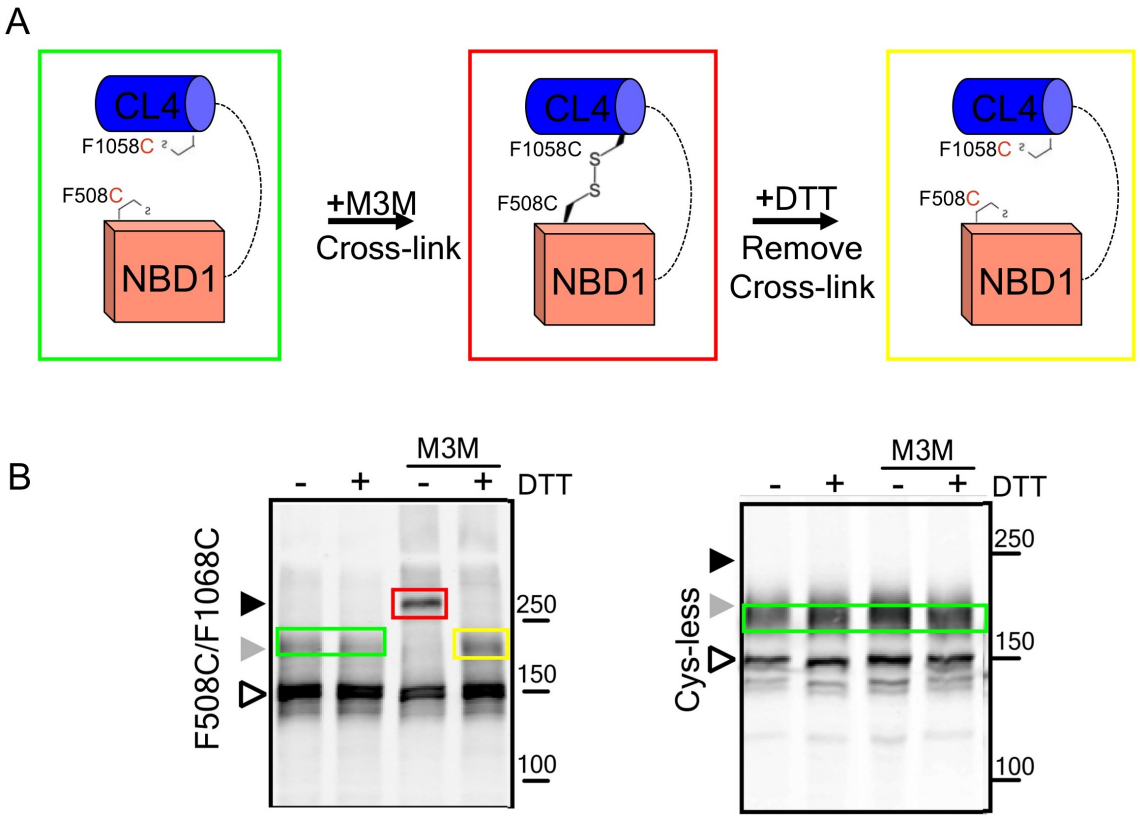

Figure 5.5. Cross-linking schema. (A) To verify if two residues are spatially close, they are mutated to cysteines and then cross-link using a bifunctional methane-thiosulfonate (MTS) reagent. The disulfide bond can be removed by adding a reducing agent (DTT). **(B)** The cross-linked species can be detected in a Western blot, as shown by the shift in band (red). As a control, no cross-linking occurs in the Cys-less CFTR[14]. (Experimental data courtesy of Lihua He, Tamas Hegedus and Liying Cui of Dr. John R. Riordan's laboratory, UNC-CH Department of Biochemistry and Biophysics).

