Cystic fibrosis transmembrane conductance regulator (CFTR)'s evolution revealed that its mutation causes cystic fibrosis

Duo Bai [1, a]

[1] The experimental high school attached to Beijing Normal University

[a] NaHCO3xyuemian@163.com

Abstract. Cystic fibrosis (CF) is a hereditary disease that can lead to death, mainly caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR). These mutations exhibit varying mechanisms and severity levels, posing a challenge for understanding their impact. Here, we hypothesized that mutations occurring in functionally important, conserved regions of the CFTR protein might lead to more severe CF phenotypes, while mutations in other domains may be less harmful. To identify these crucial regions, we conducted a survey of amino acid substitutions in CFTR across primates and non-primate mammals. Intriguingly, we found that most naturally occurring substitutions clustered within two CFTR domains, which paradoxically lacked CF-causing mutations. In contrast, CF mutations in humans are mainly located in the relatively conserved regions of CFTR, namely transmembrane domain 1 (TMD1) and nucleotide binding domain 1 (NBD1). This study presents the distribution of naturally occurring substitutions, and our findings provides a unique perspective for inferring and predicting disease-causing genetic variations by analyzing evolutionary trends.

Keywords: CFTR (Cystic fibrosis transmembrane conductance regulator), Cystic fibrosis, mutation, evolution.

1. Introduction

Cystic fibrosis (CF) is a genetic disease that is inherited in an autosomal recessive manner and poses a severe threat to life. It primarily impacts the cells responsible for producing mucus, sweat, and digestive juices. Under normal circumstances, these secreted fluids are typically thin and lubricating. But in people with CF, a defective gene causes the secretions to become sticky and thick1. Instead of acting as lubricants, the secretions plug up tubes, ducts, and passageways, especially in the lungs and pancreas.

The main organs affected were lung, intestine, pancreas, liver and other cystic fibrosis symptoms, including recurrent chest infection wheezing, cough, shortness of breath and bronchiectasis.

People with this disease also develop a number of related conditions, including diabetes, osteoporosis, and liver problems2.

Cystic fibrosis is a result of genetic mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), which is an ion channel responsible for the transportation of chloride ions (Cl-) across cell membranes. CFTR plays a crucial role in maintaining the balance of salt and water on various surfaces of the body, including the surface of the lungs. More specifically, the CFTR ion channel facilitates the movement of chloride ions from the interior of the cell to the exterior. However, functional mutations may make chloride become trapped in the cell. Without the proper movement of chloride, water cannot moisten the cell surface3. This causes the mucus covering the cells to become sticky4.

CFTR is an anion channel that maintains the correct balance of salts and fluids in the epithelium and other membranes CFTR mutations are responsible for cystic fibrosis, but these mutations can affect the function of CFTR in different ways. Therefore, some drugs used to treat the disease only partially restore the function of specific mutated forms of CFTR8.

Different types of disease-causing mutations lead to insufficient, malfunctioning, or even no production of CFTR proteins. For example, F508del is a common disease-causing mutation seen in CF patients, and it causes less stability in CFTR5. Other mutations (For example, A561E, N1303K) can influence the speed of ion passage, stability, and can lead to different degrees of disease.

The CFTR protein consists of 1,480 amino acid sequences. Once the CFTR protein chain is formed, it folds into a specific 3D shape6.

CFTR protein contains five domains : transmembrane domain 1 (TMD1), nucleotide binding domain 1 (NBD1), R domain, transmembrane domain 2 (TMD2) and nucleotide binding domain 2 (NBD2)

7. The crystal structure of CFTR has been resolved, along with structures carrying various mutations, including F508del.

At present, more than 1800 mutations have been found in CFTR. In healthy cells, the functional CFTR protein is properly transported to the cell membrane.CFTR mutations can be classified into six classes. Class I (G543X) could lead to abnormal protein synthesis. In cystic fibrosis, patients carrying the class II (ΔF508) mutation gene account for 88.5% of the total. This mutation results in a blockage of the CFTR maturation process and degradation of misfolded proteins in the endoplasmic reticulum. Class III (G551D) will affect the regulatory ability of CFTR protein, causing the channel formed by CFTR protein to be non-functional.

Class IV (R117H) reduces the conductivity of chloride ions. Class V (A455E) will reduce the number of functional CFTRs. Class VI (Q1412X) will result in a decrease in the stability of CFTR on the cell membrane. 9.

CFTR is a conserved gene through evolution and certain functional regions may witnesses more natural mutations than others. The regions that determine key functional aspects of the channel and are supposed to be more conserved, while the other regions are more subjective to changes. Therefore, by examining the sequence changes along evolution, we expect to identify the “hotspots” for accumulated natural mutations and quiet regions which are supposed to be functional conserved8. We can then correspond the locations of the disease-causing mutations to either mutation hotspots or conserved regions. Our hypothesis is that more severe CF-causing mutations reside in the evolutionary conserved regions on CFTR.

