INVESTIGATING THE EFFECTS OF PREDICTED STABILIZING SECONDARY MUTATIONS IN NUCLEOTIDE BINDING DOMAIN 2 OF CFTR-N1303K

A thesis presented to the faculty of the Graduate School
of Western Carolina University in partial fulfillment of
the requirements for the degree Master of Science in Biology.

By

Joseph Dakota Taylor

Committee Chair: Dr. Robert Youker, Department of Biology
Committee member: Dr. Heather Coan, Department of Biology
Committee member: Dr. Jeffrey Brodsky, Department of Biological Sciences, University of Pittsburgh

June 2023

ACKNOWLEDGEMENTS

I extend sincere thanks to the following people, without whom this thesis would not have been possible. My advisor, Dr. Robert Youker for his continuing support and helpfulness in all circumstances. Lab members Gabby Moses, Jonathan Zirkel, Tony Yost, Hunter Cabe, and Margaret Bray for all their help in the lab. I am indebted to Jeffrey Brodsky for his ideas that helped save the project. My fellow graduate students Melissa Rogers, Hannah Noel and Connor Larmore for their support and help bouncing ideas off each other. Brittania Bintz for her assistance in DNA sequencing and the use of the Forensic Science sequencing equipment. The Graduate Office for summer funding and Sigma Xi for a Grant in Aid of Research (GIAR). Finally, Heather Coan for guidance and training of the tissue culture facility and her continued support. Lastly, I offer my warmest regards and thanks to my parents, brother, sister, grandmother, and friends for their continued support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC- ATP-binding cassette

CFTR- Cystic fibrosis transmembrane conductance regulator

CF- Cystic Fibrosis

DMEM- Dulbecco's modified eagle medium

ERAD- Endoplasmic Reticulum associated degradation

FBS- Fetal bovine serum

GFP- Green fluorescent protein

kBp- kilobasepairs

kDA- Kilodaltons

NBD- Nucleotide binding domain

NMA- Normal mode analysis

NPD-Nasal potential difference

PA- *Pseudomonas aeruginosa*

PCA- Principal component analysis

ppFEV- Percent predicted forced expiratory volume

RI- regulatory insertion

RMSD- Root mean square deviation

RMSF- Root mean square fluctuation

SASA- Solvent accessible surface area

SDS- Sodium dodecyl sulfate

TMD- Transmembrane domain

WGA- Wheat germ agglutinin

ABSTRACT

INVESTIGATING THE EFFECTS OF PREDICTED STABILIZING SECONDARY MUTATIONS IN NUCLEOTIDE BINDING DOMAIN 2 OF CFTR-N1303K

Dakota Taylor

Western Carolina University June 2023

Program Director: Dr. Beverly Collins

Cystic Fibrosis is a disease caused by mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) protein. This membrane protein selectively transports chloride ions into epithelial cells and is important in a wide array of excretory functions. Mutations that hinder CFTR biogenesis, transport, or gating are the underlying cause of Cystic Fibrosis (CF). Impairment of CFTR function can lead to numerous health problems such as lung fluid buildup, pancreatitis, and bowel blockages. The most common mutation is a deletion of F508 (ΔF508) that resides in NBD1 of CFTR but rarer mutations in NBD2 can also lead to disease. For example, N1303K is a rarer mutation in Caucasian populations, but is more common in Mediterranean and Middle Eastern populations. Recent computational experiments from the Youker lab have identified potential stabilizing mutations in the mutant N1303K. The second site mutations L1346F and L1254A were analyzed for this thesis. The MutPred2 analyses of N1303K-CFTR predicted that L1346F mutation would greatly reduce pathogenicity, while L1254A would only have a modest effect. FoldX corroborates this as well, with a ΔΔG value of 2.3 kcal/mol upon mutation to L1254A and -1.26 kcal/mol change upon mutation of L1346F. The results of molecular dynamic (MD) simulations using NBD2 domain revealed different results with the L1254A having a closer RMSF, RMSD, hydrogen bonds and solvent accessible surface area similar to wildtype CFTR, while L1346F had characteristics closer to N1303K, suggesting that

L1346F might not be beneficial. Principle component analysis corroborates MD results, with L1254A having a similar protein movement pattern to wildtype CFTR. Site-directed mutagenesis was used to introduce L1254A and L1346F mutations into a plasmid containing GFP-Tagged N1303K-CFTR. The mutated plasmids were transiently transfected into HEK-293 cells and in-gel fluorescence performed on cellular lysates. The L1346F mutant was slightly stabilized as determined by C/B band ratio. However, a higher amount of L1254A mutant protein was present in the plasma membrane compared to L1346F and N1303K-CFTR as determined by image colocalization analysis. The L1346F mutant had more internal staining and this could indicate aggregated protein, but further cellular and biochemical experiments are needed to confirm. Overall, these results suggest the L1254A mutant has enhanced trafficking to the cell surface compared to N1303K or L1346F, but there is enhanced proteolysis of this double mutant as seen in protein gels. It is not clear if the proteolysis occurs in the cell, or post-lysis. Further cellular experiments with lysosomal and/or proteasomal inhibitors may shed light on when proteolysis occurs.

INTRODUCTION

Pathophysiology of Cystic Fibrosis

Cystic Fibrosis (CF) is an autosomal recessive disease that affects roughly 30,000 people in the USA and 80,000 people worldwide (Brown et al., 2017). In the United States, 2,500 children are born with Cystic Fibrosis each year (Childrens Hospital of Philadelphia, 2014). There are greater than 10 million carriers for CF in the United States (Miller et al., 2020). Cystic Fibrosis typically impacts systems that involve the secretion of mucus, sweat or digestive fluids; therefore, many systems are damaged, such as the digestive, respiratory, reproductive, and integumentary systems. CF patients usually have thick mucus in the pancreas and lungs that can occlude ducts and airways, causing many of the symptoms of the disease (Mayo Clinic, n.d.). In the lungs, the thick mucus buildup causes decreased mucociliary movement and reduced ion transport of chloride and bicarbonate ions; the thick mucus cannot be cleared effectively and persists in the lungs leading to recurrent bacterial infections (Brown et al., 2017). The mucus is an ideal breeding ground for many pathogenic species such as *Pseudomonas, H. influenza,* and *S. aureus* (Brown et al., 2017) . The persistence of these pathogens in the lungs creates a substantial inflammatory response and overtime can cause permanent airway destruction and even death (Brown et al., 2017). Common clinical manifestations of CF, include respiratory insufficiency, wheezing and thick sputum, recurring sinusitis and lung infections, and exercise intolerance (Mayo Clinic, n.d.).

The gastrointestinal system, especially the pancreas, is greatly affected by CF. Mucus is naturally produced to help with the excretion of carbonate and digestive enzymes into the small intestine from the pancreas. In CF, thick mucus occludes these ducts, and because the digestive enzymes cannot exist, the pancreas partially dissolves, causing pancreatitis. Pancreatitis comes along with excruciating pain and also poor digestion of food. Malnutrition is a common symptom in younger CF patients, and is a byproduct of this poor digestion, as well as chronic constipation, stunted growth, and intestinal blockages

(Brown et al., 2017; Mayo Clinic, n.d.). In the reproductive system, CF mutations can influence the formation of urogenital organs. CF can cause improper formation of the ductus deferens, epididymis and ejaculatory ducts, rendering a male CF patient with viable sperm with no method of delivery rendering them potentially sterile. CF in female patients can also reduce fertility and increase severity of pregnancy symptoms. Decreased ion trafficking in skin epithelial cells leads to electrolyte and fluid imbalances in the body (Brown et al., 2017; Mayo Clinic, n.d.).

While the above are the most common symptoms of CF, there can be other system wide complications due to this disease. The bronchi can become scarred due to constant inflammation caused by infection, producing blood in sputum (Brown et al., 2017; Mayo Clinic, n.d.). Persistent inflammation in the sinuses can lead to the development of nasal polyps. Mucus buildup and increased mass in the lung can be attributed to lung puncture and even respiratory failure due to improper ventilation. Pancreatitis can cause diabetes as the pancreas is destroyed, critically lowering insulin levels. The bile duct runs through the pancreas and inflamed tissue can create a blockage and reflux of bile into the liver, thus increasing the risk for liver disease. The electrolyte imbalances in CF patients can even cause osteoporosis (due to less calcium in bone) and increased incidence of fractures (Brown et al., 2017; Mayo Clinic, n.d.). Perhaps most importantly, living with this disease causes great pain, so the mental health of CF patients is negatively impacted. With the severe symptoms and complications, CF patients have a reduced lifespan. However, there are several highly effective treatments to help CF patients to extend their quality and length of life (see current management of CF).

Molecular mechanisms behind Cystic Fibrosis

Cystic Fibrosis is caused by a mutation in the cystic fibrosis transmembrane conductance regulator protein (CFTR). The purpose of CFTR is to transport ions across the plasma membrane. The membrane protein is composed of two nucleotide binding domains (NBDs), two transmembrane domains (TMDs), and a regulatory domain (R domain, Figure 1). CFTR is activated through phosphorylation of its NBDs that cause the CFTR channel to open. This opening allows for transport of chloride and

bicarbonate ions into the lumen of pulmonary and digestive tissues. Water follows the ions to the outside of the cell, providing secretions that are less viscous and more fluid. The bicarbonate ions also create a basic solution that help protect the intestines as well as activate pathogen degrading enzymes. The bicarbonate also helps deliver digestive enzymes to the small intestine (Saint-Criq & Gray, 2017). Mutations to this protein can cause improper ion trafficking, leading to the symptoms listed above.

CFTR is a member of the ATP binding cassette (ABC) transporter superfamily, a set of proteins that bind ATP and promote the transport of substances across the cell membrane. CFTR shares structural homology with bacterial multidrug resistance pumps, which pump toxins out of the cell membrane. However, The R domain is unique to CFTR and controls channel gating through protein kinase A phosphorylation triggered NBD heterodimerization (reviewed in Estabrooks & Brodsky, 2020). Heterodimerization of the NBDs occurs via the interaction of the Walker A, Walker B subdomains and ABC signature motifs of NBD1 and NBD2. This heterodimerization creates two composite interfacial sites called site one and two (Vergani et al. 2005). Site two harbors the Walker motifs of NBD2 (site 2) and is catalytically active compared to site one that is essentially enzymatically dead. The conformational change upon ATP catalyzation causes CFTR to open, allowing for ions to flow through (reviewed in Estabrooks & Brodsky, 2020).

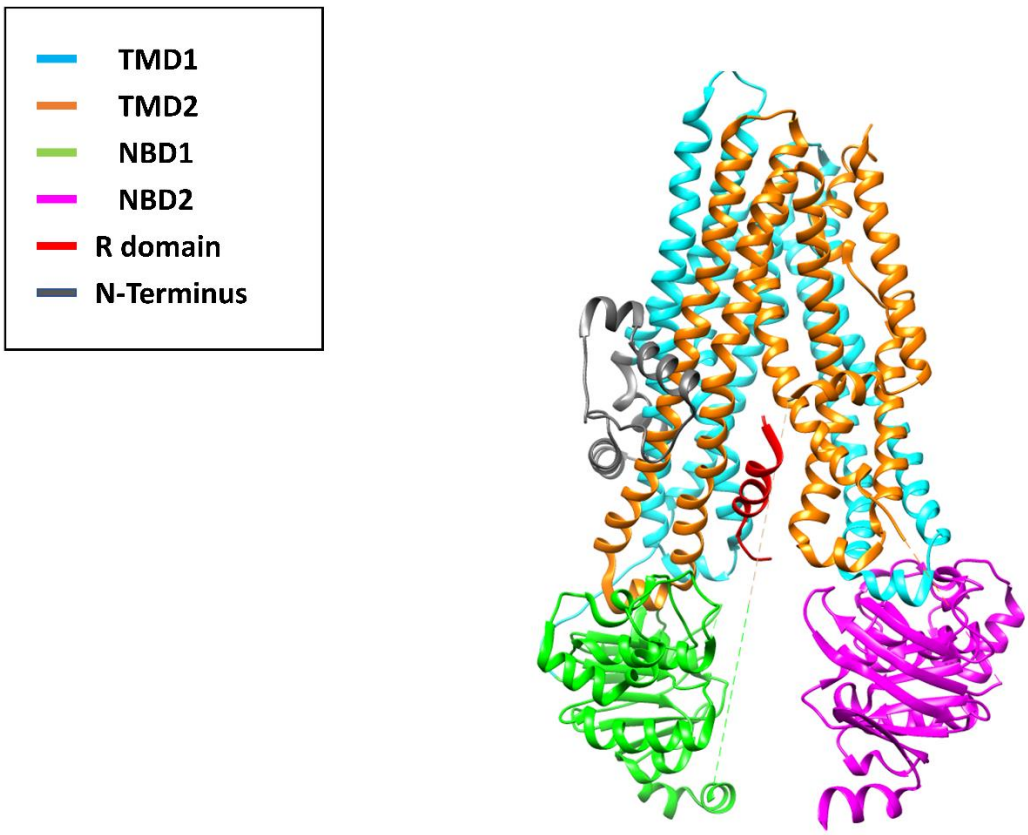


Figure 1. Cryo-EM Structure of apo-CFTR. Domains are color coded in the figure legend (PDB: 5UAK, (Liu et al. 2017)).

CFTR is a unique ABC transporter that acts as a channel instead of a transporter. Also, CFTR aids exclusively in the flow of anions but not the direction of flow, which is unique as an ABC transporter. The anions flow according to their electrochemical gradient (Locher 2016).

More than two thousand CFTR mutations have been identified, causing a wide array of phenotypes (Castellani 2013 & CFTR2 database). Originally, CFTR mutations were categorized into six classes: Class I mutations are nonsense or frameshift mutations that introduce a premature stop codon into CFTR. Class II mutations cause misfolding and impaired biogenesis resulting in ER degradation. Class III mutations impair the regulation of the CFTR channel affecting channel gating. Class IV mutations alter the positively charged core, thereby impeding the ion conduction, and leading to decreased ion flow

through the channel. Class V mutations alter the promoter of the CFTR gene, leading to decreased expression. Class VI mutations reduce the conformational stability of CFTR post-ER, or generate extra internalization signals, leading to CFTR being kept inside the cell and decreasing the plasma membrane lifetime of CFTR (Rowe et al. 2005). This classification of basic defects has been recently expanded to thirty-one categories based on new information describing the complex and combinatorial nature of the CF disease alleles (Veit et al., 2016). The most common mutation that causes CF is ΔF058, where a phenylalanine is deleted from CFTR at position 508 in NBD1. This is quite severe and causes increased morbidity and mortality in CF patients. The ΔF508 mutation is considered a class-two mutation, which means that the protein is misfolded and degraded by ER associated degradation (ERAD) (Fukuda & Okiyoneda, 2020).

Ethnic Variation amongst CFTR Mutations

The ΔF508 is the most common mutation in CFTR, occurring in 66.8% of CF patients worldwide regardless of ethnicity (Petrova et al. 2021). ΔF508-CFTR has the highest incidence in Denmark and the lowest in Algeria (Petrova et al. 2021). The second most common CFTR mutation is G542X, occurring in 2.6% of patients and this mutation is common in Mediterranean populations (Petrova et al., 2021). The third most common mutation (N1303K) occurs in 1.6% of patients and is found predominantly in Mediterranean and Middle Eastern countries, with the highest occurrence being in Tunisia (17.2% of CF patients). The G551D mutation is the fourth most common mutation (around 1.5% of CF patients) and is common in Midwest and central Europe. The W1828X mutation is the fifth most common mutation (approximately 1 percent of CF patients), with a high frequency in Israel (being 36.2% of CF cases) and is also common in Mediterranean and North African population (Petrova et al., 2021).

Genetic Origin of Cystic Fibrosis

A genetic genealogical study has estimated that ΔF508-CFTR originated between 4600-4725 years ago in northwestern regions of France (Farrell et al. 2018). The origin points towards the late

Bronze Age, and based on archaeological and genetic evidence, it is hypothesized that the Bell Beaker people were the original carriers of the mutation. As they were a nomadic tribe that sold many copper artifacts, they could be responsible for the spread of ΔF508 to multiple regions throughout Europe. However, ΔF508 was only recently introduced into the southwestern populations such as the Mediterranean around 1000 years ago (Farrell et al., 2018). N1303K is predominant in the Mediterranean/Middle East, and has high incidences in the countries of Lebanon, Tunisia, Algeria, and Southern Spain, Italy, and France (Farrell et al., 2018) The high occurrences of N1303K in these areas also correspond to the distribution of the Phoenicians (Farhat et al., 2019). It is hypothesized that the Phoenicians were the first carriers of N1303K, and spread it as they traveled, giving introduction of CF mutations multiple points of origin in human history (Farhat et al., 2019).

Current Management of CF

Airway inflammation is common in CF patients and anti-inflammatory medications are a common drug used in a combination treatment plan (Davies et al., 2014). Treating children six years and older with ibuprofen has been shown to reduce the inflammation of the lungs and increase the amount of air inhaled/exhaled (Davies et al., 2014). However, the increased consumption of anti-inflammatory has been shown to lead to other issues such as liver disease, so the dosage must be monitored closely. Mucolytics can also be used to help reduce the thickness of mucus in the lungs of CF patients. Dornase alfa (Pulmozyme) cleaves human DNA outside of epithelial cells in the lungs and helps with mucociliary clearance and mucus thickness (Davies et al., 2014). This drug is usually used on CF patients older than 18, but recent studies showing that CF patients aged 6-18 medicated with this compound also had reduced inflammation of the airway, so there is consideration on using this medication more early in a patient's life (Davies et al., 2014). Another common course of treatment for CF is antibiotics. Antibiotics have been proven effective for treating lung infections of CF patients, especially *Pseudomonas aeruginosa (PA)*, which is the most common pulmonary infection in older CF patients. Antibiotics delivered via nebulizer reduce the presence of PA. Powdered antibiotics can be used but are associated with a higher rate of

pertussis (Davies et al., 2014). Treating patients with hypertonic saline can also be effective for CF patients, as having a high amount of extracellular salt leads to water migrating to body cavities, including the lungs. The increased amount of water in the lungs makes the mucus less viscous and more easily able to be cleared. Mannitol, when inhaled, also increases the amount of water in the lungs (Davies et al., 2014).

The most potent CF treatments available are the classes of drugs called correctors and potentiators. Potentiators are compounds that help regulate CFTR gating and partially restore the flow of ions through the channel. This increases the amount of operable CFTR proteins and thus lessens the effects of CF. Ivacaftor is arguably the most popular potentiator and has been shown to significantly increase the amount of air inhaled and exhaled for the patient (Davies et al, 2014). Correctors are compounds that help CFTR fold properly and traffic to the cell surface. Lumacaftor is a compound that has been shown to help the ΔF508 mutation fold properly and be transported to the plasma membrane, increasing the amount of air the patient can inhale/exhale (Davies et al 2014). While these potentiators and correctors are potent, these drugs are not FDA approved for every CF mutation (Allen et al. 2023).

The most efficacious treatment for CF patients to date is the medicine Trikafta. Trikafta is a combination of two correctors and one potentiator. Trikafta has been exceedingly useful in ΔF508 CF patients, providing an 10% increase in absolute lung function over baseline (Ridley & Condren, 2020). It has helped reduce the morbidity and mortality of CF patients. This medicine has outperformed other dual and single therapies, but patients must have one or more ΔF508 alleles to be approved for this treatment. There are still ~10% of CF patients, who are excluded from these transformational therapies (Zaher et al., 2021).

There has been research in small molecules that could help increase the half-life of CFTR mRNA, this would be helpful for patients with type I mutations. Ataluren, a ribosomal read-through reagent meant to help the ribosome fully translate CFTR, has been shown to help with type I CF mutations (the addition of a premature stop codon). Ataluren does this by binding to the ribosomes and preventing the ribosome

from binding to the premature stop codon, producing CFTR that is fully translated. It has been investigated for other diseases (such as Duchenne muscular dystrophy) and was promising in preclinical trials. However, in larger trials there was not a significant difference between the percent predicted forced expiratory volume (ppFEV) between placebo and treatment groups. The ppFEV is a metric to analyze how well a patient can move air out of the lungs and is a popular way to assess lung function. There also was no significant difference of pulmonary exacerbations (increased cough, mucus production, shortness of breath) between the placebo and treatment group (Allen et al., 2023; Konstan et al., 2020). Another method to overcome truncation of CFTR is by using anticodon edited transfer RNAs (ACE tRNAs). The ACE tRNAs is one such molecule that has been engineered to bind to the ribosome and override the stop codon, as well as preventing premature degradation of CFTR mRNA (another issue of type I CF mutations). The use of this ACE tRNA in HEK-293 cells containing common CF mutations (W1282X, G542X, R1162X) has shown that there is an increase in CFTR mRNA half-life and a rescuing effect on CFTR, with activity levels of CFTR above the threshold to be considered therapeutic (Allen et al., 2023; Ko et al., 2022). Using oligonucleotides to repair CFTR mRNA is also being considered, namely Eluforsen. Eluforsen is a 33 base antisense oligonucleotide that had promising results in ΔF508 patients. *In vitro* and animal studies showed restored CFTR function using electrophysiological assays. Clinical trials with Eluforsen show a statistically significant difference in the nasal potential difference (NPD) between treatment and control groups in patients that are homozygous for ΔF508-CFTR. The NPD was increased in the treatment group, and this increase in NPD correlates with increased function of CFTR (Allen et al., 2023).

mRNA and DNA delivery of the CFTR gene to epithelial tissue in the lungs is also being considered. This gene therapy treatment would be considered mutation "agnostic," as it would benefit any CF patient irrespective of mutation. Translate Bio is investigating a special lipid-based nanoparticle carrier for CFTR mRNA that is aerosolized; it was shown to have promising effects in the first cohort, but not the second, and other companies are trying a similar approach (Allen et al., 2023). The UK CF Gene

Therapy Consortium (GTC) developed a lipid based CFTR DNA vehicle, showing statistically significant but clinically low differences between treatment groups. The GTC is also working with Boehringer Ingelheim and Oxford Biomedica to develop a lentivirus based CF DNA transfer (Allen et al., 2023).

Another method of gene therapy, with the potential to cure all CF mutants, is genome editing. The main method of genome editing goes to CRISPR-CAS 9, a bacterial genome editing tool used to help fend off bacteriophages (Jiang and Doudna 2017). CRISPR has the potential to cut out and insert genes, as well as edit individual bases in the DNA template strand. The majority of CF mutations can be rescued using a single mutation, so this would be ideal. However, base editing can only change sequences of about 3-4 base pairs and the assembly is quite large. If the CF-causing mutation is more complex than a single base change, prime editing (using an RNA template and polymerase) can be used to edit the entire CFTR gene (Scholefield and Harrison 2021). With these techniques, it makes CRISPR attractive for curing CF. There have been promising preclinical results with CRISPR in organoids derived from CF patient cell lines, as well as with animal models (Valley et al., 2019). However, there is a current ethical dilemma and safety concern with editing the human genome in patients; this can cause unwanted mutations in the genome, and this can be detrimental (Ledford 2018).There are currently no clinical studies that assess CRISPR use for CF. Nonetheless, CRISPR-CAS 9 seems to be the future of CF treatment (Lee et al., 2021).

The N1303K Mutation of CFTR

The N1303K Mutation affects folding in the NBD2 domain of CFTR and is more prevalent in Mediterranean/Middle Eastern populations. The N1303K mutation affects NBD2 in an analogous manner to ΔF508 in NBD1, making N1303K a class II mutation. However, the correcting pharmaceuticals that aid ΔF508-related CF have a limited effect on N1303K (Ensinck et al., 2022). Therapeutic effects are only seen with the combination of four potentiators (teza-elexa-ivacaftor and apigenin), and not to the level that is seen in ΔF508-CFTR, making the treatment of N1303K less ideal (Ensinck et al., 2022; Lee et al.,

2021). Creating protein correctors and potentiators for N1303K-CFTR that work analogously to correctors for ΔF508-CFTR would be beneficial, as gene editing is still largely ineffective.

Computational studies of CFTR mutants

Computational simulations of protein are widely used across biology due to their ease of use and the information that they reveal. Computational studies in protein are used to ascertain the effects of molecules or mutations on protein folding. These predictions of folding can be used to ascertain whether or not a mutation has a positive, negative, or neutral effect on folding. It can be used to ascertain the overall stability and/or function of a protein compared to wildtype. It can also be used to test the efficacy of certain molecules in correcting protein folding and be used to design pharmaceuticals to help rescue the protein. However, these computational studies are just predictions, and the larger the protein is, the more difficult it is to simulate properly. Nonetheless, computational studies serve as a proficient basis for creating hypotheses and aid in conducting real world experiments (Gershenson et al. 2020). Computational studies have a major presence in CFTR research and have been used to understand CFTR folding and the effects of certain correctors, being used as a basis for CF drug development. Several key computational studies for this thesis are highlighted below (Rusnati et al. 2020).

The majority of previous computational studies on CFTR have focused on ΔF508, as this is the most common mutation in CF patients. One protein dynamic study performed by Odera et al. analyzed full-length CFTTR in a plasma membrane. The study predicted alteration of a hydrophobic cluster between NBD1 and intracellular loop 4 upon deletion of F508, and disruption of NBD1 and NBD2 interactions (Odera et al., 2018). These alterations could account for the gating defect seen in ΔF508 CFTR. An increase in root mean square deviation (RMSD) was also noted between wildtype and ΔF508. RMSD serves as an indicator for the overall movement of a protein throughout the simulation; the higher the RMSD, the more deviations the protein backbone has gone through. It can be used as a basis for stability, as more stable proteins have lower RMSDs (Aier et al. 2016). The increase in ΔF508 RSMD can thus be a predictor of instability, and these results support the increase in disorder of the NBD1 domain,

and the entire protein, upon removal of F508. A protein dynamics study by Wieczorek and Zielenkiewicz was performed solely on NBD1, and indicated that the deletion of F508 increased the RMSD, which is predicted to predispose ΔF508-CFTR to degradation, agreeing with the results from Odera et al (Odera et al., 2018; Wieczorek & Zielenkiewicz, 2008). Another protein dynamics simulation tool that can be used to ascertain how ΔF508-CFTR effects protein stability is through solvent accessible surface area (SASA). SASA is a geometric measure of protein structure- the more exposed a protein's hydrophobic core is, the less stable it is. It is also worth noting that with the hydrophilic core in CFTR, variation in SASA can be a tool to understand CFTR stability (Durham et al. 2009). A study by Zhenin et al., The NBD1 simulations predicted the SASA of ΔF508-CFTR and wildtype CFTR were not significantly different; however, ΔF508-CFTR exposed hydrophobic residues in NBD1 (e.g., M498 and R560) while reducing exposure to certain hydrophilic residues (e.g., E543 and S511). The Zhenin lab predicts the change in exposure of hydrophobic/hydrophilic residues contributes to the tendency for ΔF508 to aggregate in the cell, therefore SASA could be a tool to predict the stability of CFTR (Zhenin et al., 2015). Another method of computationally predicting how CF-causing mutations affect CFTR is through ΔΔG values, or the change in overall energy of the protein after mutation. A higher positive ΔΔG value correlates with a destabilization effect by the mutation. A study by Bahia et al. compared the predicted ΔΔG values of various mutations in NBD1 (e.g., A534P and S492P) from FoldX, a protein stability prediction algorithm, to *in vitro* ΔΔG values of mutated NBD1 protein. The authors found the FoldX ΔΔG values correlated well with experimental determination of protein stability for the majority of NBD1 and NBD2 mutations tested (Bahia et al., 2021).

A study involving the alteration of ΔF508-CFTR via the removal of the regulatory insertion (RI) by the Aleksandrov lab has also predicted that root mean square fluctuation/deviation (RMSF/RMSD) can be used to measure stability of CFTR. RMSF can be used to predict the overall flexibility of a protein and can be used as an indicator for protein stability much like RMSD (Martínez 2015). The removal of the RI had a more similar RMSF to wildtype CFTR than ΔF508-CFTR, and upon analyzing the CFTR *in vitro*,

the CFTR missing the RI had increased activity and presence in the cell membrane compared to ΔF508-CFTR (Aleksandrov 2010). Proctor and colleagues investigated the effects of additional mutations on NBD1-ΔF508 and identified secondary mutations (S492P) between the 508 loop and the ATP binding subdomain of NBD1 partially restored function (A. Proctor et al. 2015). The ΔF508-S492P NBD1exhibited an RMSF closer to wildtype CFTR than ΔF508-CFTR in MD experiments. Introduction of S492P into CFTR-ΔF508 resulted in a similar protein glycosylation pattern and iodide efflux to wildtype CFTR, indicating the mutation had a rescuing effect on ΔF508-CFTR (A. Proctor et al., 2015). Computational data also shows that second-site mutations can change the arrangement of hydrogen bonds within NBD1. A study by Prins et al. concluded that Wildtype CFTR has a compact NBD1 structure, but the intracytoplasmic loop 4 (ICL4) forms transient hydrogen bonds with residues in the NBD1 domain, such as ΔF508. Deletion of F508 in NBD1 increases domain rigidity and causes a loss of interaction with ICL4 (Prins et al. 2022). The addition of R1070W in ΔF508-NBD1 restored several hydrogen bonds between ICL4 and NBD1, providing evidence that flexibility and transient hydrogen bonds between NBD1 and ICL4 are vital for CFTR function (Prins et al., 2022). The observation of hydrogen bond formation and disruption could be a vital tool to predict how secondary mutations affect CFTR function (Prins et al., 2022).

A previous computational study by Ivey and Youker predicted the addition of secondary mutations to N1303K-CFTR could also rescue folding/stability defects similar to mutations with ΔF508-CFTR (Ivey & Youker, 2020). In this study, the algorithm MutPred2 was used to determine the pathogenicity of the introduced mutation. MutPred2 is a machine learning method that consists of 30 neural networks that utilize both genetic and molecular data to make predictions about the consequences of mutations on protein structure and function (Pejaver et al. 2020). The algorithm was trained >50,000 pathogenic and >200,000 neutral variants from multiple protein databases (e.g., HGMD and SwissVar). The impact of these predicted stabilizing mutations on N1303K-NBD2 have not been investigated and are the focus of this thesis (see Thesis Objectives).

Biochemical studies of CFTR

The CFTR protein contains two sites for N-linked glycosylation (N894, N900). The addition and modification of these N-glycans can be visualized by western blotting, or native in-gel fluorescence to monitor protein folding and trafficking in the cell (He et al., 2021; O'Ryan et al., 2012). The CFTR C-band corresponds to complex glycosylated CFTR that has trafficked through the Golgi stacks, while B-band corresponded to CFTR that has been core-glycosylated in the ER. In previous studies it has been shown that wildtype CFTR has more mature C- than B-band, while ΔF508 has only B band (Amico et al., 2019). Certain correctors such as Lumacaftor and Ivacaftor have been shown to increase the amount of C band for ΔF508 (so called, C-B band ratio) and this measurement can be used as an assay to gauge if mutations, or drug treatments have a beneficial effect on CFTR biogenesis (Amico et al., 2019). Treatment of N1303K-CFTR with C4, C3 and C18 (correctors for ΔF508-CFTR created by Vertex laboratories and Pedimonte 2009) can increase C-B band ratio, but the effect was less pronounced compared to ΔF508 (Rapino et al., 2015). While correctors have been proven to increase the amount of C-band in disease-causing mutants, the addition of second-site mutations can increase the C-B band ratio as well (Vernon et al., 2017; Yang et al., 2018) Several mutations in NBD1 such as S492P and I539T have been proven to increase the C band as compared to deletion of the regulatory insertion (Yang et al., 2018). The implementation of secondary mutations to NBD2 in ΔF508-CFTR has also shown to have a positive effect on protein folding (Vernon et al., 2017). There has been some study on the effects of secondary mutations on NBD2 and how it affects ΔF508-CFTR folding, but it would be beneficial to explore the effects at additional sites in N1303K-NBD2 (Bahia et al. 2021)

Measurement of the amount of CFTR inserted into the plasma membrane can be monitored using a variety of imaging techniques. Klein *et al.* analyzed the localization of CFTR with actin on the cell surface using GFP-tagged CFTR and Texas Red-tagged actin (Klein et al., 2016). The study by Klein *et al.* concluded that the interaction between actin and CFTR can be monitored on the surface of the plasma membrane using colocalization analysis of GFP-CFTR and dsRed-Actin.(Klein et al., 2016) This could be

used as a method to analyze the amount of CFTR in the cell membrane and the effect of secondary mutations on CFTR trafficking.