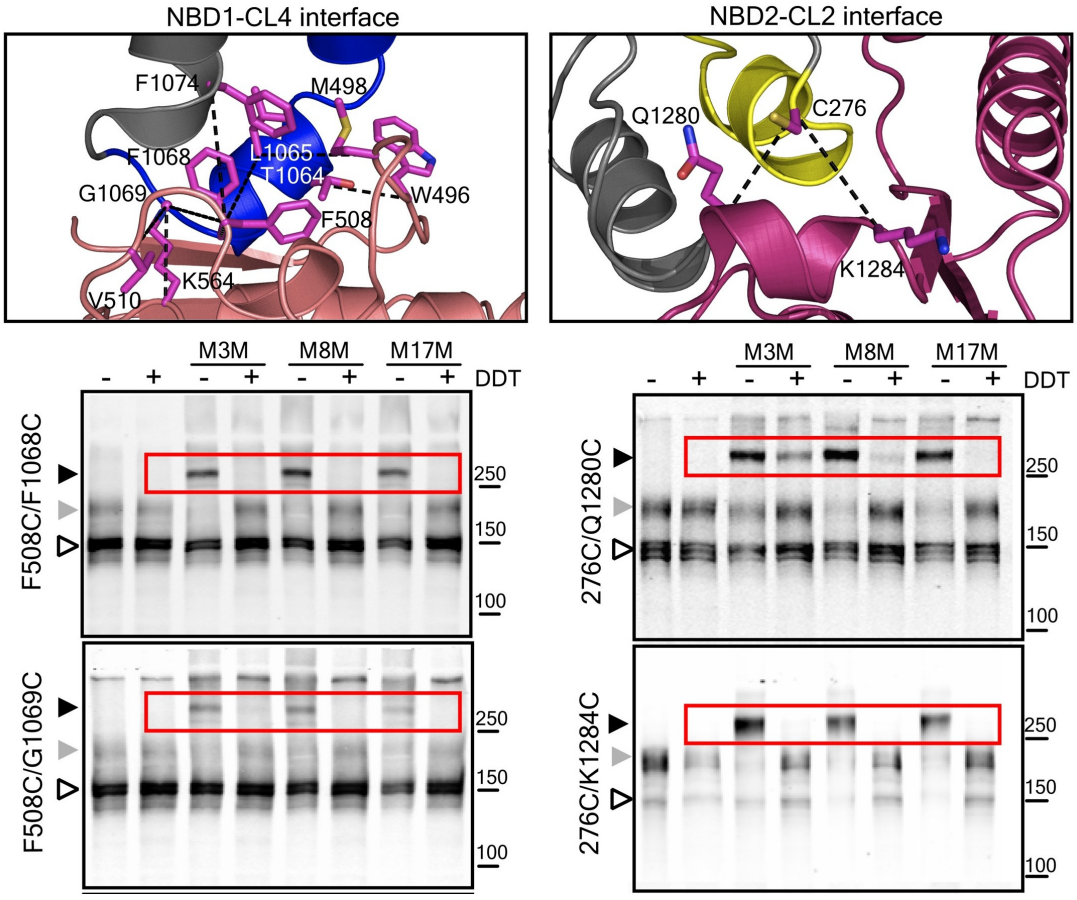

Figure 5.6. Validation #1: Cross-linking of interfacial residues. Close-up view of the interfaces formed between NBD1/CL4 and NBD2/CL2. Cross-linking of Cys pairs F508C/L1065C, F508C/F1068C, F508C/G1069C, and F508C/F1074C confirms that Phe-508 in NBD1 associates with CL4 in MSD2. Cross-linking of C276/Q1280C and C276/K1284C confirms interaction of CL2 and NBD2. *(Lower panels)* Shown in red are species of CFTR where a disulfide bridge is formed between the two cysteines, that is, the two residues are cross-linked. [Image adapted from (3)]

