Overexpression and characterization of the cystic fibrosis protein CFTR using AFM and MALDI-TOF-MS
by David Guire Wooster

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biochemistry
Montana State University

Abstract:
A need exists for the development of additional methods capable of providing structural information about large transmembrane proteins such as the protein responsible for cystic fibrosis, the cystic fibrosis transmembrane conductance regulator, (CFTR), where traditional methods like NMR and X-ray crystallography have not been applicable. In the study reported here, CFTR was initially tagged with a poly(histidine) tail and overexpressed in a baculovirus system, allowing for the only one-step purification of CFTR that has been reported, using a nickel-affinity chromatography column.

Sufficient CFTR was obtained using this method to enable the mapping of 80 proteolytic peptides, identified by MALDI-TOF MS, onto the primary sequence of CFTR. This same method was then used to obtain 3D structural information concerning CFTR by exposing native CFTR, still within cellular membranes, to the hydrophilic, cysteine-alkylating reagent IAA and observing changes in mass. One cysteine located within the predicted pore region of CFTR, Cys-343, was found to be IAA-accessible and therefore accessible to the solvent. However, 3 additional cysteine residues, one each within the NBD1 domain, R-domain, and C-terminal tail region, were found to be inaccessible to IAA and therefore predicted to be buried within the 3D structure of CFTR. This is the first method concerning the use of mass spectrometry on whole CFTR in its native environment. In another study, AFM was used to characterize purified and liposome-reconstituted CFTR from this lab for the purpose of detecting protein-protein interactions involving CFTR. Here, the cytoskeleton protein F-actin was found to interact with CFTR without the need for accessory proteins. Additionally, this study constituted the first reported attempt at using atomic force microscopy to characterize protein-protein interactions involving CFTR, or any other transmembrane protein, in a lipid environment.

OVEREXPRESSION AND CHARACTERIZATION OF THE CYSTIC FIBROSIS PROTEIN CFTR USING AFM AND MALDI-TOF-MS

by

David Guire Wooster

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY-BOZEMAN
Bozeman, Montana

April 2002

APPROVAL

of a dissertation submitted by

David Guire Wooster

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Martin Teintze, Ph.D. _____ 4/19/02
(Signature) Date

Approved for the Department of Chemistry and Biochemistry

Paul A. Grieco, Ph.D. _____ 4-19-02
(Signature) Date

Approved for the College of Graduate Studies

Bruce R. McLeod, Ph.D. _____ 4-19-02
(Signature) Date

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University-Bozeman, I agree that the Library shall make it available to borrowers under the rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Signature David G. Woods

Date 4/19/02

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A need exists for the development of additional methods capable of providing structural information about large transmembrane proteins such as the protein responsible for cystic fibrosis, the cystic fibrosis transmembrane conductance regulator, (CFTR), where traditional methods like NMR and X-ray crystallography have not been applicable. In the study reported here, CFTR was initially tagged with a poly(histidine) tail and overexpressed in a baculovirus system, allowing for the only one-step purification of CFTR that has been reported, using a nickel-affinity chromatography column. Sufficient CFTR was obtained using this method to enable the mapping of 80 proteolytic peptides, identified by MALDI-TOF MS, onto the primary sequence of CFTR. This same method was then used to obtain 3D structural information concerning CFTR by exposing native CFTR, still within cellular membranes, to the hydrophilic, cysteine-alkylating reagent IAA and observing changes in mass. One cysteine located within the predicted pore region of CFTR, Cys-343, was found to be IAA-accessible and therefore accessible to the solvent. However, 3 additional cysteine residues, one each within the NBD1 domain, R-domain, and C-terminal tail region, were found to be inaccessible to IAA and therefore predicted to be buried within the 3D structure of CFTR. This is the first method concerning the use of mass spectrometry on whole CFTR in its native environment. In another study, AFM was used to characterize purified and liposome-reconstituted CFTR from this lab for the purpose of detecting protein-protein interactions involving CFTR. Here, the cytoskeleton protein F-actin was found to interact with CFTR without the need for accessory proteins. Additionally, this study constituted the first reported attempt at using atomic force microscopy to characterize protein-protein interactions involving CFTR, or any other transmembrane protein, in a lipid environment.