Thesis Objectives

Multiple studies also have shown that secondary mutations increase the stability of ΔF508-NBD1 and ΔF508-CFTR, but there are scant published studies on N1303K-CFTR and the effects of secondary mutations. The objective of this thesis is two-fold: 1) To characterize protein dynamics of N1303K-NBD2 domain and determine the effects of secondary mutations on N1303K-NBD2 protein. There is a lack of data characterizing the RMSF, RMSD, radius of gyration and backbone flexibility of N1303K-NBD2 domain compared to wildtype. It is known that computational data has been proven to yield valuable insight on how ΔF508-NBD1 affects protein folding, therefore performing simulations on N1303K-NBD2 would be informative. 2) To create these predicted secondary mutations in N1303K-CFTR using site-directed mutagenesis and determine their effect on protein trafficking in HEK293 cells.

Objective 1: Computational Studies. Secondary mutations in N1303K-CFTR were chosen from the study by the Youker Lab. One mutation that was predicted to reduce N1303K pathogenicity is L1346F and the double mutant N1303K-L1346F-CFTR (*hereafter referred to as L1346F*) was compared to the double mutant N1303K-L1254A-CFTR (*hereafter referred to as L1254A*) that was predicted to have a minimal effect on pathogenicity (Ivey & Youker, 2020). MutPred2 analysis was carried out on the mutants in the N1303K-CFTR, as the Youker lab previously analyzed N1303T-CFTR (Ivey & Youker, 2020; Pejaver et al., 2020). NBD2 (residue 1207-1436) from the cryo-EM structure of apo-CFTR (PDB:5UAK (Bank, n.d.)) was used for all protein dynamic simulations. Mutations in NBD2 were made using UCSF Chimera, a protein visualization software (Pettersen et al., 2004). Protein dynamics simulations of wildtype, N1303K, and double mutants were performed using the University of Arkansas Medical School (UAMS) all-atom simulation server WebGro (BEKKER et al., 1993). The collected RMSF, RMSD, radius of gyration, and number of hydrogen bonds were compared between wildtype and mutants. The DΔG values of all mutants and the number of hydrogen bonds formed/broken between mutants were calculated from FoldX, a protein stability prediction software (Schymkowitz et al. 2005). Trajectory data from UAMS were subjected to principle component analysis (PCA) and comparative PCA via MDM-Tracker, an online PCA analysis software (Amamuddy et al., 2021).

Objective 2: Cellular Studies. GFP-N1303K-CFTR was mutated with the selected mutations via site-directed mutagenesis. The GFP-wildtype-CFTR, GFP-ΔF508-CFTR, GFP-N1303K-CFTR and double mutants were transiently transfected into HEK293 cells. The cells were lysed, and the lysate was run on an SDS-Page Gel under native-like conditions (SDS and reducing agent present but no boiling) (O'Ryan et al., 2012). The C and B bands were imaged via in-gel fluorescence, as the CFTR proteins are GFP-tagged. Concurrently, coverslips containing HEK293 cells expressing GFP-tagged CFTR constructs were fixed and co-stained with Wheat Germ Agglutinin-Texas Red (a plasma membrane stain) and imaged via confocal microscopy using a Stellaris 5 microscope. The amount of CFTR in the plasma membrane was calculated via colocalization analysis using Mander's correlation coefficient (MCC).

METHODS

COMPUTATIONAL METHODS

MutPred2 Analysis of CFTR Mutants. The degree of pathogenicity for N1303K and L1346F and L1254A was predicted using MutPred2 (Pejaver et al., 2020). The MutPred2 server uses a machine learning algorithm and 30 feed-forward neural networks to score a missense mutation as pathogenic, or benign based on molecular and genetic information from multiple protein databases, such as HGMD and SwissVar. The human wildtype CFTR protein sequence (P13569, (*CFTR - Cystic Fibrosis Transmembrane Conductance Regulator - Homo Sapiens (Human) | UniProtKB | UniProt*, n.d.)) was used for all analyses and the statistical p-value was set to ≤ 0.05 with a pathogenic threshold of 0.8 corresponding to a 5% false positive rate, or 0.68 corresponding to a 10% false positive rate.

Calculating ΔΔG values for mutations using FoldX. In order to ascertain the change in ΔG between wildtype and mutants, FoldX and Yasara were used (Land & Humble, 2018; Schymkowitz et al., 2005). A FoldX plugin was installed on the Yasara protein visualizer. The FoldX suite analyzes the change in the free energy of the protein (ΔΔG) upon mutation of a specific residue (mutant ΔG - wildtype ΔG) and also displays hydrogen bonds that are formed and broken with each mutation. Figures of the altered hydrogen bonds and ΔΔG values are included in result.

Simulating and mutating NBD2 models in UCSF Chimera. The Nucleotide Binding Domain 2 (NBD2) was taken from a full-length cryo-EM unphosphorylated structure of CFTR (5UAK (Protein Data Bank, 2016)). The PDB file was loaded onto UCSF Chimera, a protein simulation software (Pettersen et al., 2004). NBD2 residues 1207 to 1436 from the full structure were saved into a new PDB file. The NBD2 structure was minimized using Chimera's structure editing settings. Mutations to the wildtype NBD2 were done via the rotamer menu in Chimera. N1303K, L1346F, and L1254A were created, and each mutation was saved to their respective PDB files and had their structures minimized. The resulting PDB file was then submitted to University of Arkansas Medical School WebGro Server (BEKKER et al., 1993) in order to perform an all atom simulation of the protein. The forcefield used was GROMOS96 43a1, with a TIP4P water model. The box type was triclinic and salt type was NaCl. For the energy minimization parameters, the integrator used had the steepest descent at 5,000 steps. The Equilibration and MD run parameters used an equilibration type of NVT/NPT, with a temperature of 300 K. The pressure was set to 1 bar, and a leapfrog MD integrator was used. The simulation time was 50 nanoseconds, with a total of 5,000 frames per simulation (BEKKER et al., 1993). WebGro provides RMSF, RMSD, total hydrogen bonds, radius of gyration, solvent accessible region area (SASA), and trajectory files of the simulation in its output. Plots comparing wildtype CFTR, N1303K, L1254A, and L1346F are given in the results. The change in RMSF (ΔRMSF) between mutants and overall averages of all simulations were calculated from raw results.

PCA and NMA Analysis of CFTR Mutants. The structure and trajectory files of wildtype CFTR, N1303K and the double mutants (L1346F and L1254A) from WebGro were analyzed using an online program called MDM-TASK-web (Amamuddy et al., 2021). This program can perform normal mode analysis of proteins to analyze the movement of the backbone and can also perform PCA, and comparative PCA on the protein dynamics. For each mutant, the PCA 1 vs PCA 2, PCA 1 vs PCA 3, and PCA 2 vs 3 graphs were obtained. The settings that were used were one step and 3 N and C terminal residues ignored. Comparative PCA was done between wildtype CFTR and N1303K, and between N1303K and the double mutants L1346F and L1254A, with the 3 C and N terminal residues ignored. Normal mode analysis was performed on the mutants, with 3 N and C terminal residues ignored and default settings. The percentage explained variance of the protein backbone was compared to the first 50 modes- it was found that over fifty percent of the variation was from the first three modes, so the first three modes were compared between mutants. The variation of the protein backbone movement was visualized using a porcupine diagram.

CELLULAR METHODS

PCR Mutagenesis of N1303K-CFTR. Mutations to N1303K-CFTR were made using the NEB Q5 Site-Directed Mutagenesis Kit (Catalog #E0554S). CFTR that was ligated into PEGFP-C1 plasmid was provided to the Youker lab via the Stanton lab (Moyer et al. 1998). The N1303K mutation was created in wildtype CFTR by Dr. Youker & Dr. Storm's Biotechnology class (Fall 2020); this modified plasmid was used as the basis for the addition of secondary mutations (*Addgene: Vector Database - PEGFP-C1*, n.d.). DF508-CFTR and wildtype CFTR plasmids sharing the same profile were donated to the lab via Dr. Jeffrey Brodsky. Forward and Reverse Primers (25-40 nucleotides) used for site directed mutagenesis were created using NEB Base Changer (*NEBaseChanger*, n.d.). The primers were ordered via IDT and are listed in supplemental table A1 (*Custom DNA Oligos | IDT*, n.d.). Half reactions of the PCR kit were used to save on reagents. 7.25 µL of the Q5 Hot Start High-Fidelity 2X Master Mix is mixed with 0.625 µL of 10 µM Forward and Reverse Primers, 4.5 µL of nuclease-free water, and 0.5 µL of template DNA (concentration ranging from 1-25 ng/µL). The PCR mixture was placed in an Applied Biosystems Thermal Cycler. The cycling conditions were used from the Q5 NEB Site-Directed Mutagenesis protocol. The initial denaturation step was at 98°C for 30 seconds, then the PCR mixture underwent 25 cycles (98°C for 10 seconds, Tm for 10-30 seconds, 72°C for 250 seconds). The Tm was predicted using NEB BaseChanger and the Tms for each primer are listed in supplemental Table A1 The final extension incubated the PCR reactions at 72°C for 2 minutes, followed by an hold at 4°C.

Once the PCR is done cycling, the PCR product was run on an 0.8% agarose gel using SYBR Safe DNA Gel Stain (Catalogue S33102). The gel was made with Biorad 1x TAE buffer (Catalogue 1610743) and 5 µL of DNA Ladder (Thermo Scientific GeneRuler 1kb plus, Catalogue SM1331) was added to the first well of the gel. 5 µL of the PCR mixtures were mixed with a 6x loading dye (NEB Gel Loading Dye Purple 6x #B7024S) and then the mixtures were loaded onto the gel. The gel was run at 200 v for 30 minutes, or until the loading dye was halfway down the gel. The gel was imaged using a Biorad Chemidoc MP scanner, using the SYBR Safe settings. If the PCR mixture was shown to have the desired

product (~10 kb in size), then the mutagenesis procedure continued. The PCR reaction was treated with KLD reaction as per the Q5 Site-directed mutagenesis protocol (1 µL of PCR product, 5 µL of 2x KLD reaction buffer, 1 µL of 10x KLD enzyme mix and 3 µL of nuclease-free water). The mixture is then pipetted up and down and left to incubate at room temperature for five minutes. A tube of the NEB 5-alpha competent E. Coli (from the Q5 Kit) was thawed on ice, then 5 µL of the KLD mixture was added to the tube. The tube was flicked 4-5 times, and then incubated on ice for 30 minutes. The mixture was then heat shocked at 42 °C for 45 seconds, and then incubated on ice for five minutes. 950 µL of LB Broth was added to the cells and pipetted up and down gently. The mixture was then placed in a shaker at 37 °C and 250 rpm for an hour. 750 µL of the mixture was then placed on LB/Kanamycin plates and mixed with glass beads. The plate was then incubated overnight at 37 °C. The next day, 4 isolated bacterial colonies from the CFTR plates were inoculated into 15 mL LB Broth with 15 µL Kanamycin. The broth is shaken overnight at 37°C and 250 rpm.

Miniprep and isolation of mutant plasmid. Plasmid was isolated from the LB broth using Omega Biotech's miniprep kit (SKU: D6942-OOS). The protocol was used as written. Millipure Water was preheated to 70°C to maximize elution of the DNA. One mL of the LB broth was added to an Eppendorf tube and was then centrifuged at 10,000 g for 1 minute. The supernatant was poured off after centrifugation, leaving a pellet behind. Another 1 mL of broth is added to the tube, and this is repeated until 5-10 mL of broth has been condensed into a pellet. 250 µL of Solution I/RNase A was added to the pellet and pipetted up and down to mix thoroughly. 250 µL of solution II was added and the tube was inverted and then incubated at room temperature for 2-3 minutes. 350 µL Solution III was added to the tube and inverted several times, then centrifuged at max speed 14,000 g for 10 minutes. The clear supernatant was then transferred to a spin column whilst avoiding the pellet. The columns were then centrifuged at 14,000 g for 1 minute. Filtrate was discarded then 500 µL HBC buffer was added, and then centrifuged at 14,000 g for 1 minute. The filtrate was discarded again, and 700 µL wash buffer was added to the column and spun at 14,000 g for 1 minute. The column was spun at maximum speed for 2 minutes

to dry out the column, then 100 µL of the preheated DI water was added to the column and spun at 14,000 g for 1 minute. This was repeated twice for a yield of 200 µL.

Determining Concentration of DNA. The concentration of DNA was determined via two methods: the Nanodrop system, and Qubit Fluorometer 3.

Nanodrop. The nucleic acid option was selected and 5 uL of DI water was used for a blank. The blank was loaded, and then the arm of the nanodrop was lowered. Once this was done, the blank setting was selected. Once done, the arm was lifted, and the pedestal was cleaned with ethanol and a Kimwipe. The lysates are loaded in a similar method onto the nanodrop, using only 5 uL of product and then lowering the arm. The results of the nanodrop were recorded.

Qubit. The Invitrogen Qubit dsDNA BR Assay Kit (Q32850) was used with the Qubit to ascertain the concentration of DNA. Two mL of the buffer was mixed with 10 µL of 200x DMSO dye. Ten µL of standard 1 and standard 2 were placed in a Qubit tube, and 190 µL of the buffer/DMSO mixture was added to the standards. One µL of the miniprep DNA was added to 199 µL of the buffer/DMSO mixture. The tubes were vortexed 3-5 seconds, and then incubated at room temperature for 2 minutes. After this, the standards were loaded into the Qubit to obtain a standard curve, then the DNA samples were analyzed as well. The concentration of DNA was recorded.

Sequencing of Isolated DNA. The four replicates of the plasmids were sent to Plasmidsaurus, a sequencing company, in order to perform whole plasmid sequencing and to determine if the mutation had been made (*PlasmidSaurus Sequencing*, n.d.). The samples had to be 30 ng/µL and at least 10 uL of samples were needed. The plasmids used were 4 replicates of wildtype CFTR, ΔF508-CFTR, N1303K, L1254A, and L1346F. Replicates that had the correct mutation and sequence were used for future experiments. Results of sequencing are included in supplemental figures.

Transfection of HEK293 cells using CFTR plasmids

Maintenance of HEK293 Cells. HEK293 cells were a gift, courtesy of Dr. Heather Coan of the biology department. The cells were incubated at 37°C in 5% CO2 in a 10 cm tissue culture dish. The media used was 10% Fetal bovine serum (FBS, Gemini Bioproducts Catalogue # 900-108) in Dulbecco's Modified Eagle Medium (DMEM, Corning, High-Glucose 25 mM HEPES, L-glutamine, Catalogue 10013CV). Cells were split every other day to maintain proper cell density and growth rate. The splitting procedure is listed below.

Splitting and counting HEK293 cells. Twenty-five mL of 10% FBS solution in DMEM was prepared and then incubated at 37 °C in the water bath for 10 minutes. Five minutes after the solution was placed in the water bath, the media in the HEK 293 plate was pulled off the HEK293 plate and put in a waste container. Five mL of Trypsin (Gibco TrypLE Express, REF 12604-021) was placed on the plate and the plate was then incubated at 37 °C for 5 minutes. The plate was inspected under a brightfield microscope to determine cell detachment. Once this was confirmed, 5 mL of the FBS/DMEM solution was added to the plate and the plate was washed thoroughly to remove remaining cells. The media was pipetted into a 15 mL conical tube and centrifuged at 800 rpm at 22°C for 5 minutes. The supernatant was pulled off the pellet, the pellet was resuspended in 10 mL of FBS/DMEM solution. Once resuspended, 15 uL of the solution was pulled off and loaded onto a hemocytometer. The hemocytometer was placed onto a microscope and the microscope is focused to count the 4x4 squares. The cells were counted per square and averaged amongst the 5 squares. The average was then multiplied by 10,000 to determine the number of cells per milliliter. One mL of the resuspended cell pellet mixture was added back to the plate with 10 mL of fresh FBS/DMEM solution (stock plate). The plate was incubated at 37°C with 5% CO2. Cells were split every other day.

Seeding six well plates. Six well plates were used for seeding for transient transfection of HEK293 cells. While splitting the HEK293 cells, the cells are counted as stated above. The desired number of cells for each well is 250,000 cells per well. 250,000 is divided by the number of cells per mL in solution to obtain milliliters of resuspended cells needed for each well. The cells were added to 2 mL of FBS/DMEM solution per well, and slightly shaken. The plates were incubated at 37°C and 5% CO2 overnight. The wells were checked to see if the wells were ~80% confluent before transfection. If not, the wells are incubated longer until the desired confluency is achieved.

Transfection of HEK293 cells. The plasmid DNA was complexed with Turbofect transfection reagent (ThermoFisher, Catalogue 0533). Plasmid DNA (1-4 µg) was added to 400 µL Optimem (Gibco, Catalogue 31985088). The Turbofect was vortexed, then 4-8 uL of the polymer was added to the DNA and Optimem. The solutions were vortexed and then allowed to complex for thirty minutes at room temperature. The amount of DNA and Turbofect varied depending on plasmid and was determined empirically. A list of mutants used for transfection and experimental conditions per mutant is listed in table 1.

Table 1. Amount of Turbofect and DNA needed for CFTR Transfections.

Mutant	DNA (µg)	Turbofect (µL)
Wildtype CFTR	2	6
ΔF508-CFTR	4	6
N1303K	2	6
L1346F	2	6
L1254A	2	6
GFP	1	6
Mock	0	6

After thirty minutes, the mixtures were dropped into the 6 well plates evenly, and the plates were incubated for 24 hours. After 24 hours, widefield and GFP pictures of the wells were taken using the

EVOS microscope to determine percent cells transfected. If the cells fluoresce under the GFP filter, then the transfection was considered effective (except mock well should have no fluorescence). Cells were lysed between 24-48 hours after initial transfection. Three experiments were done for all mutants, and one experiment for ΔF508-Corrected. One well of ΔF508-CFTR was incubated at 30°C 24 hours post-transfection for 12 hours. This was done to stabilize and correct ΔF508 and should lead to increased C band and act as a positive control for correction of CFTR folding.

Preparation of HEK293 cell lysates. After 48 hours of transfection, the cells were lysed using 0.1% Triton X-100 lysis buffer (0.1% Triton-X, 25 mM Tris HCl pH 7.5, 150 mM NaCl, and one tablet protease inhibitor (Millipore Sigma Complete EDTA-Free protease inhibitor cocktail tablets, SKU: 11873580001) per 50 mL solution) The lysis buffer was made and separated into 100 µL aliquots. Due to the protease inhibitor, the lysis buffer was only thawed 1-2 times before use to prevent inactivation of protease inhibitors.

The lysis buffer was thawed on ice at the start of the experiment. The media in the wells was removed, and the wells were washed with one mL of PBS (HyClone, Catalogue SH30256.01) solution. Next, 200 uL of Lysis buffer was added to the wells, and a pipette tip was used to scrape the cells off the bottom of the dish. The 6 well plates were covered in ice and gently rocked in the cold room for 30 minutes. After 30 minutes, the wells were scraped again and then the lysis buffer was pulled off. The lysis buffer was spun at 14,000 rpm for 10 minutes in the cold room, and then 25 uL aliquots of the supernatant were aliquoted into 1.5 ml Eppendorf tubes. The Eppendorf tubes were flash frozen in liquid nitrogen and stored at -80°C for future analysis.

In-gel fluorescence of CFTR protein lysates.

Casting SDS gel. A 10% SDS gel with 4% stacker was created. The resolving portion of the gel was cast at 10% acrylamide, and was composed of 4.85 mL Water, 2.5 mL Acrylamide/Bis (National Diagnostics Accugel 40%, 29:1 Catalogue #EC-852), 2.5 mL of 1.5 M, pH 8.8 Tr-s-HCL, 100 µL of 10% SDS, 5 µL of TEMED (Bio-Rad, catalogue #1610800), and 50 µL of 10% Ammonium persulfate (APS). The 4% stacker was composed of 3.18 mL Water, 0.5 mL Acrylamide/Bis, 1.26 mL of 0.5 M, pH 6.8 Tris-HCl, 50 µL of 10% SDS, 10 µL of TEMED and 25 µL of 10% APS. The 10% gel solution was poured ~75% to the top of the cassette and allowed to solidify. A mixture of butanol and water was placed on top of the gel to ensure the gel solidifies with no jagged edges. Once solidified, the butanol/water mixture was poured off, and the cassette was rinsed with DI water. The 4% stacker was then poured and the comb was inserted into the top of the gel.

Running the SDS gel. The loading dye consisted of 125 mM Tris-HCl, pH 7.6, 10% SDS, 12.5% glycerol, 0.02% bromophenol blue, 12.5 mM of EDTA (pH 8.8) and was supplemented with fresh DTT to a final concentration of 0.125 M. The protein lysates were taken from the -80°C freezer and allowed to thaw on ice. Next, 20 uL of lysate was added to 5X SDS loading dye containing DTT. 100 µL aliquots of the dye were taken and 25 µL of 0.5 M DTT was added just before experiment to give a DTT concentration of 12.5 mM. At the same time, 1X SDS running buffer was placed in the 4°C cold room. 1L of SDS running buffer was made by diluting 10x SDS running buffer (30 g/L Tris-HCl, 144 g/L glycine, 10 g/L SDS) in Millipure water.

The loading buffer and lysate mixture was incubated at room temperature for ten minutes (Remington, 2014). While the lysates were incubating, the gel cassette was placed in the cassette holder and the cold SDS running buffer was poured between the cassettes until the liquid covered the comb. The comb was then pulled out, allowing the wells to fill with the running buffer. SDS running buffer was injected into the wells to clear debris, and the gel was run at 200 V for five minutes. After the lysates were

done incubating, 7.5 uL of Spectra Broad range protein ladder (Thermofisher, Catalogue #26634) was added to the gel, and then 25 µL of the protein lysate mixtures were added. Once loaded, the gel was run at 115V in the cold room for 3 hours, until loading dye has been run off. The gel was then run at 150 V for 1 hour.

After running, the cassette was gently pried open, and the gel is pulled off and placed in water in a small container. The imager used is a Bio-Rad ChemiDoc MP. The protein blots-> Alexa 488 setting was used to image the GFP-CFTR with a 30 second exposure time. The ladder was imaged using blots-> Alexa 688 for 10 seconds. The ladder and CFTR images were combined in Image J. After imaging the gel, the gel was stained with Coomassie stain (450 mL methanol, 100 mL acetic acid, 450 mL water, 0.25 g Coomassie Brilliant blue dye) overnight while being gently rocked. The gel was then de-stained using Coomassie de-stain (75 mL acetic acid, 75 mL methanol, 850 ml Water) overnight. The de-stained gel was then imaged using a white light filter (settings-custom-Coom.) on the ChemiDoc for 0.0005 seconds to determine total protein loaded in each well.

Analysis of C-B Band Ratio using ImageJ. The C and B bands were measured from the fluorescent images taken with the Biorad ChemiDoc using a method similar to Schindelin and colleagues (Schindelin et al., 2012). Measurements were performed in ImageJ and a box was drawn around the C band for the CFTR proteins. The measure function was used to record the intensity of the band. It is important to keep the box the same size for all C bands. A blank well was used to capture the background intensity. The same procedure was performed on the B bands for the mutants.

The intensity of the background was subtracted from the C and B bands, and then the (C)/(C+B) ratio was calculated. A graph displaying the average band ratios was made in Excel. A one-way ANOVA analysis was performed on the band ratios to determine if there is a significant difference in the data sets, and then a Tukey post-hoc test was used to see if the data sets were different between each other.

Confocal Imaging of Cells.

Setup of Glass coverslips for Confocal. Autoclaved glass coverslips were placed in 6 well plates, and then coated with 1 mL of Poly-L Lysine (Millipore Sigma, Catalogue A-0050C) for about 30 minutes. After 30 minutes, the Poly-L lysine was decanted, and the coverslips were covered in 10% FBS/DMEM media and seeded with 250,000 HEK293 cells. The cells were transfected as stated previously in methods, and after 48 hours, the cell media is pulled off the coverslip. The coverslips were washed with 2 mL PBS (HyClone, Catalogue S530256.01) by dribbling the solution down the side of the well in order to not disturb the cells. The PBS was removed and then 1 mL of 4% Formaldehyde in PBS solution (prewarmed at 37 °C) was added to the wells. The plate was put in the incubator at 37 °C for ten minutes, and then the plate was taken out, and formaldehyde solution removed. The 6 well plate was incubated 4 °C overnight in the fridge. The 6-well plate was removed from the fridge the next day, and PBS was removed. 400 µL of PBS was added to each well with 2 µL of Wheat Germ Agglutinin-Texas Red stain (Invitrogen, Catalogue W21405) to stain the plasma membrane of the HEK293 cells. The six well plate was rocked gently back and forth for ten minutes. The PBS was pulled off and washed with 1 mL of PBS and rocked for five minutes three times. 40 µL of mounting media (Invitrogen Prolong Gold antifade reagent, Lot P36930) was added to a microscope slide. After this, the coverslips were taken out of the well. Excess liquid was dabbed off, then the coverslip was placed cell-side down on the mounting media and microscope slide. The coverslip was gently lowered onto the media and pressed gently to get rid of as many bubbles as possible. The coverslip was dried in a dark cabinet for 30 minutes. The coverslips were then sealed with finger-nail polish and stored at 4 °C until imaging. Three experiments were done per mutant, with the exception of ΔF508-CFTR and GFP-PM1 with an experiment number of two (Ward et al. 2015).

Imaging Cells using Confocal Microscope. The stained and mounted coverslips were imaged using a Leica Stellaris 5 confocal microscope. 563 nm and 488 nm Lasers were used to excite Texas Red and GFP, respectively. The images were exported and processed and analyzed in ImageJ.

Analysis of Confocal Images using ImageJ. The colocalization of pixel intensities from GFP (CFTR) and Texas Red (membrane staining) images were performed to determine if the CFTR mutations enhance trafficking of the protein to the cell membrane. GFP attached to PM1, a plasma membrane targeting sequence, was used for a positive control, and mock was used as a negative control (Ward et al. 2015). Fifteen cells per experiment were analyzed per mutant using thresholded Mander's correlation coefficient (MCC). The higher MCC value is interpreted as increased CFTR in the plasma membrane. The images from the confocal were loaded into ImageJ, and then the red and green channels were split. Background subtraction was performed by creating duplicates of each channel and applying a 32-pixel median filter to the duplicate. The duplicates were then subtracted from the original images to get an image with lower background (Dunn et al., 2011). The subtracted images were then used for colocalization analysis. Borders were drawn around individual cells, and the images were cropped to the cell for each channel. The colocalization threshold option in ImageJ's plugin menu was used with default settings. Channel 1 was the Texas Red cell image, and channel 2 was GFP. If there was any colocalization, a correlation plot and an image of colocalized pixels were generated. White pixels in colocalization image indicated GFP and Texas Red overlap. TM1 and TM2 values were generated from these images. TM1 is red to green overlap, and TM2 is green to red. For the purposes of this thesis, TM2 was used for analysis. A violin plot of the TM2 values was created using Plots of Data (Postma & Goedhart, 2019). One-way ANOVA analysis was performed on the MCC values, and Tukey post-hoc tests were used to determine statistical significance between samples.

RESULTS

Computational Results

Mutpred2 results of CFTR mutants.

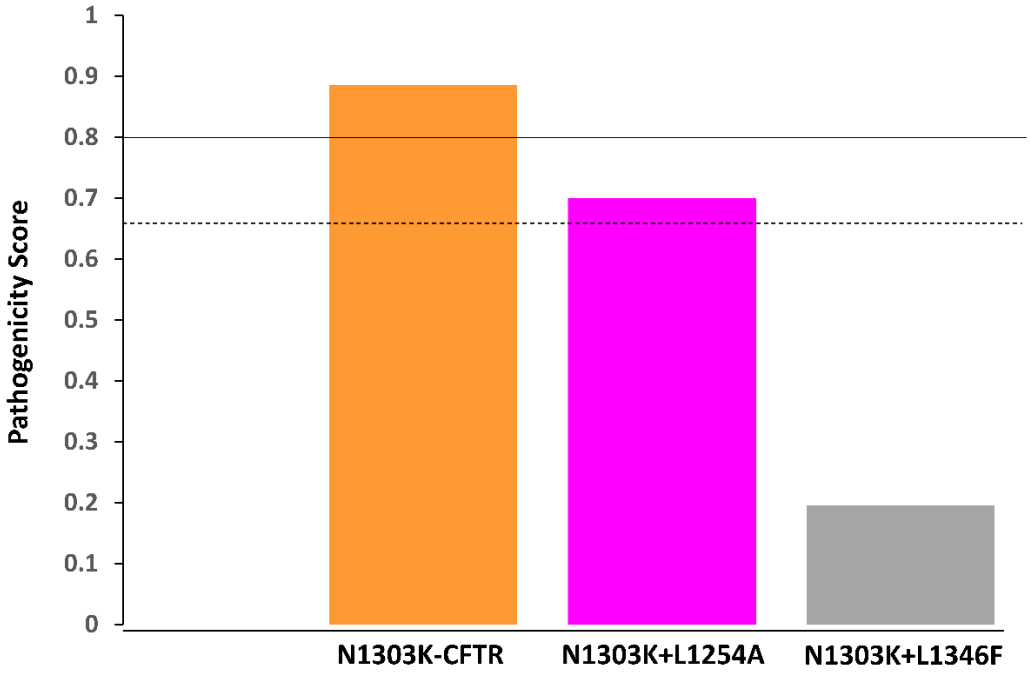


Figure 2. MutPred2 results for N1303K, L1254A, and L1346F. Dashed line at 0.68 denotes a false positive rate of 10%, while solid line denotes a false positive rate of 5% (Pejaver et al., 2020).

MutPred2 results predict that N1303K-CFTR has a pathogenicity score of 0.886. The L1254A double mutant has a slightly lower score of 0.7 and is predicted to be pathogenic. In contrast, the L1346F mutant has a 0.196 pathogenicity score that is substantially lower compared to the single N1303K, suggesting this secondary mutation could aid in restoration of CFTR structure/function.

FoldX analysis of NBD2-CFTR.

