Available online at www.elixirpublishers.com (Elixir International Journal)

Thesis Publications – Life Science

Elixir Thesis Life Sci. 65TP1 (2013) 1-54

ISSN: 2229-712X

EXPRESSION, PURIFICATION AND CHARACTERIZATION OF CFTR C-TERMINAL REGION OF KILLIFISH

A Dissertation submitted to The University of Manchester for the Degree of M.Sc Bio Chemistry in the Faculty of Life Sciences

Submitted by
SAMYAH ALANAZI

Under the Supervision of
Professor: Robert C Ford

MANCHESTER 1824
The University of Manchester

2012

ARTICLE INFO
Article history:
Received: 17 November 2013;
Published: 30 November 2013;

Tele:
E-mail addresses: saalanazi@ksu.edu.sa

List of Contents:

List of Tables:

List of Figures:

List of Abbreviations

ABC	ATP-Binding Cassette
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
MSD	Membrane Spanning Domain
NBD	Nucleotide Binding Domain
cAMP	cyclic Adenosine Monophosphate
R	Regulatory
PDZ	PSD-95 (a 95 kDa protein involved in signaling in the post-synaptic density), Dlg (the *Drosophila* discs large protein), and ZO1 (the zonula occludens 1 protein involved in maintaining epithelial cell polarity).
COPD	Chronic obstructive pulmonary disease
CAL	CFTR-associated Ligand
NHERF	Na/H Exchanger Regulatory Factor 1
CAP70	CFTR Associated Protein 70
Shank2	SH3 and multiple ankyrin repeat domains protein 2
TGN	Trans-Golgi Network
ERC	Endosomal Recycling Compartment
EBP50	Ezrin-radixin-moesin Binding Phosphoprotein 50
E3KARP	NHE3 Kinase A Regulatory Protein
IKEPP	Intestinal and Kidney enriched PDZ Protein
PDZK1	PDZ Domain Containing Protein in Kidney 1
ERM	Ezrin Radixin Moesin
MERM	Merlin Ezrin Radixin Moesin
F508	Phenylalanine 508
MALDI-TOF	Matrix-assisted laser desorption/ionization
CD	Circular Dichroism
NMR	Nuclear magnetic resonance
IMAC	Immobilised Metal Affinity Chromatography
PCR	Polymerase Chain Reaction
3D	Three Dimensional
E.coli	Escherichia coli
dNTPs	Deoxyribonucleotide triphosphates
DMSO	Dimethyl sulfoxide
MQ	Milli-Q Purification System
EDTA	Ethylenediaminetetraacetic acid
DNA	Deoxyribonucleic acid
Bp	Base pair
Kb	Kilobyte
TAE	Tris Acetate EDTA
CIP	Calf Intestinal alkaline phosphatase
SOC	Super optimal broth with catabolic repressor
LB	Luria Bertani

IPTG	Isopropyl-β-D-thio-galactoside
DTT	Dithiothreitol
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
PMSF	Phenylmethylsulfonylfluoride
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
KDa	Kilo Dalton
His	Histidine
OD_{600}	Optical density at 600nm wavelength
BSA	Bovine serum albumin
DA	Dalton

Abstract

The transportation of specific molecules across lipid membranes is an essential function of all living organisms and a large number of specific transporters and channels have evolved to carry out this function. One of the largest transporter gene families is the ATP-binding cassette (ABC). CFTR, Cystic fibrosis trans-membrane conductance regulator, is a member of this family and its mutation cause cystic fibrosis. In order to create more effective drugs to treat cystic fibrosis, understanding the mechanism of this disease and structure of CFTR will be needed. Therefore, obtaining the 3D structure will lead to a greater understanding of CFTR. Studying the structure of the C-terminal part of CFTR is the aim of this project. This study also allowed a comparison between killifish and human orthologs. By conducting the same conditions on human orthologs of the CFTR C-terminus, Killifish C-terminal was cloned into pET-28 a (cloning and expression vector), expressed in *E.coli* and purified using Talon immobilized metal affinity chromatography. The purified protein then characterized using MALDI-TOF analysis, and Circular Dichroism. Killifish gave a higher yield compared to human ortholog, which made it more suitable for further structural and functional studies.

Declaration

No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Intellectual Property Statement

The author of this dissertation (including any appendices and/or schedules to this dissertation) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this dissertation, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered into. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the dissertation, for example graphs and tables ("Reproductions"), which may be described in this dissertation, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this dissertation, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=487), in any relevant Dissertation restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's Guidance for the Presentation of Dissertations.

Acknowledgements

I am heartily thankful to my supervisor Professor Robert C Ford for his guidance and help during the project.

It is a pleasure to thank Natasha Cant, Tracy Rimington and Naomi Pollock for their guidance and support.

Thanks to my friend Vinoja Ratnasingam for her help and encouragement.

Many thanks to Emma, Dr Thomas from the Biomolecular Analysis and Paul Fullwood from DNA sequencing in the Core research Facility of the Faculty of life science in Manchester University.

Thanks to The King Saud University in the Kingdom of Saudi Arabia for funding my MSc study.

I would like to express the deepest appreciation to my husband Naif and my children Rayan and Reema for their love and understanding.

Lastly but most importantly I would like to express my gratitude to my father, mother, brothers and beautiful sisters for their support during my educational life.

1. Introduction and Aims:

1.1 The ABC Transporter Family:

The Adenosine triphosphate (ATP) - binding cassette (ABC) transporter family consists of transporter proteins that are expressed on intracellular organelles or on the cell's outer membrane and which mediate the transport of different substances across and within the cell. The ABC transporter family is coded by the ABC genes. So far, 48 ABC genes have been described. Structurally, the ABC transporters in eukaryotes consist of 2 domains that span the membrane (MSD domains) and 2 domains that bind to nucleotides (NBD domains) (Riordan et al, 1989).

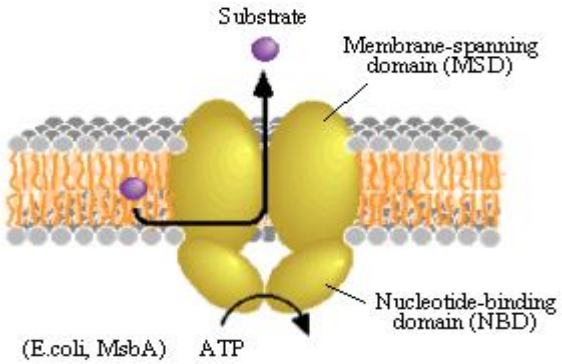


Figure 1: Structure of Eukaryotic ABC transporter.

Source: KEGG (2012b)

In prokaryotes, ABC transporters consist of 3 components namely 2 integral membrane proteins composed of 6 transmembrane segments, a lipoprotein or periplasmic subtsrate-binding protein, and 2 peripheral proteins for the binding and hydrolysis of ATP (KEGG, 2012a).

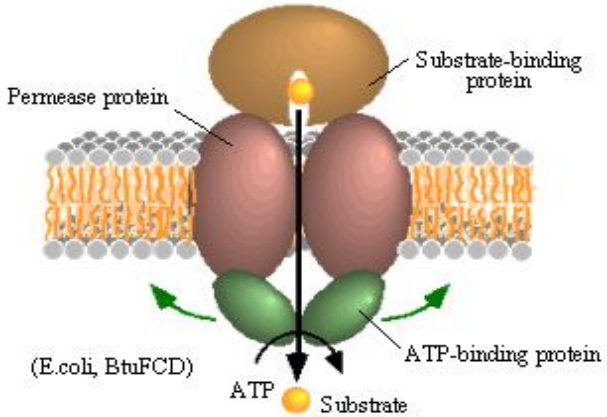


Figure 2: Structure of prokaryotic ABC transporter.

Source: KEGG(2012b)

Binding of the ATP onto the cytoplasmic domain of the ABC transporters induces the transmembrane domains to adopt an outward facing configuration. This leads to an overall movement of substances against a concentration gradient across the membrane (Kos & Fod, 2009).

1.2. The CFTR and Cystic Fibrosis:

The cystic fibrosis trans-membrane conductance regulator (CFTR) is an ion channel that is located at the apical membrane of exocrine epithelial cells and which is activated by cAMP (Mickle et al, 1998; Guggino, 2003; Dean, 2008). It is a member of the ABC transporter family and consists of 2 halves which are connected together by a regulatory (R) domain (figure 3). Each of the halves consists of 2 domains, a nucleotide binding domain (NBD) and a membrane-spanning domain (MSD). The MSD comprises of 6 transmembrane helices. The MSD forms the chloride-selective channel, NBD regulates channel gating through ATP binding and hydrolysis while The R domain regulates the activity of the channel through phosphorylation of threonine and serine residues by the cAMP-PKA pathway (Haardt, 1999; Chen et al, 2001). Functionally, CFTR mediates the transport of chloride ions and water across epithelial barriers (Fuller et al, 1992; Guggino, 2003). It also regulates bicarbonate ions transport and collaborates with inwardly rectifying potassium channels, outwardly rectifying chloride channels (ORCCs), epithelial sodium channels and a host of other proteins (Choi et al, 2001; Guggino, 2003).

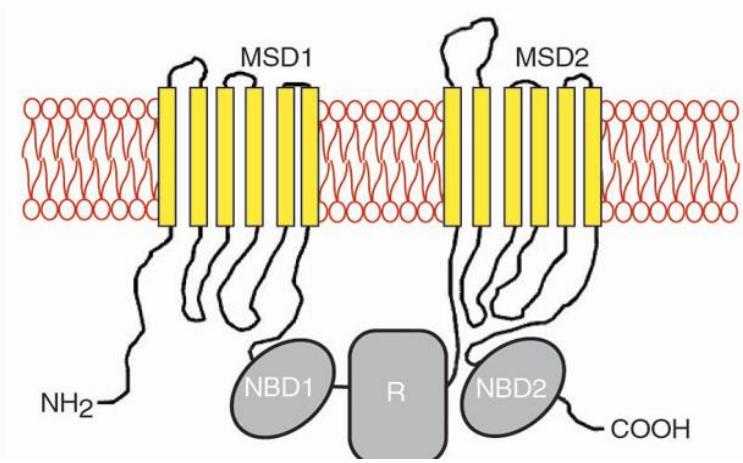


Figure 3: Structure of CFTR. Source: CFTR (2011)

In cystic fibrosis, the permeability of epithelial cells to chloride ions is drastically reduced due to the defects in the CFTR protein leading to a more negative potential difference across cell membranes causing impaired homeostasis of intracellular chloride and sodium. This leads to the production of thick, sticky mucus which obstructs all exocrine glands except sweat ducts resulting in chronic airway obstruction. This obstruction

predisposes patients to atopy, asthma, allergic bronchopulmonary aspergilosis, bacterial infections, hemoptysis, fecal impaction, nasal polyps and sinusitis, azoospermia, and pneumothorax. It can lead to complications such as malabsorption, rectal prolapse, focal billiary cirrhosis, small bowel obstruction, diabetes mellitus, COPD and exocrine pancreatic insufficiency (Sedlack et al, 2008).

1.3 Length and Boundary of the CFTR C-terminus:

There is no universal agreement on the exact length and boundary of the C-terminal region of CFTR. According to Gentzsch & Riordan (2000), the C-terminal region of CFTR is 100 residues long and stretches from residues 1380 to 1480 in humans (Gentzsch & Riordan, 2000). According to Lu & Pedersen (2000) and Walker et al (1982) however, the NBD2 and C-terminal region overlap and functionality of the boundary residues of the C-terminus may run into the NBD2 region. This issue has however been resolved conclusively. Clinical and evolutionary data and observations clearly show that the carboxyl terminus of NBD2 is E1417. In humans therefore, the length of the C-terminus spans from E1418 to L1480 and is 62 residues long. Studies supporting this assertion prove that the sequence divergence in the transitional region starts at E1418 and that the shortest deletion that is linked with cystic fibrosis ends at Q1412. In addition, there is evidence from a number of functional studies demonstrating that the synthesis, maturation and stability of CFTR is considerably affected by truncations occurring at or upstream of V1415 and that truncations occurring at N1419 or downstream of N1419 have no such effects (Zhang et al., 1998; Haardt et al., 1999; Gentzsch & Riordan 2001; CFTR, 2011). The new boundaries of the C-terminus are further supported by the observation that mutations in the region do not play a big role in the pathogenesis of cystic fibrosis. This is consistent with the relocation of the regions important for synthesis, processing, and stability of the protein to the NBD2 region (Chen et al, 2001).

