The University of Manchester

Faculty of Life Sciences

Structural biology of Cystic Fibrosis Transmembrane Conductance Regulator, an ATP-binding cassette protein of medical importance

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD)

By

ATEEQ AHMED AL-ZAHRANI

Supervisor: Professor Bob Ford

Advisor: Dr. Ewan Blanch

2012

Word Count: 49468

Abstract

The cystic fibrosis transmembrane conductance regulator (CFTR) is a transmembrane protein that functions as an ion channel. Mutations in this protein cause Cystic Fibrosis. For this reason, it is important to study the structure and function of CFTR. In this study, constructs of CFTR (C-terminii), a CFTR-interacting protein and full-length CFTR were cloned, expressed and purified for structural and functional studies. The purified C-terminal polypeptides of CFTR were soluble and shown to interact with NHERF1 PDZ 1 (a CFTR-interacting protein). The CFTR C-terminus and NHERF1 PDZ 1 domain were co-expressed and co-purified. The purified complex showed a strong interaction that might induces a conformational change. Site-directed mutation of the C-terminus of CFTR was performed in order to examine the effect of removing a potentially flexible amino acid (Arginine) on protein crystallization.

Pull-down assay experiments with full-length CFTR demonstrated an interaction between CFTR (in DDM detergent) and NHERF1 PDZ $1^{(+)}$. No interaction was observed for CFTR in LPG (a relatively denaturing detergent) and NHERF1, implying that the interaction between the PDZ motive of CFTR and NHERF1 requires a stable folded structure for both proteins. In addition, full-length CFTR in DDM has been studied by electron microscopy and Single Particle Analysis in the presence of NHERF1 PDZ $1^{(+)}$. A 3D structure was generated for the CFTR-NHERF1 PDZ $1^{(+)}$ complex at a resolution of ~ 18 Å. This 3D structure showed a new open conformation of CFTR (V shape). In comparable studies with CFTR alone, a 3D structure was generated at a resolution of 27 Å and this structure showed a closed state as previously reported. This new data suggest a possible role for NHERF1 in terms of CFTR channel gating or activation.

Acknowledgements

I would like to express my gratitude to the following people who have helped me in order to write this thesis:

- My supervisor, Professor Robert Ford who helped me in all the time of research for and writing of this thesis. Also, I would like to thank him for sharing me ideas to start a new project or to improve an existing project. (I write poetry!), but I find myself unable to find words to thank polite and gentle personality like him. Thank you Bob.
- My advisor, Dr Ewan Blanch who supported me in my study and gave me all advices that I need.
- Dr James Birtley who shared me his know-how about Biomolecular techniques and I would like to thank him for the time he spared during the first year of this project while being extremely busy himself. I would not have been able to do any of this work without the support of Dr James. His generous help will stay in my mind forever.
- All the members of my lab group who provided a friendly and relaxed atmosphere for the lab work. Dr Mark Rosenberg, Natasha Cant, Naomi Pollock, Tracy Rimington. Thank you guys. Special thanks to Dr Richard Collins for TEM work and EMAN training. And for Dr Stephen Prince for answers and comments that he always provides. I would like to express my thankful to my lab mate, John Chipperfield for his kind help.
- My family (My father, my mother, my brothers and sisters) for their love that always encourages me to do more. Also, I would like to thank my friends, Mesfer Al-ghamdi, Hussain Al-malki and Ahmed Al-essi for their appreciated advices and support.
- My small family, my wife and my children for their understanding these past years.
- My government, for the financial support throughout this project.
- Finally, there are lots of other people who supported me and I would like to express my grateful to them but they were not mentioned in this limited chance.

Table of Contents

List of Tables:

List of Figures:

List of Abbreviations

Å	Angstrom
aa	Amino acid
ABC	ATP-binding cassette
ADP	Adenosine diphosphate
ALD	Adrenoleukodystrophy
AMD	Age-related macular degeneration
APS	Ammonium persulfate
AQP	Aquaporin water channel
ATP	Adenosine triphosphate
a.u	Arbitrary unit
BCRP	Breast cancer resistance protein
BHK	Baby hamster kidney
BOG	n-octyl--D-glucopyranoside
bp	Base pair
BR	Bacteriorhodospin
***c*AMP**	Cyclic adenosine monophosphate
CAL	CFTR-associated ligand
CBAVD	Congential bilateral absence of the vas deferens
CD	Circular dichroism
cDNA	Complementary DNA
CF	Cystic fibrosis
CFTR	The cystic fibrosis transmembrane conductance regulator
CHO	Chinese hamster ovary
CRD	Cone-rod dystrophy
CTF	Contrast transfer function
Da	Dalton

DDM	n-dodecyl- β- D- maltoside
DMPC	Dimyristoyl phosphatidyl choline
DNA	Deoxyribonucleic acid
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EBP50a	Beta-catenin-associating protein
EM	Electron microscope
ENaC	Epithelial Na^+ channel
ER	Endoplasmic reticulum
ERM	Ezrin, radixin and moesin
FBHHI	Familial persistent hyperinsulinemic hypoglycemia of infancy
FDA	Food and Drug Administration
FFM	Fundus flavimaculatis
FHDLD	Familial hypoapoproteinemia
FSC	Fourier shell correlation
GST	Glutathione S-transferase
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
His	Histidine
hNBD	Human nucleotide-binding domain
kDa	Kilodalton
ICOR	Intermediate-conductance outwardly rectifying
ICP	Intrahepatic cholestasis of pregnancy
LPC	Llysophosphatidylcholine
LPG	Lysophosphatidylglycerol
LPR	Lipid to protein ratio
M	Membrane spanning segment
MALDI-MS	Matrix-assisted laser desorption/Mass spectrometry

MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MALS	Multi-angle light scattering
MDCK	Madin-Darby Canine Kidney
mg	Milligram
µg	Microgram
ml	Milliliter
µl	Microliter
MQ water	(MQ water) was obtained by using a Milli-Q purification system
MRP	Multidrug resistance-associated protein
MsbA	Lipid A export ATP-binding/permease protein
MSD	Membrane-spanning domain
NBD	Nucleotide-binding domain
ng	Nanogram
NHERF	Na+/H+ exchanger regulatory factor
nm	Nanometer
NMR	Nuclear magnetic resonance
OD	Optical density
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pdb	Protein data bank
PDBTM	Protein data bank of transmembrane proteins
PDZ	PSD-95/Dlg/ZO-1
PFIC	Progressive familial intrahepatic cholestasis
PKA	Protein kinase A
R- domain	Regulatory domain
ROMK	The Renal outer medullary potassium channel
RP	Retinitis pigmentosum 19

rpm	Revolution per minute
SDS	Sodium dodecylsulphate
SDS-PAGE	Sodium dodecylsulphate Polyacrylamide Gel
SPA	Single particle analysis
SPR	Surface plasmon resonance
TAE	Tris-acetate-EDTA
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TMD	Transmembrane domain
TNP-ATP	Trinitrophenyl-ATP
TRIS	Tris (hydroxymethy) amino methane
UV	Ultraviolet
XLCA/A	X-linked sideroblastosis and anemia
2D	Two dimensional
3D	Three dimensional

1. Introduction

Proteins are macromolecules composed of amino acids which are joined together by peptide bonds. They can be found in all living organisms, from the simplest such as viruses to more complex like humans. The role of proteins in all cells is complicated. Most biological processes are performed by functional and structural proteins inside and outside cells. Antibodies, transporters, receptors, enzymes, structural elements and much more are examples of proteins that have functional and structural roles. Among protein functions, transport of ions and other substances through the cell membrane is considered a worthy target for medical intervention. The mechanism of ion transport is related to many diseases. For this reason, capturing the three dimensional structure of transporter proteins is important in order to find out the relationship between defective proteins and their clinical aspects. In general, there are three main steps that are required to purify proteins to grow 2D or 3D crystals. First of all, expression of the targeted protein by adding a plasmid containing the corresponding gene sequence to host cells. Secondly, purifying the protein in order to separate it from other molecules and cellular substances. Finally, crystallizing the pure protein. All three steps will be explained and discussed later.

1.1 Transmembrane Proteins

The plasma membrane is a selective biological barrier located between the extracellular and intracellular cell fluids. Two main classes of biomolecules form the plasma membrane, proteins and lipids as shown in Figure 1. The location of transmembrane proteins usually influences their functions. Proteins located on the extracellular membrane surface interact with proteins or molecules and serve as active mediators between the cell and its surrounding environment. The proteins which are located on the intracellular membrane surface have many functions from anchoring cytoskeleton proteins to the membrane to triggering intracellular signaling pathways (Lodish et al, 2000). Proteins which are attached or embedded in the plasma membrane have several tasks such as carriers, receptors and adhesion sites. Beside these functions, membrane proteins convert the energy of sunlight into chemical energy by the contribution to ATP synthesis. Furthermore, membrane proteins

play an important role as a signal receptor that receives signals such as neurotransmitters, growth factors, hormones, light or chemotactic stimuli (Branden and Tooze, 1998). Because membrane proteins are involved in essential biological processes as revealed previously, they are considered as major drug targets. The study of this sub-class of proteins could lead to cures for several illnesses such as heart disease, cystic fibrosis and depression (Lunn, 2010).

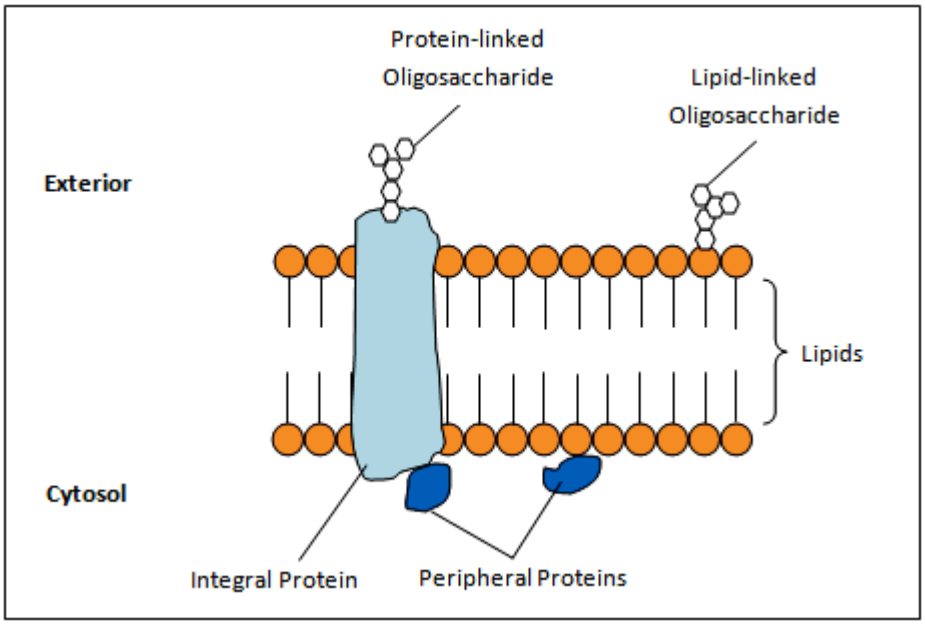


Figure 1: The two main biomolecules (proteins and lipids) found in the plasma membranes. Both can also be linked to carbohydrate moieties, as shown.

There are two kinds of protein interacting with the plasma membrane, integral or peripheral. The latter can transport ions and molecules across the plasma membrane (Thiriet, 2007). Because of the unique and essential physiological role of transporter proteins, mutations in genes encoding these proteins result in human diseases (Opella, 2003). Despite the biological importance that 25 % of proteins encoded by the genome are membrane proteins, no more than 1 % of the total files in the pdb (Protein Data Bank) are membrane proteins (Loll, 2003). According to the recent PDBTM record (Protein Data Bank of Transmembrane Proteins); there are 1550 transmembrane structures of 79538 structures. PDBTM: http://pdbtm.enzim.hu/? Last accessed 01-02-2012.

1.2 ABC Proteins

The ATP-binding cassette (ABC) transporters are the largest family of transmembrane proteins. They can be found in prokaryotes and eukaryotes as integral proteins embedded in the plasma membrane (Dean *et al*., 2001). Several ABC proteins act as transporters by which many substances, small, large, charged and hydrophobic are moved into or-out cells and this explains the diversity of their physiological roles (Higgins, 2001). They are divided into two groups as full transporters consisting of two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs), or half transporters consisting of one TMD and one NBD (Hyde *et al*., 1990). In the latter case, two half transporter proteins must associate in the cell to form an entire transporter. Despite the structural similarity among ABC proteins, there are some exceptions. For examples, the MRP (multidrug resistance-associated protein) (ABCC) subfamily contains an additional multispanning TMD which is found in a high majority of these family members (Hipfner et al, 1999**)**. In addition, many ABC proteins have NBD1-TMD2 linker polypeptides of variable length which in some cases is involved in regulating gating by phosphorylation (Schwiebert *et al*., 1999b). This linker is present in CFTR and is known as the regulatory domain (R domain) (Senior and Gadsby, 1997).

This family of proteins is divided into 7 subfamilies (ABCA, ABCB….ABCG) consisting of 48 members in humans (Dean, 2005; Kaminski *et al.*, 2006). Furthermore, ABC pseudogenes were identified in the human genome (Piehler et al., 2006, 2008). The role of these proteins depends on using ATP energy so as to drive the transport of a large number of molecules, such as sugars, lipids, peptides, ions, through cell membranes. ABC proteins can be found inside the cells of all living organisms. For example, the *Escherichia coli* genome has nearly 5 % of its genes as ABC protein genes (around 70 transporters) (Holland *et al*., 2003). And this big number of transport proteins reflects the need to adapt to a wide range of environments that bacteria live in. One aim of studying this family of proteins is to discover the relationship between defective ABC protein genes, subsequently their proteins, and many hereditary diseases associated with defective transport functions. During the past years, this relationship has been explored and it is known that complex

diseases are related to defective ABC proteins (Wenzel *et al.*, 2007). Mutations in 14 human ABC transporter genes are the main cause of different inherited diseases (Kaminski *et al.*, 2006). Table 1 shows ABCs associated with human diseases.

GENE	MONOGENIC DISORDER	COMPLEX DISEASE
ABCA1	Tangier disease, FHDLD	
ABCA4	Stargardt/FFM, RP, CRD, CD	AMD
ABCB1	Ivermectin susceptibility	Digoxin uptake
ABCB2	Immune deficiency	
ABCB3	Immune deficiency	
ABCB4	PFIC3	ICP
ABCB7	XLCA/A	
ABCB11	PFIC2	
ABCC2	Dubin-johnson Syndrome	
ABCC6	Pseudoxanthorma elasticum	
ABCC7	Cystic Fibrosis, CBAVD	Pancreatitis, bronchiectasis
ABCC8	FBHHI	
ABCD1	ALD	
ABCG5	Sitosterolemia	
ABCG8	Sitosterolemia	

Table 1: Disease and phenotypes caused by ABC genes. The Table shows the relationship between ABC genes (and their proteins as a final product) and diseases. Some diseases can spread their effects to other locations as it is with ABCA4, ABCB1, ABCB4 and ABCC7. FHDLD, familial hypoapoproteinemia; FFM, fundus flavimaculatis; RP, retinitis pigmentosum 19; CRD, cone-rod dystrophy; AMD, age-related macular degeneration; PFIC, progressive familial intrahepatic cholestasis; ICP, intrahepatic cholestasis of pregnancy; XLCA/A, X-linked sideroblastosis and anemia; CBAVD, congential bilateral absence of the vas deferens; FBHHI, familial persistent hyperinsulinemic hypoglycemia of infancy; ALD, adrenoleukodystrophy. The table was obtained from (Moitra, 2012).

The 48 members of the ABC family are expressed in different locations inside the human body (Venter *et al*., 2001). Despite the fact that some ABC transporters reduce the amount of harmful substances, they also reduce some useful substances such as chemotherapeutic drugs (Choudhuri and Klaassen, 2006). Physiologically, the absorption, metabolism and toxicity of pharmacological agents are balanced via these proteins (Glavinas *et al*., 2004). The resistance to chemotherapy is a physiological process that is related to ABC proteins. Sharom (2008) has studied this process in three ABC proteins, P-glycoprotein (ABCB1), MRP1 (ABCC1) and BCRP (breast cancer resistance protein) (ABCG2). These transporters, which can be found in overexpressing human tumors, are involved in a common physiological process in which tissues are protected from xenobiotics and endogenous metabolites. The hydrophobic regions of many cancer drugs interact with the hydrophobic and H-bonding regions of any binding sites of theses tranporters. These interactions result in transport of the drugs and more resistance to chemotherapy. P-glycoprotein (Pgp) is one of the ABC proteins that transports various xenobiotics (including anticancer drugs) out of cells by hydrolysing ATP (Di Pietro et al., 1999).

1.3 The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

Defective CFTR is the cause of two major diseases: cystic fibrosis (CF) (Cheng et al, 1990) and secretory diarrhea (Gabriel et al, 1994) . In secretory diarrhea, an extremly high CFTR activity due to the effects of bacterial toxin (e.g cholera toxin and heat stable *E. coli* enterotoxin) induces secretory diarrhea (Chao et al, 1994). Cystic fibrosis (CF) is an inherited (Mendelian autosomal recessive) chronic disease that is more frequent in the caucasian race (Damas et al, 2008) and is associated with loss of CFTR activity. CF affects around one in 2500 newborns. There are about 30,000 children and adults in the United States (70,000 worldwide) affected by CF. Cystic Fibrosis Foundation [homepage on the Internet]. About Cystic Fibrosis. Last update (08-09-12). Available from: http://www.cff.org/AboutCF/. Mutations in CFTR gene cause the body to produce thick and sticky mucus that prevent the organs from doing their biological functions. The defect CFTR has been associated with decreased in chloride ion transport across the apical membrane of the epithelial cells

(Widdicombe et al, 1985). The main organs which can be affected by cystic fibrosis (CF) are the lungs, the gastrointestinal tract, liver, exocrine pancreas and the male reproductive tract (Sheppard and Welsh, 1999). Patients who are diagnosed with CF-lung, suffer from a serious fatal disease compared to other types of CF (Davis, 2006). common symptoms of CF include lung infections mainly by *Pseudomonas aeruginosa* which live in the thick mucus and can cause damage to lung tissue. In addition, two-thirds of the patients suffer from pancreatic ducts covered by mucus which prevents necessary digestive enzymes reaching the gastrointestinal tract. Mucus also clogs small bile ducts and that can cause liver damage and slow down digestion. The reproductive system of male patients can also be affected by CF leading to blocking of fine ducts and the vas deferens, accounting for the 95% infertility rates. The blocking of sperm entry to the uterus can lead to infertility in women (Gout, 2012). Cystic fibrosis (CF) was first distinguished from celiac disease in 1938 by studies on malnourished infants that reported a disease of mucus plugging of the glandular ducts (Andersen, 1938). Defective sweat glands were linked to CF in 1953 and the sweat test in 1959 opened the possibility that CF is not just a disorder of mucus (Davis, 2006). In 1989, The CFTR gene was isolated from epithelial cell libraries and then was identified and sequenced (Riordan et al, 1989).

The cystic fibrosis transmembrane conductance regulator (CFTR) (Figure 2) is the only member of the ABC family that is recognized for its physiological function as an ion channel. CFTR is located mostly in the apical membrane of epithelia to function as a gate that controls chloride ion movement out and into the cell. CFTR is expressed in different locations, at the apical membrane of polarized cells (Dalemans et al., 1992), at random locations in the plasma membrane of non-polarized cells (Cheng et al., 1990). CFTR can also be located at endoplasmic reticulum (ER) where it is synthesized (Pasyk & Foskett, 1995), at Golgi where it matures, at endosomes where it interchanges with plasma membrane (Lukacs et al., 1992) and at lysosomes where it is degraded (Barasch et al., 1991). CFTR plays an important role in movement of chloride ions by using the energy of ATP in order to open and close its pore (Holland *et al*., 2003).). Both phosphorylation of the regulatory domain (R-domain) and ATP-binding control the opening and closing of the CFTR channel. Nucleotide binding domain 1 (NBD1) and NBD2 bind ATP to open the channel. PKA (Protein kinase A) and phosphatases add and

remove phosphate to the R-domain making potential conformational changes in the domain which leads to regulation of the channel activity.

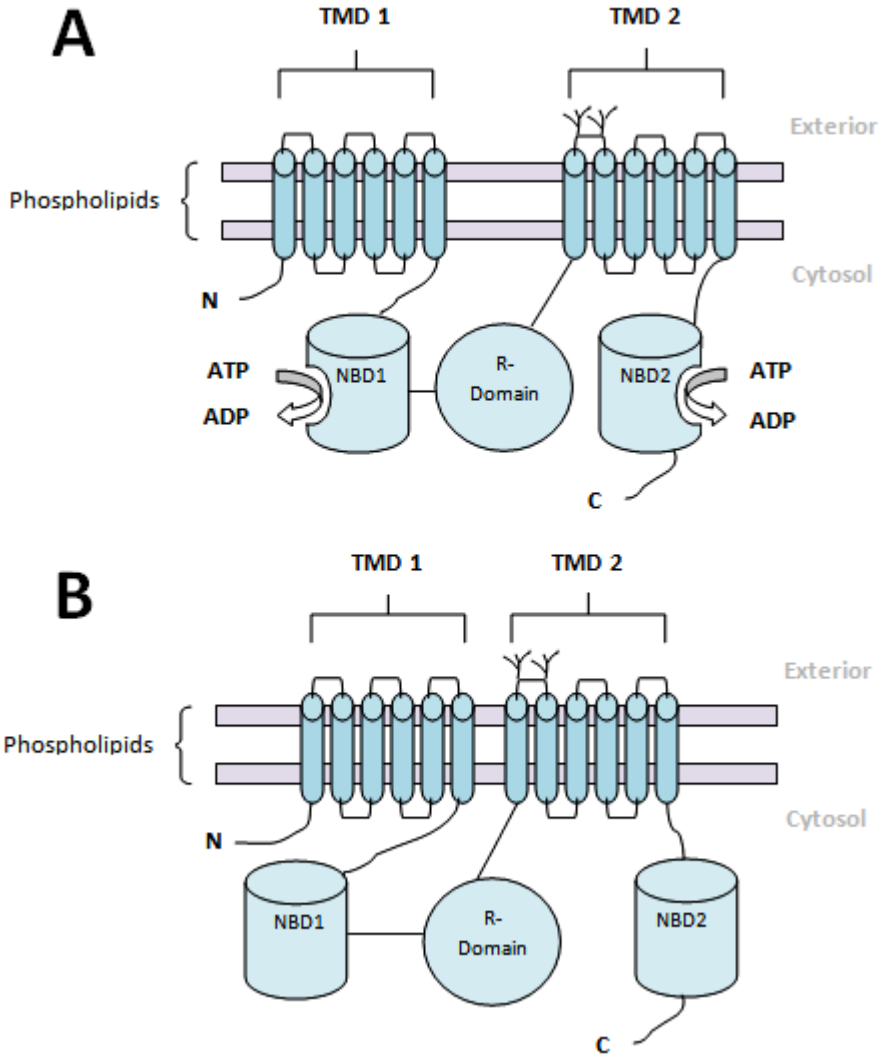


Figure 2: The proposed structure of CFTR and the ATP sites on NBD1 and 2 domains. This redrawn model is based on a research by Higgins and Linton (2004). A and B show the proposed ATP cycle which involves in CFTR gating. A, the closed state of the channel where is no ATP binding. B, the binding to ATP makes conformational changes lead to an open channel. TMD is transmembrane domain. NBD is nucleotide-binding domain. R-Domain is regulatory domain

The wide distribution of epithelia cells in the organs makes the defective CFTR a cause of several physiological problems. In humans, the gene encodes an integral membrane glycoprotein (CFTR)

consisting of 1480 amino acids. Generally, human CFTR mutations number more than 1300 and can be categorized into four classes (Welsh and Ramsey, 1998). These are Class I, where the CFTR loses its full function because it becomes shorter than usual. Class II, the most common one where mutations lead to in-frame amino acids change or deletion of the CFTR. 70 % of CF patients suffer from the F508 deletion which is one of the Class II mutations. Class III, causes changes in the nucleotide binding fold or regulatory (R) domain of CFTR resulting in errors in the control of channel opening. Class IV mutations affect membrane-spanning domains of the protein preventing the correct conduction of ions across the pores. In general, class I and II are considered a cause of serious medical problems and lead to an irregular pancreatic function (Goodman and Percy, 2005).

Cystic fibrosis is a disease that has no known cure and the treatments (physiotherapy, exercise, nutrition or even medication) are to reduce symptoms and slow the development of the disease. But more recently, in January 2012, FDA (The U.S. Food and Drug Administration) has approved Kalydeco (ivacaftor previously known as VX-770) for the treatment of a rare form of cystic fibrosis, involving patients with the G551D mutation in at least one copy of the gene. CFTR in people with the G551D mutation has the ability to locate to the right place in the plasma membrane but it acts like a blocked channel. Kalydeco can unlock the defective protein and restore its secretory function (Van Goor et al, 2009). WebSite: http://www.cff.org/treatments/Therapies/ provides most recently details about CF treatment.

1.3.1 CFTR-Interacting Proteins

A large number of studies reported interactions between CFTR and many proteins, including transporters, ion channels, receptors, kinases, phosphatases, signaling molecules, and cytoskeletal elements. The main function of CFTR is to transport ions through the plasma membrane. Beside this function, CFTR regulates other channels and transporters (Schwiebert *et al*., 1999b). CFTR can perform many physiological roles such as, mediating cAMP regulation of amiloride-sensitive Na^+ channel (Stutts *et al*., 1995), the Cl^-/$HCO3^-$ exchanger (Lee *et al*., 1999), and the ROMK K^+ channel (renal outer medullary potassium) (Schwiebert *et al*., 1999a). CFTR interactions with transporters and channels affect all epithelial tissues (Kunzelmann, 2001). Table 2 shows some these effects.

Transport	Interaction	NBD1 Required?
ENaC	Inhibition of ENaC	Yes
Na^+/H^+ exchanger	Inhibition of Na+/H+ exchanger	Unknown
HCO_3^-/Cl^- exchanger	Activation of HCO_3^-/Cl^- exchanger	Unknown
ICOR Cl^- channels	Activation of ICOR, conferring glibenclamide sensitivity	Yes
Ca^{2+}- and volume-activated Cl^- channels	Activation/inhibition of Ca^{2+}- and volume-activated Cl^- channels	Yes
ROMK2, Kir6.1	Conferring glibenclamide sensitivity to ROMK2 and Kir6.1	Yes
K_vLQT1	Activation of K_vLQT1 in oocytes	Yes
AQP3	Activation of AQP3	Yes
Gap junction channels	Activation of gap junctions	Unknown
Mucus secretion	Exocytosis and mucus secretion	Unknown
ATP transport	ATP release by epithelial cells	Yes
Glutathione transport	Activation of glutathione efflux	Unknown

Table 2: Some effects of the interaction between CFTR with other transport proteins. The Table shows some effects of CFTR-ABC protein interactions on cell functions. NBD1 of CFTR is required in some of these interactions. NBD1, nucleotide binding domain; ENaC, epithelial Na^+ channel; ICOR, intermediate-conductance outwardly rectifying Cl^- channel; ROMK2, rat distal nephron H^+ channel; K_vLQT1, delayed activated voltage-dependent N^+ channel; AQP, aquaporin water channel. This table was obtained from (Kunzelmann, 2001).

There are several direct interactions between CFTR and other proteins as follows: Syntaxin 1A interacts with the cytoplasmic N-terminal domain of CFTR and inhibits channel activity (Naren *et al*., 1997). A family of PDZ-binding domain proteins, including the Na^+/H^+ exchanger regulatory factor (NHERF1) (Hall *et al*., 1998) and ezrin-binding protein 50 (Short *et al*., 1998) interact with the C-terminal region. CFTR is also closely associated with protein phosphatase 2C (Zhu *et al*., 1999).

1.3.1.1 CFTR-NHERF Interactions

PDZ (PSD-95/Dlg/ZO-1) proteins are the most abundant proteins that regulate and maintain eukaryotic cell function. There are more than 500 PDZ domain-containing proteins which include at least 900 PDZ domains which are involved in this process (Sheng and Sala, 2001). Among all proteins that interact with CFTR, PDZ-domain-containing proteins (NHERF family) are the most common in mammalian cells (Wang et al., 1998). The NHERF family can be found in colon, renal proximal tubules, small intestine (In or near brush border) and other mammalian tissues. It is not present in bacteria, plants and yeast. The main role of the NHERF family is the formation of a multiprotein signaling complex to regulate the activities of other proteins (Donowitz et al. 2005). This family of proteins consists of four related PDZ domain-containing proteins and the total number of PDZ domains in the four proteins is 12 domains (PDZ 1 and 2 domains in NHERF 1 and NHERF 2, PDZ 1,2,3 and 4 domains in NHERF 3 and NHERF 4) (Donowitz et al. 2005) (As shown in Figure 3). Each domain has a length of 80-100 amino acids, (Guggino and Stanton, 2006) and regulates protein-protein interactions by binding to short peptides, most often in the carboxyl terminal of target proteins (Karthikeyan et al., 2001). To date, six PDZ proteins (NHERF1, NHERF2, NHERF3, NHERF4, Shank2 and CAL) have been reported to interact with the C-terminus of CFTR. The first five proteins interact with CFTR in the apical membrane of epithelial cells, while CAL (CFTR-associated ligand) interacts with CFTR in Golgi (Singh et al, 2009; Li and Naren, 2011). A single PDZ domain can target different proteins with different affinities. Moreover, a single PDZ motif can be recognized by different PDZ domains. There are four types of PDZ motifs: Type I (S/T-x-Φ), Type II (Φ-x-Φ), Type III (Ψ-x-Φ) and Type IV (D-x-V) where x: any amino-acid; Φ: hydrophobic amino-acid; Ψ:

hydrophilic amino-acid. The last four amino acids of CFTR (DTRL) are classified under Type I (Bossard et al, 2012).

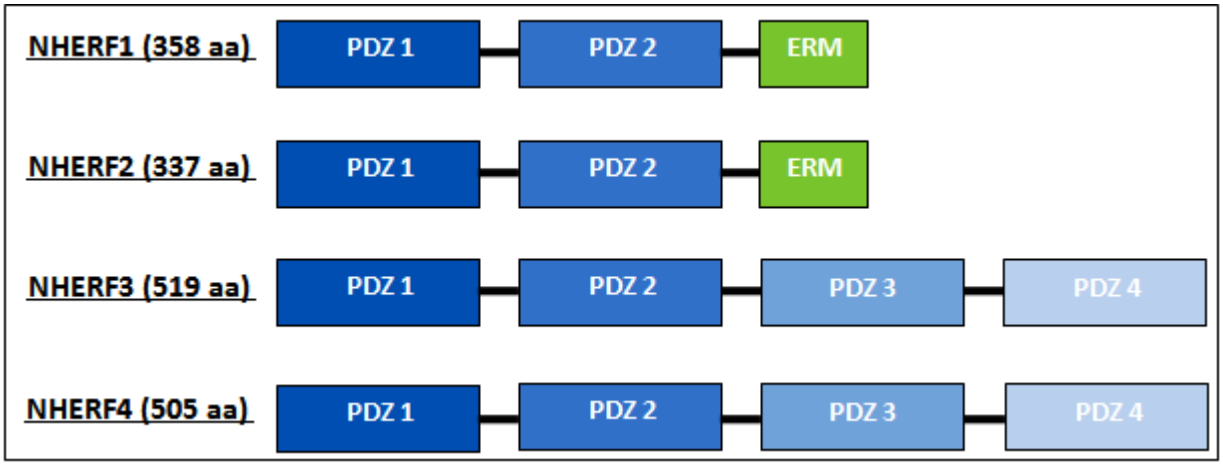


Figure 3: The four members of NHERF family and their domains. PDZ1 and 2 of NHERF 1 and 2 interact with the C-terminus (QDTRL) of CFTR (Brône and Eggermont, 2005; Li and Naren, 2010). Both NHERF3 and 4 contain four tandem PDZ domains, and PDZ3 and PDZ4 of PDZK1 are reported to bind to the C-terminus of CFTR (Wang et al, 2000; Li and Naren, 2010). NHERF1 (NHERF, EBP50). NHERF2 (E3KARP, SIP-1, TKA-1). NHERF3 (PDZK1, CLAMP, CAP70, DIPHOR-1, NaPi-CaP1). NHERF4 (PDZK2, IKEPP, DIPHOR-2, NaPi-CaP2). ERM, ezrin, radixin and moesin. The redrawn Figure was obtained from (Donowitz et al. 2005). NHERF1 PDZ 1 was involved in this study.

In this study, the C- terminus of CFTR-NHERF1 PDZ 1 complex was investigated. The last four amino acids in the C-terminus of human CFTR have the ability to interact with PDZ-domains in the NHERF family (Guggino and Stanton, 2006). A CFTR-NHERF 1 interaction (Figure 4) was reported in many studies (Lambright et al, 1996; Hall et al, 1998; Raghuram et al, 2001). Wang et al (1998) studied the interaction between NHERF 1 PDZ 1 and 2 with three different peptides (1211 to 1481, 1387 to 1481 and 1466 to 1481) of the C-terminus of human CFTR. They found that the three peptides were able to interact with NHERF 1 PDZ 1 but not to PDZ 2.

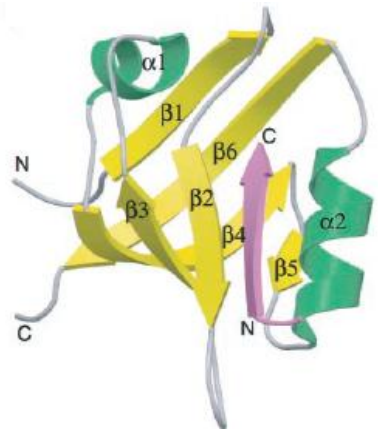


Figure 4: Crystal structure of NHERF1 PDZ1-QDTRL. Ribbon diagram of the human NHERF1 PDZ 1 domains (yellow and green) bound to the QDTRL (pink) amino acids of CFTR. NHERF1 PDZ 1 contains 6 beta strands (in yellow) and two alpha helices (in green). The peptide ligand QDTRL is shown in pink. This structure was solved by (Karthikeyan et al, 2001). QDTRL which is fused to one molecule of NHERF1 PDZ 1 interacts with NHERF1 PDZ 1 from the other molecule. (PDB code is 1G9O).

The interaction between NHERF1 PDZ 1 and many C-terminal sequences allow slight conformational changes in NHERF1 structure and this flexibility can explain why NHERF1 (PDZ 1 or 2) binds many protein targets (Shenolikar et al, 2004). The solved structure (Figure 5) of QDTRL (C-CFTR) and NHERF1 PDZ 1 was reviewed by (Ladias, 2003). As shown in the figure, the NHERF1 PDZ 1 structure consists of six beta-strands (in brown) and two alpha-helices (in yellow) and this structure is identical to other PDZ structures. Two β-sheets form an antiparallel β–sandwich (a pocket) where QDTRL (revealed as a ball-and-stick model) are inserted in. The core of PDZ 1 is stabilized by hydrophobic interactions that were formed between amino acids 17, 26, 28, 39, 53, 59, 76, 79, 86 and 88. Table 3 gives more details about the chemical bonds between QDTRL-NHERF1 PDZ 1.

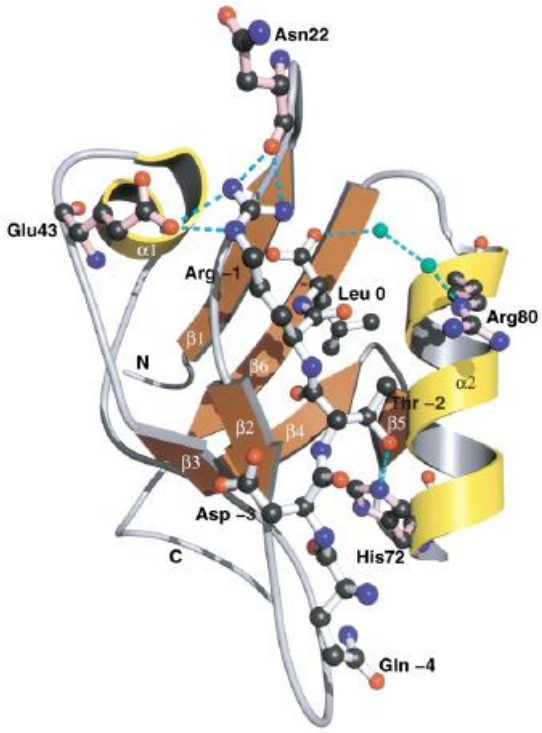


Figure 5: Crystal structure of NHERF1 PDZ1-QDTRL. Ribbon diagram of the human NHERF1 and ball and stick representation of the five amino acids of CFTR along with specific interacting residues of NHERF1 (Asn22, Glu43,His72, Arg80). Gln -4, Asp -3, Thr -2, Arg -1 and Leu 0 are C-CFTR residues and the rest of amino acids belong to the PDZ 1 domain of NHERF1. Water molecules are shown as green spheres and hydrogen bonds as cyan dashed lines. The diagram was obtained from (Ladias, 2003).

C-CFTR residue	Bond	Description
Q (Gln -4)	A hydrogen bond	The carbonyl oxygen interacts with the amide nitrogen of Gly30
D (Asp -3)	A hydrogen bond	The side chain interacts with His27
	A salt bridge	The side chain interacts with Arg40
T (Thr -2)	A hydrogen bond	The amide nitrogen carbonyl oxygen interacts with the amide nitrogen carbonyl oxygen of Leu28
	A hydrogen bond	The side chain interacts with the imidazole group of His72
R (Arg -1)	Two salt bridges	The guanido group interacts with the side chain of Glu43
	Two hydrogen bonds	The guanido group interacts with the carbonyl oxygen of Asn22
L (Leu 0)	Hydrophobic contact	The side chain and carboxylate group interact with Tyr24, Gly25, Phe26, Leu28, Val76 and Ile79
	Hydrophobic contact	The isobutyl group interacts with Phe26 and Ile79
	A hydrogen bond	The carboxyl oxygen interacts with the amide nitrogen of Gly25, Phe26, tyr24 and Arg80

Table 3: Specific interactions between QDTRL of C-CFTR and the NHERF1 PDZ 1. The Table shows the weak chemical bonds between the last five amino acids of CFTR and NHERF1 PDZ 1. Leu and Gln form the strongest and weakest interactions respectively.

The CFTR C-terminus-NHERF1 PDZ 1 complex was reported as a linker that links CFTR to the apical actin cytoskeleton in polarized epithelial cell through ezrin, radixin, or moesin proteins. This complex is required for stabilizing CFTR to perform its function (Short et al, 1998; Sun et al, 2000). A part of this complex (CFTR- NHERF1 PDZ $1^{(+)}$) was investigated in this study and a new possible conformational change of CFTR was obtained (as revealed in section 3.2.2). In addition, it was suggested that CAL protein can reduce CFTR expression at the cell surface. This suppression can be disrupted by CFTR-NHERF1 binding by which NHERF1 enhances CFTR expression in the plasma membrane. Thus, CFTR is able to interact with at least two PDZ proteins (Cheng et al, 2002) and this hypothesis is compatible with SPR data (Appendix A) in which two monomers of NHERF1 PDZ 1 were observed to interact with CFTR. Moreover, the dimerization of NHERF1 PDZ $1^{(+)}$ was confirmed using multi-angle light scattering (MALS) as shown in section (Appendix B).

1.3.1.1.1 Does NHERF-C-terminus of CFTR binding have an important role in CF?

An answer for this question was addressed by Mickle et al (1998) who found that a mother and her daughter, both heterozygous for CFTR-S1455X and deletion of exon 14a (del14a) (the deletion of the last 26 amino acids of CFTR including the PDZ motive) had no effect on CFTR expression and activity in lung and pancreas. They also reported that this mutation generated a mild disease as detected by high sweat chloride concentrations in the absence of the CF lung disease. The same study was reported that a second daughter homozygous for del14a mutant had severe CF. In vitro, CFTR constructs with the deletion of the PDZ motif (ΔTRL-CFTR) were studied. Moyer and colleagues (1999; 2000) found that the PDZ binding motif is an apical polarization signal and its deletion decreases CFTR function in MDCK (Madin-Darby Canine Kidney) epithelial cells. In addition, (Li and Naren, 2011) have shown a physiological importance of NHERF-CFTR interactions suggesting that NHERF-CFTR binding is not only important in ion transport, but it has multiple roles in regulation of other protein`s activities in epithelial cells. On the other hand, other studies suggested that the deletion of the PDZ binding motif has no influence on its apical membrane localization nor its activity in several cells such as BHK-21 (Baby Hamster Kidney epithelial cells), COS-1 (African green monkey kidney epithelial cells), MDCK, CaCo-2 (adenoCarcinoma of the Colon epithelial cells), PANC-1 (Human pancreatic carcinoma epithelial cells) and primary human airway epithelial cells (Benharouga et al., 2003; Milewski et al., 2005; Ostedgaard et al., 2003).

In general NHERF family proteins regulate CFTR by three means. First, they link CFTR to cytosolic regulatory proteins. Secondly, NHERF proteins bring other transmembrane proteins closer to CFTR for reciprocal regulation. Thirdly, NHERF1, NHERF2 and NHERF3 have the ability to enhance the formation of CFTR dimerization. This dimerization may lead to an increase in CFTR activity (Wang et al, 2000; Li et al, 2004; J. Li et al, 2005). All these mechanisms used by NHERF proteins to regulate CFTR, make the family of NHERF a potential therapeutic target for CF. In 2005, (Guerra et al., 2005) found that mouse NHERF1 (but not NHERF2) overexpression increased F508del-CFTR plasma membrane expression and activity in human bronchial epithelial cell. Four years later, Bossard

et al (2007) reported that F508del-CFTR apical plasma membrane expression and chloride channel activity were restored by human NHERF1 overexpression. These and other results make the study of NHERF family proteins an important goal in terms of CF treatment.

1.4 Structural Biology of Transmembrane Proteins

Most transmembrane protein structures (and as mentioned in Table 4 for CFTR) have been determined by X-ray crystallography, NMR and electron crystallography. X-ray crystallography requires a sufficient concentration of the protein (> 10 mg/ml) for crystallization. This requires significant amounts of protein to be purified, and as explained before, CFTR has a low level of expression because of degradation and misfolding of the protein. Beside this limitation, the presence of hydrophobic regions in CFTR is another reason that no 3D crystal of CFTR has been obtained so far. NMR techniques also require a high concentration of protein and are usually limited to proteins of mass up to about 100 kDa (Wider and Wuthrich, 1999). In addition, detergents that cover hydrophobic amino acids in transmembrane proteins can alter the conformation of the protein (Frishman and Mewes, 1997). Electron crystallography is considered as a method to determine protein structure and with the advantage that transmembrane proteins can be reconstituted into lipid. However, there is a lack of large and good quality 2D crystals (several square-micrometers in size and one unit cell thick) (Walz and Grigorieff, 1998). In comparison with the three previously mentioned protein structure determination techniques (X-ray crystallography, NMR and electron crystallography), SPA (single particle analysis) seems to be a faster and less expensive approach because it is not necessary to generate 2D crystals and only small amounts of protein are needed.

1.4.1 Electron Microscpy

3D structures of the cell components that are created using electron microscopy can be divided into two categories. First, cell components in the the size range of 100-1000 nm. These components possess a unique structure that differs from one to another. An example of this, is the mitochondria. In contrast, macromolecules ranging from 50-500 nm may have identical structures such as folded

proteins which can be treated with imaging software to classify them into similar orientations in order to create a 3D model. Electron crystallography was established by (Cowley et al, 1957). Later, Hoppe et al (1968) used crystallographic methods to obtain structural information from electron micrographs. In 1975 the first membrane protein structure was determined at intermediate resolution (7 Å) by (Unwin and Henderson). But the first high resolution structure by electron crystallography methods was achieved by Henderson et al (1990). They used electron cryo-microscopy to record high resolution micrographs of bacteriorhodopsin (2D crystals). As any other techinque, there are some advantages and disadvantages in the use of transmission electron microscopy (TEM) in order to image transmembrane proteins (as illustrated in Table 4).

Advantages	Disadvantages
No need for protein crystals	Non-crystalline samples subject to high vacuum
Small amount of protein (≈ 10 ng / µl) is required	Very thin samples are required to allow electrons pass through
Phases are already recoreded on micrographs (no phase problem compared to X-ray crystallography)	Image distortions that are caused by microscope are usually corrected (corrections of amplitude and phase)
Structural data can be obtained rapidly	Significant expertise/training required
Proteins can be studied in solution at the instant they were frozen	Resolution is usually limited, preventing tracing of the polypeptide chain
Proteins can be stained and labelled prior to imaging	Weak scattering of electrons by the the protein atoms (proteins contain light atoms e.g C, O and H)
No upper size limitation for protein complexes	Lower size limit of ≈ 100 kD for SPA
Electrons interact strongly with matter (10000 greater than X-ray) – individual protein can be imaged	Electron damage to the sample must be minimised by colling the sample or by minimizing the electron dose

Table 4: Advantages and disadvantages of electron microscopy for structural biology of transmembrane proteins. SPA (single particle analysis). This table was adapted from (Ford et al, 2011).

In general, there are two common methods to generate a 3D model of protein by electron microscopy. One method requires the formation of 2D crystals with a good quality for image processing. The other is to use non-crystalline specimens and SPA (single particle analysis) by which thousands of projections are combined by image processing to produce a complete 3D reconstruction at a low, medium (or in few cases) high resolution. However, most structures that were obtained by electron crystallography were at a resolution around 10 Å (Koeck et al, 2007).

1.4.1.1 2D Crystals of CFTR

The formation of 2D protein crystals is still unpredictable (Knol et al, 1998; Ford et al, 1999; Lu et al, 2012). This is because of the lack of understanding of the background behind the formation of 2D crystals. Three requirements have been identified that can lead to successful 2D crystal formation. Firstly homogeneous protein. Secondly a high local concentration of protein, and thirdly avoiding any condition that can denature the protein (Lu et al, 2012). The production of 2D crystals is achieved by many biophysical methods. The most common method is reconstitution by which the protein is reconstituted into lipid bilayers at a high local concentration (high protein:lipid ratio) to form 2D crystals (Figure 6). A successful reconstitution can be optimised by adjusting several parameters including the pH, salt concentration, temperature, types/compositions of lipid and detergents, lipid to protein ratio (LPR) and rate of detergent removal (Mosser, 2001). Crystallization is induced by using a lipid : protein ratio that is low enough to promote protein/protein interactions. An important aspect of this approach is that the solubilized protein must be stable for a period that is sufficient to remove detergent by dialysis. A successful 2D crystallization requires trial-and-error attempts to find the appropriate lipid:protein ratio. Too high a concentration of lipid will form a high number of lipid bilayers that can not be filled by a sufficient concentration of protein and no 2D crystals will be induced. On the other hand, too high a concentration of protein causes aggregation of protein on removal of the detergent (Yeager et al, 1999).

Up to the present date, no 3D CFTR crystals have been reported. However, 3D structures of CFTR were generated using 2D crystals. Rosenberg *et al*. (2004) studied purification and crystallization of

the protein. This study reported the first crystallization of human CFTR (Figure 7). By using electron crystallography for negatively stained 2D crystals of CFTR, the structure of this protein was visualised and then analysed for two different conformational states.

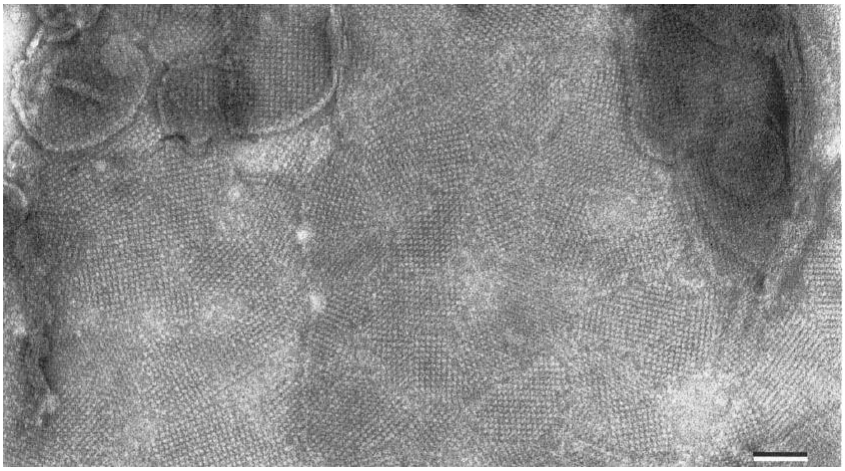


Figure 6: 2D crystals of photosystem I complexes. The complexes were reconstituted in the presence of DMPC lipid (dimyristoylphosphatidylcholine) and $MgCI_2$. The scale bar is 100 nm. This figure was obtained from (Ford et al, 1990).

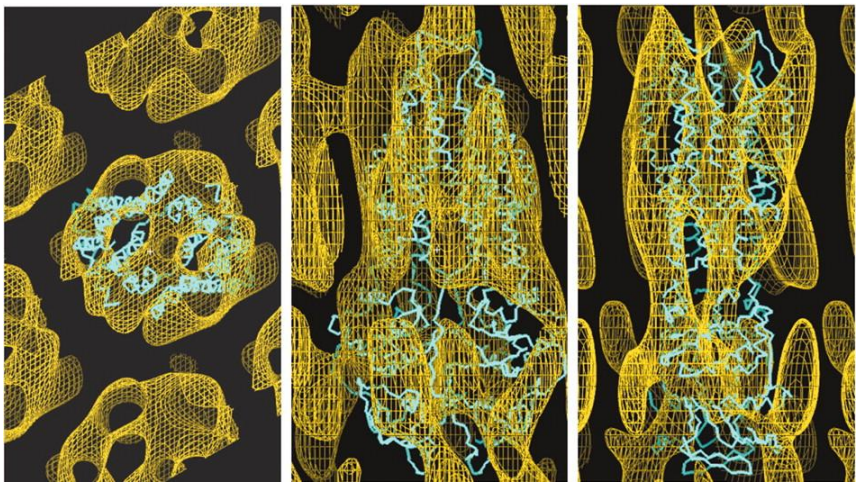


Figure 7: An atomic model of MsbA protein (blue) is fitted into a three-dimensional map of human CFTR (yellow). This comparison between the two proteins showed similar structural features. Both show a good match in overall size and shape with small differences that CFTR has additional R-domain and posttranslational modification (glycosylation of extracellular loop 4). However the MsbA model was later corrected (Ward et al., 2007). This figure was obtained from (Rosenberg et al, 2004). Left panel, a top view perpendicular to the crystal plane with a slab through the high density region. Center panel, an orthogonal view along the crystal plane. The density on the left and right edges of the panel are because of the neighboring unit cells. The right panel shows a second side view orthogonal to the first.

1.4.1.2 Single Particle Analysis of CFTR

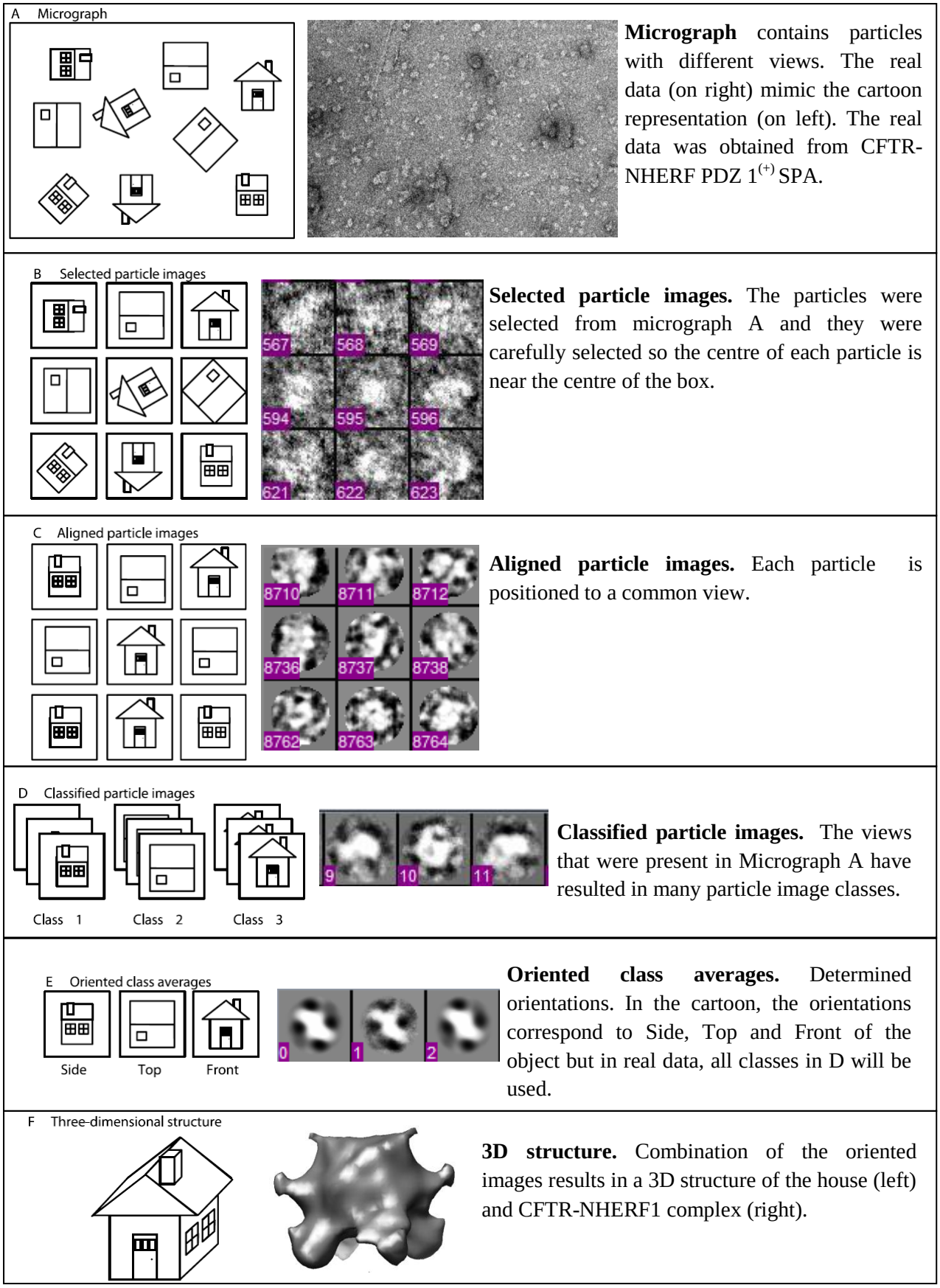
In principle, SPA is designed to perform averaging methods in order to overcome the fact that micrographs contain high-resolution information contaminated by a much higher intensity of noise. With large soluble proteins, the single particle method is able to obtain information at a medium resolution (10−20 Å) and to high resolution for some highly symmetric particles (Frank, 2002). With transmembrane protein particles, the resolution is usually lower because of a lower signal-to-noise ratio that is produced by a nonstructured detergent layers in hydrophobic regions. Single particle samples can be imaged in a native-condition (protein embedded in vitreous ice) or with negative staining. Negative staining can provide a protection for biological samples from the electron beam and improve electron scattering. Also, the negative stain is adsorbed well by biological matter and this forms dark (electron-scattering) areas around the sample which enhance the image contrast (Bozzol and Russell, 1999). Negative staining is fast and easy and a low resolution (15-20 Å) structure for preliminary investigations can be gained, but the resolution is limited by the size of the stain granularity. Furthermore, despite the fact that the stain can cause artefacts for the sample due to the staining procedure, negative staining is still useful for the analysis of small proteins (50-200 kD) which could not be visualized by cryomicroscopy due to the lack of contrast (Fernández and Valpuesta, 2009). For the previous reasons, negative staining is a suitable techinque to investigate samples in the early stages followed by cryomicroscopy for a higher resolution structure.

There are several software packages designed specifically for single-particle reconstructions, PHOELIX (Whittaker et al, 1995), SPIDER (Frank et al, 1996), SUPRIM (Schroeter and Bretaudiere, 1996), IMAGIC (van Heel et al, 1996) and EMAN (Ludtke et al, 1999). EMAN (EMAN 1 v1.9 was the software used in this study) is a software package that has a variety of tools for a complete data processing with two features. First, EMAN is an accessible software for users with a little experience and for those users who are advanced. Second, CTF (contrast transfer function) is nearly fully automated. Generally, The creation of a final 3D model by EMAN requires three main steps. The selection of particles that are collected by TEM (Transmission electron microscopy) on micrographs.

And then the selected particles are used to create an initial 3D model. This preliminary model functions as a template for several successive iterations of data to generate the final 3D structure (3D map).

In single particle 3D reconstruction, a critical step is to precisely allocate the correct orientation of each single particle. Several methods are used to allocate the view for each particle, including the common lines approach used here (van Heel, 1987; Penczek et al., 1996) and random conical tilt (Radermacher, 1988). In the method used in this study, the relative angles (Euler angles) between the particle views are assigned using mathematical relationships for different projections of the same object. The relative angles in the random conical tilt approach can be determinted by setting up a defined data collection geometry. In contrast, angular reconstruction has no restriction on orientation determination. The orientations determined using this method are based on the central projection theorem (DeRosier and Klug, 1968). This theorem states that two particle iamges of the same object share common lines. If the common lines in two particles can be located then the relative angles between those two particles can be calaculated. Figure 8 shows the main steps performed by SPA.

Several 3D structures of ABC proteins have been generated by single particle methods. A P-glycoprotein (P-gp) structure was determined at 25 Å and was reported as a monomer form (Rosenberg et al, 1997). This structure of P-gp was the first structural data for any ABC transporter. A viral protein 6 (VP6) was the first protein solved at close to atomic resolution by SPA (Zhang et al. 2008). The obtained structure at 5 Å was fairly similar to the X-ray crystallography structure at 3.8 Å for the same protein (Figure 9) (McClain et al, 2010). In theory, the minimum number of particles that are necessary to achieve a 3D structure at a 3-Å resolution ranges between 1,400 and 12,600 (Henderson, 1995; Glaeser, 1999). To date, no full-length structure of CFTR was obtained using X-ray crystallography or NMR. The only structures that are available for full-length CFTR were generated by 2D Electron crystallography and SPA.



Micrograph contains particles with different views. The real data (on right) mimic the cartoon representation (on left). The real data was obtained from CFTR-NHERF PDZ $1^{(+)}$ SPA.

Selected particle images. The particles were selected from micrograph A and they were carefully selected so the centre of each particle is near the centre of the box.

Aligned particle images. Each particle is positioned to a common view.

Classified particle images. The views that were present in Micrograph A have resulted in many particle image classes.

Oriented class averages. Determined orientations. In the cartoon, the orientations correspond to Side, Top and Front of the object but in real data, all classes in D will be used.

3D structure. Combination of the oriented images results in a 3D structure of the house (left) and CFTR-NHERF1 complex (right).

Figure 8: The main steps (A-F) of SPA. The diagram was adapted from (Thuman-Commike, 2001).

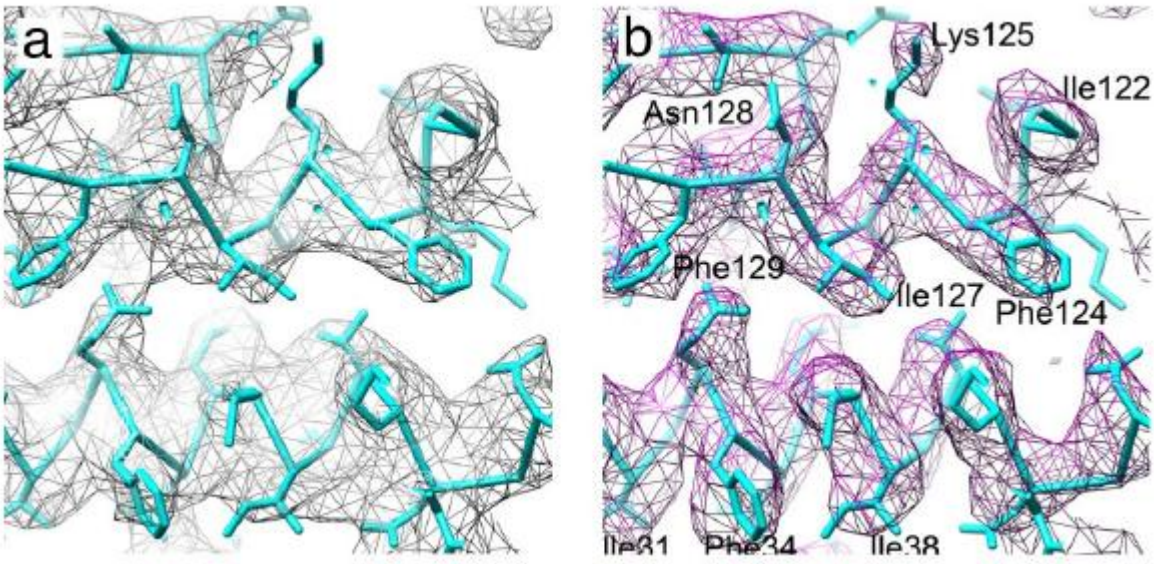


Figure 9: A comparison between the same region of two density maps. A, the map was generated by X-ray crystallography (PDB code is 3KZ4). B, the map was generated by SPA. Both maps were fitted by an atomic model. This figure was obtained from (Zhang et al. 2008).

1.4.2 The Structure of CFTR Domains

All identified ABC transporters consist of two nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs) so that they can perform their physiological functions (Higgins, 1992). In bacterial ABC transporters, the four domains are mostly encoded by different genes whilst NBDs and TMDs of CFTR are expressed by a single gene (Gadsby and Nairn, 1999). Basically, the structure of CFTR is different from the other ABC proteins in the presence of the R-domain. This domain has several sites in which phosphorylation by cAMP-dependent protein kinase (PKA) and protein kinase C occur. Both terminal sides of CFTR, the NH2- and COOH-terminal halves, include a TMD comprising 6 supposed membrane-spanning α helices linked to the NBDs which are located in the cytoplasm and connected to a regulatory (R) domain (Gadsby and Nairn, 1999). A large portion of CFTR is found in the cytoplasm (77 %) while smaller portions of the protein are found in the plasma membrane (19 %) and 4 % in extracellular loops, which (with exception for the M1-M2 and M7-M8 loops) are extremely small (Akabas, 2000).

The diameter of the CFTR channel has been determined by many studies. Linsdell *et al*. (1997) reported that the minimum channel diameter is ~5.3 Å estimated by comparison to the size of the largest permeant anion. Another study by Linsdell and Hanrahan (1998) has suggested that the diameter of the channel can be as large as 10-13 Å based on a slight transport of large anions (sized, 10-13 Å in diameter) across the channel pore. There are six positively charged amino acids within the TMDs, K95 (M1), R134 (M2), R334, K335, R347 (M6) and R1030 (M10) (Riordan et al, 1989) which have been proposed to line the pore. This basic region of the CFTR pore suggests that there is an interaction between the six positively charged amino acids and anions and this interaction facilitates anions movement through the channel.

The Walker A, Walker B and LSGGQ (or signature) motifs are located in each NBD (Figure 10). They are highly conserved sequences that can be found in many nucleotide-binding proteins. Many functional studies suggested that the three motifs play a key role of binding to ATP and consequently controling the opening and closing of the ion channel (Sheppard and Welsh, 1999).

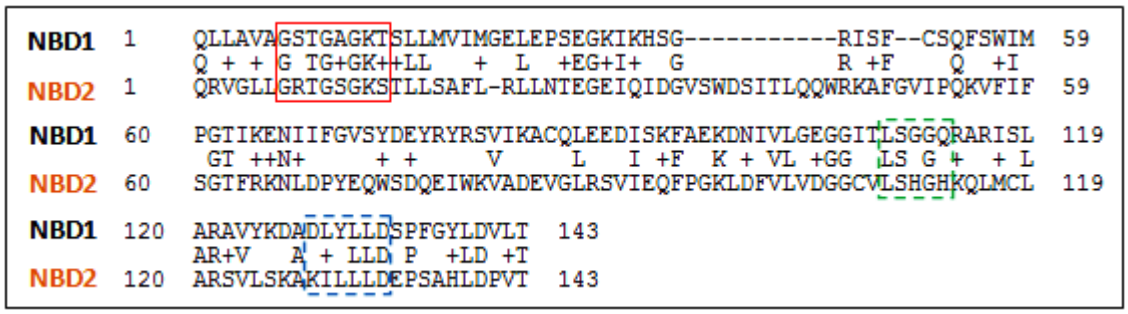


Figure 10: The three conserved motifs in NBDs of CFTR. The Walker A (red lined box), Walker B (blue dashed box) and LSGGQ motif (green dashed box). The alignment of the NBD1 (452 – 594 residues) and NBD2 (1238 – 1380 residues). BLASTP 2.2.26+ (online tool) was used for alignment (Altschul et al, 1997).

NBD1 and NBD2 form a a head-to-tail dimer in order to control the CFTR gating (Lewis et al, 2004). This dimerization is required to compose the ATP binding pockets in which the Walker A and B motifs from one NBD interact with the signature motif from the partner NBD as shown in Figures 11 and 12.

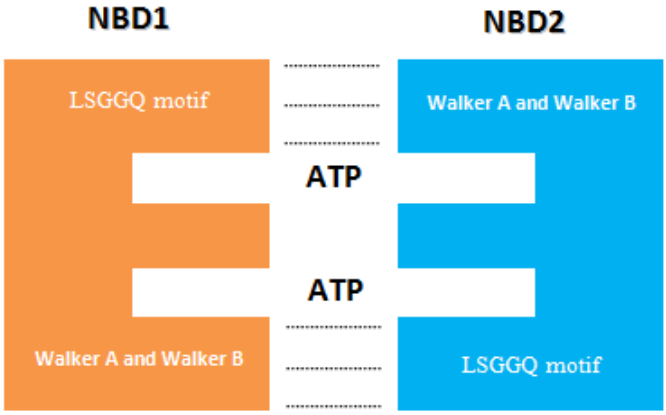


Figure 11: NBD dimerization. Two ATP binding pockets are formed. ATP binding pocket 1 is formed by an interaction of Walker A and B motifs of NBD1 and signature sequence of NBD2. ATP binding pocket 2 is formed by an interaction of Walker A and B motifs of NBD2 and signature sequence of NBD1.

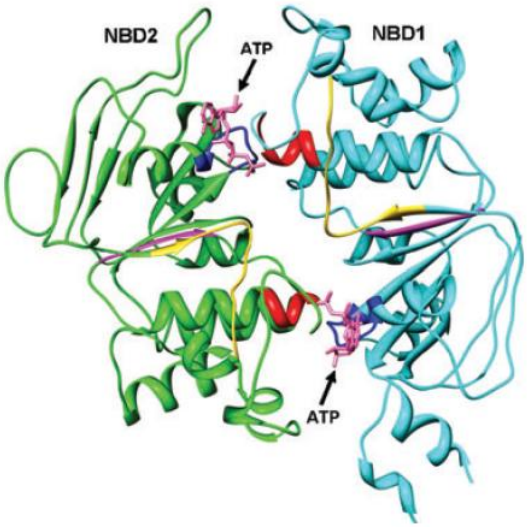


Figure 12: A model of a heterodimer of NBD1 and 2 of human CFTR. NBD1 is colored in cyan and NBD2 is in green. The Walker A (blue), Walker B (magenta) and LSGGQ (red) motifs are shown as well as the Q loop (yellow). ATP molecules (pink) are represented in stick mode. The model was obtained from (Hwang and Sheppard, 2009). The two domains are linked by an interaction between two amino acid, R555 (NBD1)/ T1246 (NBD2). A hydrogen bond is formed between Arg 555 and Thr 1246 when the CFTR gate is open or in the transition state (Vergani et al, 2005).

