# Chapter 23

# Molecular Modeling Tools and Approaches for CFTR and Cystic Fibrosis

**Adrian W.R. Serohijos, Patrick H. Thibodeau, and Nikolay V. Dokholyan**

## Abstract

Cystic fibrosis is a multi-faceted disease resulting from the dysfunction of the CFTR channel. Understanding the structural basis of channel function and the structural origin of the defect is imperative in the development of therapeutic strategies. Here, we describe molecular modeling tools that, in conjunction with complementary experimental tools, lead to significant findings on CFTR channel function and on the effect of the pathogenic mutant F508del.

**Key words:** Cystic fibrosis, CFTR, ABC proteins, molecular modeling, homology modeling, discrete molecular dynamics, protein stability estimations.

## 1. Introduction

Because of the significance of understanding CFTR function and the CF disease, solving the crystal structure of the full-length CFTR has been a Holy Grail in the field since the discovery of the CF gene (1), but still remains elusive. In the absence of an experimentally determined structure for the whole protein, structural models of the protein, albeit at lower resolution and accuracy, guide functional and experimental investigations. Modeling efforts already lead to insights and discoveries such as the identification of the molecular interface-mediated Phe508 and other deleterious mutations and the misfolding and aberrant folding kinetics of NBD1. Models for the folding and structure of CFTR have been experimentally tested using a variety of biochemical, functional, and cell biological approaches. The results of these assays both inform and experimentally test models of CFTR

M.D. Amaral, K. Kunzelmann (eds.), *Cystic Fibrosis*, Methods in Molecular Biology 741,
DOI 10.1007/978-1-61779-117-8_23, 

structure and folding. While these data, and the models generated and refined with them, are of relatively low resolution, they are informative for experimental design and interpretation and facilitate the generation of new hypotheses.

### 1.1. Modeling of the CFTR Structure

There is a strong incentive to model the structure of the CFTR protein, and ABC proteins in general, because of the available experimental structures that can serve as reasonable templates for comparative modeling. The goal is to build a model of reasonable accuracy that will enable detailed understanding of the CFTR function and the defect at the molecular level. From structural models of the complete CFTR derived from recently experimentally determined structures of ABC transporters, we found that Phe508 in NBD1 interacts with the second membrane-spanning domain through the fourth cytoplasmic loop (CL4) (2). This interface was predicted (and subsequently experimentally verified) to be perturbed upon Phe508 deletion. Other interfaces between the membrane and the cytoplasmic region that were predicted by the model have been validated experimentally (3).

### 1.2. Investigation of NBD1 Folding Kinetics Using Molecular Dynamics Simulations

Protein folding is the process by which the nascent chain eventually assumes its native three-dimensional structure. It is a broad experimental and theoretical scientific discipline and the reader is referred to numerous excellent reviews for a broader introduction (4, 5). Here, we focus on the folding of NBD1 primarily because of the misfolding defect induced by the Phe508 deletion. It has been found experimentally that the deletion of the Phe508 backbone may shift a fraction of the nascent NBD1s of the CFTR-F508del off the wild-type folding pathway, causing misfolding and eventual rapid degradation of the protein (6–8). Solubilizing and rescue mutations both in the wild type and F508del background have also been shown to attenuate the folding kinetics of NBD1 (9). Thus, there is a need for understanding the detailed structural origin of the perturbed kinetics in NBD1. Using molecular dynamics simulations and simplified protein models, we identified putative metastable folding intermediates and the folding pathways of the wild type and mutant NBD1. We likewise reproduced the experimentally observed difference in folding kinetics. Moreover, from the structures of the intermediate states, we found that this difference in kinetics could be attributed to the conformation of specific loop regions in NBD1.

## 2. Materials

In modeling the CFTR structure, we used the Sav1866 crystal structure as the template for the transmembrane domains. To eliminate clashes, short molecular dynamics simulation was

performed on the model structure using discrete molecular dynamics DMD (http://dokhlab.unc.edu/tools/ifold/) (*see* below for details). DMD was likewise used in investigating the folding kinetics of NBD1. Apart from molecular dynamics simulations, all other computing may be performed in a standard desktop computer.

## 3. Methods

### 3.1. Modeling the CFTR Structure

The basic protocol for homology modeling involves constructing a structural model of protein of unknown structure (the target protein) by copying the coordinates of a related protein of known experimental structure (the template protein) (**Fig. 23.1a**).

Several servers automatically perform homology modeling of single-domain cytosolic proteins without human interventions. However, modeling of multi-domain transmembrane proteins such as CFTR is more nuanced. We outline below the procedure for modeling the structure of CFTR and other transmembrane proteins.

#### 3.1.1. Choosing a Template

The most crucial step in homology modeling is choosing the template structure since it sets the upper bound for the accuracy of the model; choosing an unrelated protein as template leads to an altogether wrong model. In general, proteins sharing sequence identity of ~25% or greater will obey the same overall topology, although members of more ubiquitous protein families can exhibit lower sequence identity. In searching for homologs, several servers are available; most notable is the 3DJURY meta-server, which combines the results of independent servers that each implement different homology search algorithms, thus increasing the accuracy of the prediction (10).

CFTR consists of two nucleotide-binding domains (NBD1 and NBD2), two membrane-spanning domains (MSD1 and MSD2), and a regulatory region (R domain) (**Fig. 23.1b**). The two nucleotide-binding domains follow the canonical folds of the NBDs in ABC transporters. In fact, several structures of NBD1 already exist (11, 12), which subsequently inspired the construction of homology model for NBD2 (13). A putative model of the R domain was also derived from ab initio folding (3). Thus, the outstanding task for modeling the complete CFTR structure was building the membrane-spanning domains and determining the quaternary organization of the protein.