1.4 Comparison of Amino Acids Contents and Functional Motifs in The C-terminus of the CFTR:

Comparison of the amino acids and functional motifs for the C-terminal region of CFTR of humans (*Homo sapiens*), Rhesus (*Macaca mulatta*), Macaque (*Macaca fascicularis*), Rabbit (*Oryctolagus cuniculus*), Sheep (*Ovis aries*), Bovine (*Bos taurus*), Rat (*Rattus norvegicus*), mouse (*Mus musculus*), frog (*Xenopus laevis*), Dogfish (*Squalus acanthias*), killifish (*Fundulus heteroclitus*), Puffer (*Tetraodon nigroviridis*) and salmon-one and salmon-2 (*Salmo salar*) was done using the NCBI database. Among the motifs indentified in the C-terminal region are the PDZ motif, the negatively charged motif, the YDSI motif, and the di-leucine motif.

1.4.1 The PDZ Motif:

The PDZ motif is formed by the terminal 4 residues and is conserved across species. PDZ domains are protein modules that mediate the interactions between proteins by binding to their C-terminals (Hall et al, 1998). Structurally, PDZ domains comprise of a sandwich structure containing 2 α-helices and 6 β-strands. These form a hydrophobic cleft that can accommodate other short peptides containing PDZ domains. The PDZ domains, also known as scaffolding proteins, are largely multivalent with several PDZ domains. This makes them capable of mediating heterotypic and homotypic interactions between proteins in different tissues (Li & Naren, 2010). The PDZ domain of CFTR is formed by the final 4 residues of its sequence namely DTRL. These residues attach to the PDZ domains of other proteins and help to place CFTR adjacent to multi-molecular regulatory complexes (Wang et al, 1998; Short et al, 1998). Substitution or deletion of the amino acid residues forming the PDZ domain of CFTR leads to impairment of apical localization in epithelial cells (Moyer et al, 1999; Moyer et al, 2000).

Except for *Salmo salar*, the terminal 3 residues are TRL. In one of the isoforms of *Salmo salar*, the terminal sequence ends in NTHL. The sequence of the PDZ motif is DTRL in 6 of the species and these are *Homo sapiens, Macaca mulatta, Macaca fascicularis, Xenopus laevis, Fundulus heteroclitus, Tetraodon nigroviridis* and one of the isoforms of *Salmo salar*. In 5 species, E replaces D in the fourth last residue and the PDZ motif comprises of the ETRL sequence. This sequence is present in *Bos taurus, Mus musculus, Rattus norvegicus, Squalus acanthias,* and *Ovis aries*. The terminal sequence of *Oryctolagus cuniculus* is GTRL while that of one isoform of *Salmon salar* is NTHL (Table 1).

Table 1: Last 26 sequences constituting the carboxyl terminal of the species above

Species	Sequence
>gi|51095110|gb|EAL24353.1|/1455-1480 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) [Homo sapiens]	SSKCKSKPQIAALKEETEEEVQDTRL
>gi|68270990|gb|AAY89007.1|/76-101 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) [Macaca mulatta]	SSKCKTQPQIAALKEETEEEVQDTRL
>gi|130501009|ref|NP_001076185.1|/1456-1481 cystic fibrosis transmembrane conductance regulator [Oryctolagus cuniculus]	SSKHKSRPQITALKEEAEEEVQGTRL
>gi|8886465|gb|AAF80467.1|AF162177_1/1456-1481 cystic fibrosis transmembrane conductance regulator [Macaca fascicularis]	SSKCKTQPQIAALKEETEEEVQDTRL
>gi|57526399|ref|NP_001009781.1|/1456-1481 cystic fibrosis transmembrane conductance regulator [Ovis aries]	SSRQRSRANIAALKEETEEEVQETKL
>gi|38322708|gb|AAR16263.1|/1456-1481 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) [Bos taurus]	SSRQRSRSNIAALKEETEEEVQETKL
>gi|192832|gb|AAA18903.1|/1451-1476 cystic fibrosis transmembrane conductance regulator [Mus musculus]	SSKHKPRTQITALKEETEEEVQETRL
>gi|91982740|ref|NP_113694.1|/1451-1476 cystic fibrosis transmembrane conductance regulator [Rattus norvegicus]	SSKQKPRTQITAVKEETEEEVQETRL
>gi|3015540|gb|AAC41271.1|/1478-1503 cystic fibrosis transmembrane conductance regulator [Fundulus heteroclitus]	KRPQPQTTKISSLPEEAEDEIQDTRL
>gi|38322781|gb|AAR16330.1|/1485-1510 cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7) [Tetraodon nigroviridis]	KRAPPQAAKISSLPEEAEDEVHDTRL
>gi|1617482|gb|AAC60023.1|/1460-1485 cystic fibrosis transmembrane conductance regulator [Xenopus laevis]	SSKRKSRPQISALQEETEEEVQDTRL
>gi|213870|gb|AAA49616.1|/1467-1492 cystic fibrosis transmembrane conductance regulator [Squalus acanthias]	SSKRKTRPKISALQEEAEEDLQETRL
>gi|185132957|ref|NP_001117006.1|/1493-1518 cystic fibrosis transmembrane conductance regulator II [Salmo salar]	RPPQSQQPKITALQEEAEDEVQDTRL
>gi|5052017|gb|AAD38404.1|AF155237_1/1494-1519 cystic fibrosis transmembrane conductance regulator I [Salmo salar]	RPPQSQQPKITALREVAEDEVQNTHL

Notes: The PDZ domain is typed in red and the negatively charged motif in green

1.4.2 The Negatively Charged Motif:

In *Homo sapiens, Macaca mulatta, Macaca fascicularis, Ovis aries, Bos taurus, Mus musculus, Rattus norvegicus,* and *Xenopus laevis*, the sequence of the negatively-charged motif is EETEEE. The tyrosine (T) is replaced by alanine (A) in *Oryctolagus cuniculus*. The sequence is EEAEDE in *Fundulus heteroclitus, Tetraodon nigroviridis* and one of the isoforms of *Salmo salar* and EEAEED and EVAEDE in *Scalas acanthias* and the second isoform of *Salmo salar* (table 1). In humans, the negatively charged motif spans residues 1469-1474 while in *Macaca mulatta, Oryctolagus cuniculus, Macaca fascicularis, Ovis aries,* and *Bos taurus* it spans residues 1470-1475. The motif consists of E1465-E1470 in both *Mus musculus* and *Rattus norvegicus*, E1492-E1497 in *Fundulus heteroclitus*, E1499-E1504 in *Tetraodon nigroviridis*, E1474-E1479 in *Xenopus laevis*, and E1481-E1486 in *Squalus acanthus.*

1.4.3 The YDSI Motif:

The YDSI motif is an internalization motif that is conserved across species. It plays an important role in the endocytic recycling of CFTR by interaction with the clathrin adaptor complex AP-2 (Weixel & Bradbury, 2001). In humans, this motif spans residues 1424 to 1427 while in rat and mouse the motifs comprise of residues 1420-1423. The C-terminal residues that make up the YDSI motif are located between residues 1425-1428 in 5 of the species (*Macaca mulatta, Oryctolagus cuniculus, Macaca fascicularis, Ovis aries*, and *Bos Taurus*). In *Oryctolagus cuniculus*, aspartic acid (D) is replaced by glutamic acid (E) .In *Rattus norvegicus*, isoleucine (I) is also replaced by leucine (L) and aspartic acid (D) is replaced by glutamic acid (E). The C-terminal region of *Squalus acanthias* has no YDSI motif (Table 2).

Table 2: Locations of the YDSI motif

Species	Motif	Residues
Homo sapiens	YDSI	1424-1427
Macaca mulatta	YDSI	1425-1428
Oryctolagus cuniculus	YESI	1425-1428
Macaca fascicularis	YDSI	1425-1428
Ovis aries	YDSI	1425-1428
Bos taurus	YDSI	1425-1428
Mus musculus	YDSL	1420-1423
Rattus norvegicus	YESL	1420-1423
Fundulus heteroclitus	YDSI	1442-1445
Tetraodon nigroviridis	YDSI	1449-1452
Xenopus laevis	YDSI	1427-1430
Squalus acanthias	-	-
Salmo salar	YDSI	1457-1460
Salmo salar	YDSI	1458-1462

1.4.4 The Di-leucine (LL) Motif:

Together with the YDSI motif, the di-leucine motif forms the internalization motifs. All the C-terminal regions of the species under investigation have a dileucine motif except *Xenopus laevis* which has no such motif in its c-terminal region. In *Ovies aries* and *Bos Taurus* however, the leucine at position 1430 is replaced by M to form an ML motif. There are 2 dileucine motifs in *Squalus acanthus*.

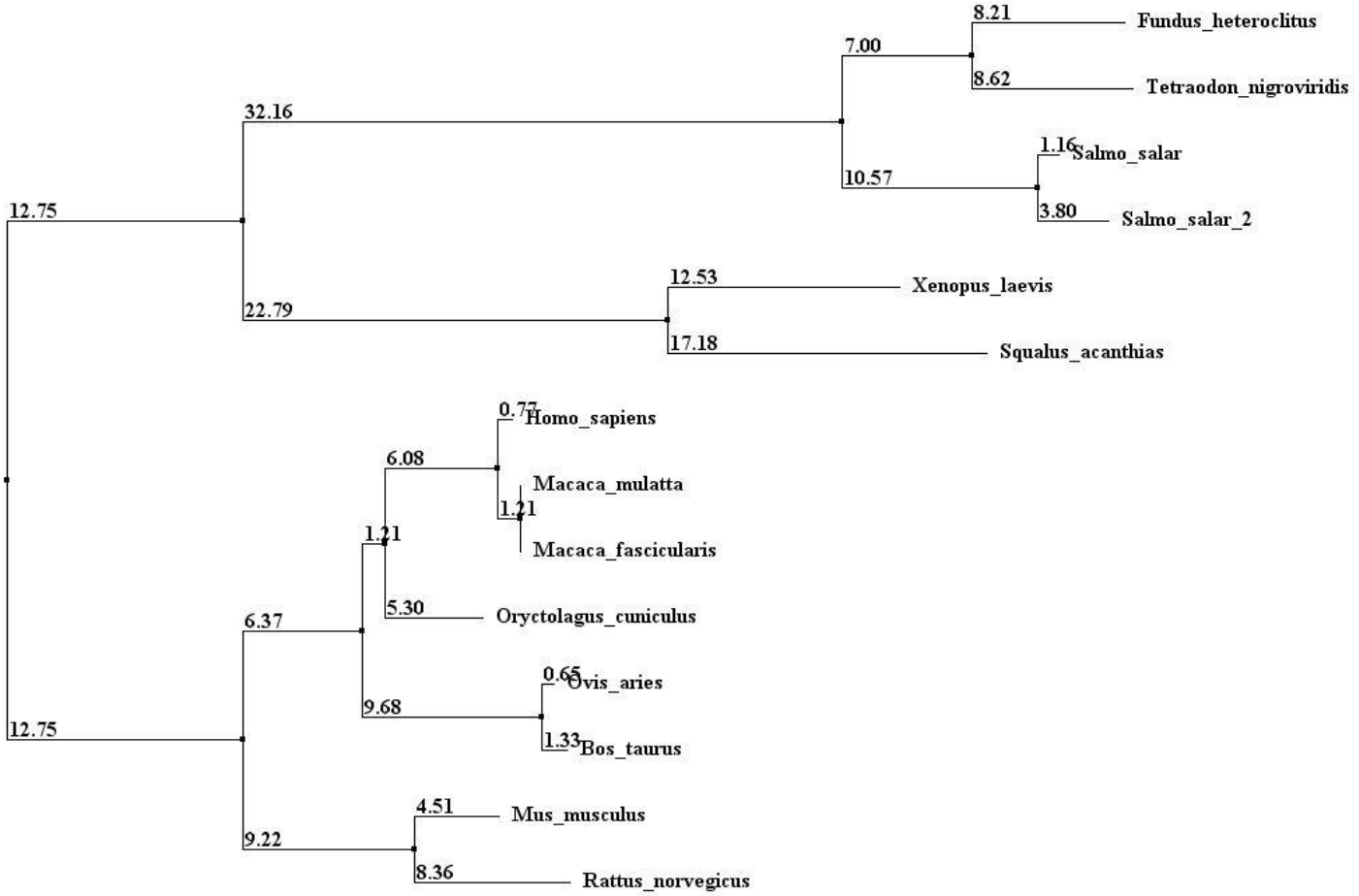
Table 3: Location of c-terminal di-leucine motifs

Species	**Motif**	**Residues**
Homo sapiens	LL	1430-1431
Macaca mulatta	LL	1431-1432
Oryctolagus cuniculus	LL	1431-1432
Macaca fascicularis	LL	1431-1432
Ovis aries	ML	1431-1432
Bos taurus	ML	1431-1432
Mus musculus	LL	1426-1427
Rattus norvegicus	LL	1420-1423
Fundulus heteroclitus	LL	1425-1426
Tetraodon nigroviridis	LL	1432-1433
Xenopus laevis	LL	1427-1430
Squalus acanthias	LL	1417-1418
Salmo salar	LL	1440-1441 and 1463-1464
Salmo salar	LL	1441-1442 and 1464-1465

1.4.5 The FLVI Motif:

In humans, the FLVI is a hydrophobic motif that spans residues 1413-1416. However, some researchers consider it to be part of the NBD2 region and not the C-terminus (Gentzsch et al, 2002).