CHAPTER 1

INTRODUCTION

Cystic Fibrosis and CFTR

The most common lethal genetic disease in North America is cystic fibrosis, with approximately one in every 2,500 infants in the U.S being born with it (1). There are currently over 730 separate mutations within the cystic fibrosis gene identified as causing the disease (2), making carrier frequency for CF in the U.S. high enough that about one in every twenty Americans of Northern European descent are capable of passing on this autosomal recessive disease (1). The first known reference to this disease can be traced to folklore of the Middle Ages, where children's songs suggest midwives were able to notice a connection between salty sweat of infants in their care and their premature death. These afflicted children were often thought to be bewitched (3). Today, through the use of techniques such as microsattelite haplotyping, it has been found that the most common CF-causing mutation, ΔF508, arose over 50,000 years ago, and later became prevalent in the population perhaps as protection against many of the bacteria-induced diarrheal diseases such as cholera that have no doubt been frequent since the beginning of human civilization, due in large part to inadequate sanitation (4). The role of the cystic fibrosis protein in these bacterial-induced diseases is now known to be due to the gene product's importance in the regulation of fluid volume in the intestines, not unlike how this protein is involved in mucus hydration in the lungs. Although CF is a significant genetic

disease in North America, with just over 30,000 known cases in the United States alone, secretory diarrhea due to the overactive CF gene product expressed in the large intestine is the second leading cause of infant mortality in the developing world, estimated to be responsible for up to 3 million deaths per year of children under the age of 5 (5,6). Today, due in large part to the fields of molecular biology, physiology, and biophysics, the ancient disease marker for cystic fibrosis of salty-tasting sweat on a newborn infant's forehead is now understood to be caused by a defect in a chloride ion channel, the cystic fibrosis transmembrane conductance regulator, or simply CFTR.

The CF gene, which codes for CFTR, was discovered in 1989 by a lengthy process involving positional cloning (7). Fifty years before this discovery, in the 1940s, chronic bacterial infections in the lungs of CF patients were, for the first time, being recognized as a major contributor to CF mortality. In 1900, an infant born with CF could have been expected to live less than 5 years. Following the introduction of antibiotics and enzyme replacement therapy after WWII, life expectancy climbed significantly, but has since leveled off to where it is today, at approximately 30 years of age (8). A major disappointment in the field of CF research in the last 12 years has been the realization that, in spite of the knowledge of the exact location of the CF gene, no new effective treatments have emerged utilizing this hard-won information. This is unfortunate because cystic fibrosis is one of only a handful of diseases that can be attributed to mutation in a single gene and is therefore an ideal candidate for molecular therapies, including those involving protein and/or gene replacement. Despite often being described as one of the most studied proteins in the last ten years, structural information

concerning CFTR protein has been scarce, due in large part to difficulties inherent in the characterization of large transmembrane proteins like CFTR. Most significant among these obstacles are poor solubility, low and inconsistent yields using eukaryotic over-expression systems, and poor signal to noise ratios inherent in spectroscopic techniques (due to light scattering of lipid membranes). Additionally, transmembrane proteins are too large for NMR study, and are difficult to crystallize for X-ray diffraction studies.

Several theories have been put forward in an attempt to explain precisely how a defect in a single chloride channel could cause cystic fibrosis, with no single one of these theories being totally accepted. The most commonly repeated theory revolves around the notion that the CFTR chloride channel is necessary for the proper hydration of the mucus coating secretory epithelia, primarily those found in the lungs (9). Because tissues do not possess pumps for moving water from one compartment to another, plants and animals have instead come to rely on the formation of osmotic salt gradients to encourage the movement of water between various compartments. Normally, water in the airway surface fluid of the lung lumen finds its way there from the bloodsteam by following sodium chloride gradients set up by ion channels as shown in figure 1. One of the most vital of these ion channels is CFTR. When CFTR is not found at the apical membrane of lung epithelial cells in active form, such as what occurs in CF, water cannot move from the bloodstream and into the airway surface fluid of the lung to hydrate the mucus and allow it to be removed, along with dust and microorganisms, in a timely fashion. Instead, the mucus becomes overly viscous and difficult to remove from the

lungs, setting the stage for the well-known chronic bacterial infections and inflammation that afflict CF patients during most of their short lifetimes.

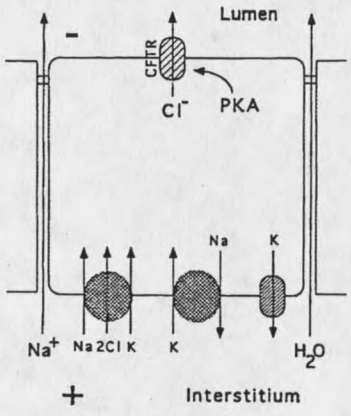

Figure 1: **Secretory epithelial cell** (e.g. lung, intestine) indicating the role of CFTR in directing net H_2O flow from the interstitium (bloodstream) into the lumen for hydration of the mucus layer (not shown). Note that Na+ flows around cells in order to neutralize negative charge built up by Cl- flow through cell via the CFTR chloride channel. **PKA**: a kinase activated by increases in cAMP. When activated, PKA phosphorylates serines within the R-domain of CFTR triggering channel opening. The driving force for Cl- secretion is the build-up of negative charge due to the flow of K+ out of the cell via the potassium channel on the basolateral side of the cell.

