



**République Algérienne Démocratique et Populaire**



**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

**Université de Blida 1**

**Faculté des Sciences de la Nature et de la Vie**

**Département de biologie et physiologie cellulaire**

**Laboratoire de biotechnologie Environnement et santé**

**Mémoire de fin d'étude en vue de l'obtention**

**Du diplôme de Master**

**Option : Génétique**

**Thème**

**Étude in silico des mutations du gène CFTR provoquant  
la variabilité des formes de mucoviscidose**

**Présenté par**

Mlle DJIDJELI Imène

**Devant le jury**

Mme ROUAKI .F	MCB	USDB1	Présidente
Mme MOKRANE.A	MAB	USDB1	Examinatrice
Mme ABDUL HUSSAIN.A.S	MCB	USDB 1	Promotrice
Mme AMOKRANE.A	MAA	USDB 1	Co-promotrice

**Promotion 2018-2019**

# **Remerciements**

*Je tiens tout d'abord à remercier **Dieu** le tout puissant qui m'a ouvert les portes du savoir et qui m'a donné la force et la volonté de poursuivre mes études.*

*Je remercie tout particulièrement ma promotrice **Mme ABDUL HUSSAIN. A. S** ; pour la confiance qu'elle m'a accordé en acceptant d'encadrer ce travail. Je lui suis reconnaissante pour sa patience, ses encouragements, sa sympathie et sa disponibilité à tout moment. Ses conseils, ses commentaires, ses corrections et ses qualités scientifiques ont été très précieux pour mener à bien ce travail. Je suis infiniment heureuse et honorée d'avoir fait ce travail sous sa direction. Aucune expression de gratitude ne sera suffisante pour vous exprimer mon respect et ma reconnaissance.*

*J'adresse mes plus sincères remerciements à ma Co promotrice **Mme AMOKRANE .A** pour avoir accepté de co-diriger ce mémoire et surtout pour sa gentillesse, Cela été un réel plaisir de travailler avec vous.*

*Mes vifs remerciement à **Mme ROUAKI .F** d'avoir voulu me faire l'honneur de présider ce jury.*

*Mon remerciement va également à **Mme MOKRANE. A** pour avoir aimablement accepter d'examiner ce travail et d'en faire l'analyse critique.*

# *Dédicaces*

*Je dédie ce travail*

*À l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde à moi, à toi mon père.*

*À la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur ; maman que j'adore*

*Aux personnes dont j'ai bien aimé la présence dans ce jour, à tous mes frères **Adel, Mohamed, Islem** à ma belle-sœur **Asma**, ma cousine **Hanane** je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements.*

*Pour les petits de la famille ma nièce **Alae**, et mon neveu **Abd -elrahim**.*

*Aux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés ,mes aimables amie : **Ahlem, Sara ,Latifa** , **Sarah, kaouther, Hassina, Imène**.*

# Liste des figures

<b>Figure 1:</b> Arbre généalogique d'une famille atteinte d'une maladie autosomique dominante. ....	4
<b>Figure 2:</b> Arbre généalogique d'une famille atteinte d'une maladie autosomique récessive .....	4
<b>Figure 3:</b> Arbre généalogique d'une maladie recessive liée à X .....	5
<b>Figure 4:</b> Caryotype humain normal.....	6
<b>Figure 5:</b> Schéma représente le principe de la FISH.....	7
<b>Figure 6 :</b> Représentation schématique de la stratégie de diagnostic par séquençage capillaire dit sanger. ....	7
<b>Figure 7:</b> Schéma représente le principe de RFLP .....	8
<b>Figure 8:</b> Principe de la CGH array sur puce à ADN .....	8
<b>Figure 9:</b> Différentes étapes du typage NGS .....	9
<b>Figure 10:</b> Prévalence approximative de la mucoviscidose dans certaines pays	13
<b>Figure 11:</b> Schéma représente l'organisation du gène CFTR. ....	14
<b>Figure 12:</b> Structure de la protéine CFTR.....	15
<b>Figure 13:</b> Organes touchés par la mucoviscidose .....	17
<b>Figure 14:</b> Différentes classes des mutations CFTR. ....	19
<b>Figure 15:</b> Utilisation de GenBank.....	29
<b>Figure 16:</b> Tableau de bord de la requête sur NCBI.....	30
<b>Figure 17:</b> Utilisation de BLAST sur la plateforme NCBI. ....	31
<b>Figure 18:</b> Utilisation d'ORF Finder .....	32

<b>Figure 19:</b> Utilisation d'Ensembl .....	33
<b>Figure 20:</b> Utilisation de swiss-model.....	33
<b>Figure 21:</b> Séquence d'ARNm du gène CFTR sous format FASTA.....	34
<b>Figure 22:</b> Représentation topographique des différents isoformes du gène CFTR sur la base de données swiss-model .....	36
<b>Figure 23:</b> Séquence protéique du gène CFTR. ....	37
<b>Figure 24:</b> Structure 3D de la protéine CFTR. ....	37
<b>Figure 25:</b> Structure moléculaire 3D de cholestérol sur le logiciel swiss-model.	38
<b>Figure 26:</b> Représentation 3D des ions de magnésium sur swiss-model .....	39
<b>Figure 27:</b> Représentation 3D de la liaison des deux molécules d'ATP sur les deux domaines NBD1-NBD2 sur swiss-model.....	40
<b>Figure 28:</b> Représentation 3D des deux domaines transmembranaires de la protéine CFTR. ....	40
<b>Figure 29:</b> Variantes du gène CFT dans swiss model .....	41
<b>Figure 30:</b> Domaines de la protéine CFTR et les deux molécules d'ATP.....	41
<b>Figure 31:</b> Modèle théorique de la structure de la protéine CFTR sauvage.....	42
<b>Figure 32:</b> Dot matrix view .....	43
<b>Figure 33:</b> Délétion des trois nucléotides CTT dans la position 1521-1523 sur l'exon 10 .....	44
<b>Figure 34:</b> Traduction de la séquence non mutée de l'exon 10 .....	45
<b>Figure 35:</b> Traduction de la séquence mutée de l'exon 10 .....	45
<b>Figure 36:</b> Alignement de la séquence protéique de l'exon 10.....	46
<b>Figure 37:</b> Résultats d'alignements nucléotidique et protéique de l'exon 10 avec Anagène.....	47