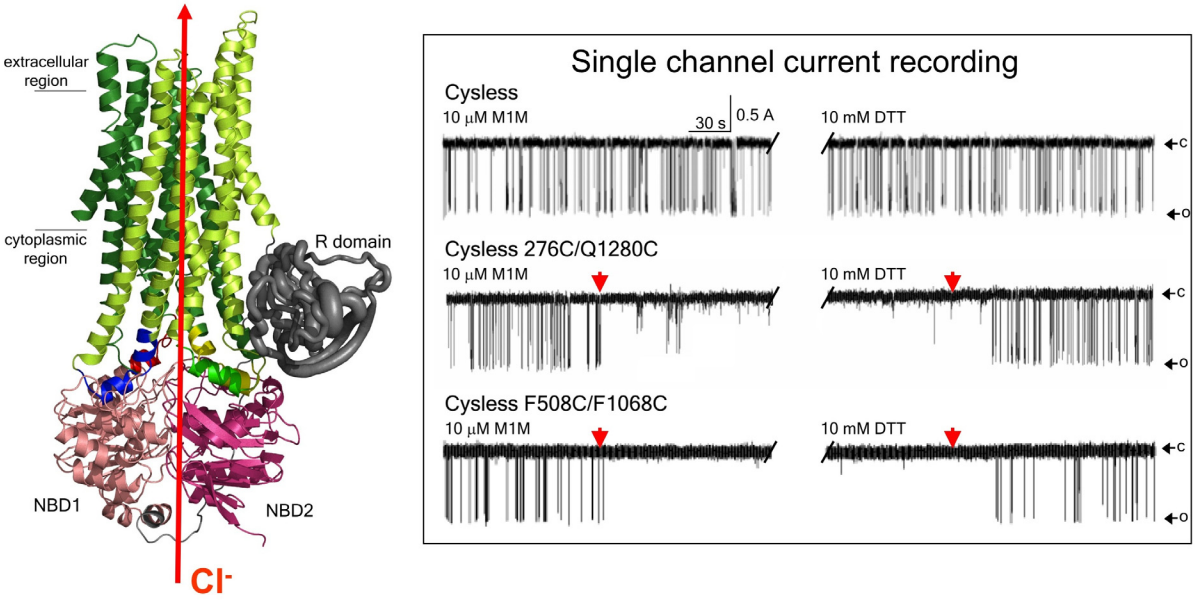

Figure 5.7. Validation #2: Cross-linking at the interface abrogates channel function. (Left) CFTR is a chloride channel whose gating behavior can be observed at single molecule level using patch clamp experiments. (Right) Current vs. time recording for single CFTR channels. "c" denotes the closed state, while "o" the open state. When cross-linking is induced upon addition of a M1M cross-linking arm (left red arrow), the channel function is inhibited. Upon addition of a reducing agent DTT (right red arrow), the channel functions again. Top trace is for the Cys-less CFTR which serves as a control. Middle trace is the cross-linking of the NBD2-CL2 interface. Bottom trace is the cross-linking of the NBD1-CL4 interface, which contains the Phe508 residue. (Experimental data courtesy of Dr. Andrei Aleksandrov and Liying Cui of Dr. John R. Riordan's laboratory, UNC-CH Department of Biochemistry and Biophysics).

Chapter 6

Conclusion and Outlook

Biology and all the phenomena that define it are inherently complex. One aspect of this complexity arises from the fact these processes span multiple length and time scales. For example processes involving proteins range from 10^{-15} s (chemical reaction) to 10^{4} s (aggregation) and from 10^{-11} m (chemical bond) to 10^{-6} m (protein complexes) (Fig. 1.1). A more specific example is the processive walking of the molecular motor dynein along the cytoplasmic dynein along the microtubule track: dynein's run length has been measured to be several millimeters with typical velocities in the order of a few nanoseconds, a time scale that is several orders of magnitude larger than the typical protein side chain and backbone movements (~10^{-9} s and 10^{-5} s, respectively). Thus, it is clear that to understand the mechanism of these biological phenomena, there is a need to develop multi-scale computational and theoretical modeling approaches. This work was an attempt to achieve this end.