Threading the sequences CFTR MSDs did not result in statistically significant hits to any known protein structure because there were few known structures of ABC proteins. Also, the sequences of the membrane-spanning regions of ABC proteins are

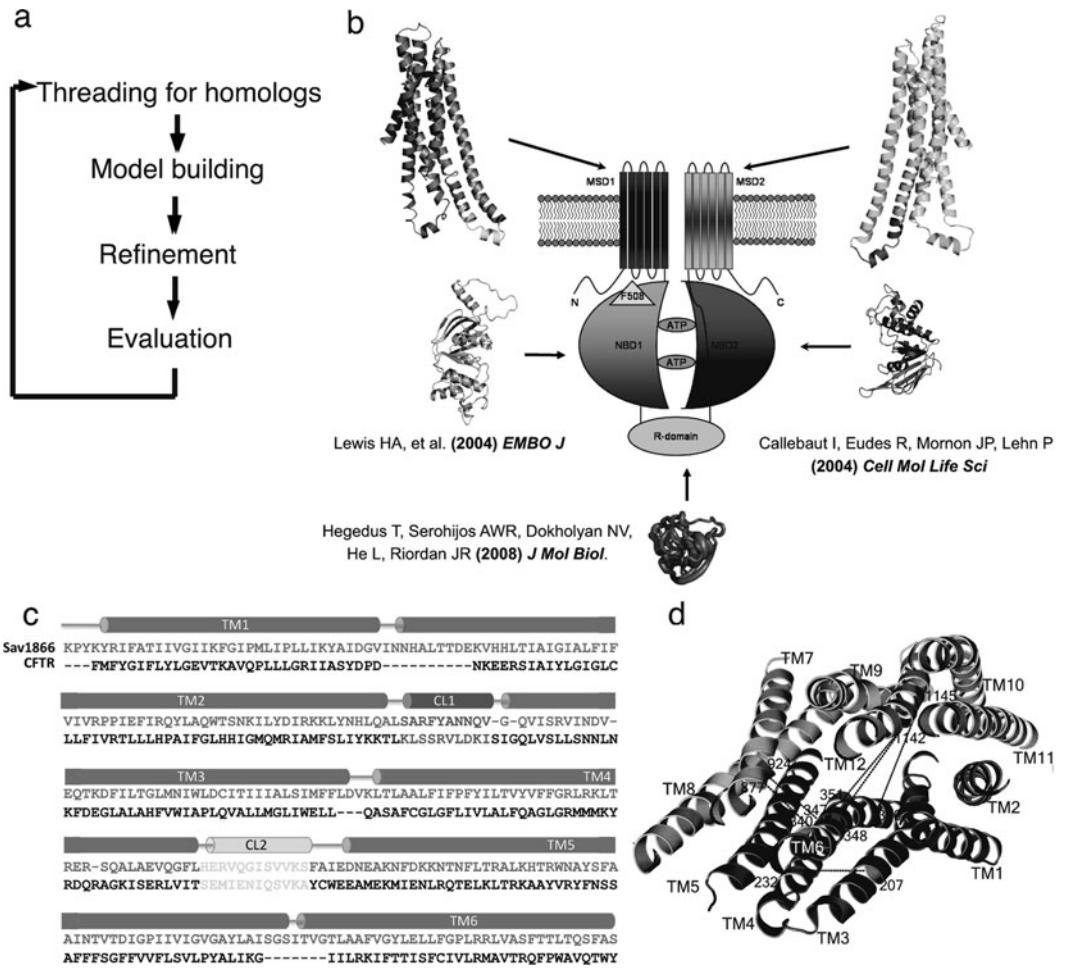


Fig. 23.1. (**a**) Homology modeling. To construct a model, the sequence of the protein of unknown structure is mapped to the sequence of the protein with known structure. The mapping optimizes the alignment of conserved residue regions. An all-atom model is then constructed by copying the backbone topology (and when possible, the rotameric states of the side chains) of the known structure. The model structure is refined using molecular dynamic simulations. The quality of the model structure is then evaluated for correct geometry and sterics. (**b**) CFTR domains. Prior to the modeling of the full CFTR structure (2, 43), there were existing model structures of the cytoplasmic domains: NBD1 from X-ray crystallography (11, 12), NBD2 from homology modeling (13), and R-domain from ab initio protein folding (44). (**c**) Alignment between CFTR and Sav1866 MSD. Alignment between the membrane-spanning domains was guided by the boundaries of putative membrane-embedded regions of each transmembrane helix and by the conserved coupling loops (CLs). (**d**) Experimental constraints. The model is evaluated as to whether it satisfies known data such as those derived from cross-linking experiments and accessibility studies.

not strictly conserved, even between those known to share similar topologies such as Sav1866, PgP, and MsbA (14–18). Nonetheless, several reasons suggest the bacterial multi-drug transporter Sav1866 (14) as a reasonable starting point for modeling the CFTR MSDs. Both CFTR and Sav1866 contain 12 transmembrane helices that are of similar length, indicating that CFTR and Sav1866 are of equal distance from the membrane-plane.

### 3.1.2. Model Construction

In our CFTR model construction, the starting point is an alignment of the sequences of the template and target proteins (Sav1866 and CFTR, respectively) (**Fig. 23.1c**). 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. 23.1c**). The membrane-embedded regions of the Sav1866 helices were identified from the PDB_TM database (19), whereas the approximate locations of CFTR TM helices were defined using the results from glycosylation site insertion studies (20) and the HMMTOP transmembrane prediction server (www.enzim.hu/hmmtop) (21). In practice, major sources of error in this step can come from insertions (i.e., sequences in the target sequence that do not have a corresponding structure in the template) and deletions (i.e., regions in the template structure without a corresponding sequence in the target). Several models must be constructed from various possible alignments and validated against known experimental data (*see* below).