In this study, we characterize both disease-causing mutations on human CFTR and natural substitutions seen in the mammalian evolution of CFTR, and compare the position and frequency of the mutations, to infer a relationship between gene and function. Specifically, we first conducted a literature review and summarize common mutations (substitutions and indels) in human CF patients and locate them in different gene domains. Meanwhile, we collected mammalian CFTR sequences, with a focus on primates. By performing sequence alignment and constructing a phylogenetic tree, we obtained information on CFTR substitutions during evolution. By comparison, we can see most CF-causing mutations reside in evolutionarily conserved regions, while the regions harboring more naturally occurred mutations have not reported any CF-causing mutations in CF patients.

2. Methods

2.1 Obtaining data

The amino acid sequence of CFTR in healthy humans was obtained through the protein database PDB (https://www.rcsb.org/). We chose the 3D crystal structure “5UAK”4 as it had the highest resolution 3.87å.

In order to obtain the amino acid sequence of other mammalian species, we refer to the human CFTR sequence, query the NCBI database through protein blasting (Blastp), and default parameters.

We further chose 11 representative species, including six primate species (Chimpanzee, Pan paniscus; Gorilla, Gorilla gorilla; Orangutan, Pongo abelii; Gibbon, Nomascus leucogenys; Rhesus monkey; Marmoset *Callithrix jacchus*, Lemur, *Lemur catta*) and four non-primate species (Rabbit, *Oryctolagus cuniculus*; Pangolin, *Manis javanica*; Horse, *Equus caballus*; Pig, *Sus scrofa*), and an outgroup species chicken, *Gallus gallus*.

For the chosen CFTR sequences, we downloaded the FASTA data of the amino acid sequences, organized them in one file, and changed the sequence headers to a common name plus a scientific

name, eg. Human_homo_sapiens.

2.2 Sequence aligning

We visualized and aligned the 13 sequences in SeaView using the method “clustalo”. With the aligned sequences, we built the phylogenetic tree construction using PhyML (maximum likelihood) method, and add “Br support” for lineage reconstruction confidence (1.00 for very confident, 0 for not confident).

We summarized the substitutions of 12 typical mammal CFTR sequences against human CFTR sequences, and pairwise compared the mammal sequences with human CFTR sequences.

2.3 Divergence time calculation

To study the differentiation time between selected mammal species and humans, we visit the Timetree website. (http://timetree.org), and enter the species name and the human name to get the results.

It was calculated the length of each domain of CFTR protein and normalize the number of mutations per amino acid on every domain. We computed the mutation rate of each mammal of each domain, which is average number of mutations per amino acid per million years.

2.4 The Active Pocket of the CFTR Pathogenic Domain

The 3D structure of CFTR (5UAK) was obtained from the PDB database. After preprocessing steps such as water removal and elimination of small molecules using PyMol software, the sizes, surface areas, and drugability scores of potential binding pockets for each structural domain were predicted using the online tool DoGSiteScorer from proteins plus (https://proteins.plus).

In this way, we can predict the binding mode with small molecule compounds.

2.5 The ANOVA test

To find out whether there is a single variable influence the average value of mutation rate, we did the one-way ANOVA test. We downloaded the latest edition of the R studio(https://rstudio.com/) to do the statistical analysis of the locations of CF-causing mutations and natural mutations.

We organized the data of 11 different mammals into two columns: mutation rate and the domain name and performed the one-way ANOVA test in R. Last, we visualized the mean and standard deviation by CFTR domain by drawing bar plot.

3. Results

3.1 Analysis of the 3D Structure and Functional Domains of CFTR

To gain insights into disease-causing mutations in CFTR, we conducted a literature search 1 and compiled a summary of frequently observed mutations in CF patients, along with their impacts on CFTR functions (Table 1). Meanwhile, we referred to the crystal structure of healthy human CFTR (Figure 1) for the determination of the five functional domains: TMD1 (81-350), NBD1 (441-576), R domain (639-849), TMD2 (862-1147), and NBD2 (1227-1347). To serve our purpose of corresponding disease-causing mutations to conserved and less conserved regions on CFTR, we further pin down each mutation to their location on the protein (Table 1).

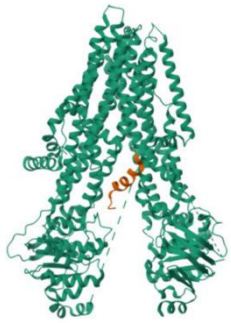


Figure 1. The human CFTR 3D crystal structure “5UAK” as it had the highest resolution 3.87å. There are 12 pathogenic mutations in human CFTR, all of which occur in the TMD and NBD domains.

