Structural Bioinformatics Final Project

Mutations in Cystic Fibrosis Transmembrane Conductance Regulator

December 2019

Maryam Khan
Kevin Mittal

Introduction

The CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) protein is a functional ATP Binding Cassette Protein (ABC) protein. It is a multidomain, integral protein made up of five domains: two membrane-spanning domains (MSD1 and MSD2) that form the chloride ion channel, two nucleotide-binding domains (NBD1 and NBD2) that bind and hydrolyze ATP and a regulatory (R) domain. It is localized in the apical membrane of epithelial cells and confers cAMP - activatable transport of chloride, bicarbonate and glutathione.

The protein functions as a channel across the membrane of cells that produce mucus, sweat, saliva, tears, and digestive enzymes such as in lungs and other tissues.. It uses energy from ATP hydrolysis to transfer. ATP-driven dimerization of CFTR's two nucleotide binding domains (NBD1 and NBD2) leads to the opening of the channel. Our focus will be on how ATP binds to NDB1. The crystal structure of human CFTR's NBD1 reveals that the side chain of the Trp-401 and Tyr-1219 residues forms ring–ring stacking interaction with the adenine ring of the bound ATP molecule (1).

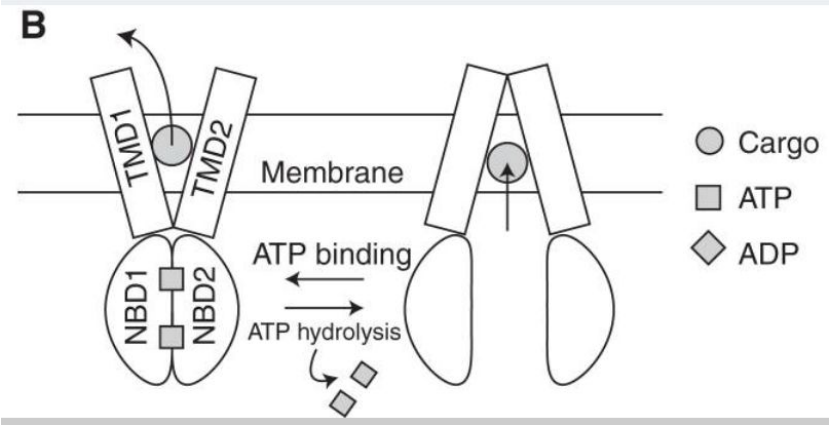

Figure 1: Cartoon description of how the CFTR protein works (2)

All people have two copies of the CFTR gene. However, mutations must be present in both copies to cause CF. More than 1,700 mutations of the CFTR gene have been identified.Mutations in the protein prevent apical membrane permeability for the chloride ion. This then leads to decreased surface liquid volume, which impairs mucociliary clearance, which in turn leads to respiratory bacterial infection. Once obstruction is established, multiple mechanisms perpetuate excessive chronic inflammation, mucus hypersecretion and chronic bacterial infection. This creates a progressive chronic

airway disease leading to end-stage lung disease, the primary cause of morbidity and mortality in this disorder.

There are a number of ways the CFTR protein can be mutated, altering the structure and function of the protein. There are five different classes of CFTR mutations, and their descriptions are listed in the table below.

Class	Description	Mutations	% of individuals with Cystic Fibrosis who have at least one mutation in that class
Class I	“Nonsense, splice, or deletion” mutations result in the production of an unstable, shortened RNA molecule that is easily degraded. No functional CFTR mutation is created	G542X, W1282X, R553X	22 %
Class II	The CFTR protein is created but it misfolds, which prevents it from moving to the cell surface	F508del, N1303K, I507del	88%
Class III	CFTR protein is created and it moves to the cell surface, but the channel gate does not open properly.	G551D, S549N	6%
Class IV	The CFTR protein is created and it moves to the cell surface. However, the function of the channel is faulty. Here, the ion conduction pore is impeded, resulting in reduced unitary conductance.	D1152H, R347P, R117H	6%
Class V	The CFTR protein is created, moves to the cell surface, and	A455E	5%

	functions normally but it is present in insufficient quantities		

Table 1: Description of the five different classes of the CFTR mutation.