Using the alignment as input, structural models can be constructed using commercially available homology modeling tools such as the Homology suite of INSIGHTII (Accelrys, Inc.) or homology modeling servers such as HOMER (http://protein.cribi.unipd.it/homer/). The MSD models of CFTR were constructed in the Homology suite of INSIGHTII.

To arrive at the quaternary structure of CFTR, the MSDs and NBDs were superimposed with the corresponding domains in Sav1866 (**Fig. 23.1b**). To eliminate any major steric clashes, a short MD simulation was performed where the protein backbone is constrained in the neighborhood of the model structure. We used the discrete molecular dynamics (DMD), a fast sampling algorithm, and the Medusa force field (22, 23) (*see* **Section 4** for more detailed description of DMD). We likewise optimized the rotameric states of the side chains using Medusa (23–25).

### 3.1.3. Model Validation

A homology model needs to be evaluated for its geometry and sterics. An excellent tool used primarily to evaluate experimental protein structures is Molprobity (26), a server that determines all-atom contacts and unusual backbone and side chain conformations by identifying Ramachandran and rotamer outliers. The second set of validation entails examining whether the model agrees with known experimental data on specific structural features. For example, residue pairs known to be cross-linked or form a salt-bridge or a hydrogen bond should have a conformation that favors the formation of these interactions. Residues known to be accessible to MTS reagents using SCAM (substituted cysteine accessibility mutagenesis) are expected to face the pore region (**Fig. 23.1d**). The CFTR model we constructed was evaluated

against these experimental constraints (**Fig. 23.1d**). Expectedly, since neither the experiment nor the model is perfect, there may be contentious data points and model features. These specific contentious points and other model features could then be validated or resolved with further experiments.

The primary utility of structural models is that they allow for positing specific features that could be tested experimentally. For example, the CFTR model predicted a cross-over of the two CFTR MSDs, similar to that in Sav1866. This cross-over predicted that the Phe508 in NBD1 interacts with CL4 of MSD2, as subsequently validated by cross-linking experiments and functional experiments (2). Other inter-domain interfaces in CFTR model have also been verified (3).

It must be noted that while the protocol is described linearly, the modeling and validation is a recursive process – additional experiments increase the accuracy of the model, and more accurate models lead to better experiments. Since the publication of the CFTR computational model, several of its features have been explored by other groups.

### *3.2. Folding Simulations of NBD1*

Single domain proteins of approximately ~100 residues or less have been successfully folded using traditional molecular dynamics with protein models that use physical force fields and all-atom models of proteins (27, 28). Several successful folding trajectories have been reported for these small proteins. However, elucidating the CFTR NBD1 kinetics and folding pathways is intractable for traditional methodologies because the domain is large (~200 residues long). More importantly, folding is a stochastic event and determining the dominant pathways requires multiple folding simulations to acquire significant statistics. However, using discrete molecular dynamics and simplified protein models surmounts this difficulty.

#### *3.2.1. Folding Simulations*

The two major components of molecular dynamics are the system (protein model and their interactions) and the engine (calculator of the system evolution). Accessing longer time scales for MD simulations entails simplifying the protein model by using fewer atoms or simplifying the description of the interaction potentials between the various atoms. We apply both simplifications to the NBD1 folding problem.

##### 3.2.1.1. Simplified Models of Proteins

Several simplified models have been developed by various groups through the years elucidating important aspects of various protein folding problems (29) (*see* **Note 1**). Obviously, the type of model depends on the specific phenomena under consideration. In our experience, the following protein model can recapitulate the difference in folding kinetics of wild type and mutant NBD1: glycines are represented as three beads (–N, Cα, C′); aromatic

residues phenylalanine, tyrosine, tryptophan, and histidine by five beads (–N, Cα, C′, Cβ, Cγ); and all other residues by four beads (–N, Cα, C′, Cβ) (30, 31). This simplified protein model maintains important features of the protein such as side chain packing and has been successfully employed in studying protein aggregation (31). The bonded interactions exist between pairs of atoms that are consecutive, next-nearest neighbors under angular constraints and atoms linked by dihedral interaction.

To model the long-range interaction between atoms in the model, we also used a structure-based potential, called the Go-model, where particles that are within a predefined distance (typically less than 4.5 Å) in the native structure are assigned an attractive potential, and zero otherwise. Since the Go-potential is structure based, we needed the native structure of the protein to infer the non-native and native interactions for the long-range interaction. We used the following structures for wild type and mutant NBD1s: wild type (PDB ID: 2BBO), F508del (PDB ID: 1XMJ), and F508A (PDB ID: 1XMI) (*see* **Note 2**). The F508A mutant has been shown to exhibit intermediate folding defects compared to F508del (7), and thus is an interesting control for the folding simulations.

#### 3.2.1.2. Discrete Molecular Dynamics

The calculation of the system's governing state equations in MD drives the evolution of the system. In traditional MD, the interactions are usually defined by continuous potentials, and the system evolution involves the time integration of Newton's laws to obtain. On the other hand, in discrete molecular dynamics (DMD), the potentials are discretized, maintaining only the significant features of the interactions (29). Since the potentials in DMD are discretized, all atoms move at constant velocity unless they encounter a collision with other particles or a discrete step in the potential function. Because atoms move at constant velocity between collision events, the system is essentially updated only during collisions. This algorithm allows for the rapid update of the systems state. In the limit of sufficiently large number of steps, DMD becomes equivalent to the traditional MD based on Newtonian dynamics. Simulations of using DMD can be performed using the iFold server (http://dokhlab.unc.edu/tools/ifold/) (32).