Specifically, we found that all the disease-causing mutations occurred in the transmembrane domain and nucleotide-biding domain, and the most common F508del in NBD1 causes no Cl- transported through CFTR and less stability of the channel. Additionally, three mutations reported to cause no function of CFTR all occur on the nucleotide-binding domains where S549R and G551D belong to NBD1 and G1329D resides in NBD2. Also, two mutations that cause no CFTR protein all occur on the nucleotide-biding domains, where G542X happens on NBD1, and W1282X happens on NBD2.

Table 1. A summary of reported CF-causing mutations and CFTR defect types

CFTR mutations	CFTR domain	CFTR defect type
R117H	TMD1	Less function
120del23	TMD1	Less stable
R334W	TMD1	Less function
A455E	NBD1	Less function
F508del	NBD1	No ion traffic, less stable
G542X	NBD1	No protein
S549R	NBD1	No function
G551D	NBD1	No function
A561E	NBD1	No ion traffic
W1282X	NBD2	No protein
N1303K	NBD2	No ion traffic
G1329D	NBD2	No function

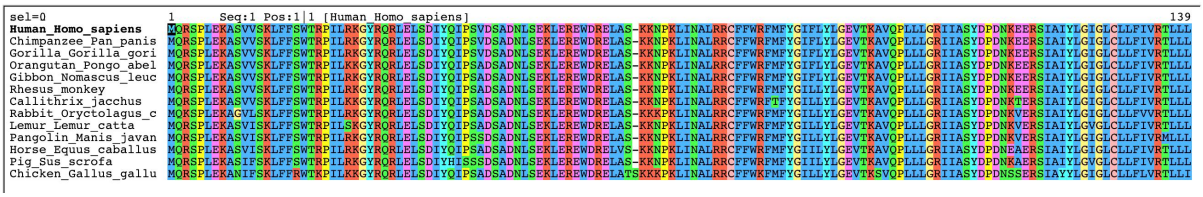
3.2 Results of sequence comparison of human CFTR with various mammalian sequences

After aligning the sequences in SeaView（Figure 2A）and building a phylogenetic tree（Figure 2B）,we summarized the mutations results by comparing the amino acid sequence of CFTR of each animal with that of the human.

After that, we compare the sequence tree with the evolutionary species tree.

The distribution of naturally occurred mutations in chosen mammals basically follows the trajectory of evolution, which is reflected in the fact that there are fewer differences between the natural mutations that occurred in primates and that of humans.

A



B

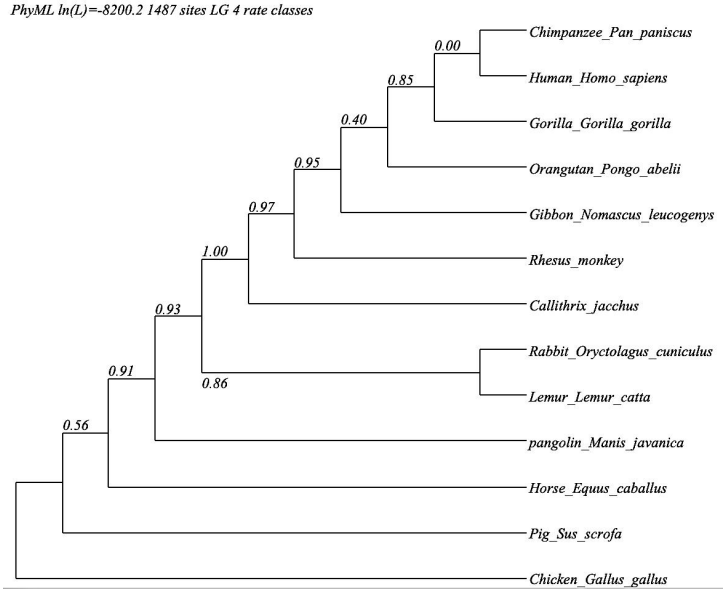


Figure 2. sequencing alignment and the phylogenetic tree. (A) The sequencing alignment in SeaView.(B) The phylogenetic tree of the amino acid sequences of CFTR in healthy human and other selected mammals.

3.3 The situation of CFTR structural domain mutations

From the pattern in the Table 2, we found that NBD 1 and NBD 2 carries few natural mutations. Specifically, in primates, there is only one amino acid V471M that is different from humans. NBD1 also harbors the least number of natural mutations among all the domains, seconded by NBD2.