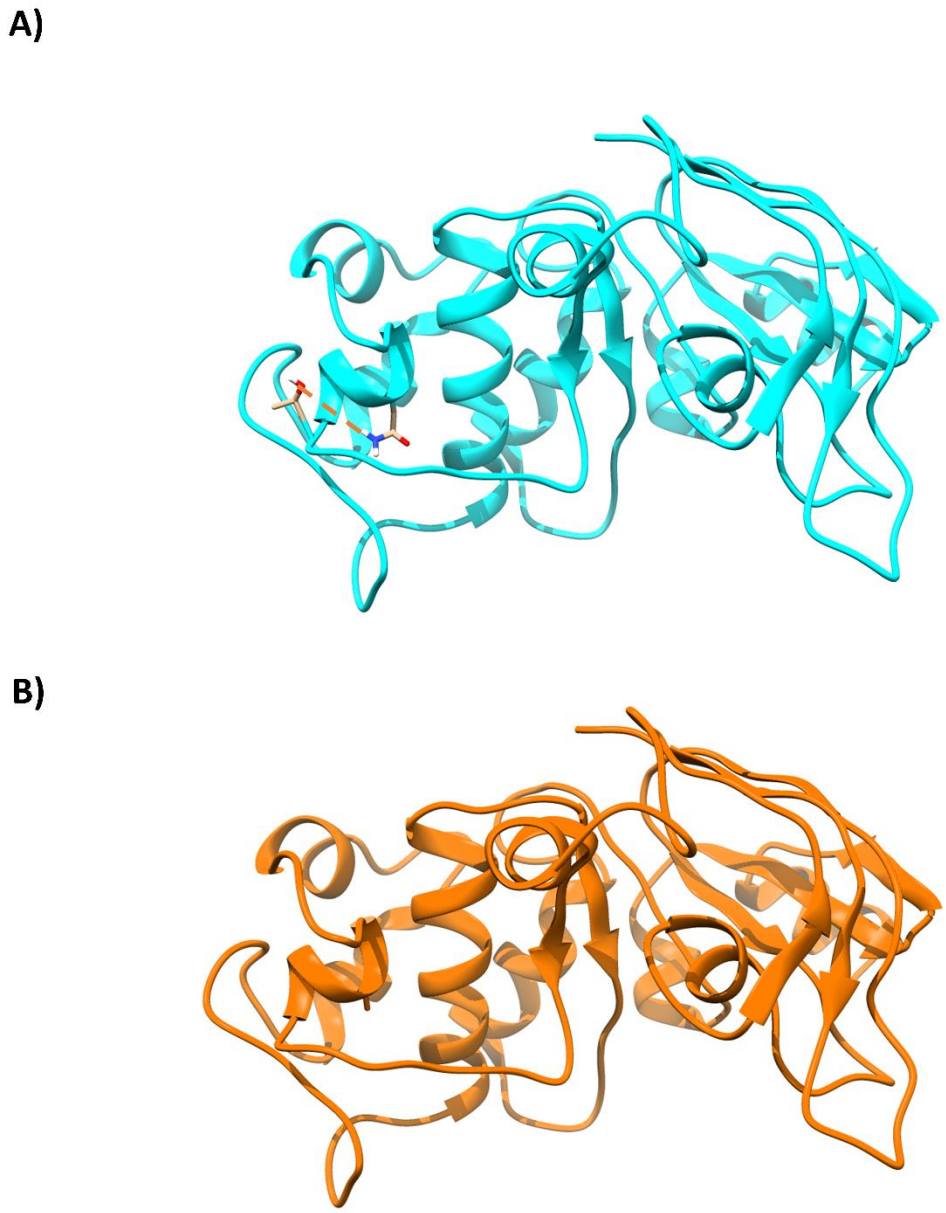


Figure 3. A) hydrogen bonds predicted to be broken and b) new hydrogen bonds predicted to be formed by mutating wildtype NBD2 to N1303K. PDB File: 5UAK, Apo Full-length Human CFTR. Residues 1207 to 1436 were used for analysis.

.

Bahia and colleagues determined the FoldX algorithm had an 86% and 75% correct classification rate for predicting the stabilizing/destabilizing effects of mutations in CFTR NBD1 and NBD2 domains, respectively (Bahia et al., 2021). They observed formation of new H-bonds usually led to stabilization, while destabilizing Van Der Waals interactions had negative consequences. The FoldX algorithm returned an ΔΔG value for the N1303K of 4.4 ± 1.1 kcal/mol confirming its destabilizing effect on CFTR. Furthermore, a predicted loss of a hydrogen bond between ASN-1303 and THR-1299 upon mutation of ASN to LYS at position 1303 was also predicted (Figure 4).

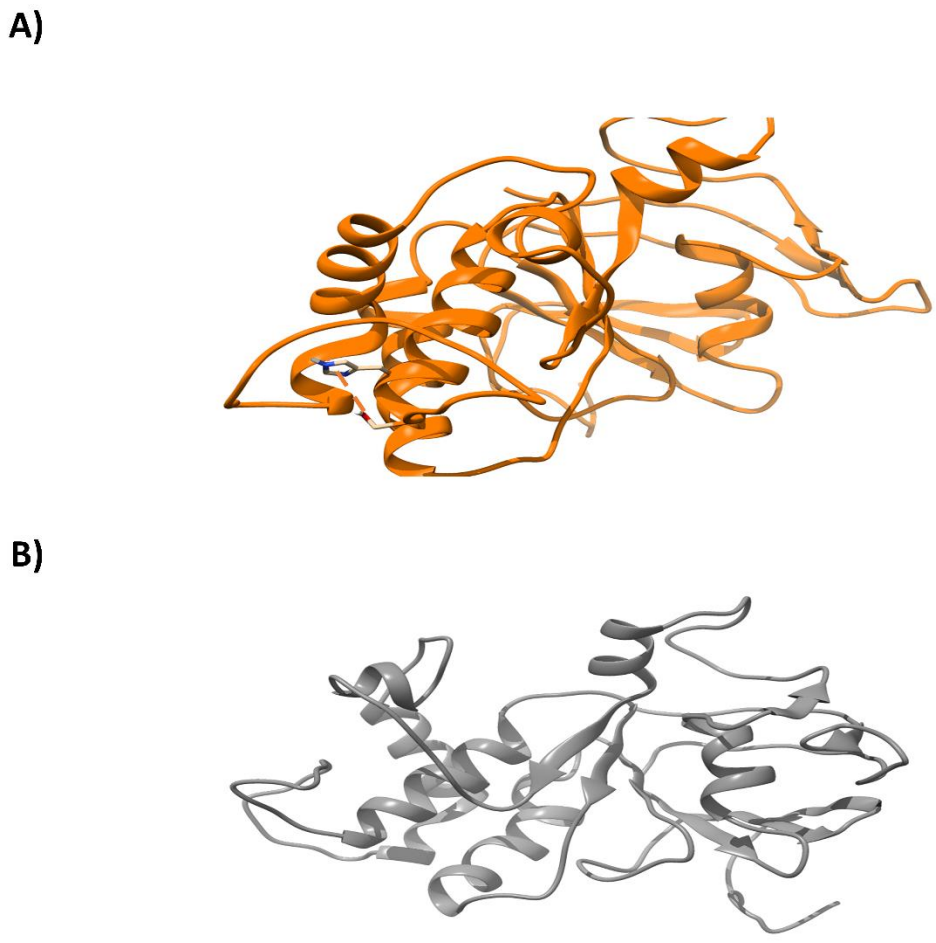


Figure 4. A) hydrogen bonds being broken and b) hydrogen bonds being formed by mutating N1303K NBD2 to L1346F. PDB File: 5UAK, Apo Full-length Human CFTR. Residues 1207 to 1436 were used for analysis.

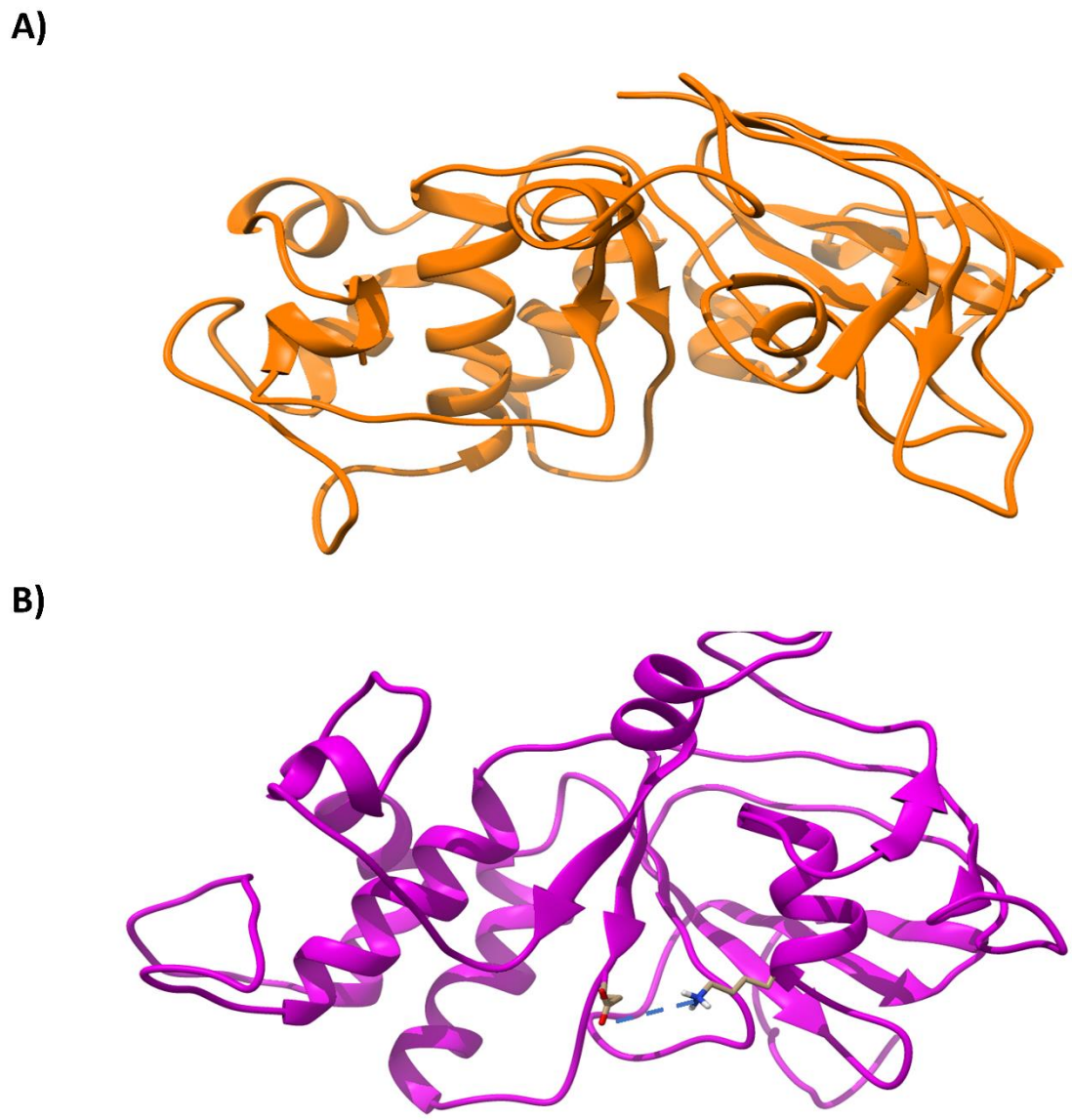


Figure 5. A) hydrogen bonds being broken and b) hydrogen bonds being formed by mutating N1303K NBD2 to L1254A. PDB File: 5UAK, Apo Full-length Human CFTR. Residues 1207 to 1436 were used for analysis.

The FoldX program predicts a ΔΔG value of -1.2606 ± 0.02763 kcal/mol upon addition of mutation L1346F to N1303K and a ΔΔG value of 2.3051 ± 0.5869 kcal/mol for L1254A (Table 2). The

SER-1347 to HIS-1350 hydrogen bond is predicted to be broken when mutating N1303K to L1346F and no new hydrogen bonds are predicted to be formed. It is predicted that no hydrogen bonds are broken when mutating N1303K to L1254A, but a hydrogen bond between LYS-1250 to GLU-1271 is predicted to be formed (Table 3).

Table 2. The ΔΔG values predicted by FoldX for each mutant.

Mutant	**ΔΔG (kcal/mol)**
Wildtype NBD2 to N1303K	4.4 ± 1.1
N1303K-NBD2 to N1303K-L1254A	2.3051 ± 0.5869
N1303K-NBD2 to N1303K-L1346F	-1.2606 ± 0.02763

Table 3. Summary of predicted hydrogen bonds broken and formed upon mutation of NBD2-CFTR

Mutation	**H-bonds broken**	**H-bonds formed.**
Wildtype NBD2 to N1303K	ASN 1303 to THR 1299	NA
N1303K-NBD2 to N1303K-L1346F	SER 1347 to H1350	NA
N1303K-NBD2 to N1303K-L1254A	NA	LYS 1250 to GLU 1271

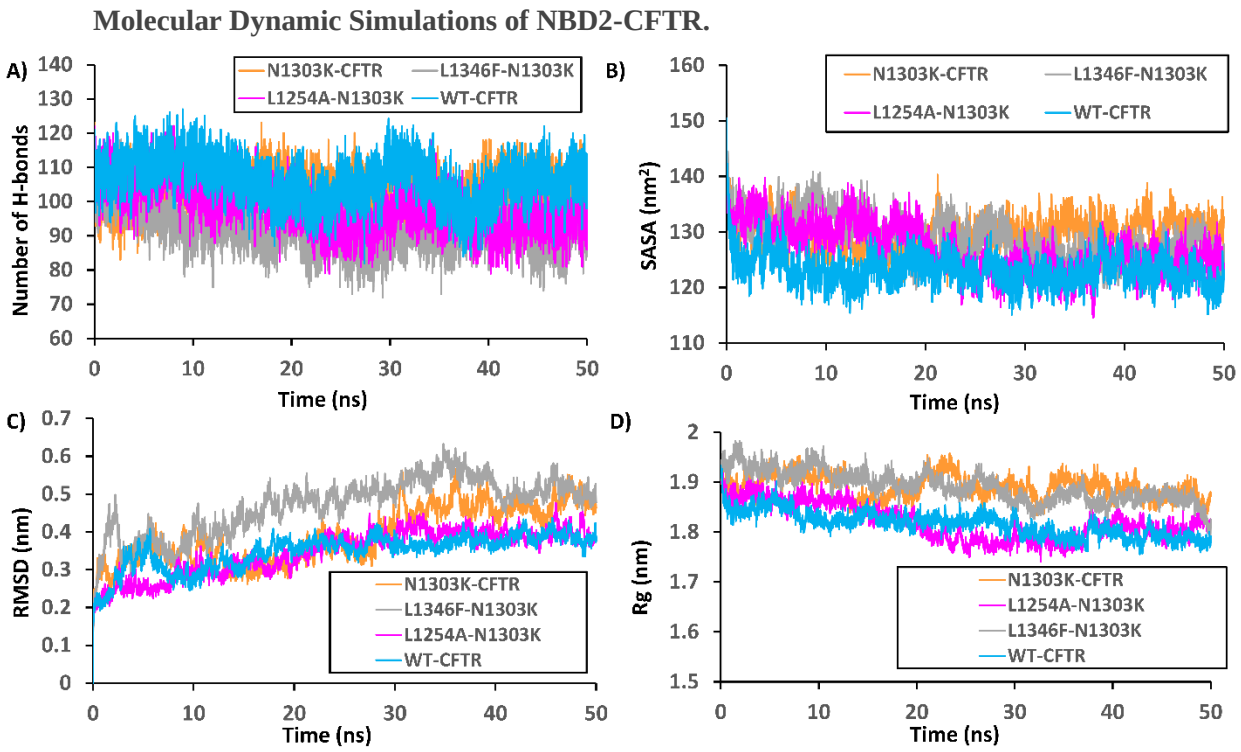


Figure 6. Results of Molecular Dynamic Simulations A) number of hydrogen bonds, B) solvent accessible surface area (SASA), C) Root mean square deviation (RMSD), and radius of gyration (Rg) for wildtype NBD2 and all mutants for 50 nanosecond all-atom simulations.

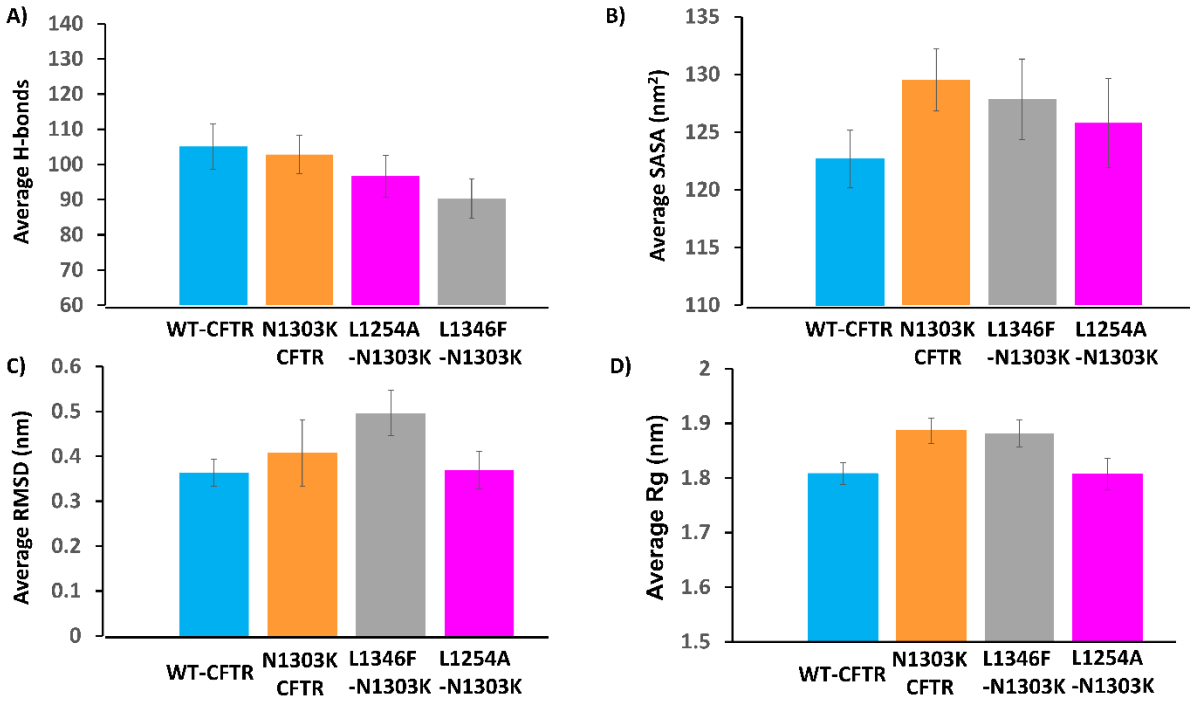


Figure 7. Averaged results of Molecular Dynamic Simulations from Figure 7 for A) hydrogen bonds, B) Solvent accessible surface, C) RMSD, and D) radius of gyration (Rg) for all mutants. Averages are taken between 10 - 50 nanoseconds from the simulation. Error bars are standard deviation.

The L1254A and L1346F double mutants exhibited ~10-15% reduced hydrogens compared to N1303K (Figure 8A). In addition, the N1303K, L1254A, and L1346F mutants had increased solvent accessible surface area (SASA) compared to wildtype (Figure 8B). Wildtype, N1303K, and L1254A had similar average root mean squared deviation (RMSD) values of 0.3-0.4nm, but the L1346F mutant had increased RMSD >0.4nm (Figure 8C). The radius of gyration (Rg) for N1303K and L1346F was increased compared to wildtype and L1254A (Figure 8D). In summary, these results suggest that N1303K and L1346F mutants are not folded as tightly as wildtype and L1254A NBD2 domains based on RMSD, SASA, and Rg values.

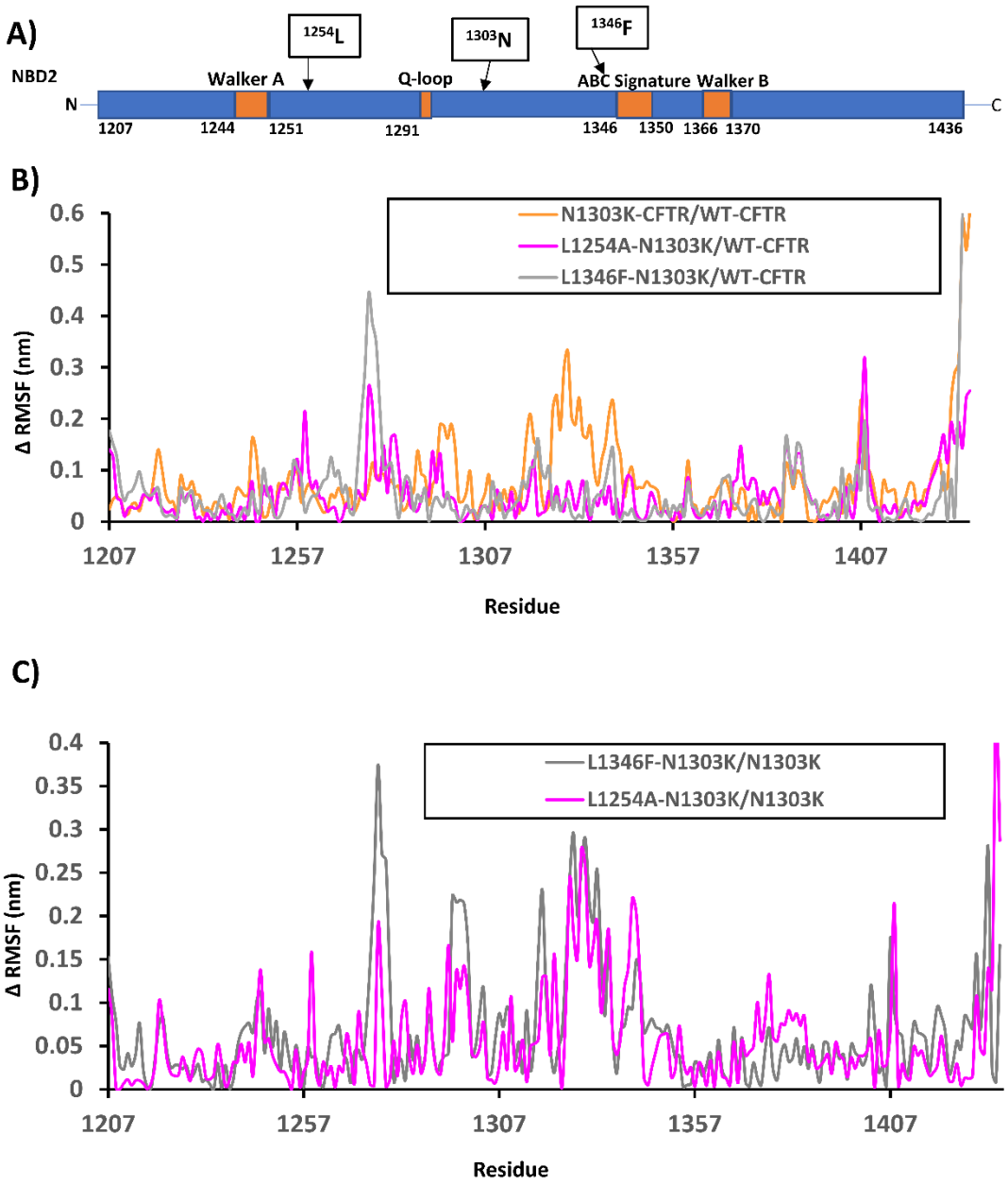


Figure 8. A) cartoon of motifs in NBD2-CFTR, B) ΔRMSF between wildtype CFTR and N1303K, L1254A, and L1346F C) ΔRMSF between L1254A, or L1346F compared to N1303K.

Differences in root mean square fluctuations (ΔRMSF) among the different NBD2-CFTR proteins was used to identify changes of interest in protein backbone dynamics (Figure 9B & 9C). There were notable changes in the backbone motions of 0.1-0.3 nm from residue 1307 to 1357 when comparing N1303K to wildtype CFTR. In contrast, the double mutants (L1254A, L1346F) did not have this large

change in RMSF compared to wildtype (Figure 9B). There are slight variations before the Q-loop at residue 1276 for L1254A (0.25 nm) and L1346F (0.4 nm). A large fluctuation is apparent for all three mutants around residue 1407. Comparing the double mutants to N1303K reveals a large difference (0.2-0.3 nm) between the ABC signature and Q-loop from residue 1307 to 1357. Differences are also apparent before and after the Q-loop. Generally, the L1346F has an overall higher ΔRMSF than L1254A when compared to the single mutant N1303K. There are minor differences in the Walker A and B subdomains for the double mutants compared to N1303K.

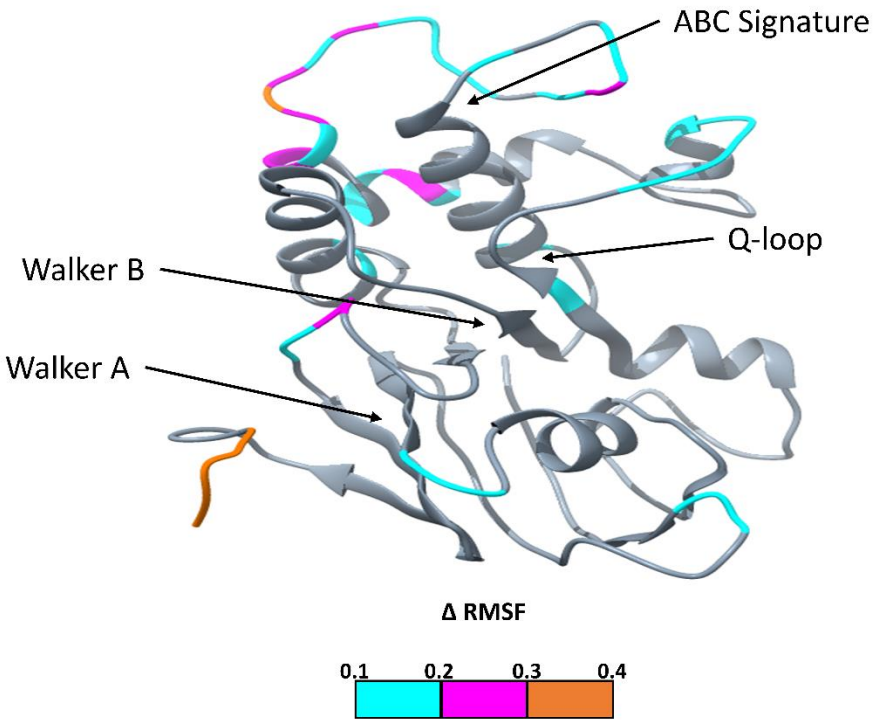


Figure 9. Overlay of ΔRMSF between wildtype NBD2 and N1303K-NBD2 on the protein structure of N1303K-NBD2. Cyan represents ΔRMSF between 0.1 to 0.2, magenta is 0.2 to 0.3, and orange is 0.3-0.4. Subdomains labeled.

Overlaying ΔRMSF values onto the N1303K-CFTR NBD2 protein structure help to visualize the increased backbone fluctuations in the ABC structure and Q-loop. Most residues affected are part of highly flexible looping structures, but there seems to be a difference in flexibility in some alpha helices

and beta sheets as well (Figure 9). Overall, these differences in RMSF indicate N1303K-CFTR NBD2 has different backbone motions than Wildtype primarily around the ABC signature and Q-loop that could contribute to the mutation's pathogenicity.

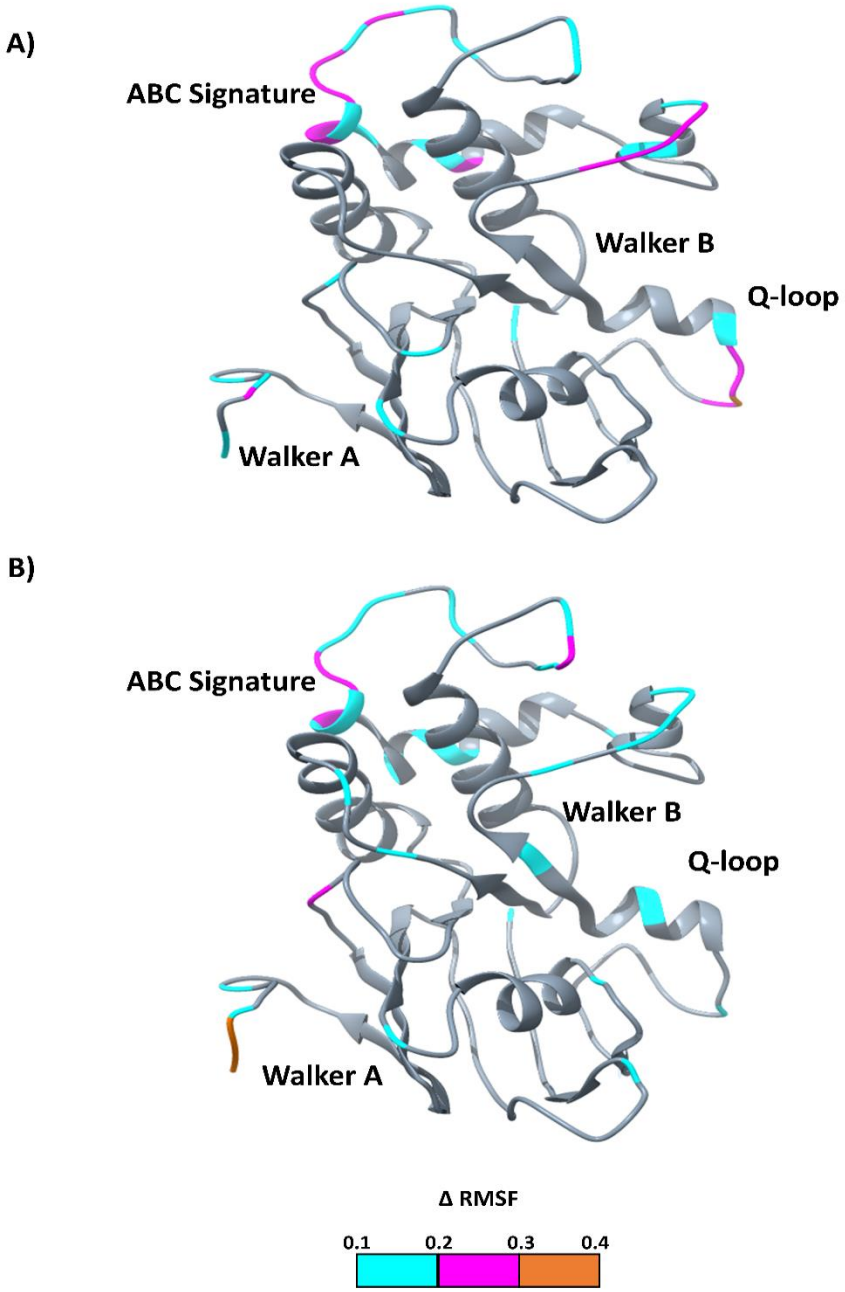


Figure 10. Overlay of ΔRMSF between N1303K-NBD2 and A) L1346F and B) L1254A on respective protein structures. Cyan represents ΔRMSF between 0.1 to 0.2, magenta is 0.2 to 0.3, and orange is 0.3-0.4. Subdomains labeled.

L1346F has a higher ΔRMSF on the alpha helices and flexible regions near the ABC signature sequence compared L1254A (Figure 11AB). The overall change in RMSF near the ABC signature and Q-loop subdomains mirror the change in RMSF between wildtype CFTR and N1303K, and according to Figure 9B, both double mutants are lower in difference from wildtype.

PCA Analyses of NBD2-CFTR. Protein folding is considered to occur through a series of interconversions between ensembles of protein conformations and not a simple single-state linear pathway. Protein dynamic simulations can be analyzed based on the diffusive properties of the protein backbone (Maisuradze 2009). Principal component analysis (PCA) is a covariance matrix based mathematical calculation that can reduce complex multivariable data sets to a lower dimension for ease of analysis. Thus, PCA can be used to identify essential protein motions that occur at all stages of protein folding and reduce the complex 3D motions from simulations into two-dimensional PC plots.