<b>Figure 38:</b> Structure 3D de la protéine de l'exon 10 sauvage et mutée (F508 del).....	47
<b>Figure 39:</b> Délétion de la Phénylalanine à la position 508.....	48
<b>Figure 40:</b> Représentation tridimensionnelle de la protéine CFT sauvage et mutée (F508del). ....	49
<b>Figure 41:</b> Substitution d'une Thymine à la position 680 de l'exon 6.....	50
<b>Figure 42:</b> Traduction de la séquence non mutée de l'exon 6 .....	50
<b>Figure 43:</b> Traduction de la séquence mutée de l'exon 6. ....	51
<b>Figure 44:</b> Alignement de la séquence protéique de l'exon 6.....	52
<b>Figure 45:</b> Structure 3D d'exon 6 et la substitution de la Leucine en Arginine à la position 227 .....	52
<b>Figure 46:</b> Substitution d'une Leucine à la position 227 .....	53
<b>Figure 47 :</b> Substitution d'une Guanine à la position 1624 de l'exon 12.....	54
<b>Figure 48:</b> Traduction de la séquence non mutée de l'exon 12 .....	54
<b>Figure 49:</b> La traduction de la séquence mutée de l'exon 12 .....	55
<b>Figure 50:</b> Alignement protéique local de la séquence protéique de l'exon 12(G542X).....	56
<b>Figure 51:</b> Structure 3D de l'exon 12 et le changement de Glycine par un codant Stop.....	57
<b>Figure 52:</b> Substitution d'une Cytosine à la position 3909 sur l'exon 24.....	58
<b>Figure 53:</b> Traduction de la séquence non mutée de l'exon 24 .....	59
<b>Figure 54:</b> Traduction de la séquence mutée de l'exon 24 .....	59
<b>Figure 55:</b> Alignement protéique de l'exon 24(N1303K). ....	60
<b>Figure 56:</b> Structure 3D de l'exon 24 et le changement d'une Asparagine en	

Lysine. ....	61
<b>Figure 57:</b> Substitution d'une Guanine à la position 1625 de l'exon 12 .....	62
<b>Figure 58:</b> Traduction de la séquence non mutée de l'exon 12 .....	63
<b>Figure 59:</b> Traduction de la séquence mutée de l'exon 12 .....	63
<b>Figure 60 :</b> Alignement local de la séquence protéique G551D .....	65
<b>Figure 61:</b> Structure 3D de l'exon 12 et le changement d'acide aminé Glycine en Acide Aspartique. ....	65
<b>Figure 62:</b> Représentation graphique linéaire des relations phénotype génotype de la mucoviscidose.....	68
<b>Figure 63:</b> Représentation graphique circulaire des relations phénotype génotype de la mucoviscidose.....	69

# Liste des tableaux

<b>Tableau I :</b> Répartition des patients atteints de mucoviscidose dans les pays d'Europe .....	12
<b>Tableau II:</b> Caractéristiques de différentes classes des mutations de gène CFTR .....	18
<b>Tableau III:</b> Nomenclature et classification des mutations .....	20
<b>Tableau IV:</b> Phénotype associé à chaque classe de mutation .....	22
<b>Tableau V:</b> Diagnostique paraclinique de la mucoviscidose .....	23
<b>Tableau VI:</b> Séquences des exons étudiés .....	28
<b>Tableau VII:</b> Résultats de BLAST de gène CFTR .....	35
<b>Tableau VIII:</b> Résultats d'une étude sur les mutation du gene CFT dans la population algérienne .....	66
<b>Tableau IX :</b> Résultats d'ensembl .....	67

# Sommaire

<b>Introduction.....</b>	<b>1</b>
<b>Partie théorique .....</b>	<b>3</b>
<b>Chapitre I : I. Généralités sur les maladies génétiques .....</b>	<b>3</b>
I .1 .Aperçu historique.....	3
I.2.Classification des maladies génétique.....	4
I..2.1.maladies génétique autosomiques dominantes.....	4
I .2.2.Maladies génétiques autosomiques récessives .....	4
I .2.3.Maladies liées au sexe .....	5
I .3.Méthodes de dépistage des anomalies génétiques .....	5
<b>II: La mucoviscidose.....</b>	<b>10</b>
II.2. Épidémiologie de la mucoviscidose .....	10
II.3. Épidémiologie Dans le monde.....	12
II.4. Épidémiologie En Algérie .....	13
<b>III .1. Aspect génétique de la mucoviscidose.....</b>	<b>14</b>
III.1.1. Gène CFTR .....	14
III .1.2. La protéine CFTR .....	14
III .1.3. Fonction de la protéine CFTR.....	15
III .1.3.1.Canal du chlore .....	15
III.1.3.2.Autres fonctions .....	16
III .2 .Description et classification des mutations.....	17
III .2 .1.Classification des mutations.....	17

<b>III.2-Mutations spécifiques de la mucoviscidose .....</b>	<b>21</b>
III.2.Mutations du gène CFTR et différentes formes de mucoviscidose .....	21
<b>IV: Étude in Silico: Contribution de la bioinformatique... ..</b>	<b>24</b>
<b>IV.1. Historique .....</b>	<b>24</b>
<b>IV.2. Description de quelques outils bioinformatiques utilisés dans les recherches biomédicales.....</b>	<b>24</b>
<b>Partie pratique.....</b>	<b>27</b>
<b>Chapitre II : Matériels et méthodes.....</b>	<b>27</b>
I.1. Séquence du gène CFTR .....	27
I.2. Différents outils et logiciels bioinformatique .....	27
<b>Chapitre III : Résultats et discussion.....</b>	<b>34</b>
<b>Chapitre III : Conclusion.....</b>	<b>71</b>
<b>Références bibliographiques</b>	

# Liste des abréviations

**ABC** : transporteurs à ATP-binding Cassette

**AD** : Autosomique Dominant

**AMPc** : Adénosine Cyclique Mono Phosphate

**AR** : Autosomique Récessif

**ATP** : Adenosine Tri-Phosphate

**BLAST** : Basic Local Alignment Search Tool

**BPCO** : Broncho-Pneumopathie Chronique Obstructive

**CF** : Cystic fibrosis

**CFTR** : Cystic fibrosis Transmembrane conductance Regulator

**COSMIC**: Catalogue Of Somatic Mutations In Cancer

**DDBJ** : DNA Data Bank of Japan

**DDP** : Différence De Potentiel transépithélia

**EBI** : European Bioinformatics Institute

**EMBL** : European Molecular Biologie Laboratory

**ENaC** Amiloride-sensitive sodium channel

**Phe** : Phénylalanine

**ID**: Identifiant

**INSDC** : International Nucleotide Sequence Database

**Kb** : kilobases

**KDa** : kilodalton

**MLPA**: Multiplex Ligation-dependant Probe Amplification

**NBD** : Nucleotide Binding Domain

**NCBI** : National Center for Biotechnology Information

**NIH** : National Institute of Health

**NMD** : Nonsense-Mediated mRNA Decay

**ORF**: Open Reading Frame

**OMIM**: Online Mendelian Inheritance in Man

**PA** : Pseudomonas Aeruginosa

**PCR** : Polymerase Chain Reaction

**PIR** : Protéine Information Ressource

**PKA** : protéine kinase A

**PKC** : protéine kinase C

**PTC** : Premature Termination Codons

**R** : Regulator

**TIR** : TrypsineImmuno-Réactive

**TMD** : Trans Membranaires hydrophobes Domain

## Résumé

La mucoviscidose est une maladie génétique grave affectant principalement les voies respiratoires et pouvant conduire à une insuffisance respiratoire chronique. Elle est transmise selon le mode autosomique récessif. La maladie est la plus fréquente au sein de la population d'origine caucasienne. Le gène responsable de la maladie appelé "gène CFTR" a été identifié en 1989, est situé sur le bras long du chromosome 7(7q31). La déficience du canal CFTR, entraîne des perturbations hydro-électrolytiques qui sont à l'origine de l'hyper viscosité des sécrétions muqueuses. La gravité de la maladie dépend de la ou des mutations du gène CFTR.