What's more, whereas domain R and TMD2 see the most numbers of naturally occurring mutations, there is no reported CF-causing mutations in human in these two domains. For those disease-causing mutations that happened in humans, only S257N occurs in natural mutations in some mammals, including rabbit, pangolin, and pig.

Table 2. Natural Mutations in Mammalian CFTR Compared with Human CFTR

	Chimpanzee	Gorilla	Orangutan	Gibbon	Rhesus monkey	Marmoset	Lemur	Rabbit	Pangolin	Horse	Pig	CF-causing mutations in humans
TMD1（81-350）			S257N I333V	I333V	F88L I333V	M83T F88L E116T S257N	F88L E116V T165I A210M I216L L234I S257N	F88L E116V I133V V182I A205S A210T I214L G229A F237V S257N I329T	F88L E116V T136M A210T I216L S223F A224T G229A F237L A253G S257N M285I L321F F343Y	F88I K115E E116A A210T E218D G229A S256N I333V	F88I E116A I126V L146P A218T G160A I232V R243K S257N D29L I333M	R117H R334W S257N
NBD1（441-576）	V471M	V471M	V471M	V471M	V471M	V471M	V471M V511I	V471M R519K A535T I540T	V471M R519T I547V	N446S V471M T548Q	N446S V471M	A455E F508del G542X S549R G551D A561E
R（639-849）		H668R	H668R S687P L735F T764M H779Q	H668R Q717P R755H V782I N787S	C648Y H668R N785S T791A	C648Y A656S H668R P706S S758N T762A A765V L774M N785S	C648Y H668R P677A T681N T683P F694I P706S I710M I715V N724S S731P L733S I756S S757N S760N L765F A767G Q786P K791I S795A V799M L801I Q804P T808S L811M T822S G823S M843V S845N	C648Y H668R P677S V678I T681N E682D K684R T691N F694L I707V T710M Q717P E733A P734S L735I I756S S757N V758M S760N T764M A767G R768C N783S Q786P N787S H789Y K791R A794T T796A V799M A805T T822S F840I M843V E844D A848T	C648F D649E H668R P677A S679T T681N K685R D730N I756S S757N S760N L765F A767G L776M H779C I786T T792K T793A V799M L811T T822S M843V A848T	C648Y H668R P677T T681N E696D I756S S757N S760N L765F A767R H779R N787S T793A V799M L811M E821D T822S M843V A848T	C648Y S650T L663I H668R P677S T681N I726F D732G Q748H I756S S757N S760N T761A L765F A767G H779R N787S T793A A794T V799M A802V L811E E821D K836R M843V A848T	
TMD2（862-1147）			H956N	I1115M V1135I	L887F L893P H901Y S913R R132T	L893F H901Y S913N M933L A1013S V1021L V1028A I1031V M1032L E1050A R132T	A882V S883F V886L L888S L890F N892K T893A L895P Q896Y D897N K898N S901N H903V S904G R905P S918A P922L M935L F938I Q964R T972A I1011V A1012V A1015S V1023L A1031S M1034I Q1041H R1134Q V1135I S1147G	K863R L890F P894A L895P G899E H903K R905G N907S S915N V921F M935L F937L A1015S V1023L V1030A M1034L Q1041H T1121A	A878V G891R N892K L895K H903K S904N R905A Y919F M935L F937V G977A I1011V A1015S V1023L V1030A M1034L Q1041H P1078S A1087T R1134T	G891K N892E L895P K898S H903K S904G R905A M935L F937L I945M A1015S V1023L V1030A M1034I Q1041H R1134T	I875V S883C W888C G891K N892K P894S L895P H903K S904G R905A S918A T931G M935L F937L H956R I1011V A1015S Q1038K V1023L M1034L Q1041H V1062I R1134T	
NBD2（1227-1347）				F1265L	A1230P	S1302T	T1228I E1229D G1230S Q1276H	T1228I E1229D G1230S	E1229D I1242V F1265L I1277V Q1289E E1316G S1319N K1325T	E1229D E1316G S1319N	T1228V E1229D I1234V S1263L I1277V E1316G S1319N D1328E	W1282X N1303K G1329D

Specifically, the mammal with the most natural substitutions on mammalian CFTR compared with humans is the lemur, with 73 substitutions; and the animal with the least number of substitutions compared with humans is the chimpanzee's, with only 1 substitution.

The results of divergence time calculation are shown in Table 3 and Table 4.