The mutation which we have decided to focus on is the *G551D* mutation, which is the most prevalent gating mutation. It is the third overall most common CF mutation with a worldwide frequency of approzimately 3%. In this mutation, glycine is substituted by aspartate at position 551 in the nucleotide binding domain-1 of the CFTR gene. The substitution by the larger amino acid, aspartate blocks the ATP from binding to the CFTR protein. This results in an open probability that is approximately 100-fold lower than that of wild-type channels. In simpler terms, the chloride current across the membrane is significantly reduced due to the inability of the channel to open up. The G551D mutation results in a severe phenotype that is characterised by pulmonary dysfunction and pancreatic insufficiency.

Hypothesis

By studying the G551D mutation, we observe that the large side chain residue of Aspartate, blocks tyrosine at position 1219 from interacting directly with the adenine ring of ATP. Previous studies have shown that this is due to a size change, not necessarily an electrostatic change.

Our hypothesis is that the same result should be observed when glycine is substituted by glutamate at the same position, G551E. Glutamate, which also contains a larger side chain residue that glycine, will block ATP binding to Tyr-1219, which will ultimately prevent the two NBDs from dimerizing and the channel from opening.

Our control study will consist of mutations within CFTR but of some other amino acids, which still preserve the function of CFTR. We found two mutations which do not alter the function: the substitution of Arginine to Cysteine at position 31 and the substitution of Isoleucine to Threonine at position 148 (3).

Dataset

The wildtype CFTR protein was obtained from the Protein Data Bank. In particular, we used **6msm**, which is a phosphorylated, **ATP-bound human cystic fibrosis transmembrane conductance regulator (CFTR).** The rest of the mutant structures were developed using Swiss-Model, with details outlined in the Testing Process.

Testing Process

Developing our mutant structures

No mutant structures were available at PDB or other online available resources. Initially, we tried making use of Scwrl to generate mutant structures but ran into issues with unexpected changes to other parts of the protein. Not only would the glycine not be changed to our desired amino acid, turning to isoleucine instead of aspartate, the amino acids before and after position 551 would be changed. It was no longer a point mutation.

We then utilized SWISS-MODEL, a fully automated protein structure homology-modelling server (4). The purpose of this server is to make protein modelling accessible. We used it to create the following mutant structures and control structures:

- Substitution of glycine to aspartate at position 551 - Known mutation
- Substitution of glycine to glutamate at position 551 - Hypothesis mutation
- Substitution of arginine to cysteine at position 31 - control study
- Substitution of Isoleucine to Threonine at position 148- control study

Each of our models developed used the pdb file, 6msm, as its template and had a 99.86% sequence identity match with the template.

The aligned structure of 6msm and our mutation 6-aspartate is shown below:

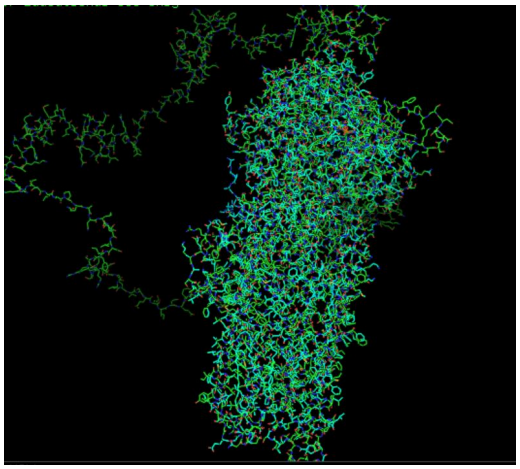

Figure 2: Aligned pdb files of 6msm with mutated G551D CFTR

It is important to note that the extra chain in the mutant structure, that looks like a strap, are ~200 amino acids that were not present in the 6msm pdb file. Since the active sites

and cavities were aligned and remained the same in the WT and mutant structures, we decided to proceed ahead, ignoring the “strap” around the protein structure.