Figure 4: Phylogenetic tree constructed using the neighbour joining method with the C-terminal sequences



1.5 The C-terminus Role in Apical Polarization and Targeting of CFTR

Rather than being distributed in a diffuse manner, ion channels are clustered and localized at specialized sub-cellular sites. These channels are organized into 3D complexes by a raft of proteins referred to as scaffolding proteins. The scaffolding proteins have PDZ domains which they use to bind onto the ion channels for the reorganization. Localization of the channels into their sub-cellular sites is mediated by scaffolding proteins with PDZ domains (Sheng, 1996; Guggino, 2004). The PDZ domain in the carboxyl terminus of CFTR plays an important role in the apical localization of the protein (Milewski et al, 2000). In CFTR, the PDZ domain possesses the (D/E) T(R/K) L motif which helps in localization of the channel (appendix 1). This motif is a type-1 domain-binding motif that is highly conserved across several species. The scaffolding proteins with PDZ domains that have been shown to bind to the PDZ motif of CFTR include the CFTR-associated ligand (CAL), the CFTR-associated protein-70, (CAP-70), and the sodium–proton exchange regulatory factor (NHE-RF) (Songyang et al, 1997).

Milewski et al (2000) demonstrated that localization of the CFTR to the apical plasma membrane can take place even without the presence of trans-membrane domains. Their study found out that the C-terminal tail alone is sufficient for this localization to occur and that the PDZ binding motif D-T-R-L> and the C-terminal regions spanning amino acids 1370-1394 and 1404-1425 are necessary for this localization (Milewski et al, 2000). Findings by Moyer et al (1999) indicate that substitution of the terminal leucine (L) of the PDZ domain of CFTR with alanine (A) or deletion of the TRL sequence of the domain leads to the mislocalization of CFTR. This finding implies that the PDZ motif plays a crucial role in the localization of CFTR at the apical membrane of epithelial cells. Milewski et al (2000) found out that localization of CFTR that has no PDZ domain is on both apical and basolateral membranes (Milewski et al, 2000).

However, recent studies demonstrate that the PDZ domain is not the only absolute required for CFTR to localize to the apical membrane of epithelial cells. Neither deletions in the C-terminal region, nor attachment of epitope tags to the region have any apparent effect on CFTR localization in tracheal, intestinal, pancreatic duct, or renal distal tubules epithelial tissue. No effect was also observed with over-expression of a dominant mutant negative for NHERF1 (Benharouga et al, 2003). These findings were corroborated by Ostedgaard et al (2003) who discovered that CFTR constructs with deletions in the C-terminal region could still express the CFTR channel in the apical membrane of airway epithelia and did not

interfere with the generation of chloride currents. This led to the conclusion that the terminal 26 residues of the PDZ domain are not the only determining factors in the localization of CFTR and that there are other sorting signals required for the localization to take place. These other signals multivalent scaffolding proteins such as NHERF1, NHERF3, CAL, CAP70 and Shank2 (Li & Naren, 2010).

Localization of the CFTR on the apical membrane can be explained using 2 models. In the first model, it is assumed that direct apical targeting is the cause of the steady-state localization of CFTR at the apical membrane. The apical targeting comprises of effective sorting of proteins in the TGN and probably a process for regulating the docking and fusion of vesicles at the apical membrane. Due to the involvement of vesicle trafficking, this model suggests that association of the membrane with the transported protein may be vital for the efficacy of the localization signal. Another aspect of the model is the transport of CFTR to the plasma membrane in a non-specific manner and through the default pathway. This aspect rationalizes the localization of CFTR on the apical and basolateral membranes following disruption of the signal responsible for apical localization (Milewski et al, 2000). The second model proposes that transportation of CFTR to different plasma membrane domains is random. The random transportation is then followed by selective stabilization of CFTR in the apical membrane. This model depends on interactions between proteins (Milewski et al, 2000).

The test carried out by Milewski et al (2000) involved construction of green fluorescent protein tagged CFTR (GFP-CFTR) plasmids and their mutants using site-directed mutagenesis. Cultivation of cell lines was in Madin-Darby canine kidney cells type II (MDCK II) and MDCK, HeLa, LLC-PKI, HEK293, and IB3-I cells used for analysis of CFTR localization. Immunostaining was done using anti-ZO-1 and anti-Na+/K+ ATPase. Analysis was done using western blot analysis, and electron and confocal laser microscopy (Milewski et al, 2000).

1.6 The Role of the C-terminus of CFTR in Recycling and Endocytosis

Apart from its role in localization of CFTR in the apical membrane, the C-terminal region also plays an important role in recycling and endocytosis. The dynamism of mature CFTR is underscored by the rapid alternate processes of endocytosis and recycling which it undergoes. Once it reaches the plasma membrane, CFTR is encased into clathrin-coated vesicles and rapidly sent back to the surface of the cell through endocytic recycling (Lukacs et al, 1997; Prince et al, 1999). The recycling and endocytic processes are important as they

help to ensure that CFTR is maintained at steady state levels, it is localized correctly at the apical membrane, and that proteins that are misfolded or defective are effectively removed. Recycling occurs at the endosomal recycling compartment (ERC) and is mediated by clathrins. This clathrin-mediated endocytosis is under the regulation of AP-2 and RabGTPases while transport to the early endosome form the plasma membrane is under the regulation of dynamin, Rab5, and clathrin (La Rusch, 2007). Endocytosis is mediated by signals from the Y-D-S-I motif and the di-leucine motif which are both present in the C-terminal region of CFTR (Prince et al, 1999; Weixel & Bradbury, 2000). Proteins which regulate endocytosis include the AP-1 and AP-2 complexes, dynamin, and clathrin. Ap-1 traffics CFTR from the TGN to the apical membrane while Ap-2 interacts directly with CFTR through its C-terminal to cause exocytosis of CFTR from the apical membrane.

Regulators of early endocytosis include Myosin VI and Rab5 which mediate the entry of CFTR into the early endosome. Thereafter Rab11 may recycle CFTR back to the membrane or Rab7 may send it to the late endosome. While in the late endosome, Rab9 can save the protein from being degraded by lysosomes and shuttle it to the TGN from where it is sent back via the recycling endosome to the apical membrane in a process that is under the regulation of Rab11. The role of Rab4 in endocytosis is yet to be conclusively determined but may control recycling of the early endosome back to the membrane or endocytosis (Gentzsch et al, 2004; Swiatecka-Urban et al, 2004; Saxena et al, 2006). Dynamin 2 K44A is a GTPase that inhibits clathrin-induced endocytosis and formation of vesicles in the Golgi apparatus (Li & Naren, 2010). Cheng et al (2004) and Cheng et al (2005) show that when dominant-negative dynamin 2 K44A is over-expressed, the expression of CFTR on the surface of cells is considerably increased and restored in cells over-expressing CAL.

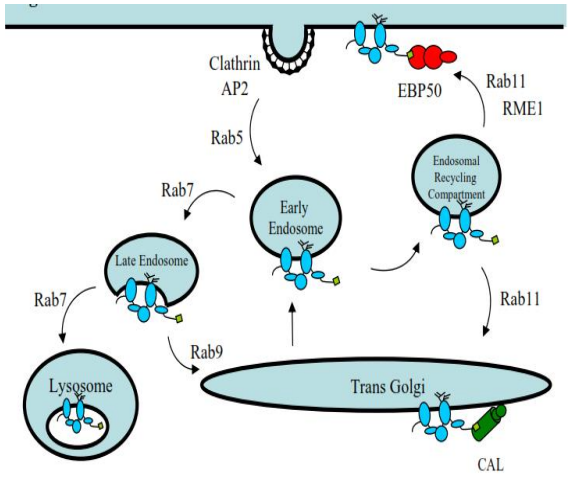


Figure 5: Endocytic recycling of CFTR

Source: LaRusch (2007)

1.7 The Interaction between the C-terminus of CFTR and Scaffolding Proteins:

The C-terminus of CFTR can interact with a number of scaffolding proteins which bind its C-terminal end and which are localized at the apical membrane of epithelial cells. These scaffolding proteins include the NHERF1 (Na/H exchanger regulatory factor 1 or ezrin-radixin-moesin binding phosphoprotein-50, EBP50), and NHERF2 (or the NHE3 kinase A regulatory protein, E3KARP). Others are the CFTR-associated protein 70 (Cap70), intestinal and kidney-enriched PDZ protein (IKEPP) and Shank 2 (SH3 and Ankyrin repeats containing protein 2). CAP-70 is also known as NHERF3 or PDZ domain-containing protein in kidney 1 (PDZK1), IKEPP as NHERF4 or PDZK2 and Shank 2 as Cortactin-binding protein 1 (CortBP1) (Li & Naren, 2010). These interactions are important for the correct localization of the ion channel (Songyang et al, 1997)

1.7.1 NHE-RF1 (EBP50):

Also referred to as or ezrin/radixin/moesin-binding phosphoprotein-50 (EBP50), NHE-RF1 has a molecular weight of 42kDa. It preferably interacts with CFTR through its PDZ2 domain. It can also interact with CFTR via the PDZ1 domain although CFTR has a markedly low affinity for PDZ1 (Short et al, 1998; Raghuram et al, 2001; Guggino, 2003). Expression of NHERF2 is ubiquitous and is not limited to any particular tissues. Localization of NHERF2 is on the apical membrane. Structurally, NHERF2 consists of 2 PDZ domains and a C-terminal ERM domain. The ERM domain induces the interaction between merlin-ezrin-radixin-moesin (MERM) proteins and CFTR leading to the linkage of the ion channel to the cytoskeleton (Biber et al, 2005; Brone & Eggermont, 2005; Lamprecht & Seidler, 2006).

Figure 6: Diagrammatic illustration of the NHERF1 scaffolding protein. The NHERF1 protein consists of 2 PDZ domains (PDZ1 and PDZ2) and an ERM domain.

Source: Li & Naren (2010)

It has been demonstrated that the ERM domain in the C-terminal region of NHERF1 ties NHERF1 to the elements of the cortical cytoskeleton through binding to ezrin and this has been interpreted to mean that interaction between CTFR and NHERF1 can attach the CFTR channel to the apical membrane cytoskeleton (Reczek & Bretscher, 1998; Short et al, 1998).

1.7.2 NHE-RF 2:

NHERF2 has a 52% amino acid sequence homology with NHE-RF1 and its expression is ubiquitous and is not limited to any particular tissues. Localization of NHERF2 is on the apical membrane. Structurally, NHERF2 consists of 2 PDZ domains and a C-terminal ERM domain. The ERM domain induces the interaction between merlin-ezrin-radixin-moesin (MERM) proteins and CFTR leading to the linkage of the ion channel to the cytoskeleton (Biber et al, 2005; Brone & Eggermont, 2005; Lamprecht & Seidler, 2006).

Figure 7: Diagrammatic illustration of the NHERF2 scaffolding protein. The NHERF2 protein consists of 2 PDZ domains (PDZ1 and PDZ2) and an ERM domain. It has 52% sequnce homology with NHERF1.

Source: Li & Naren (2010)

Like NHERF1, NHERF2 can also form homodimers and heterodimers

1.7.3 NHE-RF3 (CAP70):

Unlike NHERF1 and NHERF2, expression of CAP70 is not broad but is limited to hepatic, renal, and intestinal tissues. It is however also localized to the apical membrane of epithelial cells (Li & Naren, 2010). Structurally, CAP70 consists of 4 PDZ domains named PDZ1, PDZ2, PDZ3, and PDZ4 (figure 7).

Figure 8: Diagrammatic illustration of the CAP70 scaffolding protein. The CAP70 protein consists of 4 PDZ domains (PDZ1, PDZ2, PDZ3, and PDZ4).

Source: Li & Naren (2010)

1.7.4 PDZK2 (IKEPP):

Expression of PDZK2 is not broad but is limited to the kidney and intestine. Localization of PDZK2 is on the apical membrane of epithelial cells. Structurally, PDZK2 consists of 4 PDZ domains named PDZ1, PDZ2, PDZ3, and PDZ4.

Figure 9: Diagrammatic illustration of the PDZK2 scaffolding protein. The PDZK2 protein consists of 4 PDZ domains (PDZ1, PDZ2, PDZ3, and PDZ4).