Table 3. The number of mutations per amino acid in each animal and each domain

Table 3	Chimpanzee Pan pa	Gorilla gorilla	Orangutan Pongo abelii	Gibbon Nomascus leucogenys	Rhesus mulatta	Macaca	Marmoset Callithrix jacchus	Lemur Lemur catta	Rabbit Oryctolagus cuniculus	PigSus scrofa	Pangolin Manis javanica	Horse Equus caballus
Divergence time (Mya)	6.4	8.6	15.2	19.5	28.8	43	74	87	94	94	94	
TMD1 (270 aa)	0.00E+00	0.00E+00	7.41E-03	3.70E-03	7.41E-03	1.48E-02	2.59E-02	4.07E-02	4.07E-02	5.19E-02	2.96E-02	
NBD1 (137aa)	7.30E-03	7.30E-03	7.30E-03	7.30E-03	7.30E-03	7.30E-03	1.46E-02	2.92E-02	1.46E-02	2.19E-02	2.19E-02	
R (211aa)	0.00E+00	4.74E-03	2.37E-02	2.37E-02	1.90E-02	4.27E-02	1.37E-01	1.71E-01	1.23E-01	1.09E-01	9.00E-02	
TMD2 (287aa)	0.00E+00	0.00E+00	3.48E-03	6.97E-03	1.74E-02	3.83E-02	1.08E-01	6.27E-02	8.01E-02	6.97E-02	5.57E-02	
NBD2 (121aa)	0.00E+00	0.00E+00	0.00E+00	8.26E-03	8.26E-03	8.26E-03	3.31E-02	2.48E-02	6.61E-02	6.61E-02	2.48E-02	

Table 4. The average mutation rates on each domain in each mammal

Table 4	Chimpanzee Pan paniscus	Gorilla Gorilla gorilla	Orangutan Pongo abelii	Gibbon Nomascus leucogenys	Rhesus Macaca mulatta	Marmoset Callithrix jacchus	Lemur Lemur catta	Rabbit Oryctolagus cuniculus	PigSus scrofa	Pangolin Manis javanica	Horse Equus caballus	CF-causing mutations in humans	Average mutation rate
Divergence time (Mya)	6.4	8.6	15.2	19.5	28.8	43	74	87	94	94	94 N/A		
TMD1 (270 aa)	0.00E+00	0.00E+00	4.87E-04	1.90E-04	2.57E-04	3.45E-04	3.50E-04	4.68E-04	4.33E-04	5.52E-04	3.15E-04	3	0.0003
NBD1 (137aa)	1.14E-03	1.14E-03	4.80E-04	3.74E-04	2.53E-04	1.70E-04	1.97E-04	3.36E-04	1.55E-04	2.33E-04	2.33E-04	6	0.0004
R (211aa)	0.00E+00	7.41E-04	1.56E-03	1.22E-03	6.58E-04	9.92E-04	1.86E-03	1.96E-03	1.31E-03	1.16E-03	9.58E-04	0	0.0011
TMD2 (287aa)	0.00E+00	0.00E+00	2.29E-04	3.57E-04	6.05E-04	8.91E-04	1.46E-03	7.21E-04	8.53E-04	7.41E-04	5.93E-04	0	0.0006
NBD2 (121aa)	0.00E+00	0.00E+00	0.00E+00	4.24E-04	2.87E-04	1.92E-04	4.47E-04	2.85E-04	7.03E-04	7.03E-04	2.64E-04	3	0.0003

Through the analysis of Table 3 and Table 4, we found that the average mutation rate during evolution is inversely proportional to the number of mutations that cause CF in humans.

Specifically, the most CF-causing mutations in humans happens on NBD1 domain. Also, R domain and TMD2 domain has no CF-causing mutations.

To analyze the divergence trend of mammals in comparison to humans, we created a trend plot using the data from Table 3. We performed regression analysis using the function available in MS Excel, resulting in the creation of Figure 4.