We used an online server for DALI for protein structure comparison to make sure all our structures were aligned. This was developed by the OLM group at the University of Helsinki (5).

Developing a python script to reduce protein size

We were running into numerous errors due to the large protein size of CFTR, which is approximately 1500 amino acids long. The large size of the protein significantly slowed the ability of the programs used to analyze it, making some functions, such as difference or intersection operations unfeasible. To counteract the issue, we developed a python script which removed all of the lines from a pdb file including atoms that were more than 25 angstroms away from the alpha carbon of our known mutation. This step was completed with a short script that reads in each line of a pdb file, isolates the x, y, and z coordinates of each atom, and calculated the distance from the original 551 alpha carbon using the standard distance formula. If the distance was greater than 25, the line would be removed, and if not, it would be copied to a new pdb file. This simple method allowed visualization of only certain portions of a protein centered around a defined radius.

Creating the molecular and envelope surfaces

The molecular surfaces of each structure served the most utility throughout this experiment as the difference calculations with vasp used those surf files. The rest of the visualization was completed through pymol with the original pdb files. The surf files were constructed with the provided surfaceExtractor package using the command -surfOutput and a radius of 1.4.

Using VASP to find differences between different protein structures

The vasp package was used to visualize the difference between different molecular surfaces. Two types of comparisons were made - between wild type and mutant types, and between mutant types and the ATP bound pdb files. These calculations required the written python script to reduce the size of the files to create a reasonable computation time.

Irregularities with VASP and surface generation

After numerous trials and errors, we decided not to pursue our efforts with VASP and use surfView to see our output because the resulting shortened pdb files from our reduction method ended up distorting the protein structure, and making the python script unnecessary. The new structures would not properly align with the ATP bound structures, and thus an overlay to prove steric hindrance was not feasible. Because of this, using PyMol was deemed sufficient to fulfill experiment parameters.

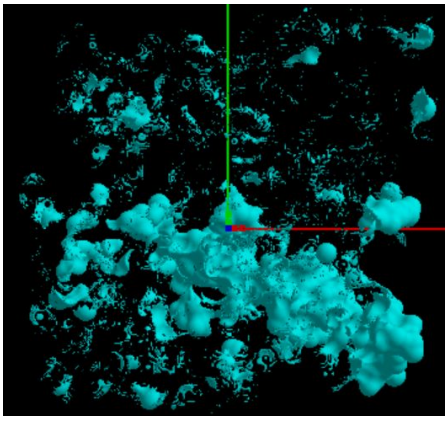

Figure 3: Surf file representing difference between mutated CFTR G551D and wildtype.

Visualizing through pymol

By making use of the 6msm pdb file, we aligned our mutant structures in PyMol on top of the 6msm structure, and observed whether the mutated amino acid (aspartate or glutamate) sterically hindered ATP from fitting into the cavity.

Results

Known Mutation - Aspartate

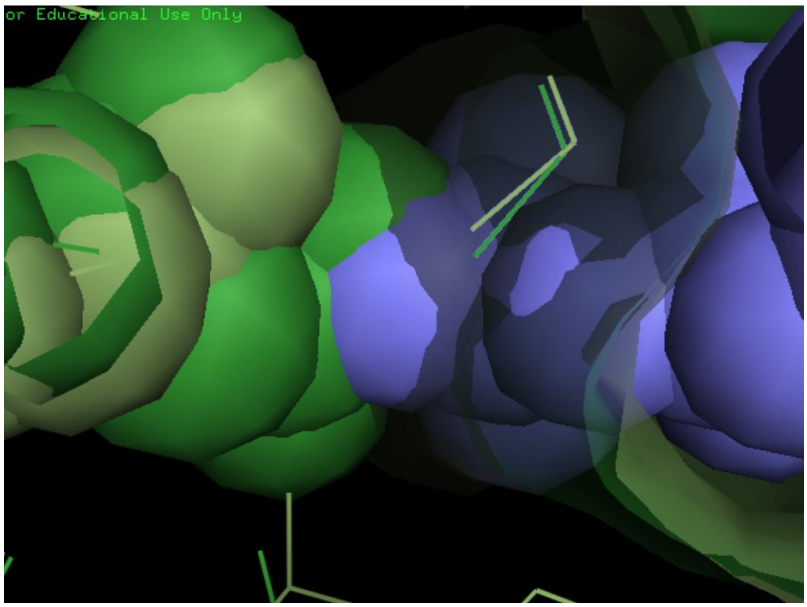

Figure 4: Steric hindrance of ATP (blue) is observed with the alpha carbon of

substituted aspartate at position 551 (dark green)