Source: Li & Naren (2010)

1.7.5 Shank2 or CorttBP1:

Shank2 is a multimodular adaptor isoform that plays an important role in the localization of CFTR at the apical membrane (Sheng & Kim, 2000). Localization of Shank 2 scaffolding protein is in the apical membrane and its expression is not broad but is limited to epithelial cells of the liver, intestine, kidney and pancreas and in the brain (Li & Naren, 2010). Structurally, shank2 consists of a SH3 domain, a single PDZ domain, a long region that is rich in proline, and a sterile alpha motif (SAM) domain (Kim et al, 2004; Lee et al, 2007).

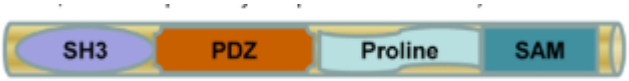


Figure 10: Diagrammatic illustration of the Shank2 scaffolding protein. The Shank2 protein consists of an SH3 domain, a PDZ domain, a region rich in proline, and a sterile alpha motif (SAM) domain. Source: Li & Naren (2010)

Kim et al (2004) have demonstrated that Shank2 interacts with CFTR through its PDZ domain. Studies by Lee et al (2007) demonstrate that over-expression of Shank2 causes the cAMP-mediated phosphorylation and activation of CFTR to be reduced. Plasmon resonance assays and successive patch clamping were used to demonstrate that the interactions between Shank2 and CFTR and NHERF1 and CFTR are competitive and that both scaffolding proteins not only compete for binding at the same site but also regulate the activity of the CFTR channel through their competitive binding. This competition has been proposed to regulate the transport of fluid and ion transport in epithelia (Lee et al, 2007).

1.7.6 CAL (NHE-RF4):

The CFTR-associated ligand (CAL) is a protein with a molecular mass of 50 kD whose expression in human tissues is ubiquitous. Its location is in the perinuclear region of the cell. Expression of CAL is broad and is not limited to any particular tissues (Li & Naren, 2010). Structurally, it is composed of 2 coiled-coil domains and a single PDZ domain.

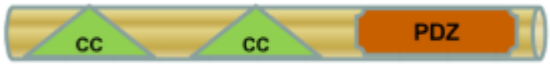


Figure 11: Diagrammatic illustration of the CAL scaffolding protein. The CAL protein consists of 2 coiled coil regions (CC) and a single PDZ domains. Source: Li & Naren (2010)

CAL largely associates with the Golgi apparatus at the trans-golgi network (TGN) but can be observed within the sub-apical or at the apical membrane compartments in certain conditions such as during its over-expression in MDCK1 cells that have been polarized and

cultivated on permeable supports (Guggino, 2003). Fuller & Benos (1992) suggest that the function of CAL is to tie CFTR to the TGN through its coiled-coil domain (Fuller & Benos, 1992). CAL has a PDZ domain which it uses to bind to CFTR and they travel together to the plasma membrane. When it is over-expressed, CAL causes the expression of CFTR on the membrane surface to be reduced. It also reduces the chloride ion currents that are produced following the cAMP mediated activation of CFTR and may control the movement of CFTR from the Golgi apparatus to the plasma membrane (Guggino, 2003).

Based on the interaction between the CC domain and Syntaxin 6 which is a N-ethylmaleimide-sensitive factor attachment receptor (SNARE) protein and a small GTPase known as TC10, it has been suggested that CAL takes part in vesicle trafficking (Neudauer et al, 2001). Attachment of CAL to the C-terminal of CFTR is inhibited by NHERF1 (Cheng et al, 2000). Surface-plasmon resonance analysis, isothermal titration calorimetry, and fluorescence polarization studies have shown that CFTR has a lower affinity for the PDZ domain of CAL than it has for the PDZ domains of NHERF1 and NHERF2 (Cushing et al, 2008). This helps CFTR to avoid premature lysosomal degradation since CAL functions to degrade wild type CFTR by targeting it to lysosomes after endocytosis. Studies also show that CAL reduces the expression of CFTR mutants with the F508 deletion (Cushing et al, 2008).

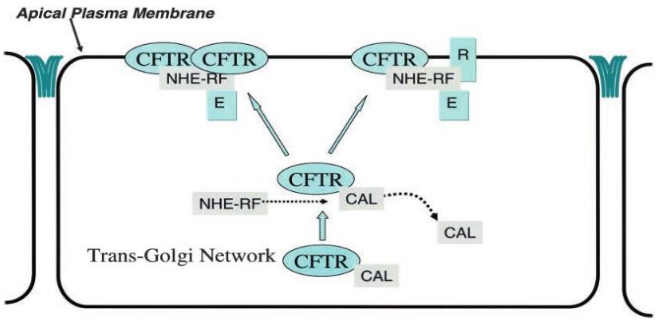


Figure 12: CAL binds to the PDZ domain of CFTR and uses its coiled-coil domain to tie the channel to the trans-Golgi network as well as traffic the CFTR to the plasma membrane. It is also possible that CAL binds to CFTR within the ER. Other scaffolding proteins such as E3KARP and CAP-70 may also bind to CFTR in a competitive manner during the trafficking of the channel protein to the plasma membrane. Once in the plasma membrane, the CFTR can be connected to other proteins in a macromolecular complex (R) or to itself to form dimers. These linkages are mediated by PDZ-binding partners within the plasma membrane. Linkage of CFTR to the cytoskeleton through Ezrin (E) to can also occur and is mediated by NHE-RF. Source: Guggino (2003).

1.8 β2 adrenergic and P2YI Receptors in CFTR Trafficking:

Trafficking involves moving the CFTR to the apical membrane. Trafficking of CFTR to the apical membrane is under the control of its PDZ motif. During trafficking, the PDZ motif of the C-terminal region of CFTR can interact with different proteins including the β2 adrenergic receptor and the Purinergic P2YI receptor, among others. The P2YI receptor has a PDZ motif made up of DTSL and which is capable of binding to the PDZ domain of NHERF. The NHERF PDZ domain has a good affinity for binding the P2Y1 receptor (Hall et al, 1998). The P2Y2 PDZ motif is DIRL and binding of this motif to NHERF is suboptimal since it lacks a hydroxyl amino acid at position 2. Hence, the NHERF PDZ domain has a poor affinity for the P2Y2 receptor (Hall et al, 1998). The PDZ domain of β2 adrenergic receptor has the sequence DSLL. Mutation of the terminal leucine (L) and of alanine at positions 2 and 3 of the domain results in inability of NHERF to bind to the β2 receptor (Hall et al, 1998).

Table 4: Proteins that can potentially bind to NHERF and their PDZ domain sequences

Source: **Hall et al (1998)**

Protein	Type	PDZ motif
β2 adrenergic receptor	Receptor	DSLL
Platelet-derived growth factor receptor (PDGFR)	Receptor	DSFL
Purinergic P2YI receptor	Receptor	DTSL
CFTR	Transporter	DTRL
Endomembrane proton pump	Transporter	DTAL
Copper transporter	Transporter	DTAL
Podocalyxin		DTHL
Phospholipase C-β1		DTPL
Kunjin virus-specified protein		DTVL

NHERF1 can bind to the tail of the β2 adrenergic receptor and affect the adrenergic control of Na+/H+ exchange (Hall et al, 1998). The binding involves attachment of the PDZ1 domain of NHERF1 and NHERF2 to the last few amino acids contained in the tail of the β2 adrenergic receptor (Hall et al, 1998). The β2 adrenergic receptor predominates in the epithelia of airways and can colocalize with CFTR at the apical membrane. It can stimulate CFTR when it is activated. The β2 adrenergic receptor can activate CFTR in sweat duct tissues through the cAMP-PKA pathway (Thelin, 2006). This happens when the β2 adrenergic receptor, NHERF, and CTRF complex together. The NHERF- β2 adrenergic receptor-CFTR macromolecular complex induces the activation of adenylyl cyclase through

the Gs protein leading to elevated levels of cAMP which in turn activates PKA. PKA is closely located to the CFTR in the complex and it causes CFTR to be activated by the β2 adrenergic receptor by phosphorylating the R domain of CFTR. phosphorylation of this domain inhibits the formation of the β2 adrenergic receptor-CFTR-NHERF (Naren et al, 2003).

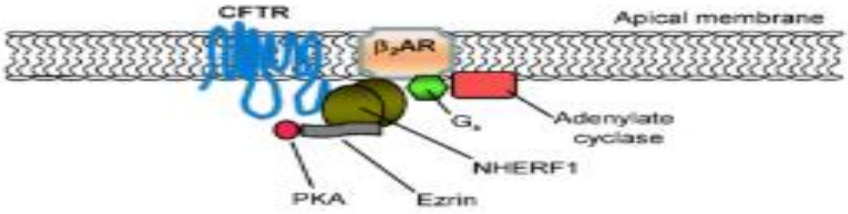


Figure 13: Mediation of CFTR activity by the NHERF- β2 adrenergic receptor-CFTR complex

1.9 Aims and Objectives:

The C-terminus of CFTR plays an important role in the localization of the protein in human airway epithelial cells, coupling it to other regulatory proteins in the cell through its -TRL motif. Understanding of this function requires information at molecular level, which will be pretty achieved by obtaining 3D structure data for the C-terminus .In order to obtain such structural data, the C-terminus has been over expressed in a foreign host, *E.coli*, and different orthologs have been generated. Expression trials should allow a comparison of yields of different orthologs for farther structural studies such as X-ray crystallography, CD, MALDI-TOF and NMR.

Accordingly, the objective of this project is to compare the yield and purity of the killifish CFTR ortholog with human ortholog, which has been previously, expressed using the same system. The comparison will take into account their amino acids composition, solubility, expression yields and predicting studies of their C-terminus CFTR structure, which were carried out using MALDI-TOF and CD. Thus, the Killifish C-terminus has been cloned, expressed, purified, concentrated and characterized by applying the same conditions as used for the human ortholog.

2. Materials and Methods:

2.1 Materials:

2.1.1 Materials used for sub-cloning the C-terminus of killifish CFTR:

The materials used for the sub-cloning were included a Polymerase chain reaction (PCR) setup which contained: CFTR Killifish gene sample, provided by Dr. Ina Urbatsch (Texas Tech University Health Sciences Center, Lubbock, Texas), forward and reversed primers, 5X Phusion HF Buffer, dNTPs, DMSO, Phusion DNA polymerase enzyme, MQ – water .A PCR purification kit was required for PCR purification. Agarose Gel electrophoresis was carried out with the use of Agarose powder, IX TAE buffer (20 mM acetic acid, 10 mM EDTA, 10 mg/ml Ethidium Bromide solution) and DNA Ladders (1KB, 100bp).

For plasmid preparation, pet 28(a), cloning and expression vector (Novagen), was amplified and transformed into Top 10 competent *E.coli* cells. In the transformation step, growth media included SOC media (1 L): 20g Bacto Tryptone, 5g Bacto Yeast Extract, 2ml of 5M NaCl, 2.5ml of 1M KCl, 10ml of 1M $MgCl_2$, 10ml of 1M $MgSO_4$, 20ml of 1M glucose) and LB broth (1L) 10g Enzymatic Digest of Casein, 5g Yeast Extract, 10g NaCl) were required and 50 mg Kanamycin. QIA prep Spin Mini prep Kit was provided as well.

For restriction digestion, the materials needed were NdeI enzyme, EcorI-HF enzyme, Plasmid DNA, MQ water and CIP enzyme (For Dephoshorelation) .For purification purposes Qiagen PCR purification kit was used. In Plasmid extraction, (500 ng-1μg) of DNA was loaded into 1% agarose gel. QIA quick gel extraction kit was essential for gel cleaning. In Insert digestion, the following materials were used: NdeI enzyme, EcorI-HF enzyme, and 10 X Buffer 4, DNA and MQ water. For insert purification, Qiagen PCR purification kit was used.

Materials included in the ligation reaction were: 50 ng of Plasmid (5.369 KD), 8 ng of Insert (144bp), 10x Ligation buffer, and MQ water. For Bacterial Transformation, Top 10 competent *E.coli* cells, SOC media, LB broth, LB Agar, and Kanamycin was needed. In ligated plasmid purification, QIA prep Spin Mini-prep Kit was needed.

2.1.2 Materials used for the C-terminus gene Expression

Materials used for the C-terminus gene Expression included: *E.coli* BL21 (DE3), Ligated plasmid, 250 ml LB broth (small scale expression), 4 L of LB broth media (large scale expression), 1 mM IPTG for 1 L of LB, Sonication buffer (500 mM NaCl, 50 mM HEPES PH 7.0, 0.01 % sodium azide (NaN3), 1 mM DTT, 0.1 % triton X-100, 0.1 mg/ml

lysozyme, 20 mg/ml PMSF), 1 mg/ml Protamine sulfate, 15 % Tris glycine Gel and Instant blue stain, and SDS-load dye (700 µl blue dye, 100 µl DTT stock, 200 µl SDS stock).