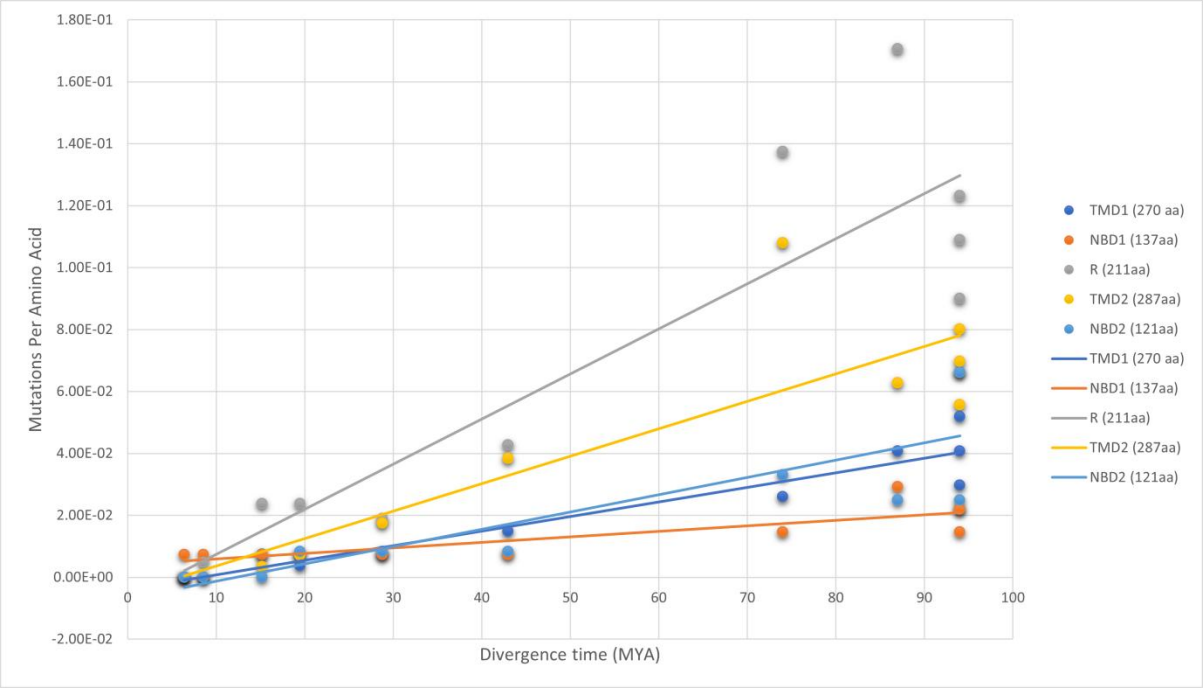


Figure 4. The mutation rate and linear regression in each domain.

In this linear regression graph, the slope of linear regression of R domain is the biggest, and then TMD2's, NBD2's, TMD1's, and that of the NBD1. The smallest slope is the line of NBD1. Thus, we can see that through the evolution progress in chosen mammals, the R domain accumulates more genetic variations, and NBD1 changes more slowly. Hence, it shows that NBD1 is relatively more conserved in those 5 domains in mammal CFTR. Also, most disease-causing mutations occur on NBD1 and TMD1. This is reasonable because mutations should happen in the most conserved area to cause disease. For TMD2 and R domains, there are no disease-causing mutations because the change of those two domains might be deleterious.

3.4 The analysis on CF-causing mutations and natural mutations

To find out whether there exists a single variable that makes the results significant different with the average value, we did a hypothesis test analysis of the locations of CF-causing mutations and natural mutations using the data in Table 4:

H_0: There doesn't exist a domain that influence the average mutation rate.

H_A: There exist a domain that influence the average mutation rate.

From the test, we found out that the P value of the domain equals to 1.71×10^{-5}, much smaller than $\alpha = 0.001$, which means that we should reject H0, that there is one domain significant enough to influence the whole test result. Then we calculated the mean and standard deviation of mutation rates by domain, and making into the Bar plot. (Figure 5)

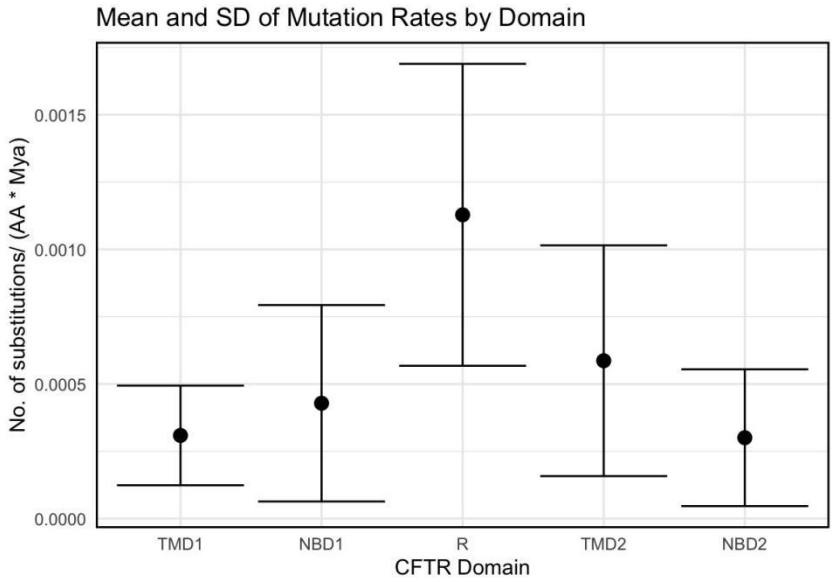


Figure 5. Average and standard deviation of mutation rate divided by domain.The mean value of R domain is the biggest, which is 0.0011, with the standard deviation=0.00056. That testified the results that the mutation rate of R domain is the biggest among the five domains, and with the biggest range of variation.