### *3.2.2. Simulation Protocol*

Using the simplified models described above, in our previous studies, we performed folding simulations for each NBD1-WT, NBD1-F508del, and NBD1-F508A. Starting from fully unfolded chains, the temperature of the system is progressively reduced to allow NBD1 to fold to its native structure. Folding simulations proceeded until a time $t_{max}$, which is chosen to be longer than the typical folding time of the protein (33). For a particular trajectory, a structure is considered folded when (1) its energy is less than or

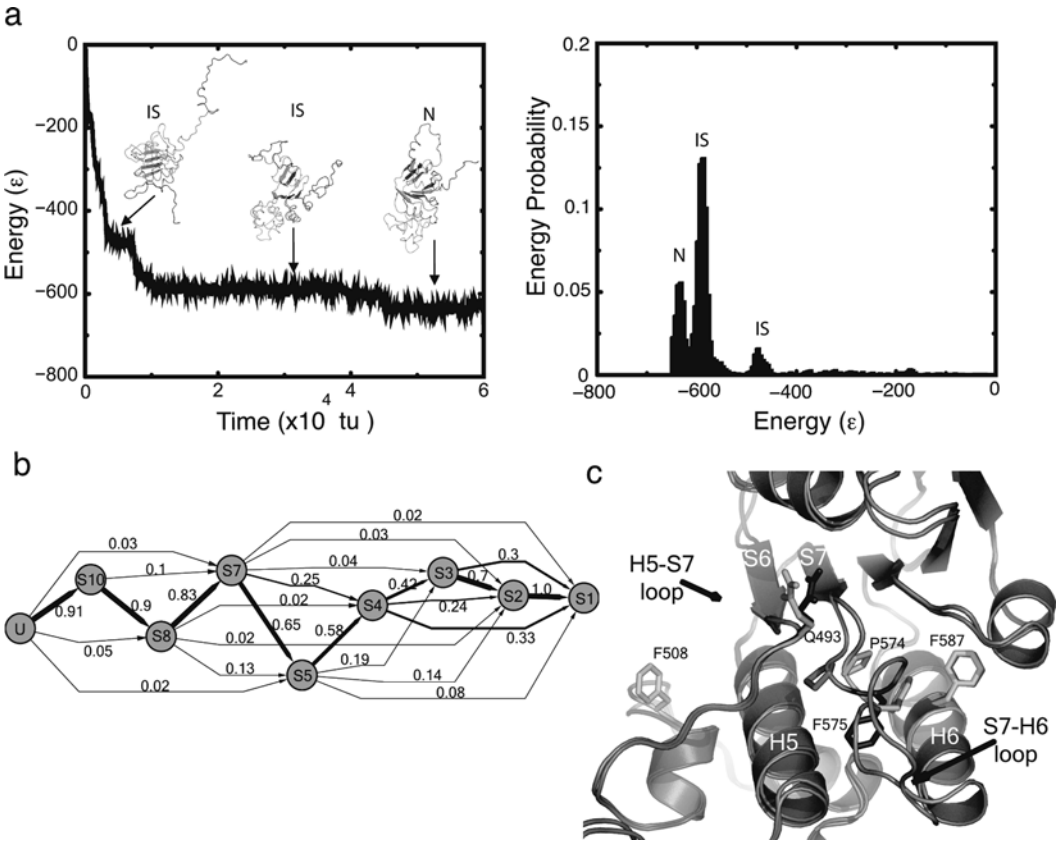


Fig. 23.2. (**a**) Folding trajectory of NBD1. Annealing simulations are performed to facilitate folding of NBD1. Shown is a representative time evolution of energy starting from the unfolded to the native (N) state. As the protein proceeds toward its native state, it goes through metastable folding intermediate states (IS), which are observed as peaks in the energy probability distribution. (**b**) Folding pathways. Probability of kinetic transitions between intermediate states of wild-type NBD1. The probability of exiting a state is normalized to 1 and the thickness is rendered proportional to the probability. (**c**) Contacts in NBD1-WT that perturbed in the F508A and F508del mutants. Difference between average contact frequencies of structures within intermediate states shows malformed contacts in NBD1–F508del (*light gray*) compared to NBD1-WT (*black*). These identified malformed contacts in the mutants are critical determinants of NBD1 folding kinetics. In particular, P574 interacts with Q493 in wild type but not in the mutant. Also, F575 interacts with F587 in mutant but not in wild type. Redesigning these contacts to their wild-type interactions in the F508del background can potentially rescue F508del–NBD1.

equal to the energy of the native state (2), its structure is within 2.5 Å RMSD (root-mean-square deviation) from the native state, and (3) the secondary structural elements possess correct topological wiring. Shown in **Fig. 23.2a** is a typical folding trajectory that reached a folding structure within $t_{max}$. As the protein folds toward the native state, it proceeds through metastable folding intermediates.

*3.2.3. Analysis*

1. Folding probability. Addresses the question of whether there is intrinsic folding difference between the wild type and the mutant NBD1s, as suggested by experiments. We calculated

the folding probability defined as the ratio of the number of trials that successfully folded to the native state and the total number of folding trials. Assuming that the folding simulation is a Bernoulli process with a binary outcome, folded or unfolded, one can estimate the error associated with the folding probability as $\sigma^2 = p(1-p)/n$, where $p$ is the probability of folding and $n$ is the total number of folding simulations. We found that at the given $t_{max}$, wild type exhibited higher folding probability than the F508del, and F508A folding probability is intermediate to that of wild type and F508del.