L'objectif principal de cette étude est l'analyse bioinformatique des mutations les plus fréquentes du gène CFTR en Algérie : c.1521-1523delCTT(F508del), c.680T>G(L227R),c.1624G>T(G542X);C.3909C>G(N1303K) et c.1652C>A(G551D).

et leurs effets sur la protéine et le phénotype de la maladie, en utilisant quelques outils bioinformatiques tel que la plateforme NCBI, la base de données GenBank pour la requête des séquences, ainsi que les différents algorithmes tels que BLAST pour les alignements, ou ORF Finder pour l'identification des différents cadres de lecture.

Les résultats obtenus après l'analyse de ces différentes mutations : La mutation F508del caractérisé par une dégradation totale de la protéine avec un phénotype sévère de la maladie.

Le caractère pathogène de la mutation L227R a été confirmé par différents logiciels ; la mutation G542X caractérisé par un phénotype sévère de la maladie ;les résultats de l'analyse des mutations N1303K , G551D montrent une perte totale de la fonction protéique avec un phénotype sévère de la maladie.

**Mots clé** : mucoviscidose, mutation ; bioinformatique, phénotype.

## Abstract

Cystic fibrosis is a serious genetic disease that primarily affects the respiratory tract and may lead to chronic respiratory failure. It is transmitted according to the autosomal recessive mode. The disease is most common in the Caucasian population. The gene responsible for the disease called "CFTR gene" was identified in 1989, is located on the long arm of chromosome 7 (7q31). The deficiency of the CFTR channel causes hydroelectrolytic disturbances which are at the origin of the hyper viscosity of the mucous secretions. The severity of the disease depends on the mutation (s) of the CFTR gene.

The main objective of this study is the bioinformatic analysis of the most frequent mutations of the CFTR gene in Algeria: c.1521-1523delCTT (F508del), c.680T> G (L227R), c.1624G> T (G542X); C.3909C> G (N1303K) and c.1652C> A (G551D).

and their effects on the protein and the phenotype of the disease, using some bioinformatic tools such as the NCBI platform, the GenBank database for querying sequences, as well as different algorithms such as BLAST for alignments, or ORF Finder for the identification of different reading frames.

The results obtained after the analysis of these different mutations: The F508del mutation characterized by a total degradation of the protein with a severe phenotype of the disease.

The pathogenicity of the L227R mutation has been confirmed by different software; the G542X mutation characterized by a severe phenotype of the disease, the results of the analysis of mutations N1303K, G551D show a total loss of protein function with a severe phenotype of the disease.

**Key words:** cystic fibrosis, mutation; bioinformatic, phenotype.

## ملخص

التليف الكيسي هو مرض وراثي خطير يؤثر بشكل رئيسي على الجهاز التنفسي وقد يؤدي الى قصور مزمن في التنفس تنقله عوامل وراثية متنحية. هذا المرض هو الاكثر شيوعا بين السكان من اصل قوقازي . تم التعرف على الجين المسؤول عن المرض المسمى "جين CFTR في عام 1989. يقع على الذراع الطويلة للكروموسوم 7 (7q31). يؤدي نقص قناة CFTR الى حدوث اضطرابات كهرومائية التي تتمثل في الزوجة المفرطة للإفرازات المخاطية. تعتمد شدة المرض على طفرة الجين .

الهدف الرئيسي من هذه الدراسة هو التحليل المعلومات الاحيائية للطفرات الاكثر شيوعا في الجين CFTR في الجزائر c.1521-1523delCTT(F508del) c.680T>G(L227R),c1624G>T(G542X);C.3909C>G(N1303K).1652C>A(G551D).

وتأثيرها على البروتين و النمط الظاهري للمرض. باستخدام بعض الادوات المعلوماتية الحيوية مثل NCBI. قاعدة البيانات Genbank للاستعلام عن التسلسل ORF Finder لتحديد اطارات القراءة المختلفة. النتائج التي تم الحصول عليها بعد تحليل هذه الطفرات المختلفة: طفرة F508del تتميز بانحطاط كامل للبروتين مع النمط الظاهري الشديد للمرض. تم تأكيد التسبب في طفرة L227R بواسطة برامج مختلفة ؛ طفرة G542X التي تتميز بنمط ظاهري حاد للمرض ، تظهر نتائج تحليل الطفرات N1303K ، G551D فقدان كلي لوظيفة البروتين مع النمط الظاهري الحاد للمرض. الكلمات المفتاحية: التليف الكيسي ، الطفرة .المعلومات البيولوجية ، النمط الظاهري.

## **Introduction**

Les maladies respiratoires touchent les voies aériennes, y compris les voies nasales, les bronches et les poumons. Elles vont d'infections aiguës comme la pneumonie et la bronchite à des affections chroniques telles que l'asthme et la broncho-pneumopathie chronique obstructive et la mucoviscidose. **(OMS,2018)**. Les maladies respiratoires chroniques représentent un large éventail de maladies graves, qui touchent des centaines de millions de personnes de tous âges. Elles représentent quatre millions de décès par an. **(Akiki, 2017)**, Par exemple l'asthme est défini comme une inflammation chronique caractérisé par manque du souffle épisodique avec le sifflement. Il est caractérisé par obstruction intermittente de flux d'air, hyperréactivité de voie aérienne, et plus grande production de mucus. **(Jamie L. Sturgill.2018)**

La broncho-pneumopathie chronique obstructive (BPCO) est une maladie chronique définie par une obstruction lente et progressive, non complètement réversible, des voies aériennes. Cette obstruction est associée à une réponse inflammatoire pulmonaire anormale à des toxiques inhalés dont le principal est le tabagisme **(Khayath et al, 2016)**. (La BPOC a été identifiée il y'a plus de deux cents ans, les première référence correspondent à Bonet (1679) qui a décrit l'emphysème de poumon volumineux, un siècle plus tard Morgani(1779) a donné la description de 19 cas de malades ayant un poumon turgescents d'air. **(Oujidi ,2018)**. Par contre La mucoviscidose est une maladie génétique affectant principalement les voies respiratoires et pouvant conduire à une insuffisance respiratoire chronique et au décès précoce. Elle est la plus fréquente des maladies génétiques héréditaire rare dans les populations d'origine caucasienne, caractérisée par des mutations du gène codant le canal chlorure CFTR, le dysfonctionnement de ce canal aboutit à la fabrication d'un mucus anormalement visqueux et obstructif à l'origine des manifestations cliniques Bien que ce gène soit connu depuis 1989. **(Lafoste et al, 2018)**

La biologie est définie comme l'étude du vivant, depuis l'étude des interactions entre espèces et populations jusqu'aux études des fonctions des tissus et des cellules au sein d'un organisme .dans le cadre de leurs travaux, les biologistes ont à collecter et à interpréter un grand nombre de données. Au début de XXIe siècle, un ensemble de technologies expérimentales complexes permet de collecter des données plus vite qu'elles ne peuvent être interprétées.