3.5 Drug Combination Active Pocket Prediction

We used DoGSiteScorer to predicts the potential binding pockets of the CFTR structural domain crystal structure (Figure 6), ranks them based on their drugability scores, and conducts a detailed analysis of the positions and related properties of the top five ranked active pockets (Table 5)."

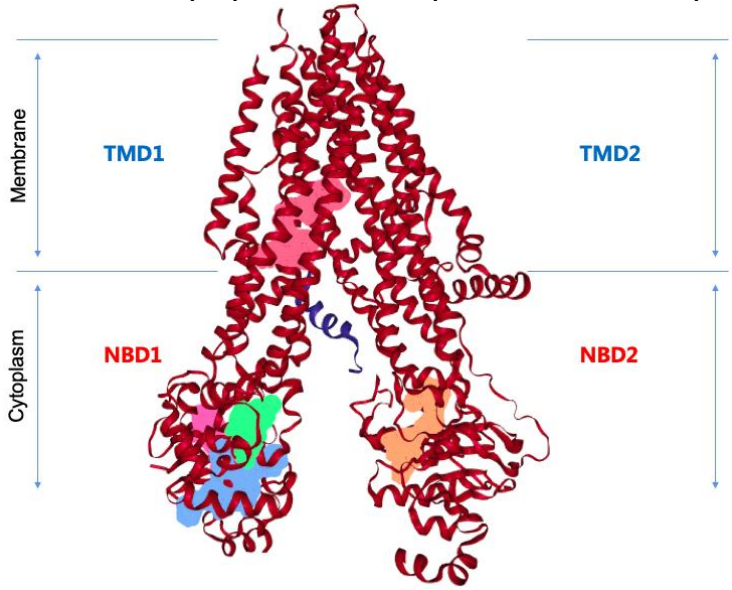


Figure 6. The predict drug pocket of CFTR

Table 5. The properties of the top five active pockets of each domain predicted by DoGSiteScorer

Pocket No.	color	Volume Å³	Surface Å²	Drug Score	Simple Score
1		708.75	1090.02	0.88	0.44
2		660.95	865.99	0.88	0.4
3		663.52	853.77	0.87	0.36
4		466.01	619.41	0.87	0.18
5		436.81	566.48	0.84	0.17

We can see that the Pocket No.1, which locates at NBD1, has the highest drug score. Thus, it may be possible to develop drugs based on pocket No.1.

4. Discussions

4.1 The significance of studying CF

Cystic fibrosis (CF) is a hereditary disorder characterized by the autosomal recessive inheritance pattern, primarily impacting mucus and sweat cells, resulting in the involvement of multiple organs.Among these organs, the lungs are most severely affected, resulting in death for 90% of patients. CF is associated with various complications, including respiratory system complications, infertility, osteoporosis, and electrolyte imbalances[10].

4.2 CFTR mutations in human

CF is the most common genetic disorder in Caucasians, so the CFTR protein, which is responsible for cystic fibrosis, is one of the most widely studied mutation profiles. Disease-causing exonic mutations in CFTR are distributed extensively throughout the coding sequence of the protein. These mutations are present in the coding region of all five domains of the CFTR protein, including two transmembrane domains (TMD 1 and 2), two nucleotide binding domains (NBD1 and 2), and regulatory (R) domains. Mutations in the coding region of NBD1 and TMD1 may have a higher likelihood of causing the disease. 11-12. Understanding the specific CFTR mutations in a patient's genome is critical for tailoring treatment. Some mutations respond well to specific therapies, such as modulator drugs that correct the defective CFTR protein's function. Studying on CF and CFTR is essential for developing new and more effective treatments.

4.3 Potential of future drug development of CFTR and the active pocket

Understanding the structure of CFTR's active pockets enables scientists to develop more targeted drugs. Medications can be designed to either restore or enhance the normal function of the CFTR protein to treat CF or to promote the normalization of the defective CFTR protein. Additionally, many CF patients carry mutated CFTR genes.

Through computer simulation, we have the capability to identify potential drugs that target CFTR through molecular docking. This approach allows us to make predictions regarding the binding affinity between small molecules and CFTR proteins, offering valuable insights into potential drug candidates.By virtually screening a vast number of compounds, we can efficiently narrow down the pool of potential drugs for further experimental validation. Computer simulations offer a cost-effective and time-efficient method for drug discovery, enabling researchers to explore a wide range of chemical space and accelerate the development of novel therapies for CFTR-related diseases. By comprehending the impact of different mutations on the active pockets, scientists can

tailor drug correction strategies specific to particular mutations, offering personalized treatment options to a broader range of patients.