2.1.3 Materials used for the C-terminus Purification

Materials used for the C-terminus Purification consisted of: Talon resin -Wash Buffer (500 mM NaCl, 50 mM HEPES pH 7.0, 0.01 % sodium azide (NaN3), 1 mM DTT, 50 mM imidazole), Elution Buffer (500 mM NaCl, 50 mM HEPES pH 7.0, 0.01 % sodium azide (NaN3), 1 mM DTT, 250 imidazole), Dialysis Buffer 100 mM NaCl, 50 mM Tris pH 7.9, 0.01 % sodium azide (NaN3), 1 mM DTT), 70 units of Thrombin (1.3 mg), Thrombin Buffer (50 mM Tris-HCL pH 7.7, 300 mM Potassium acetate, 7 mM Mg acetate, 10 % glycerol), Additives (0.01 % sodium azide, 10 mM DTT, 2mM EDTA), 15 % Tris glycine gel and Instant blue stain and SDS-load dye.

2.1.4 Materials used for the C-terminus peptide concentration

In concentration step, Materials used were: 50 KDa cut-off filter, 3 KDa cut-off filter, 15 % Tris glycine gel, BSA and Instant blue dye and SDS-load dye.

2.1.5 Materials used for the C-terminus mass determination

Materials used for the C-terminus mass determination for MALDI-TOF-MS analysis were 50 mM Tris Exchange Buffer pH 7.5, 0.02 mg/ml of the C-terminus of CFTR and 3 KDa cut-off filter.

2.1.6 Materials used for Circular Dichroism

The following materials were used for Circular dichroism to carry secondary structure analysis.1 mg/ml of the C-terminus CFTR, 10 mM Phosphate buffer pH 7.5 and 3 KDa cut-off filter.

2.2 Methods:

Standard procedures were used for cloning purposes. Primers design protocol was obtained from the General Concepts for PCR Primer Design (Dieffenbach, et al, 1993, p.48) while PCR reaction protocol, Phusion DNA Polymerase, was adapted from Analytical Biochemistry (Chester & Marshak 1993, p.14), *Bio-Chemica* (Frey, et al.1995, p.24) and (Breslauer et al.1986). Other protocols including extraction, digestion and ligation were followed the protocols on the *New England BioLabs* website (www.neb.com). PCR purification, plasmid Mini prep and gel extraction were done according to manufacturer's instructions.

2.2.1 Sub-cloning of the C-terminal of Killifish CFTR into pET28a

2.2.1.1 Designing primers

In the sub-cloning, two primers were designed to amplify the C-terminus of killifish CFTR into pET 28a plasmid. The desired length of the C-terminus was 47 amino acids.

N-terminal C-tail

QAISPADRLHLFPTPHRLNSIKRPQPQTTKISSLPEEAEDEIQDTRL

The N-terminal end of the C-terminus of Killifish CFTR was ligated into pET28 at 6-histidine (His) and thrombin site. Therefore, Ndel (CATATG) restriction enzyme, which is near to thrombin site, was chosen. The second restriction enzyme is EcoRI-HF (GAATTC) .It is preferred because it is away from pET28a C-terminus His tag which may interfere with further structural studies. Then, forward Primer was modified to fit a start cordon and reverse primer was modified to fit a stop codon .For both of them, mismatched bases were calculated. Finally, primers sequence was obtained as well as their melting temperature by using primer temperate, online software, (http://depts.washington.edu/bakerpg/primertemp/primertemp.html accessed on 29/11/2011). The sequence for forward primer is **GAGATGTCTCACCATATGCAAGCAATTTCTCC** and the sequence for reverse primer is **GATGATGGAATTCTTAAAGTCTGGTATCCTGAATTTC.**

2.2.1.2 PCR reactions setup and insert amplification

PCR reactions setup was used for insert amplification. It was done with reference to information in the following table.

Table 5: PCR reactions set-up

Source *(New England Biolabs, 2012)*

Component	**50µl Reaction**
MQ water	To 50 µl
5X Phusion HF buffer	10 µl
10 mM dNTPs	1 µl
10 µM Primer mix	2 µl
Template DNA	100 ng -250 ng
%100 DMSO	1.5 µl
Phusion DNA Polymerase	0.5 µl

Reagents were mixed and transferred to PCR tube for thermocycling.

- Thermocycling conditions were programmed as follows:

Table 6: Thermocycling conditions

Source (*New England Biolabs*, 2012)

Initial denaturation (1cycle)	98°C	30 seconds
Annealing (25-35 cycles)	98°C	5-10 seconds
	45-72°C	10-30 seconds
	72°C	15-30 seconds
Final extension	72°C	5-10 minutes
Hold	4°C	

PCR product was purified using Qiagen PCR purification kit .The size of the PCR product was measured by using 3% agarose gel electrophoresis in which the size of product was compared to a DNA size marker (100 bp.) .The DNA concentration was measured using a Nano Drop spectrophotometer.

2.2.1.3 Plasmid preparation

The preparation steps of the plasmid (pET28a) involved: amplification, transformation and mini-preparation procedures. Therefore, (2-5 µl) from pET 28a were transformed into 50 µl Top 10 competent *E.coli* cells and incubated for 30 minutes on ice. The vile containing the mixture was then heat shocked in 37 °C water bath for 30 seconds and placed again on ice .The SOC media afterwards was added to the vial at a volume of 250 µl. It was then incubated at 37°C, and shaken at 225 rpm for one hour. 20-200 µl, from the transformation vile, were spread on the top of LB agar plate with Kanamycin. The plate was inverted, and incubated at 37 °C overnight. Individual Colonies were then selected and inoculated in LB broth media with Kanamycin for a small-scale culture. These cultures were incubated at 37°C and shaken at 225 rpm overnight. The next morning, the growing cells were centrifuged and pelleted. Cells then were subjected to the Qiagen mini-prep protocol. The size of the plasmid was checked by 1% agarose gel electrophoresis and the concentration was measured using the Nano Drop spectrophotometer.

2.2.1.4 Double Digestion conditions

Several reaction conditions were checked before the double digestion with restriction enzymes. The conditions included were digestion temperature and time, buffer type, heat inactivation temperature and time and whether supplements were required or not. All conditions were checked using Double Digest Finder tool on (*New England Biolabs*) website.

2.2.1.5 Plasmid digestion

For Plasmid digestion, pET28a was digested first with EcoRI-HF enzyme for 30 minutes at 37 °C and then heat inactivated at 65 °C for 20 minutes. A Vile containing the plasmid mixture then placed on ice to cool down for two minutes.

Table 7: Plasmid digestion with EcorI-HF enzyme

MQ water	To 25 µl
Buffer 4 (X10)	2.5 µl
Plasmid	500 ng - 4 µg
EcoRI-HF enzyme	1 µl

Before proceeding to the second digestion step, 2 µl from the first digestion were loaded into 1% agarose gel to check linearity. If the plasmid was linearized, 1 µl of NdeI enzyme would have been added to the mixture which incubated at 37 °C for 1 hour. Digestion mixture then was heat inactivated at 65 °C for 20 minutes. To prevent plasmid religation, CIP was recommended (for dephosphorylation). Per 1 µg of vector, 0.5 unit of CIP was added to the mixture .The mixture then was incubated at 37 °C for 1 hour. After that, the plasmid was purified using Qiagen PCR purification kit.

2.2.1.6 Insert digestion

The digestion protocol for insert digestion is similar to plasmid digestion one except for digestion time. Since the insert was only 144 base pair, increasing incubation time may lead to un specific digestion. Dephosphorylating step is not required after insert digestion.

2.2.1.7 Plasmid extraction

In the Plasmid extraction step, the purified vector then was loaded into 1% agarose gel and gel electrophoresed for 30 minutes at 100 V. The gel afterwards was visualized under UV for a few seconds and the DNA fragment was excised and purified using QIA quick gel Extraction kit. Extraction step is not required for the digested insert.

2.2.1.8 Ligation reaction protocol

Calculating the concentration of DNA needed for Ligation reaction is a critical step for successful cloning. Different molar ratios were used for sticky end ligation. 3:1 or 6:1 insert/vector molar ratios were recommended since the size of the insert was smaller than the size of the plasmid. Insert concentration then was calculated using the following formula:

Insert mass in ng = (6 or 3) X (insert length in bp/vector length in bp) X vector mass in ng.

The standard concentration of 3 KD vector is 50 ng and 1KB insert is 50 ng (*New England Biolabs).* Using this information, the volume of the plasmid and the insert can be calculated.

- For 10µl reaction the 6:1 insert/vector molar ratio was:

Table 8: Components of ligation reaction

Vector (5 Kb)	50 ng
Insert (144 bp)	7-8 ng
10X ligation buffer	1 µl
MQ water	Up to 10µl
Ligase enzyme	0.5 µl

The reaction then was incubated at RT for 1 hour. (1-5) µl of ligation mixture then was transformed into Top 10 competent *E.coli* cells. The transformation, culturing, mini-preparation methods were described in (2.2.1.3)**.** The success of ligation reaction was checked afterwards by digesting the ligation mixture, using double digestion protocol, and loading it into a 3% gel to see if the insert would be released again at the size of 144 bp. Another check was done using DNA sequence reactions to sequence DNA into nucleotides. This sequence then was translated into the amino acid sequence as a final check.

2.2.2 Protein expression of the C-terminus of Killifish CFTR

For protein expression, ligated vector was transformed into *E.coli* BL21 (DE3) competent cells following the Top10 transformation protocol (described in 2.2.1.3).

For Small-Scale Expression, 1-5 colonies were picked from overnight plate and inoculated into 250 ml LB Broth media with kanamycin. The culture then incubated for overnight at 37 °C and shaken at 225 RPM.

For large-scale expression, 4 liters of LB media were prepared and autoclaved .To calculate the volume to transform from small-scale culture into the large one, the OD_{600} of overnight culture was measured. The OD_{600} of overnight culture was more than 2 and therefore cells dilution was required. 100 µl of the cells were added into 900 µl of LB broth. Since the starting OD_{600} for each liter should be 0.05, the following formula was used to calculate the volume of overnight cells to transfer to each liter:

0.05 / OD_{600} of overnight culture * 1 liter = volume of overnight culture to add to 1 liter.

The calculated volume then added to the LB media with kanamycin. The cultures afterwards were incubated at 37 °C, and shaken at 200 RPM. The OD_{600} was checked every hour and samples can be taken to check the growing level of cultures. When the OD_{600} reached (0.6-0.8), cultures were placed on ice for 5 minutes. After that, each culture was induced with 1mM IPTG and grown at 25 °C and harvested after four hours. Samples also can be taken every hour to check expression levels. Induced cultures were then centrifuged at 6000 RPM for 20 minutes in 4 °C Beckman JLA8.1000. After centrifugation, supernatants

were poured off and pellets were re-suspended into 40 ml of Sonication buffer and then were sonicated three times for 20 seconds using Vibra-Cell Auto tune series High intensity Ultrasonic processor (90% amplitude). The sonicated samples then were transferred into centrifuged tubes, and centrifuged at 17000 RPM using a SS34 Rotor, Beckman, for 20 minutes at 4 °C. After the centrifugation, protamine sulfate (Sigma) was added at concentration of 1 mg/ml to bind and precipitate DNA. Samples then were re-centrifuged as previously .The pH was checked and adjusted to 7. Supernatants then were added to Talon resin in a falcon tube for binding.

2.2.3 Protein Purification using metal affinity chromatography

For protein purification purposes, talon immobilized metal affinity chromatography (*CLONTECH*) resin was equilibrated with the wash buffer (1.2 ml talon: 1.2 ml wash buffer) before mixing it with supernatant. Talon was then mixed with supernatant and was incubated at 4 °C for 1.5 hour with 1 RPM shaking to allow binding. Purification steps were all carried out at 4 °C. The above mixture was loaded into 20 ml gravity flow column *(BioRad*) in which unbound materials were passed through the column and collected. Then the resin was washed with wash buffer (10 times the volume of resin used) and eluted with 20 ml elution buffer. The elutions then were collected, which had the purified protein. Before loading the purified protein into a 5000 MWCO dialysis membrane *(BDH chemicals Ltd),* thrombin cleavage was required.

Therefore, 70 units of thrombin and 2 mM of $CaCl_2$ were added to the purified protein. The sample then was loaded into a dialysis membrane, which was sealed with two plastic clips on both ends, and placed in 1 L dialysis buffer with small stirrer bar at a very low speed for overnight incubation at 4 °C. The next day, the purified protein was collected from the dialysis membrane into a falcon tube. New resin was equilibrated with dialysis buffer and added to the purified protein. Binding, washing and elution steps were repeated following the same order and volumes that were described above. Except for the washing step, dialysis buffer was used instead of wash buffer .The cleaved protein then was collected in the unbound materials. 0.01 % of Sodium Azide, 1 mM DTT and 2 mm EDTA were added to promote stability of purified protein.