PCA was performed on NBD2-CFTR of Wildtype, N1303K, L1346F and L1254A to analyze the movements of the protein backbone during all-atom simulations and to compare these motions to wildtype. The first three principal components (PC1-3) were compared to each other for wildtype and all mutants to ascertain the global motions each protein exhibited. The axes of the PCA range from -4 to 4, where 0 denotes the median movement of that component state. The PCA was used to compare the overall movement patterns between mutants, and it is predicted that a double mutant that has a more similar movement pattern in the PCA to wildtype NBD2 than N1303K will have more similar folding and increased stability (Amamuddy et al. 2021).

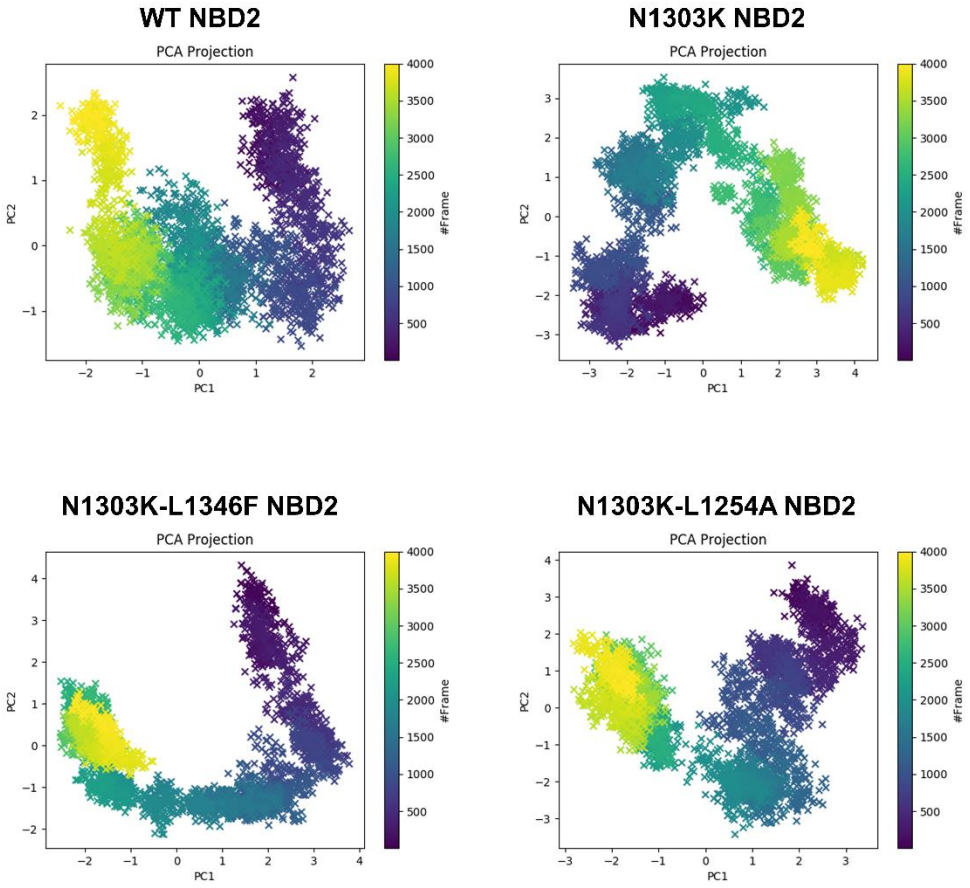


Figure 11. Principal Component Analysis (PCA) of wildtype NBD2, N1303K, L1346F, and L1254A. PCA is between PCA 1 and PCA 2. Points are individual protein conformations across the simulation. PCA spans 1,000 to 5,000 picoseconds. Each frame is denoted by a gradient ranging from blue to yellow as the simulation proceeds.

Wildtype NBD2 and N1303K protein dynamics are substantially different, with the wildtype-CFTR simulation starting around 2 on PC1 and N1303K starting on -3. Wildtype-CFTR simulation ends on -2,2 on the graph, and N1303K simulation ends at 4, -2. Both L1346F and L1254A have similar simulation distribution across the simulation as compared to wildtype NBD2. L1346F started around (2,4) and ended around (-2,1), while L1254A started the simulation around (3,3) and ended it around (-2,2).

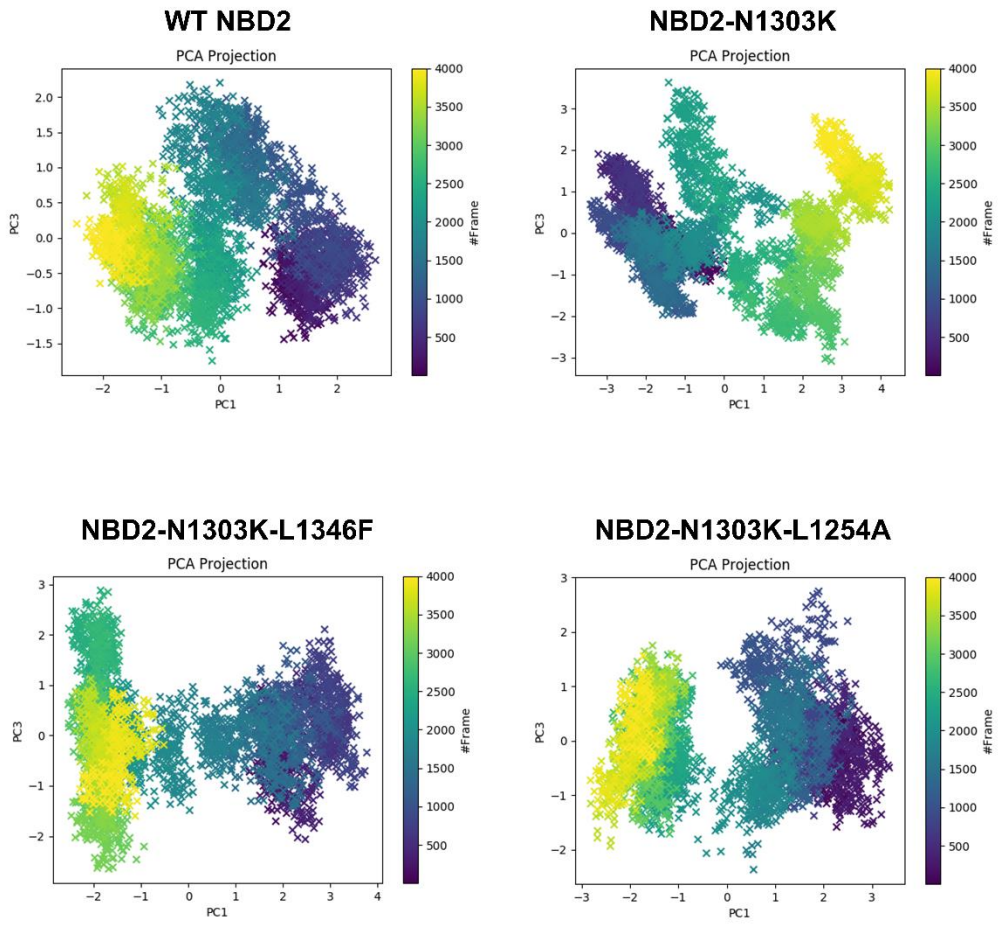


Figure 12. PCA analysis of wildtype NBD2, N1303K, L1346F and L1254A. PCA is between PCA 1 and PCA 3. Points are individual protein conformations across the simulation. PCA analysis spans 1,000 to 5,000 picoseconds. Each frame is denoted by a gradient ranging from blue to yellow as the simulation proceeds.

The wildtype NBD2 simulation starts around (1, -1), and ends at (-2,0). While the movement towards negative 2 on PCA 1 is similar to PCA 1 vs PCA 2, while the distribution on PCA 2 and 3 is opposite. N1303K also exhibits this movement pattern, with PCA 1 axis being the same between PCA 1vs PCA 2 and PCA 1 vs PCA 3, but has opposite movement between PCA 2 and PCA 3. The movements for the double mutants are similar between PCA 1 vs 2 and PCA 1 vs 3. There is a slight separation in the simulation data points in N1303K and the double mutants. Both double mutants have

similar movements to wildtype CFTR and opposite to N1303K. Wildtype NBD2 have opposite movements as also seen in PCA 1 vs 2.

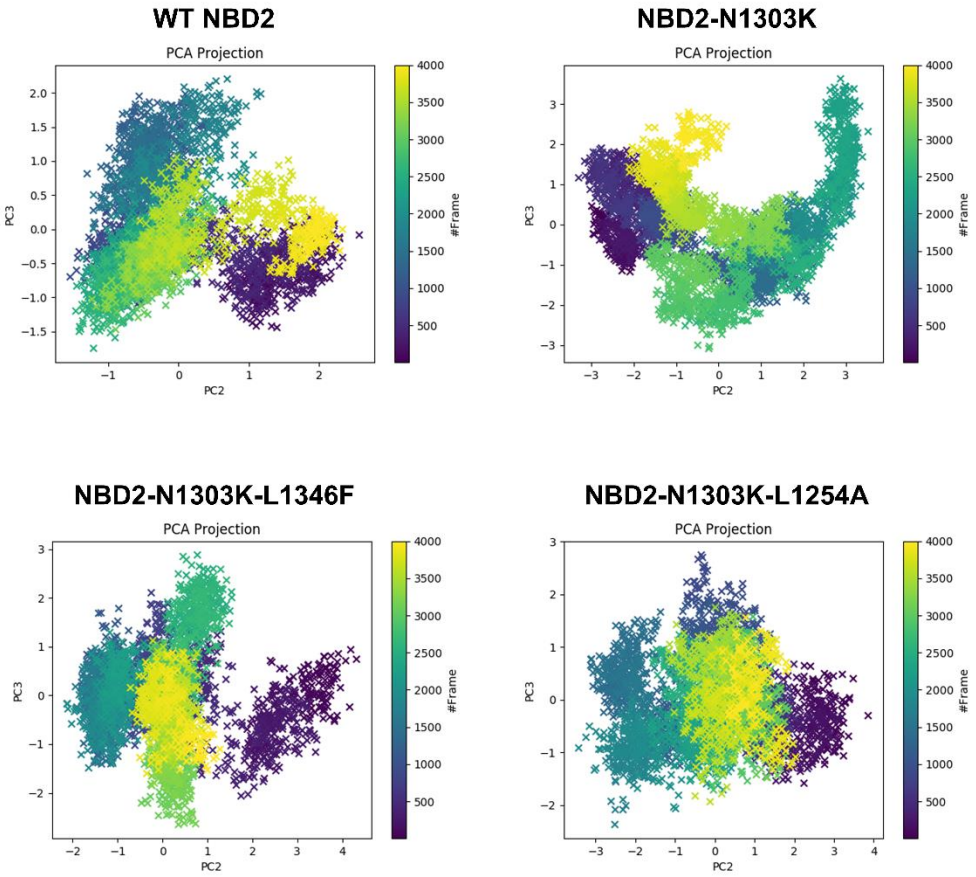


Figure 13. PCA analysis of wildtype NBD2, N1303K, L1346F and L1254A. PCA is between PCA 2 and PCA 3. Points are individual protein conformations across the simulation. PCA analysis spans 1,000 to 5,000 picoseconds. Each frame is denoted by a gradient ranging from blue to yellow as the simulation proceeds.

PCA 2 vs 3 exhibits different movement patterns as compared to PCA 1 vs PCA 2 and PCA 1 vs PCA 3 for each mutant. Wildtype CFTR has a wide distribution, but both start and end around (1.5, -0.5). N1303K also has similar start and end patterns, but the start and stop data points are on the opposite side of the graph, around (-2,2). L1346F has the furthest separation between start and stop, with a start around

(4, 0) and an end around (0,0). L1254A has a similar distribution to wildtype NBD2, with a start and stop around (1.5,0).

Comparing all of the PCAs, there are stark differences between N1303K and wildtype NBD2. N1303K has very different protein movements, with N1303K beginning and ending exactly opposite of wildtype. The protein movement distributions are almost perfect mirrors of wildtype that reinforces the fact that N1303K behaves very differently than wildtype NBD2 experimentally. L1254A follows very similar protein movements compared to wildtype, whereas L1346F has a more similar movement to N1303K. The overall similarity between wildtype and double mutant PCAs indicates that the double mutants may have restorative effects on protein movement/folding, as the double mutants have more similar movements to wildtype. Comparative PCA was done on MDM-Task as well, and it was found that wildtype NBD2 and N1303K had very little similarity in protein movement, and L1254A had some overlap with wildtype, indicating that the proteins have somewhat similar protein movements. L1346F, however, had increased separation between the double mutant PCA and wildtype (Supplemental Figures A1-A5, (Amamuddy et al. 2021)).

Normal Mode Analysis of NBD2-CFTR. Normal mode analysis (NMA) was performed on NBD2-CFTR to assess how the addition of double mutants affects the flexibility and movement of the protein backbone. NMA is used to assess the flexible states available to a protein around an equilibrium point, characterizing the pattern of movement of the protein into different modes that represent the collective motions of the protein (Bauer 2019). Upon performing NMA it was concluded that the first three modes of movement comprised over 75% of protein movement, so only the first three modes were used for analysis (Figure A6). The first three modes will be displayed in a porcupine plot, which is a visual representation of the movement of the protein backbones in arrows. The larger the arrows, the higher the amount of movement/flexibility in this region of the protein (Bauer et al. 2019a; Amamuddy et al. 2021).

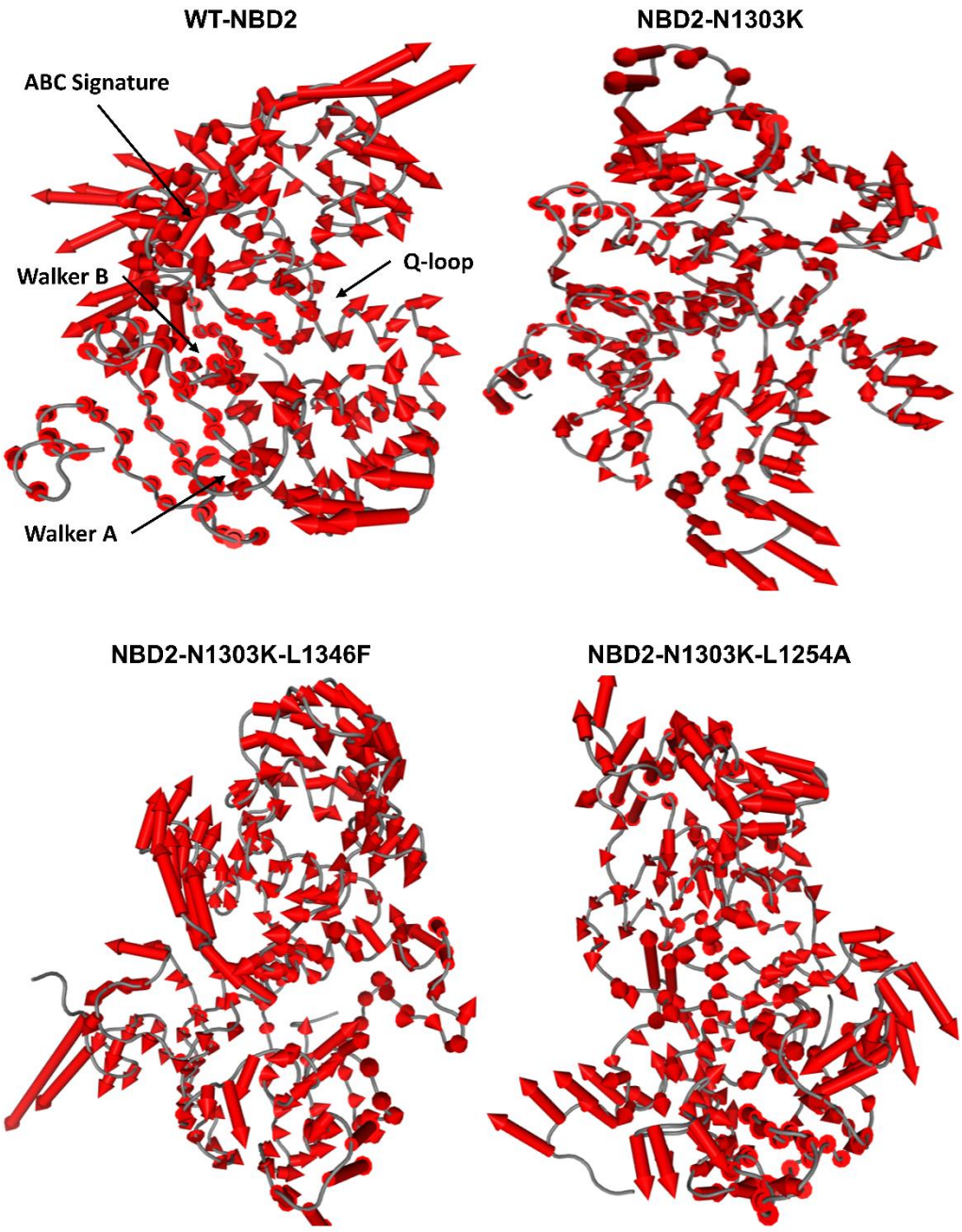


Figure 14. Porcupine plots depicting protein backbone fluctuation using mode 1 of normal mode analysis. Subdomains are labeled on wildtype NBD2 (5UAK: residues 1207 to 1436). All proteins are positioned similarly.

Walker A/B and ABC signature motifs display the largest movements. In contrast, N1303K exhibits reduced movement in ABC Signature and Walker B compared to wildtype. L1346F exhibits

motions around the ABC signature comparable to wildtype, while L1254A displays movement around the ABC signature and near the Q-loop.

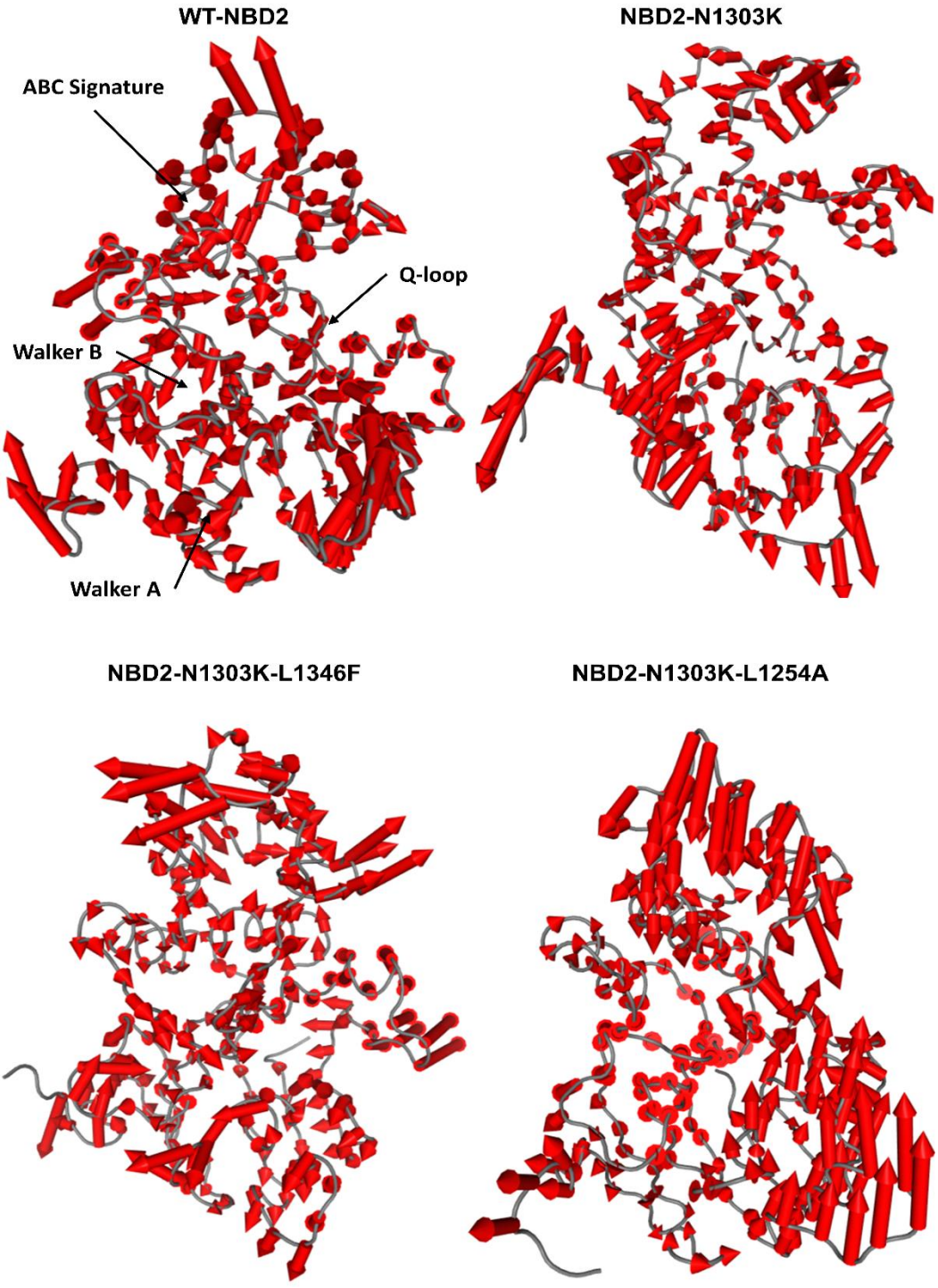


Figure 15. Porcupine plots depicting protein backbone fluctuation using mode 2 of normal mode analysis. Subdomains are labeled on wildtype NBD2 (5UAK: residues 1207 to 1436). All proteins are positioned similarly.

Inspecting mode two reveals wildtype NBD2 displays large movements of the flexible residues between the ABC signature subdomain and the Q-loop. In contrast, the N1303K mutant has larger movements in the Walker A subdomain. The L1346F mutant exhibits greater movements between ABC Signature and Q-loop. Finally, the L1254A mutant has similar movements in location and magnitude as wildtype mode two.

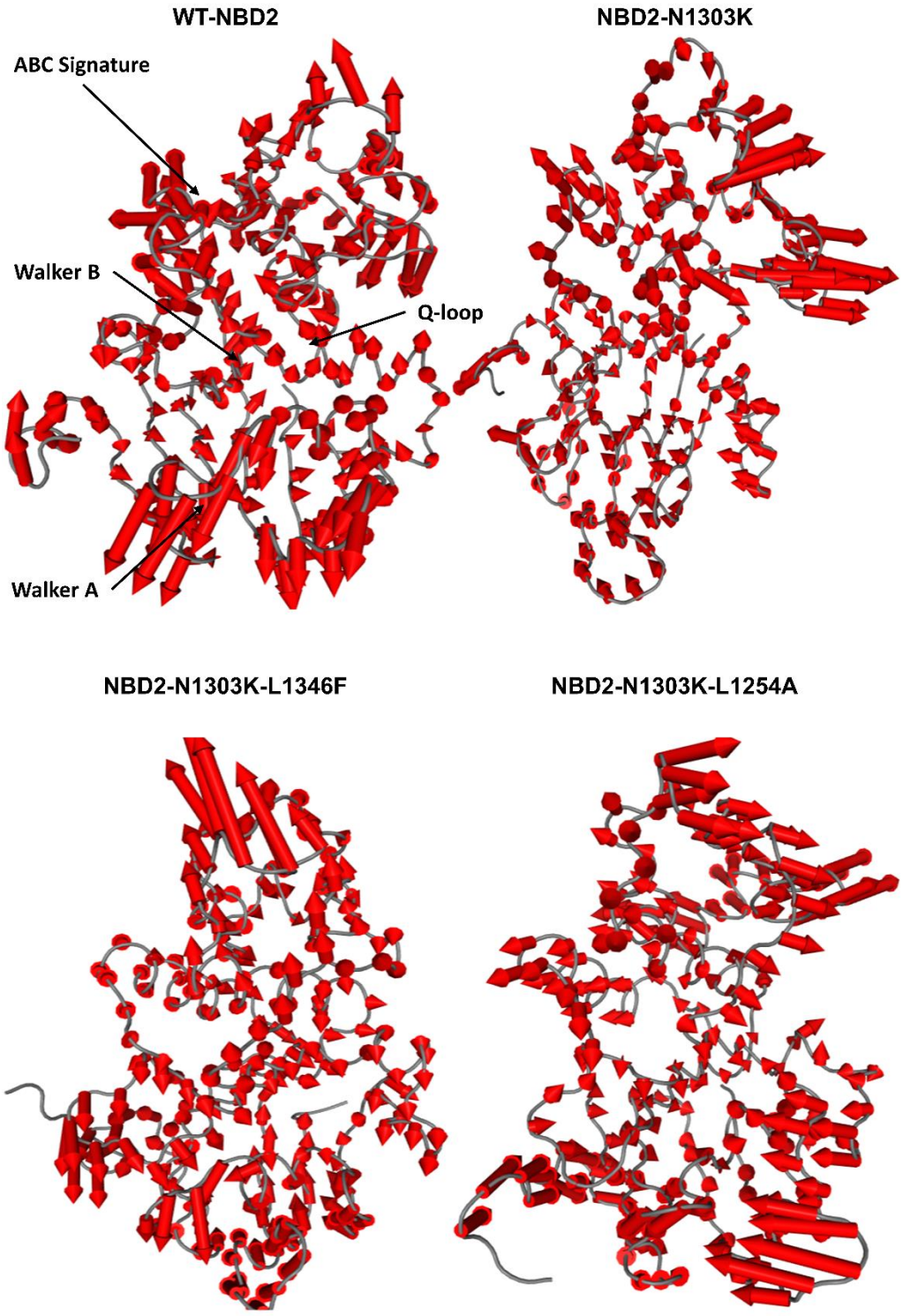


Figure 16. Porcupine plots depicting protein backbone fluctuation using mode 3 of normal mode analysis. Subdomains are labeled on wildtype NBD2 (5UAK: residues 1207 to 1436). All proteins are positioned similarly.

Inspecting mode 3, Wildtype NBD2 displayed prominent movements in the backbone in Walker A and near Q-loop. In contrast, N1303K had large movements in the backbone between the ABC signature and Q-loop. L1346F exhibited movements near Walker A and between ABC signature and Q-loop. L1254A has exhibited movements in the protein backbone between the ABC signature Q-loop and Walker A, comparable to wildtype NBD2.

N1303K has key differences compared to wildtype NBD2 in protein backbone movement as determined by NMA. In mode 1, wildtype NBD2 had more movement around the ABC signature and Walker A/B Subdomains, while N1303K had large movement only around the Walker A subdomain. In mode 2, N1303K has a large amount of movement around Walker A, while wildtype NBD2 had movement around the ABC signature and Q-loop. In mode 3, wildtype NBD2 had more movement around Walker A and Q-loop, while the largest movement in N1303K is between the ABC signature and Q-loop.

Comparing all three modes, N1303K has stark differences in protein backbone movement compared to wildtype. However, L1254A had very similar collective motions to wildtype. In all modes, L1254A exhibited similar fluctuation patterns as wildtype, usually completely opposite from N1303K. L1346F also exhibited closer movements to wildtype, but in some cases, it exhibited fluctuations apparent in both wildtype and N1303K. For example, L1346F had fluctuations in Walker A and ABC signature and Q-loop, and wildtype having large movements near Walker A, butN1303K had larger movements between ABC signature and Q-loop.

Cell Experiment Results

PCR Mutagenesis results. PCR mutagenesis using the Q5 kit and appropriate primers (Table A1) was performed on the GFP-N1303K-CFTR plasmid. Gel electrophoresis was performed on the PCR reactions and DNA bands of ~10,000 bp were observed for both L1254A and L1346F indicating successful plasmid amplification, given the template DNA was ~10,000 bp (Figure A7 and A8). The PCR product was transformed and then Miniprepped plasmid DNA was sent to Plasmidsaurus for nanopore sequencing and confirmation of mutation.

Sequencing results. Four clones were sent for sequencing and one individual clone each, L1254A-2 and L1346F-4, had the desired mutation in the N1303K background. The wildtype CFTR, DF508-CFTR and N1303K-CFTR were also sequenced to confirm identity (Supplemental Figures A9-A13).

Transfection results. Wildtype, ΔF508, N1303K, L1254A, and L1346F plasmids were transfected into HEK293 cells. The CFTR proteins expressed to varying degrees in cells (Supplemental Figure A14-A20) with the DF508-CFTR and GFP control having the highest expression. The DF508 cells had visible aggregation that increased from 24-48 hours and in some circumstances lead to cytotoxicity at 48 hours (Figure A17 & data not shown). In contrast, cells expressing wildtype CFTR had minimal aggregation except in high expressing cells, and both wildtype CFTR and N1303K-CFTR had a "medium" expression, compared to DF508. The CFTR double mutants L1254A and L1346F had low yet visible expression. Higher levels of expression of L1346F appeared to be cytotoxic at the 48 hours, while L1254A was not (data not shown).

In-gel fluorescence of CFTR mutant results.

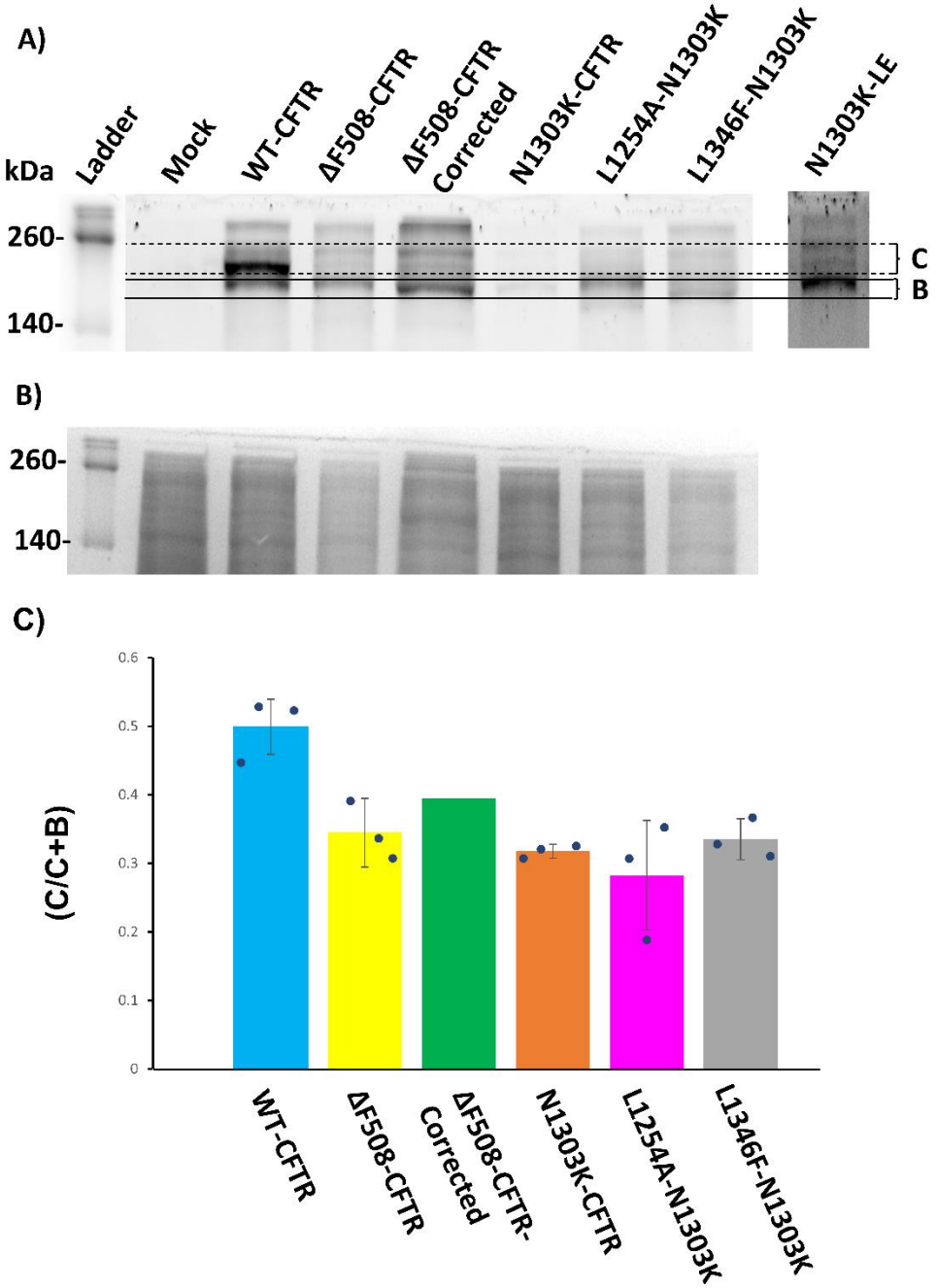


Figure 17. A) An example of In-gel fluorescence for wildtype and CFTR mutants. The location of C and B is denoted in the gel (boxes). LE stands for long exposure. B) A protein loading control via Coomassie blue. C) plot of C/B+C ratios for wildtype and CFTR mutants. Note: ΔF508 corrected (n=1) and all others (n=3).