From figure 4, we were able to confirm what we had found out coming across in our initial research: substitution of glycine to aspartate at position 551 sterically hinders ATP from binding to tyrosine-1219. With this knowledge in hand, we moved forward testing our hypothesis and our control structures.

Hypothesis - Glutamate

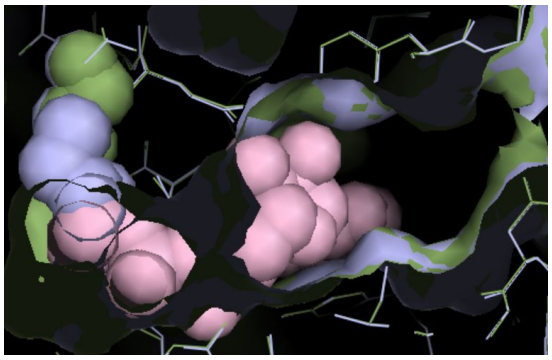

Figure 5: ATP (pink) in the CFTR cavity is shown. Steric hindrance of ATP is also observed with the alpha carbon of substituted glutamate at position 551 (purple)

Figure 5 allowed us to visualize where ATP binds to in the NB2 region of CFTR as well as to observe the accuracy of our hypothesis. Glutamate, like aspartate, is large enough to sterically hinder ATP from binding to Tyr-1219 in NBD2.

Control - Cysteine31

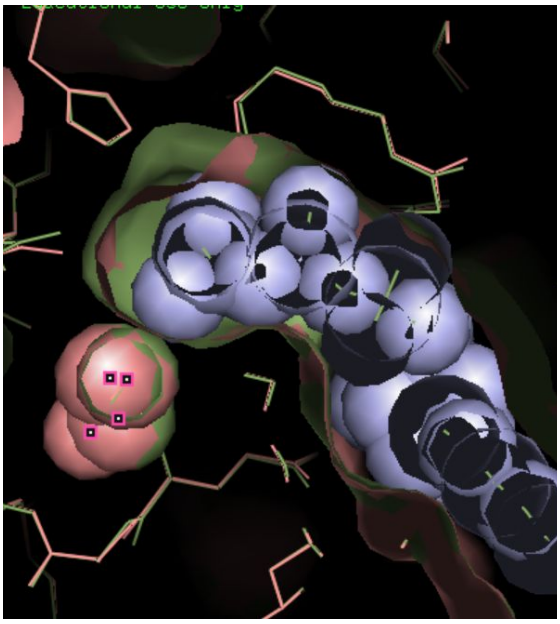

Figure 6: No steric hindrance of ATP (purple) is observed with glycine-551 (pink) in the R31C mutated structure