SDS-PAGE analysis then was used to check purity.3 µl of unbound materials, wash and elutions were loaded into 15% Tris glycine gels along with a protein marker (7-170 KD) to measure the size of purified protein.

2.2.4 Protein concentration

The purified protein was then concentrated to the volume of 1 ml using Centricon Centrifugal Filter units, 50 and 3 KD micro-concentrators, (*Millipore)*. This was done by 5000-RPM centrifugation at 4 °C .To remove DTT and EDTA, the purified protein was washed with dialysis buffer three folds for three times during concentration process. The concentrated protein then stored at -80 °C.

2.2.4.1 Determination of protein yield

The yield of protein was estimated by loading known concentrations of BSA (2µg, 5µg, 10µg) along with the concentrated protein into 15% Tris glycine gel. The gel then scanned using Fujifilm gel documentation camera. The scanned image afterwards was brought to image software (imageJ) to calculate bands intensity. The BSA bands intensities along with known concentrations of BSA were used to plot the standard curve in Excel. The straight-line equation was generated. Using this equation, the concentration of protein was estimated. Knowing the amount of the protein obtained from 2 ml was helped in estimating the protein yield per 1 liter of cell cultures.

2.2.5 Characterization of protein

2.2.5.1 MALDI-TOF-MS analysis

To characterize the C-terminus peptide, MALDI-TOF-MS analysis was used for mass determination. A final concentration of 0.02 mg/ml of purified peptide was needed to run the analysis. For this purpose, the sample was buffer exchanged with 50 mM Tris (PH 7.5) using 3KD Microcon centrifugal filter. The peptide then was sent to the bimolecular analysis unit (University of Manchester FLS) for accurate mass determination using MALDI-TOF. Peptides were dissolved in 50% acetonitrile, with TFA and mixed with an equal volume of a solution containing a α-cyano-4-hydroxycinnamic acid. 1-2 µl of the final solution was applied to the peptide sample which was allowed to dry prior to insertion into the high vacuum of the mass spectrometer. MALDI is based on the bombardment of the matrix crystals with a laser light to bring about sample ionization. Tryptic peptides normally form positively charged ions which fed into the time-of flight analyzer which separate ions according to their mass to charge ratios by measuring their time to flight (Pre-Masters Summer School Manual, University of Manchester, FLS, 2011).

2.2.5.2 Circular Dichroism

Circular Dichroism was used to determine the secondary structure and folding properties of peptide. It measures differences in the absorption of left-handed polarized light

against right-handed polarized light which arise due to structural asymmetry. Secondary structure can be determined by CD spectroscopy in the "far-UV" spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment .To run CD, 1 mg/ml of the C-terminus was buffered exchanged with 10 mM phosphate buffer, pH 7.5, at a final volume of 200 µl. Peptide then was sent to the bimolecular analysis unit for CD analysis (University of Manchester FLS) in which the analysis was carried out with a spectroplarimeter (JASCO, J-810). Cells of 0.5 mm path were used. Additives, buffers and stabilizing compounds were avoided.

3. Results:

3.1 Sub-cloning of the C-terminus of killifish CFTR:

3.1.1 PCR implication of the target gene:

The C-terminus of Killifish CFTR was successfully amplified using Phusion High-Fidelity DNA Polymerase enzyme for PCR reaction. PCR reaction was explained in the material and method section (2.2.1.2**)**. The amount of PCR fragment was measured before purification using a NanoDrop Spectrometery and it was about 300 ng. PCR product then purified using PCR purification kit as manufacture recommendation. The amount after purification was measured again using the NanoDrop and it was around 168 ng. After that, the 2 µl of PCR product before and after purification were loaded into 3% agarose gel electrophoresis to check the size of PCR product in bp against 100 bp DNA Ladder as in figure (14):

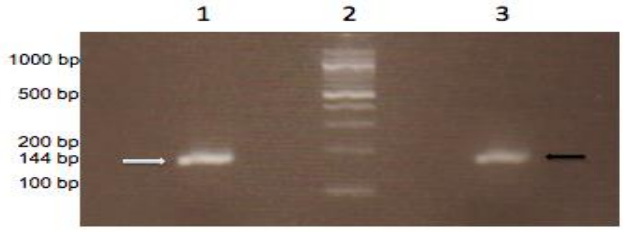


Figure 14: PCR products of the C-terminus of killifish CFTR gene before and after PCR purification. Both products were DNA electrophoresed in 3% Agarose gel. Bands in the first lane and third lane are referring to PCR product before and after PCR purification respectively. An estimation of PCR products sizes was determined by using DNA 100 bp marker in the second lane.

3.1.2 Digestion of pET28 a with EcoRI-HF enzyme

pET28 was prepared as described in methods section (2.2.1.3). The amount of mini prepped pET28a then was measured using the NanoDrop where it was about 133 ng. The digestion then started using EcoRI-HF enzyme for half an hour at 37 °C. After that 2 µl was loaded into 1% agarose gel to check plasmid linearity as in figure (15):

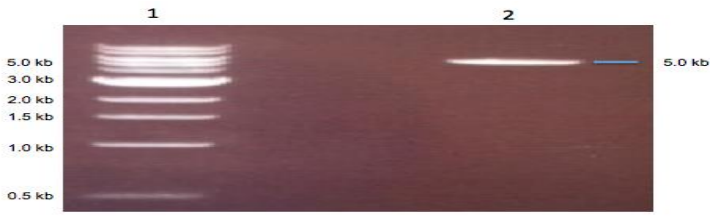


Figure 15: Plasmid digestion with EcoRI-HF enzyme. PET28a in the second lane was digested with EcoRI-HF enzyme and was run on 1% agarose gel to check linearity. 1 Kb DNA Ladder, in the first lane, was used as a molecular weight standard.

3.1.3 Digestion of pET28 a with NDeI enzyme:

NdeI enzyme was added as described in method section (2.2.1.5).

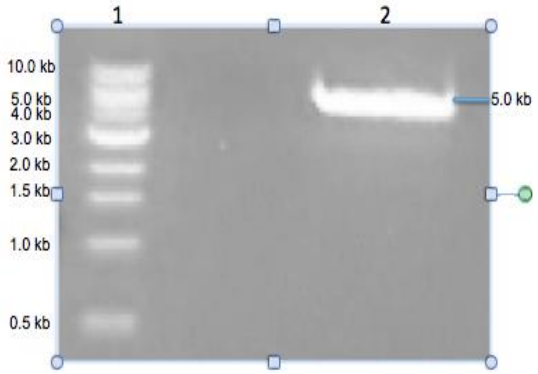


Figure 16: pET28 a after the double digestion with EcoRI-HF and NdeI enzymes in the second lane was run against the 1kb DNA Ladder in the first lane.

3.1.4 Confirmation of successful cloning via sequencing reaction:

To confirm successful cloning via sequencing reaction, 300-450 ng of minipreped ligation reaction was sent to DNA Sequencing unit (University of Manchester FLS) to confirm if the cloning was successful and if it is in the right reading frame. For sequencing reaction, the sample was subjected to different thermocycling condition for amplification. The sequencing reaction was precipitated with two volumes of 95% ethanol and 1/10 volume of 3 M sodium acetate, pH 4.6. The mixture then was incubated on ice for 10 minutes. The precipitated sequencing product was pelleted by centrifugation. The pellet then was washed with 70% ethanol, dried and suspended in 2 µl formamide loading dye, Applied Biosystems, and 1 µl was run on an ABI automated sequencer, Applied Biosystems, (Pre-Masters Summer School Manual, University of Manchester FLS, 2011).

Nucleotides sequences were then translated to amino acids sequences using an online nucleotides translator tool such as ExPASy as in Figure (17):

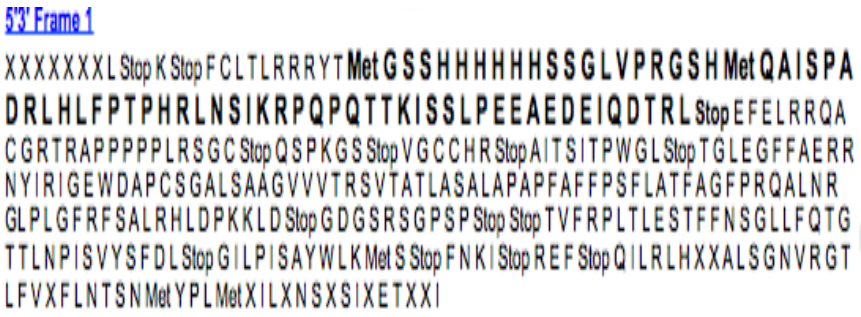


Figure 17: The translated amino acid sequence from the ExPASy server. This confirms that the C-terminus was ligated successfully and attached to 6-His and thrombin sequence on its N-terminal side.

3.1.5 Confirmation of successful cloning via double digestion of the pET28a encoding the C-terminus of killifish CFTR

The second confirmation of successful cloning was via double digestion of the pET28 encoding the target gene. 4.2 µg of ligated vector were digested using the same enzymes, which were used earlier, to see if the ligated vector would release the PCR product. 25 µl of digested mixture then loaded into 3% agarose gel and visualized under UV light as in Figure (18):

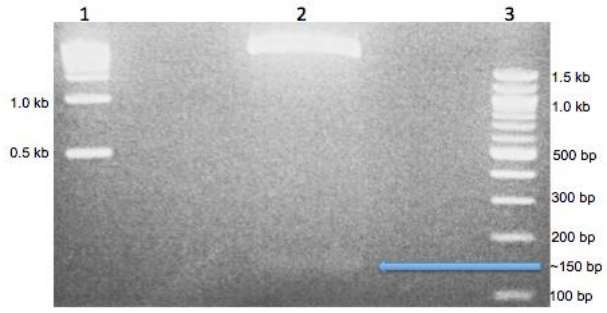


Figure 18: Conformation of successful cloning via double digestion.

The pET28 a encoding the C-terminus of the killifish CFTR in the second lane was doubled digested with EcoRI-HF and NdeI enzymes to release the insert from pET28a in order to confirm successful cloning. 1kb and 100 bp DNA markers were run in the first and third lane respectively to indicate insert and plasmid sizes.

From figure 18 it can be confirmed that the PCR product was released since the size of the released band was similar to PCR product size.

3.2 Purification of the target peptide

The purification process of the target peptide was done as explained in material and methods section (2.2.3). Purification then was checked by SDS-PAGE by loading 3 µl of unbound materials, washes and elutions from the first and the second day purification into a 15% Tricine Glycine gel. From Figure (19) it is clear that the C-terminus peptide (lane 6) migrated as a major band with a molecular weight higher than its theoretical molecular weight (5788 D). No other bands were observed in lane 6.

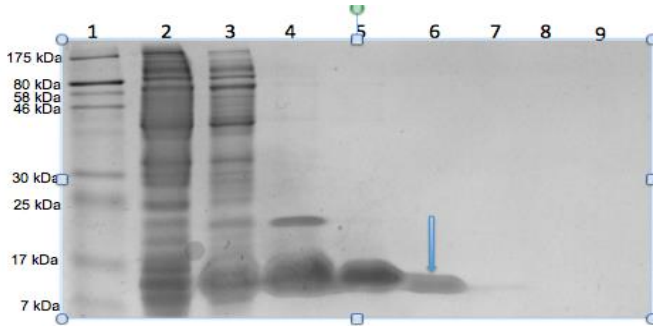


Figure 19: The Purification check of the C-terminus of Killifish CFTR. Purification was checked using 15% SDS-PAGE gel. The Prestained Protein Marker in the first lane was used for molecular weight determinations. The lanes from (2-5) refer to IMAC unbound materials; washes and elutions before thrombin cleavage while lanes from (6-9) refer to IMAC unbound materials; washes and elutions after thrombin cleavage. The arrow in lane 6 refers to the purified c-terminus peptide.

3.3 The yield of purified peptide:

The yield of the C-terminus of Killifish CFTR was measured by comparing the concentration of purified peptide with known concentrations of BSA. Figure (20) shows how the purified protein (lane 2 and 3) was concentrated (lane 6).

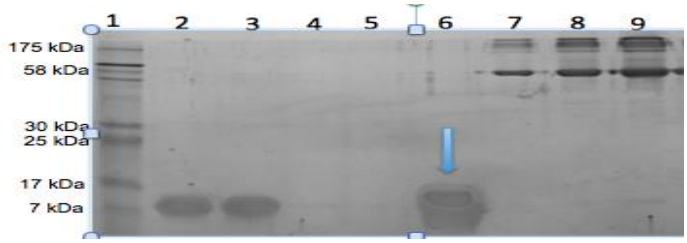


Figure 20: SDS-PAGE gel was used to check the concentration of the purified peptide. Prestaind protein marker (*BioLabs*) in the first lane was used to indicate molecular weights. The purified protein is in the second lane; the filtrate of 50 kDa is in the third lane, the filtrate of the 3 kDa is in the fourth lane, the filtrate of the 3 kDa buffered exchanged with dialysis buffer is the fifth lane, the concentrated protein is in the sixth lane and 2,5,10 micrograms of BSA are in the lanes (7-9) respectively.