However, rather than focus on developing one specific computational or theoretical method, the approach in this study was dictated by the specific biological systems and

problems under consideration. The first system I studied was the cytoplasmic dynein motor, a protein that walks along the microtubule tracks in cells, carrying cargoes (such as vacuoles, organelles, etc.) from the cell periphery towards the center. The descriptor motor is appropriate for dynein since its uses energy derived from ATP hydroysis to walk along the track. The loss of dynein function is attributed to major human diseases. The major question in the field of dynein motor is understanding its mechanism. Specifically, the questions are what is dynein's structure and how does it transduce chemical energy into mechanical work? I addressed the first question in Chapter 2 by proposing a theoretical structural model of dynein's motor unit, which is the site for ATP hydrolysis and force generation. Specifically, using homology modeling, I constructed the structural models of the AAA domains and the C-domain. To determine the organization of these individual domains, I fitted the AAA domains and the C-domain to a low-resolution EM density derived from negative staining, which was kindly provided by Dr. Michael P. Koonce (Wadsworth Center, NY) [41]. With the assistance of Dr. Feng Ding (UNC Chapel Hill), I used discrete molecular dynamics to relax the structure and eliminate steric clashes between atoms. In Chapter 3, we investigated the potential conformational changes that may accompany the force generation in dynein. I performed normal mode analysis and determined the lowest normal modes of the structural model, which suggested that the motor unit is composed of both a "rigid" and a "mobile" half. This observation was verified by molecular dynamics simulations performed using the discrete molecular dynamics package developed by Dr. Feng Ding and Dr. Nikolay V. Dokholyan [48]. These studies are described in research articles both published and under review [5,6,80].

The second system considered under this study was the CFTR channel, an ATP-binding cassette (ABC) protein that regulates ion transport in the apical membrane of

epithelial cells. Mutations in the CFTR protein, which oftentimes lead to its loss of function or absence from the epithelial membrane, are the basis of the cystic fibosis. This project focused on the most prevalent cystic fibrosis associated mutation which is the deletion of Phe508 in the first nucleotide-binding domain. In Chapter 4, using molecular dynamics simulations of simplified protein models, we explored the experimentally suggested hypothesis of abberant folding kinetics of NBD1induced by the mutation. I performed folding simulations of wild type NBD1 while Dr. Tamas Hegedus (UNC Chapel Hill) performed simulations on the ΔF508 NBD1 mutant. These simulations showed that indeed wild type and mutant NBD1 exhibit a difference in their folding kinetics; moreover, we showed these kinetic difference can be modulated by specific loop regions in the protein [13]. While these findings do not cure cystic fibrosis, they provide a consistent structural picture of how the misfolding arises. The current, albeit ambitious, goal of the project is to employ protein engineering to force the loops to their wild type conformation and show that the ΔF508 NBD1 mutant can be rescued experimentally.

This study also explored the second aspect of the ΔF508 defect that involves the misassembly of the whole CFTR protein during its biosynthesis. This misassembly results from presumably results from the perturbation of an interdomain interface when Phe508 is deleted [12]. However, the identity of this interface is still unknown. Thus, I constructed an all-atom theoretical model of the whole CFTR channel [14] from which I predicted that the Phe508 in NBD1 interacts with the second membrane-spanning domain through the fourth cytoplasmic loop (CL4). This interface has been verified biochemically by the group of Dr. John R. Riordan (UNC Chapel Hill) using cysteine cross-linking and single-channel recordings [12]. The identification of this particular interface explains the preponderance of disease-associated mutations in CL4 and NBD1. My model likewise predicted the other

CFTR cytosplamic-membrane domain interfaces, and these predictions have been similarly verified by the Riordan group [16].

REFERENCES

1. Ayton GS, Noid WG, Voth GA (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. Curr Opin Struct Biol 17: 192-198.

2. Ding F, Dokholyan NV (2005) Simple but predictive protein models. Trends Biotech 23: 450-455.

3. Vale RD (2003) The molecular motor toolbox for intracellular transport. Cell 112: 467-480.

4. Hirokawa N (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport. Science 279: 519-526.

5. Serohijos AWR, Chen Y, Ding F, Elston TC, Dokholyan NV (2006) A structural model reveals energy transduction in dynein. Proc Natl Acad of Sci USA 103: 18540-18545.