Control - Threonine148

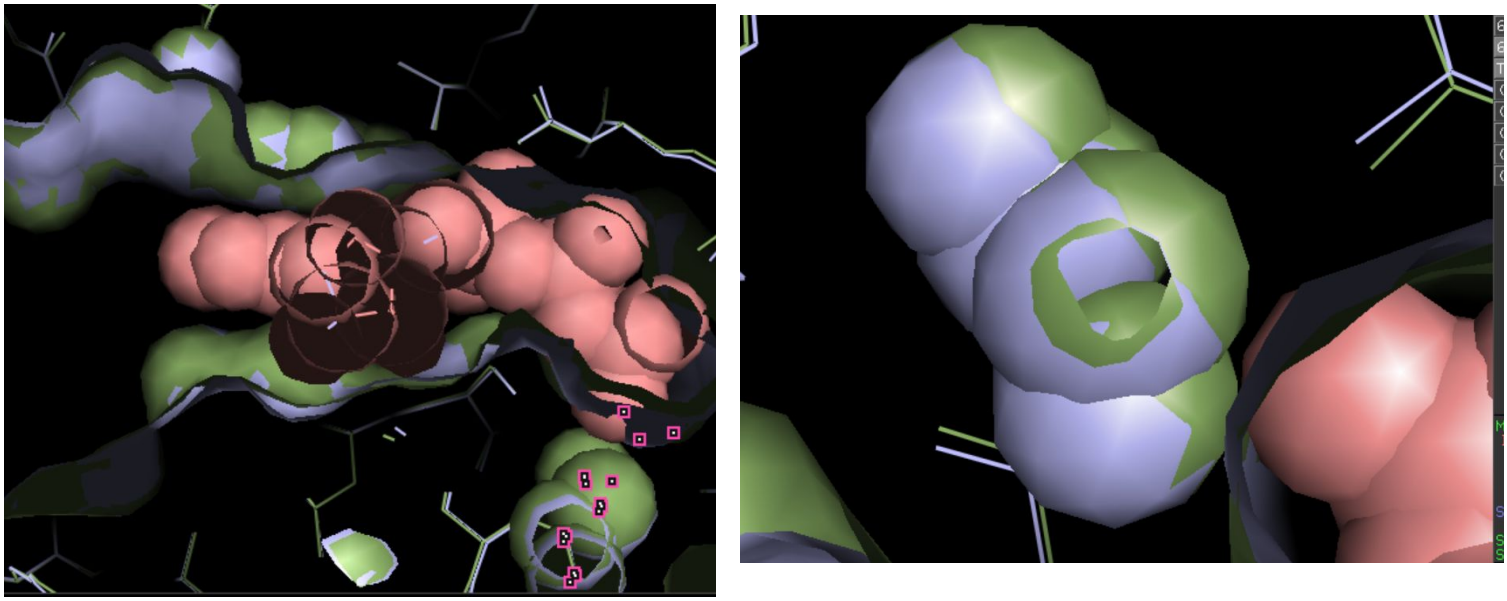

Figure 7: No steric hindrance of ATP (pink) is observed with glycine-551 (green/blue) in the I148T mutated structure

It is important to note that the glycine or rather, the protein with the 551 amino acids, usually has an overlap of 2 colors because it is the 6msm (green) surface structure aligned with the mutant structure of interest. It was necessary to align those two structures since 6msm had the ATP bound structure and we needed to observe whether or not ATP was sterically hindered.

Volumetric Analysis

The volumetric analysis conducted throughout the experiment resulted in the below values. As the volumetric analysis is a measure of the cavity space within these protein structures, these results directly correlate with the hypothesis. The wild type, as the structure with the most empty space for ATP has the largest cavity volume, while the aspartate mutation has the least as the aspartate takes up substantially more space than the glycine.

The control cases reveal that the cavity spaces are tighter, that there is less volume. However it is important to note here that the space that is being affected is not the

functional area that will bind with ATP. The volumetric difference between the wild type and the glutamate mutation is about 450 cubic Angstroms. Considering the volume of a carbon atom is around 120, these results imply a 4-5 atom difference between the structures. The difference between the wild type and the aspartate volume was closer to 1200 cubic Angstroms, indicating a substantially larger difference.

Test Cases

1. Wild type empty space volume: 179679.966616 cubic Angstroms
2. Aspartate mutation volume: 178482.129124 cubic Angstroms
3. Glutamate mutation volume: 179231.469579 cubic Angstroms

Control Cases

1. Final Volume [Cys-31.surf]: 178432.842072 cubic Angstroms
2. Final Volume [Thr-148.surf]: 180044.479592 cubic Angstroms

Discussion

Cystic Fibrosis is a disease caused by the impaired functioning or regulation of the CFTR channel. The channel is composed of 5 different domains and uses ATP hydrolysis to open the channel, and allow the release of chloride ions across epithelial cell membranes. ATP comes in and binds to two different regions (Trp-401 and Tyr-129), located in the NBD1 and NBD2 subunits respectively.