2. Intermediate states. The intermediate states can be inferred from the folding trajectory. Using the energy as the reaction coordinate, we first calculated the normalized energy distribution, then fitted a sum of multiple Gaussian curves $\sum_i a_i \exp[(x-b_i)^2/c_i^2]$, where $a_i$, $b_i$, and $c_i$ are height, center, and standard deviation of the $i$th Gaussian curve, respectively. Each Gaussian curve represents a putative folding intermediate. By comparing the intermediate states of the wild type and mutant, one can then determine if some intermediate states are specific only to either wild type or mutant.
3. Folding pathways. Once the putative folding intermediates are identified, one can then map the sequence of folding events by estimating the transition probabilities between intermediate states. The probability of transition between any two given states is proportional to the number of trajectories that exhibited the transition. The sum of the probabilities emanating from a given state is normalized to 1, which physically means that the system always exits from its current intermediate state (and assuming that backward transitions are ignored). Shown in **Fig. 23.2b** are the estimated transition probabilities for the set of simulations on NBD1-WT constructs. Since the transition probabilities represent independent conditional probabilities, the probability of a pathway is the product of the probabilities of traced edges, and the dominant pathways are the traced edges with the highest probability. The dominant pathways of various NBD1 constructs can then illustrate if the folding pathways for the wild type and NBD1 are distinct and which states are unique to each construct.
4. Structural characterization of intermediate states. Since energy as defined in the simplified models is not the exact reaction coordinate, there will be degeneracy in the structures populating our identified intermediate states. To identify the dominant structural features of a coordinate, the ensemble of structures can be clustered according to their

pairwise RMSD. Alternatively, one can calculate the contact frequency between all residue pairs to determine the persistent contacts in the structural ensemble. Using these analyses on our simulations of NBD1, we found that specific loop regions in NBD1 can be potential determinants of its folding kinetics (**Fig. 23.2c**).

### 3.3. Experimental Evaluation of CFTR Models

The functional and structural characterization of NBD1, apart from other CFTR domains, follows from observations of the ABC transporter protein family. Within the ABC transporter family of proteins, the cytosolic NBDs are highly conserved across single- and multi-cell organisms (35, 36). While mammalian ABC transporters are most often encoded as multi-domain proteins on one or two polypeptides, many bacterial ABC transporters are encoded as multiple single-domain polypeptides that assemble post-translationally to form a functional transporter (37). As such, the sequence of mammalian NBDs, while fused to other ABC domains, may also fold autonomously and respond independently to mutation.

The cloning of CFTR and identification of the F508del mutation first implicated a role of NBD1 in CF pathophysiology. To better understand the detailed molecular mechanisms associated with the F508del mutation, further, more refined biophysical analyses were needed. As the bacterial ABC transport systems suggested that the NBDs could fold and function autonomously, the isolated NBD domain was produced to evaluate the impact of the F508del mutation on its folding and function (6, 11, 12, 38). These data could then be correlated with the trafficking and function of the full-length CFTR protein. Importantly, the isolation of this domain has greatly simplified biochemical experimentation and has allowed for solution-based studies otherwise untenable with a transmembrane protein. In addition, the development of NBD1 reagents and accessibility to milligram quantities of soluble, highly purified NBD1 protein have facilitated new studies on NBD1 folding, stability, and function.

#### 3.3.1. NBD1 Production

Methods for the production of large quantities of CFTR NBD1 were initially generated by Structural GenomiX (SGX) working with Cystic Fibrosis Foundation Therapeutics (11). From these efforts, multiple NBD1 structures have been solved from *Mus musculus* and *Homo sapiens*. Though domain boundaries and specific site mutations have been engineered for both species, expression and purification of soluble NBD1 follow similarly for the murine and human NBDs.

Expression and purification follow standard techniques for recombinant protein production in *Escherichia coli* and are based on the methods presented in Refs. (7, 11). The initial expression and purification of murine NBD1 utilized a fusion strategy,

with the NBD1 sequence fused, in frame, to a $His_6$-Smt3 tag (39). This Smt3-NBD1 protein is expressed using the pET T7 protein expression system in BL21(DE3) *E. coli*. Clonal cultures are started and grown overnight to inoculate expression cultures. Expression cultures are grown to induction at 37°C with constant agitation and antibiotic selection. At induction, the cultures are shifted to low temperature for expression (between 15 and 20°C, depending on species and mutational background). Protein expression is then induced by addition of IPTG, triggering expression of the T7 RNA polymerase and the T7-regulated Smt3-NBD1 fusion protein. Protein is expressed for 12–16 h under these conditions.

Cells from the expression culture are harvested by centrifugation and lysed by sonication or French press. The lysates are clarified by centrifugation prior to purification. NBD1 protein is found in both the supernatant and the insoluble fractions after centrifugation and the relative quantities of NBD found in either pool vary with NBD1 construct, species, and mutational background (quantity of soluble mouse wild-type NBD1 > mouse F508del-NBD1 ≈ human wild-type NBD1 > human F508del-NBD1) (7, 9, 11). The soluble fraction of NBD1 protein is bound to Ni-NTA resin in either column or batch format. The protein is washed and eluted with increasing concentrations of imidazole and eluted fractions containing the partially purified NBD1 are collected.

Following initial purification, the $His_6$-Smt3 fusion is cleaved using Ulp1 protease. The Ulp1 protease recognizes the specific tertiary structure of the Smt3 protein and cleaves at the junction between the Smt3 and NBD1 proteins. This cleavage is highly efficient, resulting in no non-specific cleavage or breakdown of the NBD1 protein. The resulting fragments are then partially separated by gel filtration chromatography and the residual $His_6$-Smt3 protein is removed by a second Ni-NTA chromatography step. As the NBD1 protein no longer contains a $His_6$-tag, it remains unbound and can be separated from its fusion partner. The resulting protein can then be concentrated and frozen prior to use. Using these techniques, milligram quantities of protein can be produced and concentrated to multiple milligrams per milliliter concentrations, sufficient for both structural and biochemical studies.