6. Serohijos AWR, Tsygankov D, Liu S, Elston TC, Dokholyan NV (2009) Multi-scale approaches for studying energy transduction in dynein. Phys Chem Chem Phys Submitted.

7. Riordan JR (2005) Assembly of functional CFTR chloride channels. Annu Rev Physiol 67: 701-718.

8. Riordan JR (2008) CFTR Function and Prospects for Therapy. Annu Rev Biochem 77: 701-26.

9. Bobadilla JL, Macek M, Fine JP, Farrell PM (2002) Cystic fibrosis: A worldwide analysis of CFTR mutations - Correlation with incidence data and application to screening. Hum Mut 19: 575-606.

10. Qu BH, Strickland EH, Thomas PJ (1997) Cystic fibrosis: a disease of altered protein folding. J Bioenerg Biomem 29: 483-490.

11. Thibodeau PH, Brautigam CA, Machius M, Thomas PJ (2005) Side chain and backbone contributions of Phe508 to CFTR folding. Nat Struct & Mol Biol 12: 10-16.

12. Du K, Sharma M, Lukacs GL (2005) The F508 cystic fibrosis mutation impairs domain-domain interactions and arrests post-translational folding of CFTR. Nat Struct & Mol Biol 12: 17-25.

13. Serohijos AW, Hegedus T, Riordan JR, Dokholyan NV (2008) Diminished self-chaperoning activity of the DF508 mutant CFTR results in protein misfolding. PLoS Comp Biol 4: e1000008.

14. Serohijos AW, 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 USA 105: 3256-3261.

15. Hegedus T, Serohijos AW, Dokholyan NV, He L, Riordan JR (2008) Computational studies reveal phosphorylation dependent changes in the unstructured R domain of CFTR. J Mol Biol 378: 1052-63.

16. He L, Aleksandrov AA, Serohijos AW, 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.

17. Marx A, Muller J, Mandelkow E (2005) The structure of microtubule motor proteins. Adv Protein Chem 71: 299-344.

18. Ahmad FJ, Echeverri CJ, Vallee RB, Baas PW (1998) Cytoplasmic dynein and dynactin are required for the transport of microtubules into the axon. J Cell Biol 140: 391-401.

19. Gerdes JM, Katsanis N (2005) Microtubule transport defects in neurological and ciliary disease. Cell Mol Life Sci 62: 1556-1570.

20. Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA(+): A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res 9: 27-43.

21. Mocz G, Gibbons IR (2001) Model for the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases. Structure 9: 93-103.

22. Mizuno N, Narita A, Kon T, Sutoh K, Kikkawa M (2007) Three-dimensional structure of cytoplasmic dynein bound to microtubules. Proc Natl Acad Sci USA 104: 20832-20837.

23. King SJ, Schroer TA (2000) Dynactin increases the processivity of the cytoplasmic dynein motor. Nat Cell Biol 2: 20-24.

24. Wang ZH, Khan S, Sheetz MP (1995) Different Patterns of Kinesin and Cytoplasmic Dynein Movement - A Single Mechanism. Biophys J 68: S328-S328.

25. Reck-Peterson SL, Yildiz A, Carter AP, Gennerich A, Zhang N, et al. (2006) Single-molecule analysis of dynein processivity and stepping behavior. Cell 126: 335-348.

26. Yildiz A, Selvin PR (2005) Fluorescence imaging with one nanometer accuracy: application to molecular motors. Acc Chem Res 38: 574-582.

27. Vale RD, Milligan RA (2000) The way things move: looking under the hood of molecular motor proteins. Science 288: 88-95.

28. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291-325.

29. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19: 1015-1018.

30. Ginalski K, Rychlewski L (2003) Detection of reliable and unexpected protein fold predictions using 3D-Jury. Nuc Acids Res 31: 3291-3292.