A list of the available structures of CFTR domains is shown in Table 5. To date, human and mouse CFTR cytoplasmic domains have been crystallized and their structures determined. All the protein sequences of solved NBD1, NBD2 and R-domain structures were mutated for successful expression and purification. The full length solved structure (2D crystal of human CFTR) was achieved by a combination of 2D protein crystallization (with low resolution) compared to 3D high resolution

structure (Sav1866). The NMR (nuclear magnetic resonance) data of R-domain has confirmed that most of the domain structure was disordered. Last update: 02-10-2012.

Protein	Experiment	Length	PubMed	PDB
F508 mutation region (human) (Massiah et al, 1999).	NMR	498-523	10360942	1CKX
Residues from the carboxyl-terminal peptide (human) (Karthikeyan et al, 2001).	X-RAY	1476-1480	11304524	1G9O
NBD1 of mouse CFTR was determined in unliganded, ADP- and ATP-bound states, with and without phosphorylation (Lewis et al, 2004).	X-RAY	389-673	14685259	1Q3H 1R0Z 1R0W 1R0X 1R0Y 1R10
Residues from the amino-terminal peptide (human) (Cormet-Boyaka et al, 2004)	NMR	30-63	15141088	not available
NBD1 of human CFTR with the F508 and F508A mutations (in ATP-bound state) (Lewis et al, 2005).	X-RAY	389–673 389–678	15528182	1XMJ 1XMI
NBD1 of mouse CFTR with F508R and F508S mutation (Thibodeau et al, 2005).	X-RAY	389–673	15619636	1XF9 1XFA
NBD1 of human CFTR with solubilizing mutations (Lewis et al, 2010).	X-RAY	389-678	19944699	2BBO 2BBS 2BBT
Minimal NBD1 of human CFTR with (del405-436) and (del405-436, delF508) mutations (Atwell et al, 2010).	X-RAY	387-646 375-646 387-646	20150177	2PZE 2PZG 2PZF
R-domain of human CFTR (Baker et al, 2007).	NMR	654-838	17660831	not available
Residues from the first extracellular domain of human CFTR (human) (Balakrishna et al, 2009).	X-RAY	103–117	19626704	3FHV
Residues from the amino-terminal peptide (human) (Smith et al, 2010).	X-RAY	5-22	20351101	3ISW
NBD2 of human CFTR with N6-Phenylethyl-ATP (P-ATP) (Atwell et al, to be published).	X-RAY	1193-1427	not available	3GD7
2D crystal of human CFTR (Rosenberg et al, 2011).	Electron crystallography	Full	21931164	4A82
NBD1 of mouse CFTR with F508 mutation. (Mendoza et al, 2012).	X-RAY	389-673	22265409	3SI7

Table 5: The existing CFTR structures in the protein data bank (PDB). No full-length of CFTR structure was obtained by X-ray crystallography or NMR to the present date. In contrast, SPA and 2D electron crystallography provided low-medium 3D structure of CFTR. These structures could fill gaps in our knowledge about CFTR.

1.4.2.1 Prediction of CFTR Topology (and the location of C-terminus)

As a result of the complexity of crystallizing large membrane proteins, only a few three-dimensional structures of ABC transporters are currently available. Several online protein tools are available to predict transmembrane topology. Tusnády et al. (2006) summarized the currently available information on the membrane topology of some members of the human ABC protein subfamilies. These topology predictions were created depending on biochemical experiments or the homology modeling of the existing crystal structure of ABC transporters. Generally, the topology of ABC transporters consists of four domains (mentioned before). The predicted structure for the membrane topology of CFTR has been confirmed experimentally by insertional mutagenesis (Chang et al., 1994). They found that CFTR contains two transmembrane-spanning domains and each domain consists of 6 membrane-spanning helix regions. The proposed structure of another ABC transporter (P-glycoprotein) has been suggested to be similar to CFTR topology by Gros et al. (1986) and Chen et al. (1986) and was confirmed by X-ray crystallography (Aller et al, 2009). Several predicted topologies of CFTR were published and all of them suggested that the C-terminus of CFTR is located out of the membrane and on the inside of the cell (as shown in Figure 13). As revealed in Figure 6, CFTR is predicted to have 12 transmembrane segments. The N- and C-terminii are located inside the cell. Both N- and C-terminal peptides are expected to act as soluble proteins because they are not embedded inside the cell membrane. TOPCONS, an online topology membrane protein prediction tool, was used to predict the location of C-terminus within the CFTR topology. The tool can be found on this website: http://topcons.cbr.su.se/index.php.

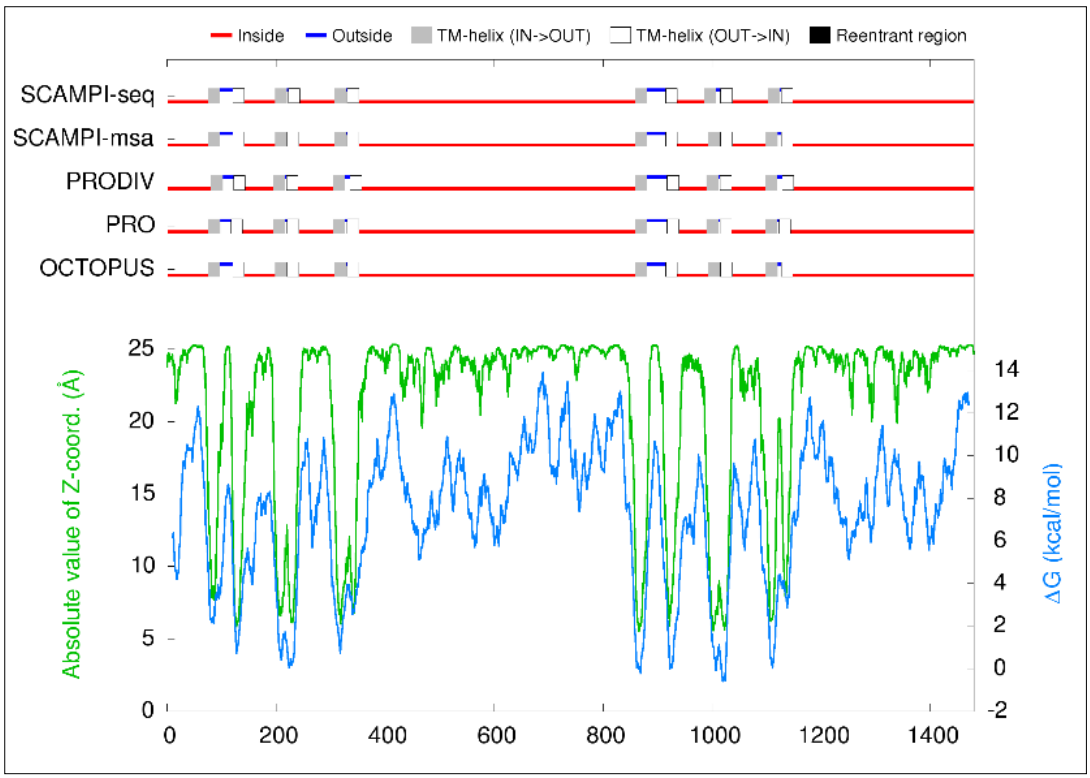


Figure 13: The predicted topology of CFTR by TOPCONS. SCAMPI-seq, SCAMPI-msa, PRODIV, PRO and OCTOPUS are individual topology prediction methods that were used to predict the locations of CFTR domains. Red lines indicate amino acids of CFTR which are located inside the cell (in the cytoplasm). Blue lines (in the top section of the figure) locate amino acids outside the cell, while gray lines locate amino acids buried in the membrane. Black lines locate turns between CFTR domains. The blue curve (in the bottom section of the figure) indicates the change of free energy for each amino acid. Amino acids with low free energy indicate that they are located into the plasma membrane and vice versa. The green curve (in the bottom section of the figure) indicates the distance (Å) between amino acids locations and the plasma membrane. Amino acids with low value indicate that they are located very close to the plasma membrane and vice versa. TOPCONS combines these methods into one consensus prediction (Bernsel et al, 2009).

1.4.2.2 The C-terminus of CFTR

Determination of the length of the C-terminus of CFTR (and by definition, the end of the NBD2 domain) is an important goal in order to investigate its structural and functional roles. As NBD2 of

CFTR is a domain that at its C-terminal end is linked to the N-terminal start of C-CFTR, the functional C-terminal boundary of NBD2 (as determined by structural studies) can be used as a guide to determine the C-CFTR length. Based on this idea, Gentzsch et al (2002) designed an experiment to investigate 8-azido-ATP binding dependence on CFTR sequence (1380-1440). Truncation of the protein at residues close to the C-terminus was carried out by site-directed mutagenesis. They found that 8-azido-ATP binding was detectable in the CFTR mutated samples truncated from 1425-1440. On the other hand, CFTR truncations from 1380 to 1420 were not labeled with azido-ATP indicating that they were not able to bind ATP and theses also lacked channel gating. This interesting result suggested that the start of the CFTR C-terminus and the C-terminal end of NBD2 is probably located between residues 1424 and 1420. The location of the C-terminus in the cytoplasm and within the overall architecture of the full-length protein was studied by single-particle electron microscopy using a C-terminal 6-Histidine tag (Nanogold labeled). The gold was located in a region predicted to be between the NBD1 and NBD2 domains of CFTR. This unique location between the two domains suggests a possible role for The C-terminus in CFTR gating activity (Zhang et al, 2011). A new NMR study was performed by (Bozoky et al, 2011a) (poster presentation NACFC 2011, Anaheim CA, USA) which suggested a strong interaction between the C-terminus^{42aa} and the phosphorylated R-domain. Previously, many other studies highlighted the importance of C-CFTR. The C-terminus of CFTR has two conserved motifs (^{1424}YDSI) (Prince et al, 1999) and (^{1430}LL) (Hu et al, 2001) which appear to function as endosomal targeting signals. The interaction between NHERF1 PDZ 1 and the last four amino acids (DTRL) of CFTR is required for the localization of the ion channel to the correct membrane of polarized epithelia. Also, amino acids (1370-1394 – which are probably in NBD2) and (1404-1425) are required for apical localization (Milewski et al, 2001). Several CF-related mutations in the C-terminus of CFTR were reported: Arg1422Trp, Tyr1424Stop, Ser1426Pro, Ser1426Phe, Ile427Thr , Asn1432Lys , Ser1435Gly, Arg1438Trp, Ala1440Ala, Asp1445Asn, Arg1453Trp, Ser1455X, Ser1456Asn, Lys1459Lys, Gln1463His, Lys1468Asn, Glu1473X, Gln1476X, Leu1480Pro. C-terminus of CFTR mutations implying that this region of the protein is crucial for

CFTR function within the cell and within the whole organism. The mutation data were obtained from: Cystic Fibrosis Mutation Database.: http://www.genet.sickkids.on.ca/cftr/SearchPage,$Form.direct.

1.4.2.3 Expression and Purification of CFTR

In order to obtain structural information for a protein it is necessary to over-express it in large quantities in a native or foreign host cell (Roth, 1994). The over-expressed target must then be purified (ideally to homogeneity) for subsequent structural analysis. The first successful trial to construct a complementary DNA (cDNA) encoding the CFTR gene was in 1990. Gregory et al used *E. coli* cells to produce a high number of cDNA encoding the full-length CFTR in order to express the CFTR protein in cell lines. CFTR has a low expression level in its natural sources and this can be a limitation to obtain adequate CFTR protein for biochemical and structural studies (Riordan, 1993). The biogenesis of CFTR includes its insertion into the ER (endoplasmic reticulum) membrane in which CFTR interacts with the cytosolic chaperones Hsc70/Hdj-2 and Hsp90, and with the ER chaperone calnexin (Yang et al, 1993). These chaperones are involved in CFTR maturation but only ~30% of the immature wild-type (wt) CFTR is released by the chaperones and placed in the plasma membrane (Lukacs et al., 1994). The low expression level of CFTR does not just exist in its natural sources but is also noticed in cells that do not naturally produce the protein such as yeast cells. In a yeast expression system, Kiser et al (2001) created mutations that prevented CFTR degradation in the ER but the expression level of the protein was still low compared to other proteins. CFTR was expressed in Chinese Hamster Ovary (CHO) cells and it was purified with low yield because it was degraded by the ER quality control system (O'Riordan et al, 1995). A higher level of the CFTR expression in baculovirus-infected insect cells was reported by (Kartner et al, 1991) and (Bear et al, 1992) but CFTR was solubilised with detergents that are probably not suitable for crystallization studies. Furthermore, CFTR was expressed in bacteria cells (*E. coli*) (Gregory et al, 1990) and (*Lactococcus lactis*) (Steen et al, 2011). The protein was detectable in the bacteria membrane with a very low yield (< 0.1% of membrane proteins) and with misfolding problems.

CFTR is an integral membrane protein that consists of hydrophilic and hydrophobic regions. One approach is to cover hydrophobic regions with detergents to allow purification and crystallization. An alternative solution for the crystallization of membrane proteins is to study only the soluble domains of the protein. Many successful experiments to express and purify such constructs of CFTR were made. A plasmid encoding amino acid sequences of NBD1 and 2 was expressed in *E. coli*. The majority of CFTR mutations are within or near these two nucleotide-binding domains (NBDs) (Ko *et al.*, 1993). Additionally, some studies showed that the interaction of ATP with NBDs is essential for opening and closing the chloride channel (Anderson *et al*., 1991). Lu and Pedersen (2000) also examined the relationship between CFTR's soluble domains (which included the two nucleotide binding domains NBD1 and NBD2) and the regulatory domain (R). They found that the C-terminus of NBD2, which is located within the C-terminal section of CFTR, may have two important roles. The first role is its ability to control CFTR function including opening and closing of the channel pore. The second role is its capability to interact with other proteins that can affect CFTR activities. A short carboxyl-terminal peptide of CFTR (QDTRL) fused to NHERF1 PDZ 1 was expressed in BL21 (DE3) cells and its structure was solved (Karthikeyan et al, 2001). Therien et al (2002) have studied small constructs of CFTR (74-146 and 194-241 amino acids) in *E.coli* and purified them by affinity chromatography with a high yield of protein for crystallization studies. The C-terminal residues 1377-1480 of CFTR was expressed in BL21 (DE3) cells and a soluble protein was obtained and used for biochemical experiments (Wolde et al, 2007). And more recently, NBD1 of CFTR (T389-A655) was expressed and purified in BL21 (DE3) as a soluble protein (Faria et al, 2011).

1.5 Characterization of Protein Using Circular Dichroism

Circular dichroism (CD) is being increasingly considered as a valuable tool to provide complementary structural information about proteins. Secondary structure and tertiary structure are detectable by far UV CD (160-280 nm) and near UV CD (260-320) (Kelly et al, 2005). The basic idea behind the CD technique depends on measuring the absorption difference (that occurs in a chiral chromophore of protein) between two components of plane polarised light, the left (L) and right (R) circularly

polarised components of radiation. Both components have equal magnitude in unpolarised light.. If the polarised components of light are absorbed by protein to different extents, this difference in light absorption creates elliptical polarisation of radiation. The right and left polarised light are not combined together by CD. In practice, they are measured as two separated components and then it will be displayed as either the difference in absorbance of the two components ($\Delta A = A_L - A_R$) or as the ellipticity in degrees (θ) in which CD instruments (known as spectropolarimeters) measure the minor and major axes that are formed by the ellipticity (θ) (Figure 14).

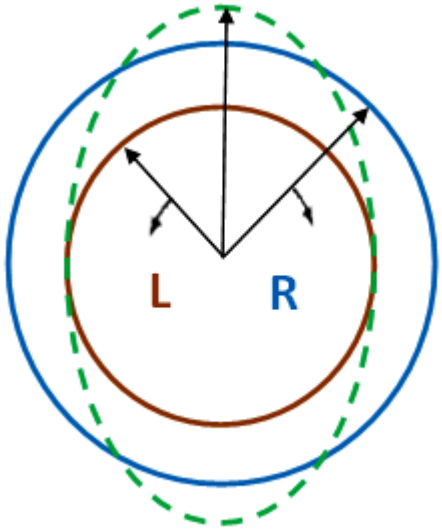


Figure 14: Plane polarised light (brown is directed to Left and blue is directed to Right) is absorbed by a chiral chromophore of protein. The dashed green line is the elliptical radiation that results from the differential absorption of left and right polarised light.

The obtained CD signal from most biological samples is small (measured in millidegrees) So it is clear that experimental conditions must be optimized to obtain a good signal: noise CD spectrum (Kelly and Price, 2000). Secondary structure composition (helix, sheet and random coil) can be determined from CD spectra as their peptide bonds absorb left and right circularly polarised radiation differently in the range from 190-220 nm. Also, some information on tertiary structure of proteins can

be determined by the CD signal in the region 260–320 nm that arises from the aromatic amino acids. The three amino acids (Phe, Tyr and Trp) have unique wavelength signatures. When protein is folded (containing secondary or tertiary structure), the peptide bonds and aromatic side chains are placed in chiral environments, and consequently they give rise to CD (Berova et al, 2000).

Several studies were performed on CFTR by CD. A construct of hNBD2 of CFTR was studied by CD in 1994 by (Ko et al). Important structural data were delivered by this experiment before the first structure of NBD2 was obtained in 2010 by X-ray crystallography. Ko et al (1994) found that the construct of 51 amino acids of hNBD2 (1228-1278) had secondary structural elements of (20-28% α-helix, 6-11% β-strand, 18-20% turns, and 43-53% random coil). Furthermore, a 51-NBD2 construct was able to bind to TNP-ATP (Trinitrophenyl-ATP), but not in the presence of urea (denaturation state of the protein) and this gave evidence that the peptide tertiary structure was essential for nucleotide binding. This experiment was the first direct data that showed an interaction between the second nucleotide-binding domain (hNBD2) and ATP. Therien et al (2001) have studied the effect of SDS (Sodium dodecylsulphate) and LPC (lysophosphatidylcholine) detergents on the secondary structure of TM 3–4 constructs of CFTR. The CD spectra of these experiments had a minima at 208 and 222 nm, respectively, indicating an α-helix structure and the protein was denatured by SDS but not by LPC. The effect of thermal unfolding on the secondary and tertiary structures of human NBD1 CFTR (with and without F508del) was investigated by far- and near-UV CD. No significant difference was noticed between the two constructs in terms of their CD spectrum (ie overall secondary structure composition) (Protasevich et al, 2010). And this result agreed with the X-ray crystallography results that the mutation of F508 creates only limited conformational changes near the site of the deleted amino acid in the crystal structures of several human NBD1 constructs (Thibodeau et al, 2005).

The CD spectrum of NHERF1 PDZ 1 (Figure 15) reported by (Sun and Mierke, 2005) showed that the protein consisted of a small percentage of α-helix (16.8%) and a higher percentage of β-strand (approximately 40%) and this result is a compatible with the X-ray crystallography structure of NHERF1 PDZ 1 which has a six-stranded antiparallel β-barrel capped by two a-helices (Ladias, 2003).

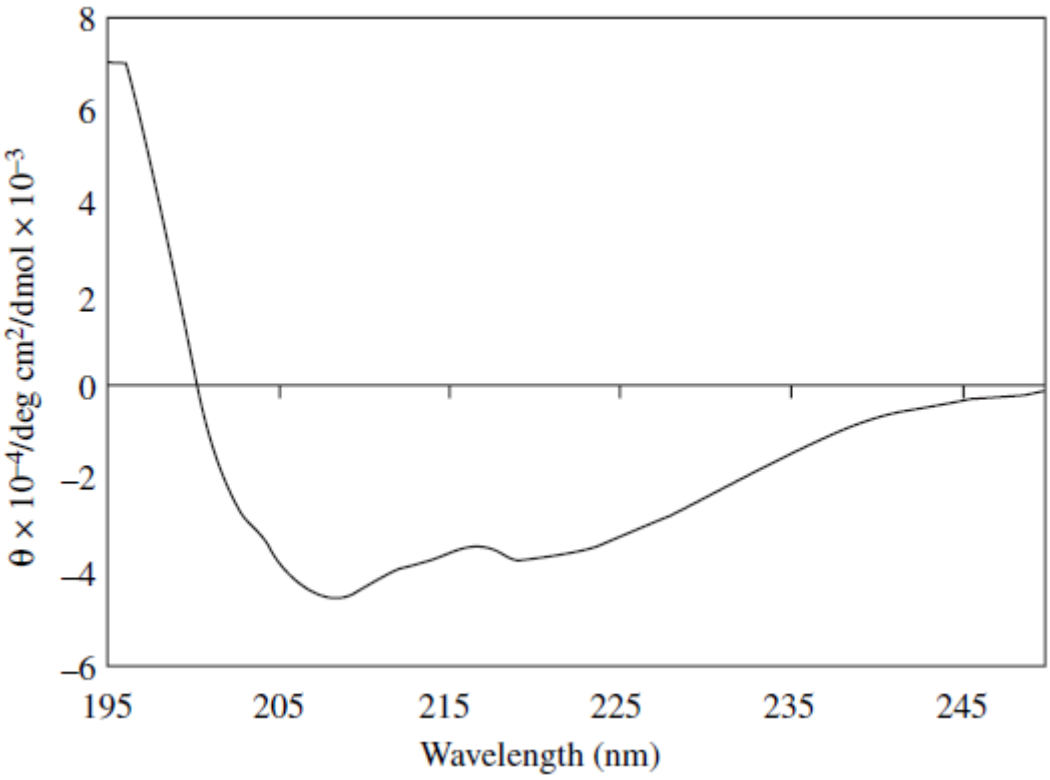


Figure 15: Circular dichroism signal of the PDZ 1 of NHERF1 protein. The CD signal was collected in 10 mM PBS (phosphate-buffered saline) at a pH 7.0 at 25^{o}C.

1.6 Characterization of Protein by MALDI-TOF Mass Spectrometry

Mass spectrometry is an important technique for the characterization of proteins and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) is one of the most common mass spectrometry methods used to investigate protein function and structure. There are two approaches to study proteins by mass spectrometry. In the first, whole proteins (undigested) are ionized and then subject to a mass analyser. In the second, proteins (in solution or isolated from SDS-PAGE gel) are digested into smaller peptides using proteases such as trypsin or pepsin (other proteolytic agents are also used). The digested proteins are ionized and analysed the same way as for undigested proteins. The final stage is to use peptide identification online software to match the masses of digested peptides with known protein sequences. MALDI-TOF (the technique that was used in this study) was introduced as an analytical technique for large biomolecules (> 10000 Dalton) in 1988 by (Karas and Hillenkamp). Protein samples in MALDI-TOF are mixed with organic compounds (matrix) to form a mixture of protein-matrix. The mixture is then illuminated by a nanosecond laser pulse in the range of 266 to 337 nm. Most of laser energy is absorbed by the matrix and this prevents fragmentation of protein sample. The vaporised and ionized protein is accelerated in an electric field and enters the flight tube. Different protein samples (or peptides) are separated according to their mass to charge ratio during the flight in the tube. The small ionized molecules with greater charge reach the detector faster than the big ones. So, MALDI-TOF detects different Mass/Charge ratios as signal peaks (Jurinke et al, 2004) representing different size and charge states of the protein. Singly charge or doubly charged protein ions are usually detectable. The study of protein-protein or protein-ligand interactions is important because there are many biological processes which are related to these interactions (Figeys et al, 2001). CFTR interactions were studied by MALDI-MS (Matrix-assisted laser desorption/Mass spectrometry) to draw a wider picture of CFTR roles. In 1997, the phosphorylated amino acids of full-length CFTR were identified by MALDI-MS (Neville et al, 1997). They reported, for the first time, phosphorylation of serine 753 which was not identified in the study by (Townsend et al, 1996) in which 13 serine residues (with at least one arginine residue) were phosphorylated by PKA (cAMP-dependent protein kinase). The reaction catalysed by PKA at serine,

threonine or tyrosine residues causes the addition of phosphate groups to the theoretical mass of unphosphorylated peptides. This comparison between phosphorylated and unphosphorylated peptide allows the detection of phosphorylation sites on protein (Townsend et al, 1996). Another interaction was detected by MALDI-MS in which filamin-A (actin-binding proteins) bound directly to the N-terminus of CFTR (1-25 amino acids). This interaction has a critical role in that CFTR is stabilized at the cell surface (Thelin et al, 2007). MALDI-MS was used to identify the interaction between CFTR and several further proteins. Wild-type CFTR, F508del-CFTR and F4RK-CFTR (substitution of arginine by lysine at positions R29, R516, R555 and R766 that allows escape from the Endoplasmic Reticulum quality control machinery and functioning of CFTR at the cell surface) were expressed in Baby Hamster Kidney (BHK) cells. The cell lysates were subjected to Western blotting and protein spots were analyzed by MALD-MS. 71 protein spots recognized as differentially expressed between the three cell lines (Gomes-Alves et al, 2010).

1.7 Aims and Objectives of this Study

Defective CFTR can cause the loss of ion transport function. As a result, many patients die at young ages. So, attempts to visualize the 3D structure of this protein are worthy from both a basic and applied science perspective since knowledge of the 3D structure could lead to the development of new drugs to treat the condition. Successful crystallization of transmembrane proteins is limited by the presence of hydrophobic regions. These regions cause protein to aggregate when they are exposed to aqueous solution so no crystals can be formed. Crystallization of predicted hydrophilic portions of a protein is an alternative solution to deal with aggregation problems.

As revealed in Table 5, the structures of NBD1 and 2, R-domain and some small portions of CFTR have been published. The aim of this study was to investigate the structure and function of the last 42 amino acids (later extended to 61 aa). This fragment of the protein has no structural data (except for the last four residues in complex with NHERF1 PDZ1). Secondly, it contains a PDZ motif which has structural and functional importance. Thirdly, it is a soluble polypeptide (a hydrophilic region located in the cytoplasm) that can be studied and potentially crystallized with fewer expected problems compared to the full-length protein. Later, the C-terminus of killifish and shark CFTR were also investigated in terms of their expression and purification as a comparative study to the human CFTR C-terminus.

Another aim of this study was to explore the effect of the interaction between the C-terminus of CFTR and NHERF1 PDZ 1 on the structure of CFTR. For this, the 3D structure of full-length CFTR-NHERF1 PDZ $1^{(+)}$ was generated by Single Particle Analysis. In addition, the full-length human, killifish and shark CFTR-NHERF1 PDZ $1^{(+)}$ interactions were studied in different conditions by using a pull-down assay. These biochemical studies confirmed the binding of NHERF1 PDZ 1 to full-length CFTR, under the conditions used for the structural studies.

This study is divided into two main themes:

(i) Studying Partial Sequences of CFTR Including:

- Cloning of the plasmid containing these sequences into host cells, for example, *E.coli* followed by large-scale expression and purification of the polypeptide.

- Following protein expression and purification, many protein characterization methods are used including SDS-PAGE, Mass spectrometry and Circular Dichroism.

- Crystallization trials performed in order to obtain 3D crystals by using techniques such as, sitting drop (Vapor Diffusion).

(ii) Studying Full-length CFTR Including:

- Single particle analysis of full-length human CFTR in the presence of NHERF1 PDZ $1^{(+)}$ (or isolated CFTR) was performed. 3D structures are compared with high resolution 3D structures of homologous proteins.
- Protein-protein interactions investigated by pull-down assay. Human, killifish and mouse CFTR (with 10His tags) are allowed to interact with NHERF1 PDZ $1^{(+)}$ (No tag) and results are analysed by SDS-PAGE.

2. Materials and Methods

2.1 Materials

2.1.1 Source of Given Proteins

1- Full-length human CFTR: For experiments that required full-length human CFTR, the samples were received from our collaborator, Professor John R Riordan (CF/Pulmonary Research & Treatment Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7248, USA).

2- Full-length mouse and killifish CFTR: Mouse and killifish CFTR samples (≈ 0.3 mg/mL) were kindly provided by Natasha Cant, another PhD student in the Ford laboratory at the University of Manchester and expression and purification of these proteins have been partly described recently (O'Ryan et al, 2012). The purity of all the three proteins can be assessed from the SDS-PAGE data shown in section 3.2.1. Human CFTR had a higher purity because it was purified by two affinity chromatography steps. Mouse and killifish protein samples have a lower purity as they were obtained from a single affinity chromatography step (Talon immobilized metal affinity chromatography (IMAC) purification).

2.1.2. Growth Media

LB Broth was used for the growth of recombinant *E.coli* strains. This medium is prepared using the Miller formula (Miller, 1972). Enzymatic Digest of Casein 10 g, Yeast Extract 5 g and Sodium Chloride 10 g. Formula / Liter. Final pH: 7.3 ± 0.2 at 25 °C.

2.1.3 Antibiotics

Kanamycin (50 mg/ml in MQ water that was obtained by using a Milli-Q Purification System) was prepared fresh and sterilized by 0.20 µm filter, and stored at -20 °C. The final concentration in the media is 50 µg/mL.

Ampicillin (200 mg/ml in MQ water) was prepared fresh and sterilized by filtration using a 0.20 µm filter, and stored at -20 °C. The final concentration in the media is 200 µg/mL.

2.1.4 Inducers

A 1M stock of IPTG (Isopropyl β-D-1-thiogalactopyranoside) was prepared in MQ water. The mixture was sterilized by filtration using a 0.20 µm filter, and stored at -20 °C. The final concentration in the media is 1 mM.

2.1.5 Buffers and Solutions

2.1.5.1 Buffer A (Unbound Materials Washing Buffer): 500 mM NaCl, 50 mM HEPES pH 7.0, 0.01 % Sodium Azide (NaN_3) and 1 mM 2-Mercaptoethanol.

2.1.5.2 Buffer B (Elution Buffer): 500 mM NaCl, 50 mM HEPES pH 7.0, 0.01 % Sodium Azide (NaN_3), 1 mM 2-Mercaptoethanol, 250 mM Imidazol.

2.1.5.3 Buffer C (Sonication Buffer): 500 mM NaCl, 50 mM HEPES pH 7.0, 0.01 % Sodium Azide (NaN_3), 1 mM 2-Mercaptoethanol, 0.1 % TritonX-100, 100mg/ml Lysozyme.

2.1.5.4 Dialysis Buffer: 100 mM NaCl, 50 mM Tris pH 7.9, 0.01 % Sodium Azide (NaN_3), 1 mM 2-Mercaptoethanol, 1 mM EDTA , 2 mM $Cacl_2$.

2.1.5.5 GST (the glutathione S-transferase) tag Purification Buffers:

- Extraction buffer: 140 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1 % TritonX-100, 100mg/ml Lysozyme. Wash buffer: 140 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4. Elution buffer: 33 mM Glutathione in 50 mM Tris-HCl pH 8.

2.1.5.6 Gel Filtration Buffer: 100 mM NaCl, 50 mM Tris pH 7.5, 0.01 % Sodium Azide (NaN_3), 1 mM 2-Mercaptoethanol, 1 mM EDTA.

2.1.5.7 CFTR Buffer (DDM): 150 mM NaCl, 40 mM Tris pH 7.5, 0.01 % sodium azide (NaN_3), 0.05 % n-dodecyl- β- D- maltoside (DDM), 5 mM 2-mercaptoethanol.

2.1.5.8 CFTR Buffer (LPG): 50 mM NaCl, 20 mM HEPES pH 7.5, 0.05 % Lysophosphatidylglycerol (LPG).

2.1.5.9 Phosphate Buffer: 10 mM Na/K phosphate pH 7.5 (80.2 ml of 1 M K_2HPO_4 was mixed with 19.8 ml of 1 M KH_2PO_4 and the solution was made up to 1L with MQ water) , 1 mM EDTA, 1 mM 2-mercaptoethanol.

2.1.5.10 Thrombin Buffer: 50 mM Tris-HCl pH 7.7, 300 mM Potassium Acetate, 7 mM Magnesium Acetate, 10 % (v/v) Glycerol.

2.1.5.11 TFB1 Buffer: 30 mM potassium acetate pH 5.8, 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$ and 15 % (v/v) Glycerol.

2.1.5.12 TFB2 Buffer: 10 mM PIPES pH 6.8,10 mM RbCl, 75 mM $CaCl_2$, 15 % (v/v) Glycerol.

2.1.5.13 Uranyl Acetate Solution: 4 % (w/v) Uranyl Acetate (Agar Scientific, UK) was prepared in MQ water.

2.1.5.14 Super Blue Coomassie Stain: 10 % Acetic Acid, 45 % MQ water, 45 % Ethanol, 0.25 % Coomassie blue, made up to 500 ml.

2.1.5.15 15 % Tricine Gel Buffers

Materials	Resolving gel buffer	Stacking gel buffer
H_2O	1.5 mL	1.6 mL
30 % (w/v) Acrylamide	2.5 mL	0.4 mL
3X Gel Buffer	2 mL	1.0 mL
APS	20 µL	20 µL
TEMED	20 µL	20 µL

Table 6: Materials that were used for preparation of 15 % Tricine gels.

3X Gel Buffer: 3M TRIS pH 8.45, 0.3 % SDS. 10X Top Buffer: 1M TRIS pH 8.25, 1M Tricine, 1 % SDS. 10X Bottom Buffer: 2M TRIS pH 8.9.

2.1.5.16 Agarose Gel Buffers

TAE Buffer (50X): Tris-Base 142 g, Glacial acetic acid 57 g, EDTA 50 mL of 0.5 M, pH 8.0, Total Volume 1 Liter. DNA Loading Buffer (6X): 0.25 % (w/v) Bromophenol Blue, 0.25% (w/v) Xylene Cyanol, and 30 % (v/v) Glycerol, 50 mM EDTA. Ethidium Bromide: 10 mg/mL (w/v) in Ethanol. DNA Marker (100 bp Ladder and 1K Ladder): DNA markers were purchased from Fermentas (UK).

2.1.5.17 M9 / Minimal Media for NMR (Nuclear Magnetic Resonance) Analysis

Solution A: Na_2HPO_4 anhyd 2.5g, KH_2PO_4 anhyd 7.5 g, 1 L was made in dH_2O at pH 7.2 and then was autoclaved.

Solution B: Glucose 4 g, NH_4Cl 1 g, $MgSO_4.7H_2O$ 0.24 g, $CaCl_2.2H_2O$ 0.02 g, Thiamine 0.01 g, Antibiotic 0.1 g. 20~ml was dissolved dH_2O and sterilized by filtering through a 0.2 μm filter.

Trace elements were prepared up to a total volume of 100 ml . $CaCl_2.2H_2O$ 550 mg, $MnSO_4.H_2O$ 140 mg, $CuSO_4.5H_2O$ 40 mg, $ZnSO_4.7H_2O$ 220 mg, $CoCl_2.6H_2O$ 45 mg, $Na_2MoO_4.2H_2O$ 26 mg, $H_3BO_4$40 mg, KI 26 mg. The previous chemicals were added to 70 ml dH_2O and adjusted to pH 8.0 and then 500 mg of EDTA was added and readjusted the pH to 8.0. A 375 mg of $FeSO_4.7H_2O$ was added and the solution was made up to 100 mls and was autoclaved.

2.1.6 *E. coli* Strains and Vectors

Strains	Genotype	Source	Function
DH5α	F$^-$ φ80*lacZ*ΔM15 Δ(□*lacZ*YA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+)*pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ$^-$	Invitrogen	Minipreparation of DNA
XL1-Blue	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB+ lacIq Δ(lacZ)M15] hsdR17(rK-mK+)	Stratagene	Ligation
T7	*miniF-lacIq(CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 D(mcrC-mrr)114::IS10*	New England Biolabs	Protein Expression
BL21(DE3)	*F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Novagen	Protein Expression

Table 7: *E.coli* strains used in this study.

Vectors	Tag	Source
pET 24(a)	N-terminal His-Tag, Thrombin site, T7 promoter, Kanamycin resistance.	Geneart
pET 52(b)	N-terminal Strep-Tag II followed by a recognition site for the human rhinovirus (HRV) 3C protease, T7 promoter, Ampicillin resistance.	Novagen
pGEX-2TJL1	Thrombin or factor Xa protease sites to cleave protein from fusion, tac promoter, N-terminal T7-Tag, an optional C-terminal His-Tag, T7 promoter, Ampicillin resistance.	Pharmacia
pGA4	N-terminal His-Tag, Thrombin site, T7 promoter, Ampicillin resistance.	Geneart

Table 8: Vectors used in this study.

Two vectors were used in this study to express NHERF1 PDZ 1. The pGEX-2TJL1 vector contains the gene for NHERF1 PDZ $1^{(+)}$. The pET 52(b) vector contains the gene for NHERF1 PDZ 1. The sequences for the two forms of NHERF1 (NHERF1 PDZ $1^{(+)}$ and NHERF1 PDZ 1) are in Appendix C. Both forms of NHERF1 proteins are human. All the vectors maps can be found in Appendix C.

2.1.7 PCR Primers (Oligonucleotides)

Table 9 shows the PCR primers that were designed for amplifying selected regions of the CFTR gene.

Name	Sequence	Description	Vector
P1For	5` CGTGGATCGTCTACCATGGGGCTGCCCCGG 3`	Forward	pET-52b
P2Rev	5` GAATTGTCAGGATCCTCATTACAGCAGAGAGTCGTTC 3`	Reverse	
Shark For	5` CGTGGCAGCATGCAGGTCTTTGGACATTTGGACCGTGC 3`	Forward	pGA4
Shark Rev	5` GTTCTTTCTCGAGTTATAACCTTGTTTCTTGAAGGTCTTCC 3`	Reverse	
Killifish For	5` CGAGATGTCTCACCGCATGCAATTTCCTCC 3`	Forward	pGA4
Killifish Rev	5` GATGATGCTCGAGTTATTAAAGTCTGGTATCCTG 3`	Reverse	
P R-A For	5` GAAGTGCAGGATACCGCTCTGTAATAACTCGGGCACCAC 3`	Forward	pET-24a
P R-A Rev	5` GTGGTGCCCGAGTTATTACAGAGCGGTATCCTGCACTTC 3`	Reverse	

Table 9: Primers used in this study.

2.2 Methods

2.2.1 Designing Primers

2.2.1.1 P1For and P2Rev Primers:

These two primers (Table 9) are designed for sub-cloning NHERF1 PDZ 1 (91 amino acids) from a pGEX-2TJL1 plasmid to a new plasmid (pET-52b). The new plasmid does not contain a tag (No thrombin site, no 6His.). The reason for this was to co-express the gene with pET24a 6-His tagged 1438-1480 C-terminus^{42aa} and co-purify them together. Two restriction enzyme sites were chosen, these are *Nco I* (CCATGG) which is present in P1For and *BamH I* (GGATCC) which is present in P2Rev. A successful PCR reaction will give rise to a 236 bp band. Digestion of this fragment (after using QIAGEN`s PCR cleaning-up kit) with *BamH I* / *Nco I* and subsequent ligation of this into a *BamH I* / *Nco I* cleaved pET-52b vector will give rise to a fragment amenable to co-expression trials.

2.2.1.2 R-AFor and R-ARev Primers:

This Primer R-A (Table 9) was designed for site-directed mutation in which Arginine at position 1479 of the C-terminus^{42aa} of CFTR is replaced with Alanine. The purposes of this mutation are the replacement of a big, flexible amino acid, Arginine with a small amino acid, Alanine, in order to try to enhance the crystallization of the C-terminus^{42aa}. An XhoI restriction site was destroyed by the mutation which then could be used for checking the construct.

2.2.1.3 Shark and Killifish Forward and Shark Reverse Primers:

The two primers (Table 9) were designed to sub-clone shark CFTR C-terminus (1436 – 1480) by introducing a SphI site and an XhoI site into the primers for efficient cloning. An additional methionine was added to the sequence. The aim of studying shark CFTR C-terminus was to perform a comparative study among human, shark and killifish CFTR C-terminii in terms of expression and purification.

2.2.2 Recombinant DNA Processing and Manipulation

General procedures for cloning and DNA manipulation were performed as described by Sambrook, *et al*., (1989) and according to product manufacturer's instructions. All initial/basic DNA manipulations and vector construction experiments were performed using *E. coli* DH5α cells.

2.2.2.1 PCR Amplification of DNA Segments of Interest.

For amplifying the desired segments of DNA using the PCR technique, a PCR mixture was prepared as described in Table 10. Failsafe™ PCR premix kit was purchased from Epicentre Biotechnologies.

Materials	Volume
Genomic DNA	1 μL
Primer-F (125 ng)	1 μL
Primer-R (125 ng)	1 μL
Failsafe PCR 2X Premix G	23 μL
Failsafe PCR enzyme Mix 1.25Units	0.5 μL
MQ water	23.5 μL

Table 10: PCR reaction mixture.

Three different temperatures were applied to the mixture for 30 cycles of: denaturation at 94 °C for 30s, annealing at 50 °C for 20s and extension at 70 °C for 30s. The PCR products were purified from unwanted enzymes and buffers using Qiagen PCR purification kit. The final concentration of the purified PCR product was 50 μL.

2.2.2.2 Insert Restriction Digestion

The amplified PCR products were mixed with the appropriate restriction enzymes to make sticky ends. The mixture was prepared as followed: 50 μL DNA, 6.5 μL Enzyme Buffer, 2 μL MQ water,

3.23 μL Enzyme A, 3.23 μL Enzyme B. The mixture was centrifuged at 13,000 rpm in a benchtop Eppendorf microfuge for a few seconds and incubated at 37 °C for 4 hrs, before being loaded onto 1 % agarose DNA mini gels (TAE) and subsequently purified by the Qiagen gel extraction kit protocol.

2.2.2.3 Vector Restriction Digestion

Matching restriction enzymes were used to cut the vector in order to create compatible sticky ends with the insert sticky ends. Identical amounts and materials of the insert restriction digestion mixture were used. Then, the vector was gel purified by the Qiagen gel extraction kit protocol.

2.2.2.4 Ligation and Transformation

After preparing the insert and vector, the ligation of the two pieces of DNA was done. The ligation mixture was prepared as follows: Mixture1: 7 μL vector, 1 μL insert, 2 μL 10X reaction buffer, 10 μL MQ water and 1 μL T4 DNA ligase. (T4 DNA ligase and reaction buffer purchased from New England Biolabs). Mixture2: 4 μL vector, 4 μL insert, 2 μL 10X reaction buffer, 10 μL MQ water and 1 μL T4 DNA ligase. Mixture3: 1 μL vector, 7 μL insert, 2 μL 10X reaction buffer, 10 μL MQ water and 1 μL T4 DNA ligase. Mixture4: 4 μL vector, No insert, 2 μL 10X reaction buffer, 14 μL MQ water and 1 μL T4 DNA ligase. The mixtures were left at 16 °C for 2 hrs. 3 μL of ligation mixture was transformed directly into a 50 μL aliquot of library efficient competent cells (*E.coli* DH5α strain, Invitrogen). The cells were incubated on ice for 30 minutes, followed by heat-shock at 42 °C for 45 seconds. Subsequently, 500 μL of LB Broth was added to the cells and incubated at 37 °C for 1hr. 200 μl of resulting cell suspension was plated onto LB agar plates containing the relevant antibiotic. The plates were incubated at 37 °C overnight. On the following day, individual colonies were picked from the plates and inoculated into 5 ml of LB Broth medium in Falcon tubes. The cultures were grown overnight at 37 °C.

2.2.2.5 Site-Directed Mutagenesis (Quick Change)

Two primers (section 2.2.1.2) were used for making the R-A mutation and for amplifying the desired segments of DNA. Site-Directed Mutagenesis mixture was prepared (1 μL DNA template, 1 μL 10× buffer, 5 μL dNTP mix, (10 mM each), 1 μL primer A, 1 μL primer B, 31 μL MQ and 1 μL high-fidelity DNA polymerase (Phusion). Three different temperatures were applied to the mixture for 18 cycles of denaturation at 98 °C for 30s, annealing at 55 °C for 30s and extension at 72 °C for 3.5 minutes. Following the PCR reaction, the PCR products were checked by 1% Agarose gel. 2 μL of DpnI restriction enzyme were added to the PCR products to destroy the methylated wild-type template DNA (samples were left for 2 hrs at 37 °C). The non-methylated nicked mutant DNA was transformed into XL1 Blue super competent *E. coli* cells (Stratagene). The new mutated plasmid was obtained using mini-prep. Kit (Invitrogen).

2.2.2.6 Gel Electrophoresis

Agarose gel electrophoresis was used for checking DNA size and for separation of DNA fragments, all agarose gel buffers and solutions were as described in 2.1.5.16. 1 % agarose gels was prepared by dissolving agarose (0.5 g) in TAE buffer (50 mL), in a 250 mL conical flask via melting in a microwave. After allowing it to cool slightly, ethidium bromide (2.5 μL of 0.35 μg /mL solution) was added, and agarose solution was poured into a casting tray of the desired size. After inserting a comb, the gel was allowed to solidify for at least 30 min, and then it was placed in an electrophoresis tank and submerged in TAE buffer. DNA samples and marker DNA mixed with DNA loading buffer were loaded into the gel wells. Gels were electrophoresed at a constant voltage in the range of 80-100 V for 40 minutes to 1hr and then visualized and photographed under UV light using a transilluminator. The amount of DNA in the samples was estimated by comparing its intensity with the known amounts of marker DNA. Two DNA ladders were used, 100 bp and 1 kb (NEW ENGLAND Biolabs).

2.2.3 Preparation of Competent *E. coli*

Luria-Bertani agar plates (LB agar) were inoculated with the commercially available cell lines, Table 1. *E. coli* cells were streaked onto LB agar plates containing 50 µg/µl kanamycin (or 200 µg/µl ampicillin). The plates were inverted and incubated at 37 °C for 16 hours. A single colony from each was used to inoculate 50 ml of LB broth containing 10 mM of sterile $MgSO_4$ (LB broth, with appropriate antibiotics) and grown overnight with shaking at 200RPM and 37 °C. The next day, 1 ml of the saturated overnight cultures was added to a sterilised conical flask containing 100 ml of LB broth (containing 10 mM $MgSO_4$). The cultures were maintained under the same conditions as the overnight cultures until the optical density at 600 nm reached 0.6 OD units. At this point the cells were divided equally between two pre-cooled (on ice) sterile 50 mL centrifuge tubes. The cells were pelleted at 3000 rpm for 10 minutes at 4 °C in a JS-25.50 rotor (Beckman centrifuge J6-HC). The supernatant was discarded and the cells were resuspended with gentle pipetting with 17 ml of ice-cold TFB1 (as described 2.1.5.11) that had been filter sterilised using a 0.2 µm pore syringe filter. The cell suspension was left on ice for 1hr and the cells re-pelleted under the same conditions as above, at 2000 rpm for 15 minutes. Cells were then resuspended in 4 ml of sterile ice-cold TFB2 (as described 2.1.5.12). The mixture was left to incubate on ice for 15 minutes. Cells were aliquoted (200 µl per chilled 1.5 ml microfuge tubes) and then snap frozen using an ethanol/dry ice bath and stored at –80 °C until required.

2.2.4 Expression and Purification

2.2.4.1 Small-Scale Expression of Targets.

DH5α and BL21 (DE3) *E.coli* cells were used for transformation of the desired plasmids as explained in section 2.2.2.4. For preparing small-scale expression of targets in LB Broth medium, on the next day, individual colonies were picked from the plates and inoculated into 250 mL of LB Broth medium in autoclaved flask. The cultures were grown overnight at 37 °C.

2.2.4.2 Large-Scale Expression of Targets.

50 mL of overnight cultures was inoculated into 1L of autoclaved LB broth containing an appropriate antibiotic (4 flasks were prepared). The 4L cultures were incubated in a shaker at 37 °C at 200 RPM. The reading of OD600 was monitored every hour. When the OD600 reached 0.6 – 0.8, the expression of the target was induced by addition of IPTG at a final concentration of 1mM. Before the induction, the four flasks were incubated on ice for 5 minutes to reduce the temperature to nearly 25 °C and the shaker was adjusted to the same temperature. The induced cultures were left to grow for 4 hrs at 25 °C and cells were then pelleted by centrifugation at 10000 rpm (F10BCI-6x500Y Rotor, Beckman), at 4 °C for 15 minutes and the spent LB Broth poured off. The pellet was kept at – 20 °C.

2.2.4.3 Protein Purification Using Metal Affinity Chromatography

Talon resin (CLONTECH) was used for purifying the Histidine tagged proteins. First, the bacterial pellet was thawed on ice. Buffer C (described in section 2.1.5.3) was added to the pellet, 20 mL of Buffer C was added per 1L of LB Broth starting culture. A serine protease inhibitor, PMSF (phenyl methane sulphonyl fluoride) was added to the pellet at a concentration of 20 μg/mL (20 mg/mL stock solution prepared in propan-2-ol). The thawing process took about 30 minutes. The pellet was resuspended by gentle aspiration with an automatic pipette before transferring the mixture into 50 mL Falcon tubes for sonication (2 Falcon tubes were used for 4L of the cultures). Each sample was sonicated for 5 seconds, allowing 10 seconds cooling period with each cycle repeated 10-15 times at 30 % amplitude (Sonics Vibra Cell). PMSF was added again as above. The sonicated samples were transferred to centrifugation tubes and centrifuged at 16000 rpm using a JA-17 Rotor, Beckman, for 20 minutes at 4 °C. After this step, Protamine Sulfate (Sigma) was added at a concentration of 1 μg/mL in order to bind and precipitate DNA, and the two tubes were re-centrifuged for 20 minutes. Finally, the pH was adjusted to 7 by adding of sodium hydroxide (0.002M). The supernatants were added to Talon cobalt resin in Falcon tubes as described below. Talon immobilized metal affinity chromatography (IMAC) resin was equilibrated with buffer B before mixing with the supernatants. Two Falcon tubes containing Talon resin were prepared. Typically 0.2 ml bed volume was used per 4

L of starting LB Broth. The Talon/supernatant mix was left at 4 °C for 1 hour with gentle rotation. Three steps followed the incubation in the cold room. First of all, the contents of two Falcon tubes were loaded into a 20 mL gravity flow column (Biorad), and unbound material that passed through the column and was collected and kept in a 200 mL beaker. Then, the resin was washed with Buffer A (as described 2.1.5.1) (10 times the volume of resin used) to remove unbound proteins. The final step was the elution of the desired protein by adding Buffer B (as described in 2.1.5.2) to the resin. The protein was eluted from Talon resin using 20 mL of Buffer B and collected in a 50 ml Falcon tube. The purity was checked by SDS–PAGE. The purified protein was pipetted into a 5000 MWCO dialysis membrane (*purchased from BDH chemicals Ltd*). The membrane was sealed with plastic clips at its two ends and placed into 1L of dialysis buffer (described in 2.1.5.4) and incubated at 4 °C overnight with movement by a small stirrer bar at low speed. If cleavage of the purified protein with thrombin was required, 100 µL of thrombin was added to the protein sample before it was placed into the dialysis membrane and 2mM $CaCl_2$ was added to the protein sample and dialysis buffer. After the dialysis step, the mixture that was placed into the dialysis membrane was loaded into a new Falcon tube that contained the same amount of Talon resin as before. The resin was equilibrated using dialysis buffer. The Falcon tube was incubated on a 1RPM rotator in the cold room (4 °C) for 1 hr. The three purification steps (described above) were repeated again in order to obtain the protein. Into the first falcon tube, unbound materials including the cleaved protein were collected by gravity flow. Into the second falcon tube, the remaining resin was washed with buffer A. Into the third falcon tube, unwanted materials were eluted by adding Buffer B. SDS PAGE analysis followed the purification. Three chemicals, 0.01 % Sodium azide (10 % stock solution), 10 mM β-mercaptoethanol and 2mM EDTA (ethylenediaminetetraacetic acid), were added to the purified protein in order to promote stability.

2.2.5 Protein Concentration

Following the above purification scheme, the purified protein was filtered through a 50kDa cutoff filter and then concentrated using a 3kDa cutoff filter (Centricon Centrifugal Filter Units, 3 and 50

KDa microconcentrator devices, from Millipore). The two concentrators were washed with dH_2O and equilibrated with Buffer A before use. The protein was concentrated to the volume of 0.5 – 0.8 mL at 4 °C at 6000 rpm and then transferred into a 3kDa Micron tube (Millipore) for further concentration to the volume of 0.1 mL. The concentrated protein was subsequently stored at –20 °C.

2.2.6 SDS-PAGE Analysis

SDS-PAGE was used for determining the purity and yield of proteins and for determining the concentration of the purified protein by comparison with known protein concentrations. 15 % TRICINE gels were prepared as described in 2.1.5.15. 2X SDS gel loading buffer was added to protein samples, spun and heated at 90 °C for 5 minutes. The samples were then loaded onto the gel and run at 200 V for 40 minutes. After this, the gels were stained with Instant Blue (Generon Ltd) and were left on a rotator for 10 minutes. The stain was kept for reuse while the gel was destained with dH_2O. The protein molecular mass markers were purchased from FERMENTAS.

2.2.7 MALDI-TOF-MS Analysis

Protein samples at 0.02 mg/ml were buffer exchanged into 50 mM Tris, pH 7.5 using a Microcon centrifugal filter (3 kDa cut-off). 0.005 g of CHCA (α-Cyano-4-hydroxycinnamic acid, MALDI matrix, Sigma) was dissolved into 250 µL of 100% acetonitrile and then 25 µL of 2 % (v/v) formic acid and 225 µL of HPLC grade water were added to the mixture. 1 µL of protein and 1 µL of the matrix solution were loaded onto a 96 × 2 MALDI plate and were allowed to dry for few minutes. The protein mass spectra were acquired with a Kratos Analytical Mass Spectrometer operated in linear mode under 70,000 V accelerating voltage. The machine was calibrated (external calibration) with ProteoMass Peptide and Protein MALDI-TOF-MS Calibration Kit from Sigma including bradykinin (a fragment 1-7, human), angiotensin II (human), ACTH (a fragment 18-39, human) and insulin (bovine). In order to prepare the protein samples for mass-fingerprinting analysis, protein at 25 µg/µL was added to 0.5 µg/µL of trypsin (in 50 mM ambic buffer) and was left overnight at 37 °C. 1 µL of

the digested protein was mixed with 1 µL of CHCA and then was scanned using the same machine and parameters that were used for undigested sample but without calibration proteins.

2.2.8 Circular Dichroism Analysis

The protein sample (0.5 mg/ml) for the circular dichroism measurement was prepared by exchanging the protein buffer into 10 mM Phosphate buffer in the presence of 1mM of EDTA and 1 mM of 2-mercaptoethanol, pH 7.5 using a Microcon centrifugal filter (3 kDa cut-off). The circular dichroism scan was carried out with a spectropolarimeter (JASCO, J-810). A 2 mm quartz cuvette with 0.2 nm path length was used and the signals were obtained in the range 190-260 nm, at 20 ^{o}C. Each spectrum was the average of 15 scans.

2.2.9 NMR Spectroscopy Analysis

For labeling of the human CFTR C-terminus^{42aa}, expression was carried out using M9 minimal media supplemented with ^{15}N ammonium chloride and ^{13}C glucose (CK gas, Hampshire, U. K.). Briefly, *E. coli* containing the pET24a CFTR C-terminus was grown in Luria Bertani media to an optical density of 0.7 and then centrifuged at 4000 RPM for 15 minutes. The pellets were resuspended in minimal M9 media (as described 2.1.5.17) and shaking at 200 RPM resumed for 1 hour before induction with 1mM IPTG, as described. Purification of induced proteins was as for the non-labeled material but with NMR buffers: Sonication Buffer (500 mM NaCl, 20mM sodium phosphate pH 7.0, 0.01 % sodium azide (NaN_3), 1 mM 2-mercaptoethanol, 0.1 % tritonX-100, 100 mg/ml lysozyme). Wash Buffer (500 mM NaCl, 20mM sodium phosphate pH 7.0, 0.01 % sodium azide (NaN_3), 1 mM 2-mercaptoethanol, 50 mM imidazol). Elution Buffer (500 mM NaCl, 20mM sodium phosphate pH 7.0, 0.01 % sodium azide (NaN_3), 1 mM 2-mercaptoethanol, 250 mM imidazol). Dialysis buffer (100 mM NaCl, 20mM sodium phosphate pH 7.9, 0.01 % sodium azide (NaN_3), 1 mM 2-mercaptoethanol). The purified protein was concentrated to 5 mg/ml for the NMR analysis.

2.2.10 Histidine-tag Pull-down Assay

Talon immobilized metal affinity chromatography (IMAC) resin was equilibrated with 15 µL of CFTR buffer (DDM) (as described in 2.1.5.7) for human CFTR and buffer (LPG) (as described in 2.1.5.8) for killifish and mouse CFTR. Deca-His-tagged CFTR and non-tagged NHERF1 PDZ $1^{(+)}$ were added to Talon resin (15 µL) in a 1.5 ml Eppendorf tube. For phosphorylation experiments, CFTR (or CFTR- NHERF1 PDZ $1^{(+)}$) was mixed with 7 mM ATP, 7 nM catalytic unit of PKA (Protein kinase A) and 10 mM $MgCl_2$. The mixture was incubated for 2 hr at RT with gentle rotation. After the incubation step, the mixture was centrifuged at 13,000 rpm for 2 min at 4 °C and the supernatant collected. The Talon resin was washed 4-6 times with 15 µL of CFTR buffer (DDM or LPG). The final step was the elution of CFTR (or CFTR- NHERF1 PDZ $1^{(+)}$) with two washes with CFTR elution buffer (DDM or LPG including 400 mM Imidazol). An additional wash step was performed by the addition of 1 % of Sodium dodecylsulphate (SDS) to remove any remaining proteins that were not eluted. All wash and elution steps were performed by centrifugation at 13,000 rpm for 2 min at 4 °C. A sample (15 µL) was collected after each centrifugation step. SDS-PAGE was used to check supernatant, wash and elution samples. The gels were stained using Silver staining kit (Thermo, FERMENTAS). Human, mouse and killifish protein samples were as described in section 2.1.1.

2.2.11 Gel Filtration of NHERF1 PDZ $1^{(+)}$

A Superdex 200 column was used on an AKTA FPLC fitted with a fraction collector (Frac-950). For purifying protein by size exclusion chromatography a method was created in the Unicorn software suite (GE Healthcare) and then the column was equilibrated with gel filtration buffer (described in 2.1.5.6) before loading the respective protein onto it. Protein samples were injected into a sample loop (500 µl) in volumes no more than half the volume of loop. The flow rate was adjusted to 0.5 ml/minute and 0.5 ml fractions were collected and the desired samples were analyzed by SDS-PAGE.

2.2.12 Vapour Diffusion Method (sitting drop)

After purifying the C-terminus^{42aa} of CFTR (or C-terminus^{42aa}-NHERF1 PDZ 1), the protein was entered into 3D crystallization trials using the sitting drop technique. The wizard I and II (Emerald BioSystems) and Screen 1 and 2 (Hampton) crystallization solutions were used as reservoir solutions. The crystallization tray consisted of 96 wells, with wizard I or Screen 1 in wells 1 – 48 while wizard II or Screen 2 filled the wells 49 – 96. Crystallization solutions were prepared from crystallants, buffers and salts with different concentrations. The sitting drop vapor diffusion technique is a popular method for the crystallization of macromolecules. The principle of vapor diffusion is straightforward. A drop composed of a mixture of a protein and reservoir solution (100 nl of C-terminus of CFTR, 5 mg/ml + 100 nl of reservoir solution) is placed in vapor equilibration with the reservoir solution. Drops of C-terminus (each 2 µl) were loaded into strips that were placed into an Automatic Protein Crystallization Robot. After loading the protein and screens into 96-well plates, the trays were sealed with sticky tape and checked using a light microscope. The drop contains a lower precipitant concentration than the reservoir. To achieve equilibrium, water vapor leaves the drop and eventually ends up in the reservoir. As water leaves the drop, the sample undergoes an increase in concentration leading to an approach to supersaturation. Both the protein sample and precipitant increase in concentration as water leaves the drop for the reservoir. Equilibration is reached when the reagent concentration in the drop is approximately the same as that in the reservoir. Two plates were prepared for a single trial. One was incubated at 4 ^{o}C and the other was at 20 °C.

2.2.13 Vapour Diffusion Method (hanging drop)

The purified C-terminus^{42aa}-NHERF1 PDZ 1 complex was entered into 3D crystallization trials using the hanging drop vapor diffusion method. Six different conditions were prepared as crystallization sparse matrices. Condition1: 50 mM potassium sodium tartrate, 60 mM sodium citrate, 2M ammonium sulfate pH 5.9, 20 mM $MnCl_2$. Condition2: 50 mM potassium sodium tartrate, 60 mM sodium citrate, 2M ammonium sulfate pH 7.5, 20 mM $MnCl_2$. Condition3: 50 mM potassium sodium tartrate, 60 mM sodium citrate, 2M ammonium sulfate pH 5.9, 60 mM $MnCl_2$. Condition4: 50 mM

potassium sodium tartrate, 120 mM sodium citrate, 2M ammonium sulfate pH 5.9, 20 mM $MnCl_2$. Condition5: 50 mM potassium sodium tartrate, 60 mM sodium citrate, M ammonium sulfate pH 7.5, 20 mM $MnCl_2$. Condition6: 100 mM potassium sodium tartrate, 120 mM sodium citrate, 2M ammonium sulfate pH 6.5, 20 mM MnCl2. Drops of 5 mg/ml C^{42}-NHERF1 PDZ 1 complex (each 1 µl) were placed on a siliconized coverslip (22 × 22 mm). Drops of reservoir solutions (each 1 µl) were mixed with the protein at the ratio of 1:1. The coverslip was immediately placed onto the well of a VDX plate (Hampton research) containing 1ml of a reservoir solution and was sealed with SANTOVAC 5 oil, taking care not to trap air bubbles. The plate was incubated at 20 °C.

2.2.14 Single Particle Analysis Electron Microscopy

2.2.14.1 Glow Discharging and Negative Staining of Samples

For the study of macromolecular complexes using an electron microscope, carbon coated grids were used. These grids are hydrophobic but glow discharging is performed to increase their hydrophilicity so that protein complexes can adhere to the surface of the grid. Copper grids 400 mesh (Agar Scientific, UK) with the shiny surface facing up were put on a parafilm wrapped glass slide. Glow discharging was performed in a Cressington carbon coater at 0.1 Torr for 30 seconds. The glow discharged grids were used within half an hour after glow discharging as the hydrophilicity of the grids decreases with the passage of time. 4 %(w/v) uranyl acetate (Agar Scientific) was prepared in advance and centrifuged in a microfuge at full speed (12,000 rpm) for 5-10 minutes immediately before use. A 3 µl protein drop containing 50-100 ng of protein, two 3 µl drops of filtered Milli-Q water and one 3 µl drop of freshly prepared uranyl acetate were placed onto a parafilm sheet in advance of glow discharging. The shiny surface of the grid was touched to the protein sample for 1 minute and excess liquid was blotted onto Whatman filter paper (no. 1 or 4). The same surface of grid was then touched to the drops of Milli-Q water for 10 seconds each and excess liquid was blotted away. Finally the same surface was touched to uranyl acetate for 30-45 seconds and excess stain was removed by blotting with the Whatman filter paper.

2.2.14.2 Electron Microscopy

Images were recorded on a Jeol electron microscope JEM-1220 which has a Gatan CCD camera for recording the images. Grids were placed in the specimen holder, and inserted into the microscope. Images were recorded at different defocus values in the range 1.5µm to 2.5 µm. Exposure time was kept at 1 second. A live FFT was used to check for any astigmatism and drift and to adjust the defocus. All the images were recorded at 100 K nominal magnification with an acceleration voltage of 100 kV.

2.2.14.3 Single Particle Analysis and 3D Reconstruction

Three-dimensional structures of proteins can be obtained by image processing using the technique of single particle analysis and 3D reconstruction. Briefly, many images of the protein oriented randomly on EM grids are obtained. The images are 2D projections of the 3D structure and theoretically, if the projection angle of each particle can be determined, it is possible to obtain the 3D structure by projecting the 2D images back along their projections angles (back-projection) (Saibil, 2000; Ruprecht and Nield, 2001). Consequently, the entire analysis starts with selecting single particles of the protein (i.e. native CFTR, or CFTR-NHEF1 PDZ $1^{(+)}$ complex).

2.2.14.4 Particle Selection

The initial data sets of a few thousand particles were interactively selected using the graphical interfaces of EMAN 1.9 (Boxer). Homogenous particles that were not overlapping or in close contact with other particles were interactively selected using a box size of (48 x 48 pixel) and Å/pix = 5.2 for all of the micrographs. The choice of the box size was made in such a way as to reduce as much background noise as possible and at the same time allow sufficient space for molecular rotation and translation during averaging and reconstruction to take place. After particle selection, the contrast transfer function CTF for each micrograph in the dataset was determined using the program ctfit from EMAN package and CTF correction was applied to the selected images of the particles.

2.2.14.5 Classification of Particles

The fact that images represent different orientations means that they might be classified according to orientation similarities. This means that images with the closest similarities are averaged together, the consequences of which is that the noise at any position varies from image to image and averages out but the signal from the molecules is the same and is reinforced. This classification procedure increases the signal of the projection while random noise is greatly suppressed. The overall implication is that class averages have high signal to noise ratio compared to the raw image data and are therefore more easily interpretable. Therefore, sets of reference-free class averages were generated representing different characteristic views or orientations of single particles using the EMAN 1.9 program. A montage of different orientation classes produced by the single particle averaging was displayed to enable recognition of distinct views. 20 to 30 characteristic views of the classes representing top, side and partial views were selected and used for the reconstruction of a preliminary 3D model (Figure 16 and 17).

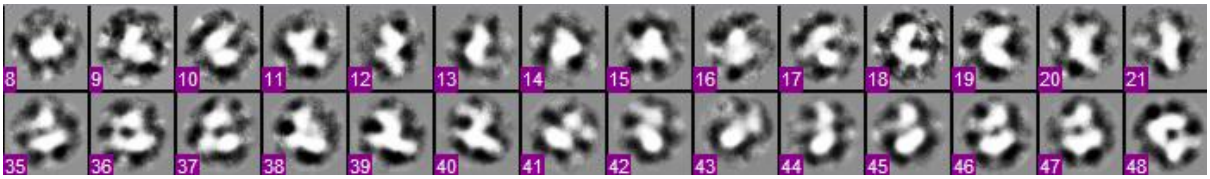


Figure 16: Typical projection classes with C2 symmetry that were generated for CFTR-NHERF1 PDZ 1$^{(+)}$ 3D structure.

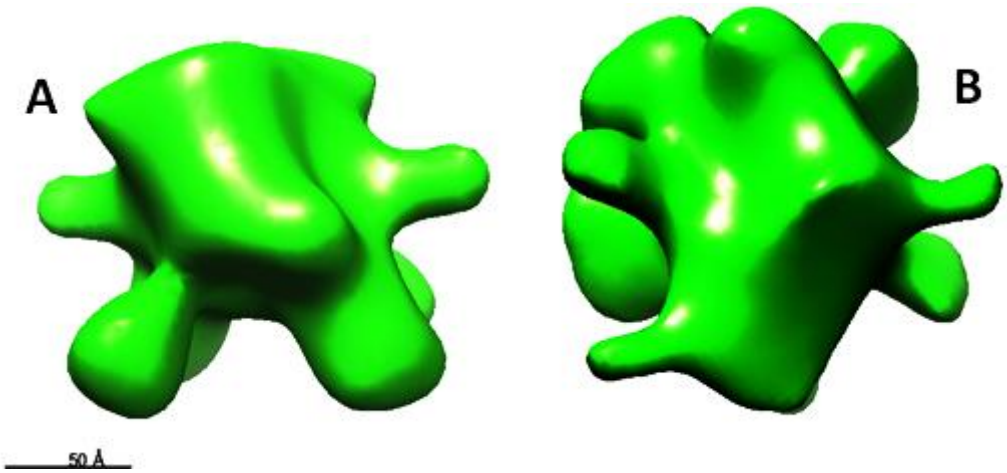


Figure 17: A side view (A) and a top view (B) of a preliminary 3D model with C2 symmetry. It was calculated from the classes in Figure 16. This preliminary 3D model was generated for CFTR-NHERF1 PDZ 1$^{(+)}$ 3D structure. The scale bar is 50 Å.