Lysates were made from HEK293 cells expressing wildtype and mutant GFP-CFTR proteins. The GFP-tagged proteins were visualized using in-gel fluorescence (Figure 18 & A20-23). After fluorescence imaging, the gel was stained with Coomassie blue to visualize total protein loaded and act as a loading control. Wildtype GFP-CFTR was predominantly C-band (with a C/B ratio of 0.50 ± 0.05) compared to ΔF508 that was mostly B-band (with a C/B ratio of 0.34 ± 0.04, Figure 18B). The N1303K mutant had a similar C/B ratio of 0.318 ± 0.009 compared to ΔF508. The L1254A mutant ratio was slightly lower (0.28 ± 0.08), while the L1346F was slightly higher ratio (0.33 ± 0.03) compared to N1303K single mutant. As a control, cells expressing ΔF508 were incubated at 30°C for 12 hours to rescue the folding defect. Corrected ΔF508 had a C/B ratio of 0.40 ± 0.02 demonstrating the ability to detect correction of CFTR folding/trafficking using the in-gel assay (Figure 18B). Interestingly, the L1254A double mutant exhibited protein bands at ~70 and ~40 kDa not seen in the other mutants (Supplemental Figures A21-A23). In addition, the L1346F had a band just below the B-band that has been reported by other researchers for N1303K, called B* that represents post-lysis proteolysis of the C terminus and is a truncated version of B-band (He et al. 2021a). Only the double mutants appeared to have the B*-band with the L1346F mutant having a more prominent B* compared to L1254A (Supplemental Figures A21 to A23). In several experiments, there were larger species observed for wildtype and the mutants that are most likely soluble higher-order oligomers that appear to be expression level dependent.

Colocalization analysis of CFTR mutants.

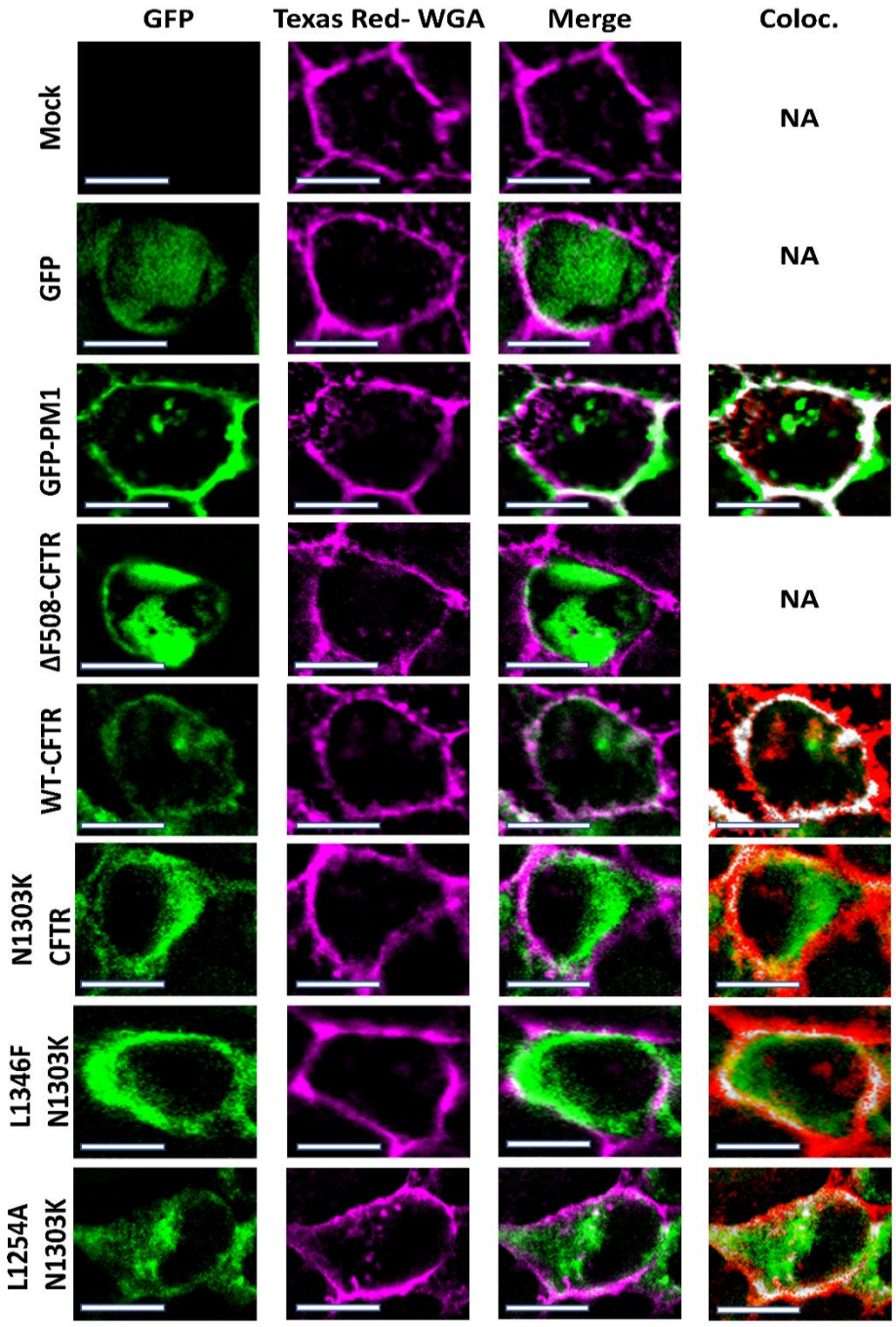


Figure 18. Colocalization of GFP-CFTR proteins with WGA-Texas Red in HEK293 cells. White color in Colocalization (Coloc.) Panels indicate GFP and Texas Red signals overlap. Scale bar is 10 µm. NA = not applicable, means no detectable colocalization measured. N=3 (45 cells) for all mutants except for ΔF508 and GFP-PM1 where N=2 (30 cells).

HEK293 cells expressing wildtype, or mutant CFTR proteins were fixed and co-stained with wheat germ agglutinin (WGA)-Texas Red to determine fraction of protein that resided in the plasma membrane (Figure 19). Monomeric enhanced GFP (GFP) and a plasma membrane attached form of GFP (GFP-PM1, (Ward et al. 2015)) were expressed as negative and positive controls for colocalization, respectively. As expected, GFP-PM1 resided in the plasma with WGA-Texas Red as determined by Mander's Colocalization Coefficient (median MCC = 0.72), while GFP did not (median MCC ~ 0, Figure 20). Wildtype CFTR partially colocalized with WGA-Texas Red with a median MCC value of 0.16. In contrast, ΔF508, N1303K, and L1346F had very little colocalization with WGA-Texas Red with median MCC values closer to zero. Interestingly, L1254A had a median MCC value approaching that of wildtype CFTR (median MCC = 0.12), suggesting enhanced trafficking to the plasma membrane compared to N1303K.

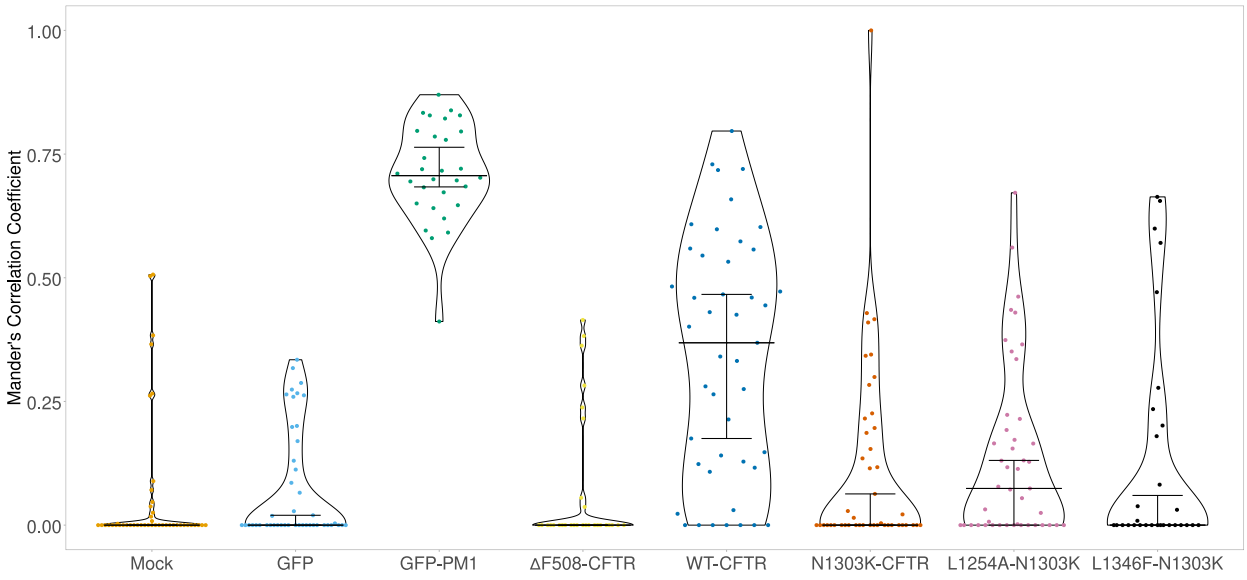


Figure 19. Colocalization analysis of GFP-CFTR proteins and controls. Violin plot created via Plots of Data (Postma & Goedhart, 2019). N=2 for GFP-PM1 and ΔF508 (30 cells) and N=3 (45 cells) for other conditions. The horizontal line indicates median value and vertical line is 95% confidence interval.

One-way ANOVA and post-hoc test indicate there is a significant difference between GFP-PM1 and the other mutants, and a significant difference between wildtype CFTR and the CFTR mutants. However, L1254A and L1346F are not significantly different from N1303K. Further experiments and increased sample size are needed to confirm these results.

DISCUSSION AND CONCLUSION

Computational Analysis

The MutPred2 algorithm predicted the L1254A would have a slight restorative effect and L1346F a major restorative effect on the N1303K-CFTR. Table 2 reiterates this as well, with the ΔΔG values from FoldX showing a more negative value for L1346F compared to L1254A. However, MD simulations and PCA support that L1254A may be more beneficial for stabilization than the L1346F mutation.

In Figure 4, N1303K was predicted to break a hydrogen bond in the mutation site upon mutation, but no new hydrogen bonds were predicted to form and the same pattern was seen for L1346F (Figure 5). However, L1254A breaks a hydrogen bond and forms a new hydrogen bond at positions LYS1250-GLU1271. These residues are between the Walker A and Q-loop subdomains, and this region is highly flexible. The ΔRMSF graph in Figure 9 shows that L1346F and L1254A are similar. However, L1254A is overall more similar to wildtype NBD2 than N1303K. The L1254A mutation is predicted to form an additional hydrogen bond, and this could lead to increased structural stability. A study by Simon and Csanády suggests that hydrogen bonds are needed to open and close CFTR efficiently. A hydrogen bond between Arginine 117 and glutamate 1124 only forms in the open state of CFTR and cannot form it when it is closed. Mutating the residue at 117 to a histidine (R117H) stops this hydrogen bond from occurring, and as such it significantly decreases the duration of the CFTR open state (Simon and Csanády 2021). With a deletion of a hydrogen bond having adverse reactions to CFTR opening, it could be inferred that formation of hydrogen bonds can have beneficial effects to mutant proteins (Figure 10). This region between Walker A and Q-loop is a flexible loop, so stabilization of this loop could help restore function. Further testing would need to be performed to ascertain whether the hydrogen bond in L1254A plays an important role in stabilization.

The MD data yields valuable insight to how mutations in NBD2 affect protein dynamics. In figure 7, the amount of hydrogen bonds, solvent accessible area, RMSD and radius of gyration are given for all

mutants over a span of 50 nanoseconds. The averages of all graphs are taken after the first 10 nanoseconds and reported in Figure 8. The L1254A mutant had an average SASA, RMSD, and radius of gyration closer to wildtype than N1303K. The average hydrogen bonds were the only caveat, with L1254A having less hydrogen bonds than both wildtype NBD2 and N1303K but having more than L1346F. N1303K conversely had similar characteristics to N1303K on the RMSD, SASA and radius of gyration. One thing to note is that L1346F had a very high RMSD compared to other mutants. The similarity between wildtype NBD2 and L1254A could possibly be an indicator that the L1254A mutation corrects the alterations imposed by N1303K and help achieve a protein movement closer to wildtype NBD2.

When observing the PCAs, N1303K behaves exactly opposite to wildtype NBD2. For example, in PCA 1 versus 2, wildtype NBD2 starts on 2,0 and ends on -2,2, while N1303K starts on -2, -2 and ends on 4,0. The overall distribution of wildtype NBD2 is in a horseshoe pattern, but N1303K is an upside-down horseshoe. The values on the PCA axes indicate variation from the median movement profile of each protein, so it is helpful to look at the PCA distribution instead of the values of the axes. With a completely different movement in the PCA, it is predicted that N1303K has dissimilar protein backbone movements to wildtype, which could partly explain why N1303K is pathogenic (interpretation of the PCAs based on MDM-Task (Amamuddy et al. 2021)). When observing the double mutants, both L1254A and L1346F had more similar PCA movements to wildtype NBD2 than N1303K. The only caveat to this is that on PCA 2 versus 3, L1346F has a more similar distribution to N1303K. It is interpreted that the similarity in PCAs between double mutants and wildtype could indicate a restorative effect on protein backbone movement, which could reduce pathogenicity. However, since L1346F has one PCA more similar to N1303K unlike L1254A, it is predicted that L1254A has a greater restorative effect. To corroborate the PCA study, a study by Orellana *et al.* has shown that if the PCA of a mutant protein is more similar to wildtype, it also exhibits the same characteristics, such as similar folding patterns and similar levels of activity. This was performed over a wide array of proteins sourced from the PDB (Orellana et al., 2016). Upon comparative PCA, it was found that L1254A is the most similar in terms of

overall protein movement to wildtype NBD2, as the two PCAs overlap slightly (interpretation based on MDM-Task (Amamuddy et al., 2021)).

Normal mode analysis also revealed similarities between wildtype NBD2 and L1254A on molecular dynamics simulations and PCA. NMA is used to assess the number of flexible states of a protein backbone around an equilibrium point, and this amount of movement availability was compared between each mutant. It is predicted that the more movement, the less stable the protein is. Much like in the PCA, N1303K behaves opposite to wildtype NBD2 (Bauer et al., 2019). An example of this is seen in mode 1 where increased movement between the ABC signature and Q loop, with some slight movement near Walker A are observed for wildtype NBD2 (Figure 15). N1303K however only has large movements near Walker A. This trend is seen across all three modes. However, the double mutants have more similar protein backbone movement to wildtype NBD2. Comparing all three modes, L1254A-NBD2 has more similarity to wildtype NBD2 than N1303K. L1346F also exhibits this trend as well, but on mode 3 it had fluctuations in Walker A and between ABC signature and Q-loop. In contrast, wildtype had large movements near Walker A and N1303K had larger movements between ABC signature and Q-loop. The overall similarity between wildtype and L1254A once again indicates that this mutation could have a beneficial effect. The most flexible regions shown across all modes are between the ABC Signature and Q-loop. Given L1254A has a closer resemblance to wildtype than N1303K, it is predicted that the Mutated CFTR should be able to open more efficiently than with the N1303K mutation alone, but this has yet to be tested.

Cellular analysis of CFTR Mutations

The cell data concurs with the computational results, with some exceptions. wildtype CFTR had the highest C-B band ratio, with N1303K and ΔF508 being significantly lower (Figure 18). The double mutant L1346F had a slightly higher C-B band ratio than N1303K. L1346F was slightly higher, this suggests that this mutation has a restorative effect that is also predicted in the simulations. However, L1254A had a lower C-B band ratio than N1303K. But there is also degradation and cleavage of the

protein with bands at ~40 and 70 kDa in L1254A lysates (see supplemental Figures A21-A23). This degradation could obscure the presence of C-band and explain the differences between the in-gel fluorescent results and the colocalization analysis. Other studies have shown that CFTR placed in a non-denaturing lysate (such as for in-gel fluorescence) is prone to cleavage in the C-terminus of NBD2 (He et al. 2021). The truncated CFTR appears as a separate band in the gel below band B and can be seen in the double mutants (He et al. 2021). In future experiments, the lysis buffer used to create cell lysates could have increased amounts of proteosome/lysosome inhibitors to prevent this degradation. Addition of these inhibitors to transfection reactions to intact cells before lysis may also minimize this degradation allowing more accurate quantification. In addition, cycloheximide chase experiments could be performed to compare the stability of B and C band of N1303K to the double mutants. The ANOVA analysis for the C-B ratio indicates that the difference between L1254A and L1346F is not significant compared to N1303K, while N1303K is significantly different than wildtype CFTR. It is recommended to have more replicates of the experiment to ascertain if the double mutants provide a statistically different C-B band ratio.

Colocalization analysis revealed that L1254A has possible restorative effects on CFTR. Figure 19 and 20 indicate that L1254A protein has increased presence in the cell membrane of transfected cells compared to N1303K, or L1346F. Wildtype CFTR was more present in the cell membrane compared to ΔF508 having almost no protein presence, which is in agreement with past publications (Varga et al. 2008). N1303K had slightly more CFTR in the cell membrane than ΔF508-CFTR, but less compared to L1254A. In contrast, L1346F did not traffic to the plasma membrane and the vast majority of CFTR was internal. The increased amount of L1254A in the cell membrane suggests increased protein trafficking, which could indicate a positive effect on folding, in agreement with our computational results. L1254A may be overcoming ER associated degradation (ERAD) and getting to the cell membrane, but then being recognized by peripheral quality control pathways based on the degradation seen in the protein gels and the lower C-B ratio (Fukuda & Okiyoneda, 2020; Sharma et al., 2004). A study by Fukuda and Okiyoneda confirmed that some amounts of ΔF508-CFTR can pass through the ER without being degraded and traffic to the cell membrane. This may be happening with the L1254A mutant, but further

testing is needed to confirm this hypothesis (Fukuda & Okiyoneda, 2020). The L1346F mutant has a high predicted RMSD, and this could indicate a less tightly folded domain, thus making the mutant trapped in the ER for degradation (Wieczorek and Zielenkiewicz 2008). Further testing is needed to determine the exact internal organelle the L1346F mutant resides (e.g., ER, Golgi, etc.). ANOVA analysis indicates that L1254A is not significantly different than N1303K but is significantly different than wildtype CFTR. A larger data set may be needed to see if there is a significant difference in L1254A colocalization.

Analysis benefits and limitations

The computational results presented contribute new information to the CF research field in several ways. First, N1303K has not been well studied in terms of protein dynamic simulations. While NBD2 is modeled in some studies, it is usually in regard to ΔF508 instead of N1303K (Odera et al., 2018). N1303K and two novel mutations, L1254A and L1346F, have been characterized using all atom protein dynamics simulations. The simulation data and PCA did in fact present L1254A as closer to wildtype CFTR than N1303K, which could mean it has beneficial effects to folding; this was supported by cellular colocalization studies (Vernon et al. 2017). Future CFTR studies could use the combination of all-atom molecular dynamic simulations in conjunction with colocalization analysis to ascertain how other mutations predicted by Ivey and Youker affect CFTR folding.

While this thesis does provide beneficial information, there are drawbacks. The all-atom molecular dynamic simulations were performed under neutral pH and with only NaCl salts present. However, in the human body this is not always the case; there are many salts that CFTR comes into contact with, such as carbonate and chloride, and this may alter folding. Also, pH differs throughout the body; the lungs will be more basic than the extremities, where the pH is more acidic (Hopkins et al. 2023). The simulations may be not as accurate due to these differences in human anatomy. Also, other studies have performed all-atom simulations on full-length CFTR, while this study focused on solely NBD2. These mutations may behave very differently in full-length CFTR (Odera et al. 2018). MutPred2 analysis also did not show the full picture of the secondary mutations; L1254A had positive effects on

CFTR shown by PCA and molecular dynamic simulations instead of deleterious as predicted by MutPred2. This is unlike other studies, which stated that MutPred2 had a high degree of accuracy to real mutants (Pejaver et al., 2020). Another key limitation to this thesis is the use of HEK293 cells, as well as transient transfections. HEK293 cells tend to overexpress protein in the cell membrane over time, and the flow of ions is significantly lower as compared to primary lines (Gibson et al. 2013). Transient transfection also has a wide array of protein expression across cells, so a stable cell line would be beneficial in making sure cell to cell variation is as low as it can be (Gibson et al. 2013). Performing this study in primary cell lines would be beneficial to gain a more accurate depiction of colocalization and C/B ratio (Domingue et al. 2014a). In addition, performing similar experiments in three dimensional organoids would also provide greater information.

Future Work

Previous studies have performed all atom simulation on CFTR in a plasma membrane instead of just NBD2, so it would be beneficial to ascertain if the L1254A mutation has long-range effects on other domains in the full-length CFTR *in silico* (Odera et al., 2018). In the future, more gels should be run to obtain higher statistical confidence; the same can be said for the colocalization analysis. The L1254A mutant appears to traffic to the cell membrane, it is unclear if the CFTR is functional. An activity assay of CFTR should be performed to see if the CFTR is functional (Ramalho et al., 2022). Proteosome inhibitors and lysosome inhibitors should be used on L1254A expressing cells to determine when truncation of the protein is occurring (pre- vs. post-lysis). A new lysis buffer with a more extensive array of anti-proteolytic reagents could be used in the future to help protect the double mutant CFTR from being degraded. In the future, computational and wet-lab methods could be used to characterize additional mutations reported by the Youker lab to potentially stabilize N1303K (Ivey & Youker, 2020). Finally, a cycloheximide study could be used to ascertain the half-life of L1254A B band, which could more precisely measure the stability of the CFTR mutant (Cebotaru et al. 2008).

RFERENCES

Addgene: Vector Database - pEGFP-C1. [accessed 2023 Apr 15]. https://www.addgene.org/vector-database/2487/.

Aier I, Varadwaj PK, Raj U. 2016. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. Sci Rep. 6(1):34984. doi:10.1038/srep34984.

Aleksandrov AA, Kota P, Aleksandrov LA, He L, Jensen T, Cui L, Gentzsch M, Dokholyan NV, Riordan JR. 2010. Regulatory Insertion Removal Restores Maturation, Stability and Function of ΔF508 CFTR. J Mol Biol. 401(2):194–210. doi:10.1016/j.jmb.2010.06.019.

Allan KM, Astore MA, Fawcett LK, Wong SL, Chen P-C, Griffith R, Jaffe A, Kuyucak S, Waters SA. 2022. S945L-CFTR molecular dynamics, functional characterization and tezacaftor/ivacaftor efficacy in vivo and in vitro in matched pediatric patient-derived cell models. Front Pediatr. 10:1062766. doi:10.3389/fped.2022.1062766.

Allen Lucy, Allen Lorna, Carr SB, Davies G, Downey D, Egan M, Forton JT, Gray R, Haworth C, Horsley A, et al. 2023. Future therapies for cystic fibrosis. Nat Commun. 14:693. doi:10.1038/s41467-023-36244-2.

Amamuddy OS, Glenister M, Bishop ÖT. 2021. MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics. :2021.01.29.428734. doi:10.1101/2021.01.29.428734. [accessed 2023 Apr 14]. https://www.biorxiv.org/content/10.1101/2021.01.29.428734v2.

Amico G, Brandas C, Moran O, Baroni D. 2019. Unravelling the Regions of Mutant F508del-CFTR More Susceptible to the Action of Four Cystic Fibrosis Correctors. Int J Mol Sci. 20(21):5463. doi:10.3390/ijms20215463.

A. Proctor E, Kota P, A. Aleksandrov A, He L, R. Riordan J, V. Dokholyan N. 2015. Rational coupled dynamics network manipulation rescues disease-relevant mutant cystic fibrosis transmembrane conductance regulator. Chem Sci. 6(2):1237–1246. doi:10.1039/C4SC01320D.

Awatade NT, Wong SL, Hewson CK, Fawcett LK, Kicic A, Jaffe A, Waters SA. 2018. Human Primary Epithelial Cell Models: Promising Tools in the Era of Cystic Fibrosis Personalized Medicine. Front Pharmacol. 9. [accessed 2023 May 2]. https://www.frontiersin.org/articles/10.3389/fphar.2018.01429.

Bahia MS, Khazanov N, Zhou Q, Yang Z, Wang C, Hong JS, Rab A, Sorscher EJ, Brouillette CG, Hunt JF, et al. 2021. Stability Prediction for Mutations in the Cytosolic Domains of Cystic Fibrosis Transmembrane Conductance Regulator. J Chem Inf Model. 61(4):1762–1777. doi:10.1021/acs.jcim.0c01207.

Bank RPD. RCSB PDB - 5UAK: Dephosphorylated, ATP-free human cystic fibrosis transmembrane conductance regulator (CFTR). [accessed 2023 Apr 25]. https://www.rcsb.org/structure/5uak.

Bauer JA, Pavlović J, Bauerová-Hlinková V. 2019b. Normal Mode Analysis as a Routine Part of a Structural Investigation. Molecules. 24(18):3293. doi:10.3390/molecules24183293.

Bekker H, Berendsen H, Dijkstra E, Achterop S, Vondrumen R, Vanderspoel D, Sijbers A, Keegstra H, Renardus M. 1993. GROMACS- A Parallel Computer for Molecular-Dynamics Simualtions: 4th International Conference on Computational Physics (PC 92). DeGroot R, Nadrchal J, editors. Phys Comput 92.:252–256.

Brown SD, White R, Tobin P. 2017. Keep them breathing: Cystic fibrosis pathophysiology, diagnosis, and treatment. JAAPA Off J Am Acad Physician Assist. 30(5):23–27. doi:10.1097/01.JAA.0000515540.36581.92.

Callebaut I, Chong PA, Forman-Kay JD. 2018. CFTR structure. J Cyst Fibros. 17(2):S5–S8. doi:10.1016/j.jcf.2017.08.008.

Castellani C, CFTR2 team. 2013. CFTR2: How will it help care? Paediatr Respir Rev. 14 Suppl 1:2–5. doi:10.1016/j.prrv.2013.01.006.

Cebotaru L, Vij N, Ciobanu I, Wright J, Flotte T, Guggino WB. 2008. Cystic Fibrosis Transmembrane Regulator Missing the First Four Transmembrane Segments Increases Wild Type and ΔF508 Processing. J Biol Chem. 283(32):21926–21933. doi:10.1074/jbc.M709156200.

CFTR - Cystic fibrosis transmembrane conductance regulator - Homo sapiens (Human) | UniProtKB | UniProt. [accessed 2023 May 18]. https://www.uniprot.org/uniprotkb/P13569/entry.

Custom DNA Oligos | IDT. Integr DNA Technol. [accessed 2023 Apr 15]. https://www.idtdna.com/pages/products/custom-dna-rna/dna-oligos/custom-dna-oligos.

Cystic fibrosis - Symptoms and causes. Mayo Clin. [accessed 2023 Apr 21]. https://www.mayoclinic.org/diseases-conditions/cystic-fibrosis/symptoms-causes/syc-20353700.

Davies JC, Alton EWFW, Bush A. 2007. Cystic fibrosis. BMJ. 335(7632):1255–1259. doi:10.1136/bmj.39391.713229.AD.

Davies JC, Ebdon A-M, Orchard C. 2014. Recent advances in the management of cystic fibrosis. Arch Dis Child. 99(11):1033–1036. doi:10.1136/archdischild-2013-304400.

DeWeerdt S. 2020. Cystic fibrosis drugs target the malformed proteins at the root of the disease. Nature. 583(7818):S2–S4. doi:10.1038/d41586-020-02106-w.

Domingue JC, Ao M, Sarathy J, George A, Alrefai WA, Nelson DJ, Rao MC. 2014b. HEK-293 cells expressing the cystic fibrosis transmembrane conductance regulator (CFTR): a model for studying regulation of Cl− transport. Physiol Rep. 2(9):e12158. doi:10.14814/phy2.12158.

Dunn KW, Kamocka MM, McDonald JH. 2011. A practical guide to evaluating colocalization in biological microscopy. Am J Physiol - Cell Physiol. 300(4):C723–C742. doi:10.1152/ajpcell.00462.2010.

Durham E, Dorr B, Woetzel N, Staritzbichler R, Meiler J. 2009. Solvent accessible surface area approximations for rapid and accurate protein structure prediction. J Mol Model. 15(9):1093–1108. doi:10.1007/s00894-009-0454-9.

Ensinck MM, De Keersmaecker L, Ramalho AS, Cuyx S, Van Biervliet S, Dupont L, Christ F, Debyser Z, Vermeulen F, Carlon MS. 2022. Novel CFTR modulator combinations maximise rescue of G85E and N1303K in rectal organoids. ERJ Open Res. 8(2):00716–02021. doi:10.1183/23120541.00716-2021.

Estabrooks S, Brodsky JL. 2020. Regulation of CFTR Biogenesis by the Proteostatic Network and Pharmacological Modulators. Int J Mol Sci. 21(2):452. doi:10.3390/ijms21020452.

Remington 2014. Expression, purification and characterisation of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) in Saccharomyces cerevisiae.

Farrell P, Férec C, Macek M, Frischer T, Renner S, Riss K, Barton D, Repetto T, Tzetis M, Giteau K, et al. 2018. Estimating the age of p.(Phe508del) with family studies of geographically distinct European populations and the early spread of cystic fibrosis. Eur J Hum Genet. 26(12):1832–1839. doi:10.1038/s41431-018-0234-z.

Fukuda R, Okiyoneda T. 2020b. Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Ubiquitylation as a Novel Pharmaceutical Target for Cystic Fibrosis. Pharmaceuticals. 13(4):75. doi:10.3390/ph13040075.

Gershenson A, Gosavi S, Faccioli P, Wintrode PL. 2020. Successes and challenges in simulating the folding of large proteins. J Biol Chem. 295(1):15–33. doi:10.1074/jbc.REV119.006794.

Gibson TJ, Seiler M, Veitia RA. 2013. The transience of transient overexpression. Nat Methods. 10(8):715–721. doi:10.1038/nmeth.2534.