#### *3.3.2. Evaluation of NBD Properties*

Classic protein folding studies have relied on biochemical experiments to quantify both kinetic and thermodynamic parameters associated with various folding and unfolding transitions in model proteins (28). These model proteins are usually small, highly soluble, single-domain proteins and folding studies are run under idealized in vitro conditions. Quantification of thermodynamic and kinetic parameters is often limited to these small proteins as

protein solubility and reversibility of the folding/unfolding reaction are often confounded by aggregation in more complicated protein folds (those with re-entrant sequences and high contact order). Further, these analyses are often modeled as two state, in part due to the difficulty in isolating and measuring intermediates, and as a means of simplifying the data analyses.

3.3.2.1. Kinetic Analyses of NBD1 Folding

Early studies of CFTR folding suggested that the F508del mutation resulted in a conformational arrest of CFTR in an intermediate state due to alterations in the kinetics of folding (40, 41). Based on the rationale for studying NBD in isolation described previously, kinetic analyses of NBD1 folding and unfolding were undertaken to ascertain the impact of mutation in this domain.

The analysis of NBD1 folding kinetics has been studied using two different kinetic methods. First, rapid mixing, or stop-flow, kinetic experiments have been performed to assess the rates of folding and unfolding of purified NBD1 (12). These experiments are completed by measuring an intrinsic spectral parameter of the purified protein as it is rapidly mixed in or out of the denaturant. These experiments are generally performed under conditions where the folding reaction is reversible and the product of the reaction is a single structural state. The spectral data from these experiments are fit to generate one or more discrete rate constants and can be used to characterize individual steps in a folding/unfolding pathway.

Second, "kinetic partitioning" experiments have been used to evaluate the relative rates of multiple pathways. These experiments are performed under conditions where competing pathways promote the formation of multiple structural species that can be separated from one another (i.e., folded and aggregated protein) (6). The relative quantities of the two separable species can then be used to assess the relative rates of the competing pathways (**Fig. 23.3a**). These experiments generally do not provide discrete rate parameters, but instead describe the overall efficiency of the reactions under evaluation.

Studies with NBD1 have suggested that the global parameters of protein folding and unfolding are unaffected by the F508del mutation. That is to say, in stop flow and rapid mixing experiments, there is no obvious change in the rate constants of folding and unfolding as a function of the F508del mutation. The aggregate folding and unfolding do not appear to be impacted by the mutation under conditions that promote protein folding (12). In contrast, under conditions where both folding and aggregation can occur, the F508del mutation impacts the formation of folded, soluble protein. Specifically, when folding reaction temperature is used as a variable to manipulate the multiple folding and aggregation pathways, the F508del protein produces less soluble protein than the wild-type NBD1. Specifically, as temperature increases

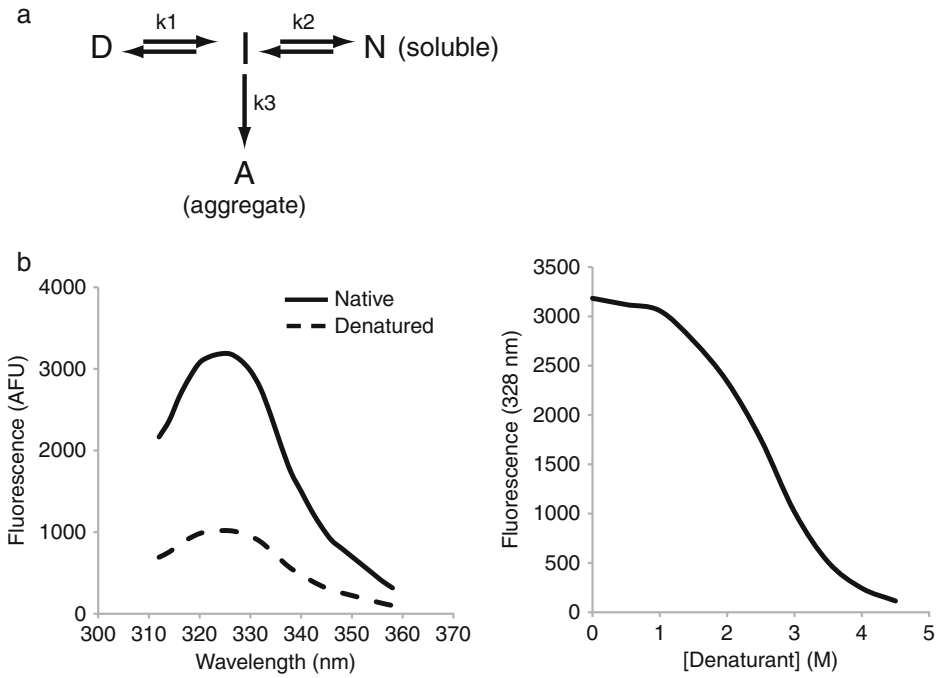


Fig. 23.3. (**a**) Simplified kinetic model for partitioning experiments. A four-state model is shown illustrating the concept behind the kinetic partitioning experiments. Denatured protein (D) is refolded under conditions where both the native state (N) and an off-pathway state (A) can be populated. The kinetic competition between $k_2$ and $k_3$ will result in portioning between the off-pathway, aggregated state, and the soluble, native state. For CFTR NBD1, the competing processes that result in the native and off-pathway states compete and are impacted by the F508del mutation (6, 42). Transition to the off-pathway aggregate is assumed to be irreversible; reverse rates are omitted for clarity. (**b**) Thermodynamic characterization in vitro. The concentration dependence of changes in biochemical parameters can be used to assess thermodynamic properties of a purified protein. Fluorescence of the native and denatured states of purified NBD protein can be monitored as a function of titration with chemical denaturant, *left*. The denaturant dependence of the transition from native to denatured, *right*, can be used to assess apparent thermodynamic parameters of the folding and unfolding transitions.

the amount of soluble F508del-NBD decreases relative to the amount of soluble wild-type NBD1 (6). From these experiments it is not possible to know exactly what steps are impacted by the F508del mutation; however, the decrease in soluble NBD1 production is consistent with local alteration of the NBD1 structure and/or alteration to its folding and aggregation pathways.