2.2.14.6 Reconstruction of a Preliminary 3D Model

Having generated sets of class averages of molecules in different orientational states, it was possible to construct preliminary 3D models of the individual sets of class averages. A preliminary 3D model was gained by back-projecting the class averages. By following established procedures in the EMAN 1.9 software suite (StartAny command), preliminary 3D models were determined from class averages that represented characteristically distinct views/orientations of the protein with C2 symmetry applied on the basis of previously published studies (Zhang et al, 2009; Zhang et al, 2011). The relative orientations of the characteristic views were determined using a Fourier common-lines routine, and the resulting averages were combined to generate the preliminary 3D model .

2.2.14.7 Refinement of the 3D Model

3D structures were refined subsequently using 6-16 rounds of iterative projection matching with each refinement evaluated by examining the convergence by comparison of the Fourier shell correlation (FSC) of the 3D model generated with the previous iteration using the EMAN 1.9 software suite. In essence, FSC compares 3D models instead of class averages. Ideally, each round of iterative projection matching improves the model as less noise would be expected as confirmed by the Fourier shell correlation. In instances where the final 3D volume was not converged fully after six rounds of iterative refinement, attempts were made to delete visually noisy images. In addition, other approaches such as tightening the mask of the selected particles, centering the selected particles were also undertaken to improve the degree of convergence between iterative cycles. The final 3D structures were low pass filtered to the resolution indicated by the shell corresponding to a correlation coefficient of 0.5 in the FSC plot calculated between two structures, each calculated from half the dataset (often termed the even-odd resolution test).

2.2.14.8 Processing commands used in SPA

After particle selection and contrast correction, a Linux operating system was used to run further processing commands. These commands were:

- *proc2D Xfix.hed start.hed hp=1 lp=20*

  The "proc2D" command is used to combine all fixed files into one file 'start.hed'. Fixed files are the selected particles that their contrast was corrected by ctfit software. X is replaced with the fixed file name. "hp" is Arctan Highpass filter (radius in Fourier pixels). "lp" is a Gaussian Lowpass filter (radius in Fourier pixels).

- *cenalignint start.hed frac=0/4 mask=20 maxshift=4*

  The "cenalignint" command is used to take a set of boxed particles and move each particle to the center of its box (+ or - 4 pixels in this example). After centering the particles it may be possible to reduce the size (in pixels) of the entire particle set. This will reduce the processing time in the refinement process. A particle that does not seem to align to a consistent center is discarded and written to 'bad.hed'. All successfully aligned images are written to 'ali.hed'. The reference image used for alignment after each iteration is stored in 'avg.hed'. This file can be examined to see if convergence really has been reached. The input file (the selected particles which were combined into 'start.hed' was aligned on four fractions (0-3) to reduce processing time. A mask (a radius in pixels) was applied to the centered images. "mask" is just larger than the expected largest axis of the particles. The "maxshift" value is the maximum shift in pixels allowed. In this example, "maxshift" specifies the 1D translation search up to 4 pixels, which actually closely related to the center of mass determination uncertainty during the boxing process.

- *startnrclasses ali.hed 200*

  The "startnrclasses" command is used to classify the centered (aligned) particles in 'ali.hed' to generate a representative set of class averages. In this example class averages composed of about 200 individual particles. The quality (centered particles with different orientations) of

the classes was checked and the good ones were selected manually. These good classes (usually 10 or 20 of these class-averages) were selected to run through the "startAny" command to generate an initial model for refinement.

- *startAny classes.nr.hed sym= c2*

  The "startAny" program uses cross common lines to determine the relative orientations of a set of class averages, then builds a 3d model from them. C2 is the symmetry used for the 3D structures based on prior studies of (Zhang et al, 2009; Zhang et al, 2011).

- *refine 16 ang=12.0 mask=19 pad=72 hard=25 classkeep=.8 classiter=8 tree=3,3 sym=c2 filt3d=14 median continue*

  "refine" was the command used for refinement. 16 cycles of refinement were performed. 12° was the angular sampling increment used to generate projections of the model. "mask=19" in refinement was tighter compared to alignment step "mask=20" and this was to reduce the noise around particles. "pad" was used to reduce Fourier artifacts during 3D reconstruction by determination an expanded box size in pixels (It is usually 25–50% larger than the normal box size). The box size was 48 and "pad" was 72. *"*hard*"* (usually 20-25) used to exclude class-averages which are not adequately consistent with the overall structure. "classkeep" of 0.7 was used to exclude "bad" particles from classaverages. A coefficient correlation of 0.5 – 3 is usually used. In this example, each particle is compared to the final average, and particles worse than 0.7 of this distribution are excluded from the final average. "classiter=8" indicated that each particle within a single class was iteratively aligned 8 times with the other particles. Values (6–8) will reduce problems with initial model bias, but will decrease resolution, so it is usual to begin with a few rounds of refinement with "classiter=8", then reduce it to 3 for final highresolution refinement. "tree" specifies faster classifications for large numbers of projections. "filt3d" applies a lowpass filter to the model after each iteration. "median" is specified when CTF correction is not being performed.

- *eotest mask=19 pad=72 hard=25 classkeep=.8 classiter=8 sym=c2 filt3d=14 3dit=2 median*

  "eotest" is used to determine the resolution of the final 3D structure. "eotest" splits each class file into even and odd halves, then proceeds to build an independent model based on each half of the data. By calculating a Fourier shell correlation between the two resulting maps, a fairly robust resolution determination can be made. "3dit" after runs several iterations of real-space model 'cleansing'.

- *proc3d threed.16a.mrc test.mrc apix=5.2 calcsf=sf.dat*

  After the estimation of resolution, "proc3d" is used to provide statistical information on the component of the structure factor (sf.dat). In addition, "proc3d" copies an input file "threed.16a" to and output file "test.mrc". "apix=5.2" is Angstrom of data per pixel of image (5.2 Angstrom/ pixel).

- *nedit sf.dat*

  "nedit" is used to perform low and high filters (based on the estimated resolution) for the final 3D structure "test.mrc" and the manually modified file was saved as "sf1.dat"

- *proc3d test.mrc test1.mrc apix=5.2 set=sf1.dat*

  The final 3D structure was filtered using the modified structure factor file and saved as "test1.mrc". This mrc file can be visualized using the Chimera software.

Some SPA commands explanation can also be obtained in more detail from (Cong and Ludtke, 2010). Further details about SPA commands can also be found in the EMAN website: http://blake.bcm.edu/emanwiki/EMAN1.

Table 11 shows a summary of SPA for the CFTR-NHERF1 $^{(+)}$ and isolated-CFTR structures.

Structure	Software	Box Size	Selected Particles	Used Particles	symmetry	Iterations
CFTR-NHERF1 PDZ $1^{(+)}$	EMAN 1.9	48	43605	35504	C2	16
Isolated-CFTR	EMAN 1.9	48	10651	9056	C2	16

Table 11: A summary of SPA for the CFTR-NHERF1 PDZ $1^{(+)}$ and isolated-CFTR structures.

The selected particles are the total of all particles used in the analysis. Used particles are the good particles that remained after alignment and classification steps. These good particles are used to generate the averaged classes and the final 3D structures. The refinement step was performed in C2 symmetry based on previous SPA studies (Awayn et al, 2005; Zhang et al, 2009). 16 cycles of refinement were performed to generate the structures.

2.2.15 CFTR Reconstitution by Freeze-Thaw

100 µl of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) chloroform stock (at 20 mg/ml) and 100 µl of *E.coli* lipid chloroform stock (at 20 mg/ml) were mixed into a 20 ml glass conical flask. The mixture was left to dry under a stream of N_2 gas for 30 minutes in fume hood. A 200 µl of (10 mM HEPES pH7.5, 50 mM NaCl and 0.01 % Sodium Azide (NaN_3)) was added to the dry materials and vortexed to obtain a cloudy mixture. The mixture was sonicated (Ultra Wave Precision Ultrasonic) in a water bath for 30 minutes at room temperature. In an Eppendorf tube (1.5 ml), CFTR at ~0.3 mg/ml in 0.05 % LPG (fraction taken from size exclusion column) was added to the mixture. As soon as the protein was added the mixture was frozen quickly on dry ice and ethanol and it was thawed quickly in a water bath at RT (Room Temperature). The freeze-thaw cycle was repeated twice. The mixture was left rotating overnight. Next day, the mixture was centrifuged at 17000 rpm (Thermo HERAEUS Fresco 17) for 20 minutes. Using a fine syringe needle (Microliter Syringe HAMILTON 100 µl), the supernatant was removed. The pellet was resuspended into 20 µl of 10 mM HEPES pH7.5, 50 mM NaCl and 0.01 % Sodium Azide (NaN_3). A sample of the reconstituted mouse and killifish CFTR samples were scanned under the electron microscopy as shown in Appendix D 5 (Figure 96).

3. Results and Discussion

The main target of this study is to examine the structure and functional properties of the cystic fibrosis transmembrane conductance regulator (CFTR). The study is divided into two main directions. In Results and Discussion 1, small fragments of CFTR (constructs) were expressed, purified, characterized and studied by biophysical methods to examine the structure and function of the protein. In the second part of this study (Results and Discussion 2), full-length CFTR (1480 amino acids) was studied by Single Particle Analysis and for the study of full-length CFTR-NHERF1 PDZ $1^{(+)}$ interactions.

3.1 Results and Discussion 1

3.1.1 Prediction of Secondary Structure of C-terminus^{42aa}

As a start point that where C-terminus of CFTR is extended, The PSIPRED secondary structure prediction method (Jones, 1999) was used to predict the secondary structure and to determine a suitable length for the desired protein. As shown in Figure 18, the last 60 amino acids of CFTR were submitted to The PSIPRED server (Bryson et al, 2005). The underlined sequences represent C-terminus^{42aa}. There are three possible secondary structure types, helical (H), strand (E), or coil (C). The prediction values range from 0 (low probability) to 9 (high probability). From the PSIPRED values, it is obvious that amino acid $I^{40\ +}$ $A^{41\ +}$ Q^{42} (**Bold**) reflect a switch region (from coil (C) to helical (H)). For this reason, this portion was selected to be the end of C-terminus^{42aa}. (This secondary structure prediction was done before the NBD2 structure appeared).

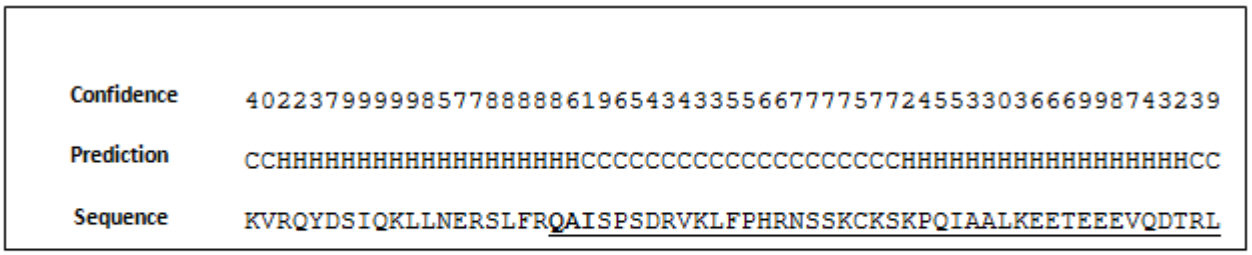


Figure 18: Secondary structure prediction values for the last 60 amino acids of CFTR using the PSIPRED secondary structure prediction method.

3.1.2 Expression, Purification, Characterization and Crystallization Trials of the C-terminus^{42aa}

A pET-24a vector including the C-Terminus^{42aa} (the last 42 amino acids of human CFTR), 6 Histidine-tag and thrombin site was synthesized by GENEART. As a routine check and to make a DNA stock, the pET-24a vector was transformed in DH5α *E.coli* cells and two colonies were grown, plasmids isolated and digested (Figure 19) by two restriction enzymes, Cla I and Nde I, as described in Materials and Methods (section 2.2.2). Linearized plasmid was also generated by adding Cla I alone. Two fragments of DNA were produced with the two enzymes, 1588 bp and 3820 bp (as shown in Figure 19, lanes 4 and 7). The linear form and the length of plasmid fragments were evidence for the presence of the correctly synthesized plasmid.

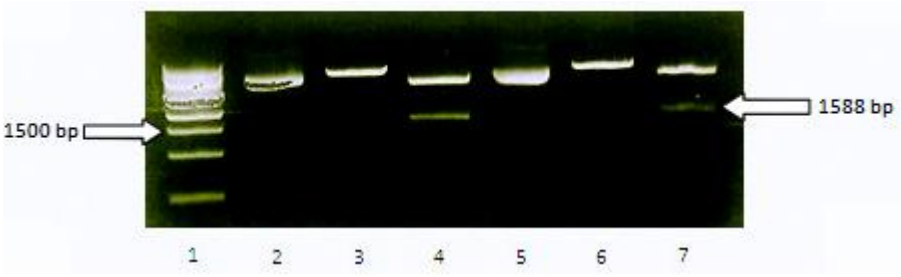


Figure 19: Double and single cut of the pET-24a vector. 1- 1 kb DNA ladder (NEW ENGLAND, Biolabs), 2- An uncut plasmid, was mini-prepped from Colony 1, 3- A single cut plasmid, from Colony 1, 4- A doubly cut plasmid, from Colony 1, 5- An uncut plasmid, was mini-prepped from Colony 2, 6- A single cut plasmid, from Colony 2, 7- A doubly cut plasmid, from Colony 2.

3.1.2.1 Expression and Purification of the C-terminus^{42aa}

The C-terminus^{42aa} of CFTR was over-expressed in BL21 (DE3) *E.coli* cells in order to obtain a sufficient yield of the protein (a few mg/ml) for crystallization and for protein-protein interaction experiments. *E.coli* cells were incubated at 37 ^{o}C until an OD600 of 0.6-0.8 was reached and then the flasks were kept on ice for 5 minutes to reduce the temperature. The shaking incubator was adjusted to 25 ^{o}C and cells were induced with 1mM of IPTG. The protein was expressed at a higher level at 25 ^{o}C compared to the expression at 37 ^{o}C. In a similar experiment, Peng *et al* (1998) observed the effect

of reduced induction temperature on protein solubility. They found that the expression of a construct of CFTR (transmembrane segments, TM3-4) was very low at 37 ^{o}C and the solubility was enhanced from 80% insoluble to less than 50% by growing the bacteria cells at 25 ^{o}C. The expression level of C-terminus^{42aa} was monitored every hour by taking a sample from the culture. Protein expression levels increased gradually from zero time to 4hr and the highest protein expression was recorded 4 hours after induction (Figure 20). Essentially all of the 6His-tagged protein was eluted at 250 mM imidazol. The protein yield was ≈ 1.5 mg/L (4L bacterial culture). A low concentration of the protein appeared in the wash buffer (contains 50 mM Imidazol). The final concentration of the protein after centrifugation filtration was ≈ 4 mg/ml with a final volume of 2 mL.

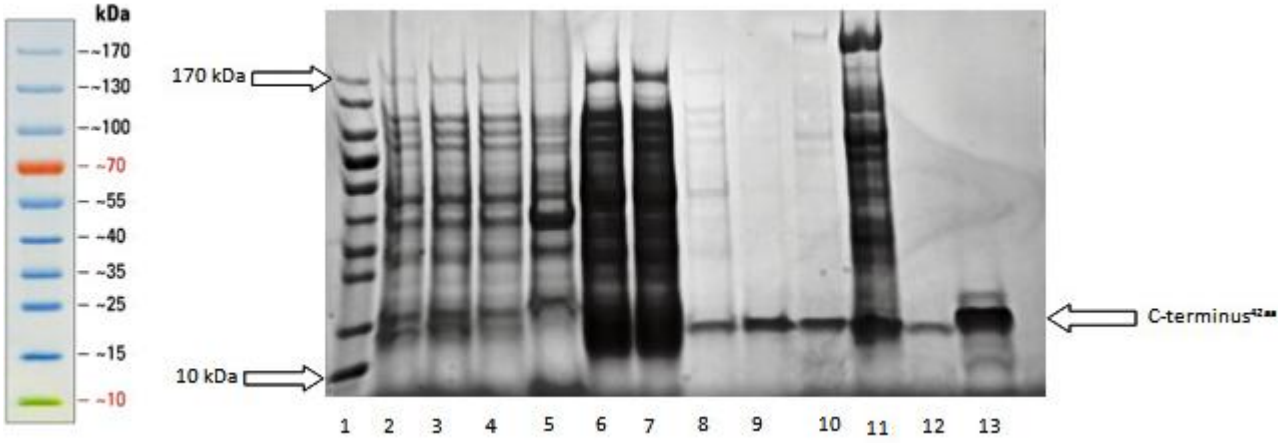


Figure 20: Expression and Purification of C-terminus^{42aa}. 1-Marker (PageRuler Prestained 4-20% Tris-glycine SDS-Page, Fermentas), 2- Uninduced cells, 3- Induced cells (4hr), 4- Induced cells (5hr), 5- Insoluble fraction after sonication, 6- Soluble fraction after sonication, 7- IMAC-Unbound material, 8- IMAC-Wash, 9- IMAC-Elution, 10- IMAC-Unbound material (after thrombin cleavage), 11- Retentate of 50 kDa concentrator, 12- Filtrate of 50 kDa concentrator, 13- Retentate of 3 kDa concentrator.

As shown in Figure 20, Lanes 2-4 show the level of expression in which protein expression was recorded 4 hours after induction indicating that 4 hours were adequate for the protein expression. Most of the proteins were obtained in the soluble fraction after sonication (Lane 6), and a low amount of proteins remained in the pellet (Lane 5). This indicates that the sonication was sufficient. Lane 7 shows all materials that did not bind to the resin. Some amount of the C-terminus^{42aa} was eluted with

50 mM Imidazol as shown in the wash sample (Lane 8). A higher amount of the C-terminus^{42aa} was eluted with 250 mM imidazol (Lane 9). Further purification (and concentration) steps were performed using 50 and 5 kDa concentrators as shown in Lanes 10-13. The purified C-terminus^{42aa} was concentrated to the concentration of ≈ 4 mg/ml (Lane 13). The C-terminus^{42aa} (≈ 5 kDa) appeared on the gel around 15 kDa. This significant shift was noticed on all gels of the protein. The shift in expected migration on SDS-PAGE may be due to some residual structure or poor SDS binding. The mass of the purified C-terminus^{42aa} was confirmed by mass spectrometry as discussed later.

3.1.2.2 Characterization of the C-terminus^{42aa}

3.1.2.2.1 MALDI-TOF-MS results indicated that the estimated molecular mass of the purified C-terminus^{42aa} was (4933.53 Dalton and 4924.17 Dalton) which were similar to the theoretical mass (4939.52 Dalton). As shown in Figure 21a, four peaks appeared in MALDI-TOF-MS data. The peak of 4924.17 Dalton has a small intensity and was probably less significant. The peak of 2461.63 was the doubly charged (M^{2+}) ion of the 4924 Dalton peak. Yergey et al (1999) reported that MALDI-TOF has the ability to doubly charge ubiquitin. The preparation of sample for MALDI analysis has an effect on the results (Gunawardena *et al*, 2005). The protein sample seems to be affected by CHCA ionization energy that can produce a higher yield of multiply charged protein ions (Frankevich *et al*, 2003). The peak of 678.92 Dalton (and its doubly charge peak, 337.87) was probably a fragment of the C-terminus^{42aa}. It is known that matrices for MALDI-TOF can cause the appearance of protein fragmentation in MALDI spectra (Karas et al, 1995). In general, MALDI matrices divide into two categories: Firstly, a soft matrix which has a minimal internal energy deposition and is essential for molecular mass determination. Secondly, a hard matrix which can be applied for structural studies, and mainly for peptide sequencing (fingerprinting) (Schulz *et al*, 2006).

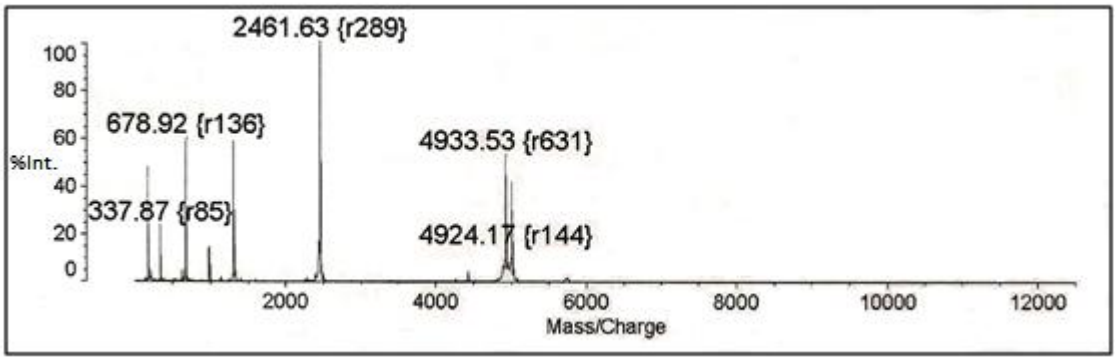


Figure 21a: MALDI-TOF mass spectrum of the C-terminus[42aa]. r is resolution. In Mass Spectrometry, resolution is the ability of instrument to distinguish two peaks of slightly different mass-to-charge ratios. The formula used to determine the resolution is: Resolution = $\frac{M/z}{\Delta M}$, where M/Z is a mass-to-charge ratio. ΔM is the width of the peak measured at a 50% (the full width at half maximum or FWHM). (Marshall and Hendrickson, 2008).

A further experiment for the identification of the C-terminus[42aa] was performed using Quadrupole Time-of-Flight (Q-TOF). The data was collected by the Protein Mass Spectrometry Core Facility, Michael Smith Building, University of Manchester. The result of Q-TOF (Figure 21b) showed one single peak (4939.565 Da) which was identical to the theoretical mass (4939.52 Dalton). Q-TOF gave better results compared to MALDI-TOF and this is expected for several reasons. First, Q-TOF is coupled to High-performance liquid chromatography (HPLC) units which increase the purity of the sample and its resolution as a result. Secondly, Q-TOF has quadrupole mass filters that allow specific ions to pass through. Finally, the ionization method (Electrospray ionization) used in Q-TOF does not cause fragmentation of the protein sample (as occurs by MALDI ionization) which results in higher resolution (Loboda et al, 2000; Chernushevich et al, 2001).

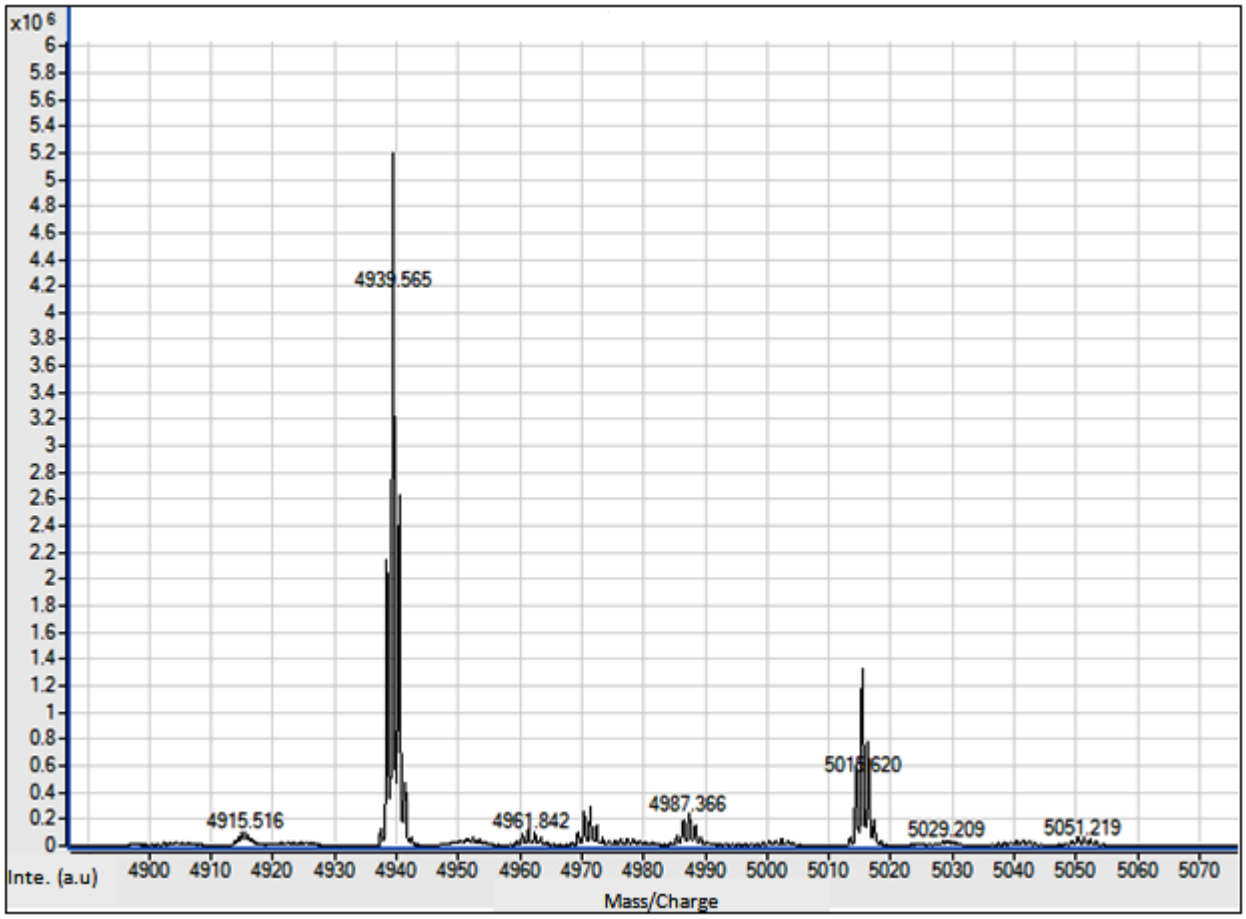


Figure 21b: Q-TOF mass spectrum of the C-terminus[42aa]. Q-TOF experiment was performed for the co-expressed C-terminus[42aa]-NHERF1 PDZ 1. The data of the C-terminus[42aa] is only presented in this graph in order to compare the peak accuracy between MALDI-TOF and Q-TOF. 4939.565 is the measured peak of the C-terminus[42aa] which is identical to the calculate peak of 4939.52 Dalton. The other peaks with the masses (4915.516, 4961.842, 4987.366, 5029.209 and 5051.219) may be fragmentations of the C-terminus[42aa]. The peak of 5015.620 gives a molecular weight difference of 76 Da and this was noticed in Mass Spectrometry experiments of NHERF1 PDZ 1-C-terminus[42aa] complex (as mentioned in Figure 28). (a.u) is an arbitrary unit.

MALDI-TOF-MS (fingerprinting) was performed for the C-terminus^{42aa}. The theoretical and estimated masses of digested peptides are shown in Table 12. MS-Digest, an online software package was used to determine the theoretical values of the digested C-terminus^{42aa}. 100 % of the protein sequence was identified by comparison between theoretical and experimental values. No MALDI signal was matched with the theoretical values of the linker peptide (thrombin site) and the 6 His-tag which confirmed the efficiency of the thrombin cleavage. Stronger MALDI signals were acquired with arginine-terminated peptides than for lysine-terminated peptides as shown in Figure 22. This may be due to the higher proton affinity of the guanidine group in arginine (Brancia *et al*, 2000; Ahn *et al*, 2007).

m/z (mi)	m/z (av)	MALDI-TOF data	C-terminus^{42aa} Sequence
955.5935	956.1803	955.466	(K)SKPQIAAL
1017.4960	1018.0775	1017.365	(R)GSQAISPSDRV
1085.5851	1086.2462	1082.359	(K)LFPHRNSSKC
1364.5812	1365.3580	1364.401	(K)EETEEEVQDTRL

Table 12: Identification of the C-terminus^{42aa} by mass spectrometry analysis.

Monoisotopic (mi): only the lowest common isotope for each element is used in the mass calculations 12C, 1H, 14N, 16O, 32S, and 31P. Average (av): All isotopes for each element are used and with their abundances reflecting their "normal" proportion in the biosphere. m/z is mass/charge.

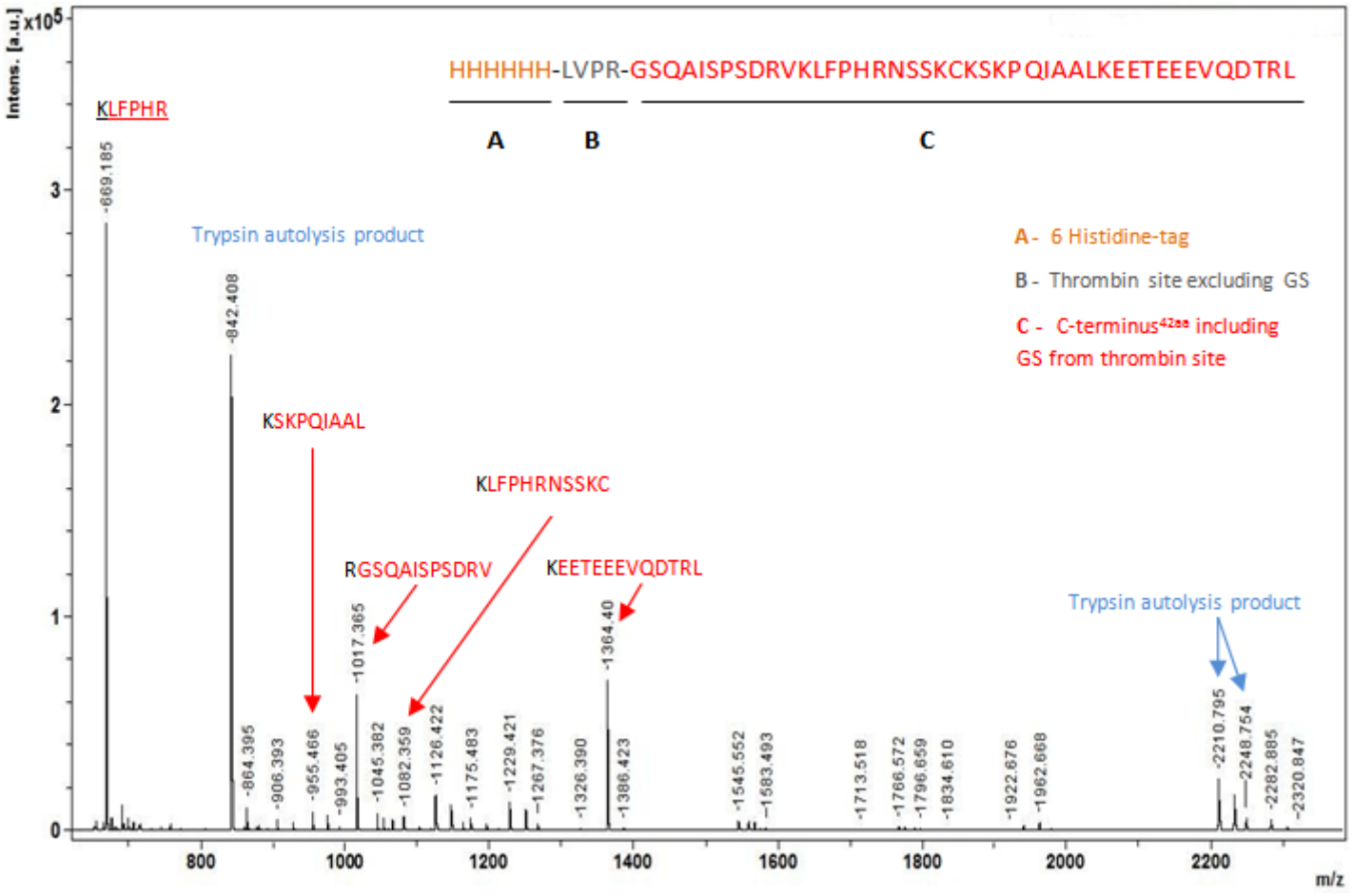


Figure 22: Identification of the digested C-terminus^{42aa} by MALDI-TOF fingerprinting analysis. The peak of 955.466 Da matched peptide (KSKPQIAAL). The peak of 1017.365 Da matched peptide (RGSQAISPSDRV). The peak of 1082.359 Da matched peptide (KLFPHRNSSKC). The peak of 1364.40 Da matched peptide (KEETEEEVQDTRL). The peak of 669.185 Da is a repeated fragment of the protein. (a.u) is an arbitrary unit.

3.1.2.2.2 Circular Dichroism Measurement of the C-terminus[42aa] revealed that there was no obvious folded state of the protein (Figure 23). The CD spectrum of C-terminus[42aa] from 190 to 250 nm shows a weak negative signal at 208–210 and a weak positive signal at 190–193 nm, indicating a transition state between random-coil and α-helical proteins. The random coil nature of the peptide could be due to insufficient length to form a stable secondary or tertiary structure; hence this result suggested the need for increasing the length of the protein. This can result in enhancing the folding by adding required amino acids for providing secondary structural elements.

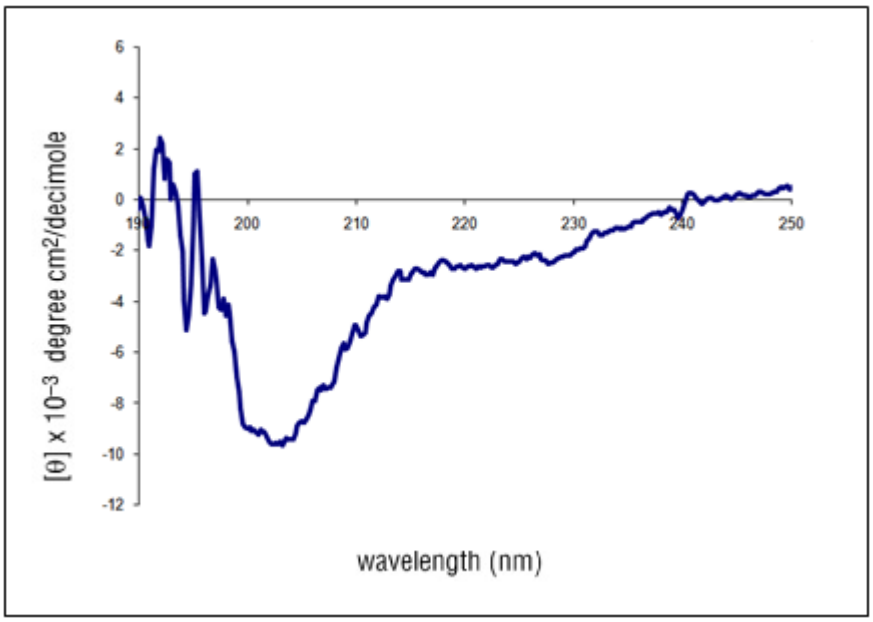


Figure 23: Far-UV CD spectrum of the C-terminus[42aa].

3.1.2.2.3 NMR Data

A sample of the C-terminus[42aa] was prepared as described in section (2.2.9) and scanned using 600 MHz Bruker DRX four-channel liquid-state spectrometer (NMR Facility, the university of Manchester). The NMR scan and analysis were kindly performed by Dr Richard Tunnicliffe.

As shown in Figure 24 revealed that most of the C-terminus[42aa] residues were partially/transiently structured as they acquired positive NOE relaxation values (~ 0.15 to 0.5). The protein residues (S8-F18, dashed box) seems the most structured section. N-terminal and C-terminal ends of the peptide

gave negative NOE relaxation values and this reflects very mobile regions. NOE negative values are likely for unstructured random coil regions (Ivancic et al, 2005). NMR data of the C-terminus[42aa] shows a relative harmony with The PSIPRED secondary structure prediction that was previously presented in this study. The two floppy ends of the protein indicate a possible reason for unsuccessful crystallization trials and the need of a further sequence extension.

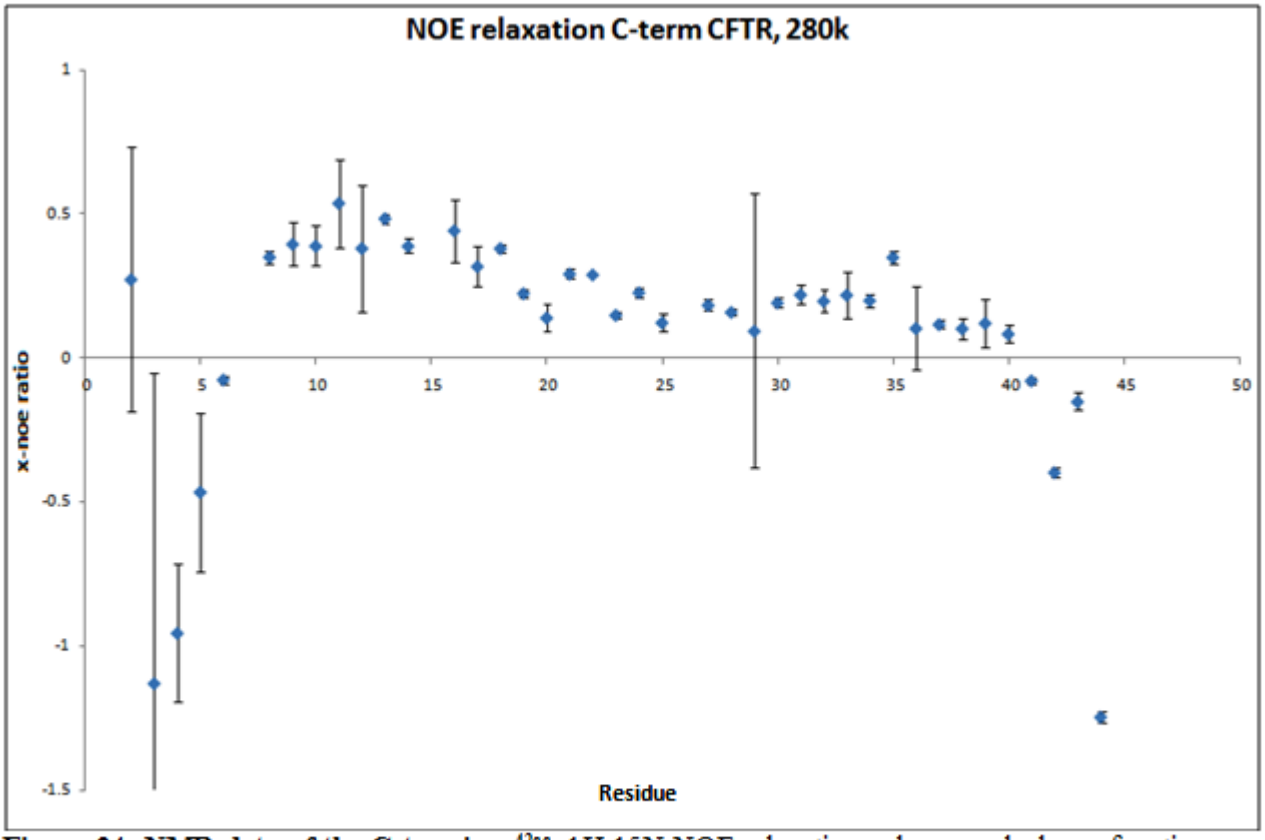


Figure 24: NMR data of the C-terminus[42aa]. 1H-15N-NOE relaxation values graphed as a function of residue number.

Several crystallization trials of C-terminus[42aa] were setup using the Wizard Classic I and II (random sparse matrix screens, Emerald Bio) and the Crystal Screen and Crystal Screen 2 (Hampton Research) but no crystals were obtained.

3.1.3 Co-expression, Co-purification, and Characterization of the C-terminus^{42aa} of Human CFTR and NHERF1 PDZ 1

3.1.3.1 Sub-cloning of NHERF1 PDZ 1 into pET-52b

Two primers (5` CGTGGATCGTCTACCATGGGGCTGCCCCGG 3`) and (5`GAATTGTCAGGATCCTCATTACAGCAGAGAGTCGTTC 3`) were designed for sub-cloning NHERF1 PDZ 1 from a pGEX-2TJL1 plasmid into a new plasmid (pET-52b). Two restriction enzyme sites were chosen, Nco I (CCATGG) and BamH I (GGATCC). One extra amino acid (GLY) was added to the sequence as a cloning artefact (MGL11-L99, 91 amino acids). A successful PCR reaction was able to give rise to a 236 bp band. Matching restriction enzymes were used to cut the vector in order to create compatible sticky ends with the insert sticky ends. Identical amounts and materials of Insert restriction digestion mixture were used. After preparing the insert and vector, the ligation of the two pieces of DNA was done with different concentrations of the insert and vector as described in (2.2.2.4). The newly ligated plasmid (pET-52b) (Figure 25) including untagged NHERF1 PDZ 1 was confirmed by sequencing.

3.1.3.2 Co-expression and Co-purification of the C-terminus^{42aa} and NHERF1 PDZ 1

A pGEX-2TJL1 vector containing NHERF1 PDZ $1^{(+)}$ was kindly provided by Dr Karthikeyan (Karthikeyan et al, 2001). NHERF1 PDZ 1 was sub-cloned into a pET-52b vector (No tag). A pET-24a vector encoding the 6His-tagged C-terminus^{42aa} of human CFTR and A pET-52b vector were co-expressed in *E.coli* BL21 (DE3) cells for 4 hr at 37 ^{o}C. The cells were grown in 4 liters of LB broth and the two proteins were co-expressed and co-purified as descibed previously with a difference that the Talon/supernatant mix was left at 4 ^{o}C for 5 hr instead of 1 hr to allow more time for Talon/6His binding. The further incubation period of Talon/supernatant increased the protein yield by a factor of 4. The incubation of supernatant with Talon for 5 hr gave ≈ 0.5 mg/L compared to 1 hr incubation period. SDS-PAGE (Figure 26) was used to determine the expression and purification levels.

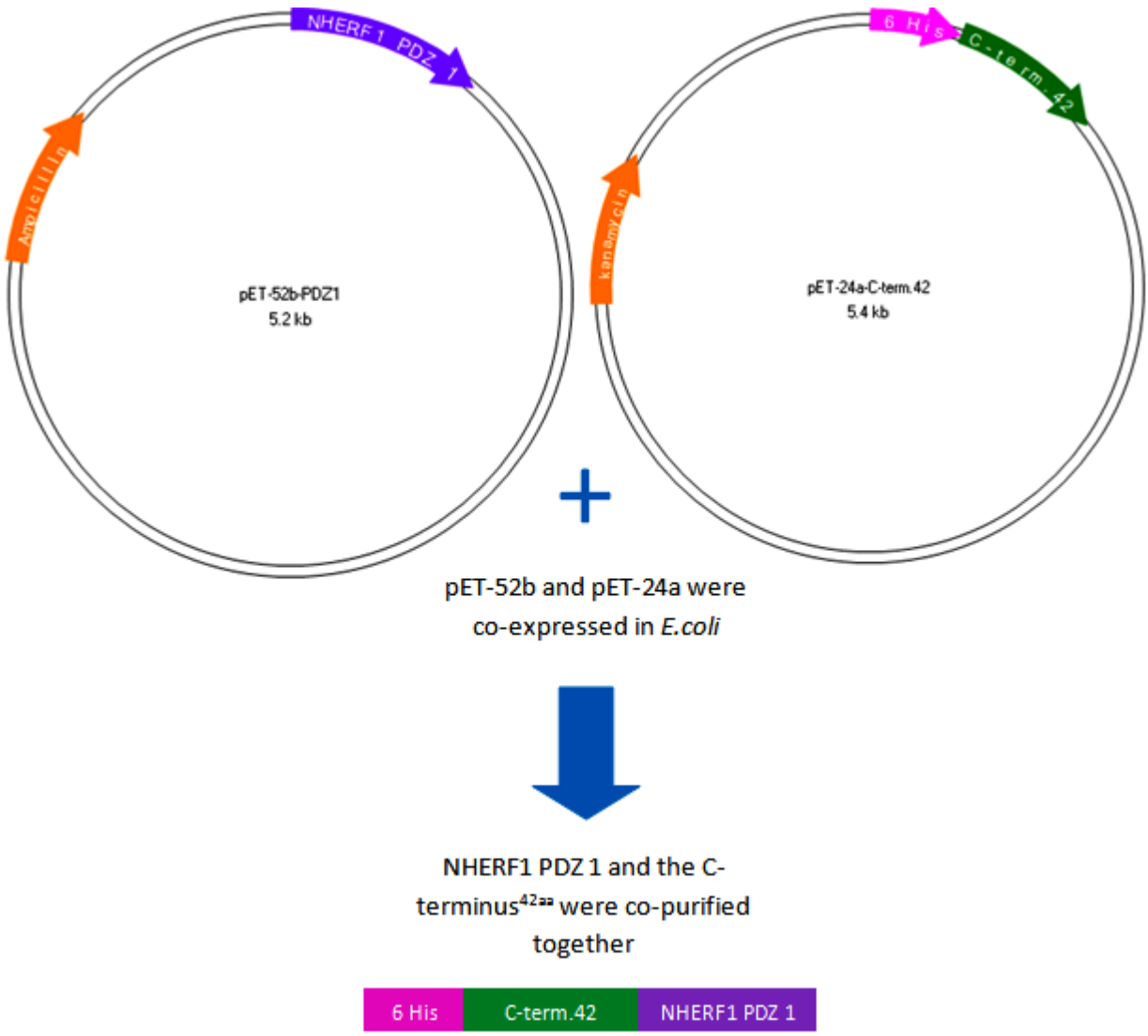


Figure 25: Strategy used to co-express and co-purify pET-25b and pET-24a. Due to the strong interaction between NHERF1 PDZ 1 and DTRL of the C-terminus^{42aa}, un-tagged NHERF1 PDZ 1 was purified from the cell lysate by binding to the 6His-tagged C-terminus^{42aa}.

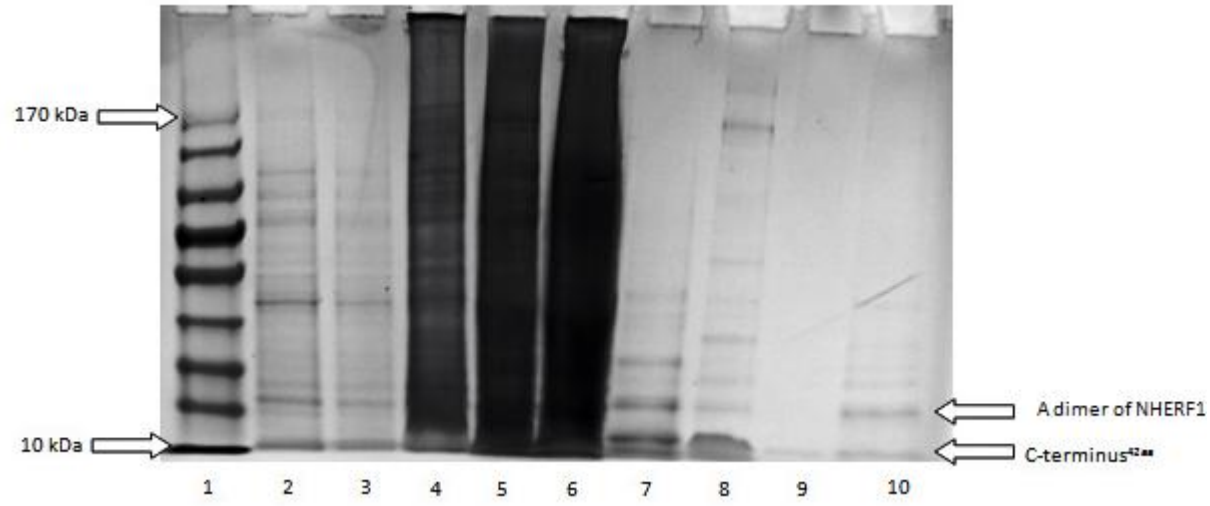


Figure 26: Co-expression and co-purification of the NHERF1 PDZ1-C-terminus^{42aa}. 1-Marker (PageRuler Prestained, 4-20% Tris-glycine SDS-PAGE, Fermentas), 2- Uninduced cells, 3- Induced cells (4hr), 4- Insoluble fraction after sonication, 5- Soluble fraction after sonication, 6- IMAC-Unbound material, 7- IMAC-Elution, 8- IMAC-Unbound material (after thrombin cleavage), 9- IMAC-Wash (after thrombin cleavage), 10-IMAC-Elution (after thrombin cleavage). It is clear that both proteins the C-terminus^{42aa} (≈ 5 kDa) and NHERF1 PDZ 1 (≈ 11 kDa) are migrating at positions that are inconsistent with their expected masses which is not unusual for low molecular mass proteins using Tris-Glycine gels.

A sample of the purified NHERF1 PDZ1-C-terminus^{42aa} is shown in Figure 26, Lane 8. This sample was collected in a Falcon tube with a final volume of 11 ml. Three additives were added to the mixture, 0.01 % Sodium Azide (NaN_3), 10 mM 2-Mercaptoethanol and 2 mM EDTA. The 11 ml mixture was concentrated using 50 and 3 kDa concentrators (as described in 2.2.5) to a volume of 1 ml and a final protein concentration of ≈ 4 mg/ml. Further concentration (to 6.7 mg/ml) was performed by Centricon Centrifugal Filter Units (from Millipore) for sitting and hanging drop experiments. No crystals of the NHERF1 PDZ1-C-terminus^{42aa} complex were obtained. Appendix D 1 and 2 show some results of sitting and hanging drop trials.

3.1.3.3 Characterization of the C-terminus^{42aa} and NHERF1 PDZ 1

3.1.3.3.1 A Dimerzation of the C-terminus^{42aa} and NHERF1 PDZ 1

A heterodimeric NHERF1-C-terminus^{42aa} band appeared on the gel as shown in Figure 27, A and B. This suggests that the complex between NHERF1 PDZ 1 and C-terminus^{42aa} was very stable despite the harsh conditions (2 % SDS and 90 ^{o}C for 4 minutes). A sample of this band (Figure 27b, Lane 4) was separated from gel and sent for MALDI-TOF-MS (fingerprinting) (Protein Mass Spectrometry Core Facility, Michael Smith Building, University of Manchester). The MALDI-fingerprinting result confirmed that the gel sample has matched peptides from both proteins as shown in Table 13. Protein identifications were made using the Mascot software (version 2.2.06; produced by Matrix Science) (Perkins et al, 1999). The number of matched peptides, 4 for NHERF1 and 2-3 for CFTR is fairly acceptable because the analysis included small constructs of NHERF1 and CFTR, 9 kD and 4 kD respectively.

Identified Proteins	Number of matched peptides	Analysis Database
NHERF1	4	SwissProt (version 2011-05)
	4	UniProt (version 2011-05)
CFTR	2	SwissProt (version 2011-05)
	3	UniProt (version 2011-05)

Table 13: Identification of digested dimeric NHERF1 PDZ1-C-terminus^{42aa} by MALDI-TOF fingerprinting analysis.

Furthermore, this stable interaction between the two proteins was confirmed by MALDI-TOF-MS data of NHERF1 PDZ1-C-terminus^{42aa} (Figure 28, C). The complex survived MALDI-TOF-MS and moreover a third peak of the complex appeared.

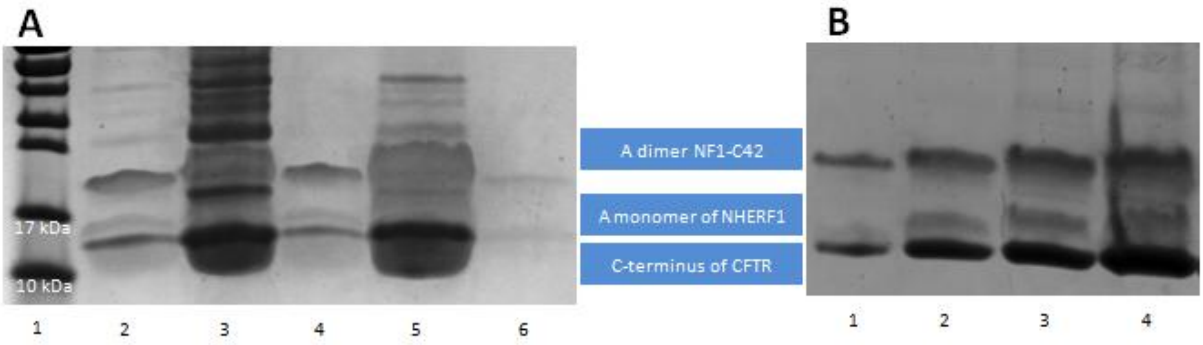


Figure 27: SDS-PAGE of the C-terminus[42aa] and NHERF1 shows the dimerization. A: 1-Marker (PageRuler Prestained, Fermentas), 2- IMAC-Unbound material (after thrombin cleavage), 3- Retentate of 50 kDa concentrator, 4- Filtrate of 50 kDa concentrator, 5- Retentate of 3 kDa concentrator, 6- Filtrate of 3 kDa concentrator. **B,** Different amounts of Retentate of 3 kDa concentrator. 1- 1 µL of Retentate of 3 kDa concentrator, 2- 2 µL of Retentate of 3 kDa concentrator, 3- 5 µL of Retentate of 3 kDa concentrator, 4- 10 µL of Retentate of 3 kDa concentrator. (Both gels, 15% polyacrylamide/Tricine gels). The samples in A were collected from a concentration step using 50 and 3 kDa concentrators. The retentate sample (Lane 5, A) was run on another gel (Figure 27b Lanes, 1-4) with different amounts loaded. A sample of this gel (Lane 4) was separated and sent for MALDI-TOF-MS (fingerprinting). The complex of NHERF1 PDZ1-C-terminus[42aa] was identified (as shown in Table 13).

3.1.3.3.2 MALDI-TOF-MS peaks of the C-terminus[42aa], NHERF1 and a complex of both proteins were obtained as illustrated in Figure 28 A,B and C. The results showed a strong interaction between the C-termnius[42aa] and NHERF1 and we hypothesized that the interaction could be regulated by phosphorylation. This hypothesis was tested theoretically by informatics and experimentally by mass spectrometry. Theoretically, NetPhose 2.0 Server (Blom et al, 1999) was used to predict serines, theronines and tryrosines that are potentially phosphorylated in the protein. A serine (position 6) on the sequence SQAI<u>S</u>PSDR of the C-termnius[42aa] gave the highest score of 0.996 and a serine (postion 38) on the sequence VEPG<u>S</u>PAEK of NHERF1 PDZ 1 gave the highest value of 0.939 (value in the range [0.000-1.000]). These data suggest that serines can be phosphorylation sites which might

essential for C-terminus^{42aa}-NHERF1 PDZ 1 interaction. In contrast, a serine residue in the β2-adrenergic receptor, a membrane protein that has a PDZ domain, was phosphorylated by GRK-5 (G-protein-coupled-receptor kinase-5) and this phosphorylation inhibited a β2-adrenergic receptor-NHERF1 interaction (Cao et al, 1999). The binding between NHERF1 PDZ 1 and DTRL of C-terminus^{42aa} and other PDZ 1 targets can be reduced by the phosphorylation of serine 77 of NHERF1 PDZ 1. This reduction of binding can be a result of a negative charge in this portion of NHERF1 (Voltz et al, 2007). Experimentally, Orbi-trap mass spectrometry was used to detect the expected phosphorylation sites on the complex of C-termnius^{42aa}-NHERF1 PDZ 1. The mass spectrometry results showed that there was no phosphorylation. These results are presented in Appendix D 3. The addition of ≈ 78 kDa that was observed in MALD-TOF (Figure 28) and was not detected in Orbi-trap mass spectrometry might be a result of modification with β-mercaptoethanol (≈ 78 kDa). This weak binding might be able to exist in the undigested complex (MALDI-TOF) but might have been detached in the fragmented sample (Orbi-trap). The absence of phosphorylation in the binding between the C-terminus^{42aa} and NHERF1 PDZ 1 is consistent with the studies of (Cao et al, 1999; Voltz et al, 2007). In addition, other studies have suggested that the phosphorylation of a serine or threonine in the C-terminus (PDZ motif) can prevent its interaction with the PDZ domain (Cohen et al, 1996; Chung et al, 2004).

Interestingly, the expected interaction between β-mercaptoethanol and the C-terminus^{42aa}-NHERF1 was detected in the complex but was not observed in the purified C-terminus^{42aa} alone. This suggests that the interaction between the two proteins may induce a conformational change that allows the interaction with β-mercaptoethanol.

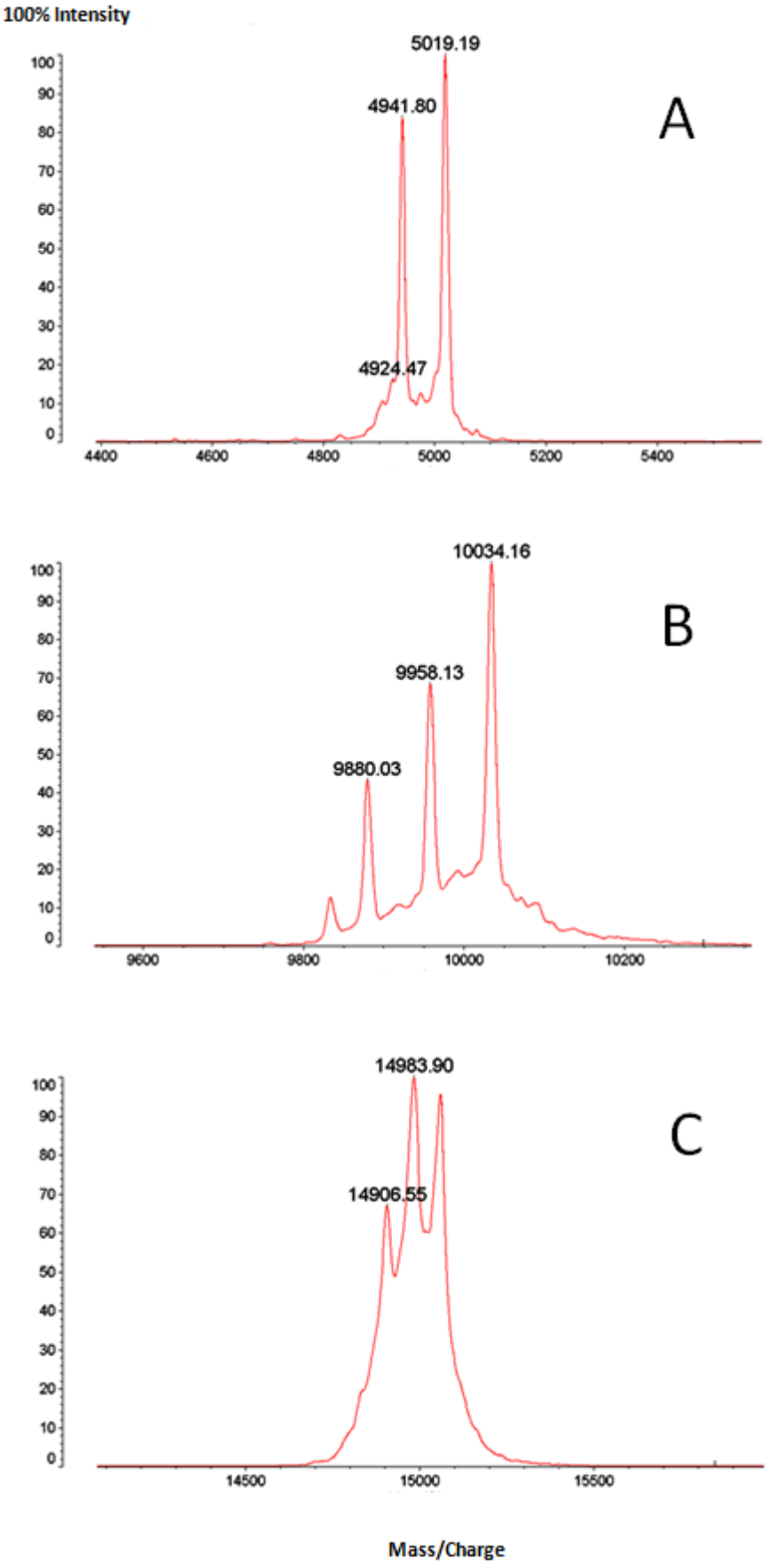


Figure 28: **A, MALDI-TOF of the band corresponding to the C-termnius^{42aa}.** The peak of 4941.80 Da matched the theoretical mass (4939.52 Da). The peak of 5019.19 Da was expected to be the molecular mass of the modified C-terminus (one site was predicted to be bound to one molecule of β-mercaptoethanol). The peak of 4924.47 Da has a small intensity and was discounted in this analysis.

B, MALDI-TOF of the band corresponding to NHERF1 PDZ1. The peak of 9880.03 Da matched the theoretical mass (9885.3 Da). The peak of 9958.13 Da was expected to be the molecular mass of the modified NHERF1 (one site was predicted to be bound to one molecule of β-mercaptoethanol). The peak of 10034.16 Da is expected to be the molecular mass of the doubly modified NHERF1 (plus two molecules of β-mercaptoethanol).

C, MALDI-TOF of the band corresponding to the C-terminus^{42aa}-NHERF1 PDZ1 complex. The peak of 14906.55 Da is expected to be the molecular mass of the modified C-terminus^{42aa} NHERF1 complex (one site was bound to one molecule of β-mercaptoethanol). The peak of 14983.90 Da is expected to be the molecular mass of the doubly modified C-terminus^{42aa}-NHERF1 complex (two molecules of β-mercaptoethanol).

3.1.3.3.3 Circular Dichroism of the C-terminus^{42aa} and C^{42aa}-NHERF1 PDZ 1

The secondary structure of the C-terminus^{42aa} and C-terminius^{42aa}-NHERF1 PDZ 1 complex were investigated by Circular Dichroism (CD). The protein samples were prepared as explained in (2.2.8). The CDSSTR program (Sreerama and Woody, 2000) was used to analyze the CD data. This program is available on DichroWeb (Whitmore and Wallace, 2008). A comparison between the C-terminus^{42aa} CD signal (as a control) and the C-terminus^{42aa}-NHERF1 PDZ 1 complex CD signal is illustrated in Table 14.

C^{42aa}	C^{42aa}-NHERF1 PDZ 1	Difference %	
0.06	0.05	1	**Helix**
0.49	0.35	14	**Strand**
0.18	0.20	2	**Turns**
0.27	0.40	13	**Unordered**
1	1	30	**Total**

Table 14: Circular dichroism data, showing a comparison between the C-terminus^{42aa} and C^{42aa}-NHERF1 PDZ 1 complex.

Compared to the previous CD data of C-terminus^{42aa} (section 3.1.2.2.2) which were analyzed based on the positions of negative and positive CD peaks, more details were provided by the CDSSTR program. As shown in Table 14 and Figure 29, some differences were noticed between the control (C-terminus^{42aa}) and the complex of C-terminus^{42aa} -NHERF1 PDZ 1. First, the interaction between the two proteins increased the content of turns and unordered regions in the C-terminus^{42aa} by 15 %. In addition, it is observed that there was an inverse relationship between the content of the secondary structure. The increase in Helix and Strand contents in the C-terminus^{42aa} was offset by a decrease in Helix and Strand contents in the complex. On the contrary, the decrease in Turns and Unordered contents in the C-terminus^{42aa} was offset by an increase in Turns and Unordered contents in the complex. However, these values were obtained from a comparative tool (CDSSTR program) and the accuracy of most of these programs is from 50-70%. On the other hand, a comparison between

experimental CD data for NHERF1 PDZ 1 (Sun and Mierke, 2005), Figure 15, and the CD data of the complex were to some extent compatible with 35% and 40% of β-sheet, and 5% and 16.8% of helix, respectively.

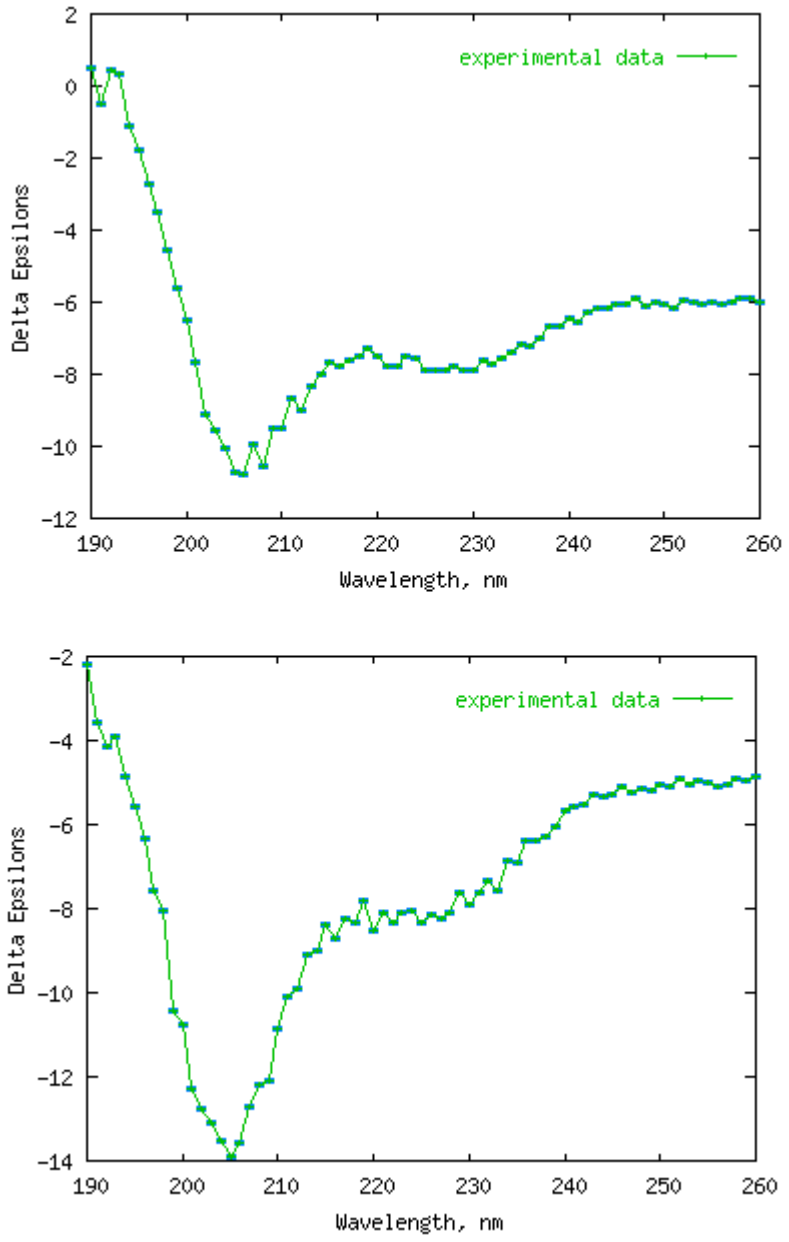
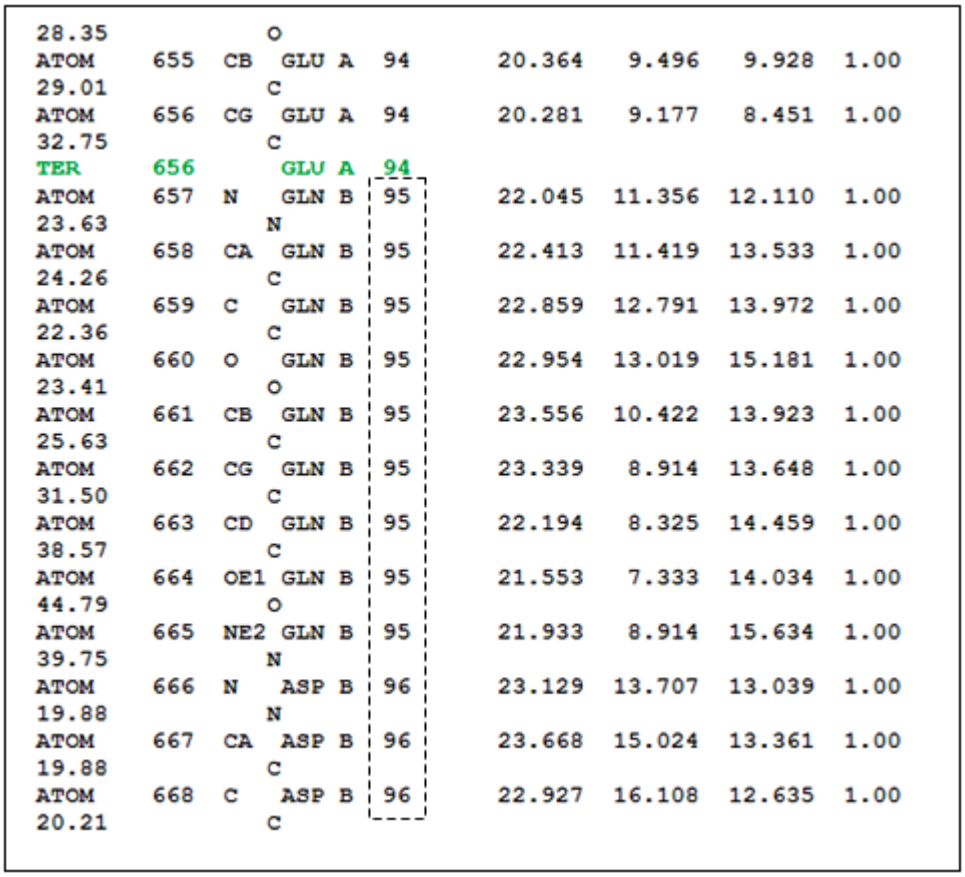


Figure 29: Circular dichroism spectra of the C-terminus^{42aa} (top) and C-terminus^{42aa}-NHERF1 PDZ 1 complex (bottom). The data was collected in 10 mM phosphate-buffered saline (PBS) at pH 7.5 and at 20 ^{o}C. These two spectra (with no corrected baseline and buffer) were obtained from the CDSSTR program. The CD raw data which was used as input with this program is presented in Appendix D 4.