He L, Kennedy AS, Houck S, Aleksandrov A, Quinney NL, Cyr-Scully A, Cholon DM, Gentzsch M, Randell SH, Ren HY, et al. 2021b. DNAJB12 and Hsp70 triage arrested intermediates of N1303K-CFTR for endoplasmic reticulum-associated autophagy. Mol Biol Cell. 32(7):538–553. doi:10.1091/mbc.E20-11-0688.

Hopkins E, Sanvictores T, Sharma S. 2023. Physiology, Acid Base Balance. In: StatPearls. Treasure Island (FL): StatPearls Publishing. [accessed 2023 Jun 8]. http://www.ncbi.nlm.nih.gov/books/NBK507807/.

Ivey G, Youker RT. 2020. Disease-relevant mutations alter amino acid co-evolution networks in the second nucleotide binding domain of CFTR. PLOS ONE. 15(1):e0227668. doi:10.1371/journal.pone.0227668.

Jiang F, Doudna J. 2017. CRISPR-Cas9 Structures and Mechanisms. Annu Rev Biophys. 46. doi:10.1146/annurev-biophys-062215-010822.

Klein H, Abu-Arish A, Trinh NTN, Luo Y, Wiseman PW, Hanrahan JW, Brochiero E, Sauvé R. 2016. Investigating CFTR and KCa3.1 Protein/Protein Interactions. PLOS ONE. 11(4):e0153665. doi:10.1371/journal.pone.0153665.

Ko W, Porter JJ, Sipple MT, Edwards KM, Lueck JD. 2022. Efficient suppression of endogenous CFTR nonsense mutations using anticodon-engineered transfer RNAs. Mol Ther - Nucleic Acids. 28:685–701. doi:10.1016/j.omtn.2022.04.033.

Konstan MW, VanDevanter DR, Rowe SM, Wilschanski M, Kerem E, Sermet-Gaudelus I, DiMango E, Melotti P, McIntosh J, De Boeck K. 2020. Efficacy and safety of ataluren in patients with nonsense-mutation cystic fibrosis not receiving chronic inhaled aminoglycosides: The international, randomized, double-blind, placebo-controlled Ataluren Confirmatory Trial in Cystic Fibrosis (ACT CF). J Cyst Fibros Off J Eur Cyst Fibros Soc. 19(4):595–601. doi:10.1016/j.jcf.2020.01.007.

Land H, Humble MS. 2018. YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. Methods Mol Biol Clifton NJ. 1685:43–67. doi:10.1007/978-1-4939-7366-8_4.

Ledford H. 2018 Jul 16. CRISPR gene editing produces unwanted DNA deletions. Nature. doi:10.1038/d41586-018-05736-3. [accessed 2023 Jun 22]. https://www.nature.com/articles/d41586-018-05736-3.

Lee J-A, Cho A, Huang EN, Xu Y, Quach H, Hu J, Wong AP. 2021. Gene therapy for cystic fibrosis: new tools for precision medicine. J Transl Med. 19(1):452. doi:10.1186/s12967-021-03099-4.

Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. 2017. Molecular Structure of the Human CFTR Ion Channel. Cell. 169(1):85-95.e8. doi:10.1016/j.cell.2017.02.024.

Locher KP. 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol. 23(6):487–493. doi:10.1038/nsmb.3216.

Maisuradze GG, Liwo A, Scheraga HA. 2009. Principal component analysis for protein folding dynamics. J Mol Biol. 385(1):312–329. doi:10.1016/j.jmb.2008.10.018.

Martínez L. 2015. Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. PLoS ONE. 10(3):e0119264. doi:10.1371/journal.pone.0119264.

Miller AC, Comellas AP, Hornick DB, Stoltz DA, Cavanaugh JE, Gerke AK, Welsh MJ, Zabner J, Polgreen PM. 2020. Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. Proc Natl Acad Sci. 117(3):1621–1627. doi:10.1073/pnas.1914912117.

Moyer BD, Loffing J, Schwiebert EM, Loffing-Cueni D, Halpin PA, Karlson KH, Ismailov II, Guggino WB, Langford GM, Stanton BA. 1998. Membrane trafficking of the cystic fibrosis gene product, cystic fibrosis transmembrane conductance regulator, tagged with green fluorescent protein in madin-darby canine kidney cells. J Biol Chem. 273(34):21759–21768. doi:10.1074/jbc.273.34.21759.

NEBaseChanger. [accessed 2023 Apr 15]. https://nebasechanger.neb.com/.

Odera M, Furuta T, Sohma Y, Sakurai M. 2018a. Molecular dynamics simulation study on the structural instability of the most common cystic fibrosis-associated mutant ΔF508-CFTR. Biophys Physicobiology. 15:33–44. doi:10.2142/biophysico.15.0_33.

Oligo Entry. [accessed 2023 Apr 15]. https://www.idtdna.com/site/order/oligoentry.

Orellana L, Yoluk O, Carrillo O, Orozco M, Lindahl E. 2016. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. Nat Commun. 7(1):12575. doi:10.1038/ncomms12575.

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

Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, Mort M, Cooper DN, Sebat J, Iakoucheva LM, et al. 2020. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 11(1):5918. doi:10.1038/s41467-020-19669-x.

Petrova N, Balinova N, Marakhonov A, Vasilyeva T, Kashirskaya N, Galkina V, Ginter E, Kutsev S, Zinchenko R. 2021. Ethnic Differences in the Frequency of CFTR Gene Mutations in Populations of the European and North Caucasian Part of the Russian Federation. Front Genet. 12. [accessed 2023 Apr 24]. https://www.frontiersin.org/articles/10.3389/fgene.2021.678374.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 25(13):1605–1612. doi:10.1002/jcc.20084.

Philadelphia TCH of. 2014 Aug 23. Cystic Fibrosis. [accessed 2023 May 16]. https://www.chop.edu/conditions-diseases/cystic-fibrosis.

Phuan P-W, Son J-H, Tan J-A, Li C, Musante I, Zlock L, Nielson DW, Finkbeiner WE, Kurth MJ, Galietta LJ, et al. 2018. COMBINATION POTENTIATOR ('CO-POTENTIATOR') THERAPY FOR CF CAUSED BY CFTR MUTANTS, INCLUDING N1303K, THAT ARE POORLY RESPONSIVE TO SINGLE POTENTIATORS. J Cyst Fibros Off J Eur Cyst Fibros Soc. 17(5):595–606. doi:10.1016/j.jcf.2018.05.010.

PlasmidSaurus Sequencing. [accessed 2023 Apr 15]. https://www.plasmidsaurus.com/index/.

Postma M, Goedhart J. 2019. PlotsOfData—A web app for visualizing data together with their summaries. PLOS Biol. 17(3):e3000202. doi:10.1371/journal.pbio.3000202.

Prins S, Corradi V, Sheppard DN, Tieleman DP, Vergani P. 2022. Can two wrongs make a right? F508del-CFTR ion channel rescue by second-site mutations in its transmembrane domains. J Biol Chem. 298(3). doi:10.1016/j.jbc.2022.101615. [accessed 2023 Apr 25]. https://www.jbc.org/article/S0021-9258(22)00055-2/abstract.

R F. Is Phoenicia the Origin of the N1303K CFTR Mutation? doi:10.23937/ijrdd-2017/1710006. [accessed 2023 May 26]. https://www.clinmedjournals.org/articles/ijrdd/international-journal-of-rare-diseases-and-disorders-ijrdd-2-006.php?jid=ijrdd.

Raëd F, Marie Claude P, Sandra C, André M, Alain K, Véronique L. 2019. Is Phoenicia the Origin of the N1303K CFTR Mutation? Int J Rare Dis Disord. 2(1). doi:10.23937/ijrdd-2017/1710006. [accessed 2023 Apr 24]. https://www.clinmedjournals.org/articles/ijrdd/international-journal-of-rare-diseases-and-disorders-ijrdd-2-006.php?jid=ijrdd.

Ramalho AS, Boon M, Proesmans M, Vermeulen F, Carlon MS, De Boeck K. 2022. Assays of CFTR Function In Vitro, Ex Vivo and In Vivo. Int J Mol Sci. 23(3):1437. doi:10.3390/ijms23031437.

Rapino D, Sabirzhanova I, Lopes-Pacheco M, Grover R, Guggino WB, Cebotaru L. 2015. Rescue of NBD2 Mutants N1303K and S1235R of CFTR by Small-Molecule Correctors and Transcomplementation. PLOS ONE. 10(3):e0119796. doi:10.1371/journal.pone.0119796.

Ridley K, Condren M. 2020. Elexacaftor-Tezacaftor-Ivacaftor: The First Triple-Combination Cystic Fibrosis Transmembrane Conductance Regulator Modulating Therapy. J Pediatr Pharmacol Ther JPPT Off J PPAG. 25(3):192–197. doi:10.5863/1551-6776-25.3.192.

Rosen BH, Chanson M, Gawenis LR, Liu J, Sofoluwe A, Zoso A, Engelhardt JF. 2018. Animal and Model Systems for Studying Cystic Fibrosis. J Cyst Fibros Off J Eur Cyst Fibros Soc. 17(2 Suppl):S28–S34. doi:10.1016/j.jcf.2017.09.001.

Rowe SM, Miller S, Sorscher EJ. 2005. Cystic fibrosis. N Engl J Med. 352(19):1992–2001. doi:10.1056/NEJMra043184.

Rusnati M, D'Ursi P, Pedemonte N, Urbinati C, Ford RC, Cichero E, Uggeri M, Orro A, Fossa P. 2020. Recent Strategic Advances in CFTR Drug Discovery: An Overview. Int J Mol Sci. 21(7):2407. doi:10.3390/ijms21072407.

Saint-Criq V, Gray MA. 2017. Role of CFTR in epithelial physiology. Cell Mol Life Sci CMLS. 74(1):93–115. doi:10.1007/s00018-016-2391-y.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods. 9(7):676–682. doi:10.1038/nmeth.2019.

Scholefield J, Harrison PT. 2021. Prime editing – an update on the field. Gene Ther. 28(7):396–401. doi:10.1038/s41434-021-00263-9.

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. Nucleic Acids Res. 33(suppl_2):W382–W388. doi:10.1093/nar/gki387.

Scotet V, L'Hostis C, Férec C. 2020. The Changing Epidemiology of Cystic Fibrosis: Incidence, Survival and Impact of the CFTR Gene Discovery. Genes. 11(6):589. doi:10.3390/genes11060589.

Sharma M, Pampinella F, Nemes C, Benharouga M, So J, Du K, Bache KG, Papsin B, Zerangue N, Stenmark H, et al. 2004. Misfolding diverts CFTR from recycling to degradation : quality control at early endosomes. J Cell Biol. 164(6):923–933. doi:10.1083/jcb.200312018.

Simon MA, Csanády L. 2021. Molecular pathology of the R117H cystic fibrosis mutation is explained by loss of a hydrogen bond. Jara-Oseguera A, Aldrich RW, Jara-Oseguera A, Hwang T-C, editors. eLife. 10:e74693. doi:10.7554/eLife.74693.

SYBR™ Safe DNA Gel Stain. [accessed 2023 Apr 15]. https://www.thermofisher.com/order/catalog/product/S33102.

Valley HC, Bukis KM, Bell A, Cheng Y, Wong E, Jordan NJ, Allaire NE, Sivachenko A, Liang F, Bihler H, et al. 2019. Isogenic cell models of cystic fibrosis-causing variants in natively expressing pulmonary epithelial cells. J Cyst Fibros Off J Eur Cyst Fibros Soc. 18(4):476–483. doi:10.1016/j.jcf.2018.12.001.

Varga K, Goldstein R, Jurkuvenaite A, Chen L, Matalon S, Sorscher E, Bebok Z, Collawn J. 2008. Enhanced Cell Surface Stability of Rescued ΔF508 Cystic Fibrosis Transmembrane Conductance Regulator by Pharmacological Chaperones. Biochem J. 410(3):555–564. doi:10.1042/BJ20071420.

Veit G, Avramescu RG, Chiang AN, Houck SA, Cai Z, Peters KW, Hong JS, Pollard HB, Guggino WB, Balch WE, et al. 2016. From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. Mol Biol Cell. 27(3):424–433. doi:10.1091/mbc.E14-04-0935.

Vergani P, Lockless SW, Nairn AC, Gadsby DC. 2005. CFTR channel opening by ATP-driven tight dimerization of its nucleotide-binding domains. Nature. 433(7028):876–880. doi:10.1038/nature03313.

Vernon RM, Chong PA, Lin H, Yang Z, Zhou Q, Aleksandrov AA, Dawson JE, Riordan JR, Brouillette CG, Thibodeau PH, et al. 2017. Stabilization of a nucleotide-binding domain of the cystic fibrosis transmembrane conductance regulator yields insight into disease-causing mutations. J Biol Chem. 292(34):14147–14164. doi:10.1074/jbc.M116.772335.

Ward RJ, Pediani JD, Godin AG, Milligan G. 2015. Regulation of Oligomeric Organization of the Serotonin 5-Hydroxytryptamine 2C (5-HT2C) Receptor Observed by Spatial Intensity Distribution Analysis. J Biol Chem. 290(20):12844–12857. doi:10.1074/jbc.M115.644724.

Wieczorek G, Zielenkiewicz P. 2008. ΔF508 mutation increases conformational flexibility of CFTR protein. J Cyst Fibros. 7(4):295–300. doi:10.1016/j.jcf.2007.11.008.

Yang Z, Hildebrandt E, Jiang F, Aleksandrov A, Khazanov N, Zhou Q, An J, Mezzell AT, Xavier BM, Ding H, et al. 2018. Structural stability of purified human CFTR is systematically improved by mutations

in nucleotide binding domain 1. Biochim Biophys Acta Biomembr. 1860(5):1193–1204. doi:10.1016/j.bbamem.2018.02.006.

Zaher A, ElSaygh J, Elsori D, ElSaygh H, Sanni A. 2021 Jul 3. A Review of Trikafta: Triple Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Modulator Therapy. Cureus. doi:10.7759/cureus.16144. [accessed 2023 Apr 24]. https://www.cureus.com/articles/58184-a-review-of-trikafta-triple-cystic-fibrosis-transmembrane-conductance-regulator-cftr-modulator-therapy.

Zhenin M, Noy E, Senderowitz H. 2015. REMD Simulations Reveal the Dynamic Profile and Mechanism of Action of Deleterious, Rescuing, and Stabilizing Perturbations to NBD1 from CFTR. J Chem Inf Model. 55(11):2349–2364. doi:10.1021/acs.jcim.5b00312.

APPENDIX

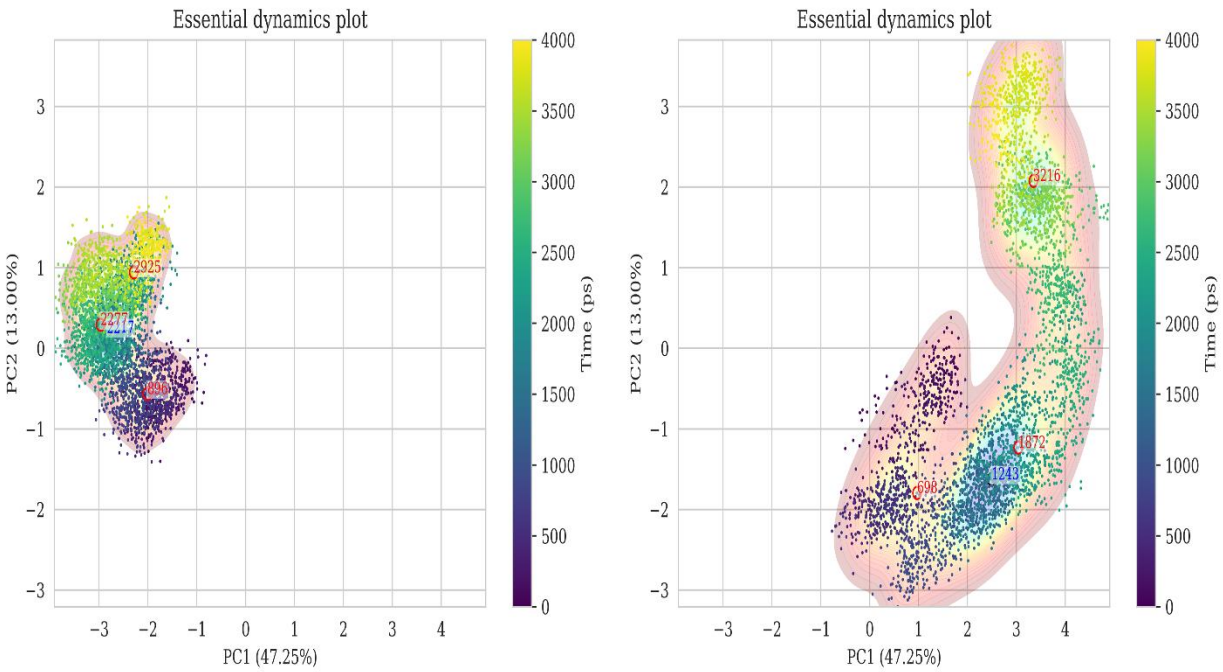


Figure A 1. Comparative essential dynamics PCA between wildtype NBD2 (left) and N1303K (right). Red circles represent K-means centroids and blue circles represent conformations with the lowest energy. The contour plot corresponds to the kernel densities of the first two PCAs. Contour plot also is comparable to the free energy of the protein surface.

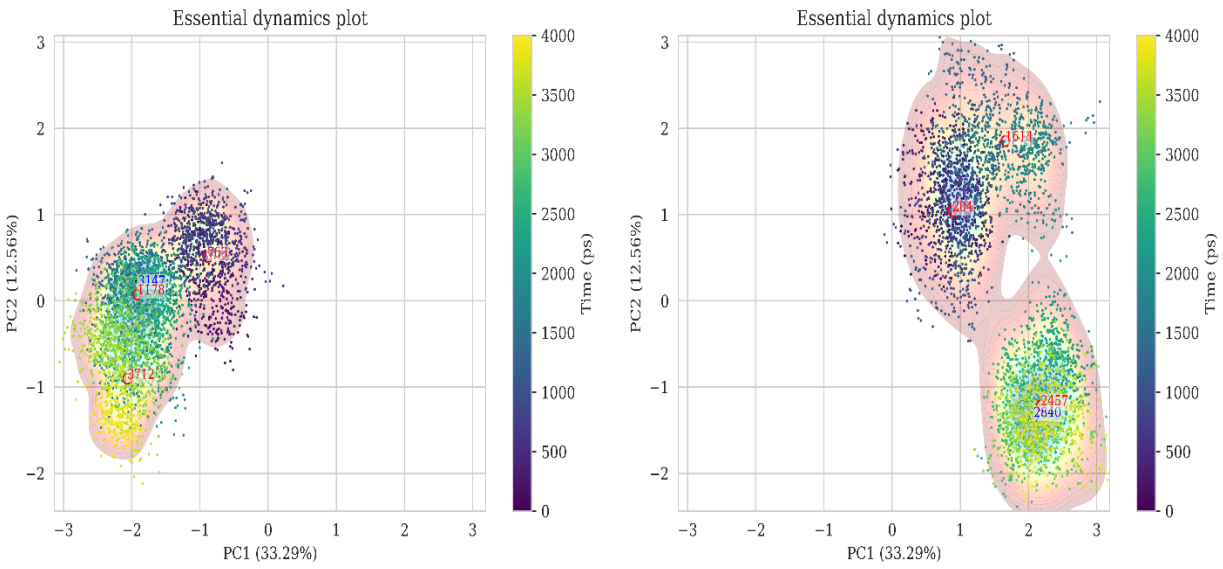


Figure A 2. Comparative essential dynamics PCA between wildtype NBD2 (left) and L1254A (right). Red circles represent K-means centroids and blue circles represent conformations with the lowest energy. The contour plot corresponds to the kernel densities of the first two PCAs. Contour plot also is comparable to the free energy of the protein surface.

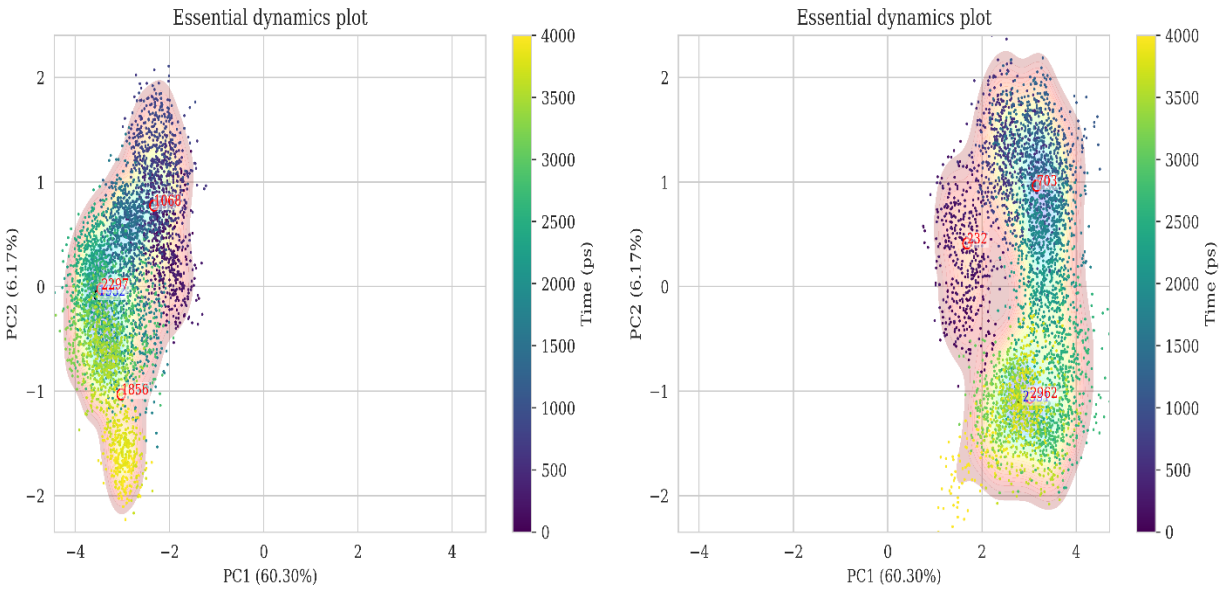


Figure A 3. Comparative essential dynamics PCA between wildtype NBD2 (left) and L1346F (right). Red circles represent K-means centroids and blue circles represent conformations with the lowest energy. The contour plot corresponds to the kernel densities of the first two PCAs. Contour plot also is comparable to the free energy of the protein surface.

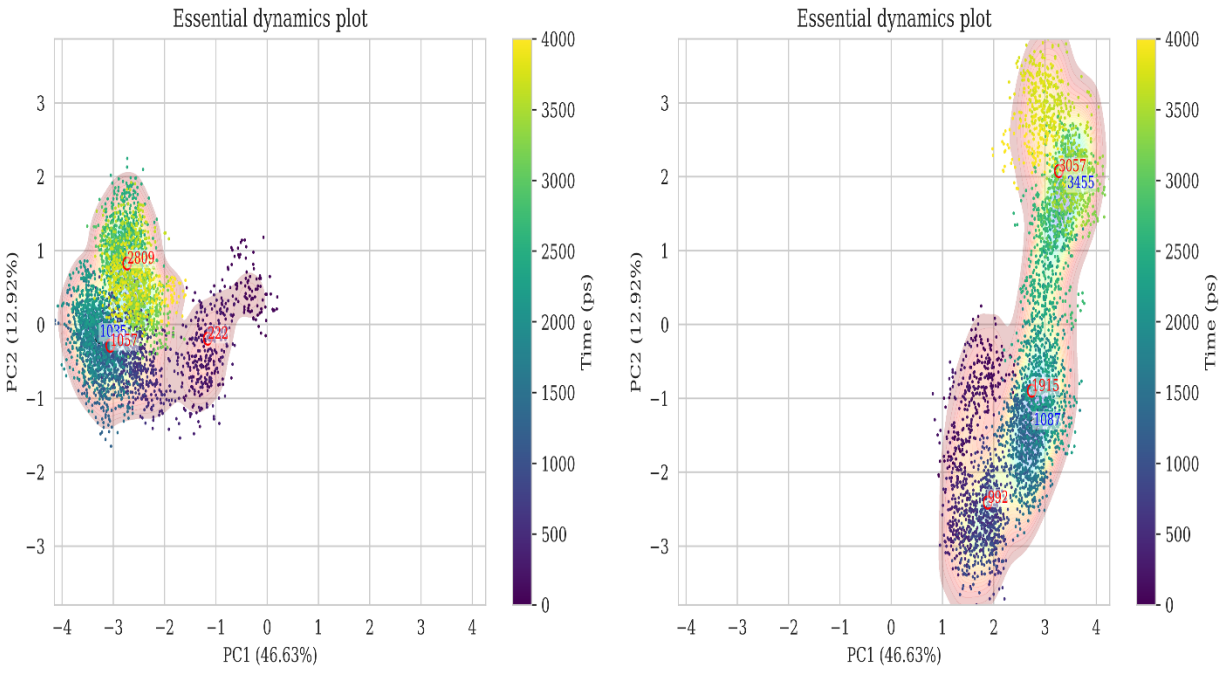


Figure A 4. Comparative essential dynamics PCA between N1303K NBD2 (left) and L1346F(right). Red circles represent K-means centroids and blue circles represent conformations with the lowest energy. The contour plot corresponds to the kernel densities of the first two PCAs. Contour plot also is comparable to the free energy of the protein surface.

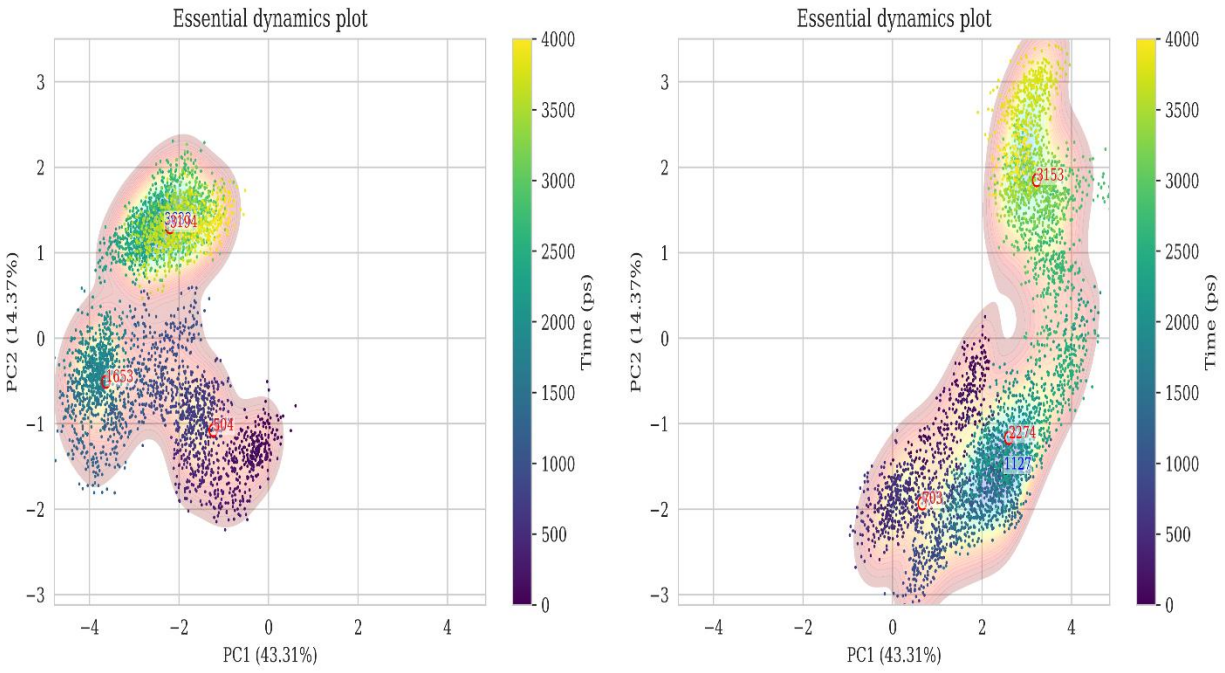


Figure A 5. Comparative essential dynamics PCA between N1303K NBD2 (left) and L1254A (right). Red circles represent K-means centroids and blue circles represent conformations with the lowest energy. The contour plot corresponds to the kernel densities of the first two PCAs. Contour plot also is comparable to the free energy of the protein surface.

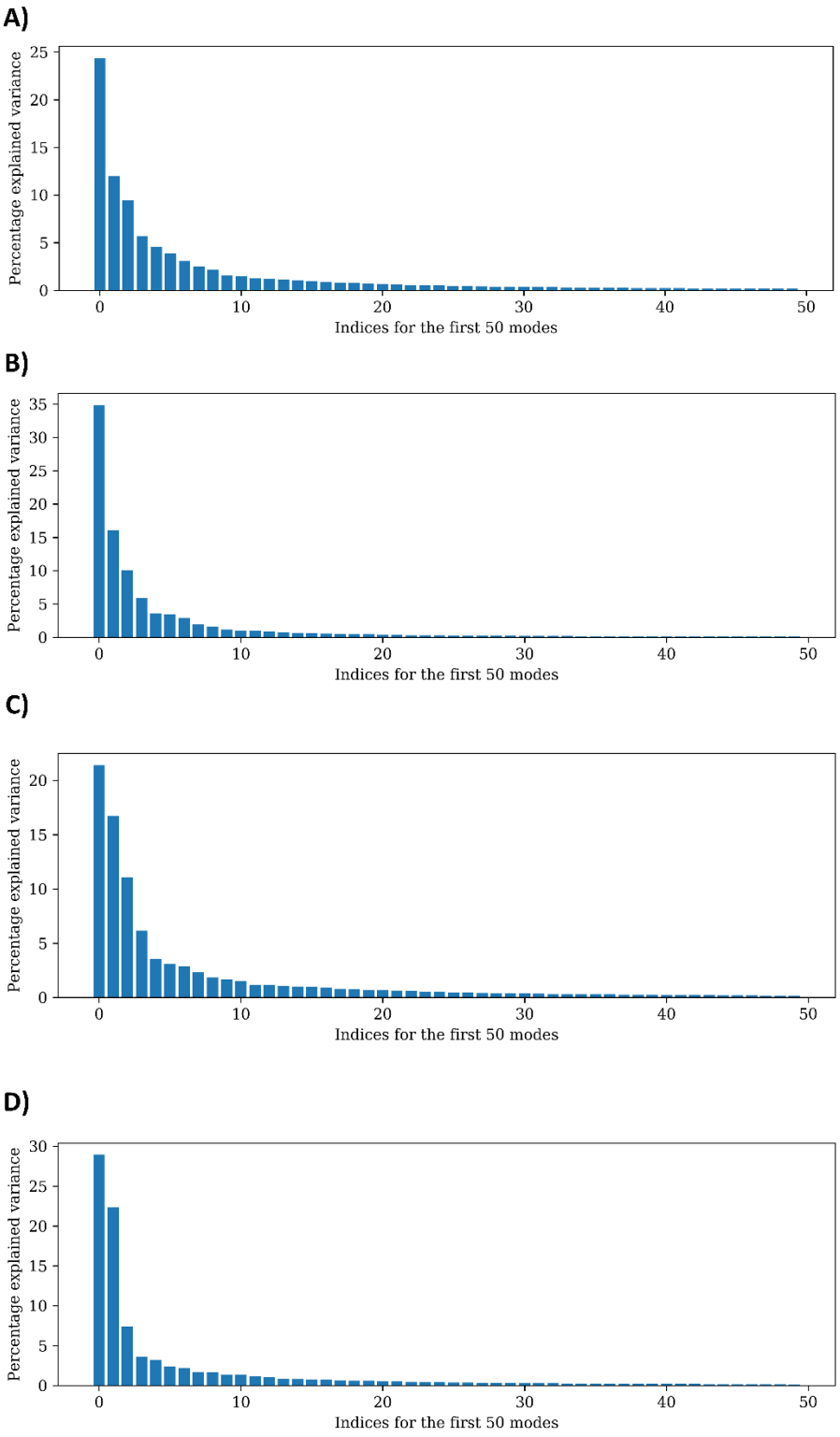


Figure A 6. Percentage explained variance across modes for NMA A) wildtype NBD2, B) N1303K, C) L1346F and D) L1254A.