After scanning the gel using Fujifilm gel documentation camera, the scanned image was brought to the imagej software to calculate bands intensities. Excel was used to generate a straight-line equation. From the equation the concentration, amount and yield of the concentrated protein were calculated. The yield of the human protein was 0.5 mg / l of cell culture, previously studied, while the yield of killifish one was 2.86 mg / l of cell culture.

3.4 Mass determination using MALDI-TOF analysis:

In Mass determination using MALDI-TOF analysis, peptide mass was determined using the analysis (figure 21). 4 peaks were produced. The first peak was the matrix (1351 D). The second peak was unidentified (1841 D), the third peak was the doubly charged ion (2897 D) and the fourth peak represents the single charged ion peptide (5798 D). The difference between the measured mass (5798 D) and theoretical mass (5788 D) is within the accuracy range (5 D -10 D for such peptide size) of linear mode of MALDI-TOF.

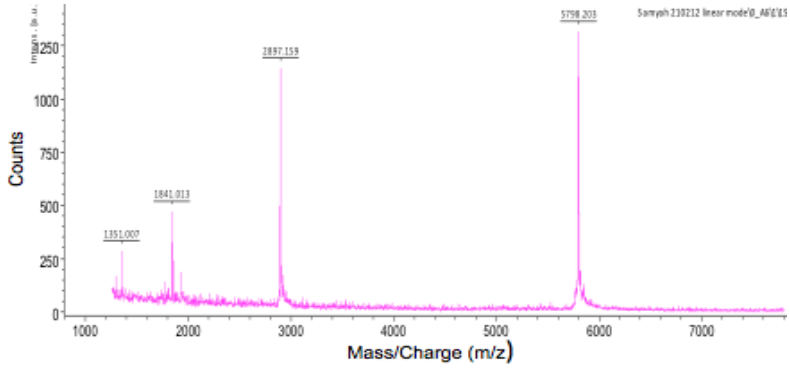


Figure 21: Linear mode of MALDI-TOF analysis. Peaks from left to right are: the matrix (1351 D), unidentified peak (1841), double charge ion (2897 D) and single charge ion peak (5798).

3.5 Prediction of peptide Secondary structure using Circular Dichroism

In secondary structure determination using Circular Dichroism, 1 mg of the purified protein was buffered exchanged with 10 mM phosphate buffer and 150 µl of the above sample was used for the analysis. The data obtained from the CD analysis, in Delta Epsilons, was compared with a referenced dataset in the prediction server DichroWeb. In figure 22, the CDSSTR algorithm with referenced dataset 4 shows a good fitness of experimental data with the reference one while the secondary structure components were estimated in table (10) as follow: 12% Helix, 32% beta and 55% unordered/turns.

Table 9: Secondary structure components of the C-terminus of killifish CFTR

Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.04	0.08	0.21	0.11	0.25	0.30	0.99

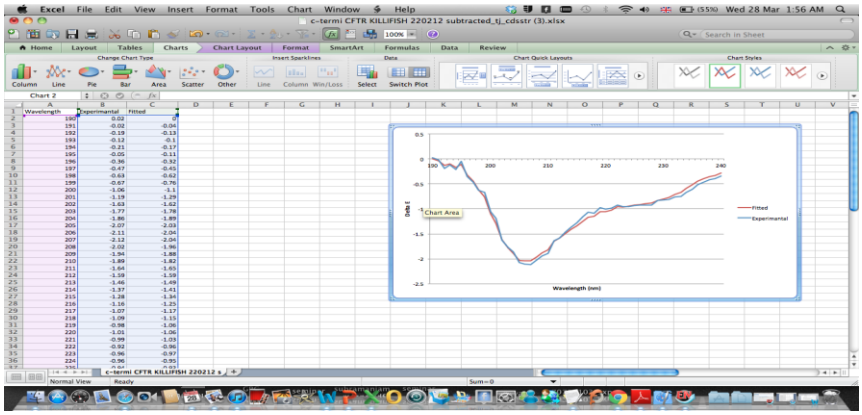


Figure 22: A reasonable agreement of Experimental data, plotted in blue, **with the calculated spectrum**, plotted in red, derived from the CDSSTR algorithm to calculate secondary structure components.

4. Discussions and Conclusions:

The main objective of studying the C-terminus of killifish CFTR was to compare its yield with the C-terminus of human ortholog in order to find which ortholog among them would be more suitable for further structural and functional studies. For this reason, protocols and experimental conditions that were used previously for expressing, purifying and characterizing human ortholog were kept the same. The cloning protocol required a complete change in which a new plasmid was introduced. The plasmid used for cloning the C-terminus of human ortholog, pET-24a, gave negative results in all cloning trials. Investigations were needed to find the reasons behind unsuccessful cloning. It was found that pET-24a was modified previously and by mistake another SphI site was introduced before the T7 promoter. Therefore in every digestion step with SphI-HF, enzyme was used to digest pET-24a, the T7 promoter was cut and therefore there was no transcription. A partial digestion protocol was suggested in order to solve the pET24a design problem. This protocol involved doing a complete digestion with XhoI enzyme first, the second enzyme that was used to digest pET-24a, and then partial digestion with SphI-HF for different incubation periods. Unfortunately, it did not succeed. These trials consumed money and resources giving negative results every time. A decision then was made for using another plasmid and design new primers. The cloning and expression vector, pET28a, was chosen to clone and express the C-terminus of killifish. It has on its N-terminal end a 6- His tag which is commonly used in attaching heterologous expressed proteins and for purification and immunological detection (R.A. Melnyk et al, 2003, p.112). The thrombin sequence was required for peptide cleavage.

To improve accuracy of cloning Phusion DNA Polymerase was used instead of Failsafe enzyme, used for amplifying the C-terminus of human ortholog, since its error rate is 50 fold lower than *Taq* polymerase (Frey & Suppmann, 1995, p.90) while Fail Safe, used in pET24a PCR setup reaction, error rate is 3 fold lower than *Taq* polymerase (Cline, J. *et al*., 1996, p.38).

For proper plasmid digestion, each enzyme was introduced separately and checked on 1% agarose gel before it was heat inactivated. The de-phosphorylation step was introduced to prevent plasmid religation. This will increase cloning efficiency (Bruce Alan white, 1997, p.67).

For ligation 1:3 or1: 6 vectors: insert molar ratio for sticky end ligation can be used to increase success rate. Optimizing each step in sub cloning protocol led to successful cloning.
The C-terminus was heterologousely expressed using BL21 competent *E.coli* cells. BL21 cells were used since they have the T7 promoter expression system, capable of producing

more protein than any other bacterial expression system (Guide to Gene Expression in BL21, biomol). IPTG standard induction protocol was used and cells were induced at log phase (O.D = 0.6) to optimize protein expression.

The purification protocol was slightly modified by adding double the amount of resin and increasing binding time to 1.5 hour. This suggestion was made after it noticed from a previous purification trial that considerable amount of peptide was lost in the wash step. This suggestion helped to recover some of the peptide. The migration of purified peptide on SDS-PAGE was higher (approximately 10 KD). This observation does not correlate with theoretical molecular weight of the C-terminus peptide (5788 D). According to (Rath, Glibowicka, Nadeau, Chen, & Deber, 2009) this observation was termed " gel shifting " and it appears to be common for membrane proteins and has not yet to be conclusively explained.

The protein concentration protocol used for the human ortholog was also used for killifish except that the C-terminus of killifish was buffered exchanged with dialysis buffer three times to remove the effect of DTT, a reducing agent. Removing the 6-His tag, DTT, imidazole was a necessary step before characterization of the peptide since these substances absorb UV light used in CD analysis. (http://www.ap-lab.com/circular_dichroism.htm: accessed at 16/3/2012).

In characterization studies, MALDI-TOF analysis was used for peptide mass determination. The C-terminus result was obtained in the linear mode and it is in the accuracy range. The 10 Da difference was expected since linear mode is limited in resolution leading to low accuracy. (http://chemwiki.ucdavis.edu/Analytical_Chemistry/Instrumental_Analysis/Mass_Spectrometry/MALDI-TOF#Modes: accessed at 22/3/2012). Various reasons can affect the accuracy of MALDI-TOF like the presence of internal standards and the selection of matrix materials (Trauger et al, 2002). Obtaining the single charge ion ($M+H^+$) should be the dominant species however, double charge ions, molecular ions at approximately half the m/z value, and matrix peak were expected in the positive ionization mode of MALDI-TOF, mode for peptide and protein (Glen L. Hortin, 2006, p.21). A further investigation of unidentified peak was ignored since MALDI-TOF was done mainly to determine the mass of The C-terminus peptide.

Circular Dichroism Spectroscopy was used to analyze secondary structure and conformation of macromolecules as proteins. The analysis was done in the far UV region from (190 – 260 nm) since this region reports on secondary structure. To run such an analysis, 1 mg of the peptide sample was buffered and exchanged with 10 mM phosphate buffer. Data obtained from Circular Dichroism spectrometer are plotted against reference

dataset from the prediction server .The C-terminal peptide fit quite well with CDSSTR algorithm but not entirely. This is expected since membrane proteins (CFTR) are not well analyzed by the existing reference databases (Lee Whitmore and B. A. Wallace 2004, p88).

In conclusion, the yield obtained from the C-terminus of human and killifish CFTR was not enough to probe protein structure. Therefore, my suggestion is to optimize cloning, expression and purification techniques. According to Raymond Stevens, from The Scripps Research Institute, "working on all aspects of the process is the strategy and the key to success is the order in which you try everything (Baker, 2010).

References:

ABCMdb. Database for Mutations in ABC Proteins. Available at <http://abcmutations.hegelab.org/mutationCompactDetails?id=4195> [Accessed 7th March, 2012).

Baker, M. (2010). Making membrane proteins for structures: a trillion tiny tweaks. [Comment]. *Nat Methods, 7*(6), 429-434. doi: 10.1038/nmeth0610-429

Benharouga M, Sharma M, So J, Haardt M, Drzymala L, Popov M, Schwapach B, Grinstein S, Du K, Lukacs GL. 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. *Biol Chem*.; 278:22079–22089.

Biber J, Gisler SM, Hernando N, Murer H. 2005. Protein/protein interactions (PDZ) in proximal tubules *Membr Biol.* 203:111–118.

Breslauer, et al., 1986. *PNAS*. Washington, DC*:* National Academy of Sciences, 83, 3746-3750.

Brone B, Eggermont J. 2005. PDZ proteins retain and regulate membrane transporters in polarized epithelial cell membranes. *Am J Physiol Cell Physiol*. 288:C20–29.

Cline, J. *et al*., 1996. Nucl. Acids Res. *Oxford Journals*, 24 (18), 3546

Cramb, G., and Ferec, C. (2001). A Combined Analysis of the Cystic Fibrosis Transmembrane Conductance Regulator: Implications for Structure and Disease Models. *Mol. Biol. Evol*. 18(9):1771–1788.

Choi, J. Y., D. Muallem, K. Kiselyov, M.G. Lee, P.J. Thomas, and S. Muallem. 2001.

Aberrant CFTR-dependent HCO3-transport in mutations associated with cystic fibrosis. *Nature* 410:94–97.

CFTR. (2011). Cystic Fibrosis Mutation Database. Available at <http://www.genet.sickkids.on.ca/app > [Accessed 7th March, 2012).

Cheng J, Wang H, Guggino WB. 2004. Modulation of mature cystic fibrosis transmembrane regulator protein by the PDZ domain protein CAL. *J Biol Chem*. 279:1892–1898.

Cheng J, Wang H, Guggino WB. 2005. Regulation of cystic fibrosis transmembrane regulator trafficking and protein expression by a Rho family small GTPase TC10.; *J Biol Chem*. 280:3731–3739

Cunningham, F., and Deber, C.M. (2006). Optimizing synthesis and expression of transmembrane peptides and proteins. *Methods* 41:370–380

Cushing PR, Fellows A, Villone D, Boisguerin P, Madden DR. 2008. The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling. *Biochemistry*. 47:10084–10098.

Dean, M. 2008. ATP-binding Cassette (ABC) Transporter Supergene Family: Genetics and Evolution. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

DieVenbach, C.W., Lowe, M.J., & Dveksler, G.S., 1993. *General Concepts for PCR Primer Design. PCR Methods Appl*, 3, 30–37.