3.1.3.3.4 Does the C-terminus^{42aa}-NHERF1 PDZ 1 Interaction Require More Contact Than QDTRL-NHERF1 PDZ 1 Interaction?

From SDS-PAGE (Figure 27 A and B) and MALDI-TOF results (Figure 28 C), in which C-termiuns^{42aa}- NHERF1 PDZ 1 interaction survived despite the exposure to harsh conditions, it is assumed that this strong interaction required more chemical bonds than were observed by the structure of QDTRL-NHERF1 PDZ 1 (Karthikeyan et al, 2001), i.e. that more amino acids of C-terminus^{42aa} interact with NHERF1 PDZ 1. To investigate this hypothesis, online prediction software PISA (Protein interfaces, surfaces and assemblies service) at European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel and Henrick, 2007) was used to estimate the change of free energy in the dissociation of the QDTRL-NHERF1 PDZ 1 assembled structure. The obtained free energy value for dissociation of the QDTRL-NHERF1 PDZ 1 structure was compared with the change of free energy of two alternative structures of NDSLL-NHERF1 PDZ 1 and EDSFL-NHERF1 PDZ 1 (Karthikeyan et al, 2002). The strategy used to express theses two complexes was to fuse NHERF1 PDZ 1 (11-94 amino acids) with two small peptides NDSLL (95-99) and EDSFL (95-99). The same strategy was used to express NHERF1 PDZ 1 (11-94 amino acids) with QDTRL (95-99). Because the three solved structures were fused proteins consisting of one amino acid chain for each structure, the pdb text file for every structure was modified to separate the small peptides (95-99) from NHERF1 PDZ 1 (11-94). A second amino acid chain was generated by renaming it as chain B for the last five amino acids (95-99) in every structure. Also, to separate the two chains, a 'TER' tag was added to the end of the NHERF1 sequence as shown in Figure 30 A. The three modified pdb files were applied to the PISA server. The 'Assemblies' option was selected in order to get quaternary structure information. The results are presented in Table 15.



```
28.35           O
ATOM    655  CB  GLU A  94      20.364   9.496   9.928  1.00
29.01           C
ATOM    656  CG  GLU A  94      20.281   9.177   8.451  1.00
32.75           C
TER     656      GLU A  94
ATOM    657  N   GLN B  95      22.045  11.356  12.110  1.00
23.63           N
ATOM    658  CA  GLN B  95      22.413  11.419  13.533  1.00
24.26           C
ATOM    659  C   GLN B  95      22.859  12.791  13.972  1.00
22.36           C
ATOM    660  O   GLN B  95      22.954  13.019  15.181  1.00
23.41           O
ATOM    661  CB  GLN B  95      23.556  10.422  13.923  1.00
25.63           C
ATOM    662  CG  GLN B  95      23.339   8.914  13.648  1.00
31.50           C
ATOM    663  CD  GLN B  95      22.194   8.325  14.459  1.00
38.57           C
ATOM    664  OE1 GLN B  95      21.553   7.333  14.034  1.00
44.79           O
ATOM    665  NE2 GLN B  95      21.933   8.914  15.634  1.00
39.75           N
ATOM    666  N   ASP B  96      23.129  13.707  13.039  1.00
19.88           N
ATOM    667  CA  ASP B  96      23.668  15.024  13.361  1.00
19.88           C
ATOM    668  C   ASP B  96      22.927  16.108  12.635  1.00
20.21           C
```

Figure 30 A: A section of a modified pdb file of PDZ motive-NHERF1 PDZ 1. The amino acids (95-99) were renamed with letter 'B' to create a new chain (dashed box). In addition, one tag was added (**TER 656 GLU A 94**) to terminate chain A.

Complex	ΔG^{int}, kcal/mol	ΔG^{diss}, kcal/mol	H-bonds	Sa-bridges	Ligands
QDTRL-NHERF1 PDZ 1	-13.2	0.4	14	7	2 Cl^-
EDSFL-NHERF1 PDZ 1	-14.6	1.2	12	4	2 Cl^-
NDSLL-NHERF1 PDZ 1	-34.6	4.1	3	1	4 Cl^-

Table 15: A summary of the quaternary structure information for the three complexes. ΔG^{int} indicates the solvation free energy gain upon formation of the assembly, in kcal/mol ΔG^{diss} indicates the free energy change upon dissociation, in kcal/M. H-bonds refers to the number of Hydrogen bonds between the two structures. Sa-bridges refer to the number of Salt bridges between the two structures.

PISA results (Table 15) indicated that $\Delta\mathbf{G}^{\mathbf{int}}$ and $\mathbf{\Delta G}^{\mathbf{diss}}$ of QDTRL-NHERF1 PDZ 1 were lower compared to the two other complexes despite that QDTRL-NHERF1 PDZ 1 quaternary structure has more hydrogen bonds and salt bridges. This theoretical weakness in the free energy required to form or disconnect QDTRL-NHERF1 PDZ 1 is a good indicator that the strong interaction between C-terminus[42aa] and NHERF1 PDZ 1 that we observed was stabilized by more chemical bonds or ligands. In addition, four chloride ions interacted with NHERF1 PDZ 1 (in NDSLL-NHERF1 PDZ 1) and these interactions seemed to increase $\Delta\mathbf{G}^{\mathbf{int}}$ and $\mathbf{\Delta G}^{\mathbf{diss}}$ values. This increase in the free energy is linked directly to the presence of more Cl^- ions regardless of the number of chemical bonds which were after all lower compared to the other two complexes. The three complexes mentioned in this section were solved structures with chloride ions bound. It was reported that α-amylase was not able to crystallize in the absence of chloride ions and this was because Cl^- ions were required for crystallization to maintain an ordered conformation (Aghajari et al, 2002). Interestingly, the molecular weight of a chloride ion is 35.453 Dalton (nearly half of PO_3, 78.9 Da), and hence this offers an alternative interpretation of the mass spectra for NHERF1 PDZ1 and the complex of NHERF1 PDZ 1-C-terminus[42aa]. To replace a PO_3 mass with chloride ion mass in MALDI-TOF results (Figure 28, C), 2 and 4 tightly bound Cl^- ions respectively would be required to form the two mass peaks of the complex. The effect of ions on protein stability in solution was revealed by several studies (Baldwin, 1996; Dempsey et al, 2011; Shukla et al, 2011). Chloride ion binding sites are presented in Figure 30 B. Two of the four chlorides are close to the small QDTRL peptide which may be consistent with a stabilization role.

The strong interaction observed between the C-terminus[42aa] and NHERF1 (as revealed by SDS-PAGE, Figure 27 A and B, and MALDI-TOF results, Figure 28 C) compared to the weak predicted interaction from the PISA calculation indicates that more regions of the C-terminus probably interact with NHERF1. This hypothesis agrees with LaRusch (2007) who found that the ^{1417}EENKVR region and a region between residues 1454 and 1476 of the CFTR C-terminus bind to EBP50, in additional to the DTRL-PDZ 1 binding.

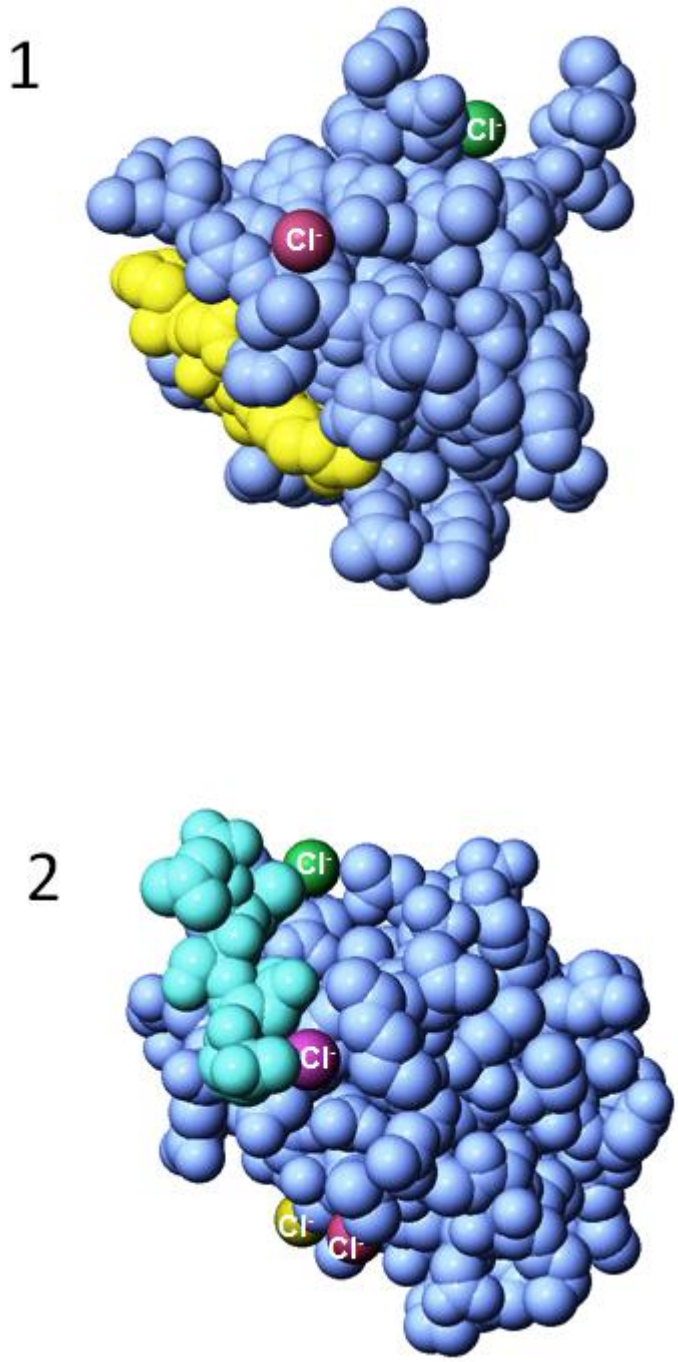


Figure 30 B: The binding sites of chloride ions (Cl^-) in the PDZ motive-NHERF1 PDZ 1. 1, the binding of Cl^- in QDTRL-NHERF1 PDZ 1 structure. NHERF1 PDZ 1 (blue). QDTRL (yellow). 2, the binding sites of chloride ions (Cl^-) in NDSLL-NHERF1 PDZ 1 structure. NHERF1 PDZ 1 (blue). NDSLL (cyan).

3.1.4 Studies of Other Constructs in terms of Expression and Purification

3.1.4.1 Expression and Purification of the C-terminus^{61aa}

A pET-24a vector including the C-terminus^{61aa} (the last 61 amino acids of human CFTR), a 6 Histidine-tag and a thrombin cleavage site was purchased from GENART. The desired sequence of the protein was the last 61 amino acids but an additional sequence (LEHHHHHH) was mistakenly added to the C-terminus^{61aa} by GENART as shown in Figure 31. Nevertheless, the protein was expressed with the additional sequence assuming that LE+6His is a small tag which has no effect of protein as reported in several studies (Carson et al, 2007). Furthermore, the sequence (LEHHHHHH) was found to be a useful portion and required for formation of the crystal structure of PH0828 protein (Tajika et al, 2004). The addition of Lys (K) after the Met (start codon) was a cloning artefact that allows high-level protein expression and a low level of methionine processing in *E. coli* (Dalbøge et al, 1990). A 10 µl sample of mini-prepped C^{61aa} was sent for sequencing and the result confirmed the presence of C^{61aa} and the additional sequence.

M	K	H	H	H	H	H	H	L	V	P	R	G	S
K	V	R	Q	Y	D	S	I	Q	K	L	L	N	E
R	S	L	F	R	Q	A	I	S	P	S	D	R	V
K	L	F	P	H	R	N	S	S	K	C	K	S	K
P	Q	I	A	A	L	K	E	E	T	E	E	E	V
Q	D	T	R	L	L	E	H	H	H	H	H	H	*

Figure 31: The amino acid sequence of the C-terminus^{61aa}.

Cloning artefact

6 Histidine tag

The thrombin cleavage site

Last 61 amino acids of CFTR

Additional amino acid sequence (The Hexahistidine tag at the C-terminal end was mistakenly added to the Geneart construct during their synthesis of the DNA)

3.1.4.1.1 Expression and Purification of C-terminus^{61aa}

A pET-24a vector encoding the C-terminus^{61aa} of Human CFTR was expressed in *E.coli* BL21 (DE3) cells by growing the cells at 37 °C until they reached an OD_{600} of 0.6-0.8, followed by induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The induced cultures were left to grow for 7 hr at 37 °C and were pelleted by centrifugation as described in 2.2.4. The yield of the protein from 2L Culture was ≈ 0.2 mg/L. The final protein concentration (≈ 0.6 mg/ml, and volume of 1.2 ml) was estimated by NanoDrop (1000 Spectrophotometer, Thermo Scientific) at wavelength of 280 nm and with an extinction coefficient of 0.149 (units: M-1 cm-1). The lower protein expression level of C-terminus^{61aa} compared to the C-reminus^{42aa} may be the result of the extended sequence length of C^{61aa}. Also, the addition of LEHHHHHH to the native protein sequence could be another explanation for the poor yield. Generally, native proteins have higher expression levels compared to mutated copies. A sample of the purified C-terminus^{61aa} was sent for MALDI-TOF and it was identified with limited degradation as shown in Figure 32. The MALDI-TOF scan was performed by the Mass Spectrometry Core Facility, Michael Smith Building, University of Manchester. No experiments were performed using this protein because it was degraded after a few cycles of freeze-thaw at -80 °C.

After thrombin cleavage (as shown in Figure 33), a significant shift appeared. This shift was not noticed in the cleaved C-terminus^{42aa} (Figure 20). This is might be an effect of the protein size. The bigger polypeptide (the C-terminus^{61aa} with 8.3 kDa) is possibly separated more accurately compared to the smaller one (the C-terminus^{42aa} with 4.9 kDa). However, both polypeptides migrated at unexpected positions on SDS-PAGE.

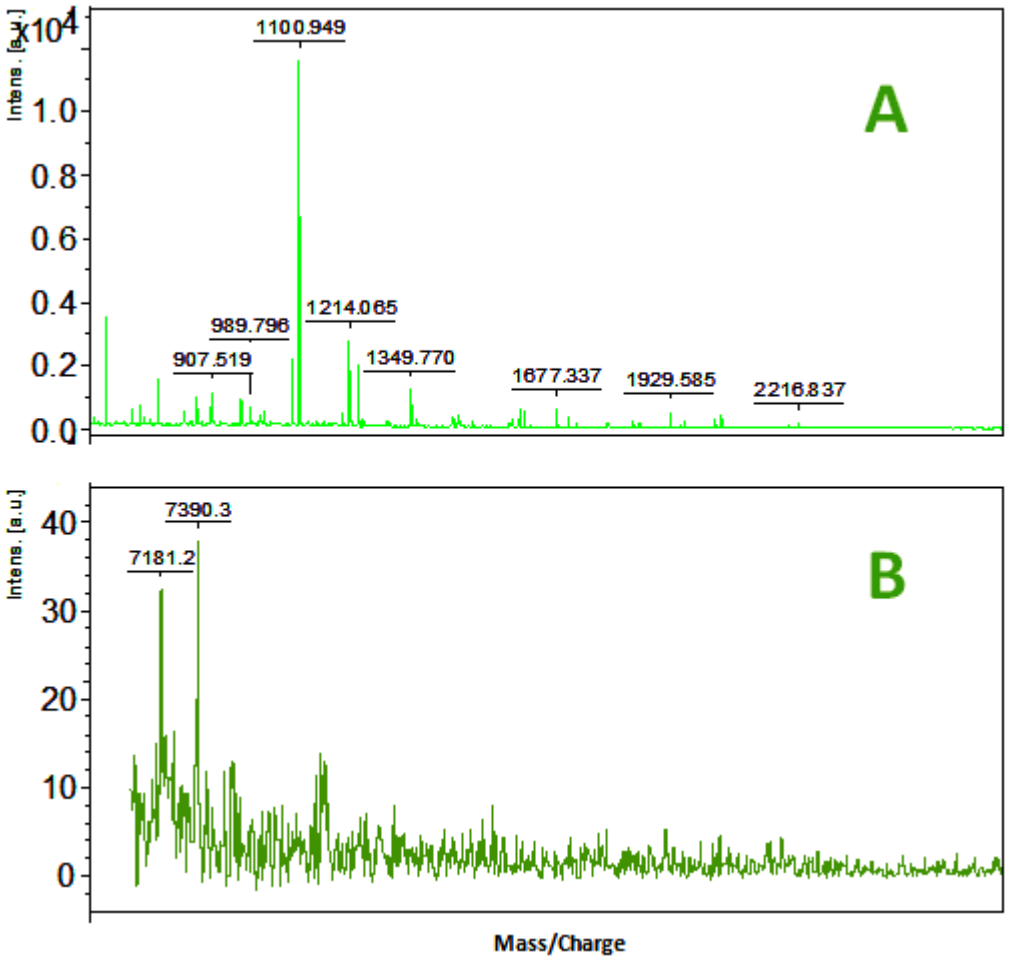


Figure 32: MALDI-TOF mass spectrum of the C-terminus[61aa]. The same C-terminus[61aa] sample was scanned 3 times using MALDI-TOF. A, the first scan which was in the range of 600-6000 Dalton. B, the third scan which was in the range of 7000-20000 Dalton. The second scan was in the range of 2000-20000 Dalton and one peak was detected with mass/charge of 7181.2 (data not shown). (a.u) is an arbitrary unit.

As shown in Figure 32, the theoretical molecular weight of C-terminus[61aa] (8380.4 Da) was roughly obtained with two fragments (989.796 + 7390.3 = 8380.09 Da) and (1214.065 + 7181.2 = 8395 Da). The addition of 15.9 Da can be a common modification where one molecule of Oxygen is added to a methionine residue. Other peaks also appeared in the scan perhaps as a result of sample degradation as mentioned before.

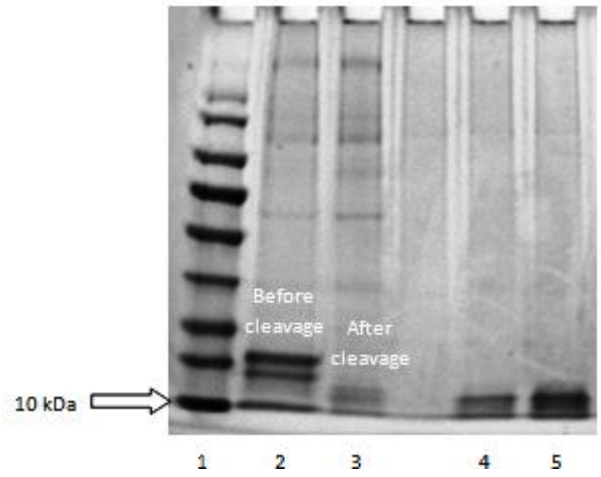


Figure 33: Thrombin cleavage of the C-terminus[61aa]. 1-Marker (PageRuler Prestained, Fermentas), 2- Elution (from previous column). 3- Unbound materials (After Thrombin cleavage), 4- 5 µL of Retentate of 50 kDa concentrator, 5- 10 µL of Retentate of 3 kDa concentrator. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

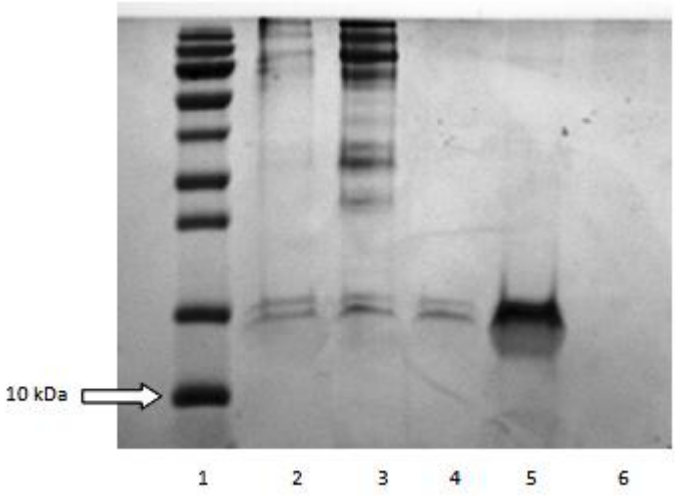


Figure 34: Concentration of the C-terminus[61aa] using 50 and 3 kDa concentrators. 1-Marker (PageRuler Prestained, Fermentas), 2- Start Sample (After thrombin cleavage). 3- Retentate of 50 kDa concentrator, 4- Filtrate of 50 kDa concentrator, some protein was lost in this step. 5- Retentate of 3 kDa concentrator indicates the high purity of the sample, 6- Filtrate of 3 kDa concentrator.

3.1.4.2 Expression and Purification of the C-terminus^{R-A}

Based on the NMR result (section 3.1.2.2) which revealed that the two sides of the C-terminus^{42aa} were floppy, two primers (described in 2.2.1.2) were designed. The primers were used for Site-directed mutation in which Arginine at position 1479 was replaced with Alanine and the enzyme restriction site (XhoI) was removed by a silent mutation for checking the construct. The aim of this mutation is the substitution of a big, flexible amino acid, Arginine with a small amino acid, Alanine, which may improve the crystallization of C-terminus^{42aa}. The mutated C-terminus is referred to herein as C-terminus$^{R-A.}$ Many studies suggest that the replacement of large flexible amino acids with smaller amino acids using site-directed mutagenesis can enhance the crystallisation (Derewenda, 2004). One of first trials of site-directed mutagenesis for crystal engineering was done by Lawson et al. (1991). They replaced Lys86, which is found in the human ferritin, with Glu, which is found in the rat homologue. This mutation created a Ca^{2+} binding connection that had previously been found to enhance crystal contacts in the rat protein. As a result, many other studies examined the effect of altering a protein's surface in related to crystallisation improvement. McElroy et al. (1992) showed that site-directed mutagenesis of the surface amino acids of thymidylate synthase changed the protein's properties for crystallisation.

3.1.4.2.1 Determination of Flexibility of Surface Amino Acids

To determine the flexible amino acids on the surface of C-terminus^{R-A}, online software was used (**PROFBval**, Avner Schlesinger & Burkhard Rost, CUBIC, Columbia University New York). https://www.predictprotein.org/. Last update (08-03-12). The prediction values are shown in Table 16. Arginine (R41) has the highest score of 6 and the highest molecular weight. Glutamine (Q1), Alanine (A2) and Leucine (L42) have equal or a higher flexibility values compared to Arg41 but their molecular weights are smaller. This was the reason why Arg41 was chosen to be mutated.

number	residue	accessibility	S	RI_s
1	Q	E	F	9
2	A	E	F	7
3	I	E	F	2
4	S	E	F	4
5	P	E	-	0
6	S	E	-	1
7	D	E	-	0
8	R	E	-	0
9	V	E	-	0
10	K	E	-	0
11	L	B	-	0
12	F	B	-	2
13	P	E	-	1
14	H	E	-	0
15	R	E	-	0
16	N	E	F	0
17	S	E	F	0
18	S	E	F	2
19	K	E	F	3
20	C	E	F	3
21	K	E	F	4
22	S	E	F	2
23	K	E	-	1
24	P	E	-	2
25	Q	E	-	2
26	I	B	-	2
27	A	E	-	0
28	A	E	-	0
29	L	E	F	0
30	K	E	F	2
31	E	E	-	1
32	E	E	F	4
33	T	E	-	2
34	E	E	F	2
35	E	E	F	0
36	E	E	-	1
37	V	B	-	0
38	Q	E	-	3
39	D	E	F	6
40	T	E	F	5
41	**R**	**E**	**F**	**6**
42	L	E	F	6

Table 16: Prediction of flexibility/rigidity from the C-terminus^{R-A} sequence using (PROFBval).

Accessibility - solvent accessibility prediction by PROFacc - 'e' marks exposed and 'b' marks buried. A residue that is predicted to be ≥16% exposed to the surface is defined to be exposed.

Reliability index - the output of the network was converted to reliability index reflecting the strength of the prediction (values 0-9). High value indicates strong prediction. RI_s marks the reliability index of the strict mode

Strict mode (S) - about a third of the residues are flexible, therefore, a stretch of residues that is predicted to be flexible might be important for the protein function.

3.1.4.2.2 Site-directed Mutagenesis of the C-terminus[R-A]

The Quick change reaction was carried out as described in (2.2.2.5). The pET-24a vector including the C-terminus of CFTR was used as a template to make the mutation. Subsequent to the PCR reaction, the product was checked by 1% Agarose gel electrophoresis as shown in Figure 35.

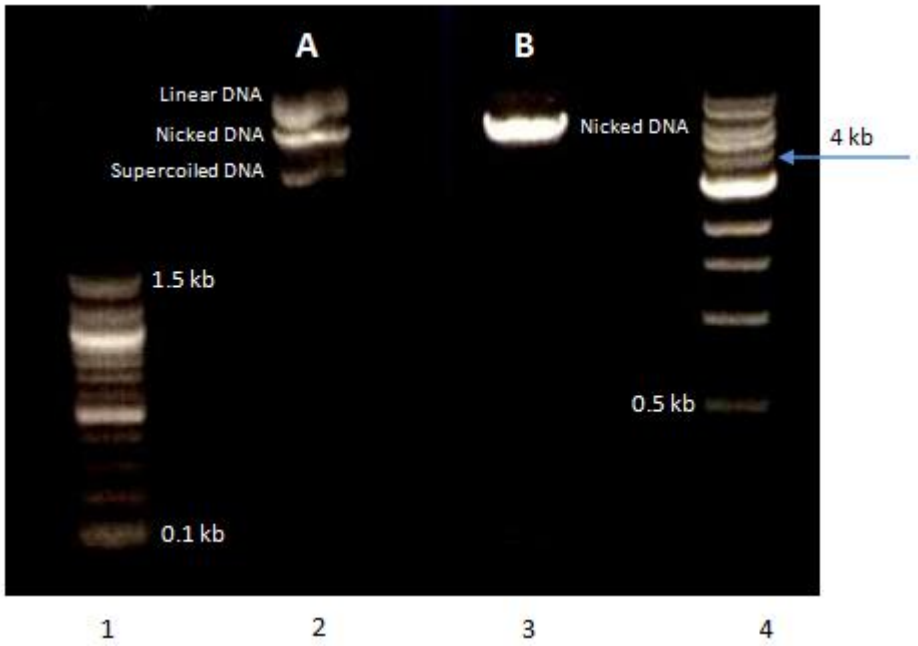


Figure 35: Check of PCR product (B) compared to the wild-type (A). 1, 100 bp DNA Ladder. 2, the wild-type. 3, the mutated plasmid containing the restriction site. 4, 1Kb DNA Ladder.

In Figure 35, track A shows 3 bands: Linear DNA, Nicked DNA and Supercoiled DNA. Track B shows Nicked DNA. It is known that if the reaction was successful, the product would be a nicked as proved on this gel. The mutated plasmid was checked using a restriction digestion enzyme (XhoI). Three mini-prepped products were restriction digested using XhoI enzyme. Because the XhoI enzyme site was destroyed, no linear band was expected to appear.

10 µl of the mutated plasmid was sent for sequencing in order to confirm the point mutation. Figure 36 shows the result of sequencing. The wild-type has been mutated from DTRL to DTAL (Boxed). Also, XhoI site has been changed from CTCGAG to CTCGGG (Blue Box)

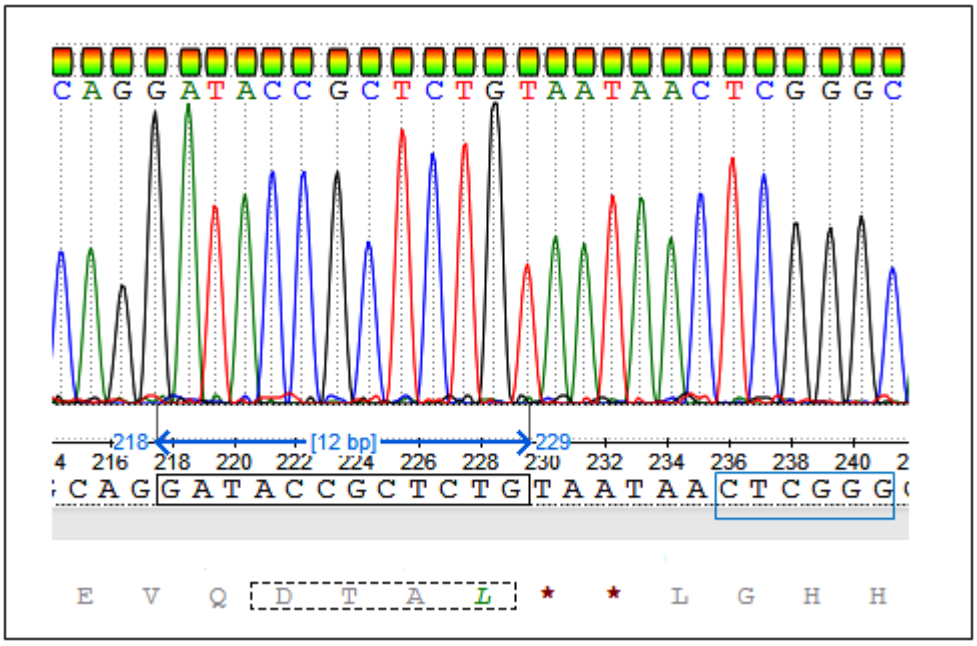


Figure 36: Confirmation of the R-A mutation by DNA sequencing.

3.1.4.2.3 Expression and Purification of the C-terminus^{R-A}

The new mutated plasmid was transferred into *E.coli* cells BL21 (DE3). A 5 L culture was prepared in order to express and purify the polypeptide as described in 2.2.4**.** The results of expression and purification are presented in Figures 37 and 38. The yield of this protein was ≈ 0.5 mg/L culture. The final concentration of the C-terminus^{R-A} was ≈ 0.59 mg/ml in 0.1 ml, i.e. a significantly lower yield than the wild-type peptide. No further work was performed using this protein because the aim of this experiment was to crystallize the protein.

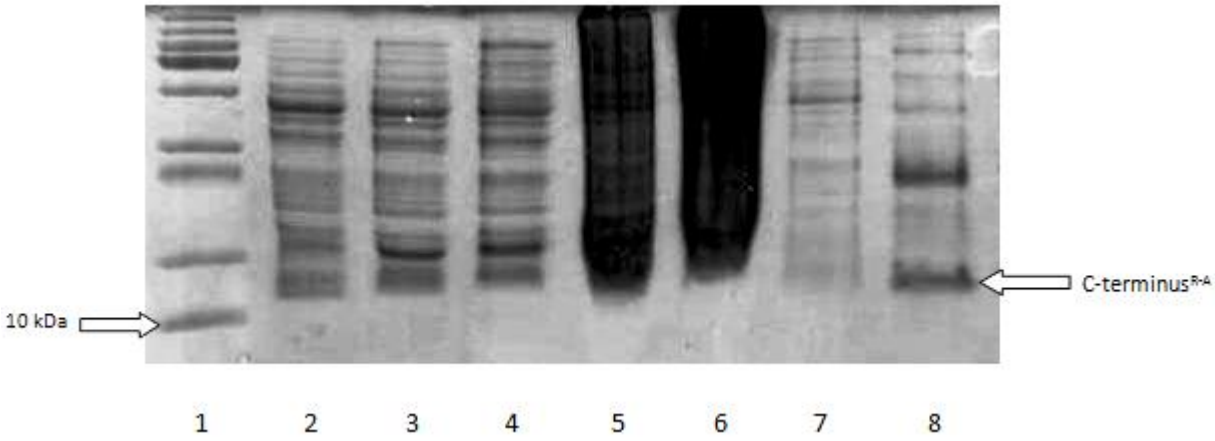


Figure 37: Expression and purification of the C-terminus^{R-A}. 1-Marker (PageRuler Prestained, Fermentas), 2- Uninduced cells, 3- Induced cells 4- Insoluble fraction after sonication, 5- Soluble fraction after sonication, 6- IMAC-Unbound material 1, 7- IMAC-Wash 1, 8- IMAC-Elution 1.

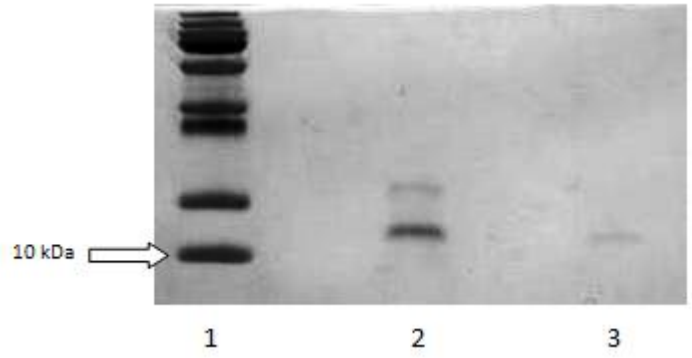


Figure 38: Concentration of the C-terminus^{R-A} using a 3 kDa concentrator. 1-Marker (PageRuler Prestained, Fermentas), 2- Retentate of 3 kDa concentrator, 3- Filtrate of 3 kDa concentrator. (lab-made, 15% polyacrylamide/Tricine gel).

3.1.4.3 Sub-cloning, Expression and Purification of Shark CFTR C-terminus

As mentioned before, the aim of the expression and purification of Shark and Killifish C-termini was to compare them with the Human C-terminus of CFTR in terms of expression, purification (or characterization as future work). The CFF (Cystic Fibrosis Foundation) 3D Structure Consortium are studying several orthologs of CFTR including human, pig, ferret, rabbit, sheep, dog, possum, platypus, chicken, mouse, rat, salmon, killifish and shark. Shark and killifish CFTR were selected because they were available for the study at an early stage. However, despite the divergence in the relationship in terms of sequence between human, shark and killifish CFTR orthologs, there are many common conserved amino acids as shown in Figure 39. The last four amino acids of the three orthologs (D/E TRL) are conserved and have importance for binding to NHERF1 as discussed in this study. This is a clear reason why the study of CFTR orthologs is important. In general, the similarities (or differences in some cases) between CFTR orthologs can lead to a better understanding of human CFTR function and structure.

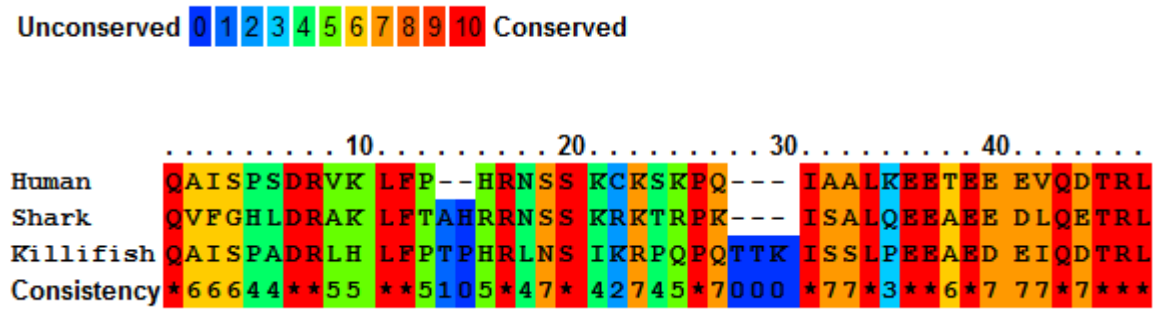


Figure 39: The conserved amino acids in the C-termini of human, shark and killifish CFTR. The conservation scoring was performed by PRALINE server (Simossis and Heringa, 2005). Zero is for the least conserved alignment positions, 10 (star) is for the most conserved alignment positions.

3.1.4.3.1 Sub-cloning of Shark C-terminus into pGA4 Vector

The two primers used to sub-clone Shark C-terminus as described in (2.1.7 and 2.2.1.3). The preparation of vector and insert were described in 2.2.2. The template for PCR was kindly provided by Professor John R Riordan (section 2.1.1). Figure 40 illustrates the cloning strategy.

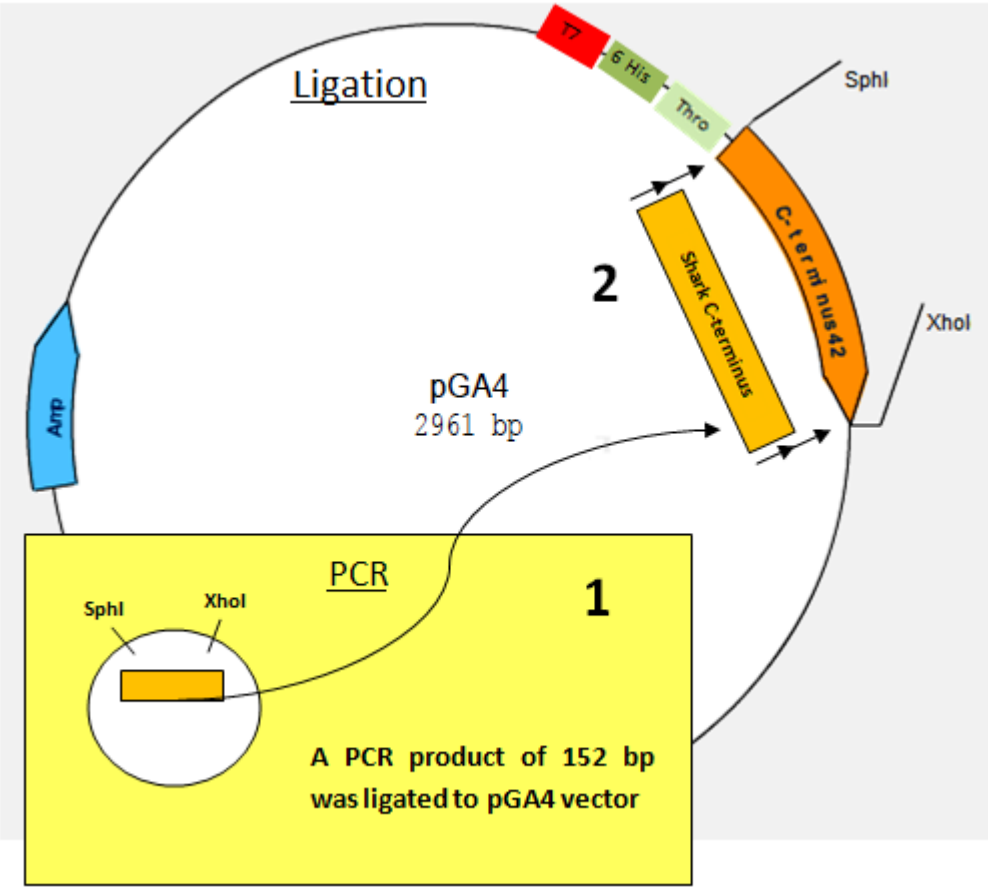


Figure 40: Strategy used to sub-clone Shark C-terminus into a pGA4 vector. Two main stages were performed. 1, PCR was used to produce the Insert. 2, the restriction enzyme digested Insert and Vector were ligated. The desired sequence (GS from thrombin site + M Cloning artefact + the last 44 amino acids of Shark CFTR, 1436-1480) in the ligated plasmid was confirmed by sequencing. The vector (pGA4) has T7 promoter, 6 His tag and Thrombin site (Thro). A pGA4 is a cloning vector but was used in this study as an expression vector.

3.1.4.3.2 Expression and Purification of Shark C-terminus

The newly ligated plasmid (with the Shark C-terminus insert) was transformed into *E.coli* cells BL21 (DE3). 3 L of culture was prepared in order to express and purify the polypeptide as described in 2.2.4**.** The results of expression and purification are presented in Figures 42 and 43. *E. coli* cells were grown at room temperature (16 ^{o}C) for 4 hr and were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The yield was very low as a result of two factors. First, the protein expression was performed at low temperature. In addition, a cloning vector (pGA4) was used for protein expression and this vector lacks regulatory sequences that act as enhancer and promoter regions for expression. Therefore, the expression and purification of shark C-terminus of CFTR (or killifish) is still under optimization and was mentioned in this thesis as a possible future work.

3.1.4.4 Sub-cloning, Expression and Purification of Killifish C-terminus

The same strategy (and vector) was used to clone the C-terminus of killifish CFTR. The expression and purification conditions were also identical. Figures 42 and 43 show a comparison between the expression and purification of the two proteins.

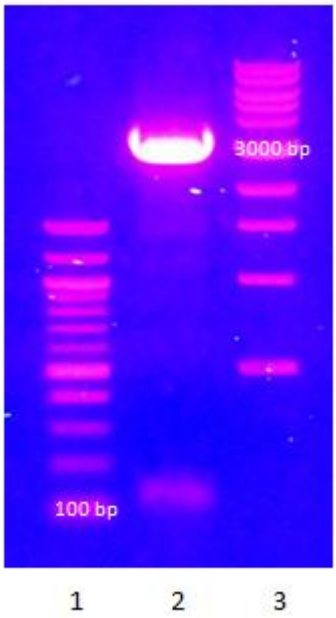
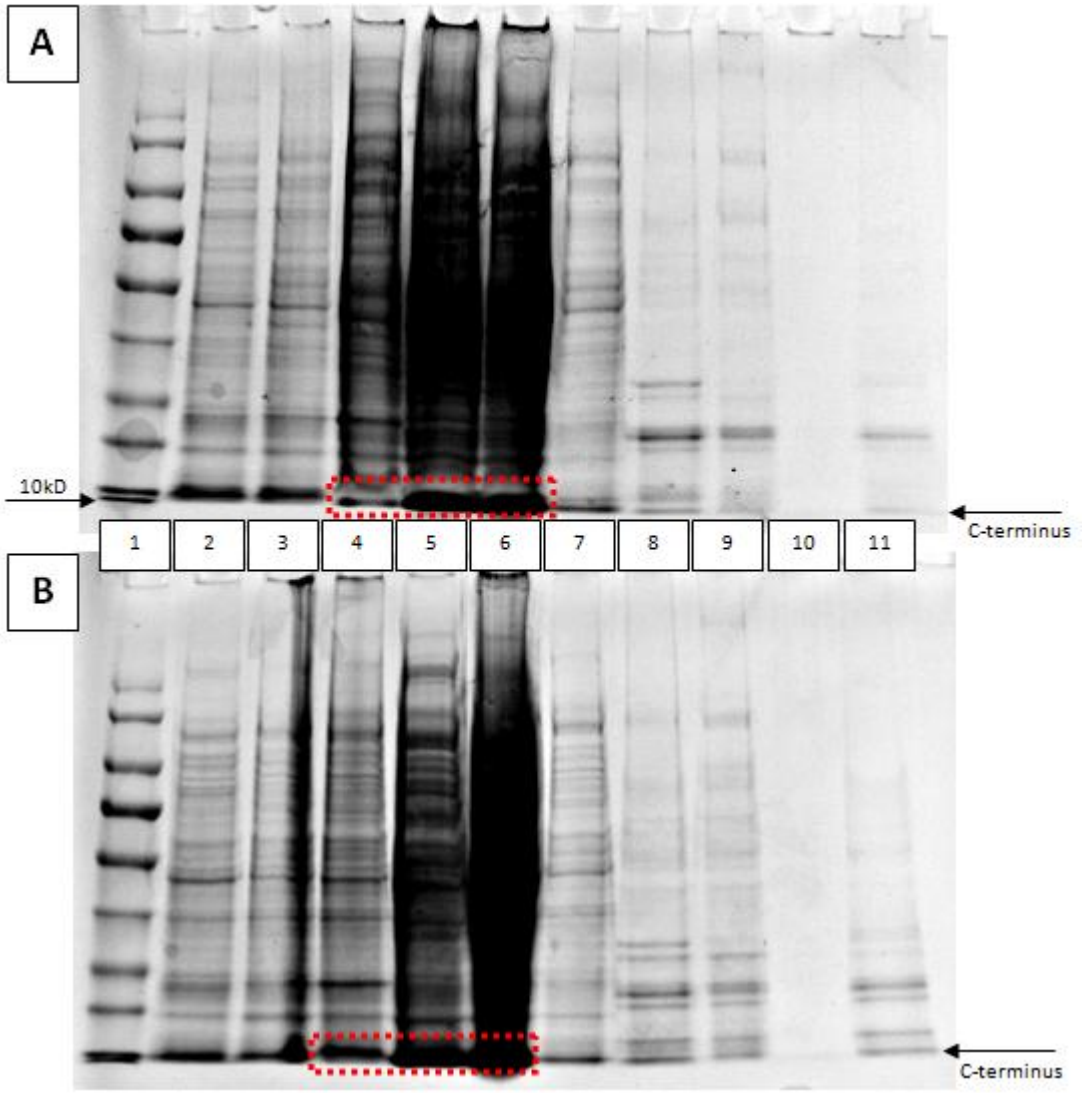


Figure 41: Check of digested vector (pGA4). The vector was digested with two restriction enzymes (SphI and XhoI). After enzyme digestion, two fragments of DNA were produced, 138 bp and 2328 bp respectively (Lane 2). 1, 100 bp DNA Ladder. 2, the digested vector. 3, 1 Kb DNA Ladder. The picture was taken by digital camera (5.00 Megapixels, Apple) before Gel Extraction to avoid additional check step.



Figures 42 (top) and 43 (bottom): Expression and purification of shark CFTR C-terminus (Top panel) and killifish CFTR C-terminus (Bottom panel). 1-Marker (PageRuler Prestained, Fermentas), 2- Uninduced cells, 3- Induced cells (4hr), 4- Insoluble fraction after sonication, 5- Soluble fraction after sonication, 6- IMAC-Unbound material (before thrombin cut), 7- IMAC-Wash (before thrombin cut), 8- IMAC-Elution (before thrombin cut), 9- IMAC-Unbound material (after thrombin cleavage), 10- IMAC-Wash (after thrombin cut), 11- IMAC-Elution (after thrombin cut). There is a band under the 10 kDa marker and it is thought to be an excess loading stain that usually appears in the bottom of SDS-page. The red boxes indicate the positions of the C-terminus on the gels.

As shown in Figure 42, the yield of both proteins was low. Killifish C-terminus showed a higher expression level as judged by the Elution sample (lane 8) despite the formation of inclusion bodies (lane 4). These expression and purification trials were performed to examine the yield and solubility for both proteins and it was concluded that higher expression temperature (25 ^{o}C or 37 ^{o}C are needed to improve the solubility and protein yield.

3.1.5 Expression and Purification of NHERF1 PDZ 1$^{(+)}$

A GST-tag vector (pGEX-2TJL1) containing NHERF1 PDZ 1$^{(+)}$ was kindly provided by Dr Karthikeyan (Karthikeyan et al, 2001) (A list of given vectors can be found in Apnedex C). This vector was expressed in BL21 (DE3) *E.coli* cells in order to obtain a sufficient yield of the protein for protein-protein interaction experiments (pull-down assay) and Single Particle Analysis. The protein expression was performed as described in 2.2.4.2 with a difference that 0.1 mM IPTG was used for induction. The protein purification was performed as described in 2.2.4.3 but with the difference that Glutathione resin (Clontech) was used for protein binding. Also different buffers were used for sonication, wash and elution as described in 2.1.5.5. Furthermore, after the collection of IMAC-unbound material and Wash (before thrombin cut), the protein was not subjected to the elution step and it was cleaved from its GST-tag by thrombin and the protein was collected in 10 ml of thrombin buffer (section 2.1.5.10). This step was required because in a previous purification the GST-tag was separated from the resin and appeared in an elution sample. After purification, the protein was concentrated as described in 2.2.5. And then the protein sample was left at 4 ^{o}C overnight for thrombin cleavage. SDS-PAGE (Figure 44 and 45) was used to determine the expression and purification levels. A further purification step was performed by gel filtration (Figure 47) as described in 2.2.11. The yield of the protein was estimated to be ≈ 0.2 mg/L (bacterial culture). The thrombin cleavage efficiency was about 50 % as estimated from Figure 44 Lane 10. The final concentration of the protein was ≈ 1 mg/ml in 100 µl.

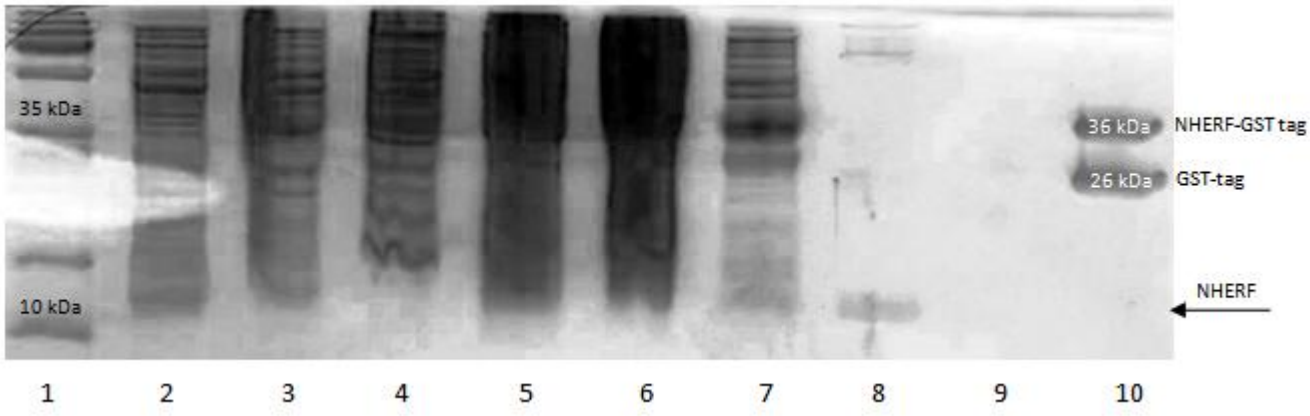


Figure 44: Expression and purification of NHERF1 PDZ $1^{(+)}$. 1-Marker (PageRuler Plus Prestained, Fermentas), 2- Uninduced cells, 3- Induced cells (4hr), 4- Insoluble fraction after sonication, 5- Soluble fraction after sonication, 6- IMAC-Unbound material (before thrombin cut), 7- IMAC-Wash (before thrombin cut). 8- IMAC-Unbound material (after thrombin cut), 9- IMAC-Wash (after thrombin cut). 10- IMAC-Elution (after thrombin cut). (Lab- made 15% polyacrylamide/Tricine gel).

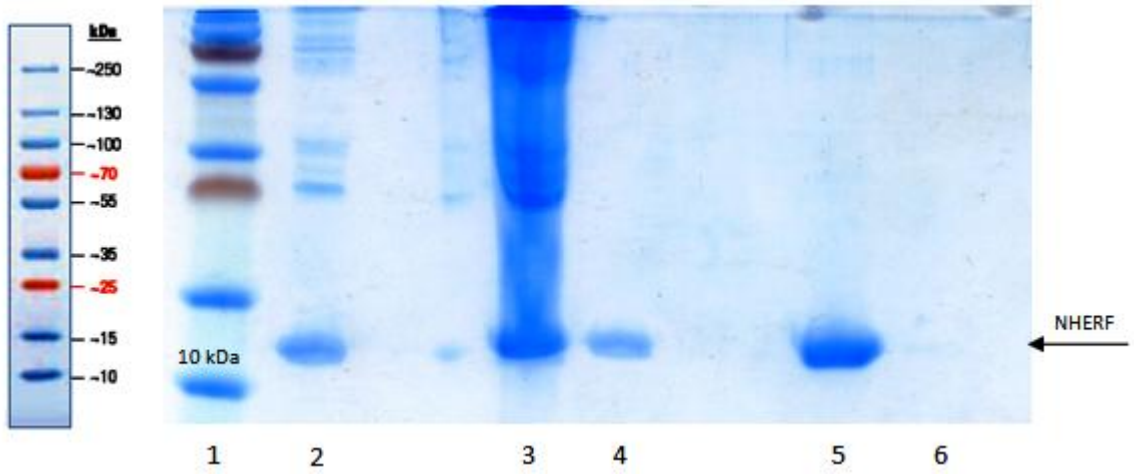


Figure 45: Concentration of NHERF1 PDZ $1^{(+)}$ using 50 kDa and 3 kDa concentrators. 1-Marker (PageRuler Plus Prestained, Fermentas), 2- IMAC-Unbound material (after thrombin cleavage), 3- Retentate of 50 kDa concentrator, 4- Filtrate of 50 kDa concentrator, 5- Retentate of 3 kDa concentrator, 6- Filtrate of 3 kDa concentrator. (Lab made 15% polyacrylamide/Tricine gel).

3.1.5.1 Gel Filtration Chromatography of NHERF1 PDZ $1^{(+)}$

The purified NHERF1 PDZ $1^{(+)}$ protein (800 µl) was further purified by gel filtration chromatography as described in 2.2.11. Two peaks appeared in the gel filtration trace (Figure 47A) and a third peak was related to the buffer components which emerged at the end of total column volume at 24 mL. Materials from the first two peaks (named NHERF1 PDZ $1^{(+)}$ and unknown) were collected in a 96-Well Gel Filtration Plate. Samples were pooled from fractions B1-B4 for NHERF1, B5-B7 for Unknown material. SDS-PAGE analysis of the final purified material revealed a single band (Figure 47B) for the samples from B1-B4 and a molecular weight of 11 kDa which was roughly compatible with NHERF1 PDZ $1^{(+)}$ size (10.215 kDa). This slight shift in the molecular weight of NHERF1 PDZ $1^{(+)}$ in SDS-PAGE might be a result of the posttranslational modification (S-glutathionylation) that the addition of glutathione to a cysteine residue. The identification and modification of NHERF1 PDZ $1^{(+)}$ was confirmed by MALDI-TOF as shown in Figure 46. However, it was reported that the mobility of NHERF1 on SDS-PAGE can be delayed and the protein appeared slightly larger because of another type of posttranslational modification (phosphorylation of Ser289 in PDZ 2) (Hall et al, 1999). In this study, the delay of the protein mobility in the gel was noticed with NHERF1 PDZ $1^{(+)}$ which was purified using GST-tag and glutathione. In contrast, the posttranslational modification (the addition of glutathione) was not observed with NHERF1 PDZ 1 which purified by 6His and imidazole.

No band appeared in the B5-B7 sample suggesting it was also a non-protein buffer component. The elution of NHERF1 PDZ $1^{(+)}$ as a monomer at 16 ml was also reported for gel-filtration chromatography of NHERF1 PDZ 1 by Nawrot et al (2006). The purity of the final product could be high as there were no contaminating peaks.

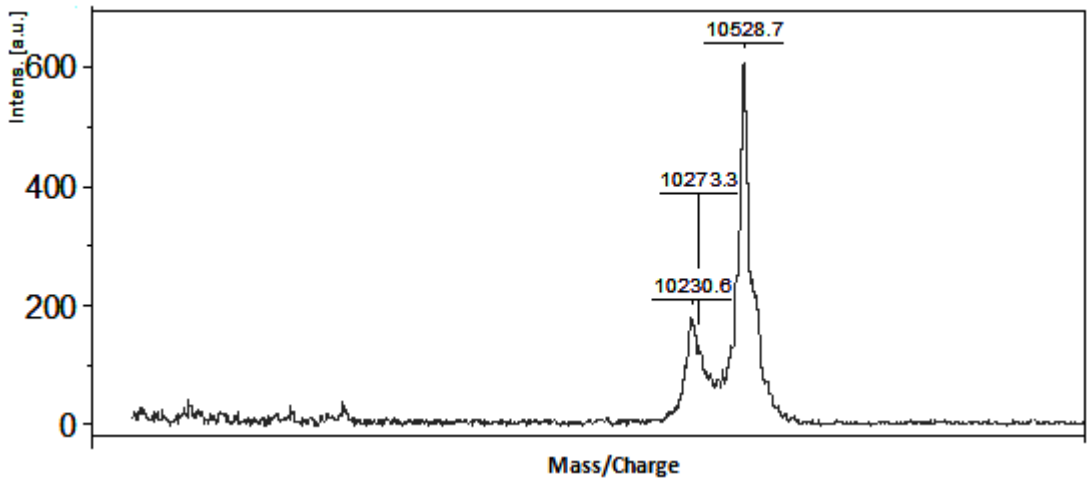


Figure 46: MALDI-TOF mass spectrum of NHERF1 PDZ 1$^{(+)}$. Three peaks were obtained and all the three peaks indicate a type of posttranslational modification. The peak of 10528.7 is the calculated molecular weight (MW) of NHERF1 PDZ 1$^{(+)}$ (10215 Dalton) plus glutathione (313 Dalton). The S-glutathionylation is a posttranslational modification by which glutathione is added to a cysteine residue. The peak of 10273.3 is thought to be Carbamidomethyl (a common posttranslational modification that can be detected by in MALDI-TOF) by which carbamidomethyl group (57.07 Da) is added to a cysteine residue. The peak of 10230.6 is thought to be oxidation (a common posttranslational modification that can be detected by in MALDI-TOF) by which one Oxygen molecule (15.9 Da) is added to a methionine residue. (a.u) is an arbitrary unit.

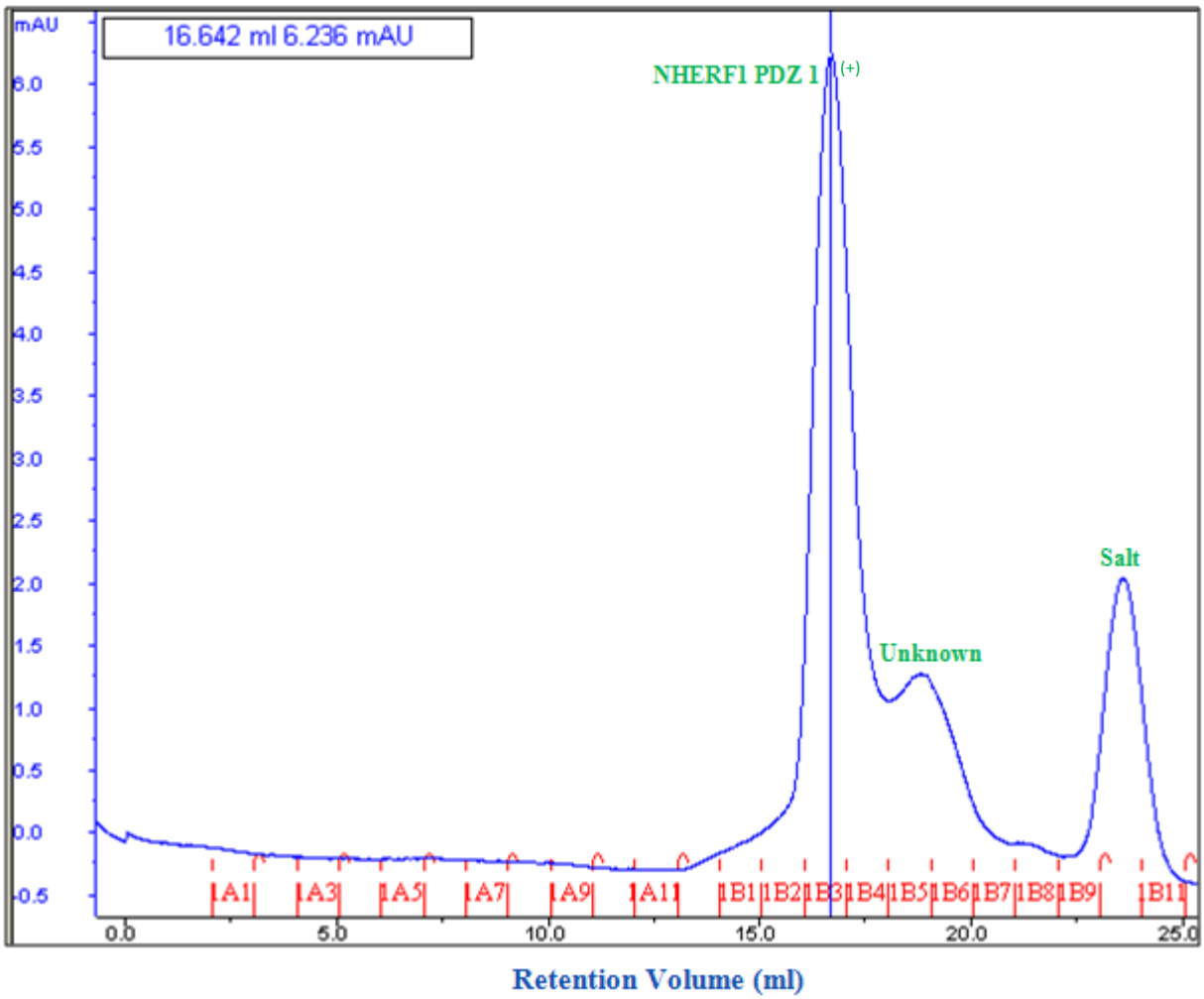


Figure 47A: Gel filtration chromatography of NHERF1 PDZ 1$^{(+)}$. The absorbance was monitored at 280 nm, mAU (ultraviolet absorbance unit). Samples from B1, B3, B4 and B5 were analysed by SDS-PAGE (Figure 47). NHERF1 PDZ 1$^{(+)}$ was appeared in B3 and B4.

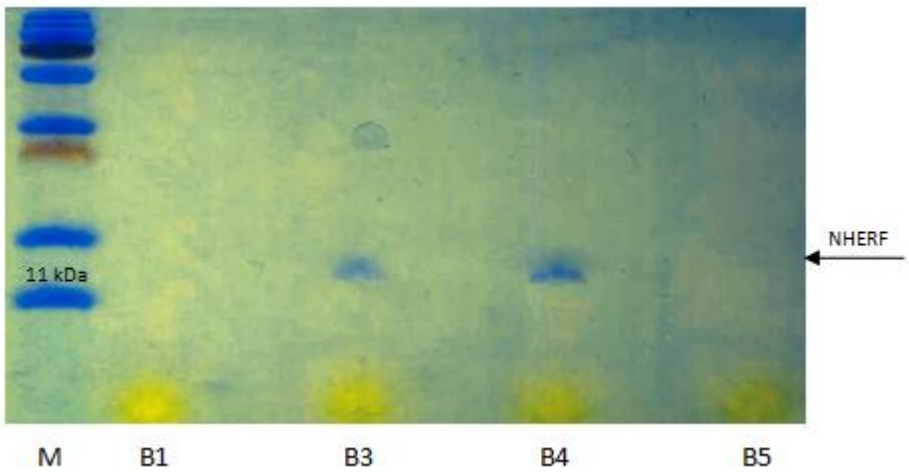


Figure 47B: SDS-PAGE of gel filtration samples (B1, B3, B4 and B5). M is marker (PageRuler Plus Prestained, Fermentas). Lab made 15% polyacrylamide/Tricine gel.

3.2 Results and Discussion 2

3.2.1 Pull-down Assay of CFTR-NHERF1 PDZ $1^{(+)}$

The protein samples used in pull-down assay experiments were His-tagged full-length human, mouse and killifish CFTR (Bait Proteins). The source of human CFTR (0.1 mg/mL) is described in section 2.1.1. Mouse and killifish CFTR samples (≈ 0.3 mg/mL) were described in section 2.1.1. NHERF1 PDZ $1^{(+)}$ (Prey Protein) (0.1 mg/mL) was expressed and purified as described in 3.1.5. Figure 48 shows the strategy used in these experiments.

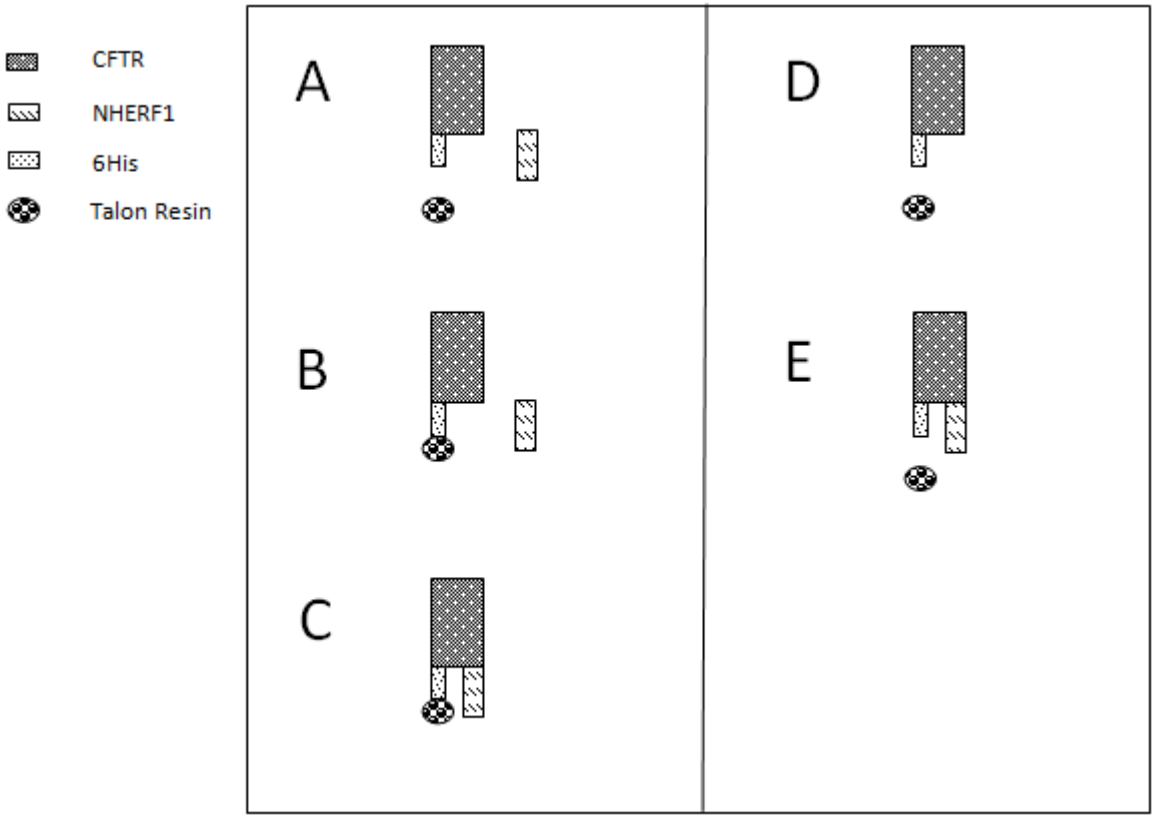


Figure 48: The diagram illustrates incubation and elution steps of the pull-down assay. A, mixture of CFTR-NHERF1 PDZ $1^{(+)}$-Talon resin which was left rotating for 2 hr at RT to allow the binding. B shows an outcome where there was no interaction between NHERF1 and CFTR while C shows the second possibility that the two proteins interact. D and E, show the corresponding elution results.