We initially sought to validate the G551D mutation, the most prevalent CFTR mutation occurring in about 5% of patients suffering from cystic fibrosis. In this mutation, glycine-551 is substituted to aspartate, a much larger amino acid. We observed how this substitution to an amino acid, with a much larger side chain, blocks ATP from binding to the NBD2 region.

Once that was proven, we also observed steric hindrance of ATP with the G551E mutation, when glycine-551 is mutated to glutamate. Again, the larger amino acid blocked ATP from binding. Since ATP cannot bind to Tyr-1219, the CFTR channel cannot use energy from ATP hydrolysis to open the channel. As a result, the channel remains closed, blocking the release of chloride ions, which in turn leads to the dysregulation of epithelial fluid transport in the lung, pancreas and other organs, resulting in cystic fibrosis. Therefore, our hypothesis that the G551E mutation will sterically hinder ATP, resulting in a cascade of events leading to cystic fibrosis was validated.

It was important to test some mutations which while might cause a substitution of an amino acid, they do not impair or harm the functionality of the CFTR protein in any way. We tested mutants, R31C and I138T as part of our control study and visualized the interaction of ATP with pymol. These two amino acid positions were chosen because neither of the residues at 31 nor at 138 were close to either of the ATP binding sites in NBD1 or NBD2. As expected, neither of these substitutions caused steric hindrance of ATP and as such, CFTR functionality was not impaired.

Conclusion

The conclusion of the experiment revealed that the initial hypothesis of the project, that the targeted mutation on protein CFTR at position 551 would prevent ATP from bonding at that position, was correct. There were a number of possibilities for this blockage, from electrical interference to steric hindrance. Assessing different types of mutations allowed us to determine that it was not an electric reaction that was preventing the ATP from binding to CFTR when these mutations occurred, but instead a lack of space caused by the slightly larger amino acids. These results were validated when protein structures were closely examined in PyMol.

The second method of assessment involved examining protein structure volumes, specifically of empty space within the structure. The hypothesis predicted that the wild type mutation, with glycine at 551 would have the most space for ATP to bind. This was validated when volumetric analysis revealed that the wild type had the most volume and the mutations had substantially less. When the differences were compared numerically, the gap appears to indicate a difference of a number of alpha carbons, which would result in blocking in the expected positions. These experimental results validate the hypothesis of the experiment and mark it a success.

References

1. Zhou Z, Wang X, Liu HY, Zou X, Li M, Hwang TC. The two ATP binding sites of cystic fibrosis transmembrane conductance regulator (CFTR) play distinct roles in gating kinetics and energetics. *J Gen Physiol*. 2006;128(4):413–422. doi:10.1085/jgp.200609622
2. Hwang TC, Kirk KL. The CFTR ion channel: gating, regulation, and anion permeation. *Cold Spring Harb Perspect Med*. 2013;3(1):a009498. Published 2013 Jan 1. doi:10.1101/cshperspect.a009498
3. *Welcome to CFTR2. CFTR2. https://www.cftr2.org/mutations_history. Accessed December 15, 2019.*
4. MODEL. SWISS. https://swissmodel.expasy.org/. Accessed December 15, 2019.
5. Home Page. Home Page. http://ekhidna2.biocenter.helsinki.fi/. Accessed December 15, 2019.
6. Bompadre SG, Sohma Y, Li M, Hwang TC. G551D and G1349D, two CF-associated mutations in the signature sequences of CFTR, exhibit distinct gating defects. *J Gen Physiol*. 2007;129(4):285–298. doi:10.1085/jgp.200609667
7. Cystic Fibrosis Foundation. Cystic Fibrosis Foundation. https://www.cff.org/. Accessed December 15, 2019. - https://www.cff.org/Care/Clinician-Resources/Network-News/August-2017/Know-Your-CFTR-Mutations.pdf