Par exemple La bioinformatique est un domaine de recherche qui propose et développe des modèles ,des méthodes et les outils afin d'analyser l'information disponible (donnée de séquences ,etc.),et produire de nouvelles connaissance pour mieux comprendre la biologie ,en mettant à la disposition des chercheurs en biologie des bases de données, permet de comparer des séquence d'ADN et de générer des résultat qui sont potentiellement significatifs. Le but de ce mémoire est : l'étude des mutations du gène CFTR et l'analyse bioinformatique rétrospective dans la population algérienne. (**Gibas et Jambeck, 2002**)

## Chapitre I : Généralités sur les maladies génétiques

Les maladies génétiques ou les maladies héréditaires est un groupe de maladies provoquées par des mutations du génome d'un individu. (Chu SY, Weng CY.2017) une maladie génétique peut ne pas être héréditaire (cas de plusieurs cancers), par contre une maladie héréditaire est réserver aux maladies génétiques. (Fagioli et al ,2013).

Il existe trois groupes de maladies génétiques :

- les maladies monofactorielles (ou *monogéniques*) liées à une anomalie sur un gène donné qui sont héréditaires et transmises d'une génération à l'autre avec une probabilité connue.
- les maladies plurifactorielles, c'est-à-dire auxquelles peuvent prédisposer plusieurs anomalies génétiques en interaction avec des facteurs environnementaux.
- les aberrations chromosomiques (modification dans la structure et le nombre des chromosomes), qui sont héritées ou non (accidentelles). (IFSI, 2015)

### I.1 .Aperçu historique

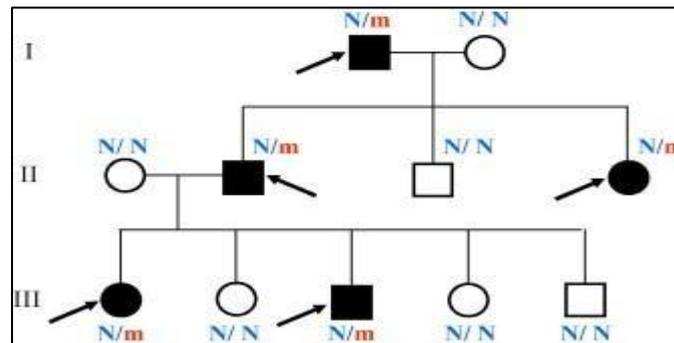
Bien que les maladies héréditaires ont été reconnues depuis des siècles, mais elles sont devenu récemment l'explication qui prévaut pour de nombreuses pathologies humaines. Avant les années 1970, les médecins ont vu la maladie génétique rare et sans rapport avec les soins cliniques. Mais, dans les années 1990, les gènes semblent être des facteurs critiques dans toutes les maladies humaines. (Lindee, 2000). De nombreuses maladies compris être génétique ont été décrites et nommées au 19<sup>ème</sup> siècle. Le syndrome de Down et la maladie de Huntington ont tous deux été décrits à la fin du 19<sup>ème</sup> siècle. Xerodoma pigmentosum en 1874. Le syndrome de Gilles de la Tourette a été décrit en 1885 par le neurologue français Georges Gilles de la Tourette. (Lindee, 2002)

## I .2.Classification des Maladies Génétiques

Sont classé par apport à leur mode de transmission

### I .2.1.Maladies génétiques autosomiques dominantes

Les gènes responsables des maladies transmises sur le mode autosomique dominant (AD) sont localisés sur les autosomes. L'allèle muté responsable de la maladie est dominant sur l'allèle sauvage donc la maladie s'exprime chez les hétérozygotes.( **Figure 1**)



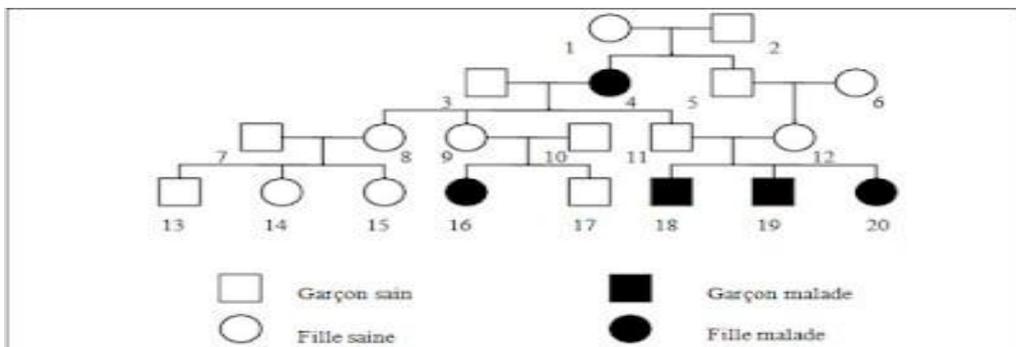
**Figure 1:** Arbre généalogique d'une famille atteinte d'une maladie autosomique dominante.

(Lamoril, 2008)

**N** : Allèle sauvage. / **m** : Allèle Muté. / **m** > **N**.

### I .2.2.Maladies génétiques autosomiques récessives

Les gènes responsables des maladies transmises sur le mode autosomique récessif (AR) sont localisés sur les autosomes. L'allèle muté responsable de la maladie est récessif sur l'allèle sauvage; les hétérozygotes sont sains et la maladie ne s'exprime que chez les homozygotes.( **Figure 2**)



**Figure 2:** Arbre généalogique d'une famille atteinte d'une maladie autosomique récessive.

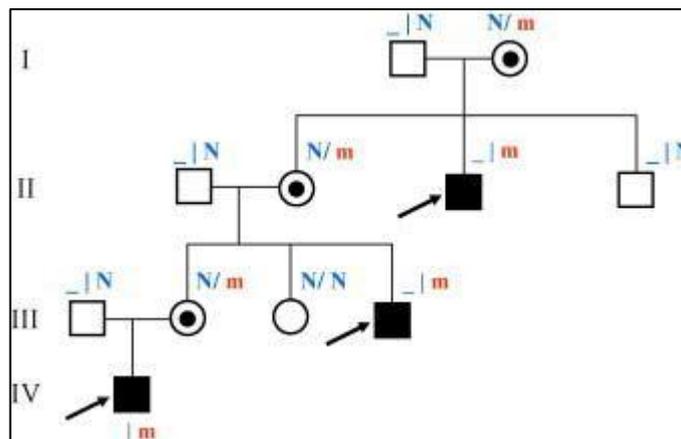
(Anonyme 1

### I .2.3.Maladies liées au sexe

Le plus souvent en rapport avec une mutation située sur le chromosome X, dans les troubles récessifs liés à l'X, les personnes touchées sont généralement de sexe masculin.

(Figure 3)

(IFSI, 2015)



**Figure 3:** Arbre généalogique d'une maladie récessive liée à X (Lamoril, 2018)

### I .3.Méthodes de dépistage des anomalies génétiques

De nombreux tests génétiques apportent des informations relatives à la santé des individus ou à celle de leur famille. Ces tests consistent à rechercher des anomalies sur la molécule d'ADN elle-même, ou à dépister des anomalies concernant le nombre ou la forme des chromosomes.