3.2.1.1 CFTR-NHERF1 PDZ 1 Interaction Analysis

Untagged-NHERF1 PDZ $1^{(+)}$ was used in pull-down assay experiments to investigate the ability of this domain to bind to human, mouse and killifish CFTR that have C-terminal 10 His-tag. The binding between the the two proteins were performed as described in 2.2.10. The aims of the pull-down assay of CFTR-NHERF1 PDZ $1^{(+)}$ were to examine the interaction between the two proteins by biochemical experiments (in vitro) before preparation of samples for Single Particle Analysis. The second aim was to try CFTR-NHERF1 PDZ $1^{(+)}$ binding assay in Eppendorf tubes and to analyze the result by SDS-PAGE. According to our information, no previous experiments were performed for the two proteins using this method.

3.2.1.1.1 Pull-down Assay for Human CFTR-NHERF1 PDZ 1 Complex

50 µg/µl of human CFTR was mixed with 100 µg/µl of NHERF1 PDZ $1^{(+)}$ (final concentration) and 15 µl of Talon. The mixture was left at room temperature for 2 hr as described in 2.2.10. A 1:1 ratio of CFTR to NHERF1 based on their molecular weight is 18 µg/µl of CFTR and 1 µg/µl of NHERF1. However, an excess of NHERF1 was used to allow for inefficient binding. The molar ration was approximately 1:36 of CFTR and NHERF1 PDZ 1, respectively. The results of this experiment are presented in Figure 49.

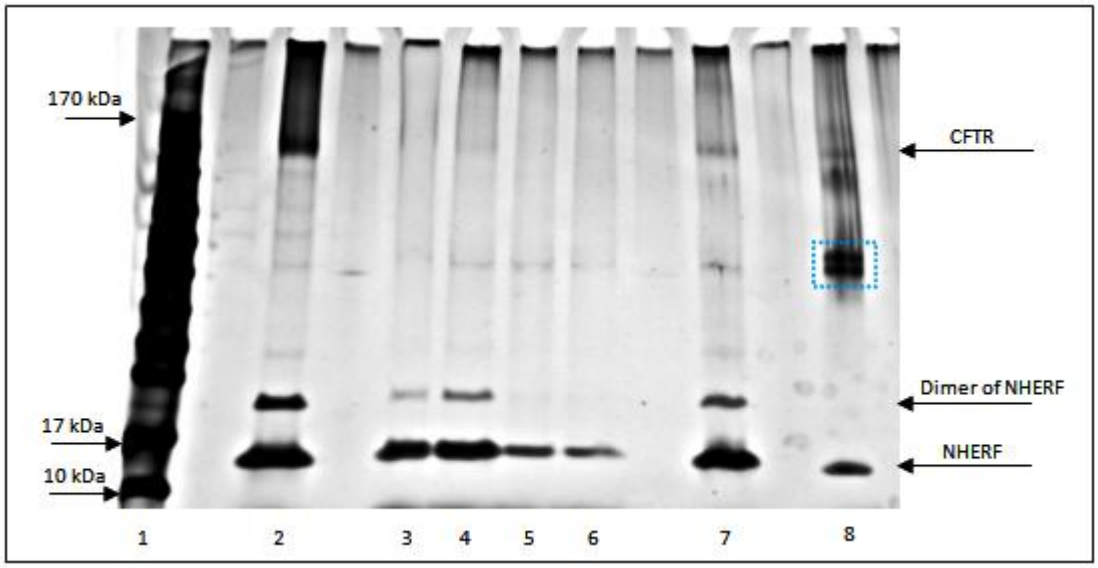


Figure 49: Analysis of human CFTR-NHERF1 PDZ 1$^{(+)}$ pull-down assay by SDS-PAGE. 1-Marker (PageRuler Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4- Supernatant following Wash2. 5- Supernatant following Wash3. 6-Supernatant following Wash4. 7- Supernatant following Elution step. 8-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad). In lane 8, after washing the beads with 1% SDS, a band (≈ 95 Da in blue dashed box) appeared. This band may be a CFTR degradation product.

As shown in Figure 49, the unbound materials were removed (Lane 2) before Wash and Elution steps. Full-length human CFTR appeared in the gel as a monomer. A homodimer of NHERF1 PDZ 1$^{(+)}$ appeared and its appearance was proportional to the protein concentration. The homodimer was nearly 30 % of the NHERF1 protein detected. The amount of NHERF1 was lowered gradually through the wash steps but a high amount of CFTR-NHERF1 appeared in the Elution sample (Lane 7) indicating that there was probably an interaction between the tagged-CFTR and untagged-NHERF1 PDZ 1$^{(+)}$. This is the first time that this interaction between the two proteins was analyzed in this way. A sample of eluted complex was negatively stained and imaged in the transmission electron microscope (TEM) but was not used for Single Particle Analysis (SPA) because the number of particles was very small.

The same strategy was used to examine CFTR-NHERF $1^{(+)}$ interaction in the presence of ATP-PKA (7 mM and 7 nM respectively) by the pull-down assay but with different protein concentration (25 µg/µl of human CFTR and 50 µg/µl of NHERF1 PDZ $1^{(+)}$). The molar ratio was approximately 1:36 of CFTR and NHERF1 PDZ 1, respectively. The result is presented in Figure 50.

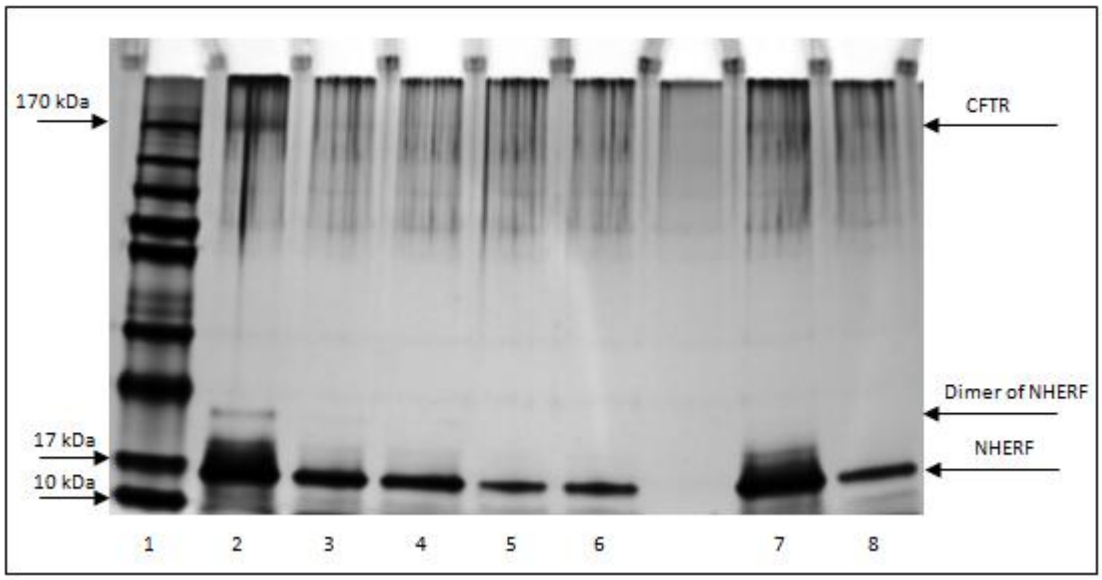


Figure 50: Analysis of human CFTR-NHERF1 PDZ $1^{(+)}$ (in the presence of ATP-PKA) by pull-down assay. 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Elution step. 8-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad). Smearing of the CFTR band was problematic with these experiments.

As shown in Figure 50, the higher amount of CFTR-NHERF in the elution sample (lane 7) compared to the wash samples indicates that CFTR was able to interact with NHERF1 in the presence of ATP-PKA. Furthermore, the homodimer band of NHERF1 was much lower (about 4-5 % of the protein concentration) in the presence of ATP-PKA. This loss of the homodimer was expected as a result of the dissociation of the dimers after phosphorylation (Shenolikar et al, 2001).

3.2.1.1.2 Pull-down Assay for Mouse and Killifish CFTR-NHERF1 PDZ $1^{(+)}$ Complex

Two separate pull-down assay experiments were performed for mouse CFTR-NHERF1 PDZ $1^{(+)}$ and killifish CFTR-NHERF1 PDZ $1^{(+)}$. Both CFTR samples were purified with LPG detergent. 163 µg/µl of CFTR and 90 µg/µl of NHERF1 PDZ $1^{(+)}$ were used in the two experiments. The molar ratio was approximately 1:10 of CFTR and NHERF1 PDZ 1, respectively. As shown in Figure 51 A, CFTR was not able to interact with NHERF1 PDZ $1^{(+)}$ in the presence of LPG. A very small amount of NHERF1 was in the Elution sample. A comparison between this small amount of NHERF1 with the eluted CFTR-NHERF in the previous experiments, highlights a huge difference in terms of protein-protein affinity. The same result was observed for killifish CFTR-NHERF1 PDZ $1^{(+)}$ (Figure 52 A). A dimer band of NHERF1 appeared in the two gels but was very weak.

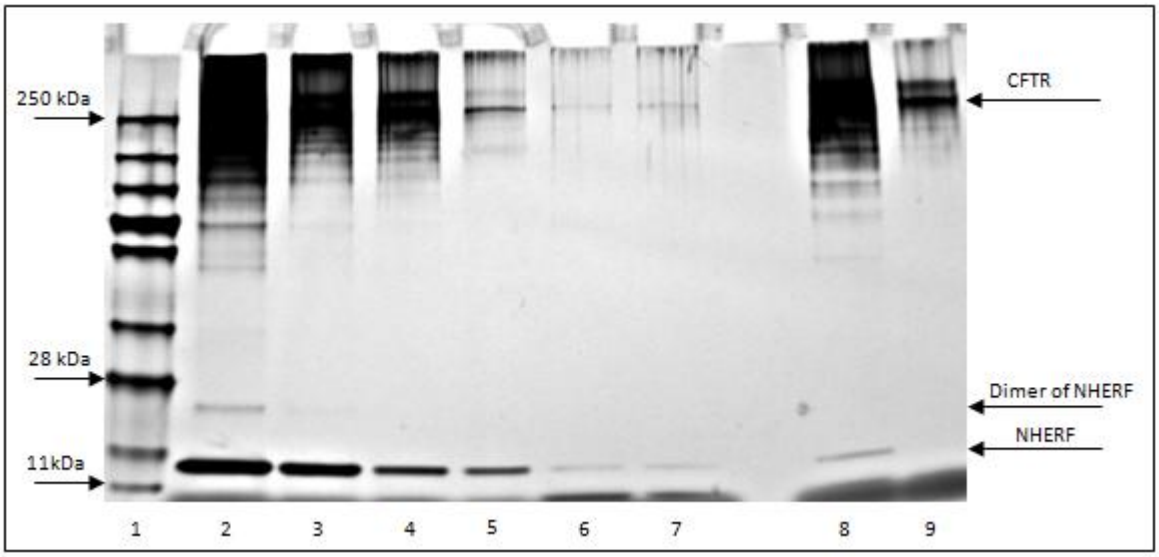


Figure 51 A: Analysis of mouse CFTR-NHERF1 PDZ $1^{(+)}$ pull-down assay in the presence of LPG. 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Elution step. 9-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

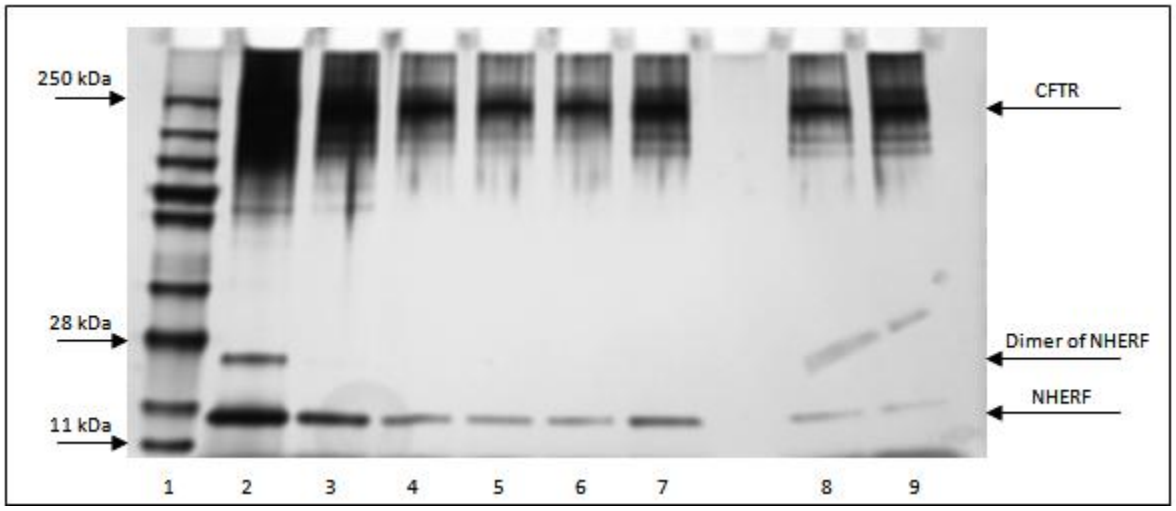


Figure 52 A: Analysis of killifish CFTR-NHERF1 PDZ $1^{(+)}$ pull-down assay in the presence of LPG. 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Elution step. 9-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

It was concluded that LPG which has a negative net charge and is a relatively harsh detergent, reduced CFTR-NHERF1 interaction. To investigate whether the anionic charge of LPG was responsible for the reduction of the interaction of CFTR with other proteins (NHERF1 in this case) (Huang et al, 1998), mouse and killifish CFTR samples were reconstituted into lipid (as described in section 2.2.15, and shown in Figure 96). The aim of reconstitution of CFTR into lipid was to try CFTR-NHERF1 PDZ $1^{(+)}$ pull-down assay in the absence of LPG. Theoretically, by this method, lipid vesicles will replace LPG detergent. Figures (51 B) and (52 B) present the reconstituted CFTR-NHERF1 results. The two experiments revealed that CFTR was able to interact with NHERF1 with a much lower affinity in the presence of LPG (lane 8). This conclusion was based on the very low amount of NHERF1 PDZ $1^{(+)}$ in the elution sample. The presence of a dimer band for NHERF1 in the reconstituted sample (Figure 51B) versus CFTR in the presence of LPG (Figure 51A) was an indication that CFTR interacted strongly with NHERF1 PDZ $1^{(+)}$ after reconstitution because the dimer band was only observed on SDS-PAGE when high concentrations of NHERF1 PDZ $1^{(+)}$ were loaded. After reconstitution, CFTR was difficulty to solubilize for SDS-PAGE and migrated as a band at the top of the gel as has been observed previously (N. Cant, personal communication).

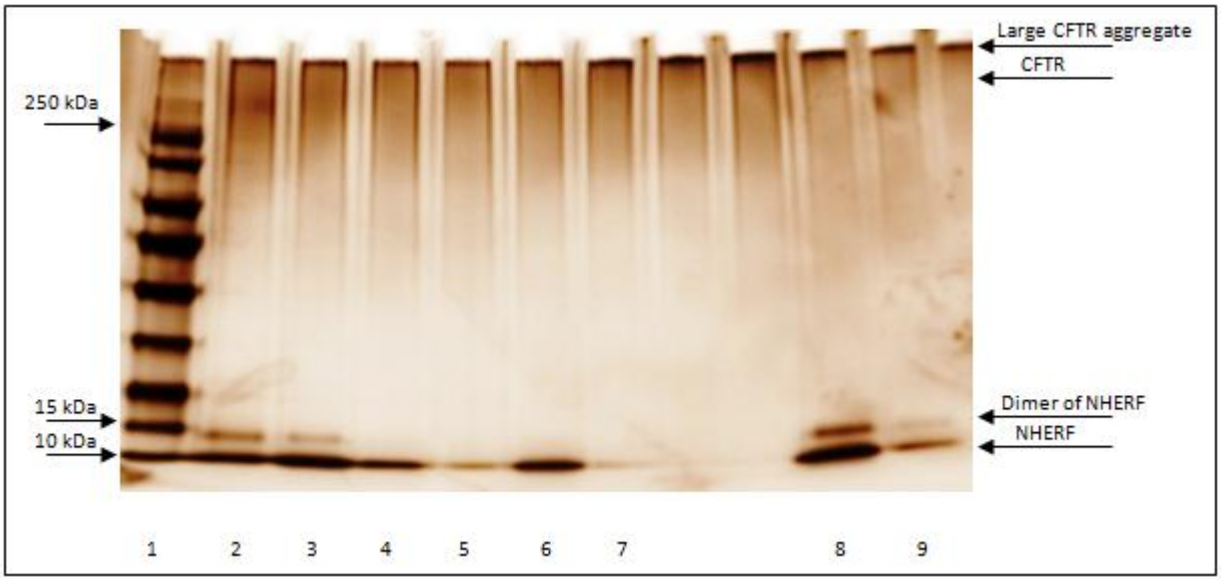


Figure 51 B: Analysis of reconstituted mouse CFTR-NHERF1 PDZ 1$^{(+)}$ pull-down assay by SDS-PAGE. 1-Marker (PageRuler Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Elution step. 9-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

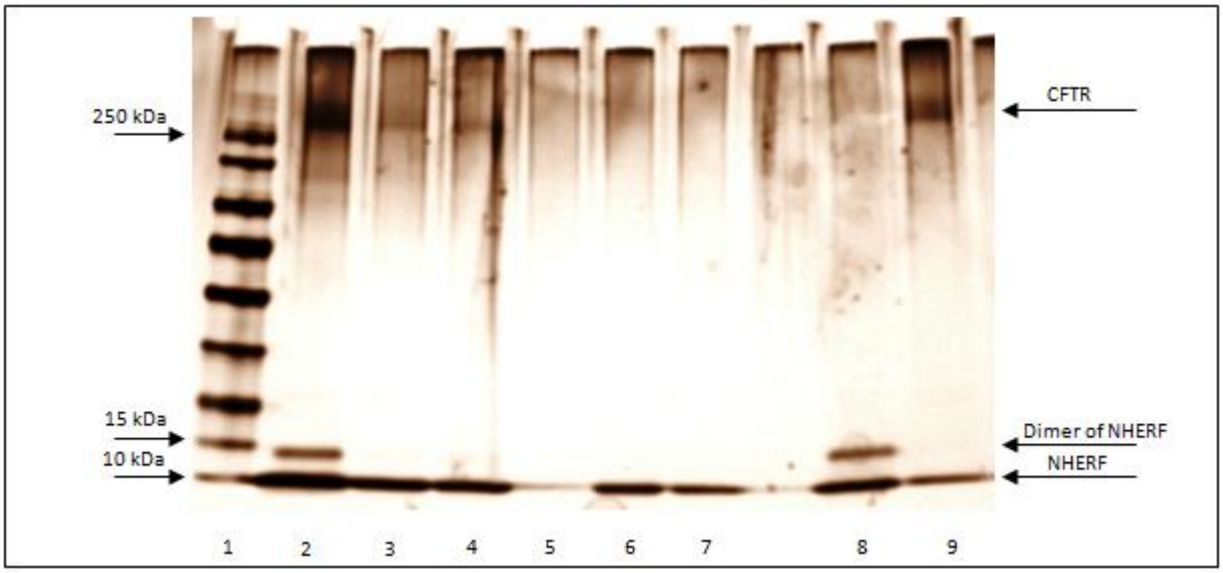


Figure 52 B: Analysis of reconstituted killifish CFTR-NHERF1 PDZ 1$^{(+)}$ pull-down assay by SDS-PAGE. 1-Marker (PageRuler Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Elution step. 9-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

As shown in Figure 51 B, after reconstitution, mouse CFTR did not enter the SDS-PAGE gel efficiently and the Elution band was smeared. After reconstitution, killifish CFTR also migrated poorly into the gel (Figure 52 B) although the samples in lanes 2, 3, 4 and 9 were just detected.

3.2.1.1.3 Pull-down Assay for Human CFTR-C-terminus^{42aa} Complex

63 µg/µl of the human CFTR was mixed with 210 µg/µl of the C-terminus^{42aa}. The molar ratio was approximately 1:100 of CFTR and the C-terminus^{42aa}, respectively. The aim of this experiment was to investigate whether the R-domain of full-length CFTR is able to bind to the C-terminus^{42aa} and also if this interaction is stronger in the phosphorylated state. This hypothesis was based on NMR data that was obtained by (Bozoky et al, 2011b) (unpublished data), Department of Molecular Structure and Function, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada. The NMR data showed that phosphorylation of residues (660, 670, 700, 712, 737, 753, 768, 795 and 813) of the isolated R-domain increased binding to the C-terminus^{42aa} . Figures 53 A and B show the result of the pull-down assay for the human CFTR-C-terminus^{42aa} complex.

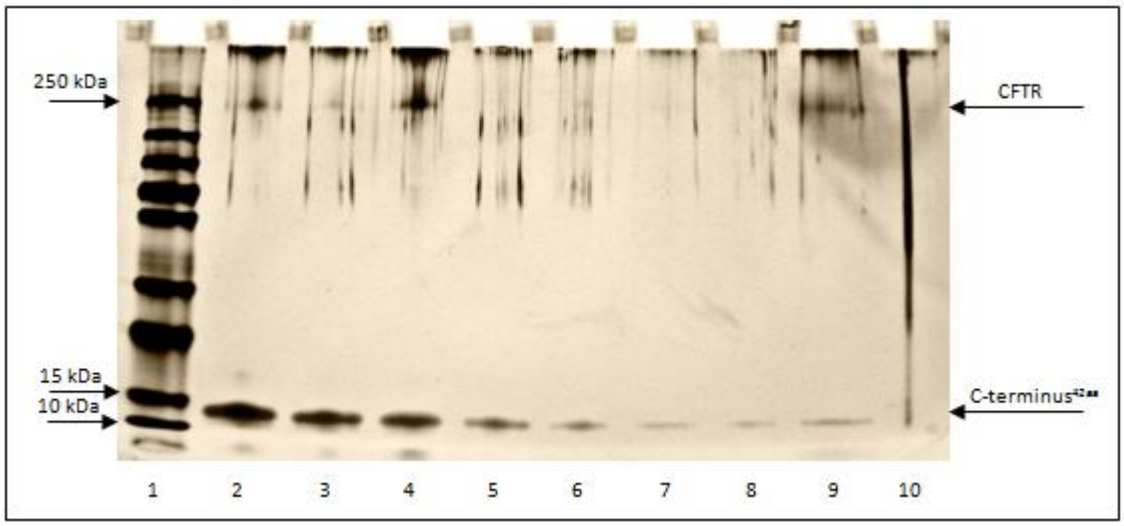


Figure 53 A: Analysis of human CFTR-C-terminus^{42aa} pull-down assay (No ATP) by SDS-PAGE. 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Wash6, 9-Supernatant following Elution step. 10-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

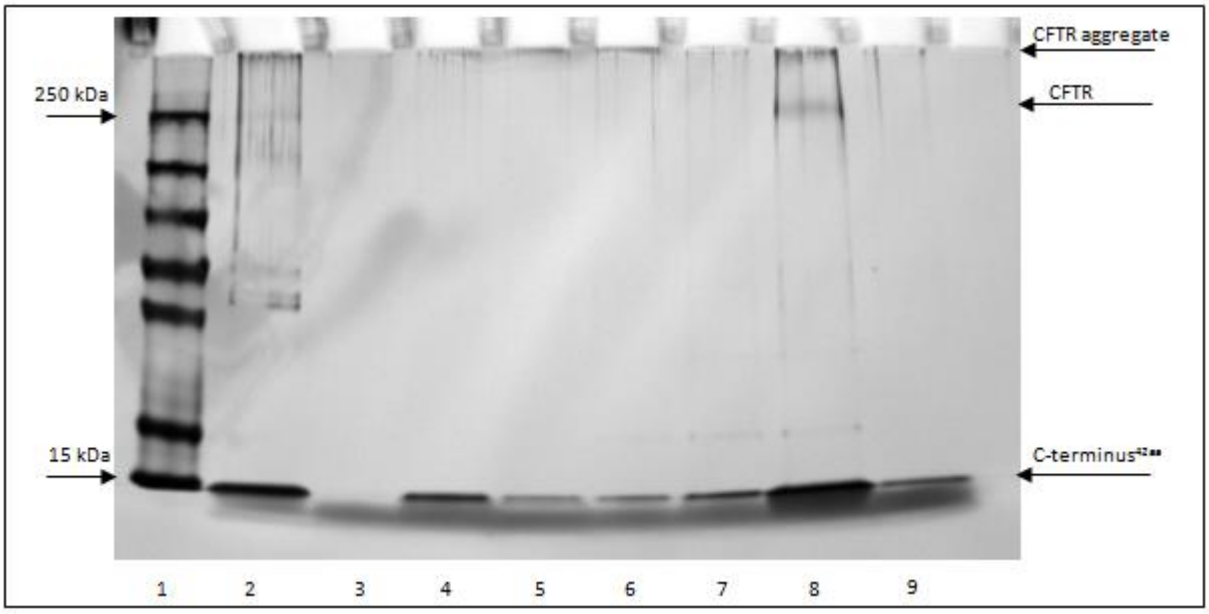


Figure 53 B: **Analysis of human CFTR-C-terminus^{42aa} pull-down assay (in the presence of ATP-PKA) by SDS-PAGE.** 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Blank lane (the sample was lost). 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Elution step. 9-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad). The C-terminus^{42aa} band overlaps with loading solvent front.

Figure 53 C is with contrast-adjusted image to show the C-terminus^{42aa} band clearly.

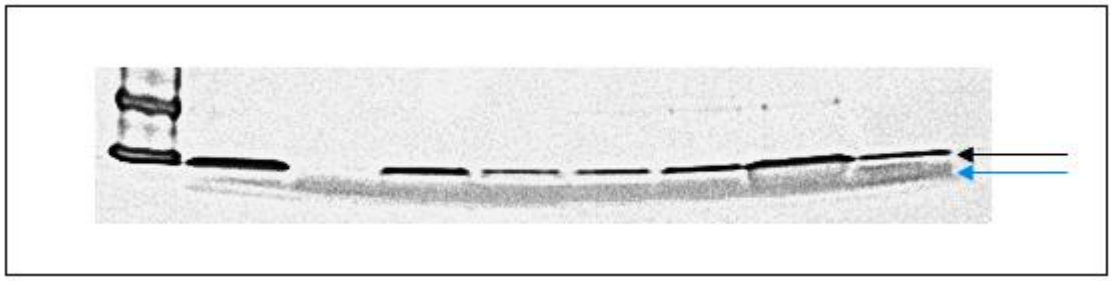


Figure 53 C: **Contrast-adjusted image of the bottom part of Figure 53 B.** This image shows the space between the C-terminus^{42aa} band and loading stain. The black arrow points the C-terminus^{42aa}. The blue arrow points the stain band.

Figure 53A showed that the C-terminus^{42aa} appeared in the Elution sample indicating that there was a weak interaction between CFTR and the C-terminus^{42aa} in the absence of ATP. This was expected from NMR data where the R-domain and the C-terminus are weakly interacting in the unphosphorylated state hence we would propose that exogenously added C-terminus^{42aa} could bind. On the other hand, no significant interaction was observed (even there was a very weak band of the C-terminus^{42aa} in Elution sample as appeared in Figure 53 B, lane 8) between CFTR and the C-terminus^{42aa} in the presence of ATP. We would propose that the R-domain and the C-terminus in the full-length of CFTR are already fully interacting and hence exogenous C-terminus^{42aa} no longer bind. These results are therefore compatible with the NMR data if one assumes a tight binding of the phosphorylated R-region to the intrinsic C-terminus of CFTR which is not competed by exogenous C-terminal peptide.

3.2.1.1.4 Pull-down Assay for killifish CFTR-C-terminus^{42aa} Complex

In a similar experiment to pull-down assay for human CFTR-C-terminus^{42aa} complex (as shown in section 3.2.1.1.4), a sample of full-length killifish CFTR (in LPG) was mixed with the C-terminus^{42aa}. The molar ratio was approximately 1:10 of CFTR and the C-terminus^{42aa}, respectively. There was not evidence of interaction as shown in Figures 54 A and B. A similar conclusion is made here that the R-domain and the C-terminus in the full-length of CFTR are already fully interacting and hence exogenous C-terminus^{42aa} no longer bind.

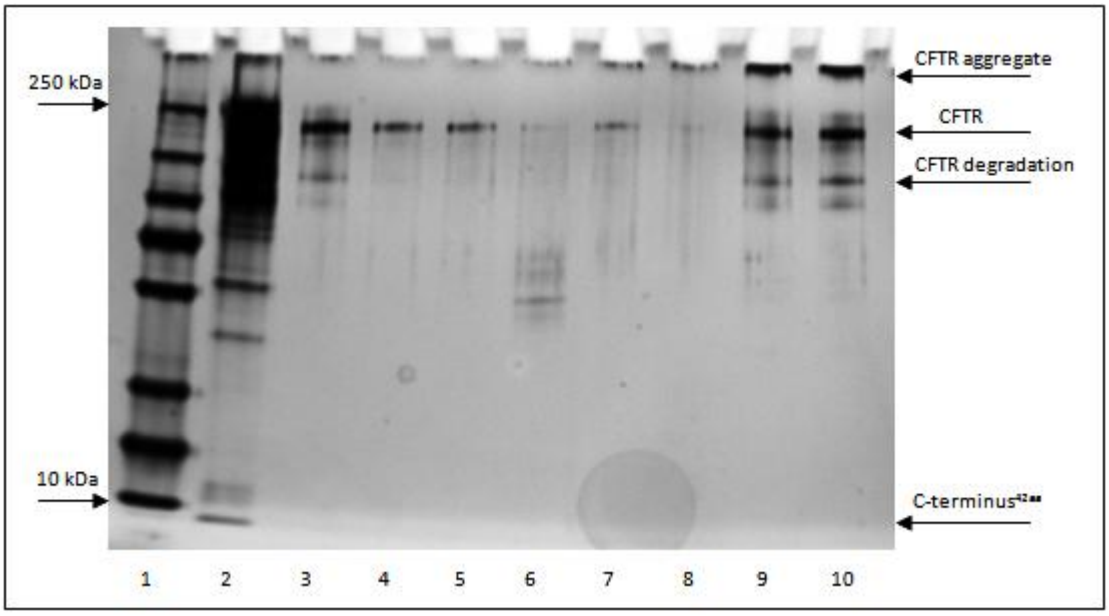


Figure 54 A: Analysis of killifish CFTR-C-terminus^{42aa} pull-down assay (No ATP) by SDS-PAGE. 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Wash6, 9-Supernatant following Elution step. 10-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

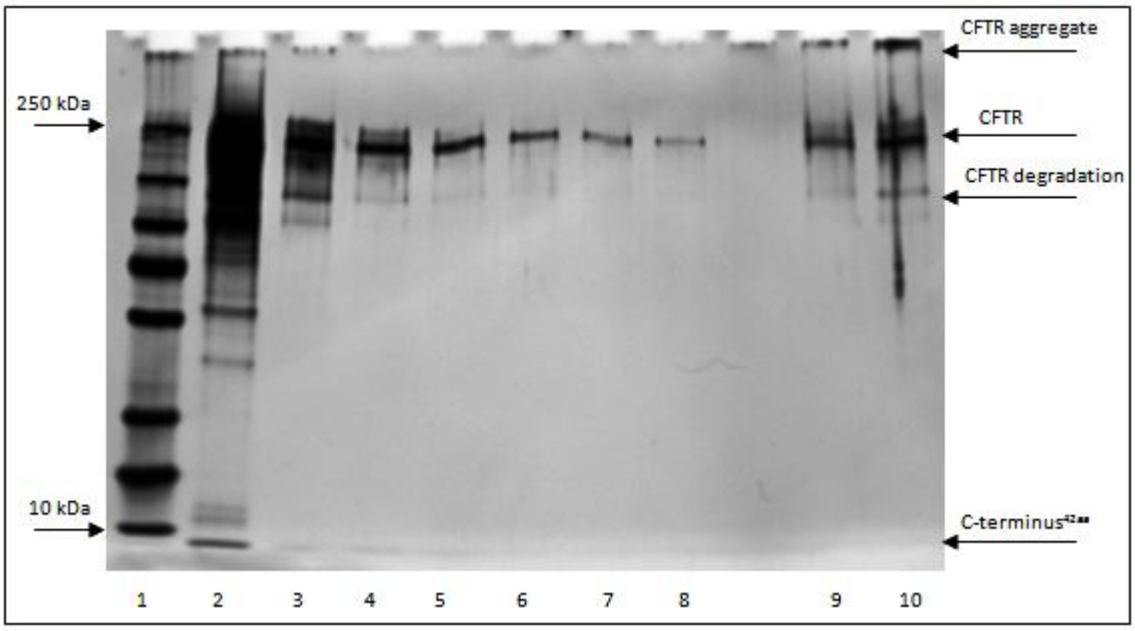


Figure 54 B: **Analysis of killifish CFTR-C-terminus^{42aa} pull-down assay (in the presence of ATP-PKA) by SDS-PAGE.** 1-Marker (PageRuler Plus Prestained, Fermentas). 2-Supernatant following the incubation step. 3-Supernatant following Wash1. 4-Supernatant following Wash2. 5-Supernatant following Wash3. 6-Supernatant following Wash4. 7-Supernatant following Wash5. 8-Supernatant following Wash6, 9-Supernatant following Elution step. 10-Supernatant following Wash with 1% SDS. (Ready Gels, 4-15 % Tris-HCl, Bio-Rad).

Interestingly, the C-terminal 10 His-tag did not prevent the CFTR-NHERF1 interaction. In a similar experiment that was performed by Li and Naren (2011), the 10 histidine residues at the C-terminus of CFTR did not disrupt CFTR-NHERF1 interaction. On the other hand, Bates et al (2006) found that the 10 histidine residues at the C-terminus of CFTR prevented CFTR-NHERF1 interaction.

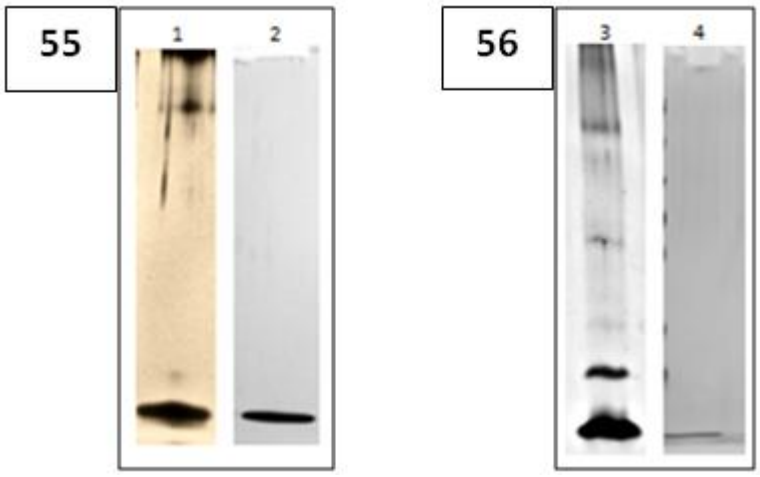
A dimer of NHERF1 PDZ $1^{(+)}$ appeared on the gel in most of the pull-down assays and this observation was also reported by (Fouassier et al, 2000). They found that NHERF1 had the ability to self-associate in which PDZ 1 binds with more affinity to PDZ 1 and PDZ 2 binds with high affinity to PDZ 2. In addition, both PDZ domains had the ability to interact with each other but with lower affinity compared to self-associate. This study by (Fouassier et al, 2000) is consistent with pull-down assay in terms of PDZ1-PDZ1 binding and the appearance of this dimerization on gel.

Table 17 shows a summary of pull-down assay experiments.

Complex	**Detergent**	**Protein-Protein Interaction 0-1-2-3**
Human CFTR-NHERF1 PDZ $1^{(+)}$	DDM	3
Human CFTR-NHERF1 PDZ $1^{(+)}$-ATP-PKA	DDM	2
Mouse CFTR-NHERF1 PDZ $1^{(+)}$	LPG	1
Mouse CFTR-NHERF1 PDZ $1^{(+)}$	No (reconstituted)	3
Killifish CFTR-NHERF1 PDZ $1^{(+)}$	LPG	1
Killifish CFTR-NHERF1 PDZ $1^{(+)}$	No (reconstituted)	3
Human CFTR-C-terminus^{42aa}	DDM	1
Human CFTR- C-terminus^{42aa}-ATP-PKA	DDM	0
Killifish CFTR-C-terminus^{42aa}	LPG	0
Killifish CFTR- C-terminus^{42aa}-ATP-PKA	LPG	0

Table 17: A summary of pull-down assay conditions (1st and 2nd columns) and results (3rd column). 0-3 protein-protein interaction scale, 3 strong interaction, 0 no interaction. Score 0-3 based on the concentration of NHERF1 PDZ $1^{(+)}$ in the Elution sample and the presence of a NHERF1 PDZ $1^{(+)}$ dimer.

The purity of the C-terminus[42aa] and NHERF1 were checked indirectly by silver staining after the pull-down assay. The purity was estimated to be 100 %. NHERF1 (from gel filtration) and the C-terminus[42aa] (purified by Talon immobilized metal affinity chromatography) showed the same level of purity (Figures 55 and 56).



Figures 55 and 56: A purity check of the C-terminus[42aa] and NHERF1. **Figure 55:** Lane1, a sample of pull-down assay, the higher band belongs the full-length CFTR, the lower is the C-terminus. Lane 2, a single band corresponding to the C-terminus alone. **Figure 56:** Lane 3, a sample of pull-down assay, the higher band belongs the full-length CFTR, the lower are a dimer and monomer of NHERF1 respectively. Lane 4, a single band corresponding to a sample of NHERF1 alone. All the gels were silver stained.

The SDS-PAGE gels used in pull-down experiments were gradient gels (4-15%) which have an increasing concentration of acrylamide and which causes a decrease of pore sizes in the direction of samples migration. This type of gel allows separation of complex mixtures with different molecular weight ranges. However, the best results in resolving two molecules from one another are achieved in single percentage gels. This disadvantage of gradient gels may explain the incorrect positions of CFTR and NHERF1 on the gels.

3.2.2 Single Particle Analysis of Human CFTR-NHERF1 PDZ $1^{(+)}$

Several 2D crystallization trials were performed with no success. Therefore, Single Particle Analysis was an alternative method to study the CFTR structure. Single Particle Analysis (SPA) is a and high (Zhang et al. 2008) resolution. In this study, a complex of human CFTR-NHERF1 PDZ $1^{(+)}$ was investigated by SPA as described in section 2.2.14. The aim of this study was to generate a reasonable powerful technique to study protein structure at low (Awayn et al, 2005), medium (Stark et al. 2001) resolution (~20 Å) 3D structure of CFTR in the presence of NHERF1. This structure has been compared with free-CFTR (control) to highlight any conformational changes in CFTR structure as a possible result of CFTR-NHERF1 interaction. A 3D map of CFTR-NHERF1 was generated using 43605 particles that were selected semi-automatically. Beside the investigation of the NHERF1 effect on the CFTR structure, there are other possible future approaches as mentioned in the Future Work section.

3.2.2.1 A Low Resolution Structure of CFTR-NHERF1 PDZ $1^{(+)}$

As a start point, biochemical experiments (pull-down assay) as described in section (3.2.1.1.1) were performed to ensure that the binding of CFTR to NHERF1 was occurring. The two proteins appeared in the elution sample. Lane 7 and 8 (Figure 49) from pull-down assay of CFTR-NHERF1 were moved to this section as (Figure 57) to show the binding between the two proteins.

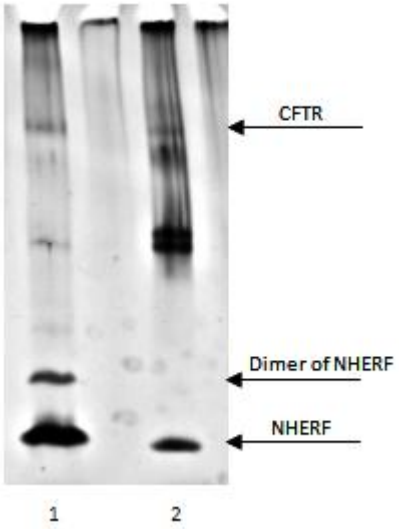


Figure 57: Pull-down assay of CFTR and NHERF1 revealed the interaction between the two proteins. 1, Supernatant following Elution step. 2, Supernatant following of wash with 1% SDS.

A mixture of CFTR-NHERF1 PDZ $1^{(+)}$ (375 µg and 150 µg respectively) was adsorbed (for 2 minutes) onto a thin carbon film rendered hydrophilic by glow-discharge. The molar ratio was approximately 1:7 (CFTR to NHERF1). This ratio ensures an excess amount of NHERF1 compared to CFTR (as the equal ratio of CFTR to NHERF1 based on their molecular weight is 18:1). The sample was then washed (for 10 s) with a 5 µl drop of MQ water and then negatively stained with 4 % uranyl acetate solution for 30 s. The sample was blotted on filter paper (Whatman No 1), and dried in air. 40 micrographs of the negatively stained particles were recorded as described in section 2.2.14.2. The 40 micrographs were assessed for data quality (suitable range of defocus and astigmatism, high contrast, homogeneous particles with enough space for selection). Live Fast Fourier Transforms (FFT) (electron microscopy tool) was used to optimize defocus (Figure 58). Good quality micrographs were analysed by EMAN. A few particles (20-30) with a diameter of 10 – 15 nm were selected from each micrograph. The CFTR diameter was based on the studies that were done by (Eskandari et al, 1998; Rosenberg et al, 2004; Awayn et al, 2005; Zhang et al, 2009). The 20-30 particles were used as a reference to select 500-1000 particles using EMAN (Boxer) autoboxing software. Figures 59 and 60 show examples of images and selected particles. After particle selection, contrast transfer function (CTF) correction was performed by the *`ctfit´* software in EMAN. The particles were aligned and then classified according to their orientation. 8101 bad particles (outliers in terms of size and shape) were removed (as shown in Figure 61). 35504 good particles were used to generate averaged classes (176 classes). 40 averaged classes from 176 classes (with top, side and partial views) were used to generate the start 3D model (Figure 17). The start 3D model was used to begin the refinement of the structure (16 iterations, C2 symmetry). A 3D structure was produced with a resolution of ~ 18 Å estimated from the FSC plot shown in Figure 64. This may be an optimistic estimation as the FSC test has been argued to be. Figure 62 shows one of the averaged classes in which raw data was compared to the back-projection images of the final 3D model. This comparison indicated that the final 3D model matched very well the individual class averages and suggested high reliability of this analysis. Figure 63 shows the final 3D model of human CFTR-NHERF1 PDZ $1^{(+)}$ before and after the low-pass and high-pass filtration. The resolution of the final 3D map was estimated using the *´eotest`* program in

EMAN. This program generates 3D maps from even and odd halves of the raw particle set with calculation of the Fourier shell correlation (FSC) curve between the two 3D models (Van Heel, 1987, Ludtke et al, 2001). The resolution of ~ 18 Å was assessed by the standard criterion using a FSC value of 0.5 (Ludtke et al, 2004) (Figure 64). Finally, the 3D structure was subjected to low-pass and high-pass filtration to reduce the background noise that frequencies below zero or above the estimated resolution (Figure 63).

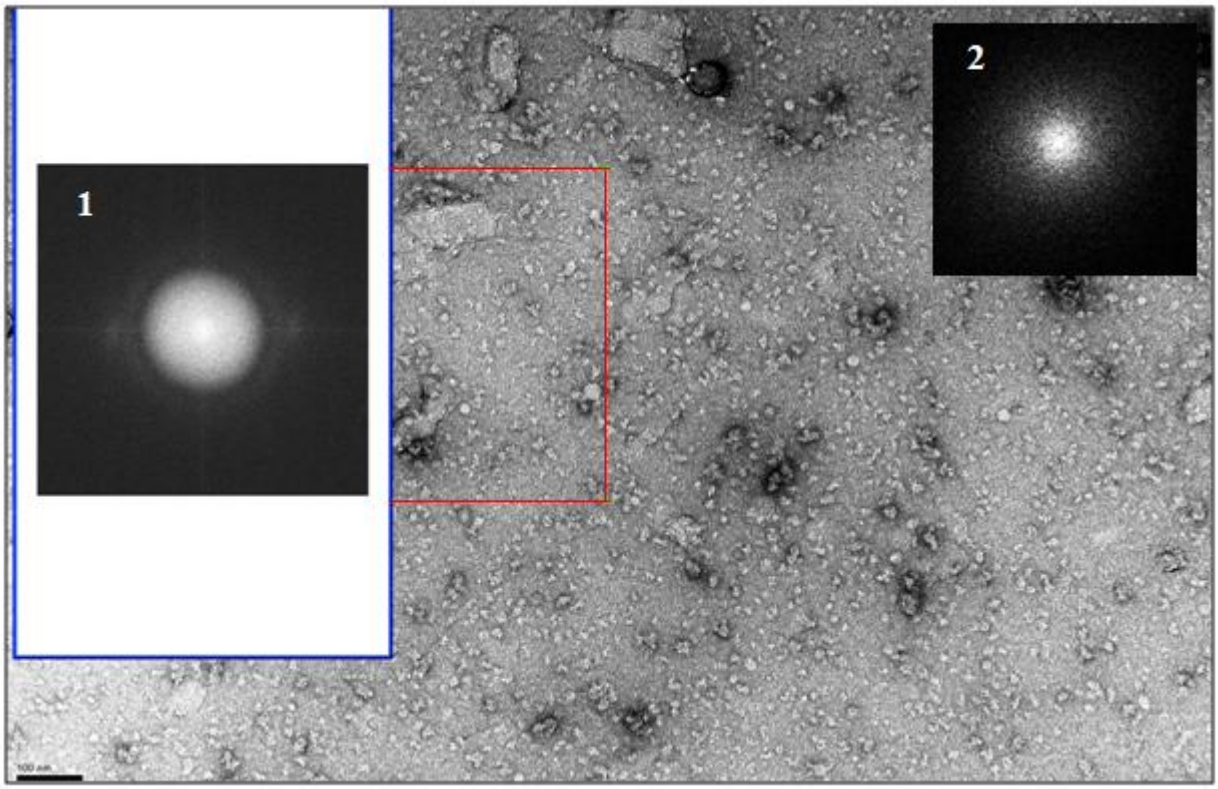


Figure 58: Assessment of defocus by FFT. The red box is the area of the micrograph that was assessed by FFT. 1and 2 are examples of FFT at strong defocus and close to focus respectively.

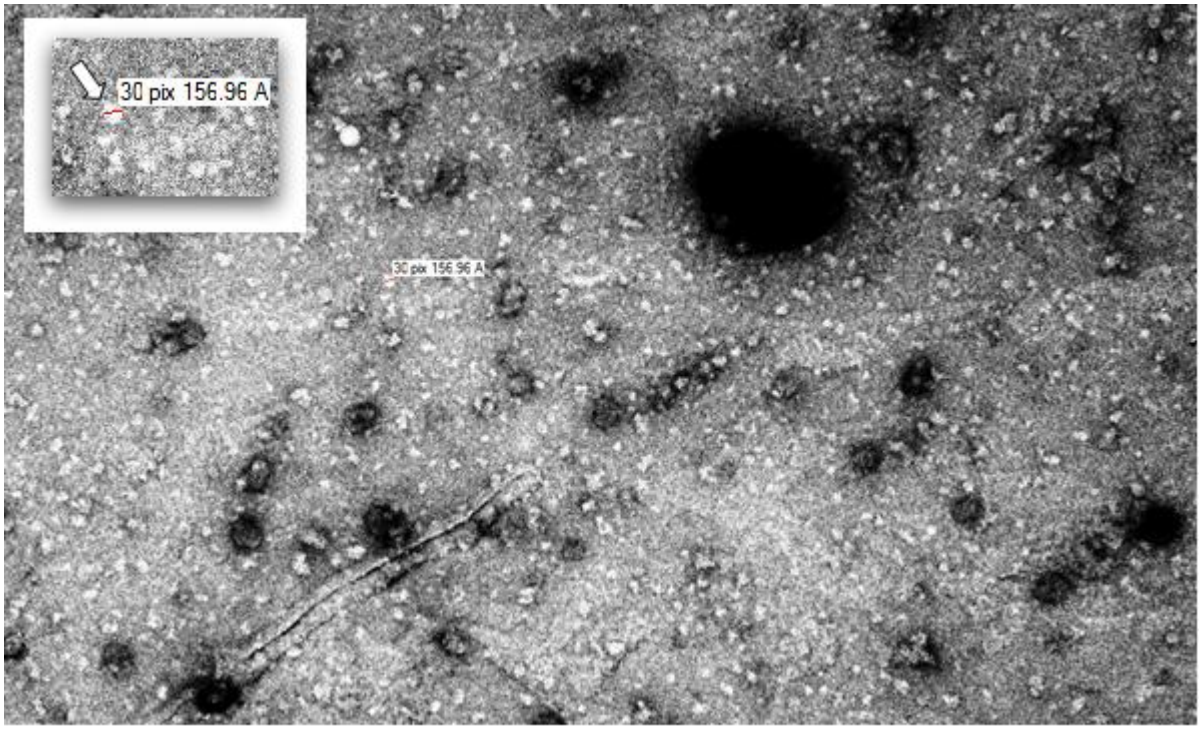
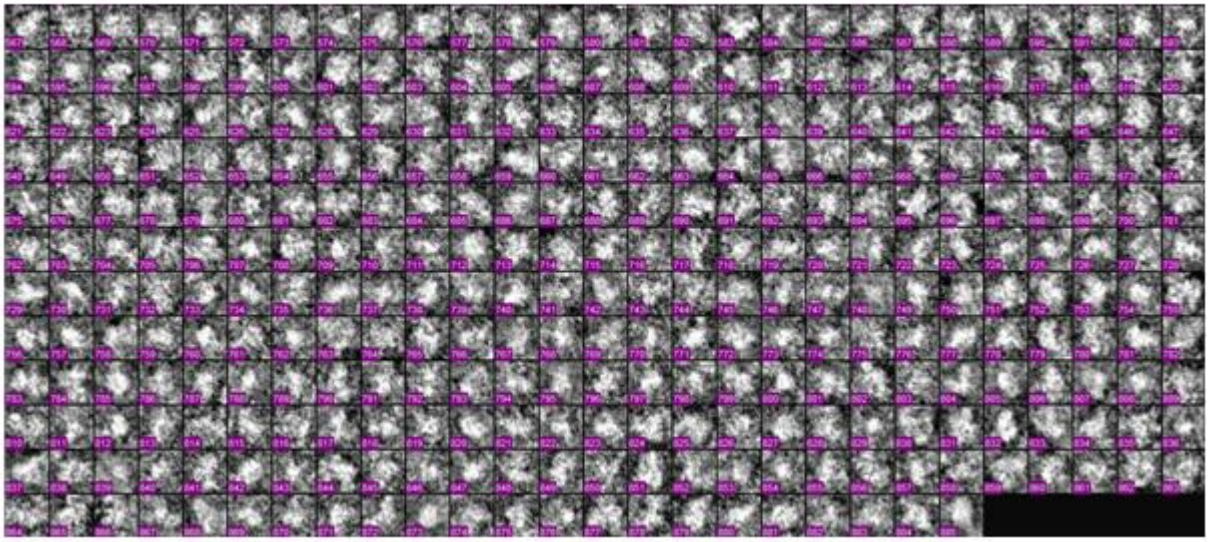


Figure 59: Particles of CFTR-NHERF1 complex. The white arrow in the white box points to one single particle of ~15 nm.

A



B

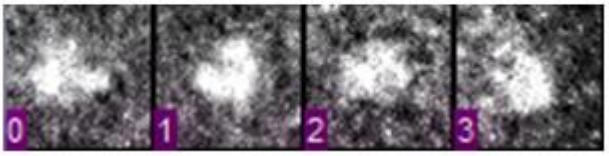


Figure 60: Selected particles from CFTR-NHERF1 complex. (A) Shows the boxed particles as they appeared in the Boxer window. The box size was 48 pixels (249.6 Å x 249.6 Å). (B) shows four boxed particles (zoomed out) of CFTR. CFTR is the white spots.

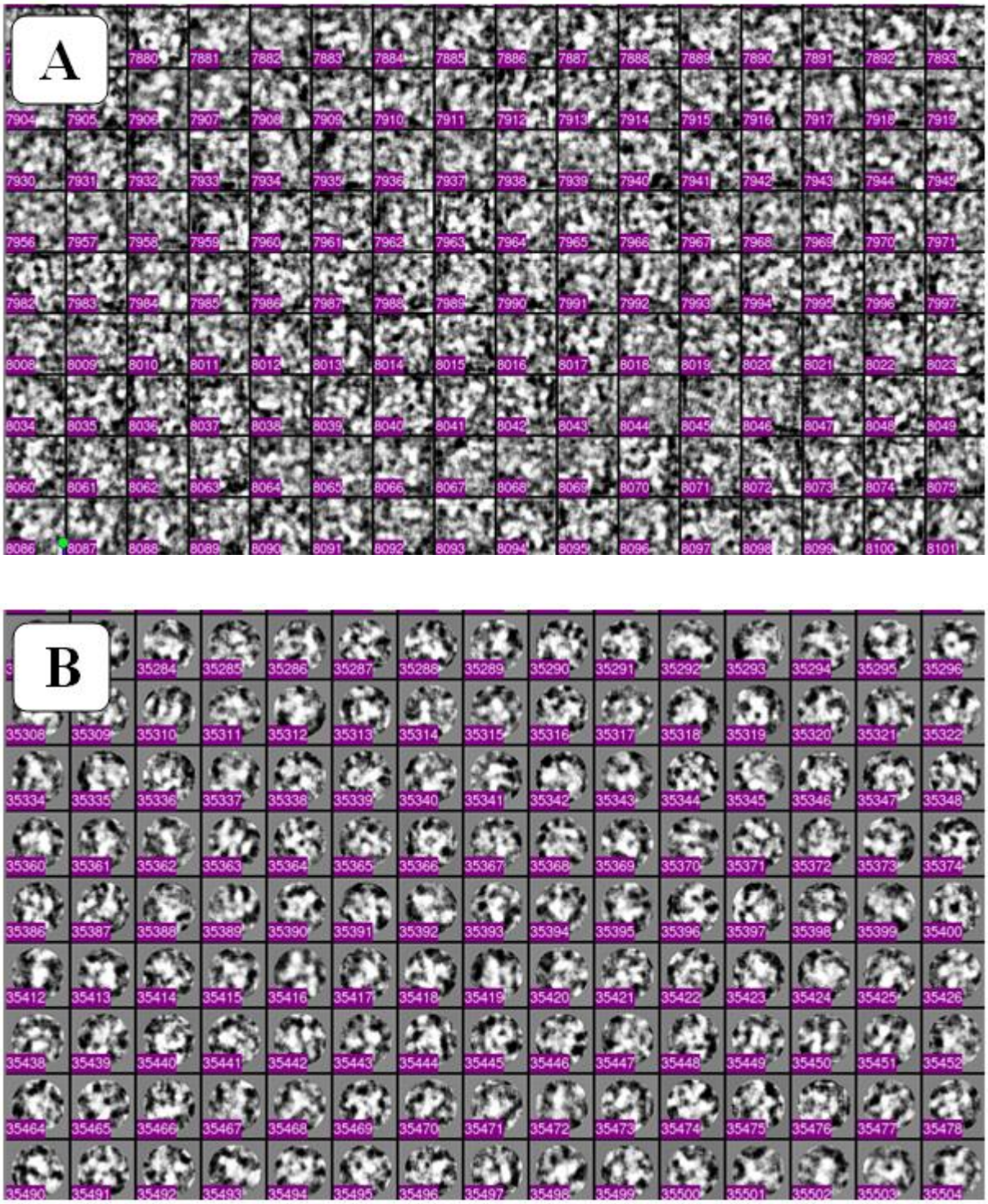


Figure 61: Bad and good particles after alignment. (A) Bad particles which are outliers in terms of size or shape. (B) Good particles.

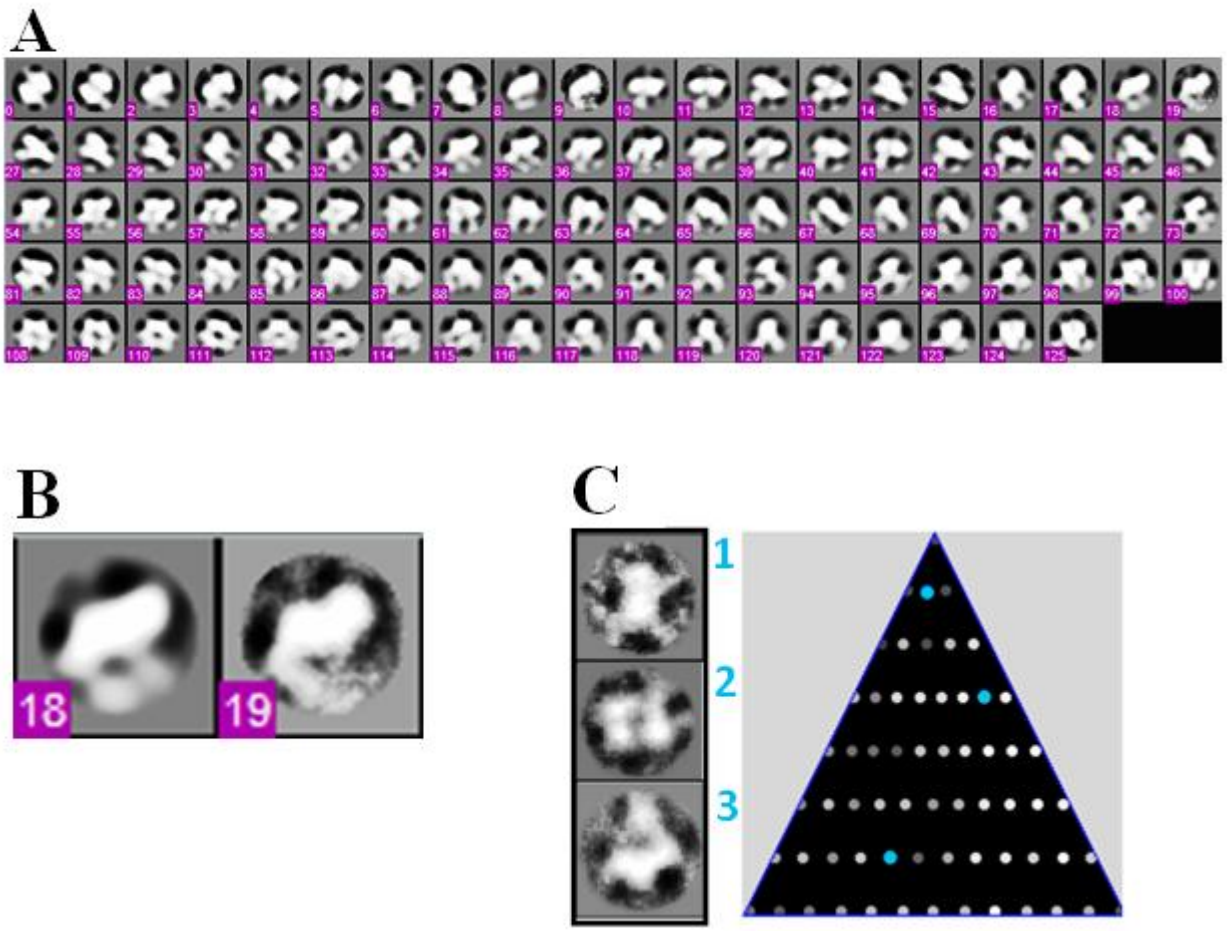


Figure 62: Back-projections of the final 3D structure of CFTR-NHERF1. (A) shows back-projections of the final 3D structure of CFTR-NHERF1 (even-numbered) and their corresponding class averages (odd-numbered). (B) Shows a comparison between a back-projection of the final 3D structure (18) and the projection class average from the raw images (19), 18 and 19 were zoomed out from the back-projections. (C) The Euler angle distribution of classified particle averages. Each dot within the triangle has a different orientation. 1, 2 and 3 are three different classes with different orientations. The three blue dots belong 1, 2 and 3, respectively from top to bottom. The brightness of each dot indicates the number of particles used in the class-average in that orientation. Brighter dots indicate a larger number of particles. The fidelity of the 3D structure was tested here by showing the similarities between class-averages and the final 3D structure as shown in panel (B). Other ways can be used to test the quality of the final 3D model e.g. using a variety of different start models (e.g. a sphere). In addition, using of FSC plots as was used in this study.

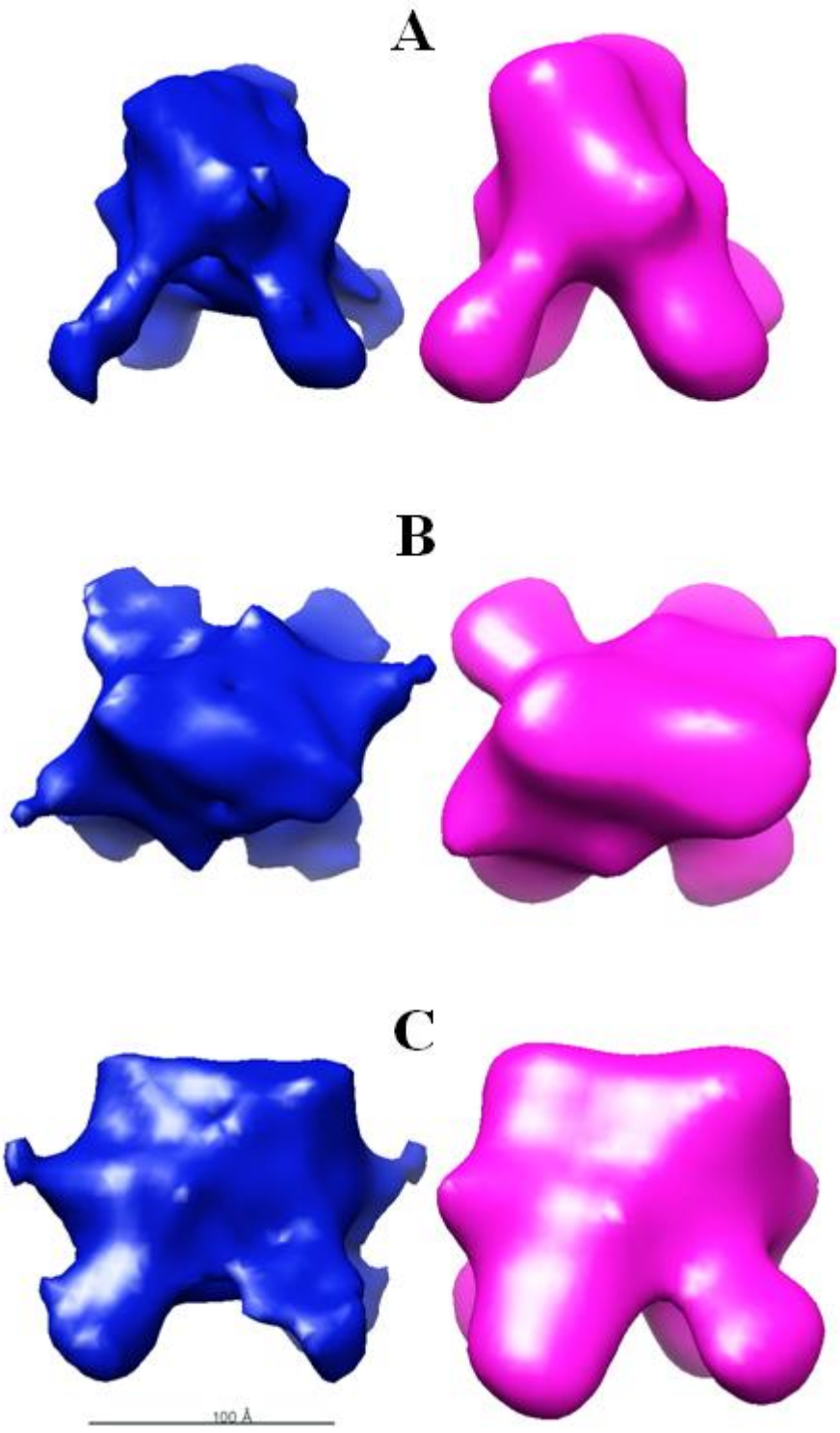


Figure 63: The final 3D structure CFTR-NHERF1. (A) Pink model presents the final 3D structure before the low-pass and high-pass filtration procedure. Blue model presents the final 3D structure after the low-pass and high-pass filtration procedure; more structural details were obtained after the filtration. A, B and C are front, top and side views, respectively. The scale bar is 100 Å.

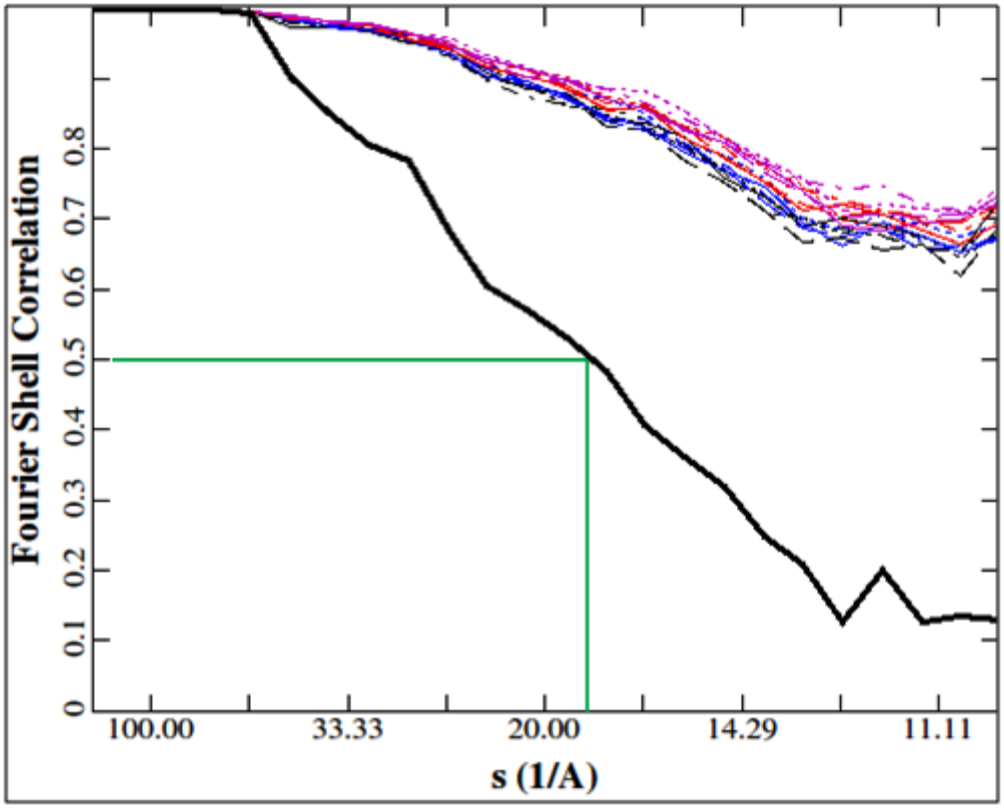


Figure 64: FSC plots for the (n)th versus the (n+1)th iterations refinement in the generation of the three dimensional reconstruction for CFTR-NHERF1 (coloured lines). These show convergence to a final refined model and do not represent a resolution test. The thick black line shows the FSC calculated between two 3D structures generated from two subsets of the data. The green lines indicate the resolution (Å) estimated at FSC = 0.5.