31. Yamada K, Kunishima N, Mayanagi K, Ohnishi T, Nishino T, et al. (2001) Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8. Proc Natl Acad Sci USA 98: 1442-1447.

32. Jeruzalmi D, O'Donnell M, Kuriyan J (2001) Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III. Cell 106: 429-441.

33. Bowman GD, O'Donnell M, Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature 429: 724-730.

34. Szakonyi G, Guthridge JM, Li D, Young K, Holers VM, et al. (2001) Structure of complement receptor 2 in complex with its C3d ligand. Science 292: 1725-1728.

35. Kristelly R, Gao G, Tesmer JJ (2004) Structural determinants of RhoA binding and nucleotide exchange in leukemia-associated Rho guanine-nucleotide exchange factor. J Biol Chem 279: 47352-47362.

36. Lee SY, De La Torre A, Yan D, Kustu S, Nixon BT, et al. (2003) Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. Genes Dev 17: 2552-2563.

37. Chacon P, Wriggers W (2002) Multi-resolution contour-based fitting of macromolecular structures. J Mol Biol 317: 375-384.

38. Wriggers W, Milligan RA, McCammon JA (1998) SITUS: A package for the docking of protein crystal structures to low-resolution maps from electron microscopy. J Mol Graph Model 16: 283-283.

39. Smith GR, Contreras-Moreira B, Zhang X, Bates PA (2004) A link between sequence conservation and domain motion within the AAA+ family. J Struct Biol 146: 189-204.

40. Kon T, Nishiura M, Ohkura R, Toyoshima YY, Sutoh K (2004) Distinct functions of nucleotide-binding/hydrolysis sites in the four AAA modules of cytoplasmic dynein. Biochemistry 43: 11266-11274.

41. Samso M, Koonce MP (2004) 25 angstrom resolution structure of a cytoplasmic dynein motor reveals a seven-member planar ring. J Mol Biol 340: 1059-1072.

42. Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146: 11-31.

43. Meng X, Samso M, Koonce MP (2006) A flexible linkage between the dynein motor and its cargo. J Mol Biol 357: 701-706.

44. Asai DJ, Koonce MP (2001) The dynein heavy chain: structure, mechanics and evolution. Trends Cell Biol 11: 196-202.

45. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15: 586-592.

46. Zhou YQ, Karplus M (1999) Interpreting the folding kinetics of helical proteins. Nature 401: 400-403.

47. Peng S, Ding F, Urbanc B, Buldyrev SV, Cruz L, et al. (2004) Discrete molecular dynamics simulations of peptide aggregation. Phys Rev E 69: 041908.

48. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI (2002) Direct molecular dynamics observation of protein folding transition state ensemble. Biophys J 83: 3525-3532.

49. Abe H, Go N (1981) Non-Interacting Local-Structure Model of Folding and Unfolding Transition in Globular-Proteins .2. Application to Two-Dimensional Lattice Proteins. Biopolymers 20: 1013-1031.

50. Go N, Abe H (1981) Non-Interacting Local-Structure Model of Folding and Unfolding Transition in Globular-Proteins .1. Formulation. Biopolymers 20: 991-1011.

51. Hook P, Mikami A, Shafer B, Chait BT, Rosenfeld SS, et al. (2005) Long range allosteric control of cytoplasmic dynein ATPase activity by the stalk and C-terminal domains. J Biol Chem 280: 33045-33054.

52. Roberts AJ, Numata N, Walker ML, Kato YS, Malkova B, et al. (2009) AAA+ Ring and linker swing mechanism in the dynein motor. Cell 136: 485-495.

53. Boucher RC (2004) New concepts of the pathogenesis of cystic fibrosis lung disease. Euro Respir J 23: 146-158.

54. Caldwell RA, Grubb BR, Tarran R, Boucher RC, Knowles MR, et al. (2002) In vivo airway surface liquid Cl- analysis with solid-state electrodes. J Gen Phys 119: 3-14.