#### I.3.1.Analyse de chromosome

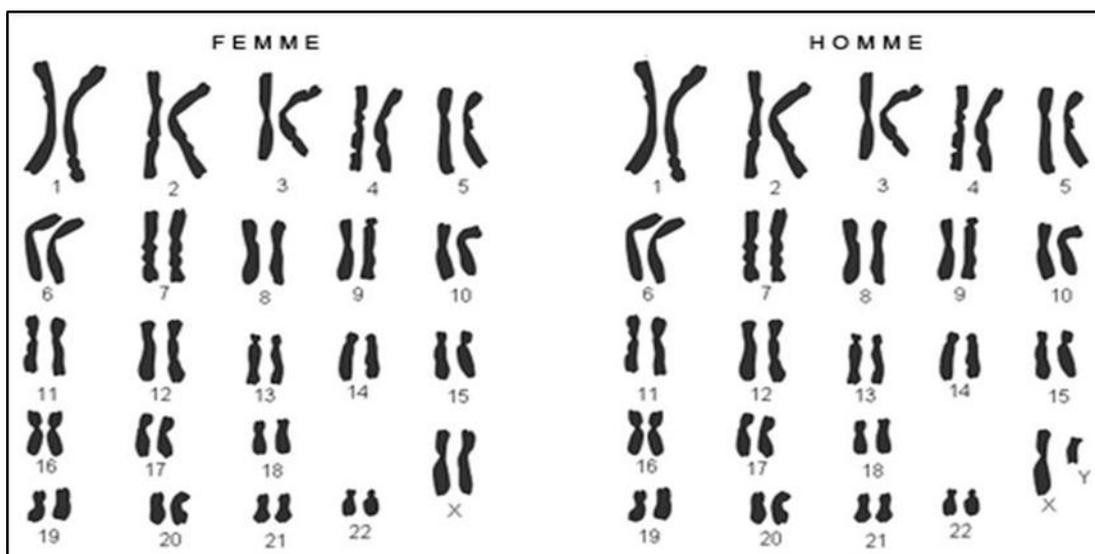
La cytogénétique est une discipline qui a pour objet l'étude de la physiopathologie des chromosomes (condensation, recombinaison, réparation, ségrégation, transmission) et de la chromatine (organisation et rôle dans la régulation de l'expression des gènes), elle a un but de détecter les anomalies chromosomiques constitutionnelles ou acquises grâce à des techniques microscopiques (techniques de bandes, techniques de cytogénétique moléculaire) ou de biologie moléculaire afin d'établir un diagnostic biologique et d'assurer un conseil génétique. Il est possible de différencier la cytogénétique avec le caryotype et la cytogénétique moléculaire qui a révolutionné

le domaine de la génétique chromosomique, dont les deux techniques majeures sont l'*hybridation in situ en fluorescence* (FISH) et l'analyse chromosomique sur puce à ADN (ACPA) ou CGH array.

### 1. a. Le caryotype

Le caryotype est le premier examen permettant l'analyse morphologique des chromosomes a pour objet soit de dépister des anomalies constitutionnelles présentes sur toutes les cellules (caryotype constitutionnel), soit de reconnaître des anomalies acquises limitées à un clone cellulaire (cancers et leucémies). **(Figure 4) (Caquet, 2010)**

- ❖ La principale limite du caryotype est son niveau de résolution qui ne dépasse pas 5—10 millions de paires de bases (Mb). **(Dimassi et al ,2017)**



**Figure 4:** Caryotype humain normal (Anonyme2)

### 1. b. L'analyse par Hybridation in situ en fluorescence FISH

L'analyse par FISH est une technique qui permet de voir une zone spécifique d'un chromosome **(Figure 5)**, contrairement au caryotype qui permet de tous les analyser, est une technique très sensible et spécifique qui peut être réalisée en routine dans un laboratoire de pathologie, elles s'appliquent sur des tissus congelés ou fixés et inclus en paraffine ou sur du matériel cytologique. **(Hamard ; 2018)**

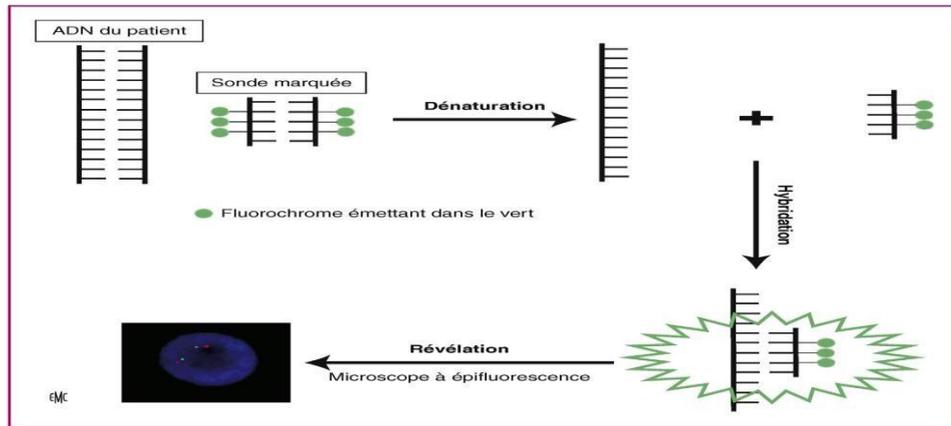


Figure 5: Schéma représente le principe de la FISH (Dimassi et al, 2017)

### 1.3.2. Analyse de gène

Si l’anomalie résulte de la modification d’un gène (ADN) on parle généralement de mutation. Ces mutations peuvent entraîner des dysfonctionnements plus ou moins graves : on parle alors de maladie génique.

#### 1.3.2. a. Séquençage de gènes

La méthode la plus utilisée dans les laboratoires de diagnostic moléculaire est le séquençage dit « sanger » d’un seul gène ou d’un groupe de gènes ,les régions ciblées font l’objet d’une amplification génique (PCR),indépendamment les unes des autres ,les fragments d’ADN obtenue sont séquencés sur un appareil de séquençage capillaire .une comparaison est alors possible entre les séquences d’ADN du patient et la séquence de référence pour mettre en évidence d’éventuels variantes nucléotidiques (Figure 6).