Table A 1. Primers Used for PCR Mutagenesis

Mutant	**Sequence**	**Annealing Temperature (Tm (°C))**
L1254A-Forward	CATCAGCTTTTTTGAGACTACTGAACACTGAAG	59
L1254A-Reverse	CCAAAGTACTCTTCCCTGATCCAGTTCTTCCCAAG	59
L1346F-Forward	TGGCCACAAGCAATTAATGTGCTTGGCTAGATCTGTTC	59
L1346F-Reverse	TGGCTGAAGACACAGCCCCCATCCACAAGGACAAAGTC	59

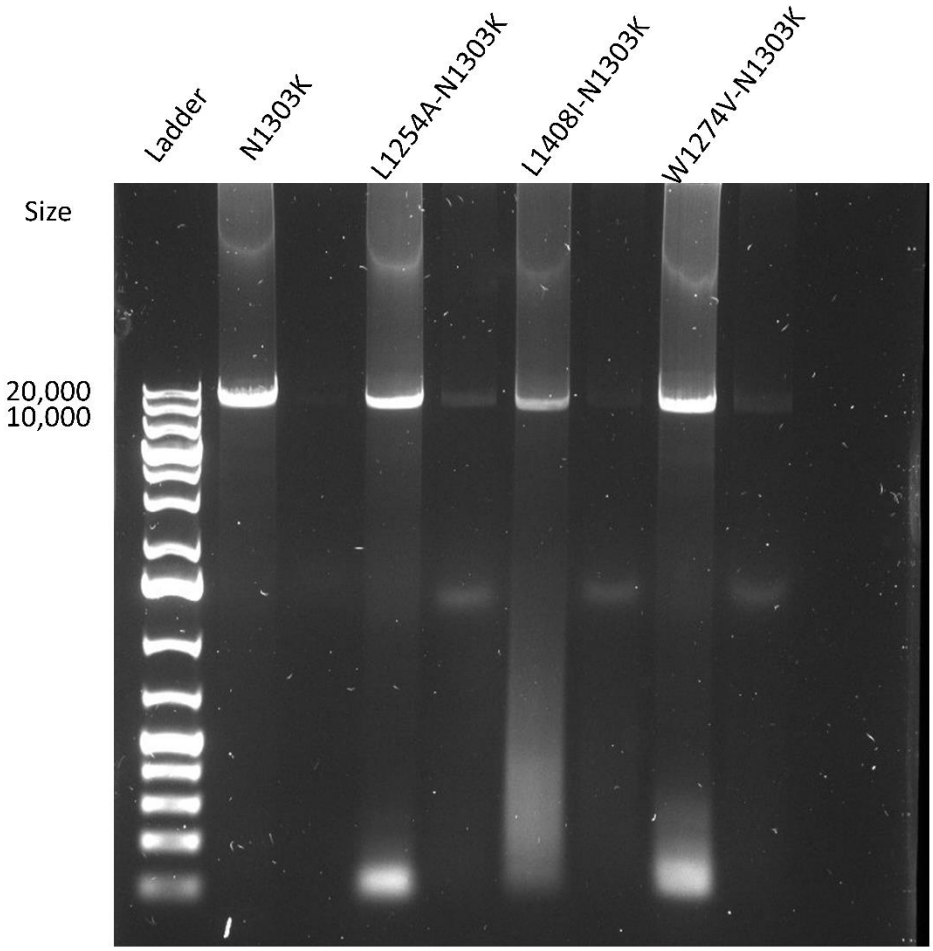


Figure A 7. PCR products of site-directed mutagenesis of L1254A. PCR products are around 10,000 bp, which is the correct size of the plasmid, suggesting plasmid was amplified correctly. L1408I and W1274V were not used for this thesis.

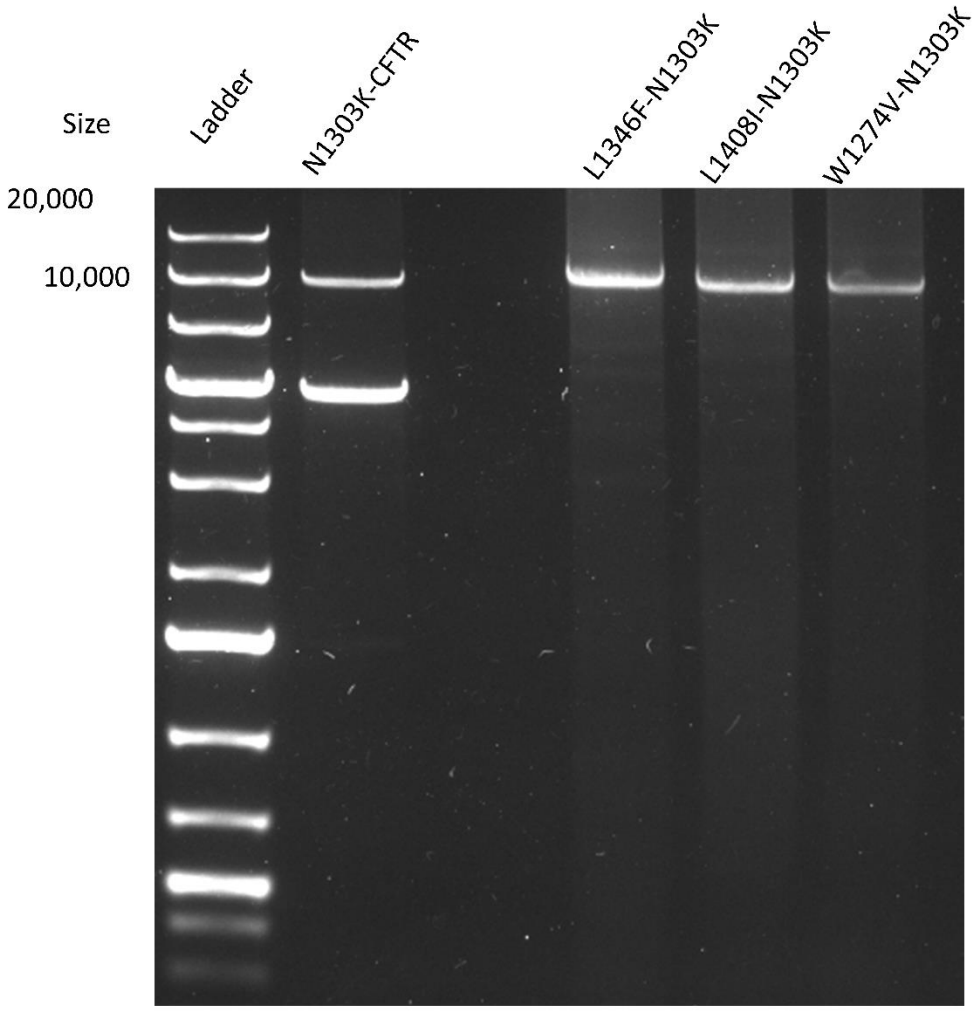


Figure A 8. PCR products of site-directed mutagenesis of L1346F. Size of PCR product is around 10,000 bp, suggesting that the mutation was made correctly, as this is the size of the N1303K plasmid. L1408I and W1274V were not used for this thesis.

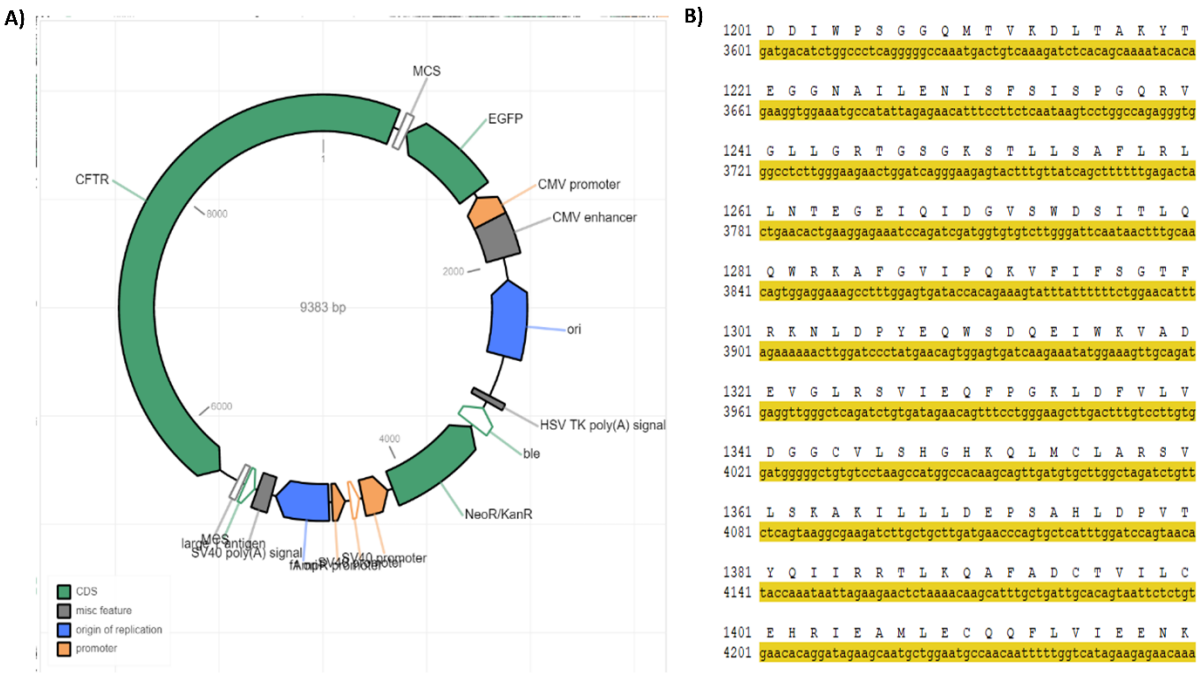


Figure A 9. Sequencing results of wildtype CFTR. A) shows plasmid map, indicating the plasmid is around the correct size (10 kbp). CFTR and GFP also appear on the plasmid map, indicating that CFTR-GFP is present. B) shows the sequences for NBD2 (between 1200 and 1400) and indicates that this plasmid is indeed wildtype CFTR. Sequencing was done via PlasmidSaurus.

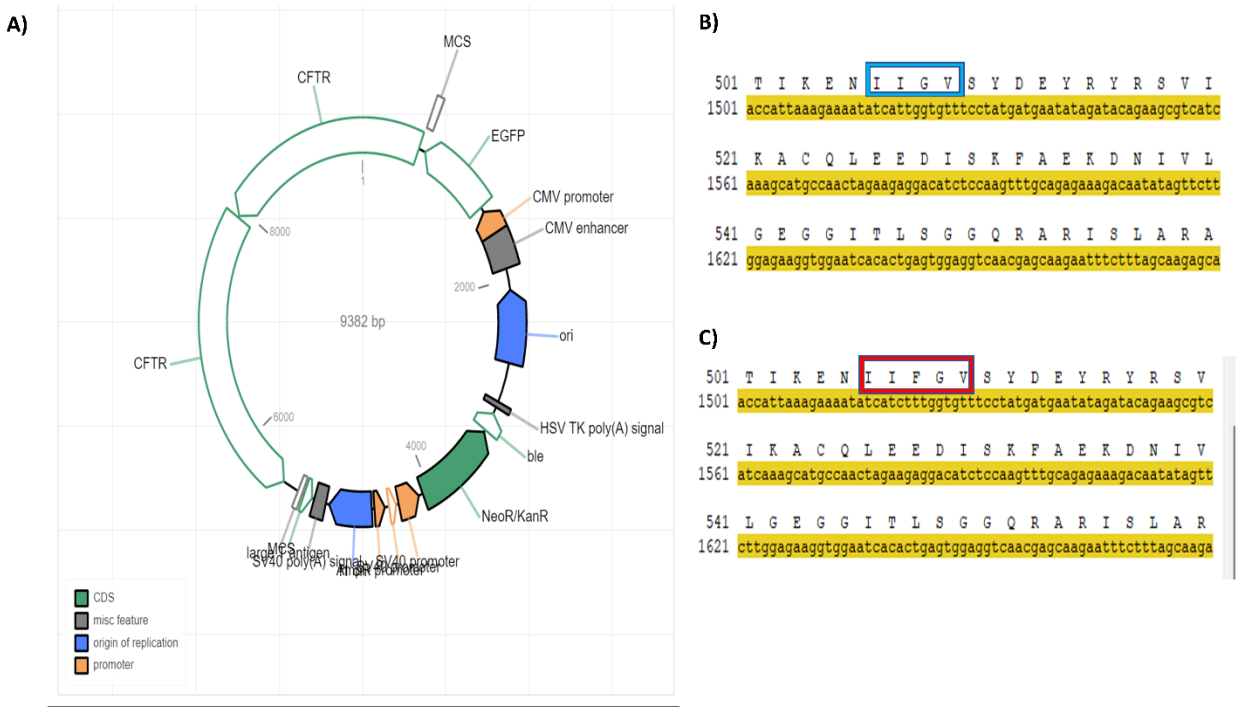


Figure A 10. Sequencing results of ΔF508-CFTR. A) shows plasmid map, indicating the plasmid is around the correct size (10 kbp). CFTR and GFP also appear on the plasmid map, indicating that CFTR-GFP is present. B) shows the sequences for NBD1 for ΔF508 (betwee500 and 550), while C) shows the same sequence for wildtype CFTR. The Red Box denotes that the sequence is indeed missing an F amino acid residue at position 508, showing that the correct mutation has been made. The blue box denotes the wildtype CFTR sequence at position 508, showing that there is an F at that position. Sequencing was done via PlasmidSaurus.

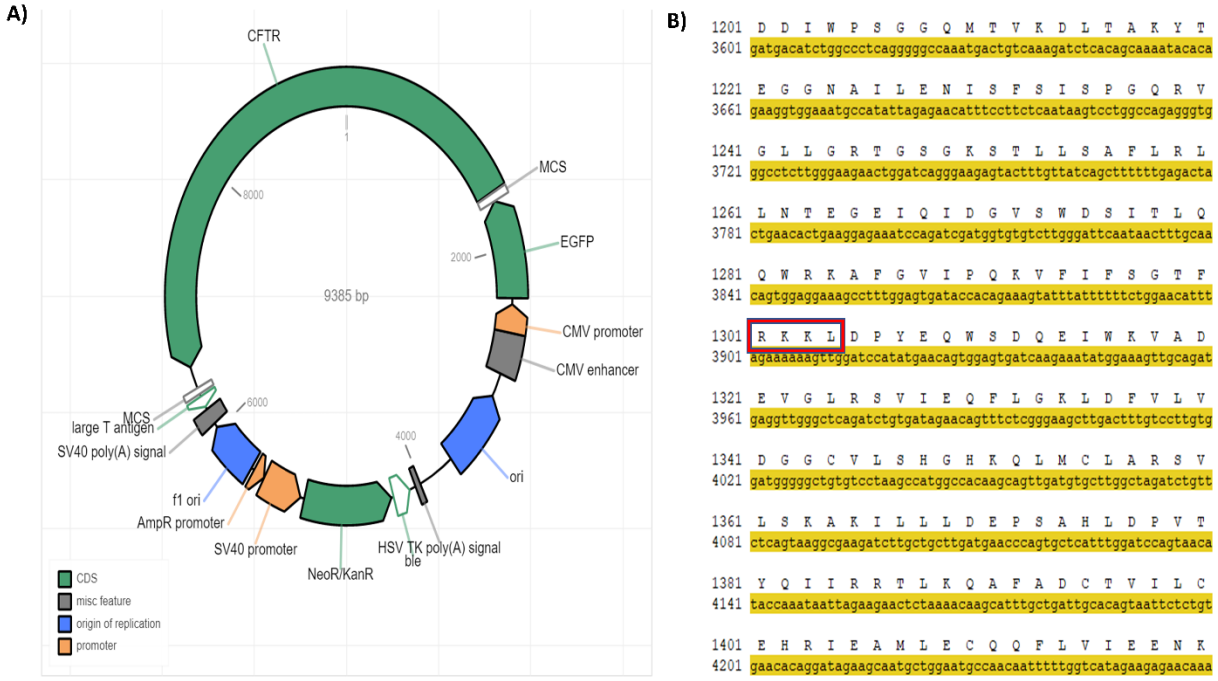


Figure A 11. Sequencing results of N1303K-CFTR. A) shows plasmid map, indicating the plasmid is around the correct size (10 kbp). CFTR and GFP also appear on the plasmid map, indicating that CFTR-GFP is present. B) shows the sequences for NBD2 (between 1200 and 1400). The Red Box denotes that the sequence does indeed have a K amino acid at position 1303, meaning the mutation was successful. Sequencing was done via PlasmidSaurus.

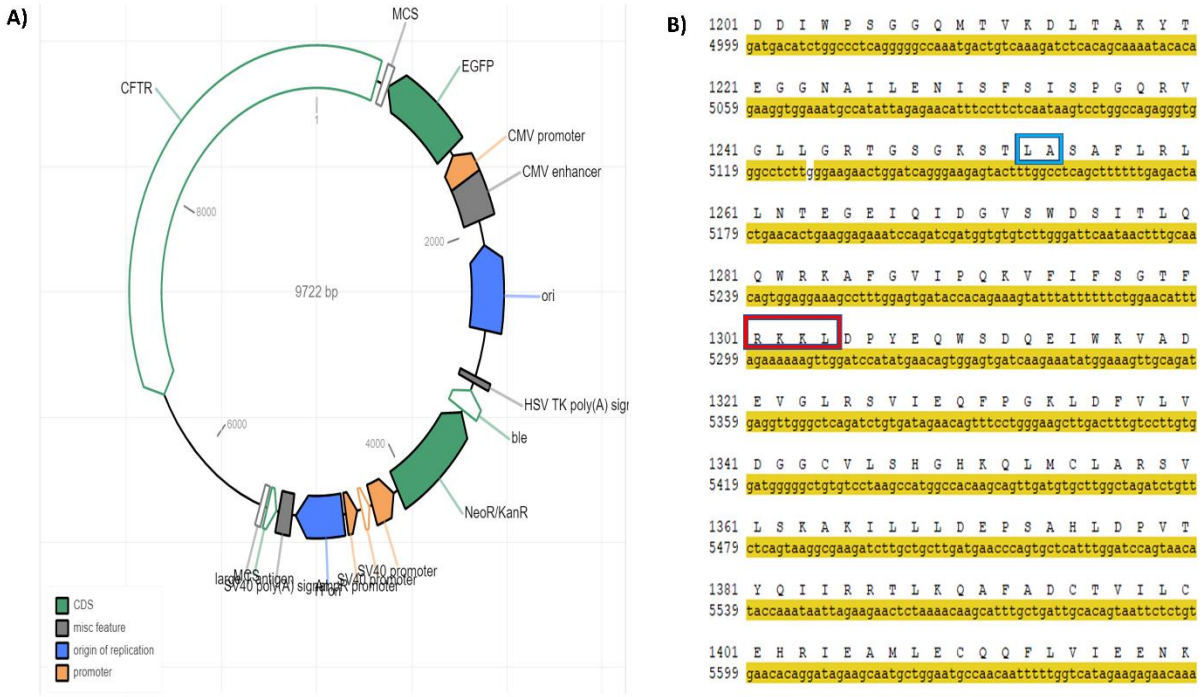


Figure A 12. Sequencing results of L1254A-CFTR. A) shows plasmid map, indicating the plasmid is around the correct size (10 kbp). CFTR and GFP also appear on the plasmid map, indicating that CFTR-GFP is present. B) shows the sequences for NBD2 (between 1200 and 1400). This sequence was spliced together from two different sequencing reads. The Red Box denotes that the sequence does indeed have a K amino acid at position 1303, meaning that it has the N1303K mutation. The blue box denotes the change of L to A at position 1254, meaning that the mutation was correctly made. Sequencing was done via PlasmidSaurus.

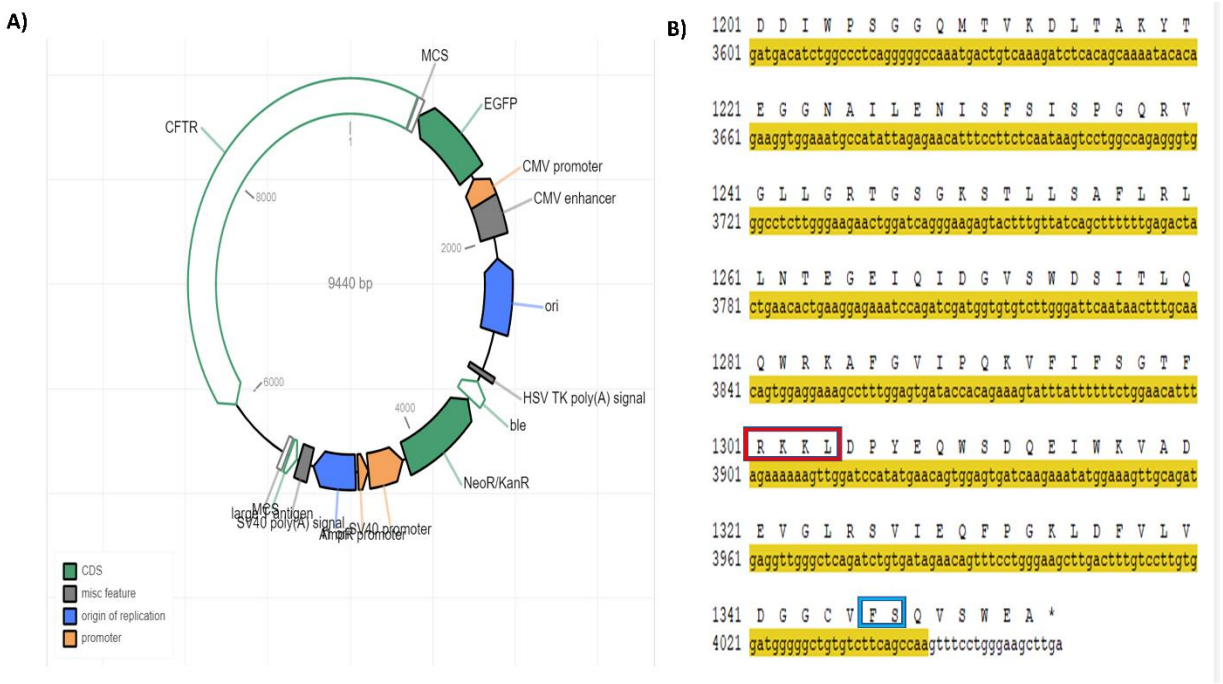
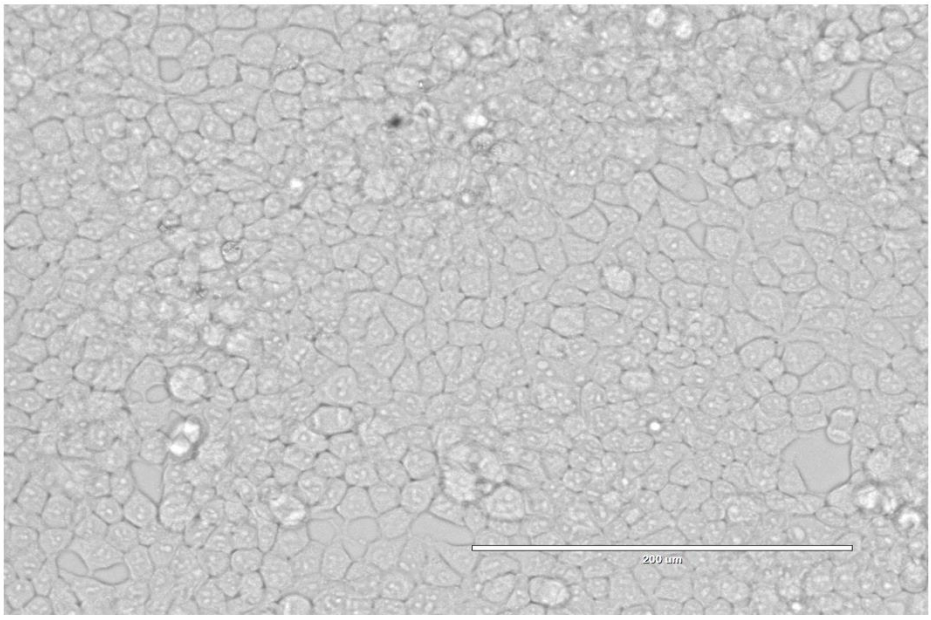


Figure A 13. Sequencing results of L1346F-CFTR. A) shows plasmid map, indicating the plasmid is around the correct size (10 kbp). CFTR and GFP also appear on the plasmid map, indicating that CFTR-GFP is present. B) shows the sequences for NBD2 (between 1200 and 1400). The Red Box denotes that the sequence does indeed have a K amino acid at position 1303, meaning that it has the N1303K mutation. The blue box denotes the change of L to F at position 1346, meaning that the mutation was correctly made. Sequencing was done via PlasmidSaurus.

A)



B)

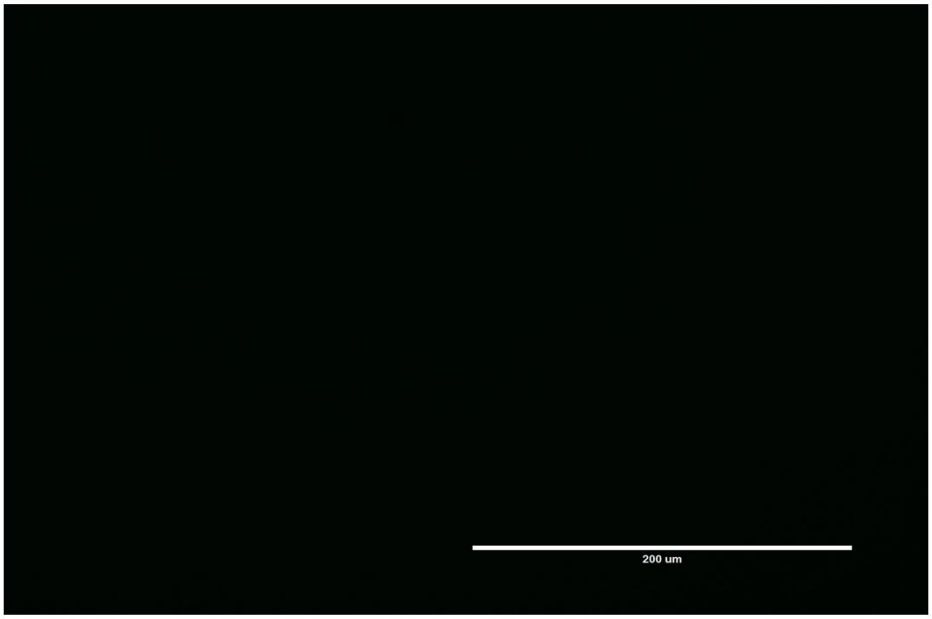
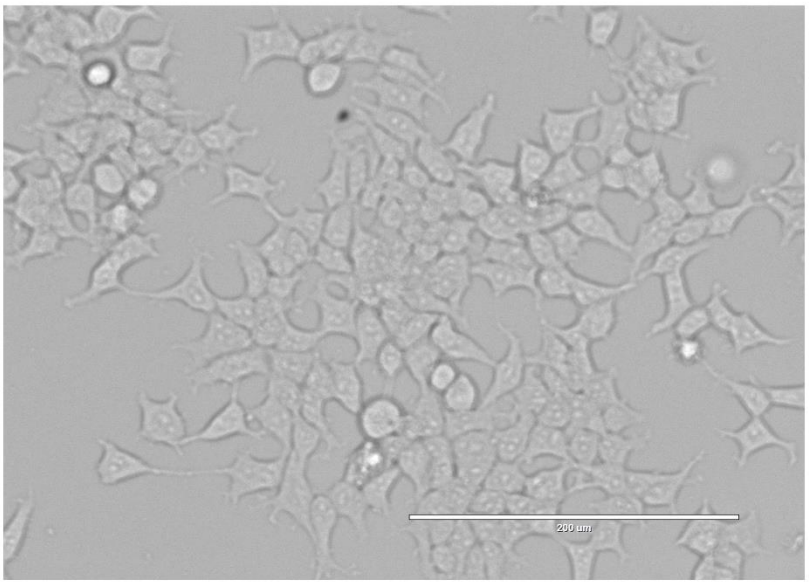


Figure A 14. EVOS images of mock transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. Lack of presence in GFP is an indicator of successful transfection, as no GFP image should be available for mock. The scale bar indicates 200 µm.

A)



B)

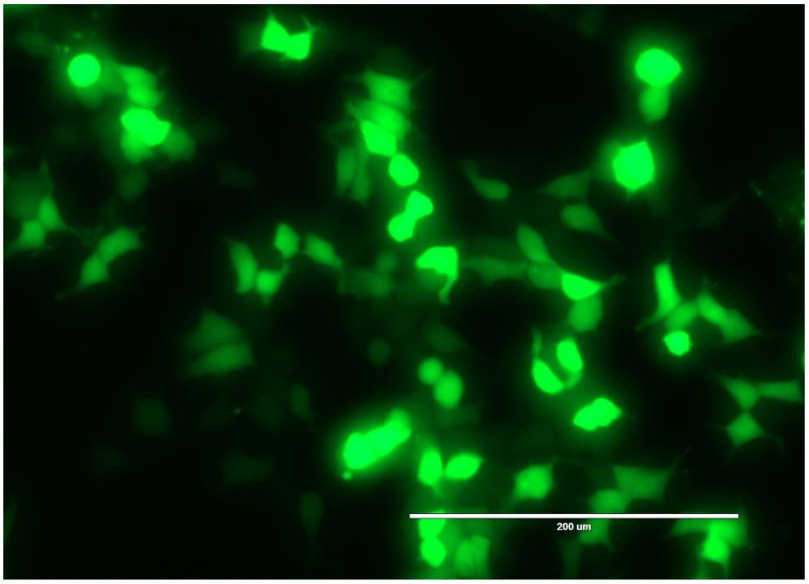
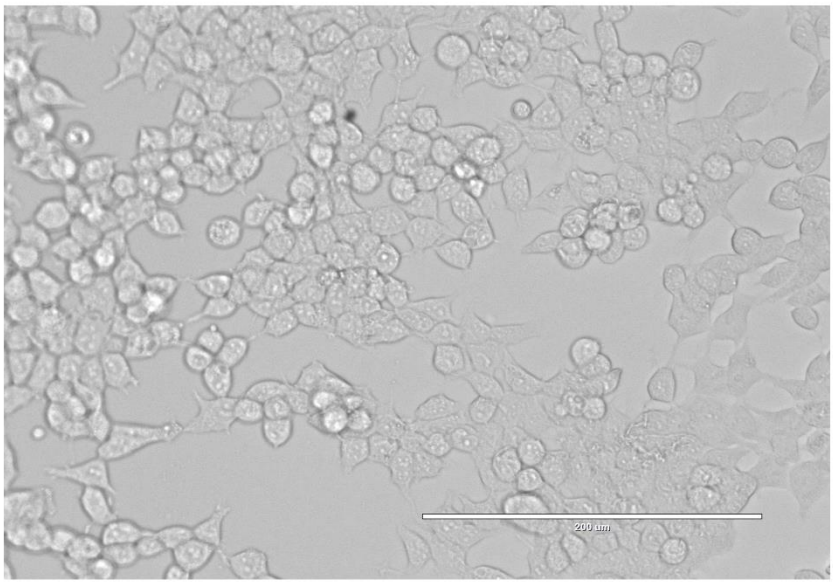


Figure A 15. EVOS images of GFP transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. Presence in the GFP filter is an indicator of successful transfection, as GFP protein should show up in this filter, indicating the plasmid was translated. The scale bar indicates 200 µm.