Search for the CF Gene

In the late 1980s, two collaborating groups, one at the University of Toronto and the other at the University of Michigan, were closing in on the location of the CF gene on the long arm of human chromosome 7 using RFLP markers obtained from families known to be carriers of CF. These researchers were aware that the technique they were using, chromosomal walking, carried with it the inherent limitation that it was impossible to narrow down the putative CF gene to less than 5 total "candidate genes" on chromosome 7 in the area between their closest RFLP markers (7). A hypothesis was therefore needed to help determine which of these 5 genes on human chromosome 7 was the CF gene. It was therefore reasoned that any single gene capable of causing a disease as devastating as CF by itself was likely to be conserved among various mammals. One way of testing this hypothesis meant the generation of a series of Zoo Blots (or, Southern Blots performed on genomes across several species) of the genomic DNA of a diverse group of mammalian species (10). Restriction digests of these mammalian genomic libraries were probed with short nucleotide sequences based on each of the 5 candidate human genes. It was soon clear from the Zoo Blotting analysis that only one of the 5 genes was located within each of the mammalian genomes tested. This gene turned out to be a 250 KB stretch of DNA on the long arm of chromosome 7, band q31, and became the new focus of their work (7). The next step towards proving that this was indeed the CF gene entailed obtaining full-length cDNA copies of the mRNA from a subject without CF and then performing Northern Blotting analysis for the purpose of locating the

expression pattern of the gene in various human tissue samples. It was reasoned that if this unknown gene was the CF gene, then mRNA corresponding to it should be detected in tissues affected in cystic fibrosis, such as the epithelial tissue of the lungs. Once this was verified by Trezise et al. in 1991 (11), full-length cDNA representing the mRNA from the proposed CF gene was tested for its ability to change the properties of an immortalized airway epithelial tissue culture (from a CF patient) back to the wild-type state. This was shown to be the case using retroviral vectors, accompined by patch clamping procedures, providing further confirmation that CFTR was indeed the long sought after CF gene (12,13). Within two years, peptides would be synthesized based on the extracellular loops of CFTR, and antibodies specific to these loops generated for the purpose of detecting the subcellular localization of the CFTR protein, which turned out to be localized to the apical, but not the basolateral, membranes of secretory epithelia in the lungs, where it was expected (14,15).

CFTR: Structure and Function

CFTR is a 1480 amino acid transmembrane glycoprotein, widely expressed in several secretory and absorptive epithelial tissues, including those of the pancreas, lung, large intestine, sweat duct, and vas deferns (16). CFTR is one of only a few members of a widely distributed superfamily of proteins called the ABC Transporters (for **A**TP **B**inding **C**assette) that is not itself a transporter, but rather an ion channel. CFTR also carries with it the distinction of being the only known ABC superfamily member that has 5 separate domains as opposed to 4 domains characteristic of all other ABC proteins, the

5th domain being the "R-domain" (so-named because its function appears to be in regulation of channel gating, along with two NBD domains)(16). However, CFTR is similar to other ABC family members in that it possesses two nucleotide binding domains (NBD1 and NBD2), both of which bind and hydrolyze ATP. CFTR also has two transmembrane domains, each consisting of 6 predicted transmembrane helices (17).

The CFTR gene itself appears to be a relatively ancient gene, believed to have arisen as the result of a gene duplication event of an unidentified ancestral prokaryotic ABC transporter gene prior to the divergence of fishes (18). This gene duplication event was, in all likelihood, then followed by fusion of the duplicated halves into a tandem arrangement along chromosome 7 forming a mirror image of the original transporter gene within the entire, newly fused-together CFTR gene (Figure 2). Some evidence for this

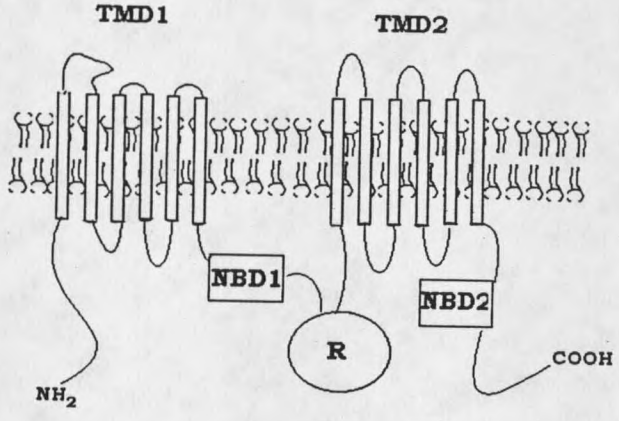

Figure 2: **Predicted topological structure of CFTR** showing the 5 domains. The cytoplasmic portion consists of the R-domain and the two nucleotide binding domains (NBD1 and NBD2), all 3 of which have been shown to be involved in channel gating. Each transmembrane domain (TMD) consists of 6 predicted helices. The most common CF-causing mutation is a deleted Phe-508 in NBD1.