- ❖ Cette approche est chronophage et ne permet d’explorer d’un nombre limité de gène. (Lacoste et al ,2017)

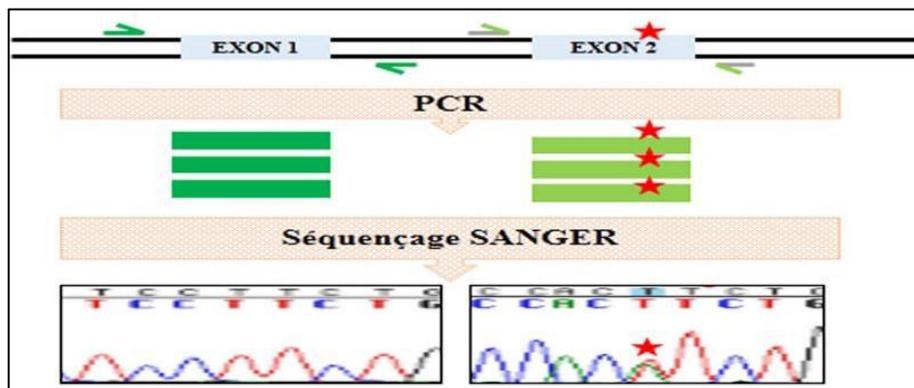
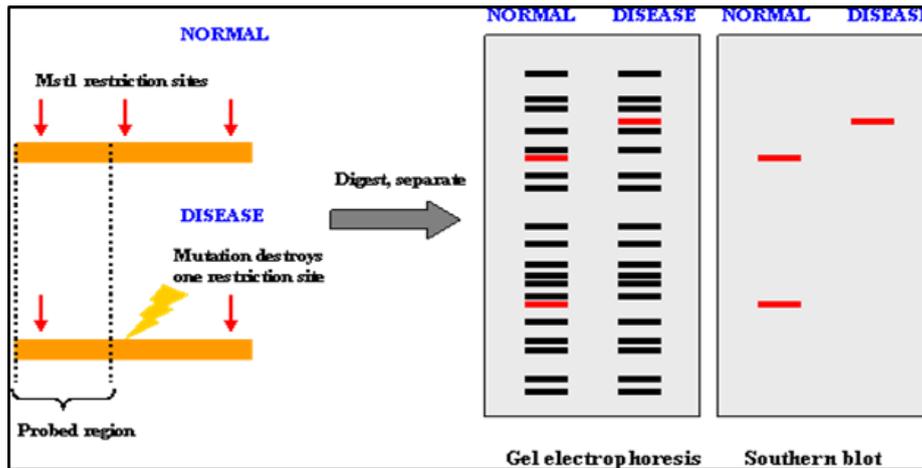


Figure 6: Représentation schématique de la stratégie de diagnostic par séquençage capillaire dit sanger. (Lacoste et al ,2017)

**1.3.2. b.RFLP (Restriction Fragment Length Polymorphism)**

Une technique qui exploite les variations dans les séquences d'ADN homologues. Il y a une différence dans les molécules d'ADN homologues qui peuvent être détectés par la présence de différents fragments de longueur après digestion de l'ADN par des enzymes de restriction (**Figure 7**). (**Dai et al 2014**)



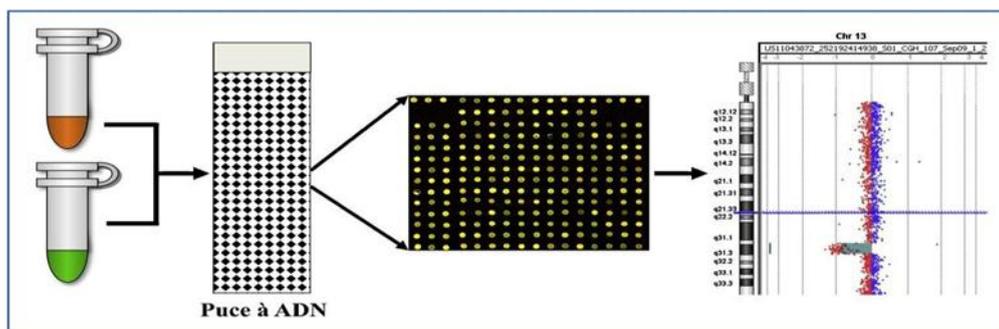
**Figure 7:** Schéma représente le principe de RFLP (**Anonyme 3**)

**1.3.2. Analyse mixte**

Certaines techniques plus récentes mélangent l'approche de cytogénétique (analyses des chromosomes) et de la génétique moléculaire (analyses des gènes).

**1.3.3. a. L'analyse chromosomique par puce à ADN (ACPA) (aussi appelée CGH array)**

La technique de CGH array (**Figure 8**), s'affranchissant d'une analyse morphologique des chromosomes, offre l'avantage d'être un examen automatisable permettant une meilleure fiabilité et reproductibilité des résultats ainsi qu'une étude d'un plus grand nombre de patients, elle permet non seulement de détecter les anomalies chromosomiques déséquilibrées détectées par le caryotype, mais aussi celles qui sont cryptiques. (**Malan, 2012**)



**Figure 8:** Principe de la CGH array sur puce à ADN (**Malan, 2012**)

### 1.3.3. b. Les nouvelles générations de séquenceurs (NGS)

Le NGS a révolutionné la connaissance du génome et du transcriptome tumoraux, ainsi que l'évolution clonale du cancer. En clinique, l'avantage principale du NGS est la possibilité de détecter un nombre important d'anomalies moléculaires potentielles à partir d'une faible quantité d'ADN et en un temps court. En plus de sa rapidité, il présente une sensibilité de détection supérieure aux techniques antérieures basées sur la PCR en temps réel ou le séquençage Sanger. Les techniques de NGS peuvent être utilisées pour différents types d'analyse :

- ❖ le séquençage du génome complet.
- ❖ de l'exome (régions codantes des génomes).
- ❖ du RNAseq (ensemble des ARN messagers, permettant de détecter les fusions de gènes) ou le séquençage ciblé d'un panel de gènes préétabli (quelques dizaines à quelques centaines). (Hamard et al, 2018).