3.2.2.1.1 Analysis of the 3D Structure of CFTR-NHERF1 PDZ $1^{(+)}$

Several 3D models of current ABC structures were fitted into the 3D map of CFTR-NHERF1. The bacterial ABC lipid flippase, MsbA (lipid A export ATP-binding/permease protein) (Ward et al, 2007) (PDB code is 3B5W) was the best atomic model that fitted into the map with a correlation coefficient of 0.8879 (ranging from 0-1) and 252 atoms of 1144 outside the map. The atomic model of NHERF1 PDZ 1 (Karthikeyan et al, 2001) (PDB code is 1G9O) was fitted in the 3D map with a correlation coefficient of 0.8145 and one atom outside. Two copies of MsbA and NHERF1 were inserted into the map (Figure 65) indicating that a possible 3D structure of a homodimer of CFTR was generated. Furthermore, Chimera was used to estimate the map volume in cubic Angstroms and it was estimated

to be 744,000 $Å^3$. To estimate the molecular weight of CFTR-NHERF1 complex from the map volume, an approximate value for globular proteins of 1.21 cubic Angstroms per Dalton was used (Harpaz 1994). The estimated molecular weight was 614.8 kDa. 360 kDa is equivalent to two CFTR molecules, 254 kDa is expected to be the mass of two molecules of NHERF1, glycosylation and detergent (Zhang et al, 2009). This value is consistent with the molecular weight of dimeric CFTR 3D structures that were generated using Single Particle Analysis (Awayn et al, 2005; Zhang et al, 2009).

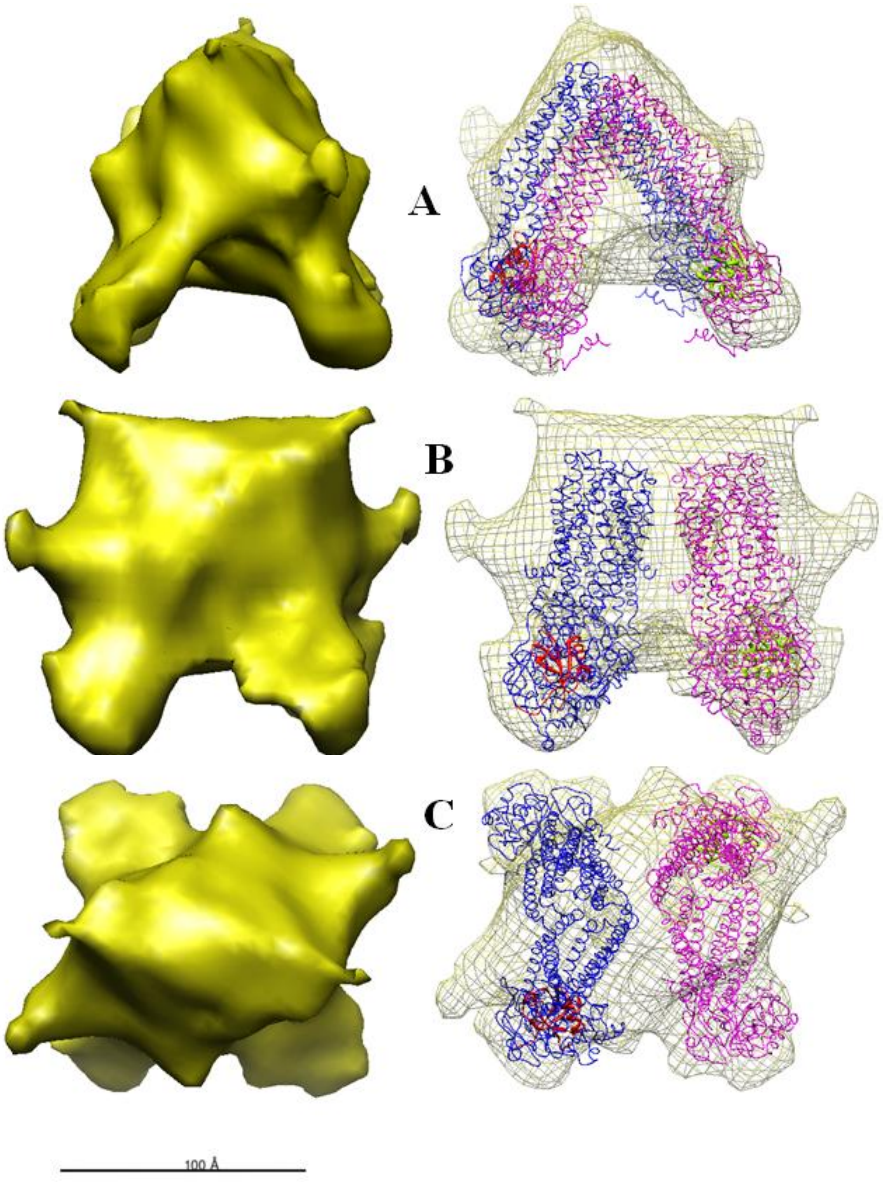


Figure 65: The 3D map of CFTR-NHERF1. A, B and C front, side and top views, respectively. Pink and Blue models are MsbA atomic structures. Yellow and Red models (NHERF1 atomic structures).

The 3D structure of CFTR-NHERF1 showed a new conformational state (an open state as shown in Figure 66) that was not observed before. As far as we are aware, this is the first 3D structure of CFTR with an open V shape conformation. We hypothesize that CFTR (in the presence of NHERF1) can undergo a wide conformational change which may be essential for regulation of ion transport. A similar open V shape of MsbA was reported by Ward et al (2007). The NBDs of CFTR are thought to be responsible for the opening and closing of the channel (along with phosphorylation of R-domain) (Sheppard and Welsh, 1999 ; Holland et al, 2003; Hwang and Sheppard, 2009). On the other hand, the TMDs compose the channel pore which is embedded in the lipids bilayer. From the 3D structure of CFTR-NHERF1, the locations of the TMDs and NBDs can be predicted (Figure 66). NBDs are linked directly to the conformational change and they are separated by a large space forming a V shape.

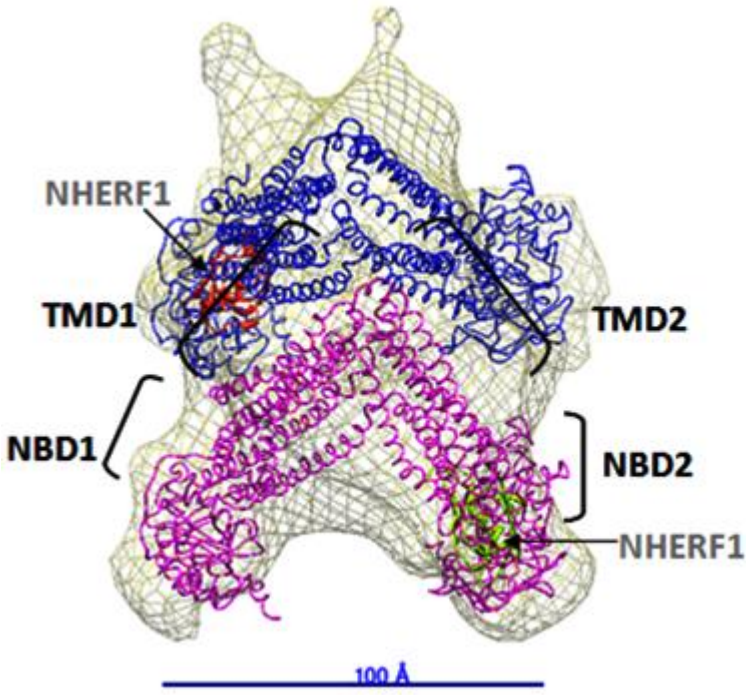


Figure 66: Locations of TMDs and NBDs. One copy of MsbA (Pink) used to locate the four domains. The yellow and red models are NHERF1. The blue model is the second copy of MsbA.

In addition, the location of the C-terminus (bound to NHERF1 with a high affinity) can be another indication to the location of NBD2. A difference map between the PDB file and the 3D map of CFTR-NHERF1 was generated to try to locate NHERF1 but the difference map was not very useful.

NHERF1 fitted with low correlation in the difference map and this may be a result of the small size of NHERF1. The association between NBD2 and the C-terminus was observed in an other 3D structure of CFTR (Awayn et al, 2005) in which a 1.8 nm diameter Ni-NTA nanogold lable binds to the 10His-tag placed at the C-terminus of the protein. The location of the Ni-NTA nanogold sphere was used to predict the postion of the C-terminus. In other experiments on a construction of CFTR, the location of the C-terminus was found to be close to the expected postion of NBD2 (Zhang et al, 2011). The 3D structure of CFTR-NHERF1 also revealed two additional structures adjacent to the TMDs which may due to detergent (Figure 67). These two structures were not observed in the previously mentioned 3D structures of CFTR which may be a result of their lower resolution compared to ~ 18 Å (obtained in this SPA). Another interesting region of the structure was a large bump attached to the expected position of NBD1. This bump could be the R-domain, which links between NBD1 and TMD2 (Figure 67). Slices of the internal structure of the CFTR-NHERF1 desity map showed the proposed CFTR domains locations as shown in Figure 68. The highest value of the density map (dark blue) indicated the potential locations of the four domains. This could be evidence for the reliability of this structure.

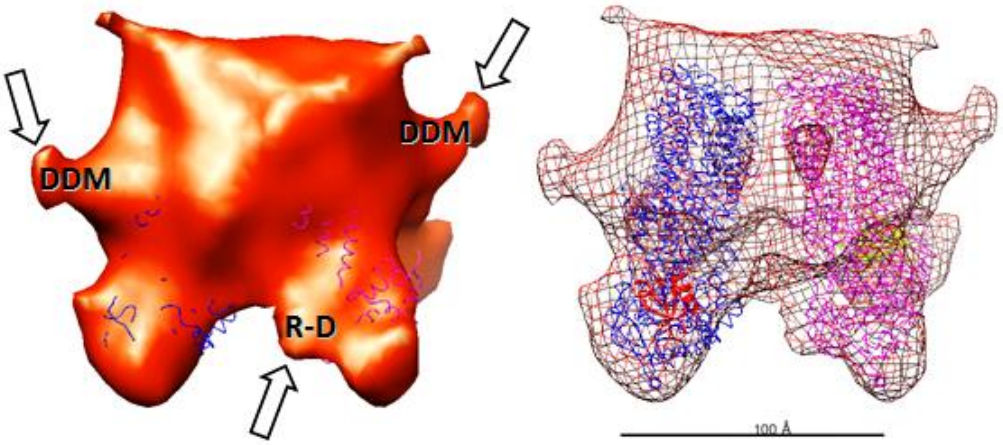


Figure 67: Side view of the 3D map of CFTR-NHERF1 shows proposed locations of R-D (R-domain) and DDM detergent density (as pointed by arrows). Mesh model shows the internal structure. Blue and pink models are two copies of the open conformational state of MsbA (Ward et al, 2007) (PDB code is 3B5W).

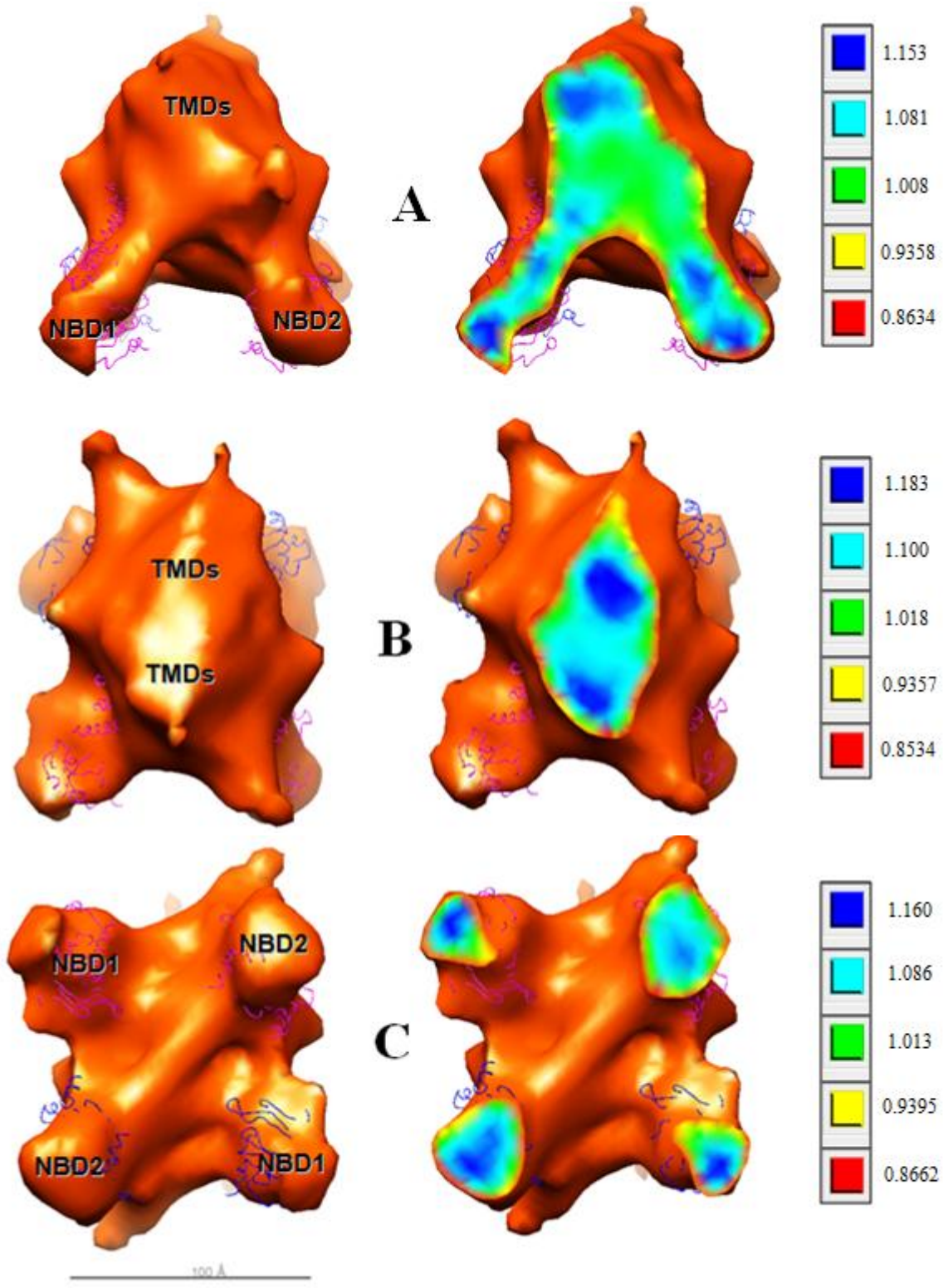


Figure 68: Slices showing the density values for the internal structure of the CFTR domains. The colour box indicates increasing density from dark blue (highest) to red (lowest).

3.2.2.2 A Low Resolution Structure of Isolated CFTR

A sample of isolated CFTR (without NHERF1 or nucleotide) was used for SPA as described in section 3.2.2.1 with some differences. Firstly, 10 micrographs were collected using the Tecnai G2 Spirit (Biotwin) transmission electron microscopy. The micrographs recorded at 13000 magnification (5.8 Å/pixel). In addition, 9056 good particles from a selection of 10651 were used for SPA. 20 averaged classes from 86 classes (with top, side and partial views) were used to generate the start 3D model as shown in Figure 69 C. Figure 69 shows some stages in the CFTR structure processing. 16 iterations were done to generate the final 3D structure of isolated CFTR with a final resolution of ~ 27 Å estimated from the FSC plot shown in Figure 70. This may be an optimistic estimation as the FSC test has been argued to be. The bacterial 'open' form of the lipid flippase, MsbA (Ward et al, 2007) (fitted into the 3D map of NHERF1-CFTR) was fit in the 3D map of isolated CFTR but a very low correlation coefficient was obtained. Another atomic model of MsbA which represented the closed conformation (Ward et al, 2007) (PDB code is 3B5X) was fitted well with a correlation coefficient of 0.8007 and 188 atoms of 1144 were outside the map. Figures, 71, 72 and 73 show different views of the isolated CFTR structure. The volume of the isolated CFTR map was estimated to be 714,000 $Å^3$. To estimate the molecular weight of the isolated CFTR from the map volume, an approximate value for globular proteins of 1.21 cubic Angstroms per Dalton was used (Harpaz 1994). The estimated molecular weight was 590.6 kDa. 360 kDa is roughly equivalent to two CFTR molecules, 230 kDa is expected to be the mass of glycosylation and detergent (Zhang et al, 2009). This value is consistent with the molecular weight of dimeric CFTR 3D structures that were generated using Single Particle Analysis (Awayn et al, 2005; Zhang et al, 2009). The molecular weight difference between the CFTR-NHERF1 (614.8 kDa) and isolated CFTR (590.6 kDa) is small. This difference is approximately compatible with two NHERF1 masses (~ 20 kDa). But it is debatable whether the calculations are sufficiently accurate to allow such an interpretation. For example, the volume will vary considerably depending on the threshold chosen. However, the 3D structure of isolated CFTR showed a closed conformation compared to the V shaped open structure of CFTR-NHERF1. This difference gave more emphasis that the interaction between CFTR and NHERF1 is being detected and moreover has a

significant role in terms of the opening and closing of the CFTR channel. This possible role will be discussed in the Conclusion section.

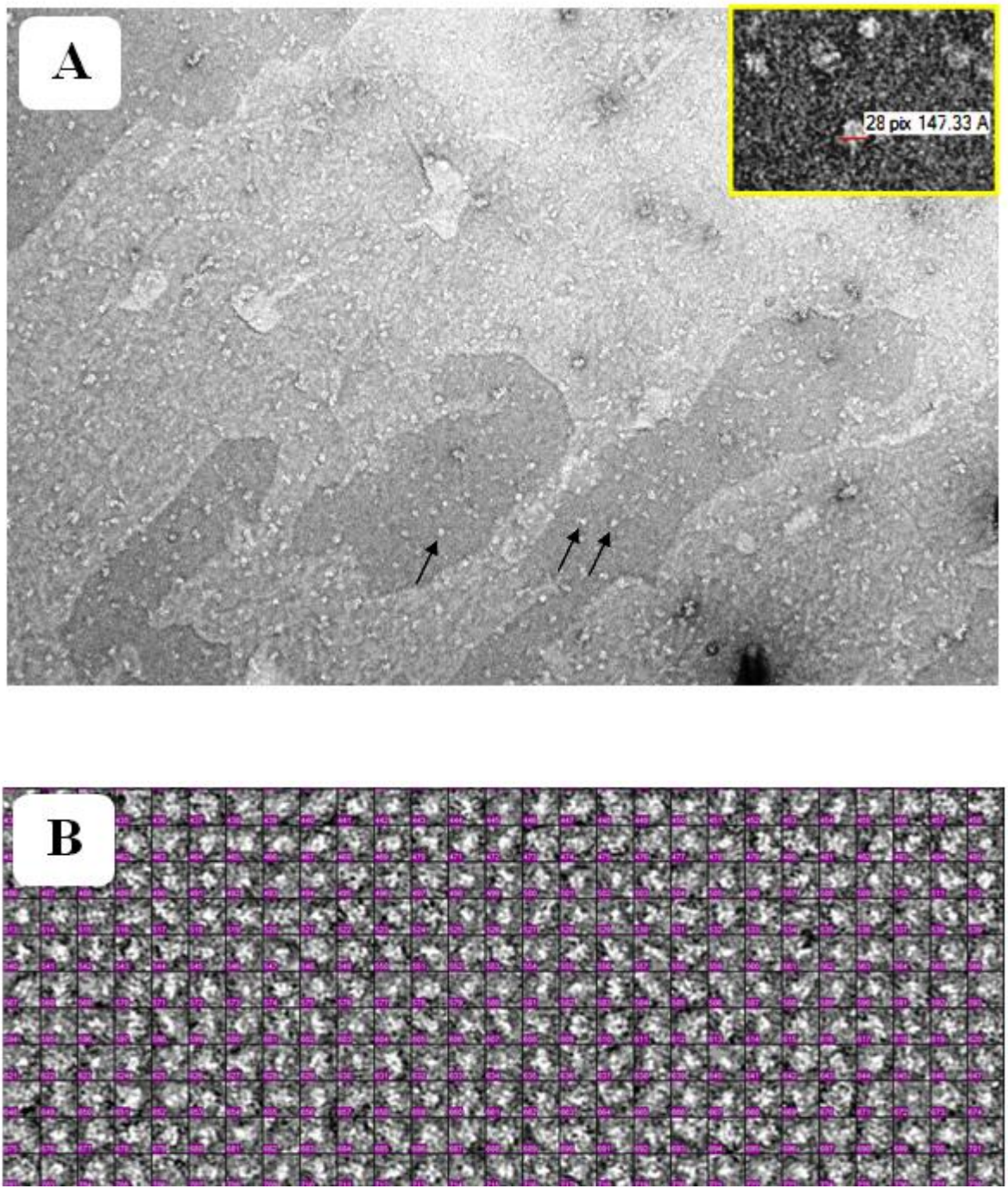
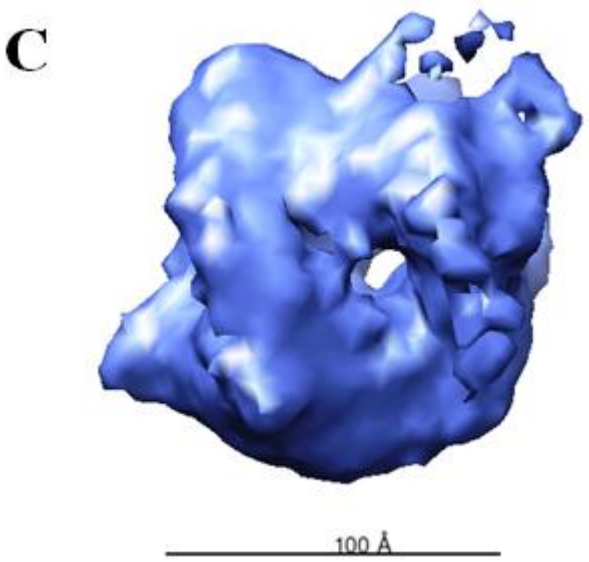


Figure 69: A: Particles of isolated CFTR. The black arrows point to one single particle of ~15 nm. The yellow box contains a measurement of a single particle of ~147 Å. **B: Selected particles from isolated CFTR.** Box size is 48 x 48 pixels.



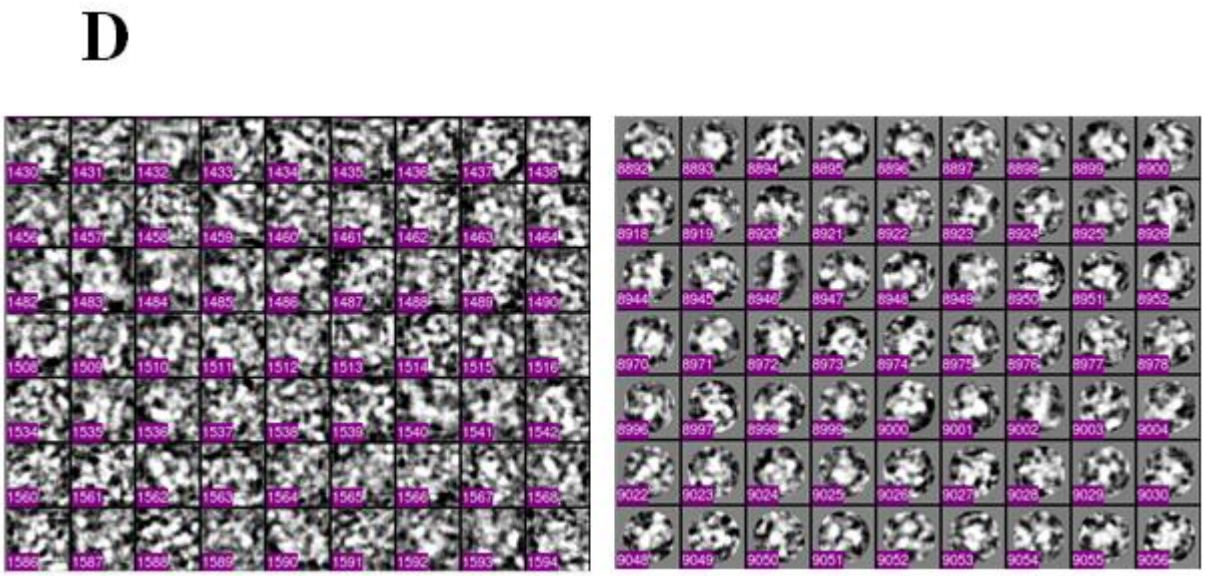
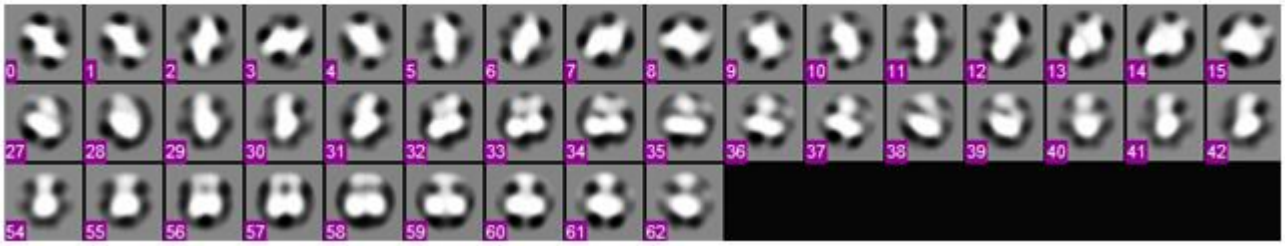


Figure 69: C: Initial model of isolated CFTR that was generated by 20 averaged classes. The scale bar is 100 Å. **D: Bad and good particles after alignment.** Bad particles (on the left) which are outliers in terms of size or shape and good particles (on the right).

E



F

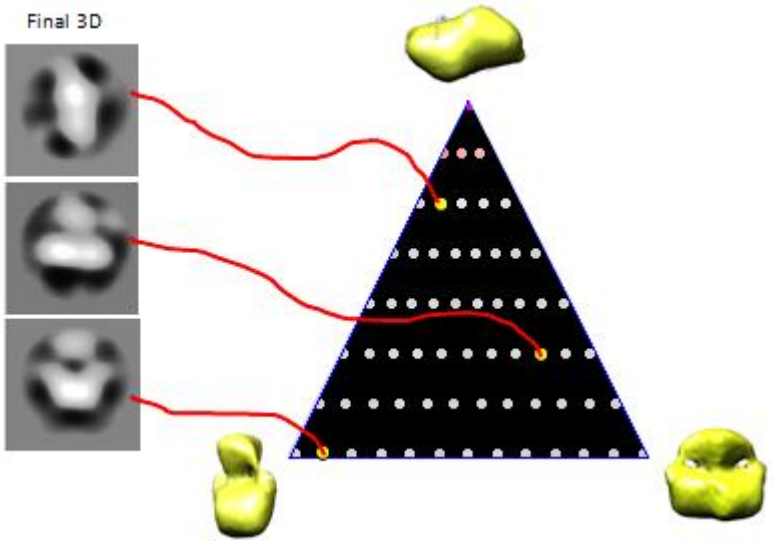


Figure 69: E: Back-projections of the final 3D structure of CFTR-NHERF1 PDZ 1$^{(+)}$. Back-projections of the final 3D structure of CFTR-NHERF1 PDZ 1$^{(+)}$ (even-numbered) and their corresponding class averages (odd-numbered). **F: The Euler angle distribution of oriented particle averages.** Each dot within the triangle represents a population with a different orientation. The three yellow dots are linked (by the red lines) to their corresponding orientation. The brightness of each dot indicates the number of particles used in that orientation-average. Brighter dots indicate a larger number of particles. Raw data is the oriented particle averages used to build the final 3D structure. The 3D structures (In yellow with different views) represent the rotation of the 3D structure equivalent to the angles at each corner of the Euler triangle.

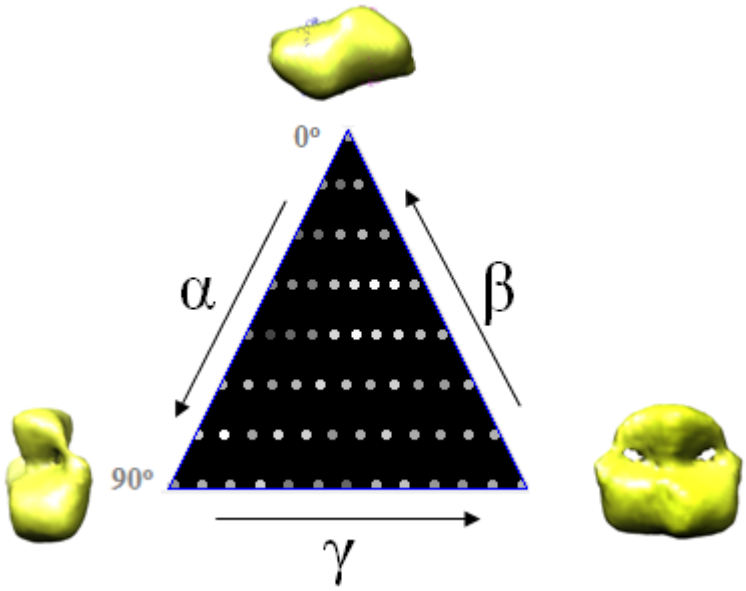


Figure 69: G: Example of the orientation distribution in Euler space represented by a triangle. There are three Euler angels (α, β and γ) used to determine the location and orientation of a rigid object in space axis (X, Y and Z). The triangle shows the distribution of these three angles. The brightness of each dot indicates the number of particles used in that orientation. Brighter dots indicate a larger number of particles.

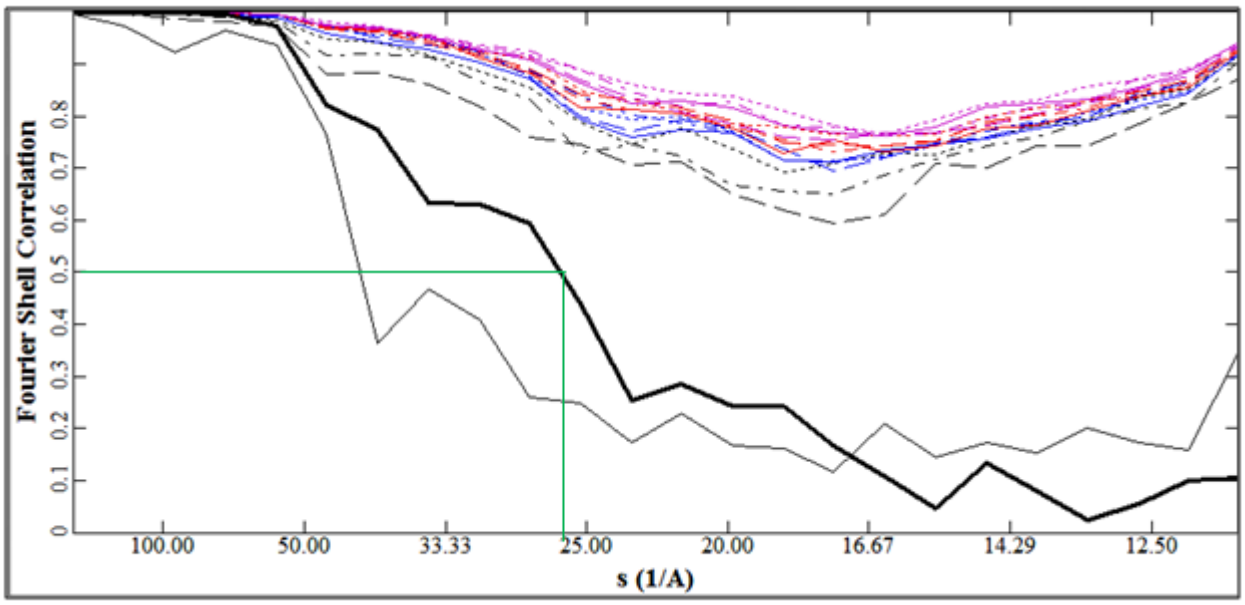


Figure 70: FSC plots for the (n)th versus the (n+1)th iterations of refinement in the generation of the three dimensional reconstruction for isolated CFTR (coloured lines). These show convergence to a final refined model and do not represent a resolution test.The thick black line shows the FSC calculated between two 3D structures generated from two subsets of the data. The green lines indicate the resolution (Å) estimated at FSC = 0.5.

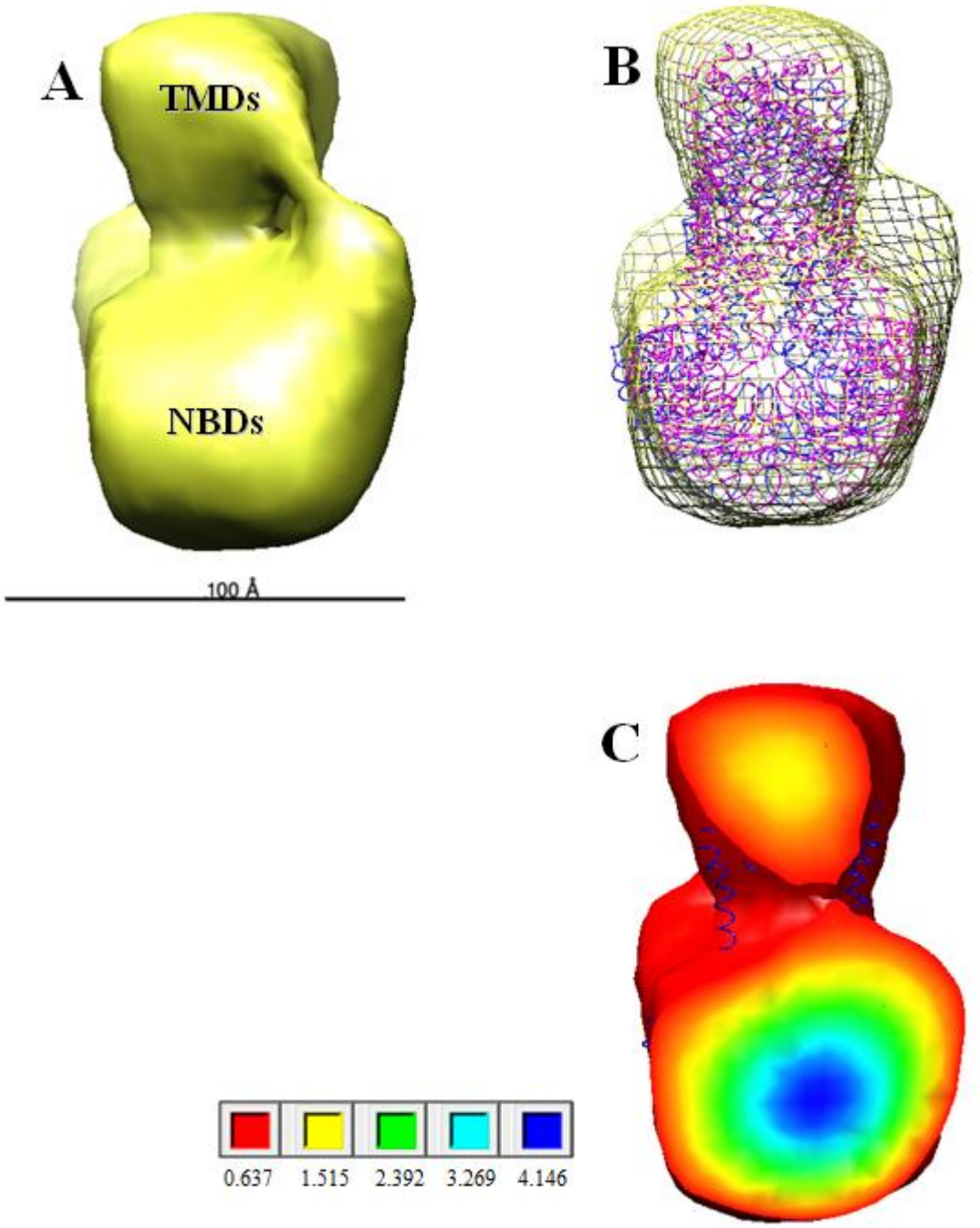


Figure 71: Front views of the 3D structure of isolated CFTR. A, Surface panel of the 3D structure shows possible locations of the CFTR domains. B, Mesh panel shows two fitted copies of MsbA (pink and blue). C, Slice shows the density values for the internal structure. Colour box indicates increasing density from dark blue (highest) to red (lowest).

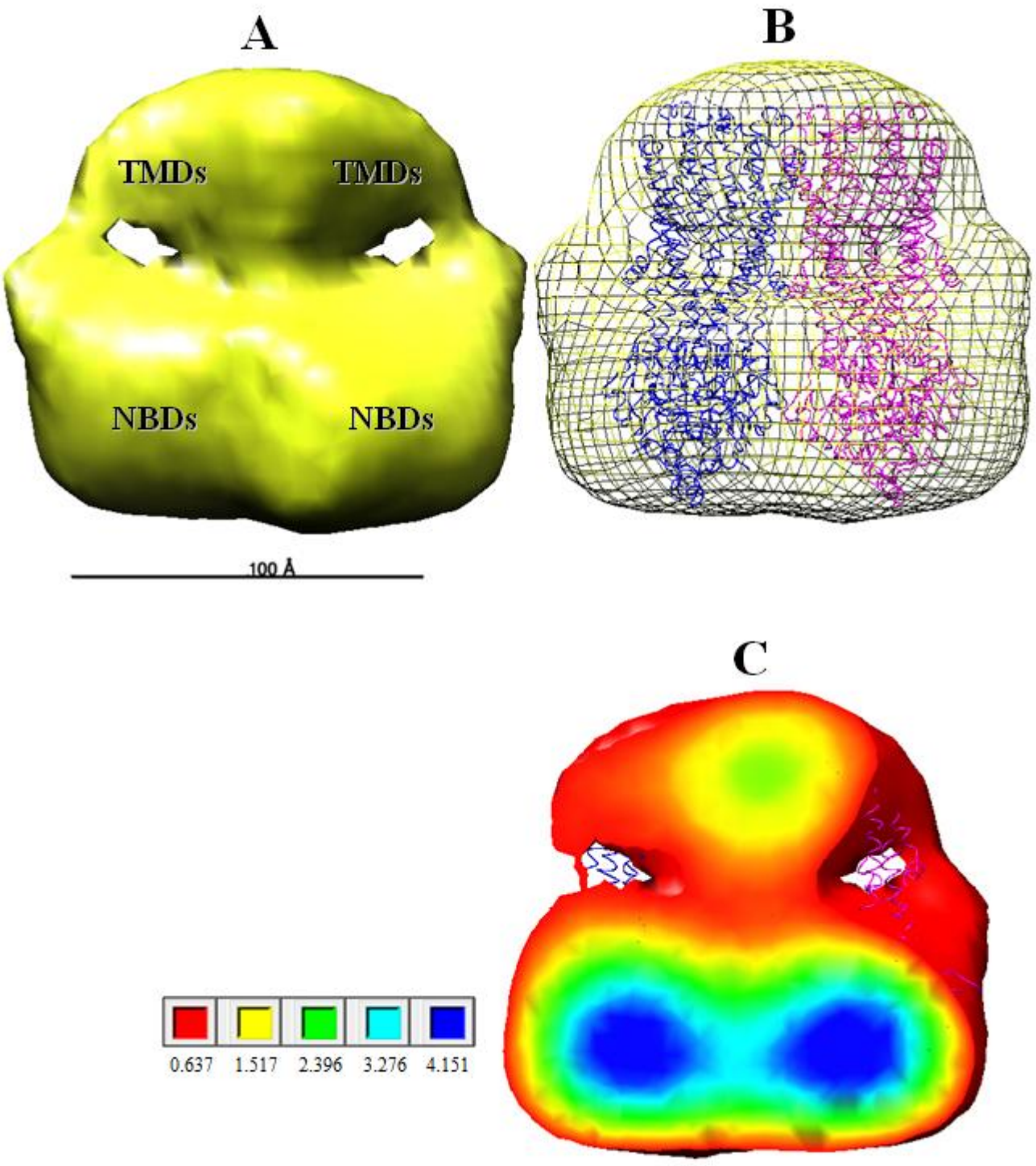


Figure 72: Side views of the 3D structure of isolated CFTR. A, Surface panel of the 3D structure shows possible locations of the CFTR domains. B, Mesh panel shows two fitted copies of MsbA (pink and blue). C, Slice shows the density values for the internal structure. Colour box indicates increasing density from dark blue (highest) to red (lowest).

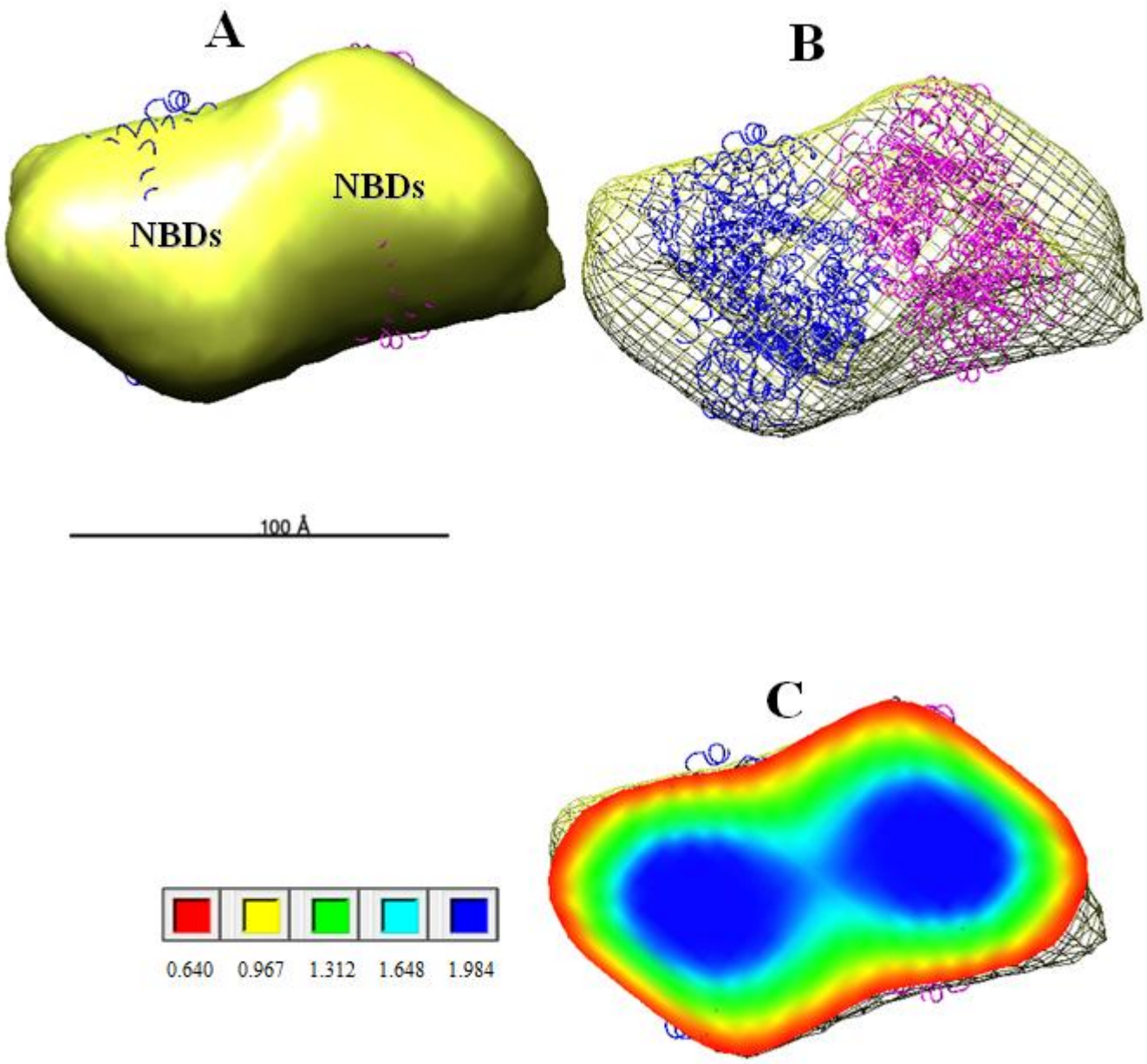


Figure 73: Bottom views of the 3D structure of isolated CFTR. A, Surface panel of the 3D structure shows possible locations of the CFTR domains. B, Mesh panel shows two fitted copies of MsbA (pink and blue). C, Slice shows the density values for the internal structure. Colour box indicates increasing density from dark blue (highest) to red (lowest).

3.2.2.3 Structural Analysis of the CFTR-NHERF1 Complex and Isolated CFTR

The two 3D structures (CFTR-NHERF1 and isolated CFTR) revealed two different conformations. The first structure of CFTR-NHERF1 showed a conformation interpreted as a widely-separated NBD state.

3.2.2.3.1 Distance Measurement: The distance between the assumed NBD1 and NBD2 was measured (Figure 74). The distance between the NBD1 and NBD2 of one molecule of CFTR was estimated to be 73 Å which was similar to the distance between the NBD1 and NBD2 of two molecules of CFTR (70 Å).

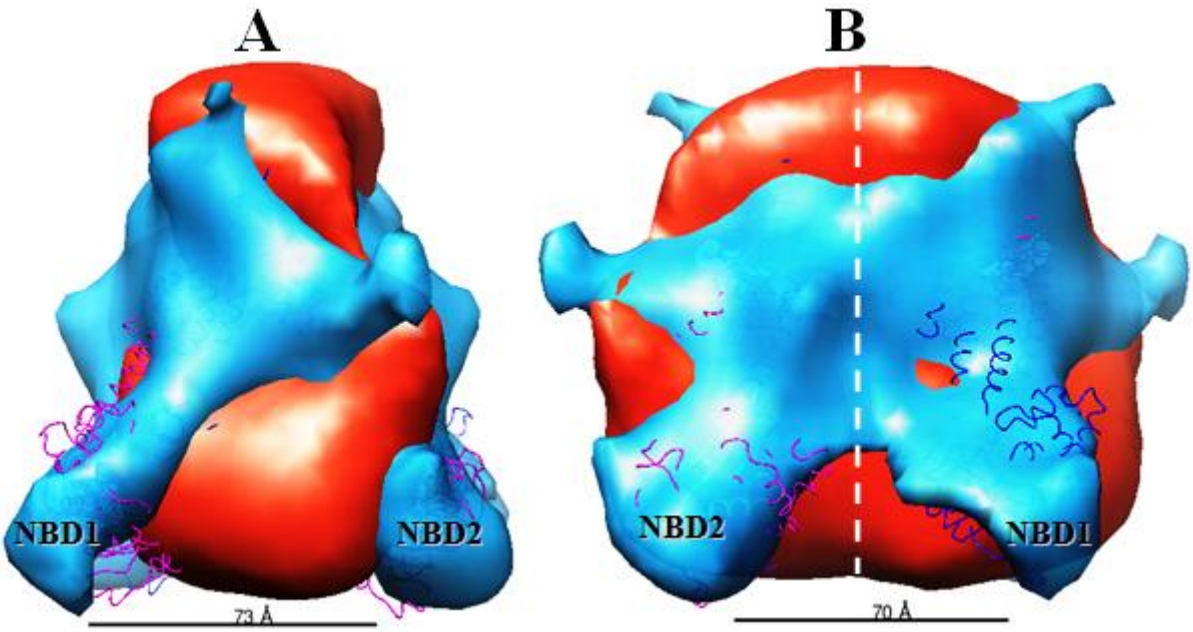


Figure 74: A distance measurement between the NBD1 and NBD2 of one CFTR (A) and the NBD1 and NBD2 of two adjacent CFTR molecules in the dimer (B). Blue 3D structure is CFTR-NHERF1. Red structure is the isolated CFTR. The dashed white line indicates the approximate interface between the two molecules of CFTR.

Borbat et al (2007) measured the amplitude of conformational motion in MsbA and they suggested that (> 70 Å) is a distance required for a complete dissociation of the dimer. In contrast, these electron microscopy measurements suggest that a stable homodimer of CFTR was formed (in the presence of NHERF1) and this dimer was stabilized (mostly) by the TMD connections of both molecules. This hypothesis agrees with the crystal structure of the inverted V shape of MsbA (which was fitted into the 3D map of CFTR-NHERF1). In this structure of MsbA, the TMD helices (TM4 and TM5), and the intracellular loop 2 (ICL2) of two monomers cross over to the other half of the protein to form a homodimer (Figure 75) (Ward et al, 2007). This association between two monomers was also observed in SAV1866 (Dawson and Locher, 2006) and was defined as intertwined interfaces which formed by an interaction of two subunit chains. These interfaces include hydrophobic regions and their interaction is required for stability (Larsen et al, 1998).

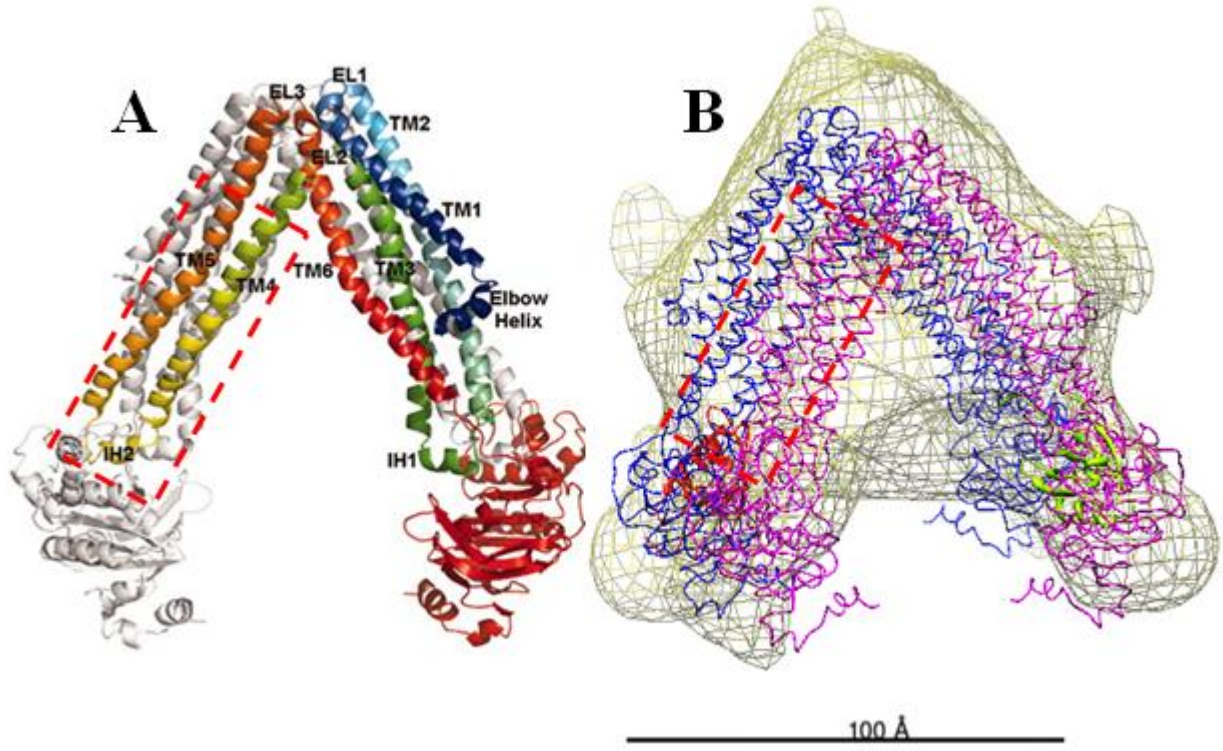


Figure 75: The proposed dimerization contacts in the 3D structure of CFTR-NHERF1. A, The open shape of MsbA structure showed intertwined interfaces between TM4, TM5 and ICL2 of one monomer (coloured) with the other (white). TM4/TM5/ICL2 (yellow and orange) (Ward et al, 2007). B, The 3D structure of CFTR-NHERF1 shows the proposed intertwined interfaces between the TMD1 (pink) and the TMD1 (blue). The red dashed lines indicate the intertwined interfaces. Pink and blue models are two copies of MsbA. Yellow and red models are two copies of NHERF1.

3.2.2.3.2 Structural Similarity: The structure of the isolated CFTR showed reasonable similarities in terms of shape and size to the 3D structure of isolated CFTR that was generated by (Awayn et al, 2005) (as shown in Figure 76).

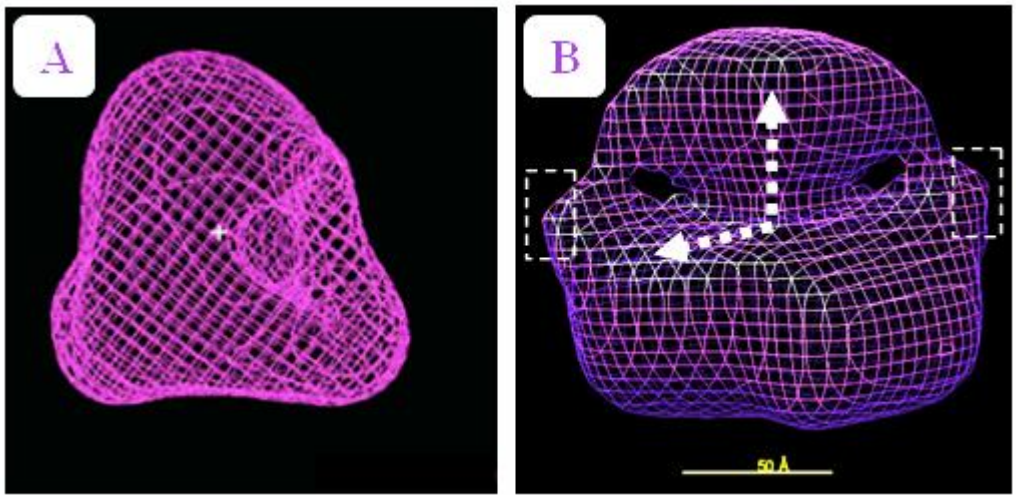


Figure 76: A comparison between the 3D map of nucleotide-free CFTR at 35 Å resolution (Awayn et al, 2005) (A) and the isolated CFTR structure presented in this thesis (B). Figure (A) was obtained from (Awayn et al, 2005). The scale bars is 50 Å. Both structures showed a closed NBDs state. The isolated CFTR structure reported herein with the resolution of ~ 27 Å revealed two structural components (white dashed boxes) which may be detergent. In addition, a twisted area (white dashed arrows) appeared in the CFTR structure.

The structure of the isolated CFTR showed similarities to the 3D structures of isolated CFTR that was generated by (Zhang et al, 2009) (as shown in Figure 77). The similarities between the isolated CFTR structure and the 3D structure of CFTR (in LPG) in terms of shape and size gives another support for this analysis. Furthermore, this very close similarity between the two structures despite that they were purified with two different detergents (DDM and LPG) suggested that there was no significant effect of detergent on the overall shape. In addition, The CFTR structure (in LPG) was obtained using electron cryomicroscopy. On the other hand, the isolated CFTR structure in this study was obtained using negatively stained specimens. These two different techniques were interestingly able to generate similar 3D structures.

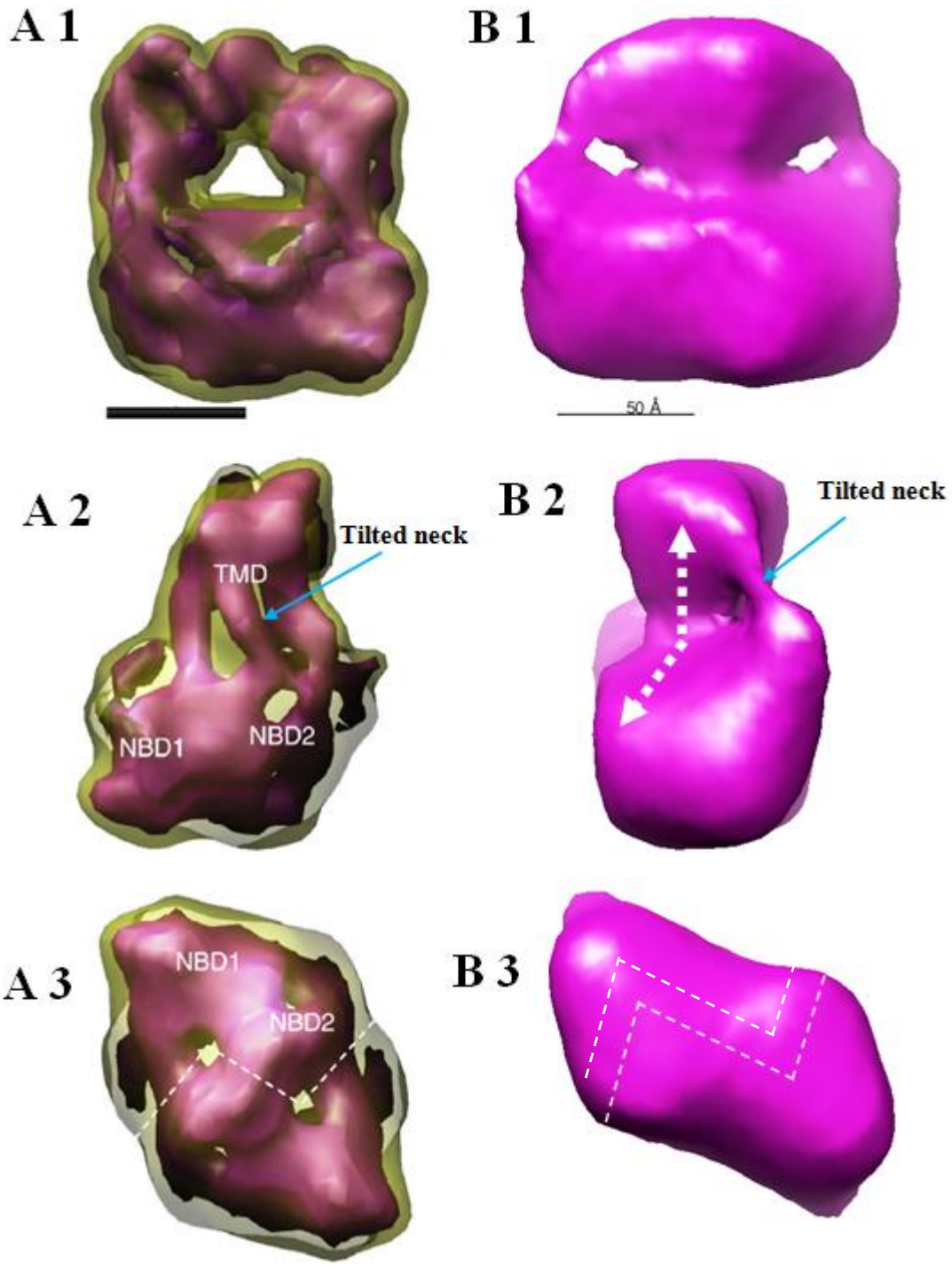


Figure 77: A comparison between the 3D map of nucleotide-free CFTR (35 Å) (A 1,2 and 3) and the isolated CFTR structure (B 1,2 and 3). Figure (A) was obtained from (Zhang et al, 2009). Both scale bars are 50 Å. The two 3D structures showed a roughly cylindrical shape with a narrow top (compared to bottom) which is expected to be the TMD region as revealed in the previous comparison (Figure 76). A tilted neck in the TMD regions appeared in the two structures (A2 and B2). Two copies

of SAV1866 (closed NBDs) were fitted into the structure of (Zhang et al, 2009) indicating a dimerization state.

3.2.2.3.3 The C-terminus Location: The location of the C-terminus in the CFTR-NHERF1 structure was assumed (as mentioned in section 3.2.2.1.1) to be in the cytoplasmic area and close to NBD2. This assumption was based on the study of (Awayn et al, 2005) by which the location of a 1.8 nm diameter Ni-NTA nanogold label was close to NBD2 (Figure 78 B). Another study of gold labeled-CFTR agrees with this assumption (Zhang et al, 2011). A difference map between the structures of MsbA and the 3D map of the structure of CFTR-NHERF1 PDZ $1^{(+)}$ was generated. NHERF1 PDZ $1^{(+)}$ fitted with low correlation in the difference map. However, NHERF1 molecules fitted very well into the 3D map space where the MsbA structures were fitted (as shown in Figure 78A and as mentioned in section 3.2.2.1.1).

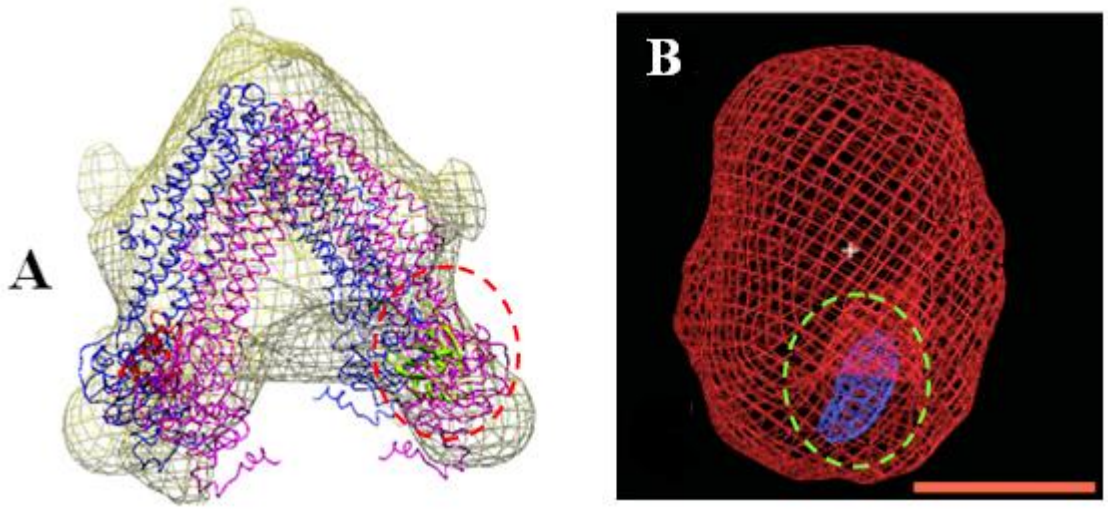


Figure 78: The proposed location of the C-terminus in the CFTR-NHERF1 structure (red dashed circle, panel A). Yellow model is one molecule of NHERF1 (PDB code is 1G9O) which binds to NBD2 (replaces the gold label in panel B). Panel B shows the gold label (blue sphere) which binds to the C-terminus of CFTR and locates close to NBD2. Panel B was obtained from (Awayn et al, 2005). The scale bar is 50 Å. The PDB code of MsbA structures (blue and pink models in panel A) is 3B5W.

3.2.2.3.4 CFTR channel gating: The opening and closing of the CFTR channel is regulated by the NBDs (Sheppard and Welsh, 1999; Gadsby and Nairn, 1999). Several models have been proposed to describe the channel gating cycle. These models were based on the assumption that the closed NBDs cause the opening of the channel and the open NBDs cause the closing of the channel (Sheppard and Welsh, 1999; Gadsby and Nairn, 1999, Zeltwanger et al, 1999; Sheppard, 2012). Figure 79 (A2 and 3) revealed two separated high densities (in the expected TMDs regions). These two density areas may be a closing state of the TMDs. On the other hand, Figure 79 (B2 and 3) showed two interfaced density areas (appeared as a one big density) which may be a partially open state of the channel.

3.2.2.3.5 Dimerization of CFTR: The two structures (CFTR-NHERF1 PDZ $1^{(+)}$ and isolated CFTR) that were obtained by SPA suggest a dimerization of CFTR. Several other SPA studies also suggested that CFTR forms a dimer (Awayn et al, 2005; Zhang et al, 2009; Zhang et al, 2011). In addition, it was reported that CFTR was purified from detergent as a dimer (but crystallized as a monomer form) (Riordan, 2008). Biochemical studies of CFTR at the cell surface have suggested that the dimerization of CFTR is required for CFTR-protein interactions (Ramjeesingh et al, 2003; Li et al, 2004). The size of CFTR particles in freeze-fracture electron micrographs showed that CFTR is a dimer with a diameter of 9 nm (Eskandari et al, 1998). Zerhusen et al (1999) suggested that two CFTR molecules interact together to form a single channel for chloride ion transport. On the other hand, several biochemical studies revealed that CFTR functions as a monomer. Western blots studies showed that CFTR was detected as a monomer (Marshall et al, 1994; Chen et al, 2002). Ramjeesingh et al (2003) have found that CFTR was purified as a monomer by gel filtration. Furthermore, single channel electrophysiological studies showed that monomric CFTR was able to regulate the channel activity (Ramjeesingh et al, 2001). Interestingly, to date, all X-ray crystal structures of ABC transporters were obtained as a monomer, whilst studies by EM on noncrystalline structures of ABC transporters showed a dimerization (Zhang et al, 2011). However, it was suggested that the CFTR channel activities are controlled by a monomer form of CFTR but CFTR-protein interactions and trafficking are regulated by a dimer of two molecules of CFTR (Stratford and Bear, 2006).

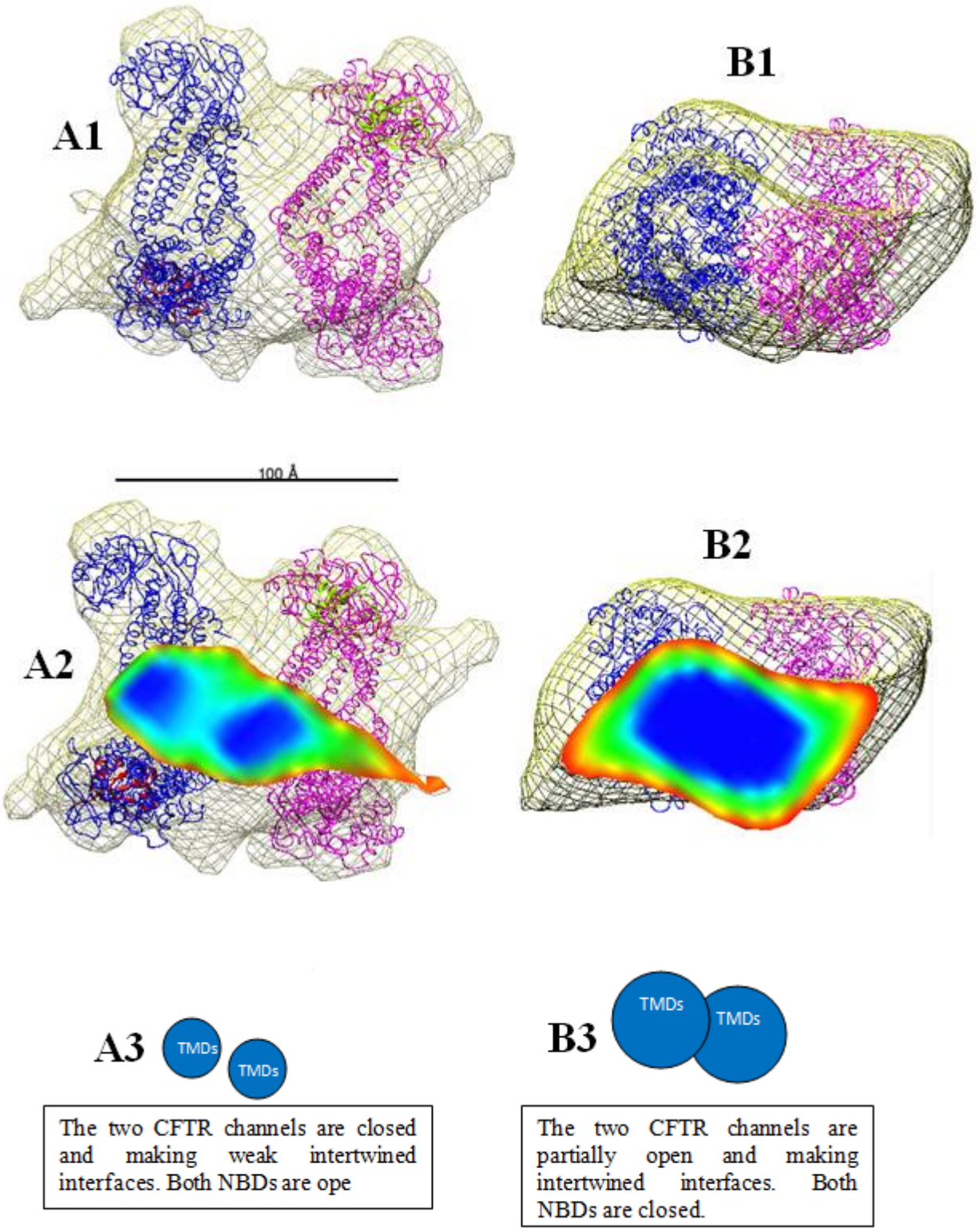


Figure 79: Slices of MSD regions of the CFTR-NHERF1 structure (A1,2 and 3) and the isolated CFTR structure (B1, 2 and 3). Both structures show top views.