Some evidence for this duplication event followed by a fusion event includes the finding of significant sequence similarity between the two halves of CFTR, as well as the fact that all other ABC transporter proteins not believed to have arisen as the result of gene fusion appear to function normally as dimers with one another (18).

While nearly all ABC transporters appear to use both NBD domains to hydrolyze ATP and transport a wide variety of substances including amino acids, peptides, proteins, small hydrophobic molecules (e.g. the multi-drug resistance protein Pgp), lipids such as cholesterol and phospholipids, sugars, inorganic ions, and polysaccharides (18) against a concentration gradient, CFTR on the other hand appears to have, at some point in its history, managed to "commandeer" its NBD domains into functioning during the process of "gating" of its anion channel, apparently as an added layer of regulation during opening and closing of the channel. One line of evidence for this is that the measured V_{max} of ATP hydrolysis by the CFTR NBD domains matches closely the measured kinetics of CFTR channel opening and closing as determined by patch clamping (~1/sec) (19).

Site-directed mutagenesis has been extensively performed on the CFTR sequence during the last decade, and much of the evidence accumulated from this has suggested that amino acid changes localized to both the NBD and R domains tend to alter the kinetics of channel gating (20, 17), while amino acid changes in the predicted transmembrane domains affect anion channel selectivity as well as conductance, the most likely explanation for this being that the transmembrane domains are where the pore is located (21, 22, 23). Three of the loops of CFTR which, like the other 7 loops, function

to join the transmembrane helices with one another, have also been selectively mutated to produce ion channels having alterations in both halide ion selectivity as well as changes in channel gating kinetics (24). Exon 13 of CFTR encodes the R domain. This exon was inserted into the ancestral CFTR gene following fusion of the duplicated halves as discussed above. Curiously, the R domain of CFTR, while having no known sequence homology to any other eukaryotic or prokaryotic protein domain, has been found by Dulhanty et al. to possess sequence similarity to DNA polymerases of viruses (25). The significance of this finding is still unknown. The CFTR gene also possesses the remnants of L1 retrotransposons within the intron sequences flanking exon 9. Exon 9 encodes a highly conserved section of the NBD1 domain of CFTR, and at various times during human evolution, these L1 elements near exon 9 of CFTR are theorized to have undergone transcription from the opposite direction of the gene. These transcripts were then reverse transcribed into a cDNA copy, and finally co-integrated along with the L1 sequences where they eventually came to rest within at least 10 different locations throughout the human genome, effectively "amplifying" CFTR exon 9 sequences within the human genome (26). This basic mechanism of exon shuffling via the reverse transcription of L1 elements is believed to have occurred in the evolution of many other human genes as well. Also of interest, the two NBD domains of CFTR have been by found by Pratt et al. to contain sequence similarity to a known actin-binding protein (see chapter 4) (27).

The NBD domains of ABC transporters are the most conserved portions of the protein (30-50% similarity), each consisting of a short Walker A and Walker B motif specialized

for the binding and hydrolysis of ATP, respectively (28, 29). Perhaps as an indication of their importance for proper ion channel function, most CF-causing mutations occur in these two NBD domains (30). In addition, there appears to be significant similarity between the NBD domains of CFTR and G-proteins, although threading of the primary sequence of the NBD1 domain of CFTR onto the crystal structure of the G-protein Ras was not considered particularly revealing. However, the glycine in the LSGGQ "signature sequence" (so-named because this sequence, or one like it, is found in all known ABC superfamily members) within the NBD domains of CFTR corresponds to a conserved glycine in G-proteins and is necessary for GTP hydrolysis (31). The presence of this short LSGGQ signature sequence near the Walker motifs of the NBD domains also serves to distinguish the ABC superfamily of proteins from other known ATP-binding protein families (e.g. G-proteins, ATPases, and kinases), and this signature sequence has been shown to be important for the transduction of free energy resulting from ATP binding and hydrolysis (which takes place at the Walker motifs), into mechanical energy for the pumping of substrates (or ion transport gating in the case of CFTR). To date, there are 4 crystal structures of NBD subunits of ABC family members in the literature (Figure 3), and these structures point to a conserved mechanism for the binding and hydrolysis of ATP in ABC transporter family members, as well as the transduction of this energy into opening and closing of the pore region [MalK (32), Rad50 (33), HisP (34), and MJ0795 from Methanococcus jannaschii (35)]. At near-atomic resolution, crystallized NBD domains can be seen as forming an overall "L" shape. The NBD domains have both Walker A and B motifs (and therefore the site of