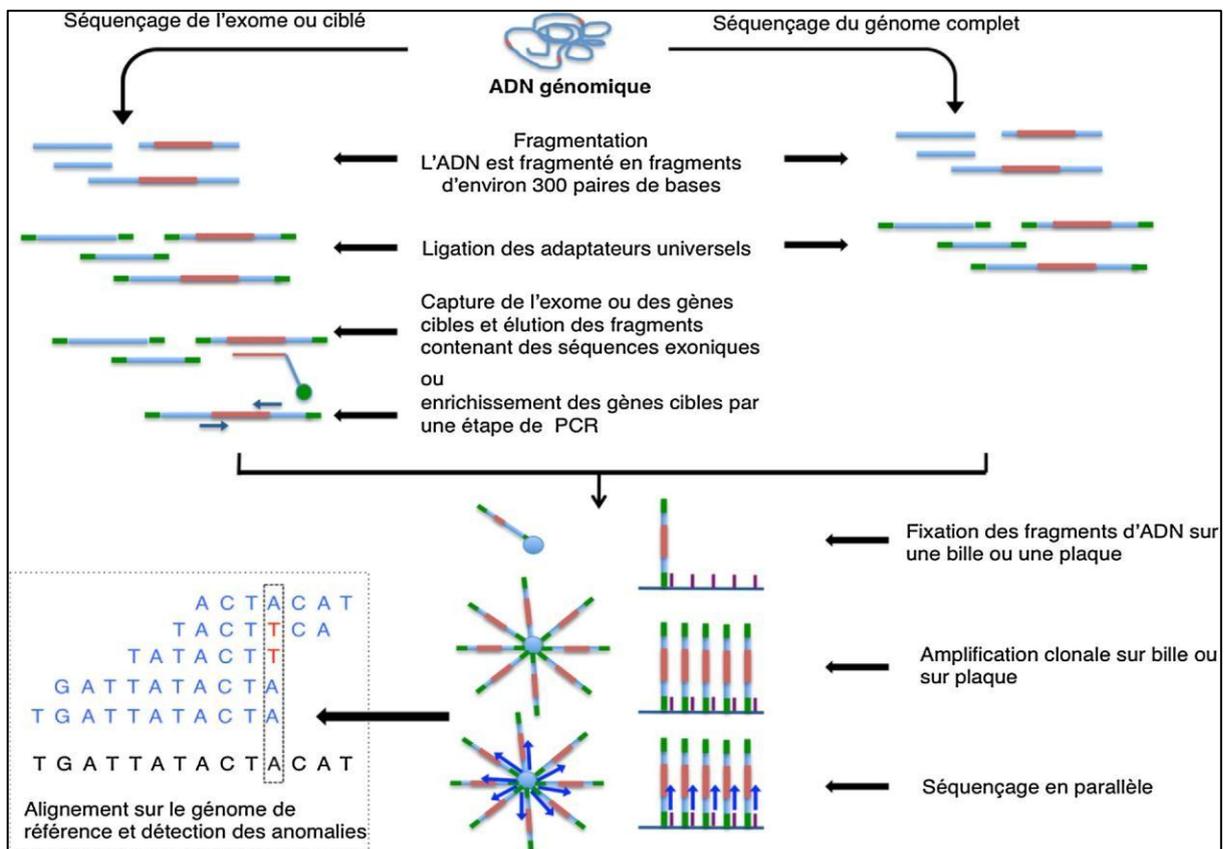


Figure 9: Différentes étapes du typage NGS (Hamard et al, 2018)

## II: La mucoviscidose

Le mot **mucoviscidose** dérive Du latin *mucus* (« mucus ») / sécrétion et *viscum* (« glu ») / gluant, visqueux.

Dès le moyen âge, les enfants atteints de mucoviscidose furent décrits comme ensorcelés, comme en témoigne ce dicton du XVIIème siècle : « Malheur à l'enfant qui laisse un goût de sel lorsqu'on l'embrasse sur le front: on lui jeta un sort et il mourra bientôt ! ».

En 1606, Juan Alonzoy de los Ruyzes de Fonteca, médecin espagnol, décrit un enfant atteint de la maladie et rapporta l'anomalie du transport ionique transépithéliale, et dont la conséquence est la production d'une sueur anormalement salée.

A la même époque Pieter Pauw, professeur d'anatomie de l'université de Leyde aux Pays-Bas, décrit via l'autopsie d'une fillette de 11 ans supposée ensorcelée, des atteintes pancréatiques. Le pancréas était élargi, dur et blanc.

En 1936, cette pathologie nommée pour la 1ère fois « fibrose kystique du pancréas et bronchectasie » dans une thèse dirigée par Guido Fanconi.

En 1938, la maladie fut réellement décrite comme une entité pathologique distincte par la pédiatre américaine Dorothy Andersen, qui donna une description des caractéristiques cliniques et histologiques de la maladie, notamment l'obstruction intestinale néonatale, les complications respiratoires et digestives et les lésions histologiques spécifiques du pancréas.

Le terme de mucoviscidosis, créé à partir des termes « mucus » et « visqueux », fut utilisé pour la première fois en 1943 par le docteur Sydney Farber.

En 1946, Hodge et Anderson suggérèrent le caractère héréditaire et le mode de transmission récessif de la maladie, qui est probablement causé par la mutation d'un seul gène.

En 1953, Paul Di Sant'Agnese, médecin hospitalier a remarqué que les enfants atteints de mucoviscidose étaient particulièrement affectés et meurent par déshydratation. Il a montré que cette déshydratation est due à une concentration anormale de chlorure de sodium et de potassium dans la sueur de ces patient

En effet en 1959, la technique du test de la sueur a été simplifiée par la méthode de l'iontophorèse à la pilocarpine décrite par Gibson et Cooke. Grâce à un diagnostic précoce basé sur ce test, des centres spécifiques de traitement ont été développés. L'origine biochimique de la mucoviscidose fut précisée en 1981 par Knowles qui remarqua que le potentiel transépithélial, mesuré au niveau de l'épithélium nasal des patients mucoviscidosiques, est largement plus électronégatif que celui des sujets sains, puis par Quinton en 1983, qui démontra que cette électronégativité accrue provient d'une faible perméabilité de l'épithélium aux ions chlorures. Cette découverte a fait le lien physiologique entre les poumons, le pancréas et les glandes sudoripares expliquant que l'atteinte des différents organes était due à des anomalies électrolytiques et que le mucus visqueux n'était que l'une des conséquences et non la cause.

En 1985, le locus du gène responsable de la maladie fut localisé sur le bras long du chromosome 7 en 7q31 grâce à un marqueur RLFP

Puis en 1989 l'anomalie génétique à l'origine de la maladie fut mise en évidence par les équipes de Lap-Chi Tsui, Collins et Riordan. Il s'agit d'une mutation d'un gène localisé en 7q31 codant une protéine transmembranaire appelée CFTR. Ce n'est qu'en 1991, qu'Anderson *et al.* Apportèrent les preuves que le CFTR était bien un canal chlore. (Ouchtati A, Dali H ,2018)

## II.2. Épidémiologie de la mucoviscidose

La mucoviscidose est la maladie la plus fréquente dans les populations d'origine caucasienne avec plus de 70 000 individus atteints dans le monde, elle est transmise selon un mode autosomique récessif et elle est due à des mutations du gène codant pour la protéine CFTR. Cette dernière est un canal ionique du chlore exprimé à la surface des cellules épithéliales des organes producteurs de mucus (voies respiratoires, pancréatiques, biliaires, intestinales, canaux déférents, glandes sudoripares). La déficience du canal CFTR, entraînent des perturbations hydro électrolytiques qui sont à l'origine de l'hyper viscosité des sécrétions muqueuses. (Regard et al.2018).

L'épidémiologie clinique signifie une enquête qui montre l'incidence de la maladie dans les populations humaines. Au cours de ces dernières années, de nombreux

chercheurs se sont concentrés sur la fréquence et / ou la prévalence de la mucoviscidose.