4.4 Future Research with Animal Experiments

In the future, it may be possible to utilize mouse models for the study of CFTR protein. Mice are a commonly used research animal model, and scientists can employ gene editing techniques to create CFTR mutant mice, mimicking human CFTR mutations, and observe their physiological and pathological characteristics. These mouse models can be employed to test new therapeutic approaches and gain insights into CFTR functionality. Animal models for studying CFTR can be instrumental in evaluating novel drug candidates, assessing their impact on the CFTR channel, and exploring potential therapeutic efficacy.This will help to improve the drug development and CF therapy.

Acknowledgement

I would like to express my heartfelt gratitude to several individuals who have played a pivotal role in the completion of this paper.

First and foremost, I extend my sincere appreciation to my dedicated biology teacher, Fang Wei. Your guidance, invaluable advice, and unwavering support have been instrumental in shaping the direction of this research. Your profound knowledge and commitment to teaching have not only enhanced my understanding of the subject but have also inspired me to strive for excellence.

I would also like to thank my classmates and colleagues for their constructive feedback and discussions, which significantly contributed to the refinement of this paper.

Lastly, I am deeply thankful to my family for their constant encouragement and belief in my abilities.

This paper would not have been possible without the collective support and mentorship of these individuals, and for that, I am truly grateful.

References

[1] Klimova, Blanka, et al. “Cystic Fibrosis Revisited – a Review Study.” *Medicinal Chemistry*, vol. 13, no. 2, 30 Jan. 2017, pp. 102–109, pubmed.ncbi.nlm.nih.gov/27292156/, https://doi.org/10.2174/1573406412666160608113235.

[2] Buckland, Danny. “Recent Advances in the Management of Cystic Fibrosis.” *Prescriber*, vol. 27, no. 11, Nov. 2016, pp. 19–24, https://doi.org/10.1002/psb.1512.

[3] Fanen, Pascale, et al. “Genetics of Cystic Fibrosis: CFTR Mutation Classifications toward Genotype-Based CF Therapies.” *The International Journal of Biochemistry & Cell Biology*, vol. 52, July 2014, pp. 94–102, www.sciencedirect.com/science/article/pii/S1357272514000764, https://doi.org/10.1016/j.biocel.2014.02.023.

[4] Liu, Fangyu, et al. “Molecular Structure of the Human CFTR Ion Channel.” *Cell*, vol. 169, no. 1, Mar. 2017, pp. 85-95.e8, https://doi.org/10.1016/j.cell.2017.02.024.

[5] NHS. “Overview - Cystic Fibrosis.” *NHS*, 16 Mar. 2021, www.nhs.uk/conditions/cystic-fibrosis/.

[6] Rafeeq, Misbahuddin M., and Hussam Aly Sayed Murad. “Cystic Fibrosis: Current Therapeutic Targets and Future Approaches.” *Journal of Translational Medicine*, vol. 15, no. 1, 27 Apr. 2017, translational-medicine.biomedcentral.com/articles/10.1186/s12967-017-1193-9, https://doi.org/10.1186/s12967-017-1193-9.

[7] Wang, Shusheng, and Shelley D Minteer. *Molecular Studies of CFTR Interacting Proteins*. 1 Jan. 2001, https://doi.org/10.1007/s004240100646. Accessed 19 May 2023.

[8] Levring, Jesper, et al. “CFTR Function, Pathology and Pharmacology at Single-Molecule Resolution.” *Nature*, vol. 616, no. 7957, 1 Apr. 2023, pp. 606–614, www.nature.com/articles/s41586-023-05854-7, https://doi.org/10.1038/s41586-023-05854-7. Accessed 21 Apr. 2023.

[9] Gawenis, Lara R., et al. “A BAC Transgene Expressing Human CFTR under Control of Its Regulatory Elements Rescues Cftr Knockout Mice.” *Scientific Reports*, vol. 9, no. 1, 14 Aug. 2019, https://doi.org/10.1038/s41598-019-48105-4. Accessed 11 Sept. 2020.

[10] Rishishwar L, Varghese N, Tyagi E, Harvey SC, Jordan IK, McCarty NA. Relating the disease mutation spectrum to the evolution of the cystic fibrosis transmembrane conductance regulator (CFTR). PLoS One. 2012;7(8): e42336. doi:10.1371/journal. pone. 0042336

[11] Rafeeq MM, Murad HAS. Cystic fibrosis: current therapeutic targets and future approaches. J Transl Med. 2017;15(1):84. Published 2017 Apr 27. doi:10.1186/s12967-017-1193-9

[12] Guggino WB, Banks-Schlegel SP. Macromolecular interaction and ion transport in cystic fibrosis. Am J Respir Crit Care Med. 2004; 170: 815–820. doi: 10.1164/rccm.200403-381WS.