Conclusion

The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) is an ion channel that regulates the secretion of Cl^- in epithelial cells. Beside this function, CFTR is recognized as a regulator of several transport proteins, including the epithelial sodium channel (ENaC), K^+ channels, ATP-release mechanisms, anion exchangers, sodium-bicarbonate transporters, and aquaporin water channels (Guggino and Stanton, 2006). Mutations in the CFTR gene cause cystic fibrosis which is linked to severe symptoms and an average age of death of ~ 30 years (Davis, 2006). The study of CFTR function and structure, and its association with several signaling pathways may lead to drug discovery. Transmembrane proteins are difficult to crystallize because of their hydrophobic regions that lead to aggregation in aqueous solutions, hence two strategies were used in this project. Firstly, by the study of small soluble fragments of CFTR which were expected to be in the cytoplasm. Secondly, by Single Particle Analysis of the full-length CFTR it was possible to study this protein in the non-crystallized state. In this study, the structure and function of small constructs of CFTR, the C-terminus^{42aa}, the C-termiuns^{61aa}, the C-terminus^{R-A}, the C-termini of Shark and Killifish (and full-length CFTR) were studied. One of most well studied CFTR-interacting molecules (NHERF1 PDZ 1) was involved in some of these experiments. The results suggested new possible roles of NHERF1 in the CFTR physiological mechanism.

The expression and purification of the CFTR C-terminus^{42aa} allowed a high concentration of pure protein for crystallization. The C-terminus^{42aa} was purified as a soluble protein in agreement with the prediction that this domain is located outside the transmembrane. The CD and NMR results of this protein revealed an overall random coil structure with two flexible sides. This flexibility of the protein may explain why several crystallization trials were not successful. However, the flexibility of the C-terminus of CFTR is linked to a well-known role of this domain as a protein-protein interactor. To overcome the possible flexibility problem, the scaffold protein (NHERF1 PDZ 1) was co-expressed and co-purified with the C-terminus^{42aa}. Contrary to what was expected, this interaction between the two proteins did not form a more ordered system as illustrated in Table 14. This is possibly because

NHERF1 PDZ 1 only binds a small portion of the C-terminal peptide (Consistent with the X-ray crystal structure). However, the MALDI and SDS-PAGE results of the co-purified proteins showed a strong interaction. This interaction showed a possible posttranslational modification (an interaction with β-mercaptoethanol) as shown in Figure 28.

The extension of the C-terminus length (the C-terminus^{61aa}); the substitution of a big, flexible amino acid Arginine with a small amino acid Alanine (the C-terminus^{R-A}) and the expression and purification of two homologues of human CFTR (the C-terminus of shark and killifish) were all tried to further investigate the C-terminus of CFTR. All these expression and purification trials were successful and soluble proteins were obtained but with low yield. More optimization experiments are needed to have a higher concentration of these constructs suitable for 3D crystal trials.

Samples of the C-terminus^{42aa} were studied by our collaborators (Bozoky et al, as mentioned in section 3.2.1.1.3, and the data (has been submitted for publication). They found that a strong interaction between the C-terminus^{42aa} and the R-domain of CFTR was observed when the R-domain was in the phosphorylated state. These data were presented in the 25th annual North American Cystic Fibrosis Conference (2011), Anaheim, California as a poster presentation and a talk. Other samples of the C-terminus^{42aa} were studied by our collaborators (Mario A. Pagano et al, the University of Padova, Italy, not published). They found that the C-terminus^{42aa} interacted with kinases such as CK2 in *E.coli* and identified a phosphorylated residue (unpublished data). Both results suggested a potential role for the C-terminus^{42aa} in the regulation of the CFTR channel gating by phosphorylation. More recently, the work of the C-terminus^{42aa}-NHERF1 PDZ 1 was selected for a poster presentation and a talk in the 26th annual North American Cystic Fibrosis Conference (2012), Orlando, Florida.

The full-length CFTR-NHERF1 interaction was investigated by pull-down assay experiments. The results revealed that CFTR (in DDM) was able to interact with NHERF1. On the other hand, CFTR did not interact with NHERF1 when LPG was present. These results highlight the role of CFTR environment (lipids or even more the type of the cell that CFTR is located in) on the protein activity. In addition, the CFTR-C-terminus^{42aa} interaction was studied and the result was compatible with the

NMR data that the C-terminus^{42aa} interacted strongly with the R-domain. NHERF1 PDZ $1^{(+)}$ has the carboxyl-terminal extension N95DSLL99 that corresponds to residues 409–413 of human β_2AR. This mutation was designed to allow a homodimer formation. The pull-down assay, surface plasmon resonance (SPR) and multi-angle light scattering (MALS) results suggested the presence of a homodimer of NHERF1 PDZ $1^{(+)}$. However, this homodimerization did not prevent the interaction between the C-terminus^{42aa} and NHERF1 PDZ $1^{(+)}$. Furthermore, SPR results showed that there was a strong interaction between the two proteins (in the presence of a possible dimerization that was enhanced by DSLL of human β_2AR) with an apparent dissociation constant (k_d^{app}) of 226 nM which is fairly similar to k_d^{app} of 220 nM of CFTR-NHERF1 (in the absence of DSLL of human β_2AR) (Wang et al, 2000).

A potential role of the C-terminus in channel transport is presented in a model in Figure 80 C. The two 3D structures (the CFTR-NHERF1 and the isolated CFTR) were helpful to link the data for isolated domains with data for full-length CFTR. A hypothesis of how the CFTR channel opens and closes is presented in Figure 80 A, B and C.

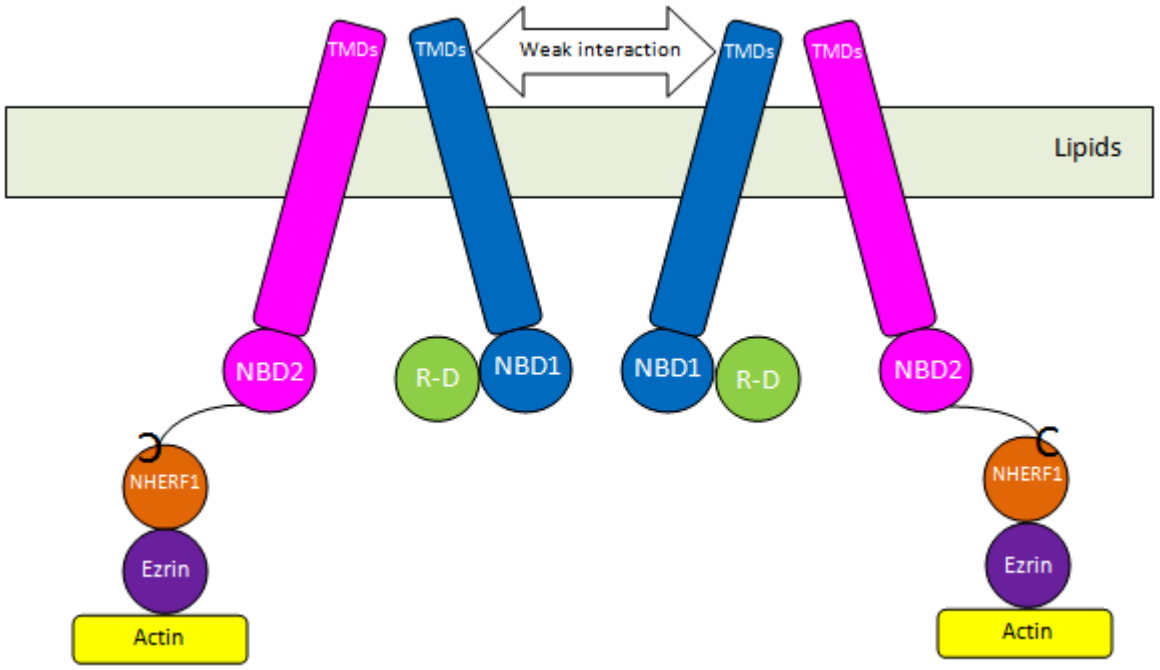


Figure 80 A: The closed state of the channel gating. NHERF1 binds to the C-terminus of CFTR to stabilize it by the binding to Ezrin which binds to actin. Two molecules of CFTR may form a dimer. The interaction between the NBD1 in one CFTR with the NBD1 in the other CFTR is weak. This figure represents the CFTR-NHERF1 3D structure with an inverted V shape.

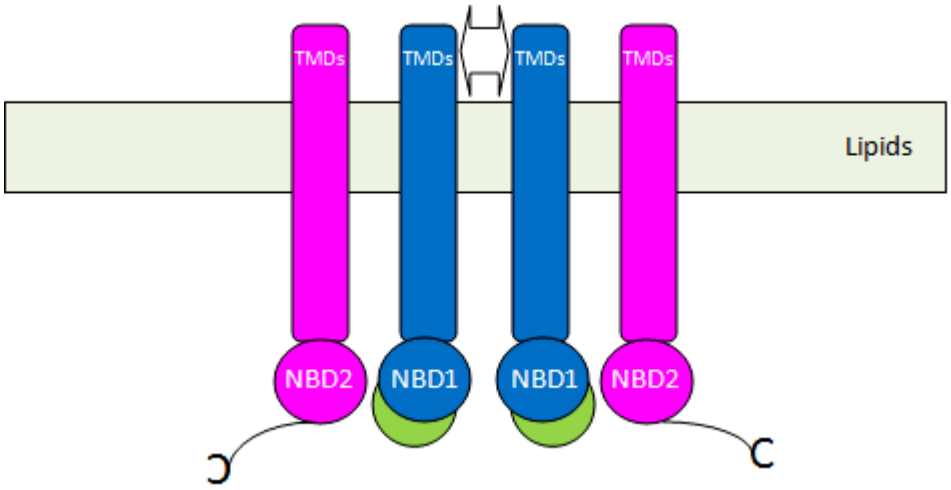


Figure 80 B: The transition state of the channel gating. CFTR in this figure represents the isolated CFTR structure in which no NHERF1 was bound. The transition state is predicted to be a stage with no channel activity. The CFTR channel is partially but not fully opened.

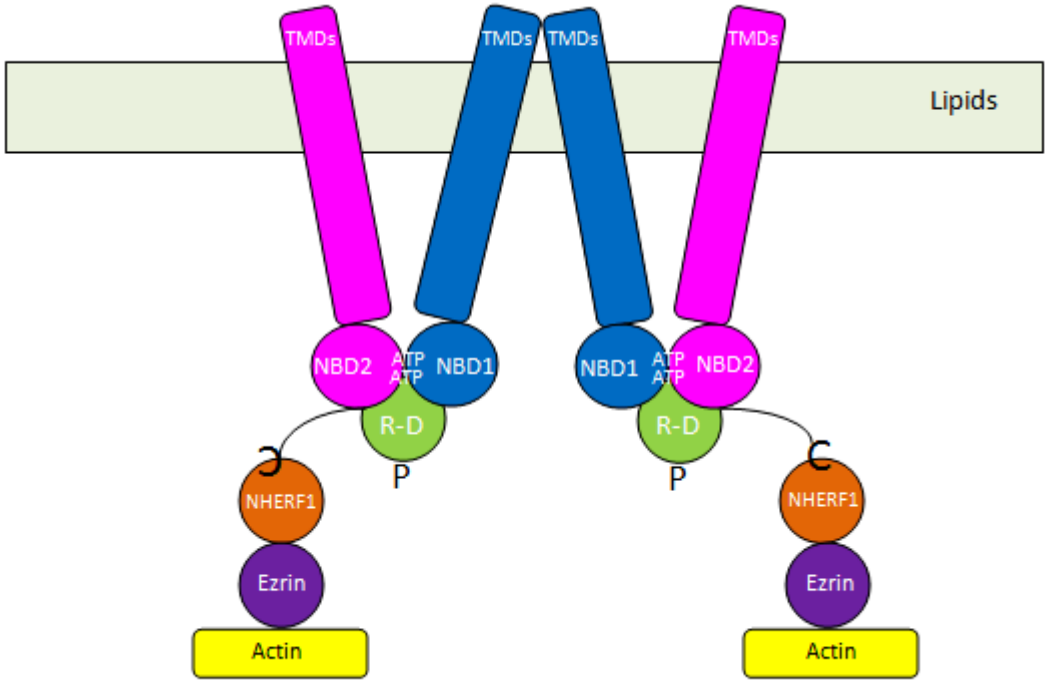


Figure 80 C: The open state of the channel gating. The NBD1 and 2 bind to ATP. This binding makes conformational changes so that the R-domain binds to PKA and the C-terminus. NHERF1 functions as stabilizer of the channel but with decreasing affinity because the whole complex is moving to the transition state (Figure 80 B). This hypothesis can explain why NHERF1 binds to the phosphorylated CFTR with less efficiency as revealed in Figure 50. Furthermore, The the conformation that followed the binding of the ATP allowed the R-domain to bind to PKA and to the C-terminus. This assumption agrees with (Bozoky et al) data as mentioned in section 3.2.1.1.3. The suggested stabilization role of NHERF1-C-terminus in this proposed model can explain why the C-terminus[42aa] interacted strongly with NHERF1 as discussed in section 3.1.3.3.1 and 3.1.3.3.2.

There are several studies that support the channel gating hypothesis. NHERF1 has a C-terminal ERM-binding domain, which interacts with ezrin, radixin, or moesin to link CFTR to the apical actin cytoskeleton in polarized epithelial cell (Short et al, 1998; Sun et al, 2000). It was reported that NHERF1 increases the CFTR channel activity and its ion transport (Wang et al, 2000; Benharouga et al, 2003). In addtion, it has been suggested that NHERF1 stimulates the formation of a dimer of CFTR. This dimerization was proposed to make conformational changes by which the channel is

activated (Wang et al, 2000; Raghuram et al, 2001). Li et al (2004) determined the stoichiometry of CFTR-NHERF1 complex (that was purified from the plasma membrane) and they found that the contribution of NHERF1 in this complex was ~ 2 % (the CFTR to NHERF1 stoichiometry was 0.058 to 0.001 pmol). This data also suggested that NHERF1 was not the major component of this high order complex. Furthermore, although NHERF1 constituted this small amount (~ 2 %) of CFTR-CFTR dimeric complex, this dimerization was capable of forming without the presence of NHERF1. This data is consistent with the isolated CFTR (where was no CFTR-NHERF1 binding) data in which a 3D structure of a dimeric CFTR was obtained as shown in Figures 71, 72 and 73 and in the proposed transition state of the channel gating (Figure 80 B).

Future Work

1- **Distance measurements for NBD1-NBD2 separation have been considered.** In a joint project with Dr Alistair Fielding (Photon Science Institute, School of Chemistry, The University of Manchester), spin-labelled ATP or ATP analogs will be bound to CFTR NBD1 and NBD2 and the spin-spin distance measured by EPR (Electron Paramagnetic Resonance). The effect of binding of NHERF1 PDZ 1 to CFTR would be investigated. Similarly, the change in the distance between the two domains upon the binding of CFTR-NHERF1 PDZ 1 can be studied and measured with fluorescent ATP analogs (TNP-ATP) by Fluorescence Resonance Energy Transfer (FRET).

2- **Gold labeling of NBD1 and 2 of CFTR.** To perform gold-labeling, an additional sequence (10 Histidine) would be fused to each NBD. The aim of this mutation is to allow Ni-NTA-Nanogold to bind to the 10His. We hypothesize that two Nanogold clusters will bind to NBDs, one 10His for each domain. As revealed in Figure 78 B, one Nanogold molecule, which was bound to the C-terminal 10His, can be located using SPA. Based on this data, the fusion of two 10His to CFTR will allow two Nanogold binding. As soon as the two Nanogold are capable of binding to CFTR (and are located by SPA), several conditions well be investigated e.g. the binding to NHERF1 PDZ 1, ATP-PKA. The effect of these ligands on CFTR can be detected by measuring the locations of the two Nanogold molecules and comparing this data to control data (no ligands).

3- Some findings that were revealed in this study could be developed further. The C-terminus of killifish CFTR was sub-cloned into pET-28a and expressed and purified as soluble protein with a high concentration. This study was designed by me as a comparative study to the C-terminus[42aa] of human CFTR. The study was performed by Samyah Al-Anazi (a Masters student in the Ford laboratory at the University of Manchester). A similar approach (sub-cloning, expression and purification of the C-termini of killifish and shark CFTR) was

performed. The two constructs were purified but with low yield. Further optimization for the expression of both constructs is needed.

Electron microscopy micrographs of CFTR-NHERF1 in the presence of ATP (and CFTR-ATP-PKA as a control) were recorded for SPA analysis. This data could be analyzed by SPA image processing software to generate 3D structures. The generated 3D structures will be compared to the structures of CFTR-NHERF1 and isolated CFTR.

4- The investigation of the C-terminus^{42aa}-NHERF1 PDZ 1 interaction showed some interesting results in this study. These could be further investigated as follows:

- A sample of the two proteins (with and without ATP) would be studied by MALS.
- A sample of the two proteins will be sent to our collaborators (Bozoky et al) for NMR studies.
- A sample of the R-domain of CFTR will be added to the complex to investigate if the R-domain will prevent the complex interaction.

Appendix A

This section reveals the results of the C-terminus^{42aa}-NHERF1 PDZ $1^{(+)}$ surface plasmon resonance experiment.

Summary: The results for one more experiment (surface plasmon resonance of the C-terminus^{42aa} and NHERF1 PDZ $1^{(+)}$) were obtained after the completion of the thesis writing. For this reason, these results were added as an Appendix section.

The aim of this experiment was to investigate the interaction between the C-terminus^{42aa} and NHERF1 PDZ $1^{(+)}$. Surface plasmon resonance (SPR) provides information such as the association (k_a), dissociation (k_d), and apparent dissociation constants (k_d^{app}). Furthermore, the interaction between the two proteins was performed in the presence of a 6His-tag to investigate whether the His-tag prevented NHERF1 binding as had previously been reported (Bates et al, 2006).

Background: Surface plasmon resonance is a biophysical method used to detect and measure protein-protein interactions or the interaction between proteins with other ligands. One component (a protein or a ligand) must be immobilized on a sensor surface (typically a gold layer). When light hits the gold layer, the gold atoms resonate and the the excited electrons produce plasmon waves which propagate along the metal interface. Plasmon waves are very sensitive to the surrounding environment (gold/protein/ligand/buffer). Any changes that occur in the protein/ligand/buffer environment (and as a result their refractive index) affect the medium (plasmon waves) which subsequently affect the intensity of the reflected light. These changes in the refractive index are directly related to the entire mass of interacting molecules at the gold chip surface (Schuck, 1997).

Methods: The proteins used to detect the interaction were the 6-His-tagged C-terminus^{42aa} (the expression and purification of this protein were as described in section 2.2.4.2 and 2.2.4.3 with one exception that the purified protein was not thrombin cleaved) and NHERF1 PDZ $1^{(+)}$ (as described in section 3.1.5). The two proteins were buffer exchanged using a 3kDa cutoff filter (Centricon

Centrifugal Filter Units, a 3 KDa microconcentrator device, from Millipore) into 10 mM Na/K phosphate pH 7.4 (80.2 ml of 1 M K_2HPO_4 was mixed with 19.8 ml of 1 M KH_2PO_4 and the solution was made up to 1L with MQ water). The buffer was sterilized by filtering through a 0.2 μm filter. The proteins concentration was ≈ 0.3 mg/mL.

In the SPR experiment the 6-His tagged C-terminus^{42aa} was covalently attached to the sensor chip (HTG, Bio-Rad) surface via the 6-Histidine residues. Un-tagged NHERF1 PDZ $1^{(+)}$ was a mobile reactant. SPR was performed in real time using ProteOn XPR36 (Protein Interaction Array System, Bio-Rad). Interactions were monitored by injecting different concentrations of NHERF1 PDZ $1^{(+)}$ in running buffer (10 mM Na/K phosphate pH 7.4). Binding was performed at 25 °C at 100 μL/min flow rate. The association phase was followed by increasing refractive index at the sensor surface. Subsequently, the dissociation phase was followed by injecting running buffer alone. Between injections, surfaces were regenerated by washing with 300 mM EDTA. All association and dissociation curves were corrected for non-specific binding by subtraction of control curves obtained from injection of the different concentrations of NHERF1 PDZ $1^{(+)}$ through the blank flow channel. The kinetic constants, k_a, k_d and k_d^{app} were calculated using the ProteOn Manager software assuming a two-state model of interactions. Experiments were performed twice. The SPR scan and analysis were kindly performed by Dr Thomas Jowitt (Faculty of Life Sciences, The university of Manchester).

Results and Discussion: The SPR results indicated that there was a strong interaction between the C-terminus^{42aa} and NHERF1 PDZ $1^{(+)}$. As shown in Figure 81, there are 4 stages for the interaction:

1- A fast association (stage 1) between the C-terminus^{42aa} and a probable dimer state of NHERF1 PDZ $1^{(+)}$.
2- A slow association (stage 2) between the C-terminus^{42aa} and a probable dimer state of NHERF1 PDZ $1^{(+)}$ in which the two proteins reached the maximum binding state.
3- A fast dissociation (stage 3) between the C-terminus^{42aa} and a (probably loosely bound) monomer of NHERF1 PDZ $1^{(+)}$.

4- A very slow dissociation (stage 4) between the C-terminus^{42aa} and a tightly bound monomer of NHERF1 PDZ $1^{(+)}$.

The ProteOn Manager software was used to fit the association and dissociation curves. A two-state model was fitted well as shown in Figure 82. The kinetic constants, k_a and k_d for the first state (A dimer of NHERF1 PDZ $1^{(+)}$) were respectively: $9.63 \times 10^5\ M^{-1}s^{-1}$, $2.1 \times 10^{-1}\ s^{-1}$. The kinetic constants, k_a and k_d for the second state (a monomer of NHERF1 PDZ $1^{(+)}$) were respectively: $7.87 \times 10^{-3}\ s^{-1}$, $2.08 \times 10^{-4}\ s^{-1}$. The apparent dissociation constant (k_d^{app}) was 226 nM and this result broadly agreed with the previously published k_d^{app} of CFTR-NHERF1 of 220 nM (Wang et al, 2000). Both k_d^{app} constants were in nanomolar concentration range and this reflects a strong binding between The C-terminus of CFTR and NHERF1.

The SPR results suggested a possible role of a NHERF1 dimerization in CFTR-NHERF binding. It was reported that NHERF1 PDZ 1 was able to form a dimer as revealed in the crystal structure (Ladias, 2003). This role may affect the CFTR channel activity because NHERF1 was suggested to stabilize CFTR and a dimeric complex may allow crosslinks to the cell skeleton as discussed previously. In addition, Cheng et al (2002) suggested that CFTR interacts with at least two PDZ proteins and this is compatible with SPR data in where two forms of NHERF1 PDZ $1^{(+)}$ might interact with CFTR.

The 6-His tag seemed to have no major effect on the interaction of the two proteins, as judged by comparison with the work of Bates and co-workers (2006). It is not possible, however to rule out the possibility that the His tag is compatible with NHERF1 binding when bound to the chip, but not when free in solution. Further experiments would be required to test this possibility.

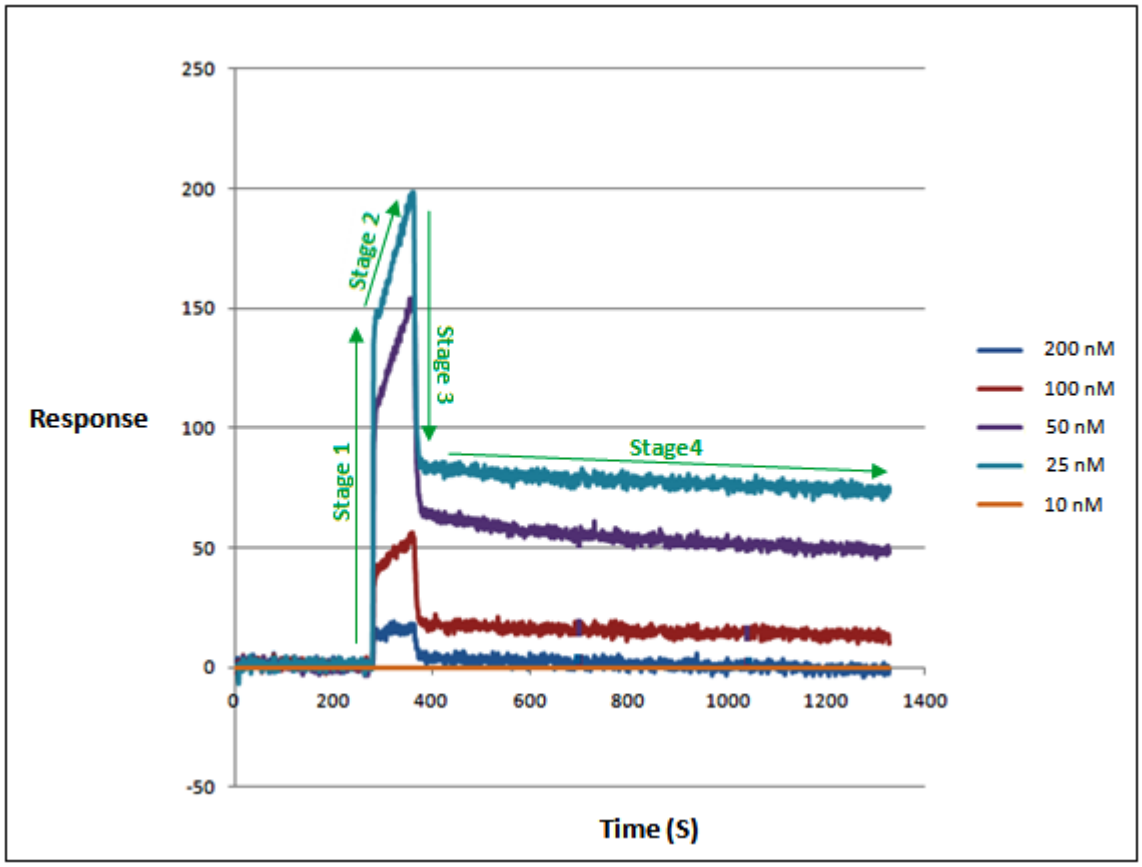


Figure 81: The SPR data for the interaction of the C-terminus^{42aa} and NHERF1 PDZ 1$^{(+)}$ showing the four stages of the binding and dissociation. The colour coding for the experiments with different concentrations of NHERF1 PDZ 1$^{(+)}$ is shown on the right.

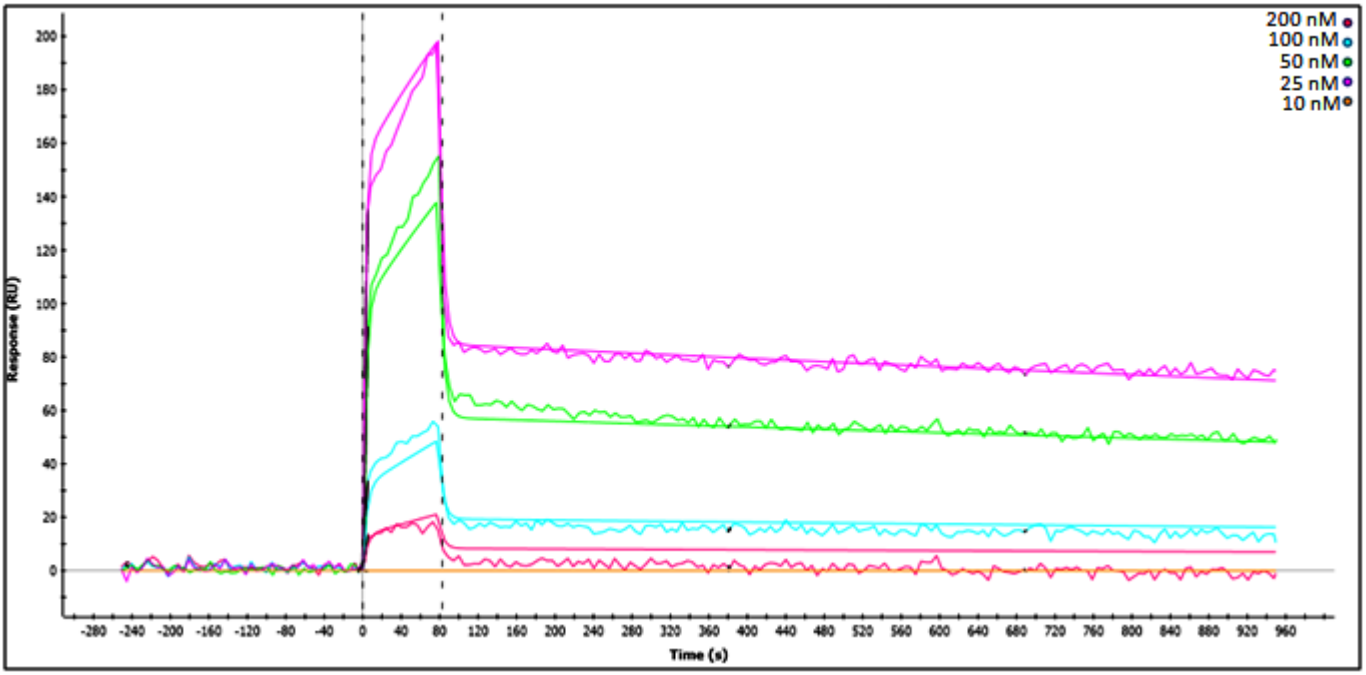


Figure 82: Fitting of a two-state model is shown for the data (smooth lines). The two-state model accounts for a complex that changes shape after interacting with the analyte. The colour coding for the experiments with different concentrations of NHERF1 PDZ $1^{(+)}$ is shown top right.

Appendix B

This section reveals the results of NHERF1 PDZ $1^{(+)}$ multi-angle light scattering (MALS) experiment.

Summary: In order to investigate the assumption that a dimer of NHERF1 PDZ $1^{(+)}$ is interacting with a monomer of C^{42aa} (as explained in the SPR results section, Figure 81), a sample of NHERF1 PDZ $1^{(+)}$ was analysed by multi-angle light scattering (MALS). MALS experiments work by measuring the amount of scattered light at different angles. The amount of scattered light is related to the molar mass of the protein. In addition, the size of the molecule can be measured using the angular distribution of the scattered light.

Methods: A NHERF1 PDZ $1^{(+)}$ sample (0.5 mg/ml) in (140 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) was subjected to gel filtration followed by MALS for detecting the distribution of monomers or dimers in the protein sample. Briefly, a HPLC system with an automated 96-well sample changer was used with a Superdex Sx75 HPLC size-exclusion column to separate proteins based on their size. After the separation of the protein sample, a DAWN HELEOS detector array (18 detectors) (Wyatt technologies) was used to measure light scattering at different angles (15 angles) at a laser wavelength of 658 nm. Refractive index (Optilab rEX Refractometer) and UV absorbance (Generic UV) were also measured.

Results and Discussion: As shown in Figure 83, one peak was obtained. The peak indicated a dimer of NHERF1 PDZ $1^{(+)}$ with an average molecular weight of 22.2 kDa. Refractive index and (and UV absorbance) indicated that most of the mass was obtained as a dimer and no peak was obtained for a monomer of NHERF1 PDZ $1^{(+)}$ (Molecular weight of 10.2 kDa).

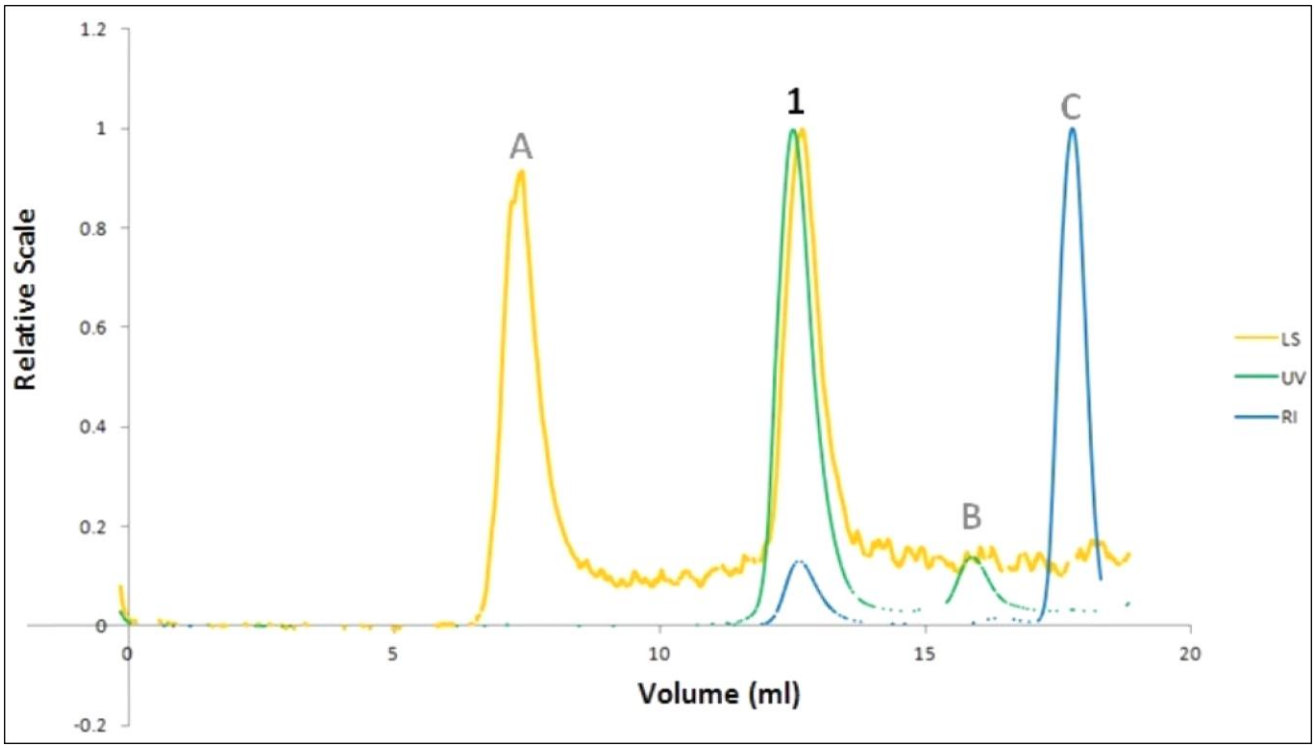


Figure 83: Multi-angle light scattering of NHERF1 PDZ $1^{(+)}$. Peak 1 illustrates three signals, Light scattering (LS), Refractive index (RI) and UV absorbance (UV) indicating a dimer of NHERF1 PDZ $1^{(+)}$. Peak A is big aggregates at the start of the gel filtration column with no RI or UV signals. Peaks B and C are buffer traces in the end of the column.

Appendix C

This section gives some details about the vectors and their maps.

1- pGEX-2T vector

A vector pGEX-2T used as a template to subclone the human NHERF PDZ1 domain (residues 11–94) and the carboxyl-terminal extension N95DSLL99 that corresponds to residues 409–413 of human β_2AR. The new created vector pGEX-2TJL1 was used in this stduy. A pGEX-2TJL1 vector containing NHERF1 PDZ $1^{(+)}$ (as shown in Figure 84) was kindly provided by Dr Karthikeyan (Karthikeyan et al, 2001).

NHERF1

```
       10         20         30         40         50         60
MSADAAAGAP LPRLCCLEKG PNGYGFHLHG EKGKLGQYIR LVEPGSPAEK AGLLAGDRLV

       70         80         90        100        110        120
EVNGENVEKE THQQVVSRIR AALNAVRLLV VDPETDEQLQ KLGVQVREEL LRAQEAPGQA

      130        140        150        160        170        180
EPPAAAEVQG AGNENEPREA DKSHPEQREL RPRLCTMKKG PSGYGFNLHS DKSKPGQFIR

      190        200        210        220        230        240
SVDPDSPAEA SGLRAQDRIV EVNGVCMEGK QHGDVVSAIR AGGDETKLLV VDRETDEFFK

      250        260        270        280        290        300
KCRVIPSQEH LNGPLPVPFT NGEIQKENSR EALAEAALES PRPALVRSAS SDTSEELNSQ

      310        320        330        340        350
DSPPKQDSTA PSSTSSSDPI LDFNISLAMA KERAHQKRSS KRAPQMDWSK KNELFSNL
```

NHERF1 PDZ 1^{+}

```
GSSRMLPRLC CLEKGPNGYG FHLHGEKGKL GQYIRLVEPG SPAEKAGLLA GDRLVEVNGE
NVEKETHQQV VSRIRAALNA VRLLVVDPEN DSLL
```

NHERF1 PDZ 1

```
MGLPRLC CLEKGPNGYG FHLHGEKGKL GQYIRLVEPG SPAEKAGLLA GDRLVEVNGE
NVEKETHQQV VSRIRAALNA VRLLVVDPEN DSLL
```

Figure 84: The amino acid sequences of NHERF1 protein. Yellow highlighted amino acids indicate the original sequence (no mutations) of NHERF1 PDZ 1. Two forms of NHERF1 PDZ 1 were investigated in this study. First: NHERF1 PDZ $1^{(+)}$ which is encoded by a gene in

a pGEX-2TJL1 vector. Secondly: NHERF1 PDZ 1 which is encoded by a gene in a pET-52b vector. Red letters indicate the mutated amino acids. NSL mutations were created by (Karthikeyan et al, 2001) for NHERF1-ligand crystallization studies. These mutations had no effect on the C-terminus of CFTR and NHERF1 PDZ 1 binding as reveald by MALDI-TOF-MS (in section 3.1.3.3.2) and surface plasmon resonance results (in Apendix A). Furthermore, a sample of co-expressed C-terminus^{42aa}-NHERF1 PDZ 1 was sent to (Protein Mass Spectrometry Core Facility, Michael Smith Building, University of Manchester). The sequences of the two proteins (100% coverage for both proteins) were confirmed by Orbitrap Elite (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. GSSRM mutations: GS are from thrombin site, SR are cloning artefact, and M is a start codon. MG mutations: M is a start codon, G is cloning artefact.

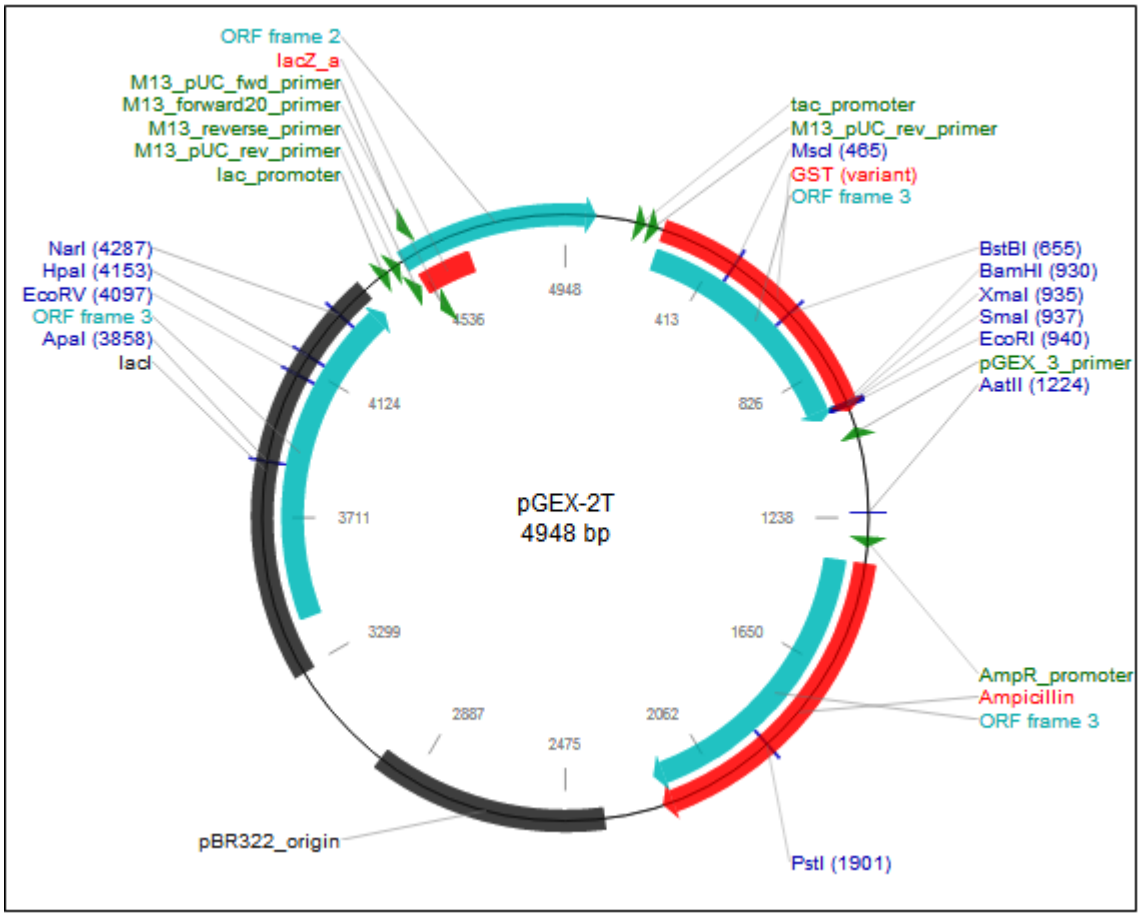


Figure 85: The map of pGEX-2T vector. NHERF1 PDZ 1$^{(+)}$ was sub-cloned into pGEX-2T vector, fused to N-terminal GST-tag.

2- pET-52b vector map

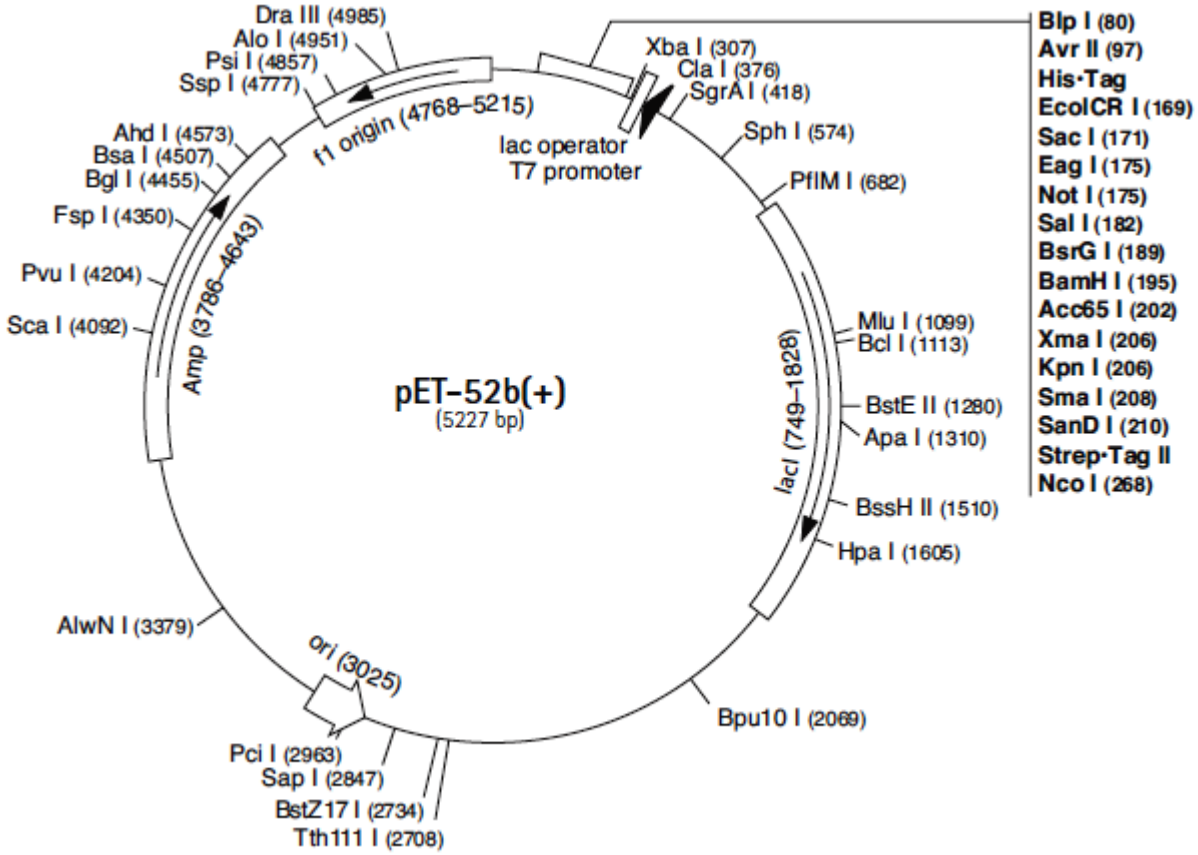


Figure 86: **The map of pET-52b vector.** NHERF1 PDZ 1 was sub-cloned using two restriction sites, BamH I / Nco I (as described in section 2.2.1.1).

3- pET-24a vector map

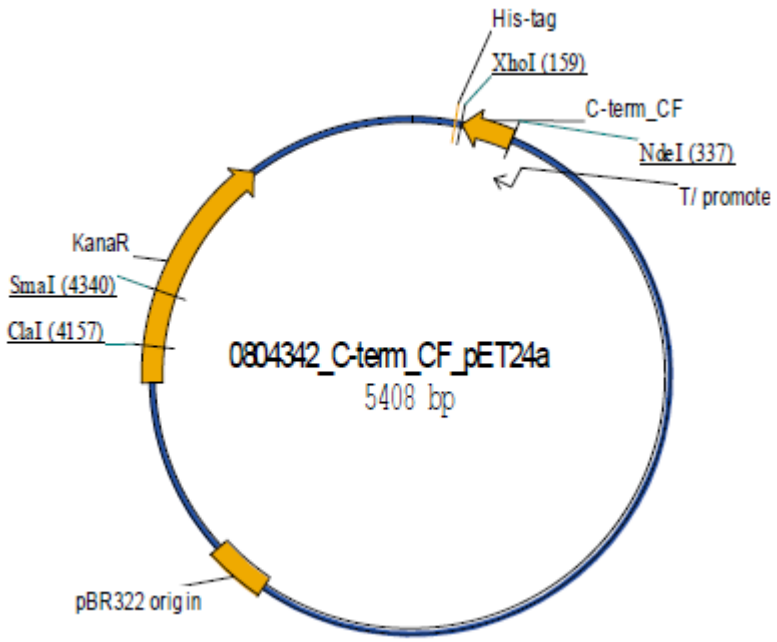


Figure 87: **The map of pET-24a vector.** The synthetic gene C-term_CF (the C-terminus^{42aa}) was assembled from synthetic oligonucleotides and / or PCR products. The fragment was cloned into pET24a (kanR) using NdeI and XhoI restriction sites. The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. The same strategy (and vector) used to synthesize the C-terminus^{61aa} (section 3.1.4).

4- pGA4 vector map

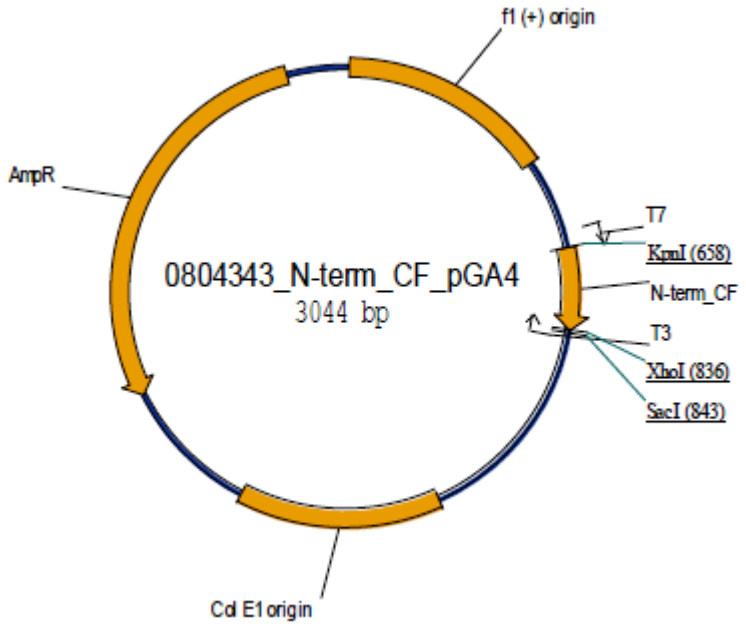


Figure 88: The map of pGA4 vector. The pGA4 vector containg the N-terminus of CFTR (was not used in this study) used to sub-clone the shark CFTR C-terminus using two restriction sites SphI and XhoI (section 2.2.1.3). The plasmid DNA was purified from transformed bacteria and concentration determined by UV spectroscopy. The final construct was verified by sequencing. The sequence congruence within the used restriction sites was 100%. The same strategy (and vector) used to sub-clone the killifish CFTR C-terminus (section 2.2.1.3).

Appendix D

This section shows additional data that were mentioned in Results and Discussion 1 and 2.

1- Sitting drop experiments

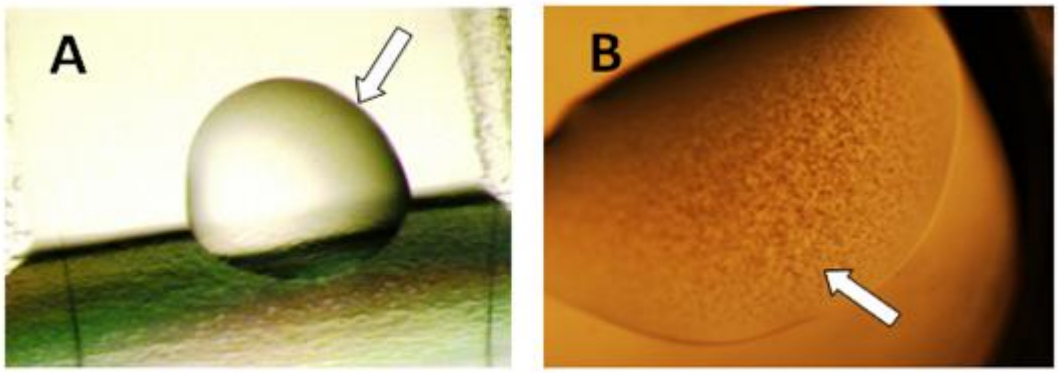


Figure 89: An example of sitting drop experiments. A, a clear drop of the C-terminus^{42aa} sits down into 96-well plate. B, precipitation of protein-reservoir buffer mixture. Precipitation can be a stage of crystal formation (within the nucleation zone) or can lead to no crystals. The arrows point the drops.

2- Hanging drop experiments

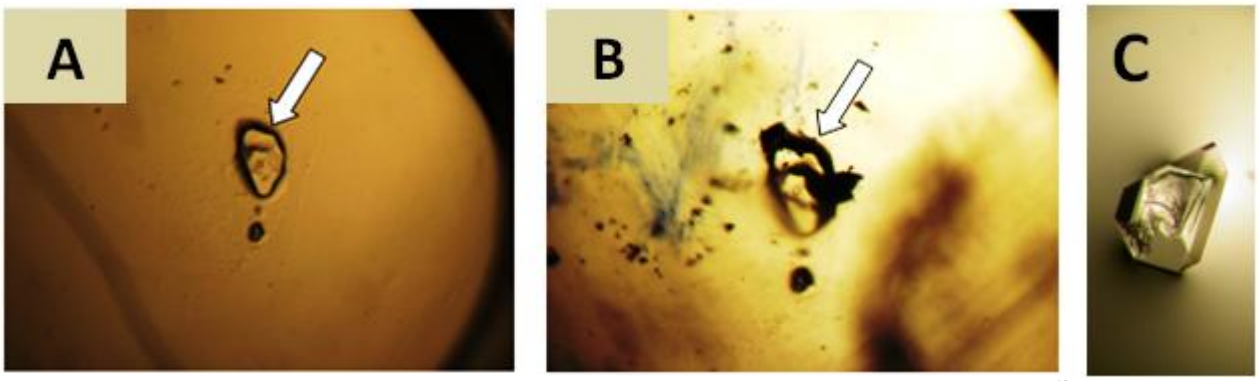


Figure 90: An example of hanging drop experiments. A, a crystal of the C-terminus^{42aa}-NHERF1 PDZ 1 complex was obtained and stained with Methylene Blue (CALBIOCHEM) as a routine check (as shown in slide B) but the crystal did absorb the stain indicating that it was a salt crystal. The arrows point the crystals. C, another crystal of the C-terminus^{42aa}-NHERF1 PDZ 1 complex and it was checked by a crush using a fine syringe. The crystal was not crushed indicating it was probably a salt crystal.

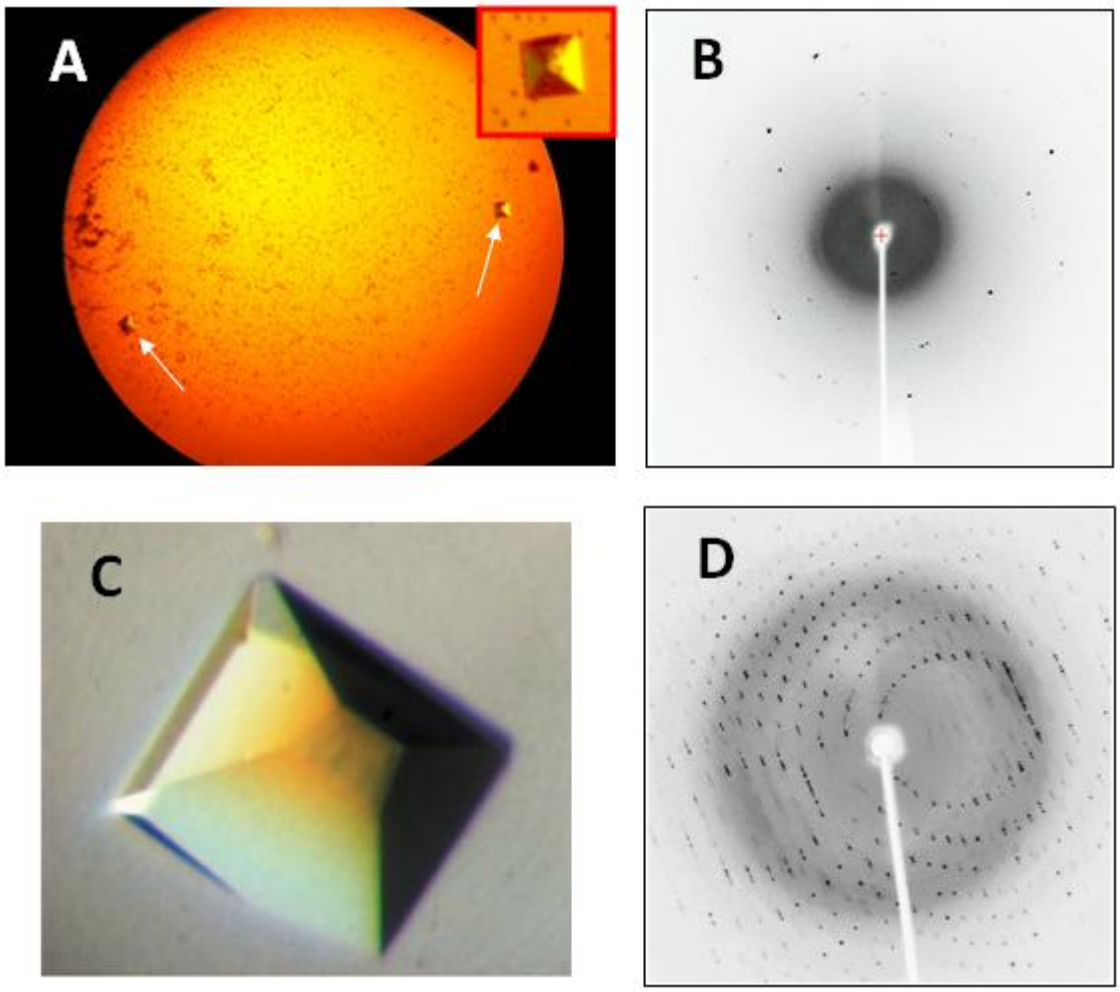


Figure 91: Comparisons between the C-terminus^{42aa}-NHERF1 PDZ 1 complex crystal and a marRAB protein crystal. A, two pyramidal crystals (pointed by the two white arrows) were formed using a complex of the C-termiuns^{42aa}-NHERF1 PDZ 1 (condition: 0.1 M sodium acetate pH 5.4, 2 M sodium chloride, 40% glycerol, were left at 20 ^{o}C), the size of both crystals was 60 µm. B, the X-ray diffraction pattern of one crystal of the two obtained crystals. The X-ray scan was performed using R-AXIS IV++ instrument (the Macromolecular Crystallography Facility, the Manchester Interdisciplinary Biocentre). C, a pyramidal crystal of the multiple antibiotic resistance operon (marRAB) protein which looks similar to the C-terminus^{42aa}-NHERF1 crystal as shown in the red box in slide A. The marRAB crystal was obtained into 0.1 M sodium acetate pH 5.5, 0.2 M MgCl2 and 22% PEG 2K MME (polyethylene glycol monomethylether) with a size of 150 µm. D, the X-ray diffraction pattern of the marRAB crystal. This x-ray diffraction is a typical diffraction that usually obtains from protein crystals. In contrast, the diffraction in slide B is a typical diffraction from salt

crystals. Salt crystals are solid and contain narrow spaces between atoms (small unit cell). For this reason, the salt diffraction spots are widely spaced as shown in slide B. On the other hand, protein crystals are large and fragile (large unit cell) which leads to a small spacing between diffraction spots as shown in slide D. The data of marRAB was obtained from (Tu et al, 2005). A comparison between The C-terminus^{42aa}-NHERF1 PDZ 1 crystal and marRAB crystal is presented in Table 18.

Protein crystal	The C-terminus^{42aa}-NHERF1 PDZ 1	marRAB
Crystallization method	Hanging drop vapour diffusion	Sitting drop vapour diffusion
Crystal shape	Pyramid-shaped crystal	Pyramid-shaped crystal
Crystal size	60 µm	150 µm
Shared additives	0.1 M sodium acetate pH 5.4	0.1 M sodium acetate pH 5.5
Protein concentration	6 mg/ml	22 mg/ml
Data collection	R-AXIS IV++	R-AXIS IV++
Resolution	No diffraction	1.8 Å

Table 18: Similarities and differences between the C-terminus^{42aa}-NHERF1 PDZ 1 and marRAB crystals. The similarities between the two crystals suggest that further optimization for the C-terminus^{42aa}-NHERF1 PDZ 1 complex crystal is needed by varying the concentration of protein and additives or trying different temperatures.

3- Orbi-trap Mass Spectrometry of the C-terminus^{42aa}-NHERF1 PDZ 1

In order to investigate the hypothesis that the interaction between the C-terminus^{42aa} and NHERF1 PDZ 1 is regulated by phosphorylation, a sample of this complex was sent for Orbi-trap mass spectrometry analysis. The scan was performed by Mass Spectrometry Core Facility, Michael Smith Building, University of Manchester.

Methods: The C-terminus^{42aa}-NHERF1 PDZ 1 sample (0.5 mg/ml) into (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) was analysed by Liquid chromatography linked to tandem mass spectrometry (LC-MS/MS) using an UltiMate® 3000 Rapid Separation LC (RSLC, Dionex Corporation, Sunnyvale, CA) coupled to an Orbi-trap Elite (Thermo Fisher Scientific, Waltham, MA) mass spectrometer. Peptide mixtures were separated using a gradient from 92% A (0.1% formic acid (FA) in water) and 8% B (0.1% FA in acetonitrile) to 33% B, in 44 min at 300 nL min-1, using a 75 mm x 250 μm i.d. 1.7 μM BEH C18, analytical column (Waters). Peptides were selected for fragmentation automatically by data dependant analysis. Data produced were searched using Mascot (Matrix Science UK), against the [UNIPROT] database with taxonomy of [HUMAN] selected with Phosphorylation as variable mode. Data were validated using Scaffold (Proteome Software, Portland, OR).

Results and Discussion: The results showed that there was no evidence of phosphorylation. The amino acids coverage was 100% for both proteins. The common modification Carbamidomethyl (+57 Da) was detected in both proteins and this type of modification is discounted in mass spectrometry experiments.

4- Analysis of CD data by DichroWeb

The CDSSTR program (Sreerama and Woody, 2000) was used to analyze the CD data (as explained in section 3.1.3.3.3). This program is available on DichroWeb and registration is required to be used (Whitmore and Wallace, 2008). In order to use the CDSSTR program, there were several steps as follows:

- The experimental CD data (raw data) was saved in ascii format (text format).
- The data in ascii format was copied and pasted to Excel file in order to be modified.
- Some header information (in red boxes) was deleted before further analysis as appeared in Figure 92.

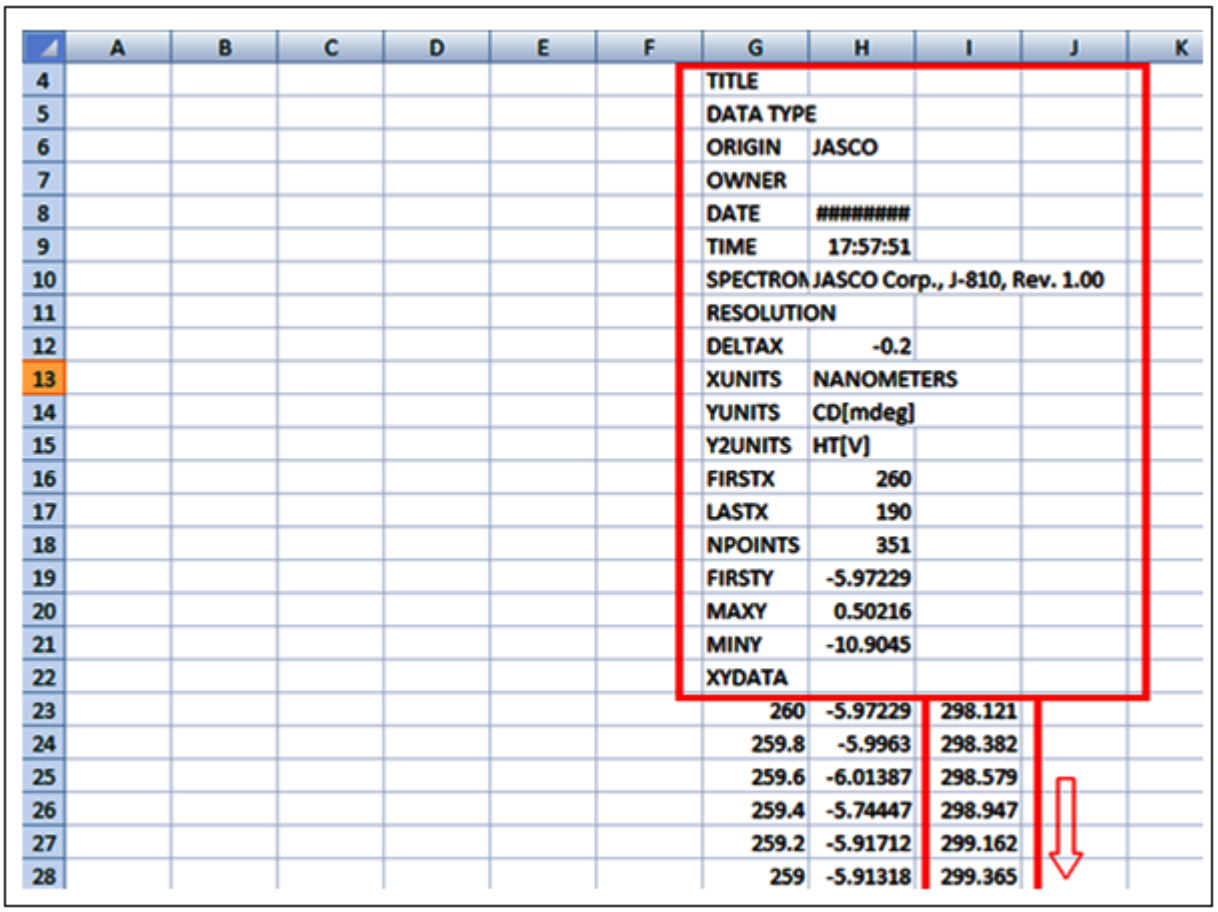


Figure 92: A section of Exel file contains the raw CD. Data in red boxes was deleted. The first and second columns of data were kept. The arrow indicates that the remained data is below (not shown).

- The data was copied and pasted into a Text document.
- The Text document was uploaded to DichroWeb to analyse the data. The DichroWeb can be found on: http://dichroweb.cryst.bbk.ac.uk/html/home.shtml.
- The parameters used to analyze the data by the CDSSTR program is shown in Table 19.

File format	Free with preview
Input units	Millidegrees / theta (machine units)
Initial wavelength (nm)	260
Analysis program	CDSSTR
Final wavelength (nm)	190
Reference set	Set 4
Output units	Delta epsilon
Mean residue weight (Da)	115 for the C-terminus^{42aa} and 110 for the C-terminus^{42aa}-NHERF1 PDZ 1
Protein concentration (mg/ml)	0.5
Path length (cm)	0.5

Table 19: Parameters used to analyze the CD data by DichroWeb.

- The CD data (in the Text doecument) was submitted to the program server using the previous parameters
- The analysis reults were obatined as explained in Table 14 (section 3.1.3.3.3).

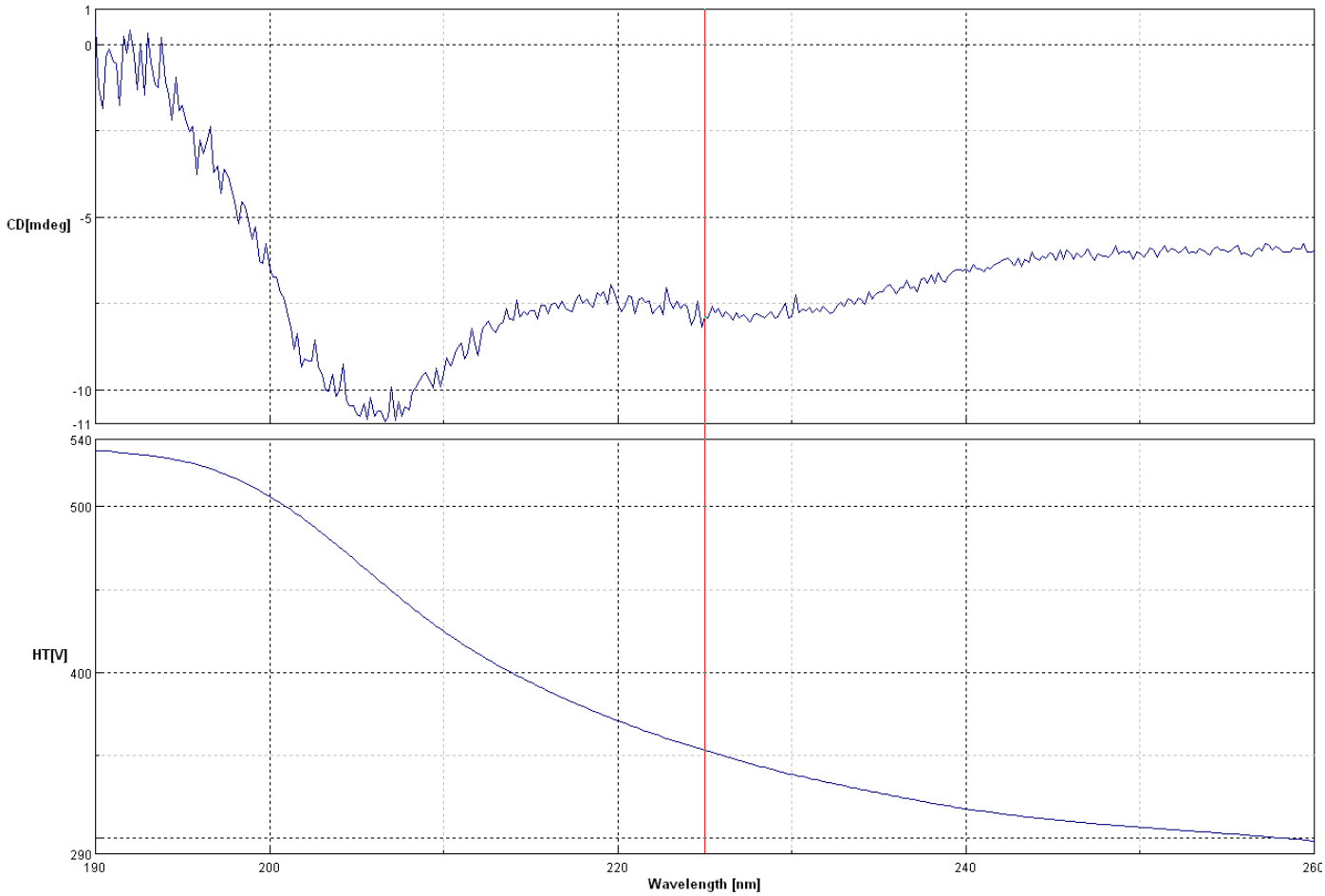


Figure 93: The raw CD data of the C-terminus[42aa]. This raw data was used for the DichroWeb analysis as mentioned in this section and section (3.1.3.3.3). HT [V] is High tension voltage. The HT [V] is proportional to absorbance. If the HT [V] is ≥ 600, this means the scanned sample is saturated and needs to be diluted. [mdeg] is millidegrees CD unit. The red line is a measurement scale.