3.3.2.2. Thermodynamic Analyses of NBD1

The second critical parameter in defining the reaction pathway for protein folding is the thermodynamic stability of the various structural states. Again, protein folding pathways are generally assumed to be two-state, though examples of metastable intermediates and multi-state reactions have been reported in the literature (28). The difference in energy between any two states (i.e., folded and unfolded, folded and intermediate) can be measured and, as in other chemical reactions, this difference in energy is the driving force for the reaction.

As with the kinetic studies described above, characterization of thermodynamic properties is often based on monitoring a spectral property of the experimental protein (fluorescence, secondary structure). The experimental sample is incubated with denaturant or heated and solution-based studies completed. For rigorous thermodynamic parameters to be calculated, the reaction must be reversible and show no hysteresis (that is to say, it must be the same in the forward and reverse directions). However, relative stability – sensitivities to perturbations by heat and/or temperature – can also be informative and does not require reaction reversibility.

Initial studies on NBD1 stability relied on the folding and unfolding of model NBD1 proteins and suggested that the stability of the fold was not sensitive to the F508del mutation (**Fig. 23.3b**) (6, 7, 12). That is to say, the F508del mutation had no significant impact on the stability of the native state relative to the wild-type protein (42). However, emerging evidence suggests that the NBD proteins undergo multiple structural transitions in their denaturation and that these transitions may be sensitive to the F508del mutation. Significantly, and consistent with prior studies, complete folding and unfolding show similar relative stabilities; however, transition through one or more intermediate states appears to be impacted by the F508del mutation. Thus, the native–denatured states transition appears to be similar in the wild type and F508del proteins, though individual transitions between the (1) native and intermediate and (2) intermediate and denatured states appear sensitive to the F508del mutation. Further experimental characterization of these structural alterations is necessary to fully elucidate the impact of the F508del mutation on domain stability.

## 4. Notes

1. Apart from the details and accuracy of the protein model or of the MD algorithm, most if not all of MD simulations of protein folding aim to recapitulate the folding of the protein in vitro. While the complete CFTR in vitro is known to fold co-translationally in vivo with the aid of chaperones (34), this effect is not included in the modeling just described. Folding simulations of NBD1 using MD investigate the intrinsic folding properties of the domain as dictated by the physico-chemical nature of the polypeptide chain. However, focusing on NBD1, both MD simulations and experimental refolding experiments already show that Phe508 deletion can impair the intrinsic folding pathway of the protein.

2. The use of the structure-based potential effectively smoothens the folding landscape of the protein and shortens the search time for the native structure to within computationally tractable realm. Since the structural information of the native structure is used as input in the modeling, it is necessary that crystal structures of the native structure be known a priori. For the case of CFTR NBD1, while there are several structures of wild type and mutants, most of these contain solubilizing and rescue mutations, which have been shown experimentally to affect the folding properties of the whole protein. In the absence of pure wild-type NBD1 and F508del, one may argue that the folding rescue effects of the accompanying mutants are embedded in the simulation and that any observation already reflects rescue effects of the protein. Nonetheless, to address this issue, we choose structures of wild type and mutant NBD1 that contain the same background of solubilizing mutations such that any observed difference in the folding properties from the simulations can be attributed to the presence or absence of the Phe508 residue.

## Acknowledgments

The authors would like to thank Drs. J. R. Riordan, A. L. Aleksandrov, L. Cui, L. He, F. Ding, and T. Hegedus. We also acknowledge the support of the Cystic Fibrosis Foundation.

### References

1. Riordan, J. R., Rommens, J. M., Kerem, B. S., Alon, N., Rozmahel, R., Grzelczak, Z., et al. (1989) Identification of the cystic-fibrosis gene – cloning and characterization of complementary-DNA. *Science* **245**, 1066–1072.
2. Serohijos, A. W., Hegedus, T., Aleksandrov, A. A., He, L., Cui, L., Dokholyan, N. V., 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.
3. He, L., Aleksandrov, A. A., Serohijos, A. W., Hegedus, T., Aleksandrov, L. A., Cui, L., 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.
4. Shakhnovich, E. (2006) Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem. Rev.* **106**, 1559–1588.
5. Bowie, J. U. (2005) Solving the membrane protein folding problem. *Nature* **438**, 581–589.
6. Qu, B. H., Strickland, E. H., and Thomas, P. J. (1997) Cystic fibrosis: a disease of altered protein folding. *J. Bioenerg. Biomembr.* **29**, 483–490.
7. Thibodeau, P. H., Brautigam, C. A., Machius, M., and Thomas, P. J. (2005) Side chain and backbone contributions of Phe508 to CFTR folding. *Nat. Struct. Mol. Biol.* **12**, 10–16.