Frey, B., & Suppmann, P., 1995. *Biochemica*. Netherlands: Bentham Science *Publishers* Ltd. 34-35.

Fuller CM, Benos DJ. 1992. CFTR. *Am J Physiol Cell Physiol*, 263:C267–C286

Gentzsch,M., Aleksandrov,A., Aleksandrov,L. & Riordan,J.R. 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

Gentzsch, M., and Riordan, J.R. 2001. Localization of Sequences within the C-terminal Domain of the Cystic Fibrosis Transmembrane Conductance Regulator Which Impact Maturation and Stability. *The Journal of Biological Chemistry*. 276(2):1291–1298.

Gentzsch,M. *et al.* 2004Endocytic trafficking routes of wild type and DeltaF508 cystic fibrosis transmembrane conductance regulator. *Mol. Biol. Cell* 15, 2684-2696

Guggino, W.B. 2003. The Cystic Fibrosis Transmembrane Regulator Forms Macromolecular Complexes with PDZ Domain Scaffold Proteins. *Proc Am Thorac Soc* 1: 28–32, Guide to Gene Expression in BL21, biomol

Hall, R.A., Ostedgaard, L.S., Premont, R.T., Blitzer, J.T., Rahman, N., Welsh, M.J., and Lefkowitz, R.J. 1998. A C-terminal motif found in the b2-adrenergic receptor, P2Y1receptor and cystic fibrosis transmembrane conductance regulator determines binding to the Na^+/H^+ exchanger regulatory factor family of PDZ proteins. *Proc. Natl. Acad. Sci*. 95: 8496–8501.

Hortin, G. L., 2006. *Clinical Chemistry* 52 (7), 1223–1237.

KEGG. (2012a). ABC Transporters. Available at <http://www.genome.jp/dbget-bin/www_bget?map02010> [Accessed 7th March, 2012).

KEGG (2012b). ABC transporters - Reference pathway. Available at <http://www.genome.jp/kegg-bin/show_pathway?map02010> [Accessed 7th March, 2012).

Kerem, B., Buchanan, J. A., Durie, P., Corey, M. L., Levison, H., Rommens, J. M., Buchwald, M., Tsui, L.-C. 1989. DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. *Am. J. Hum. Genet*. 44: 827-834.

Kim JY, Han W, Namkung W, Lee JH, Kim KH, Shin H, Kim E, Lee MG. 2004. Inhibitory regulation of cystic fibrosis transmembrane. *J Biol Chem*. 279:10389–10396.

Kos, V., and Ford, R. C. 2009. The ATP-binding cassette family: a structural perspective. *Cell Mol. Life Sci.* 66, 3111–3126

Lamprecht G, Seidler U. 2006. The emerging role of PDZ adapter proteins for regulation of intestinal ion transport. *Am J Physiol Gastrointest Liver Physiol*, 291:G766–777.

LaRusch, J.A. 2007. A Novel Internal Binding Motif In The CFTR C-Terminus Enhances Ebp50 Multimerization And Facilitates Endocytic Recycling. PhD Dissertation.

Lee J.H., Richter W., Namkung W., Kim K.H., Kim E., Conti M, Lee M.G. 2007. Dynamicregulation of cystic fibrosis transmembrane conductance regulator by competitive interactions of molecular adaptors. *J Biol Chem*. 282:10414–10422.

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

Lukacs, G. L., Segal, G., Kartner, N., Grinstein, S., and Zhang, F. 1997. Constitutive internalization of cystic fibrosis transmembrane conductance regulator occurs via clathrin-dependent endocytosis and is regulated by protein phosphorylation. *Biochem. J.* 328, 353–361

Marks, M. S., Woodruff, L., Ohno, H., and Bonifacino, J. S. 1996. Protein targeting by tyrosine-and di-leucine-based signals: evidence for distinct saturable components. *J. Cell Biol.*135, 341–354

Mickle, J.E., Macek Jr, M., Fulmer-Smentek, S.B., Egan, M.M., Schwiebert, E., Guggino, W.,

Moss, R., and Cutting, G.R. 1998. A mutation in the cystic fibrosis transmembrane conductance regulator gene associated with elevated sweat chloride concentrations in the absence of cystic fibrosis. *Human Molecular Genetics*, 7(4):729–735

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., and Stanton, B. A. 1999. *J. Clin. Invest.* 104, 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., and Stanton, B. A. 2000. *J. Biol. Chem.* 275, 27069–27074

Naren AP, Cobb B, Li C, Roy K, Nelson D, Heda GD, Liao J, Kirk KL, Sorscher EJ, Hanrahan J, *et al.* 2003. A macromolecular complex of β-adrenergic receptor, CFTR, and

ezrin/radixin/moesin-binding phosphoprotein 50 is regulated by PKA. *Proc Natl Acad Sci USA*; 100:342–346.

NCBI. (2012). cystic fibrosis transmembrane conductance regulator [Homo sapiens]. Available <http://www.ncbi.nlm.nih.gov/protein/NP_000483.3> [Accessed 7th March, 2012).

Neudauer C.L., Joberty G., and Macara, I.G. 2001. PIST: a novel PDZ/coiled-coil domain binding partner for the rho-family GTPase TC10.*Biochem Bioph Res Co*. 280:541–547.

OMIM. 2012. Cystic Fibrosis Transmembrane Conductance Regulator; CFTR. Available <http://omim.org/entry/602421> [Accessed 7th March, 2012).

Park, J.G., Aksamit, T.R., Swanson, K.L., Thomas, C.F., and Caples, S.M. 2008. Pulmonary Diseases in *Mayo Clinic Internal Medicine Review*, 8th Ed. Ghosh, A.K. (ed). Mayo Clinic Scientific Press. Pp.907-908

Pre-Masters Summer School Manual, University of Manchester FLS, 2011.

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

Raghuram V, Mak DOD, Foskett JK. 2001. Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction. *Proc Natl Acad Sci USA*; 98:1300–1305.

Reczek D, and Bretscher, A. 1998. The carboxyl-terminal region of EBP50 binds to a site in the amino-terminal domain of ezrin that is masked in the dormant molecule. *J Biol Chem*. 273:18452–18458.

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

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 [published erratum appears in Arch Biochem Biophys 2000 May 1;377(1):213]. *Arch. Biochem. Biophys.* 375, 7-20.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. [Research Support,

Non-U.S. Gov't]. *Proc Natl Acad Sci U S A, 106*(6), 1760-1765. doi: 10.1073/pnas.0813167106

Riordan, J. R. 2008. CFTR function and prospects for therapy. *Annu. Rev. Biochem*. 77, 701–726

Rosenberg, M.F., O'Ryan, L.M., Hughes, G., Zhao, Z., Aleksandrov, L.A., Riordan, J.R., and Ford, R.C. 2011. The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). Three-Dimensional Structure and Localization of a Channel Gate. *The Journal of Biological Chemistry*, 286(49): 42647–42654

Rowntree, R.K., and Harris, A. 2003. The Phenotypic Consequences of CFTR Mutations. *Annals of Human Genetics*, 67,471–485

Saxena,S.K., Kaur,S. & George,C. 2006. Rab4GTPase modulates CFTR function by impairing channel expression at plasma membrane. *Biochem. Biophys. Res. Commun.* 341, 184-191

Sedlack, R.E., Loftus, C.G., Viggiano, T.R., and Poterucha, J.J. (2008). *Mayo Clinic Internal Medicine Review*, 8th Ed. Ghosh, A.K. (ed). Mayo Clinic Scientific Press. Pp.907-908

Sheng M, Kim E. 1996. Ion channel associated proteins. *Curr Opin Neurobiol,* 6:602–608.

Short, D. B., Trotter, K. W., Reczek, D., Kreda, S. M., Bretscher, A., Boucher, R. C., Stutts, M. J., and Milgram, S. L. 1998. *J. Biol. Chem.* 273, 19797–19801

Singer, T.D., Keir, K.R., Hinton, M., Scott, G.R., McKinley, R.S., and Schulte, P.M. 2008. Structure and regulation of the cystic fibrosis transmembrane conductance regulator (CFTR) gene in killifish: A comparative genomics approach. *Comparative Biochemistry and Physiology*, Part D 3:172–185

Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*; 275: 73–77.

Swiatecka-Urban,A. *et al.* 2004. Myosin VI regulates endocytosis of the cystic fibrosis transmembrane conductance regulator. *J Biol. Chem.* 279, 38025-38031.

Thelin, W.R. 2006. The regulation of CFTR by protein-protein interactions. PhD dissertation.

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

Walker,J.E., Saraste,M., Runswick,M.J. & Gay,N.J. 1982. Distantly related sequences in the alpha and beta subunits of ATP synthase, myosin, kinases and other ATPrequiring enzymes and a common nucleotide binding fold. *EMBOJ.*1, 945-951.

Wang, S., Yue, H., Derin, R.B., Guggino, W.B, and Li, M. Regulation of CFTR Activity via Protein-Protein Interactions. S1.2

Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B., and 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

Weinman EJ, Steplock D, Tate K, Hall RA, Spurney RF, Shenolikar S. 1998. Structure–function of recombinant Na/H exchanger regulatory factor (NHE-RF). *J Clin Invest*; 101:2199–2206.

Weixel, K. M. and Bradbury, N. A. (2000). The carboxyl terminus of the cystic fibrosis transmembrane conductance regulator binds to AP-2 clathrin adaptors. *J. Biol. Chem.* 275, 3655-3660.

Wallace B.A., Lees, J.G., Orry, A.J.W., Lobley A., and Robert, W. J., 2004. *Analyses of Circular Dichroism Spectra of Membrane Proteins*. London, UK: University of London.

Appendices:

Sequence of the CFTR protein in humans:

1 mqrsplekas vvsklffswt rpilrkgyrq rlelsdiyqi psvdsadnls eklerewdre

61 laskknpkli nalrrcffwr fmfygiflyl gevtkavqpl llgriiasyd pdnkeersia

121 iylgiglcll fivrtlllhp aifglhhigm qmriamfsli ykktlklssr vldkisigql

181 vsllsnnlnk fdeglalahf vwiaplqval lmgliwellq asafcglgfl ivlalfqagl

241 grmmmkyrdq ragkiserlv itsemieniq svkaycweea mekmienlrq telkltrkaa

301 yvryfnssaf ffsgffvvfl svlpyalikg iilrkiftti sfcivlrmav trqfpwavqt

361 wydslgaink iqdflqkqey ktleynlttt evvmenvtaf weegfgelfe kakqnnnnrk

421 tsngddslff snfsllgtpv lkdinfkier gqllavagst gagktsllmv imgelepseg

481 kikhsgrisf csqfswimpg tikeniifgv sydeyryrsv ikacqleedi skfaekdniv

541 lgeggitlsg gqrarislar avykdadlyl ldspfgyldv ltekeifesc vcklmanktr

601 ilvtskmehl kkadkililh egssyfygtf selqnlqpdf ssklmgcdsf dqfsaerrns

661 iltetlhrfs legdapvswt etkkqsfkqt gefgekrkns ilnpinsirk fsivqktplq

721 mngieedsde plerrlslvp dseqgeailp risvistgpt lqarrrqsvl nlmthsvnqg

781 qnihrkttas trkvslapqa nlteldiysr rlsqetglei seeineedlk ecffddmesi

841 pavttwntyl ryitvhksli fvliwclvif laevaaslvv lwllgntplq dkgnsthsrn

901 nsyaviitst ssyyvfyiyv gvadtllamg ffrglplvht litvskilhh kmlhsvlqap

961 mstlntlkag gilnrfskdi ailddllplt ifdfiqllli vigaiavvav lqpyifvatv

1021 pvivafimlr ayflqtsqql kqlesegrsp ifthlvtslk glwtlrafgr qpyfetlfhk

1081 alnlhtanwf lylstlrwfq mriemifvif fiavtfisil ttgegegrvg iiltlamnim

1141 stlqwavnss idvdslmrsv srvfkfidmp tegkptkstk pykngqlskv miienshvkk

1201 ddiwpsggqm tvkdltakyt eggnaileni sfsispgqrv gllgrtgsgk stllsaflrl

1261 lntegeiqid gvswdsitlq qwrkafgvip qkvfifsgtf rknldpyeqw sdqeiwkvad

1321 evglrsvieq fpgkldfvlv dggcvlshgh kqlmclarsv lskakilllld epsahldpvt

1381 yqiirrtlkq afadctvilc ehrieamlec qqflvieenk vrqydsiqkl lnerslfrqa

1441 ispsdrvklf phrnsskcks kpqiaalkee teeevqdtrl

Yellow highlight = C-terminal region of CFTR. The PDZ domain is formed by the last 4 residues of the sequence (DTRL), conserved hydrophobic residues FLVI (in green) are important in the maturation and stabilization of CFTR, blue sequence represents the negatively charged motif. Source: NCBI (2012)