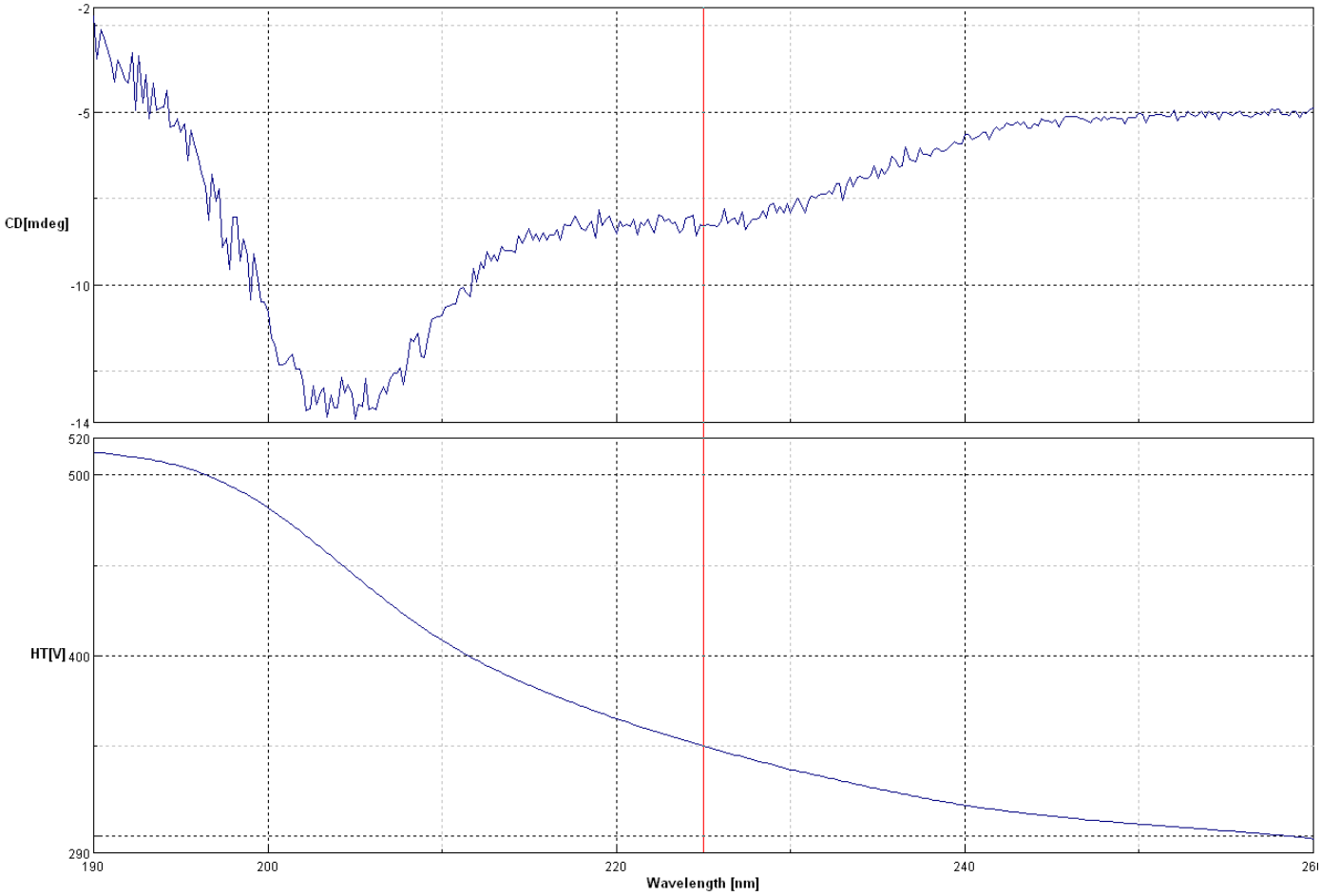


Figure 94: The raw CD data of the C-terminus^{42aa}-NHERF1 PDZ 1. Other details as in Figure 93.

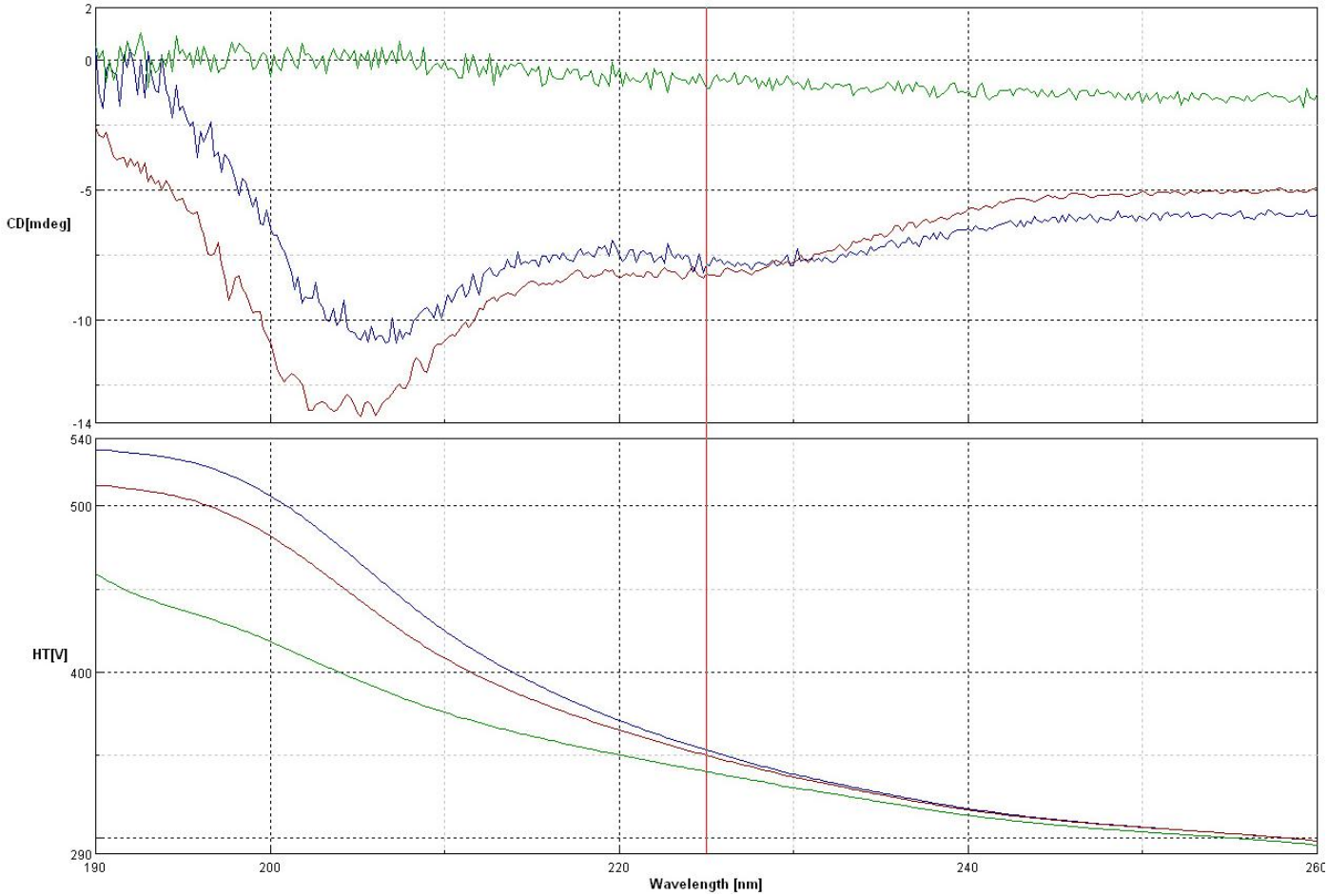


Figure 95: Superimposed CD data. The blue curve is the CD signal of the C-terminus^{42aa}. The dark brown curve is the CD signal of the C-terminus^{42aa}-NHERF1 PDZ 1. The green curve is the CD signal of buffer. Other details as in Figure 93.

5- Micrographs of the reconstituted mouse and killifish CFTR

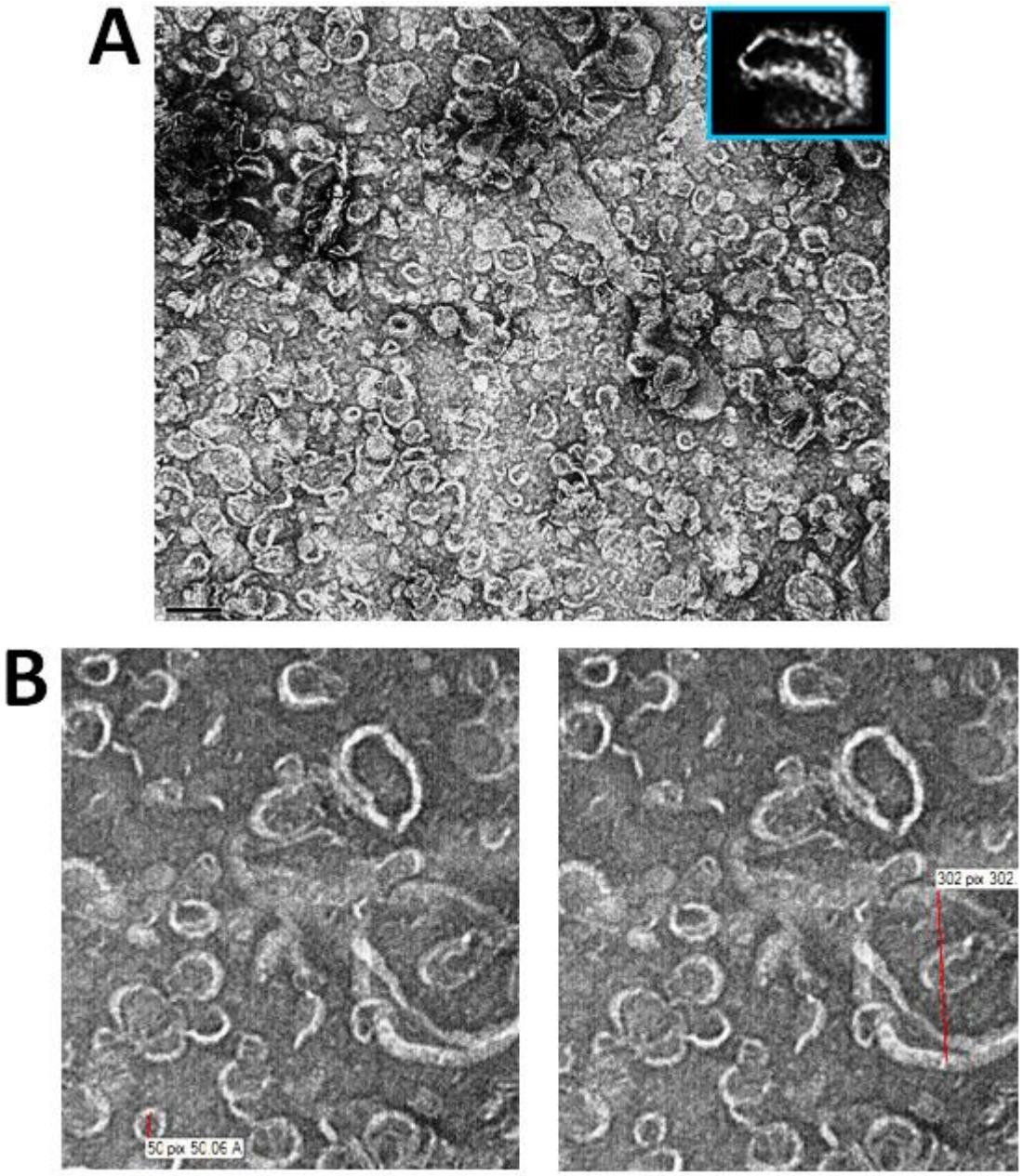


Figure 96: A micrograph of vesicles after an attempt to reconstitute mouse CFTR. A**,** The protein is not easily identifiable. The blue box contains one viscle. The scale bar is 100 nm. The micrograph was recorded at 100K magnification. B, Measurments of the dimentions of two typical viscles, the big one on the right (302 Å) and the small one on the left (50 Å).

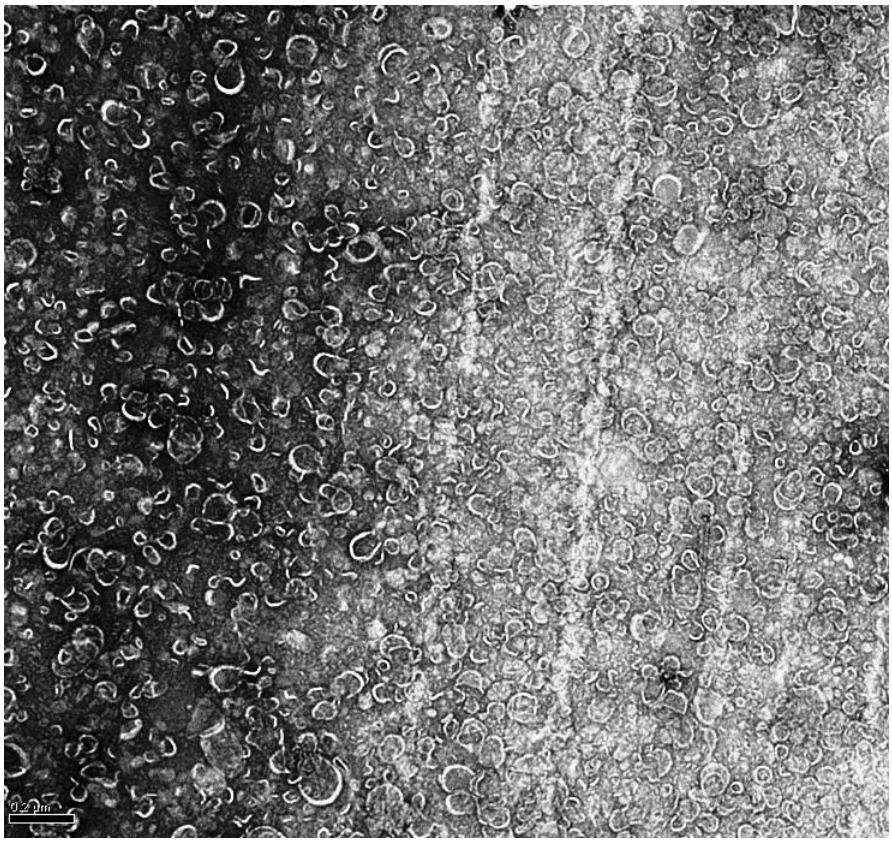


Figure 97: A micrograph of vesicles from an experiment aimed at reconstitution of killifish CFTR. The scale bar is 200 nm. The micrograph was recorded at 60K magnification. As before, there is no clear evidence for protein particles within the vesicles.

References

Aghajari, N., Feller, G., Gerday, C., Haser, R., **2002**. Structural basis of alpha-amylase activation by chloride. Protein Sci 11, 1435-1441.

Ahn, Y.H., Ji, E.S., Lee, J.Y., Cho, K., Yoo, J.S., **2007**. Arginine-mimic labeling with guanidinoethanethiol to increase mass sensitivity of lysine-terminated phosphopeptides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 21, 2204-2210.

Akabas, M.H., **2000**. Cystic fibrosis transmembrane conductance regulator. Structure and function of an epithelial chloride channel. J Biol Chem 275, 3729-3732.

Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., Chang, G., **2009**. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323, 1718-1722.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., **1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Andersen, D.H., **1938**. Cystic fibrosis of the pancreas and its relation to celiac disease. Am J Dis Child;56:344.

Anderson, M.P., Berger, H.A., Rich, D.P., Gregory, R.J., Smith, A.E., Welsh, M.J., **1991**. Nucleoside triphosphates are required to open the CFTR chloride channel. Cell 67, 775-784.

Atwell, S., Brouillette, C.G., Conners, K., Emtage, S., Gheyi, T., Guggino, W.B., Hendle, J., Hunt, J.F., Lewis, H.A., Lu, F., Protasevich, I.I., Rodgers, L.A., Romero, R., Wasserman, S.R., Weber, P.C., Wetmore, D., Zhang, F.F., Zhao, X., **2010**. Structures of a minimal human CFTR first nucleotide-binding domain as a monomer, head-to-tail homodimer, and pathogenic mutant. Protein Eng Des Sel.;23(5):375-84.

Awayn, N.H., Rosenberg, M.F., Kamis, A.B., Aleksandrov, L.A., Riordan, J.R., Ford, R.C., **2005**. Crystallographic and single-particle analyses of native- and nucleotide-bound forms of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Biochem Soc Trans 33, 996-999.

Baker, J.M., Hudson, R.P., Kanelis, V., Choy, W.Y., Thibodeau, P.H., Thomas, P.J., Forman-Kay, J.D., **2007**. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol.;14(8):738-45.

Balakrishna, A.M., Saxena, A.M., Mok, H.Y., Swaminathan, K., **2009**. Structural basis of typhoid: Salmonella typhi type IVb pilin (PilS) and cystic fibrosis transmembrane conductance regulator interaction. Proteins. :77(2):253-61.

Baldwin, R.L., **1996**. How Hofmeister ion interactions affect protein stability. Biophys J 71, 2056-2063.

Barasch, J., Kiss, B., Prince, A., Saiman, L., Gruenert, D., & Al-Awqati, Q., **1991**. Defective Acidification of Intracellular Organelles in Cystic Fibrosis. Nature, 352, 6330, pp. 70-73.

Bates, I.R., Hébert, B., Luo, Y., Liao, J., Bachir, A.I., Kolin, D.L., Wiseman, P.W. and Hanrahan, J.W., **2006**. Membrane lateral diffusion and capture of CFTR within transient confinement zones. Biophys. J., 91:1046-58.

Bear, C.E., Li, C.H., Kartner, N., Bridges, R.J., Jensen, T.J., Ramjeesingh, M., Riordan, J.R., **1992**. Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). Cell 68, 809-818.

Benharouga, M., Sharma, M., So, J., Haardt, M., Drzymala, L., Popov, M., Schwapach, B., Grinstein, S., Du, K., Lukacs, G.L., **2003**. The role of the C terminus and Na+/H+ exchanger regulatory factor in the functional expression of cystic fibrosis transmembrane conductance regulator in nonpolarized cells and epithelia. J Biol Chem 278, 22079-22089.

Bernsel, A., Viklund, H., Hennerdal, A., Elofsson, A., **2009**. TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 37, W465-468.

Berova, N., Nakanishi, K., Woody, R.W., **2000.** Circular Dichroism: Principles and Applications, 2nd edR, Wiley-VCH, New York, 877 pp.

Blom, N., Gammeltoft, S., Brunak, S., **1999**. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294, 1351-1362.

Borbat, P.P., Surendhran, K., Bortolus, M., Zou, P., Freed, J.H., Mchaourab, H.S., **2007**. Conformational motion of the ABC transporter MsbA induced by ATP hydrolysis. PLoS Biol 5, e271.

Bossard, F., Silantieff, E., and Gauthier, C., **2012**. Fine Tuning of CFTR Traffic and Function by PDZ Scaffolding Proteins. Cystic Fibrosis - Renewed Hopes Through Research, ISBN: 978-953-51-0287-8.

Bossard, F., Robay, A., Toumaniantz, G., Dahimene, S., Becq, F., Merot, J., & Gauthier, C., **2007**. Nhe-Rf1 Protein Rescues Deltaf508-Cftr Function. American Journal of Physiology - Lung Cellular and Molecular Physiology, 292, 5, pp. L1085-1094.

Bozoky, Z.; Vernon, R.; Thomas, P.J.; Frizzell, R.A.; Ford, R.C.; Forman-Kay, J., **2011a**. Dynamic intermolecular and intramolecular interactions of the CFTR R region. In: The 25th annual North American Cystic Fibrosis Conference, Anaheim, California, November 3-5. Pediatric Pulmonology. Volume 46, Issue S34, pp 212 – 428.

Bozoky, Z., Krzeminski, M., Baker, J.M., R., M., Hudson, R.P., Birtley, J., Al-Zahrani, A., Ford, R.C., Frizzell, R.A., Thomas, P.J., and Forman-Kay, J.D., **2011b**. The regulatory R region of CFTR serves as a dynamic integrator. Manuscript in Preparation.

Bozzola, John J.; Russell, Lonnie D., **1999**. "Specimen Preparation for Transmission Electron Microscopy". Electron microscopy : principles and techniques for biologists. Sudbury, Mass.: Jones and Bartlett. pp. 21–31.

Brancia, F.L., Oliver, S.G., Gaskell, S.J., **2000**. Improved matrix-assisted laser desorption/ionization mass spectrometric analysis of tryptic hydrolysates of proteins following guanidination of lysine-containing peptides. Rapid Commun Mass Spectrom 14, 2070-2073.

Branden, C., Tooze, J. **1999**. Introduction to Protein Structure. New York: Garland. pp. 224.

Brône, B., and Eggermont, J., **2005**. PDZ proteins retain and regulate membrane transporters in polarized epithelial cell membranes. Am J Physiol Cell Physiol : vol. 288 no. 1 C20-C29.

Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T., **2005**. Protein structure prediction servers at University College London. Nucleic Acids Res 33, W36-38.

Cao, T.T., Deacon, H.W., Reczek, D., Bretscher, A., von Zastrow, M., **1999**. A kinase-regulated PDZ-domain interaction controls endocytic sorting of the beta2-adrenergic receptor. Nature 401, 286-290.

Carson, M., Johnson, D.H., McDonald, H., Brouillette, C., Delucas, L.J., **2007**. His-tag impact on structure. Acta Crystallogr D Biol Crystallogr 63, 295-301.

Chang, X.B., Hou, Y.X., Jensen, T.J., Riordan, J.R., **1994**. Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. J Biol Chem 269, 18572-18575.

Chao, A.C., et al., **1994**. Activation of intestinal CFTR Cl- channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. Embo. J. 13, 1065-1072.

Chen, C.J., Chin, J.E., Ueda, K., Clark, D.P., Pastan, I., Gottesman, M.M., Roninson, I.B., **1986**. Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. Cell 47, 381-389.

Chen, J.H., Chang, X.B., Aleksandrov, A.A., Riordan, J.R., **2002**. CFTR is a monomer: Biochemical and functional evidence. J Membr Biol;188:55–71.

Cheng, J., Moyer, B.D., Milewski, M., Loffing, J., Ikeda, M., Mickle, J.E., Cutting, G.R., Li, M., Stanton, B.A., Guggino, W.B., **2002**. "A Golgi-associated PDZ domain protein modulates cystic fibrosis transmembrane regulator plasma membrane expression". J. Biol. Chem. (United States) 277 (5): 3520–9.

Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'riordan, C. R., & Smith, A. E., **1990**. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. Cell. 63, 827-834.

Chernushevich, I.V., Loboda, A.V., Thomson, B.A., **2001**. An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom.;36(8):849-65.

Choudhuri, S., Klaassen, C.D., **2006**. Structure, function, expression, genomic organization, and single nucleotide polymorphisms of human ABCB1 (MDR1), ABCC (MRP), and ABCG2 (BCRP) efflux transporters. Int J Toxicol 25, 231-259.

Chung, H.J., Huang, Y.H., Lau, L.F., Huganir, R.L., **2004**. Regulation of the NMDA receptor complex and trafficking by activity-dependent phosphorylation of the NR2B subunit PDZ ligand, i24:10248–10259.

Cohen, N.A., Brenman, J.E., Snyder, S.H., and Bredt, D.S., **1996**. Binding of the inward rectifier K+ channel to PSD-95 Is regulated by protein kinase A phosphorylation. Neuron 17: 759−767.

Li, C., Naren, A.P**., 2011**. CYSTIC FIBROSIS (Analysis of CFTR Interactome in the Macromolecular Complexes). Methods in Molecular Biology, Volume 741, Part 3, 255-270.

Cormet-Boyaka, E., Jablonsky, M., Naren, A.P., Jackson, P.L., Muccio, D.D., Kirk, K.L., **2004**. Rescuing cystic fibrosis transmembrane conductance regulator (CFTR)-processing mutants by transcomplementation. Proc Natl Acad Sci U S A.;101(21):8221-6.

Cong, Yao., and Ludtke, Steven J., **2010**. Single Particle Analysis at High Resolution. Methods in Enzymology, Volume 482. 211-235.

Cowley, J.M., Goodman, P., Rees, A.L.G., **1957**. Acta Cryst. 10: 19-25.

Dalbøge, H., Bayne, S., and Pederson, J., **1990**. In vivo processing of N-terminal methionine in E. coli. FEBS Lett. Volume: 266, Issue: 1-2, Pages: 1-3.

Dalemans, W., Hinnrasky, J., Slos, P., Dreyer, D., Fuchey, C., Pavirani, A., & Puchelle, E., **1992**. Immunocytochemical Analysis Reveals Differences between the Subcellular Localization of Normal and Delta Phe508 Recombinant Cystic Fibrosis Transmembrane Conductance Regulator. Experimental Cell Research, 201, 1, pp. 235-240.

Damas, C., Amorim, A., Gomes, I., **2008**. Cystic fibrosis: review. Rev Port Pneumol: 14(1):89-112.

Davis, P.B., 2006. Cystic fibrosis since **1938**. Am J Respir Crit Care Med 173, 475-482.

Dawson, R.J., Locher, K.P., **2006**. Structure of a bacterial multidrug ABC transporter. Nature 443, 180-185.

Dean, M., **2005**. The genetics of ATP-binding cassette transporters. Methods Enzymol 400, 409-429.

Dean, M., Rzhetsky, A., Allikmets, R., **2001**. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11, 1156-1166.

Dempsey, C.E., Mason, P.E., Jungwirth, P., **2011**. Complex ion effects on polypeptide conformational stability: chloride and sulfate salts of guanidinium and tetrapropylammonium. J Am Chem Soc 133, 7300-7303.

Derewenda, Z.S., **2004**. Rational protein crystallization by mutational surface engineering. Structure 12, 529-535.

DeRosier, D.J., Klug, A., **1968**. Reconstruction of Three Dimensional Structures from Electron MicrographsNature, 217. 130–134

Di Pietro, A., Dayan, G., Conseil, G., Steinfels, E., Krell, T., Trompier, D., Baubichon-Cortay, H., Jault, J., **1999**. P-glycoprotein-mediated resistance to chemotherapy in cancer cells: using recombinant cytosolic domains to establish structure-function relationships. Braz J Med Biol Res 32, 925-939.

Donowitz, M., Cha, B., Zachos, N.C., Brett, C.L., Sharma, A., Tse, C.M., Li, X., **2005**. NHERF family and NHE3 regulation. J Physiol 567, 3-11.

Eskandari, S., Wright, E.M., Kreman, M., Starace, D.M., Zampighi, G.A., **1998**. Structural analysis of cloned plasma membrane proteins by freeze-fracture electron microscopy. Proc Natl Acad Sci U S A 95, 11235-11240.

Faria, D., Dahimène, S., Alessio, L., Scott-Ward, T., Schreiber, R., Kunzelmann, K., Amaral, M.D., **2011**. Effect of Annexin A5 on CFTR: regulated traffic or scaffolding? Mol Membr Biol 28, 14-29.

Fernández, José-Jesús., Valpuesta, José-María., **2009**. Single Particle EM. eLS.

Figeys, D., McBroom, L.D., Moran, M.F., **2001**. Mass spectrometry for the study of protein-protein interactions. Methods 24, 230-239.

Ford, R.C., Hefti, A., Engel, A., **1990**. Ordered arrays of the photosystem I reaction centre after reconstitution: projections and surface reliefs of the complex at 2 nm resolution. EMBO J 9, 3067-3075.

Ford, Robert C., Birtley, James., Rosenberg, Mark F., and Zhang, Liang., **2011**. CFTR Three-Dimensional Structure (Methods in Molecular Biology). Volume 741, Part 4, 329-346.

Fouassier, L., Yun, C. C., Fitz, J. G. & Doctor, R. B., **2000**. Evidence for Ezrin-Radixin-Moesin-Binding Phosphoprotein 50 (Ebp50) Self-Association through Pdz-Pdz Interactions. The Journal of Biological Chemistry, 275, 32, pp. 25039-25045.

Frank, J., **2002**. Single-particle imaging of macromolecules by cryo-electron microscopy. Annu Rev Biophys Biomol Struct 31, 303-319.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., **1996**. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116, 190-199.

Frishman, D., Mewes, H.W., **1997**. Protein structural classes in five complete genomes. Nat Struct Biol 4, 626-628.

Gabriel, S.E., Brigman, K.N., Koller, B.H., Boucher, R.C., & Stutts, M.J., **1994**. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science. 266, 107-109.

Gadsby, D.C., Nairn, A.C., **1999**. Control of CFTR channel gating by phosphorylation and nucleotide hydrolysis. Physiol Rev 79, S77-S107.

Gentzsch, M., Aleksandrov, A., Aleksandrov, L., Riordan, J.R., **2002**. Functional analysis of the C-terminal boundary of the second nucleotide binding domain of the cystic fibrosis transmembrane conductance regulator and structural implications. Biochem J 366, 541-548.

Glaeser, R.M., **1999**. Review: electron crystallography: present excitement, a nod to the past, anticipating the future. J Struct Biol 128, 3-14.

Glavinas, H., Krajcsi, P., Cserepes, J., Sarkadi, B., **2004**. The role of ABC transporters in drug resistance, metabolism and toxicity. Curr Drug Deliv 1, 27-42.

Gomes-Alves, P., Couto, F., Pesquita, C., Coelho, A.V., Penque, D., **2010**. Rescue of F508del-CFTR by RXR motif inactivation triggers proteome modulation associated with the unfolded protein response. Biochim Biophys Acta 1804, 856-865.

Goodman, B.E., Percy, W.H., **2005**. CFTR in cystic fibrosis and cholera: from membrane transport to clinical practice. Adv Physiol Educ 29, 75-82.

Gout, T., **2012**. Role of ATP binding and hydrolysis in the gating of the cystic fibrosis transmembrane conductance regulator. Ann Thorac Med.;7(3):115-21.

Gregory, R.J., Cheng, S.H., Rich, D.P., Marshall, J., Paul, S., Hehir, K., Ostedgaard, L., Klinger, K.W., Welsh, M.J., Smith, A.E., **1990**. Expression and characterization of the cystic fibrosis transmembrane conductance regulator. Nature 347, 382-386.

Gros, P., Croop, J., Housman, D., **1986**. Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47, 371-380.

Guerra, L., Fanelli, T., Favia, M., Riccardi, S. M., Busco, G., Cardone, R. A., Carrabino, S., Weinman, E. J., Reshkin, S. J., Conese, M., & Casavola, V., **2005**. Na+/H+ Exchanger Regulatory Factor Isoform 1 Overexpression Modulates Cystic Fibrosis Transmembrane Conductance Regulator (Cftr) Expression and Activity in Human Airway 16hbe14o- Cells and Rescues Deltaf508 Cftr Functional Expression in Cystic Fibrosis Cells. The Journal of Biological Chemistry, 280, 49, pp. 40925-40933.

Guggino, W.B., Stanton, B.A., **2006**. New insights into cystic fibrosis: molecular switches that regulate CFTR. Nat Rev Mol Cell Biol 7, 426-436.

Gunawardena, H.P., He, M., Chrisman, P.A., Pitteri, S.J., Hogan, J.M., Hodges, B.D., McLuckey, S.A., **2005**. Electron transfer versus proton transfer in gas-phase ion/ion reactions of polyprotonated peptides. J Am Chem Soc 127, 12627-12639.

Hall, R.A., Ostedgaard, L.S., Premont, R.T., Blitzer, J.T., Rahman, N., Welsh, M.J., Lefkowitz, R.J., **1998**. A C-terminal motif found in the beta2-adrenergic receptor, P2Y1 receptor and cystic fibrosis transmembrane conductance regulator determines binding to the Na+/H+ exchanger regulatory factor family of PDZ proteins. Proc Natl Acad Sci U S A 95, 8496-8501.

Hall, R.A., Spurney, R.F., Premont, R.T., Rahman, N., Blitzer, J.T., Pitcher, J.A., Lefkowitz, R.J., **1999**. G protein-coupled receptor kinase 6A phosphorylates the Na(+)/H(+) exchanger regulatory factor via a PDZ domain-mediated interaction. J Biol Chem 274, 24328-24334.

Harpaz, Y., Gerstein, M., Chothia, C., **1994**. Volume changes on protein folding. Structure 2, 641-649.

Henderson, R., **1995**. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys 28, 171-193.

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., Downing, K.H., **1990**. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J Mol Biol 213, 899-929.

Higgins, C.F., **1992**. ABC transporters: from microorganisms to man. Annu Rev Cell Biol 8, 67-113.

Higgins, C.F., **2001**. ABC transporters: physiology, structure and mechanism--an overview. Res Microbiol 152, 205-210.

Higgins, C. F., Linton, K. J., **2004**. The ATP switch model for ABC transporters. Nat Struct Mol Biol.;11:918–26.

Hipfner, DR,. Deeley, RG., **1999**. Cole, SPC., Structural, mechanistic and clinical aspects of MRP1. Biochim Biophys Acta 1461: 359–376.

Holland, B. Cole, S., Kuchler, K. and Higgins, c., **2003**. ABC Proteins From Bacteria To Man . California : Elsevier science Ltd.

Hoppe, W., Langer, R., Knesch, G., Poppe, C., **1968**. [Protein crystal structure analysis with electron radiation]. Naturwissenschaften 55, 333-336.

Hu, W., Howard, M., Lukacs, G.L., **2001**. Multiple endocytic signals in the C-terminal tail of the cystic fibrosis transmembrane conductance regulator. Biochem J 354, 561-572.

Huang, P., Liu, Q., Scarborough, G.A., **1998**. Lysophosphatidylglycerol: a novel effective detergent for solubilizing and purifying the cystic fibrosis transmembrane conductance regulator. Anal Biochem 259, 89-97.

Hwang, T.C., Sheppard, D.N., **2009**. Gating of the CFTR Cl- channel by ATP-driven nucleotide-binding domain dimerisation. J Physiol 587, 2151-2161.

Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.E., Higgins, C.F., **1990**. Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. Nature 346, 362-365.

Ivancic, M., Spuches, A.M., Guth, E.C., Daugherty, M.A., Wilcox, D.E., Lyons, B.A., **2005**. Backbone nuclear relaxation characteristics and calorimetric investigation of the human Grb7-SH2/erbB2 peptide complex. Protein Sci 14, 1556-1569.

Jones, D.T., **1999**. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292, 195-202.

Jurinke, C., Oeth, P., van den Boom, D., **2004**. MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. Mol Biotechnol 26, 147-164.

Kaminski, W.E., Piehler, A., Wenzel, J.J., **2006**. ABC A-subfamily transporters: structure, function and disease. Biochim Biophys Acta 1762, 510-524.

Karas, M., Bahr, U., Strupat, K., Hillenkamp, F., Tsarbopoulos, A., Pramanik. B., **1995**. Matrix Dependence of Metastable Fragmentation of Glycoproteins in MALDI TOF Mass Spectrometry. Anal. Chem., 67 (3), 675–679.

Karas, M., Hillenkamp, F., **1988**. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60, 2299-2301.

Karthikeyan, S., Leung, T., Birrane, G., Webster, G., Ladias, J.A., **2001**. Crystal structure of the PDZ1 domain of human Na(+)/H(+) exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains. J Mol Biol 308, 963-973.

Karthikeyan, S., Leung, T., Ladias, J.A., **2002**. Structural determinants of the Na+/H+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors. J Biol Chem 277, 18973-18978.

Kartner, N., Hanrahan, J.W., Jensen, T.J., Naismith, A.L., Sun, S.Z., Ackerley, C.A., Reyes, E.F., Tsui, L.C., Rommens, J.M., Bear, C.E., **1991**. Expression of the cystic fibrosis gene in non-epithelial invertebrate cells produces a regulated anion conductance. Cell 64, 681-691.

Kelly, S.M., Price, N.C., **2000**. The use of circular dichroism in the investigation of protein structure and function. Curr Protein Pept Sci 1, 349-384.

Kelly, S.M., Jess, T.J., Price, N.C., **2005**. How to study proteins by circular dichroism. Biochim Biophys Acta 1751, 119-139.

Kiser, G.L., Gentzsch, M., Kloser, A.K., Balzi, E., Wolf, D.H., Goffeau, A., Riordan, J.R., **2001**. Expression and degradation of the cystic fibrosis transmembrane conductance regulator in Saccharomyces cerevisiae. Arch Biochem Biophys 390, 195-205.

Knol, J., Sjollema, K., Poolman, B., **1998**. Detergent-mediated reconstitution of membrane proteins. Biochemistry 37, 16410-16415.

Ko, Y.H., Thomas, P.J., Pedersen, P.L., **1994**. The cystic fibrosis transmembrane conductance regulator. Nucleotide binding to a synthetic peptide segment from the second predicted nucleotide binding fold. J Biol Chem 269, 14584-14588.

Ko, Y.H., Thomas, P.J., Delannoy, M.R., Pedersen, P.L., **1993**. The cystic fibrosis transmembrane conductance regulator. Overexpression, purification, and characterization of wild type and delta F508 mutant forms of the first nucleotide binding fold in fusion with the maltose-binding protein. J Biol Chem 268, 24330-24338.

Koeck, P.J., Purhonen, P., Alvang, R., Grundberg, B., Hebert, H., **2007**. Single particle refinement in electron crystallography: a pilot study. J Struct Biol 160, 344-352.

Krissinel, E., Henrick, K., **2007**. Inference of macromolecular assemblies from crystalline state. J Mol Biol 372, 774-797.

Kunzelmann, Karl., **2001**. CFTR: Interacting With Everything?. Physiology. vol. 16 no. 4 167-170

Ladias, J.A., **2003**. Structural insights into the CFTR-NHERF interaction. J Membr Biol 192, 79-88.

Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., Sigler, P.B., **1996**. The 2.0 A crystal structure of a heterotrimeric G protein. Nature 379, 311-319.

Larsen, T.A., Olson, A.J., Goodsell, D.S., **1998**. Morphology of protein-protein interfaces. Structure 6, 421-427.

LaRusch, J.A, **2007**. A novel internal binding motif in the CFTR C-terminus enhances EBP50 multimerization and facilitates endocytic recycling. Doctor of Philosophy. The Johns Hopkins University. United State.

Lawson, D.M., Artymiuk, P.J., Yewdall, S.J., Smith, J.M., Livingstone, J.C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., **1991**. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. Nature 349, 541-544.

Lee, M.G., Choi, J.Y., Luo, X., Strickland, E., Thomas, P.J., Muallem, S., **1999**. Cystic fibrosis transmembrane conductance regulator regulates luminal Cl-/HCO3- exchange in mouse submandibular and pancreatic ducts. J Biol Chem 274, 14670-14677.

Lewis, H.A., Buchanan, S.G., Burley, S.K., Conners, K., Dickey, M., Dorwart, M., Fowler, R., Gao, X., Guggino, W.B., Hendrickson, W.A., Hunt, J.F., Kearins, M.C., Lorimer, D., Maloney, P.C., Post, K.W., Rajashankar, K.R., Rutter, M.E., Sauder, J.M., Shriver, S., Thibodeau, P.H., Thomas, P.J., Zhang, M., Zhao, X., Emtage, S., **2004**. Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. EMBO J 23, 282-293.

Lewis, H.A., Zhao, X., Wang, C., Sauder, J.M., Rooney, I., Noland, B.W., Lorimer, D., Kearins, M.C., Conners, K., Condon, B., Maloney, P.C., Guggino, W.B., Hunt, J.F., Emtage, S., **2005**. Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem.;280(2):1346-53.

Lewis, H.A., Wang, C., Zhao, X., Hamuro, Y., Conners, K., Kearins, M.C., Lu, F., Sauder, J.M., Molnar, K.S., Coales, S.J., Maloney, P.C., Guggino, W.B., Wetmore, D.R., Weber, P.C., Hunt, J.F., **2010**. Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. J Mol Biol. ;396(2):406-30.

Li, C., Naren, A.P., **2010.** CFTR Chloride Channel in the Apical Compartments: Spatiotemporal Coupling to its Interacting Partners. Integr Biol (Camb).; 2(4): 161–177.

Li, C., Naren, A.P., **2011**. Analysis of CFTR interactome in the macromolecular complexes. Methods Mol Biol 741, 255-270.

Li, C., Roy, K., Dandridge, K., Naren, A.P., **2004**. Molecular Assembly of Cystic Fibrosis Transmembrane Conductance Regulator in Plasma Membrane. The Journal of Biological Chemistry, 279, 24673-24684.

Li, J., Dai, Z., Jana, D., Callaway, D. J.E., & Bu, Z., **2005**. Ezrin Controls the Macromolecular Complexes Formed between an Adapter Protein Na+/H+ Exchanger Regulatory Factor and the Cystic Fibrosis Transmembrane Conductance Regulator. The Journal of Biological Chemistry, 280, 45, pp. 37634-37643.

Linsdell, P., Hanrahan, J.W., **1998**. Adenosine triphosphate-dependent asymmetry of anion permeation in the cystic fibrosis transmembrane conductance regulator chloride channel. J Gen Physiol 111, 601-614.

Linsdell, P., Tabcharani, J.A., Rommens, J.M., Hou, Y.X., Chang, X.B., Tsui, L.C., Riordan, J.R., Hanrahan, J.W., **1997**. Permeability of wild-type and mutant cystic fibrosis transmembrane conductance regulator chloride channels to polyatomic anions. J Gen Physiol 110, 355-364.

Loboda, A.V., Krutchinsky, A.N., Bromirski, M., Ens, W., Standing, K.G., **2000**. A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance. Rapid Commun Mass Spectrom: 14(12):1047-57.

Lodish, H., Berk, A., Zipursky, S.L., et al., **2000**. Molecular Cell Biology. 4th edition. New York: W. H. Freeman

Loll, P.J., **2003**. Membrane protein structural biology: the high throughput challenge. J Struct Biol 142, 144-153.

Lu, J., Shen, S., Li, R., Lin, Q., and Zheng, M., **2012**. Two-dimensional (2D) crystallization of engineered proteins: A review. African Journal of Microbiology Research Vol. 6(1), pp. 7-15.

Lu, N.T., Pedersen, P.L., **2000**. Cystic fibrosis transmembrane conductance regulator: the purified NBF1+R protein interacts with the purified NBF2 domain to form a stable NBF1+R/NBF2 complex while inducing a conformational change transmitted to the C-terminal region. Arch Biochem Biophys 375, 7-20.

Ludtke, S.J., Baldwin, P.R., Chiu, W., **1999**. EMAN: semiautomated software for high-resolution single-particle reconstructions. J Struct Biol 128, 82-97.

Ludtke, S.J., Jakana, J., Song, J.L., Chuang, D.T., Chiu, W., **2001**. A 11.5 A single particle reconstruction of GroEL using EMAN. J Mol Biol 314, 253-262.

Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., Chiu, W., **2004**. Seeing GroEL at 6 A resolution by single particle electron cryomicroscopy. Structure 12, 1129-1136.

Lukacs, G.L., Chang, X.B., Kartner, N., Rotstein, O.D., Riordan, J.R., & Grinstein, S., **1992**. The Cystic Fibrosis Transmembrane Regulator Is Present and Functional in Endosomes. Role as a Determinant of Endosomal Ph. The Journal of Biological Chemistry, 267, 21, pp. 14568-14572.

Lukacs, G.L., Mohamed, A., Kartner, N., Chang, X.B., Riordan, J.R., 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.

Lunn, C.A., **2010**. Progress in Molecular Biology and Translational Science. Membrane Proteins as Drug Targets. Elsevier Inc. Volume 91, Pages 1-247.

Marshall, A.G., and Hendrickson, C.L., **2008.** High-Resolution Mass Spectrometers. Annu. Rev. Anal. Chem. 1:579–99.

Marshall, J., Fang, S., Ostedgaard, L.S., O'Riordan, C.R., Ferrara, D., Amara, J.F., Hoppe, H.T., Scheule, R.K., Welsh, M.J., Smith, A.E., et al., **1994**. Stoichiometry of recombinant cystic fibrosis transmembrane conductance regulator in epithelial cells and its functional reconstitution into cells in vitro. J Biol Chem;269:2987–2995.

Massiah, M.A., Ko, Y.H., Pedersen, P.L., Mildvan, A.S., **1999**. Cystic fibrosis transmembrane conductance regulator: solution structures of peptides based on the Phe508 region, the most common site of disease-causing DeltaF508 mutation. Biochemistry: 38(23):7453-61.

McClain, B., Settembre, E., Temple, B.R., Bellamy, A.R., Harrison, S.C., **2010**. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 A resolution. J Mol Biol 397, 587-599.

McElroy, H.E., Sisson, G.W., Schoettlin, W.E., Aust, R.M., and Villa-franca, J.E., **1992**. Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. J. Cryst. Growth 122: 265–272.

Mendoza, J.L., Schmidt, A., Li, Q., Nuvaga, E., Barrett, T., Bridges, R.J., Feranchak, A.P., Brautigam, C.A., Thomas, P.J., **2012**. Requirements for efficient correction of ΔF508 CFTR revealed by analyses of evolved sequences. Cell 148, 164-174.

Mickle, J.E., Macek, M.Jr., Fulmer-Smentek, S.B., Egan, M.M., Schwiebert, E., et al., **1998**. A mutation in the cystic fibrosis transmembrane conductance regulator gene associated with elevated sweat chloride concentrations in the absence of cystic fibrosis. Hum. Mol. Genet. 7:729–35.

Milewski, M.I., Lopez, A., Jurkowska, M., Larusch, J. & Cutting, G.R., **2005**. Pdz-Binding Motifs Are Unable to Ensure Correct Polarized Protein Distribution in the Absence of Additional Localization Signals. FEBS letters, 579, 2, pp. 483-487.

Milewski, M.I., Mickle, J.E., Forrest, J.K., Stafford, D.M., Moyer, B.D., Cheng, J., Guggino, W.B., Stanton, B.A., Cutting, G.R., **2001**. A PDZ-binding motif is essential but not sufficient to localize the C terminus of CFTR to the apical membrane. J Cell Sci 114, 719-726.

Miller, J.H., **1972**. Experiments in molecular genetics. New York : Cold Spring Harbor Laboratory, Cold Spring Harbor.

Moitra, K., **2012**. ABC Transporters in Human Disease . United Kingdom: Morgan & Claypool.

Mosser, G., **2001**. Two-dimensional crystallogenesis of transmembrane proteins. Micron 32, 517-540.

Moyer, B.D., Denton, J., Karlson, K.H., Reynolds, D., Wang, S., Mickle, J.E., Milewski, M., Cutting, G.R., Guggino, W.B., Li, M. & Stanton, B.A., **1999**. A Pdz-Interacting Domain in Cftr Is an Apical Membrane Polarization Signal. The Journal of Clinical Investigation, 104, 10, pp. 1353-1361.

Moyer, B.D., Duhaime, M., Shaw, C., Denton, J., Reynolds, D., Karlson, K.H., Pfeiffer, J., Wang, S., Mickle, J.E., Milewski, M., Cutting, G.R., Guggino, W.B., Li, M. & Stanton, B.A., **2000**. The Pdz-Interacting Domain of Cystic Fibrosis Transmembrane Conductance Regulator Is Required for Functional Expression in the Apical Plasma Membrane. The Journal of Biological Chemistry, 275, 35, pp. 27069-27074.

Naren, A.P., Nelson, D.J., Xie, W., Jovov, B., Pevsner, J., Bennett, M.K., Benos, D.J., Quick, M.W., Kirk, K.L., **1997**. Regulation of CFTR chloride channels by syntaxin and Munc18 isoforms. Nature 390, 302-305.

Nawrot, M., Liu, T., Garwin, G.G., Crabb, J.W., Saari, J.C., **2006**. Scaffold proteins and the regeneration of visual pigments. Photochem Photobiol 82, 1482-1488.

Neville, D.C., Rozanas, C.R., Price, E.M., Gruis, D.B., Verkman, A.S., Townsend, R.R., **1997**. Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. Protein Sci 6, 2436-2445.

Opella, S.J., **2003**. Membrane protein NMR studies. Methods Mol Biol. 307-320.

O'Riordan, C.R., Erickson, A., Bear, C., Li, C., Manavalan, P., Wang, K.X., Marshall, J., Scheule, R.K., McPherson, J.M., Cheng, S.H., **1995**. Purification and characterization of recombinant cystic fibrosis transmembrane conductance regulator from Chinese hamster ovary and insect cells. J Biol Chem 270, 17033-17043.

O'Ryan, L., Rimington, T., Cant, N., Ford, R.C., **2012**. Expression and Purification of the Cystic Fibrosis Transmembrane Conductance Regulator Protein in Saccharomyces cerevisiae. J. Vis. Exp. (61), e3860, DOI: 10.3791/3860.

Ostedgaard, L.S., Randak, C., Rokhlina, T., Karp, P., Vermeer, D., Ashbourne Excoffon, K.J. & Welsh, M.J., **2003**. Effects of C-Terminal Deletions on Cystic Fibrosis Transmembrane Conductance Regulator Function in Cystic Fibrosis Airway Epithelia. Proceedings of the National Academy of Sciences of the United States of America, 100, 4, pp. 1937-1942.

Pasyk, E.A., & Foskett, J.K., **1995**. Mutant (Delta F508) Cystic Fibrosis Transmembrane Conductance Regulator Cl- Channel Is Functional When Retained in Endoplasmic Reticulum of Mammalian Cells. The Journal of Biological Chemistry, 270, 21, pp. 12347-12350.

Penczek, P., Zhu, J., and Frank, J., **1996**. A common-lines based method for determining orientations for N>3 particle projections simultaneously. Ultramicroscopy 63: 205-218.

Peng, S., Liu, L.P., Emili, A.Q., Deber, C.M., **1998**. Cystic fibrosis transmembrane conductance regulator: expression and helicity of a double membrane-spanning segment. FEBS Lett 431, 29-33.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., **1999**. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551-3567.

Piehler, A.P., Hellum, M., Wenzel, J.J., Kaminski, E., Haug, K.B., Kierulf, P., Kaminski, W.E., **2008**. The human ABC transporter pseudogene family: evidence for transcription and gene-pseudogene interference. BMC Genomics 9, 165.

Piehler, A.P., Wenzel, J.J., Olstad, O.K., Haug, K.B., Kierulf, P., Kaminski, W.E., **2006**. The human ortholog of the rodent testis-specific ABC transporter Abca17 is a ubiquitously expressed pseudogene (ABCA17P) and shares a common 5′ end with ABCA3. BMC Mol. Biol. 7, 28.

Prince, L.S., Peter, K., Hatton, S.R., Zaliauskiene, L., Cotlin, L.F., Clancy, J.P., Marchase, R.B., Collawn, J.F., **1999**. Efficient endocytosis of the cystic fibrosis transmembrane conductance regulator requires a tyrosine-based signal. J Biol Chem 274, 3602-3609.

Protasevich, I., Yang, Z., Wang, C., Atwell, S., Zhao, X., Emtage, S., Wetmore, D., Hunt, J.F., Brouillette, C.G., **2010**. Thermal unfolding studies show the disease causing F508del mutation in CFTR thermodynamically destabilizes nucleotide-binding domain 1. Protein Sci 19, 1917-1931.

Radermacher, M., **1988**. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. J Electron Microsc Tech 9 (4), 359-394.

Raghuram, V., Mak, D.O., Foskett, J.K., **2001**. Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction. Proc Natl Acad Sci U S A 98, 1300-1305.

Ramjeesingh, M., Kidd, J.F., Huan, L.J., Wang, Y., Bear, C.E., **2003**. Dimeric cystic fibrosis transmembrane conductance regulator exists in the plasma membrane. Biochem J;374(pt 3): 793–797.

Ramjeesingh, M., Li, C., Kogan, I., Wang, Y., Huan, L.J., Bear, C.E., **2001**. A monomer is the minimum functional unit required for channel and ATPase activity of the cystic fibrosis transmembrane conductance regulator. Biochemistry ;40:10700–10706.

Riordan, J.R., **1993**. The cystic fibrosis transmembrane conductance regulator. Annu Rev Physiol 55, 609-630.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., **1989**. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245, 1066-1073.

Riordan, J.R., **2008**. CFTR Function and Prospects for Therapy. Annu. Rev. Biochem. 77:701–726.

Rosenberg, M.F., Callaghan, R., Ford, R.C., Higgins, C.F., **1997**. Structure of the multidrug resistance P-glycoprotein to 2.5 nm resolution determined by electron microscopy and image analysis. J Biol Chem 272, 10685-10694.

Rosenberg, M.F., Kamis, A.B., Aleksandrov, L.A., Ford, R.C., Riordan, J.R., **2004**. Purification and crystallization of the cystic fibrosis transmembrane conductance regulator (CFTR). J Biol Chem 279, 39051-39057.

Rosenberg, M.F., O'Ryan, L.P., Hughes, G., Zhao, Z., Aleksandrov, L.A., Riordan, J.R., Ford, R.C., **2011**. The cystic fibrosis transmembrane conductance regulator (CFTR): three-dimensional structure and localization of a channel gate. J Biol Chem.;286(49):42647-54.

Roth, M.G., **1994**. Protein Expression in Animal Cells. Academic Press.

Ruprecht, J., Nield, J., **2001**. Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. Prog Biophys Mol Biol 75, 121-164.

Saibil, H.R., **2000**. Macromolecular structure determination by cryo-electron microscopy. Acta Crystallogr D Biol Crystallogr 56, 1215-1222.

Schroeter, J.P., Bretaudiere, J.P., **1996**. SUPRIM: easily modified image processing software. J Struct Biol 116, 131-137.

Schuck, P., **1997**. Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. Annual Reviews of Biophysics and Biomolecular Structure 26:541-566.

Schulz, E., Karas, M., Rosu, F., Gabelica, V., **2006**. Influence of the matrix on analyte fragmentation in atmospheric pressure MALDI. J Am Soc Mass Spectrom 17, 1005-1013.

Schwiebert, E.M., Benos, D.J., Egan, M.E., Stutts, M.J., Guggino, W.B., **1999a**. CFTR is a conductance regulator as well as a chloride channel. Physiol Rev 79, S145-166.

Schwiebert, L.M., Estell, K., Propst, S.M., **1999b**. Chemokine expression in CF epithelia: implications for the role of CFTR in RANTES expression. Am J Physiol 276, C700-710.

Senior, A.E., and Gadsby, D.C., **1997**. ATP hydrolysis cycles and mechanism in P-glycoprotein and CFTR. Sem Cancer Biol 8: 143–150.

Sharom, F.J., **2008**. ABC multidrug transporters: structure, function and role in chemoresistance. Pharmacogenomics 9, 105-127.

Sheng, M., Sala, C., **2001**. PDZ domains and the organization of supramolecular complexes. Annu Rev Neurosci 24, 1-29.

Shenolikar, S., Voltz, J.W., Cunningham, R., Weinman, E.J., **2004**. Regulation of ion transport by the NHERF family of PDZ proteins. Physiology (Bethesda) 19, 362-369.

Shenolikar, S., Minkoff, C.M., Steplock, D.A., Evangelista, C., Liu, M., Weinman, E.J., **2001**. N-terminal PDZ domain is required for NHERF dimerization. FEBS Lett 489, 233-236.

Sheppard, D.N., **2012**. CFTR channel pharmacology: insight from a flock of clones. Focus on "Divergent CFTR orthologs respond differently to the channel inhibitors CFTRinh-172, glibenclamide, and GlyH-101". Am J Physiol Cell Physiol 302, C24-26.

Sheppard, D.N., Welsh, M.J., **1999**. Structure and function of the CFTR chloride channel. Physiol Rev 79, S23-45.

Short, D.B., Trotter, K.W., Reczek, D., Kreda, S.M., Bretscher, A., Boucher, R.C., Stutts, M.J., Milgram, S.L., **1998**. An apical PDZ protein anchors the cystic fibrosis transmembrane conductance regulator to the cytoskeleton. J Biol Chem 273, 19797-19801.

Shukla, D., Schneider, C.P., Trout, B.L., **2011**. Complex interactions between molecular ions in solution and their effect on protein stability. J Am Chem Soc 133, 18713-18718.

Simossis, V.A., Heringa, J., **2005**. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Res 33, W289-294.

Singh, A.K., Riederer, B., Krabbenhöft, A., Rausch, B., Bonhagen, J., Lehmann, U., de Jonge, H.R., Donowitz, M., Yun, C., Weinman, E.J., Kocher, O., Hogema, B.M., Seidler, U., **2009**. Differential roles of NHERF1, NHERF2, and PDZK1 in regulating CFTR-mediated intestinal anion secretion in mice. J. Clin. Invest. 119: 540-550.

Smith, L., Page, R.C., Xu, Z., Kohli, E., Litman, P., Nix, J.C., Ithychanda, S.S., Liu, J., Qin, J., Misra, S., Liedtke, C.M., **2010**. Biochemical basis of the interaction between cystic fibrosis transmembrane conductance regulator and immunoglobulin-like repeats of filamin. J Biol Chem.;285(22):17166-76.

Sreerama, N., Woody, R.W., **2000**. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287, 252-260.

Stark, H., Dube, P., Lührmann, R., Kastner, B., **2001**. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature 409, 539-542.

Steen, A., Wiederhold, E., Gandhi, T., Breitling, R., Slotboom, D.J., **2011**. Physiological adaptation of the bacterium Lactococcus lactis in response to the production of human CFTR. Mol Cell Proteomics 10, M000052MCP000200.

Stratford, F.L.L., and Bear, C.E., **2006**. Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. Prog Respir Res, vol 34, pp 29–37.

Stutts, M.J., Canessa, C.M., Olsen, J.C., Hamrick, M., Cohn, J.A., Rossier, B.C., Boucher, R.C., **1995**. CFTR as a cAMP-dependent regulator of sodium channels. Science 269, 847-850.

Sun, C., Mierke, D.F., **2005**. Characterization of interactions of Na+/H+ exchanger regulatory factor-1 with the parathyroid hormone receptor and phospholipase C. J Pept Res 65, 411-417.

Sun, F. et al., **2000**. E3KARP mediates the association of ezrin and protein kinase A with the cystic fibrosis transmembrane conductance regulator in airway cells. J. Biol. Chem. 275, 29539–29546.

Tajika, Y., Sakai, N., Tamura, T., Yao, M., Watanabe, N., Tanaka, I., **2004**. Crystal structure of hypothetical protein PH0828 from Pyrococcus horikoshii. Proteins 57, 862-865.

Thelin, W.R., Chen, Y., Gentzsch, M., Kreda, S.M., Sallee, J.L., Scarlett, C.O., Borchers, C.H., Jacobson, K., Stutts, M.J., Milgram, S.L., **2007**. Direct interaction with filamins modulates the stability and plasma membrane expression of CFTR. J Clin Invest 117, 364-374.

Therien, A.G., Grant, F.E., Deber, C.M., **2001**. Interhelical hydrogen bonds in the CFTR membrane domain. Nat Struct Biol 8, 597-601.

Therien, A.G., Glibowicka, M., Deber, C.M., **2002**. Expression and purification of two hydrophobic double-spanning membrane proteins derived from the cystic fibrosis transmembrane conductance regulator. Protein Expr Purif 25, 81-86.

Thibodeau, P.H., Brautigam, C.A., Machius, M., Thomas, P.J., **2005**. Side chain and backbone contributions of Phe508 to CFTR folding. Nat Struct Mol Biol 12, 10-16.

Thiriet, M., **2007**. Biology and Mechanics of Blood Flows. New York. Springer, 57-91.

Thuman-Commike, P.A., **2001**. Single particle macromolecular structure determination via electron microscopy. FEBS Lett: 505(2):199-205.

Townsend, R.R., Lipniunas, P.H., Tulk, B.M., Verkman, A.S., **1996**. Identification of protein kinase A phosphorylation sites on NBD1 and R domains of CFTR using electrospray mass spectrometry with selective phosphate ion monitoring. Protein Sci 5, 1865-1873.

Tu, Z.-L., Li, J.-N., Chin, K.-H., Chou, C.-C., Lee, C.-C., Shr, H.-L., Gao, F. P., Lyu, P.-C., Wang, A. H.-J. and Chou, S.-H., **2005**. Cloning, expression, crystallization and preliminary X-ray analysis of a putative multiple antibiotic resistance repressor protein (MarR) from Xanthomonas campestris. Acta. Cryst. F61, 706-708.

Tusnády, G.E., Sarkadi, B., Simon, I., Váradi, A., **2006**. Membrane topology of human ABC proteins. FEBS Lett 580, 1017-1022.

Unwin, P.N., Henderson, R., **1975**. Molecular structure determination by electron microscopy of unstained crystalline specimens. J Mol Biol 94, 425-440.

Van Goor, F., Hadida, S., Grootenhuis, P.D., Burton, B., Cao, D., Neuberger, T., Turnbull, A., Singh, A., Joubran, J., Hazlewood, A., Zhou, J., McCartney, J., Arumugam, V., Decker, C., Yang, J., Young, C., Olson, E.R., Wine, J.J., Frizzell, R.A., Ashlock, M., Negulescu, P., **2009**. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. Proc Natl Acad Sci U S A 106, 18825-18830.

Van Heel, M., **1987**. Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy 21, 111-123.

Van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., **1996**. A new generation of the IMAGIC image processing system. J Struct Biol 116, 17-24.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z.,

Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al., **2001**. The sequence of the human genome. Science 291, 1304-1351.

Vergani, P., Lockless, S.W., Nairn, A.C., Gadsby, D.C., **2005**. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. Nature 433, 876-880.

Voltz, J.W., Brush, M., Sikes, S., Steplock, D., Weinman, E.J., Shenolikar, S., **2007**. Phosphorylation of PDZ1 domain attenuates NHERF-1 binding to cellular targets. J Biol Chem 282, 33879-33887.

Walz, T., Grigorieff, N., **1998**. Electron Crystallography of Two-Dimensional Crystals of Membrane Proteins. J Struct Biol 121, 142-161.

Wang, S., Raab, R.W., Schatz, P.J., Guggino, W.B., Li, M., **1998**. Peptide binding consensus of the NHE-RF-PDZ1 domain matches the C-terminal sequence of cystic fibrosis transmembrane conductance regulator (CFTR). FEBS Lett 427, 103-108.

Wang, S., Yue, H., Derin, R.B., Guggino, W.B., Li, M., **2000**. Accessory protein facilitated CFTR-CFTR interaction, a molecular mechanism to potentiate the chloride channel activity. Cell 103, 169-179.

Ward, A., Reyes, C.L., Yu, J., Roth, C.B., Chang, G., **2007**. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proc Natl Acad Sci U S A 104, 19005-19010.

Welsh, M.J., Ramsey, B.W., **1998**. Research on cystic fibrosis: a journey from the Heart House. Am J Respir Crit Care Med 157, S148-154.

Wenzel, J.J., Piehler, A., Kaminski, W.E., **2007**. ABC A-subclass proteins: gatekeepers of cellular phospho- and sphingolipid transport. Front Biosci 12, 3177-3193.

Whitmore, L., Wallace, B.A., **2008**. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89, 392-400.

Whittaker, M., Carragher, B.O., Milligan, R.A., **1995**. PHOELIX: a package for semi-automated helical reconstruction. Ultramicroscopy 58, 245-259.

Widdicombe, J.H., Welsh, M.J., Finkbeiner, W.E., **1985**. Proc. Natl. Acad. Sci. U.S.A. 82, 6167.

Wider, G., Wüthrich, K., **1999**. NMR spectroscopy of large molecules and multimolecular assemblies in solution. Curr Opin Struct Biol 9, 594-601.

Wolde, M., Fellows, A., Cheng, J., Kivenson, A., Coutermarsh, B., Talebian, L., Karlson, K., Piserchio, A., Mierke, D.F., Stanton, B.A., Guggino, W.B., Madden, D.R., **2007**. Targeting CAL as a negative regulator of DeltaF508-CFTR cell-surface expression: an RNA interference and structure-based mutagenetic approach. J Biol Chem 282, 8099-8109.

Yang, Y., Janich, S., Cohn, J.A., Wilson, J.M., **1993**. The common variant of cystic fibrosis transmembrane conductance regulator is recognized by hsp70 and degraded in a pre-Golgi nonlysosomal compartment. Proc Natl Acad Sci U S A 90, 9480-9484.

Yeager, M., Unger, V.M., Mitra, A.K., **1999**. Three-dimensional structure of membrane proteins determined by two-dimensional crystallization, electron cryomicroscopy, and image analysis. Methods Enzymol. ;294:135-80.

Zeltwanger, S., Wang, F., Wang, G.T., Gillis, K.D., Hwang, T.C., **1999**. Gating of cystic fibrosis transmembrane conductance regulator chloride channels by adenosine triphosphate hydrolysis. Quantitative analysis of a cyclic gating scheme. J Gen Physiol 113, 541-554.

Zerhusen, B., Zhao, J., Xie, J., Davis, P.B., Ma, J. **1999**. A Single Conductance Pore for Chloride Ions Formed by Two Cystic Fibrosis Transmembrane Conductance Regulator Molecules. J. Biol. Chem. 274:7627–7630.

Zhang, L., Aleksandrov, L.A., Riordan, J.R., Ford, R.C., **2011**. Domain location within the cystic fibrosis transmembrane conductance regulator protein investigated by electron microscopy and gold labelling. Biochim Biophys Acta 1808, 399-404.

Zhang, L., Aleksandrov, L.A., Zhao, Z., Birtley, J.R., Riordan, J.R., Ford, R.C., **2009**. Architecture of the cystic fibrosis transmembrane conductance regulator protein and structural changes associated with phosphorylation and nucleotide binding. J Struct Biol 167, 242-251.

Zhang, X., Settembre, E., Xu, C., Dormitzer, P.R., Bellamy, R., Harrison, S.C., Grigorieff, N., **2008**. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. Proc Natl Acad Sci U S A 105, 1867-1872.

Zhu, T., Dahan, D., Evagelidis, A., Zheng, S., Luo, J., Hanrahan, J.W., **1999**. Association of cystic fibrosis transmembrane conductance regulator and protein phosphatase 2C. J Biol Chem 274, 29102-29107.