8. Cyr, D. M. (2005) Arrest of CFTRDeltaF508 folding. *Nat. Struct. Mol. Biol.* **12**, 2–3.
9. Pissarra, L. S., Farinha, C. M., Xu, Z., Schmidt, A., Thibodeau, P. H., Cai, Z., et al. (2008) Solubilizing mutations used to crystallize one CFTR domain attenuate the trafficking and channel defects caused by the major cystic fibrosis mutation. *Chem. Biol.* **15**, 62–69.
10. Ginalski, K., and Rychlewski, L. (2003) Detection of reliable and unexpected protein fold predictions using 3D-Jury. *Nucleic Acids Res.* **31**, 3291–3292.
11. Lewis, H. A., Buchanan, S. G., Burley, S. K., Conners, K., Dickey, M., Dorwart, M., et al. (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *EMBO J.* **23**, 282–293.
12. Lewis, H. A., Zhao, X., Wang, C., Sauder, J. M., Rooney, I., Noland, B. W., 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.
13. Callebaut, I., Eudes, R., Mornon, J. P., and Lehn, P. (2004) Nucleotide-binding domains of human cystic fibrosis transmembrane conductance regulator: detailed sequence analysis and three-dimensional modeling of the heterodimer. *Cell. Mol. Life Sci.* **61**, 230–242.
14. Dawson, R. J. P., and Locher, K. P. (2006) Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185.
15. Hollenstein, K., Dawson, R. J. P., and Locher, K. P. (2007) Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **17**, 412–418.
16. Hollenstein, K., Frei, D. C., and Locher, K. P. (2007) Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213–216.
17. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P., and Rees, D. C. (2007) An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315**, 373–377.
18. Ward, A., Reyes, C. L., Yu, J., Roth, C. B., and Chang, G. (2007) Flexibility in the ABC transporter MsbA: alternating access with a twist. *Proc. Natl. Acad. Sci. USA* **14**, 19005–19010.
19. Tusnady, G. E., Dosztanyi, Z., and Simon, I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res.* **33**, D275–D278.
20. Chang, X. B., Hou, Y. X., Jensen, T. J., and Riordan, J. R. (1994) Mapping of cystic-fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. *J. Biol. Chem.* **269**, 18572–18575.
21. Tusnady, G. E., and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850.
22. Ding, F., and Dokholyan, N. V. (2006) Emergence of protein fold families through rational design. *PloS Comput. Biol.* **2**, 725–733.
23. Ding, F., Tsao, D., Nie, H., and Dokholyan, N. V. (2008) Ab initio folding of proteins with all-atom discrete molecular dynamics. *Structure* **16**, 1010–1018.
24. Yin, S., Ding, F., and Dokholyan, N. V. (2007) Modeling backbone flexibility improves protein stability estimation. *Structure* **15**, 1567–1576.
25. Yin, S., Ding, F., and Dokholyan, N. V. (2007) Eris: an automated estimator of protein stability. *Nat. Methods* **4**, 466–467.
26. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
27. Dill, K. A., Ozkan, S. B., Shell, M. S., and Weikl, T. R. (2008) The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316.
28. Chen, Y., Ding, F., Nie, H., Serohijos, A. W., Sharma, S., Wilcox, K. C., et al. (2008) Protein folding: then and now. *Arch. Biochem. Biophys.* **469**, 4–19.
29. Ding, F., and Dokholyan, N. V. (2005) Simple but predictive protein models. *Trends Biotechnol.* **23**, 450–455.
30. Serohijos, A. W., Hegedus, T., Riordan, J. R., and Dokholyan, N. V. (2008) Diminished self-chaperoning activity of the DeltaF508 mutant of CFTR results in protein misfolding. *PLoS Comput. Biol.* **4**, e1000008.
31. Khare, S., Ding, F., and Dokholyan, N. V. (2003) Hybrid molecular dynamics studies on Cu,Zn superoxide dismutase reveal topologically important residues. *Abstr. Pap. Am. Chem. Soc.* **225**, U704–U704.
32. Sharma, S., Ding, F., Nie, H., Watson, D., Unnithan, A., Lopp, J., et al. (2006) iFold: a platform for interactive folding simulations of proteins. *Bioinformatics* **22**, 2693–2694.
33. Hubner, I. A., Shimada, J., and Shakhnovich, E. I. (2004) Commitment and nucleation in the protein G transition state. *J. Mol. Biol.* **336**, 745–761.

34. Goldberg, A. L. (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899.
35. Davidson, A. L., and Chen, J. (2004) ATP-binding cassette transporters in bacteria. *Annu. Rev. Biochem.* **73**, 241–268.
36. Dean, M., Rzhetsky, A., and Allikmets, R. (2001) The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11**, 1156–1166.
37. Wilken, S., Schmees, G., and Schneider, E. (1996) A putative helical domain in the MalK subunit of the ATP-binding-cassette transport system for maltose of Salmonella typhimurium (MalFGK2) is crucial for interaction with MalF and MalG. A study using the LacK protein of Agrobacterium radiobacter as a tool. *Mol. Microbiol.* **22**, 655–666.
38. Strickland, E., Qu, B. H., Millen, L., and Thomas, P. J. (1997) The molecular chaperone Hsc70 assists the in vitro folding of the N-terminal nucleotide-binding domain of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **272**, 25421–25424.
39. Mossessova, E., and Lima, C. D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5**, 865–876.
40. Cheng, S. H., Gregory, R. J., Marshall, J., Paul, S., Souza, D. W., White, G. A., et al. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827–834.
41. Lukacs, G. L., Mohamed, A., Kartner, N., Chang, X. B., Riordan, J. R., and Grinstein, S. (1994) Conformational maturation of CFTR but not its mutant counterpart (delta F508) occurs in the endoplasmic reticulum and requires ATP. *EMBO J.* **13**, 6076–6086.
42. Richardson, J. M., Thibodeau, P. H., Watson, J., and Thomas, P. J. (2007) Identification of a non-native state of NBD1 that is affected by ΔF508. *Pediatr. Pulmonol. Suppl.* **30**, 1.
43. Mendoza, J. L., and Thomas, P. J. (2007) Building an understanding of cystic fibrosis on the foundation of ABC transporter structures. *J. Bioenerg. Biomembr.* **39**, 499–505.
44. Hegedus, T., Serohijos, A. W., Dokholyan, N. V., He, L., and Riordan, J. R. (2008) Computational studies reveal phosphorylation dependent changes in the unstructured R domain of CFTR. *J. Mol. Biol.* **378**, 1052–1063.