A)



B)

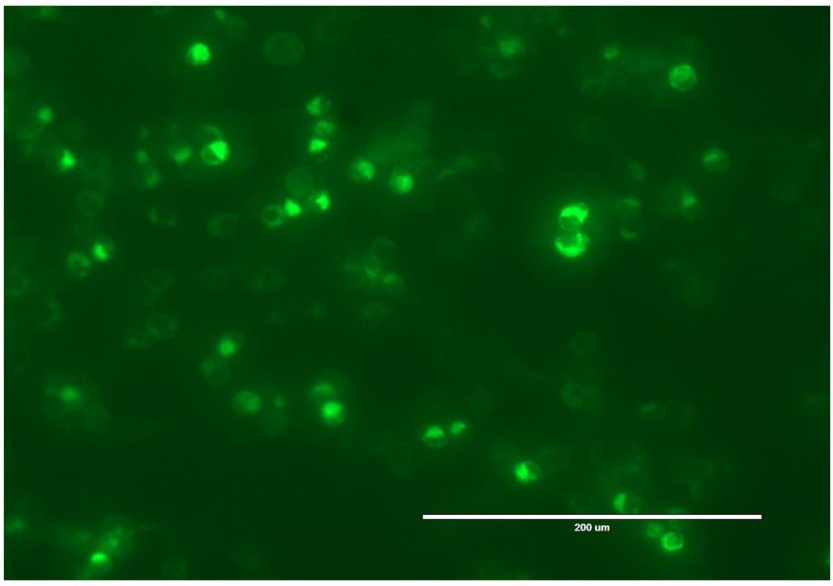
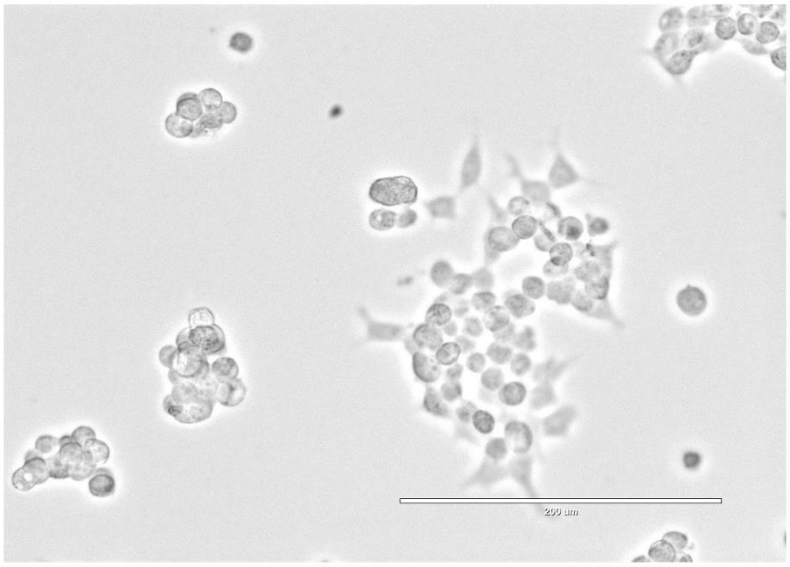


Figure A 16. EVOS images of wildtype CFTR transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. Presence in the GFP filter is an indicator of successful transfection, as GFP-CFTR protein should show up, indicating the plasmid was translated. The scale bar indicates 200 µm.

A)



B)

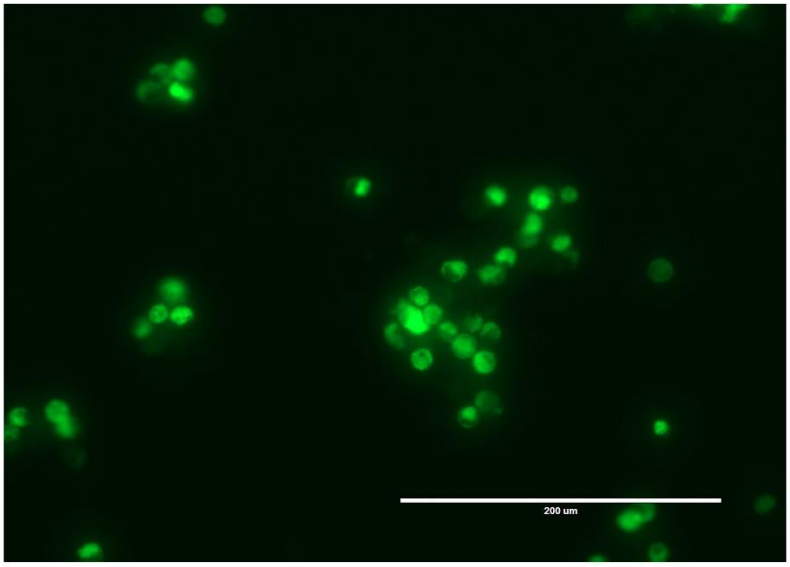
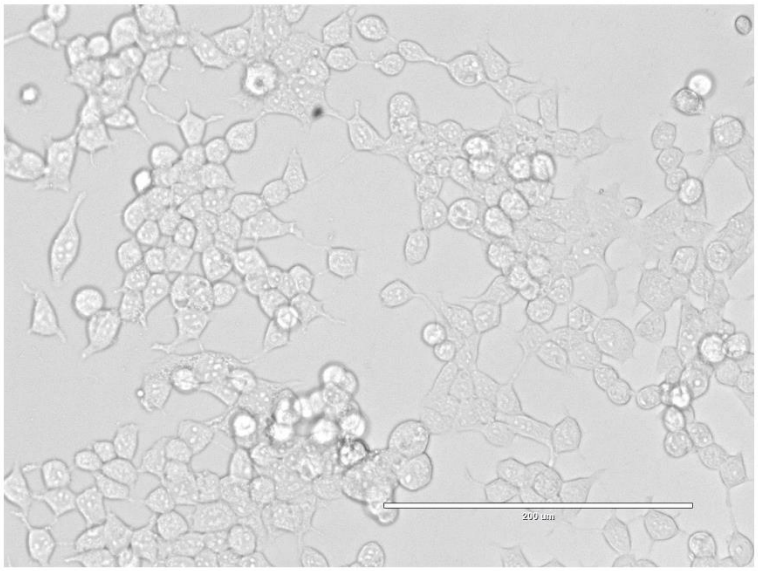


Figure A 17. EVOS images of ΔF508-CFTR transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. The presence in the GFP filter is an indicator of successful transfection, as GFP-CFTR protein should show up, indicating the plasmid was translated. The scale bar indicates 200 µm.

A)



B)

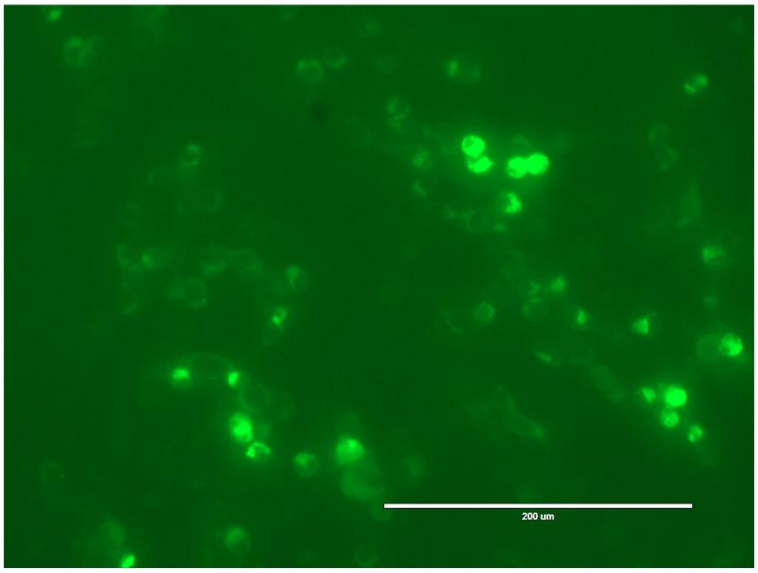
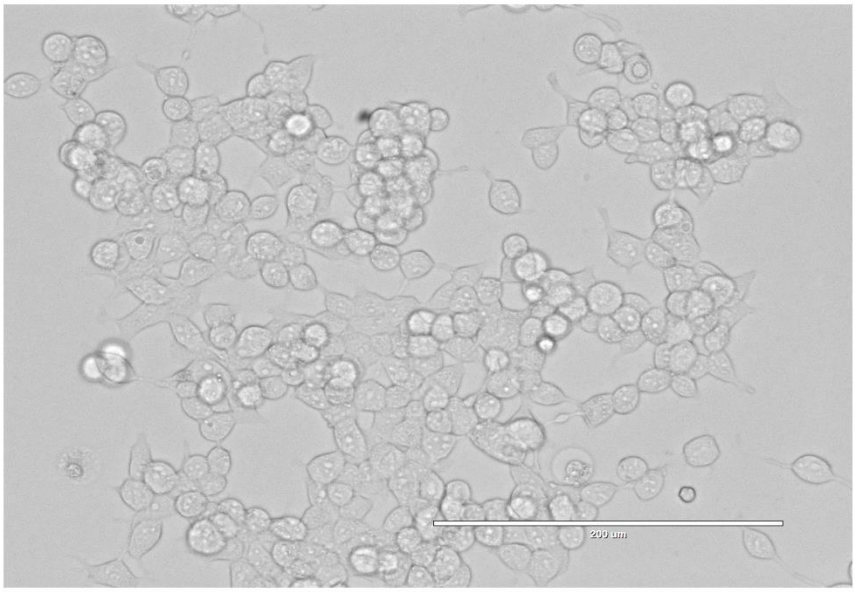


Figure A 18. EVOS images of N1303K-CFTR transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. The presence in the GFP filter is an indicator of successful transfection, as GFP-CFTR protein should show up, indicating the plasmid was translated. The scale bar indicates 200 µm.

A)



B)

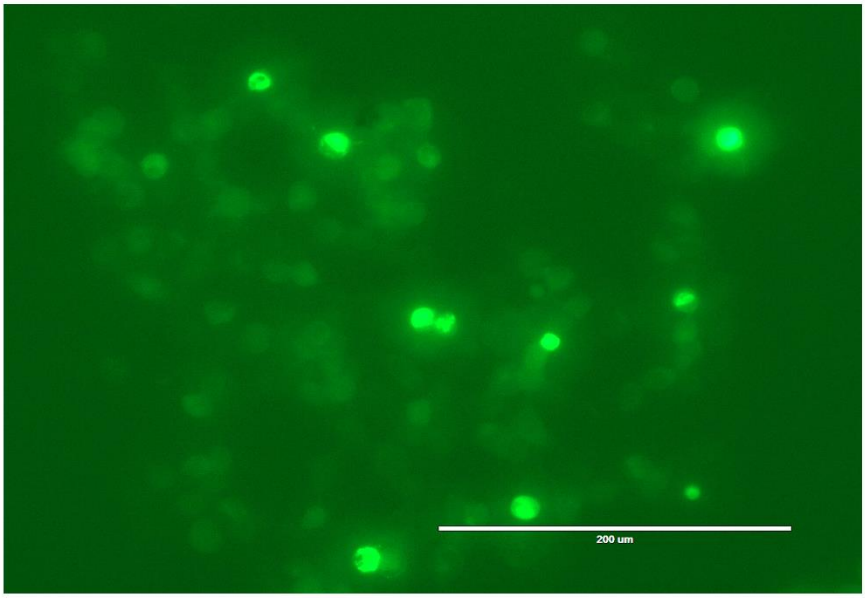
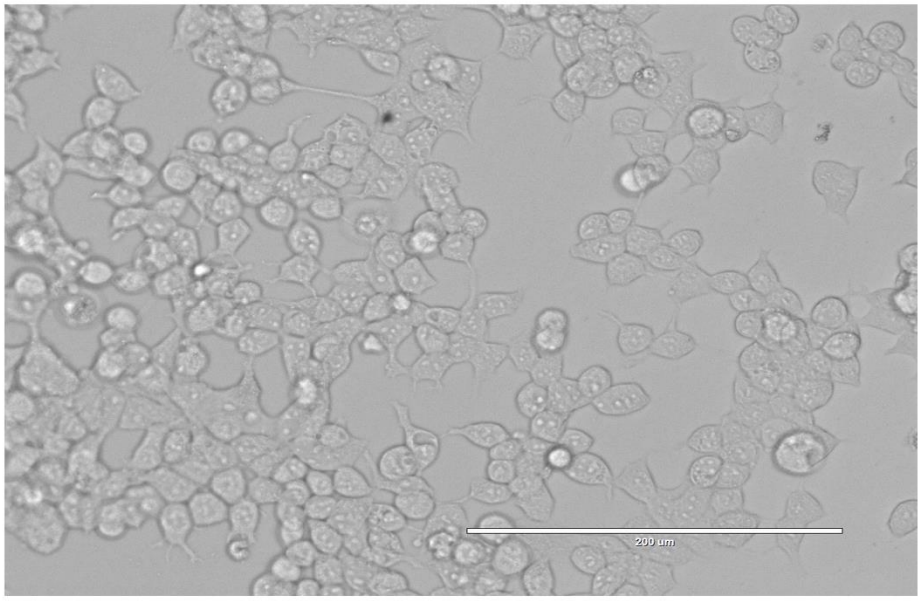


Figure A 19. EVOS images of L1254A-CFTR transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. The presence in the GFP filter is an indicator of successful transfection, as GFP-CFTR protein should show up, indicating the plasmid was translated. The scale bar indicates 200 µm.

A)



B)

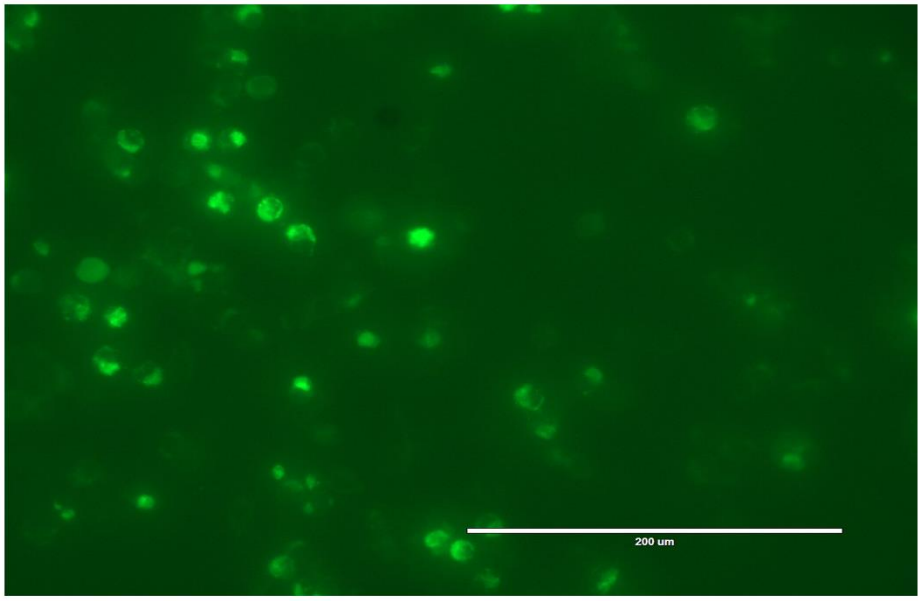


Figure A 20. EVOS images of L1346F-CFTR transfection of HEK293 cells. A) shows the transmission filter for the HEK293 cells, whereas B) shows GFP. The presence in the GFP filter is an indicator of successful transfection, as GFP-CFTR protein should show up, indicating the plasmid was translated. The scale bar indicates 200 µm.

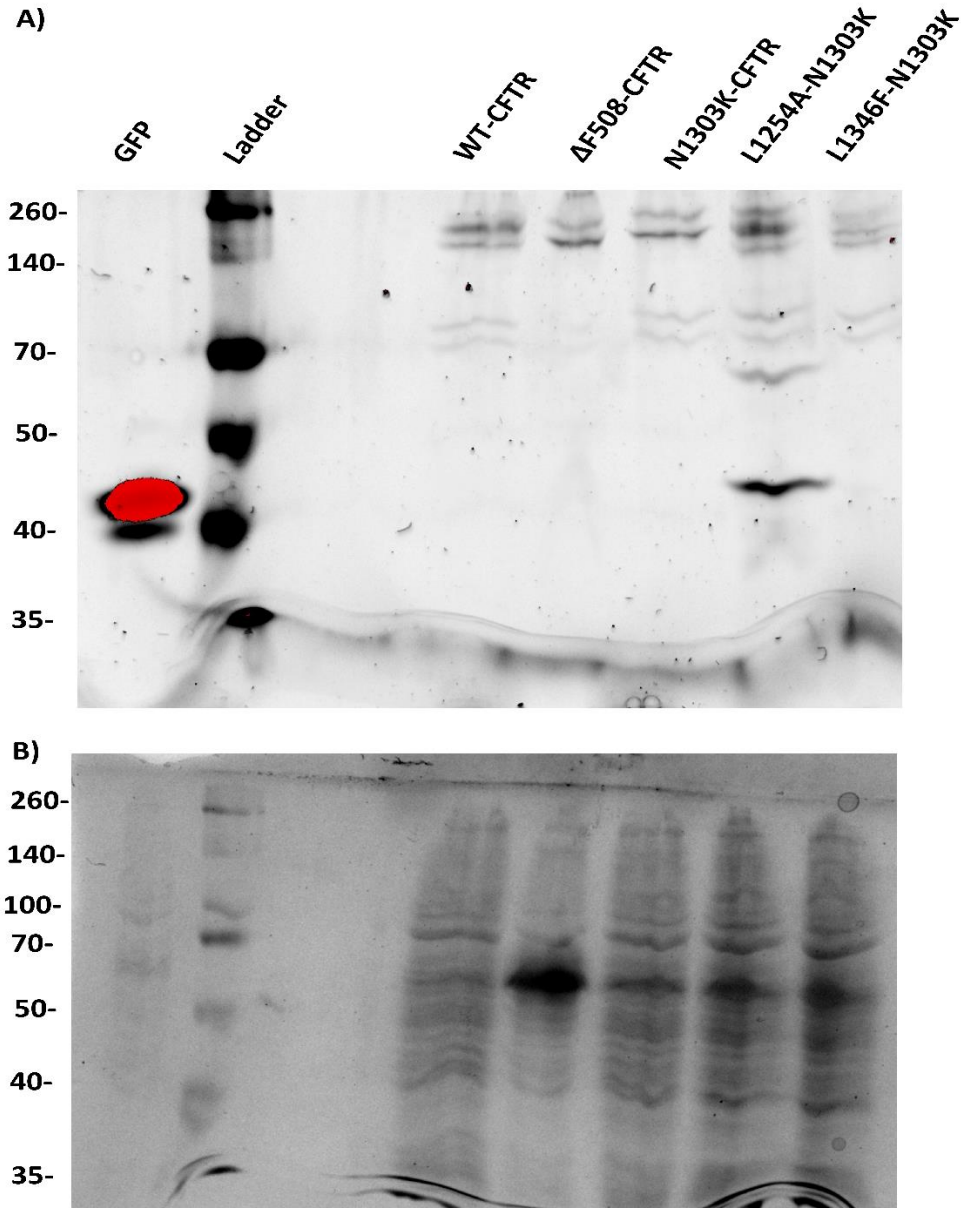


Figure A 21. A) In-gel fluorescence of CFTR proteins from transfection experiment 1. There was an acidity problem with this gel, explaining the waviness. C-B bands are still readable. L1254A shows cleavage of a GFP-sized protein. Other endogenous proteins are apparent in the gel besides CFTR. B) shows Coomassie stain of all proteins.

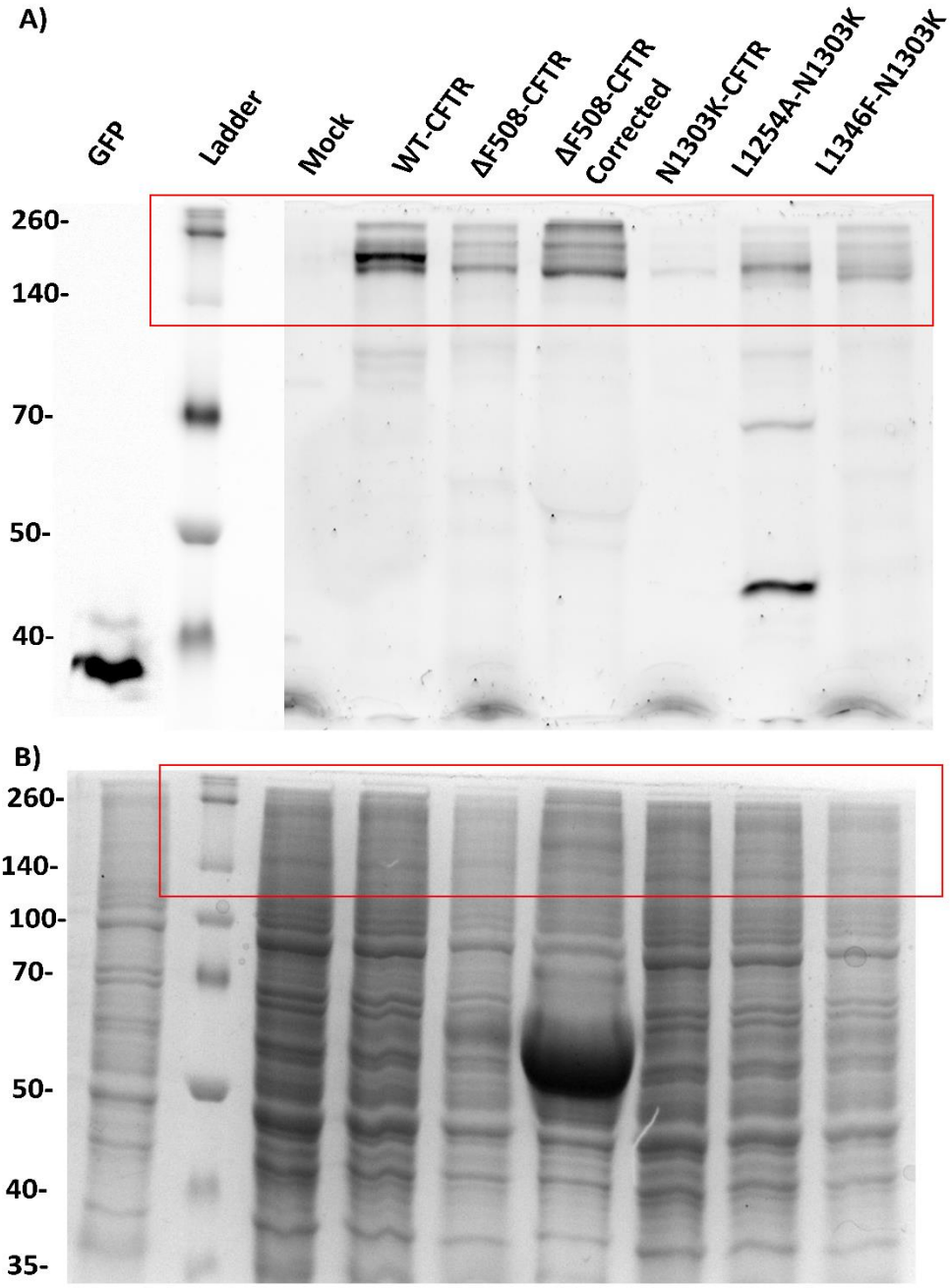


Figure A 22. A) In-gel fluorescence of CFTR proteins from transfection experiment 2. C-B bands are still readable. L1254A shows cleavage of a GFP-sized protein. Other endogenous proteins are apparent in the gel besides CFTR. B) shows Coomassie stain of all proteins. The red box indicates the area the representative gel is sourced from.

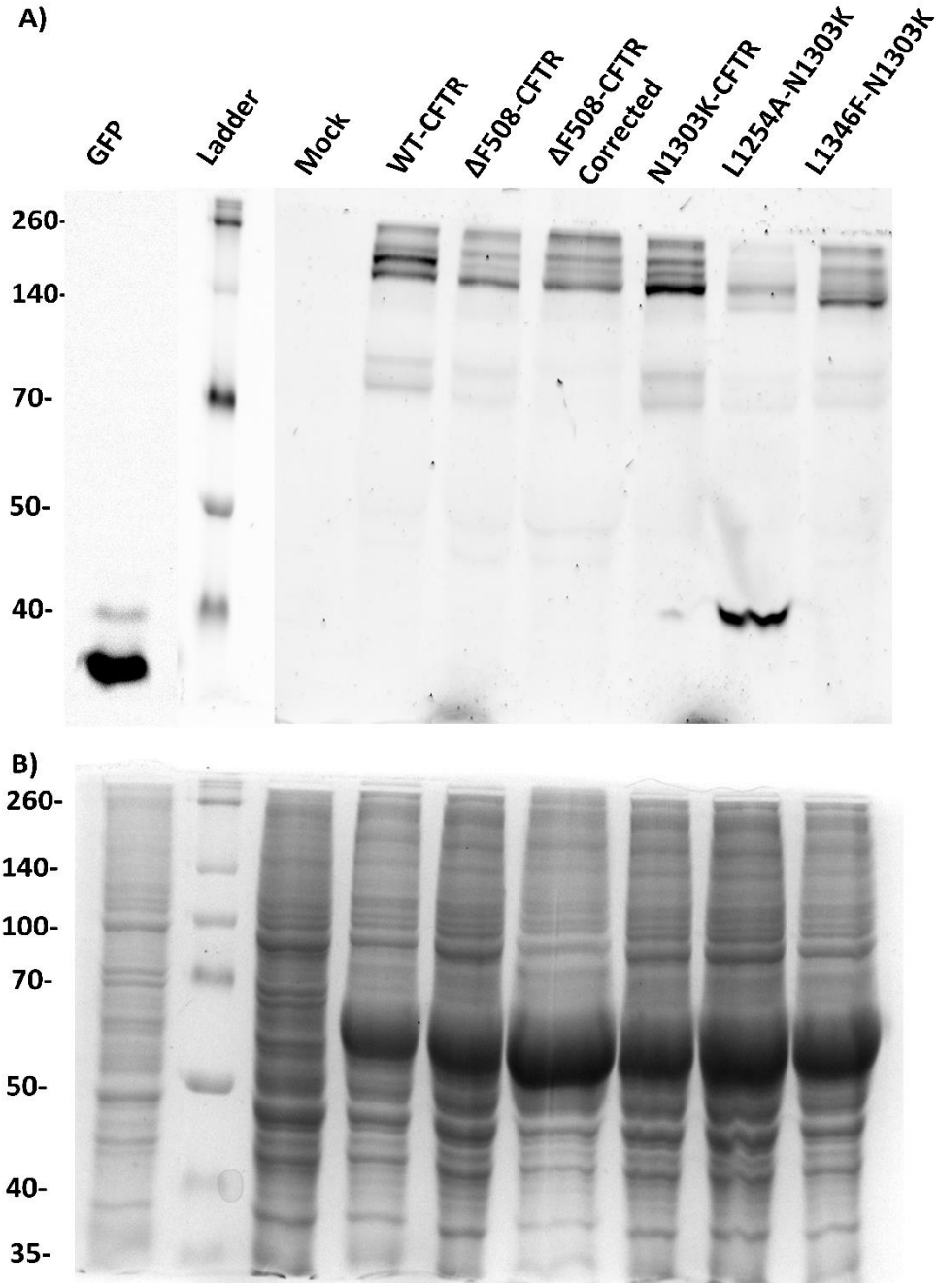


Figure A 23. A) In-gel fluorescence of CFTR proteins from transfection experiment 3. C-B bands are still readable. L1254A shows cleavage of a GFP-sized protein. Other endogenous proteins are apparent in the gel besides CFTR. B) shows Coomassie stain of all proteins.

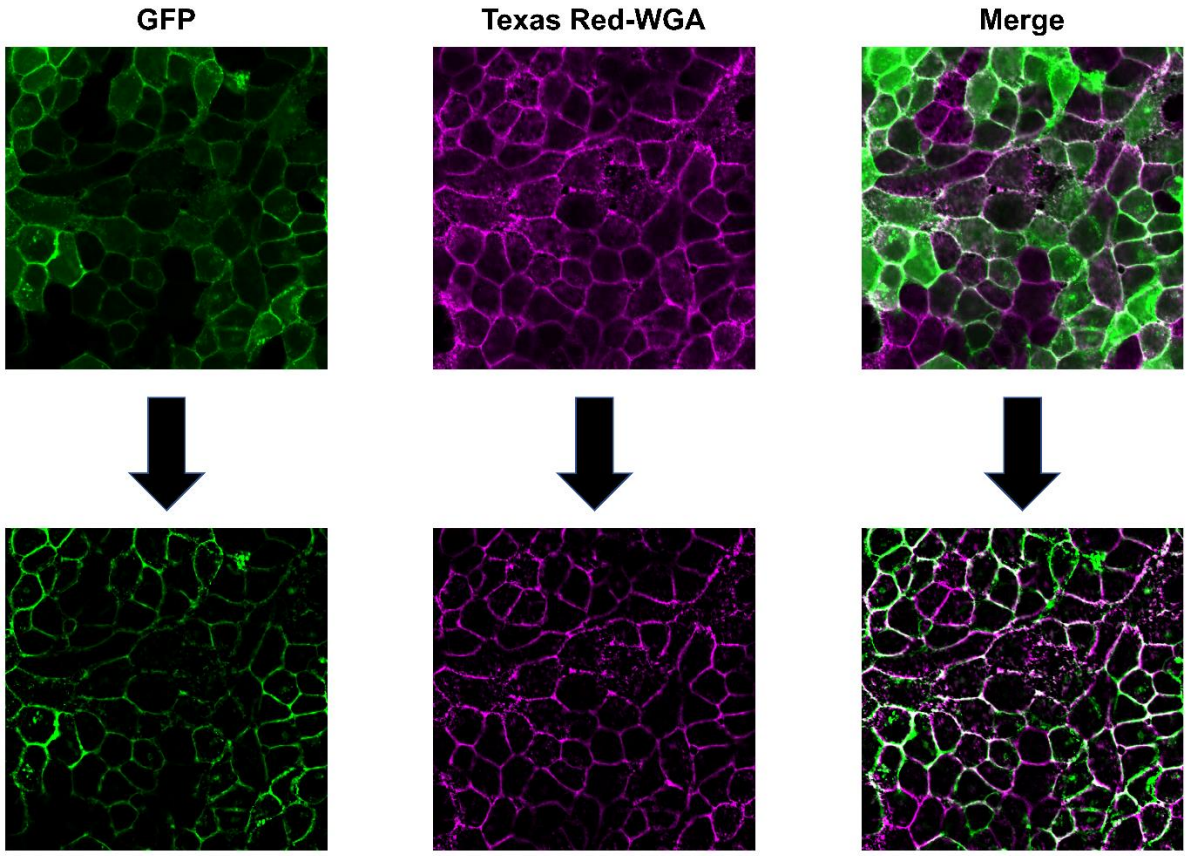


Figure A 24. Example median filter subtraction of GFP-PM1 (GFP) and Texas Red-WGA using ImageJ (Schindelin et al., 2012). The median filter is 32 pixels.