55. Antunes MB, Cohen NA (2007) Mucociliary clearance - a critical upper airway host defense mechanism and methods of assessment. Curr Opin Allergy Clin Immun 7: 5-10.

56. http://www.genet.sickkids.on.ca/cftr.

57. Gennerich A, Carter AP, Reck-Peterson SL, Vale RD (2007) Force-induced bidirectional stepping of cytoplasmic dynein. Cell 131: 952-965.

58. Khare S, Ding F, Dokholyan NV (2003) Hybrid molecular dynamics studies on Cu,Zn superoxide dismutase reveal topologically important residues. Abstr Pap Amer Chem Soc 225: U704-U704.

59. Lewis HA, Buchanan SG, Burley SK, Conners K, Dickey M, et al. (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23: 282-293.

60. Lewis HA, Zhao X, Wang C, Sauder JM, Rooney I, et al. (2005) Impact of the Delta F508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 280: 1346-1353.

61. Pande VS, Grosberg A, Tanaka T (1997) On the theory of folding kinetics for short proteins. Fold Des 2: 109-114.

62. Hubner IA, Shimada J, Shakhnovich EI (2004) Commitment and nucleation in the protein G transition state. J Mol Biol 336: 745-761.

63. Callebaut I, Eudes R, Mornon JP, Lehn P (2004) Nucleotide-binding domains of human cystic fibrosis transmembrane conductance regulator: detailed sequence analysis and three-dimensional modeling of the heterodimer. Cell Mole Life Sci 61: 230-242.

64. Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180-185.

65. Mense M, Vergani P, White DM, Altberg G, Nairn AC, et al. (2006) In vivo phosphorylation of CFTR promotes formation of a nucleotide-binding domain heterodimer. EMBO J 25: 4728-4739.

66. Tusnady GE, Dosztanyi Z, Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nuc Acids Res 33: D275-D278.

67. Chang XB, Tabcharani JA, Hou YX, Jensen TJ, Kartner N, et al. (1993) Protein Kinase-A (Pka) Still Activates Cftr Chloride Channel After Mutagenesis of All 10 Pka Consensus Phosphorylation Sites. J Biol Chem 268: 11304-11311.

68. Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17: 849-850.

69. Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. PLoS Comp Biol 2: 725-733.

70. Chen EY, Bartlett MC, Loo TW, Clarke DM (2004) The Delta F508 mutation disrupts packing of the transmembrane segments of the cystic fibrosis transmembrane conductance regulator. J Biol Chem 279: 39620-39627.

71. 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.

72. Cotten JF, Ostedgaard LS, Carson MR, Welsh MJ (1996) Effect of cystic fibrosis-associated mutations in the fourth intracellular loop of cystic fibrosis transmembrane conductance regulator. J Biol Chem 271: 21279-21284.

73. Therien AG, Grant FEM, Deber CM (2001) Interhelical hydrogen bonds in the CFTR membrane domain. Nat Struct Biol 8: 597-601.

74. Akabas MH, Kaufmann C, Cook TA, Archdeacon P (1994) Amino-Acid-Residues Lining the Chloride Channel of the Cystic-Fibrosis Transmembrane Conductance Regulator. J Biol Chem 269: 14865-14868.

75. Cheung M, Akabas MH (1996) Identification of cystic fibrosis transmembrane conductance regulator channel-lining residues in and flanking the M6 membrane-spanning segment. Biophys J 70: 2688-2695.

76. http://zlab.bu.edu/zdock/.

77. Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52: 80-87.

78. Hollenstein K, Dawson RJP, Locher KP (2007) Structure and mechanism of ABC transporter proteins. Curr Opin Struct Biol 17: 412-418.

79. Scott-Ward TS, Cai ZW, Dawson ES, Doherty A, Da Paula AC, et al. (2007) Chimeric constructs endow the human CFTR Cl(-)channel with the gating behavior of murine CFTR. Proc Natl Acad Sci USA 104: 16365-16370.