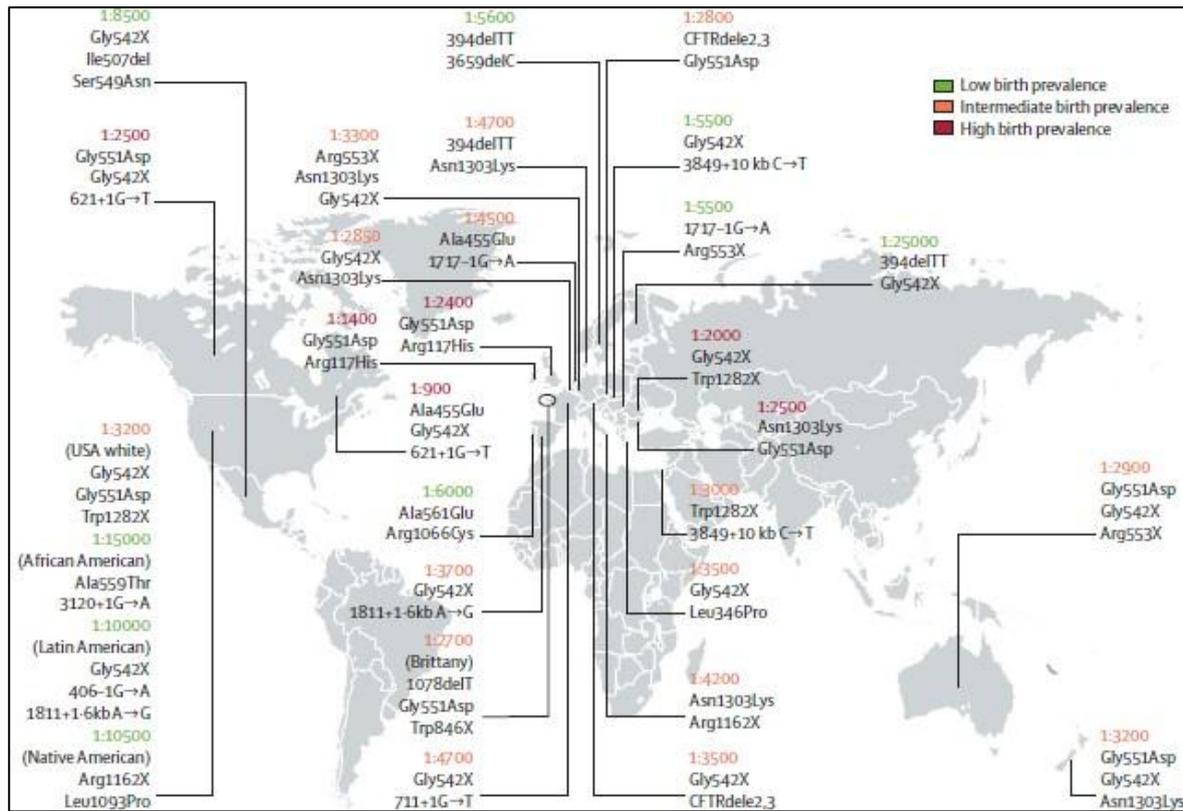
(Mirtajani et al ,2017 ). Elle est variable en fonction de la localisation géographique des populations et de la fréquence des mutations responsables de la maladie. La migration des populations et le mariage consanguin sont aussi des facteurs qui influent la répartition de cette pathologie. (Ouchtati A, Dali H ,2018).

### II.2.1. Épidémiologie Dans le monde

Une incidence plus élevée est notée dans les pays européens à un taux de 1 sur 2000 à 3000 naissances (**Tableau I**), Aux États-Unis, l'incidence de la mucoviscidose est rapportée à 1 dans chaque 3500 naissance. En Afrique et en Asie, bien que CF est sévèrement sous-diagnostiquée, les données indiquent que la prévalence des FC est faible à rare (**Figure 10**). (Bono-neri et al ,2018)

**Tableau I** : Répartition des patients atteints de mucoviscidose en Europe  
(Mirtajani et al ,2017)

	Population	Prévalence	Incidence
Danemark	5413	0/761	1 :4700
Finlande	5215	0/123	1 :25000
France	60424	0/750	1 :4700
Allemagne	82425	0/829	1 :3300
Irlande	3970	2/98	1 :1353
Italie	58057	0/872	1 :4238
Pays-Bas	16318	0/781	1 :4750
Le Portugal	10524	0/271	1 :6000
Espagne	40281	0/546	1 :3750
Royaume-Uni	60271	1/37	1 :2381



**Figure 10:** Prévalence approximative de la mucoviscidose dans certains pays

(Sediki FZ.2015)

### II.2.2. Épidémiologie en Algérie

L'incidence de cette maladie n'est pas connue de manière précise, car nous, n'avons pas de données épidémiologiques, ni de recensement, ni de notification précise, ni de registre de cette maladie. Le nombre de cas recensés correspond aux cas colligés dans les services hospitaliers de la capitale et des grandes villes d'Algérie.

Pour une population de 38 millions d'habitants et une cohorte néonatale de près d'un million de personnes, l'estimation de nombre de nouveaux cas à près de 250 par an. L'immense majorité d'entre eux n'est pas diagnostiquée. Les ressources et dépenses de santé dans notre pays sont encore très faibles.

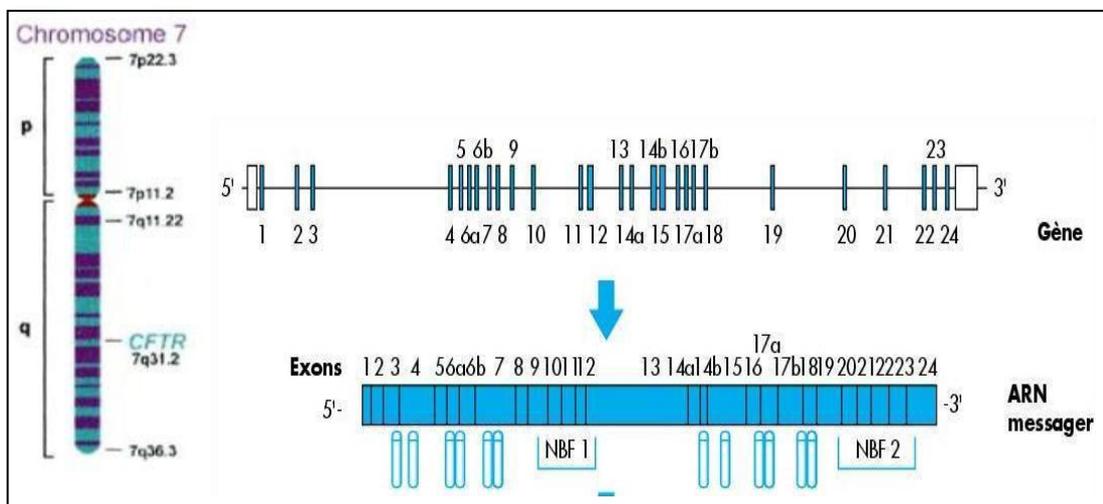
Compte tenu du taux de consanguinité élevé au Maghreb – il atteint presque 25 % en Algérie – il fait peu de doute que cette maladie est plus fréquente qu'on ne le pense, et au moins aussi fréquente qu'en France. Notre retard est donc réel en matière de diagnostic.

(Bendridi et al ,2014)

### III : 1. Aspect génétique de la mucoviscidose

#### III .1.1. Gène CFTR

Le gène CFTR localisé sur le bras long du chromosome 7 (7q31.2), avec une taille de 250 kb (kilobases), est constitué de 27 exons notés classiquement de 1 à 24 (avec des exons dédoublés : 6a et 6b, 14a et 14b, 17a et 17b). Après transcription, l'ARNm est constitué de 6 500 paires de bases et sera traduit en une protéine dont l'expression normale est située au niveau de la membrane apicale des cellules épithéliales (**Figure 11**). (**Sediki FZ.2015**)



**Figure 11:** Schéma Représentant l'organisation du gène CFTR. (**Sediki, 2015**)

#### III .1.2. La protéine CFTR

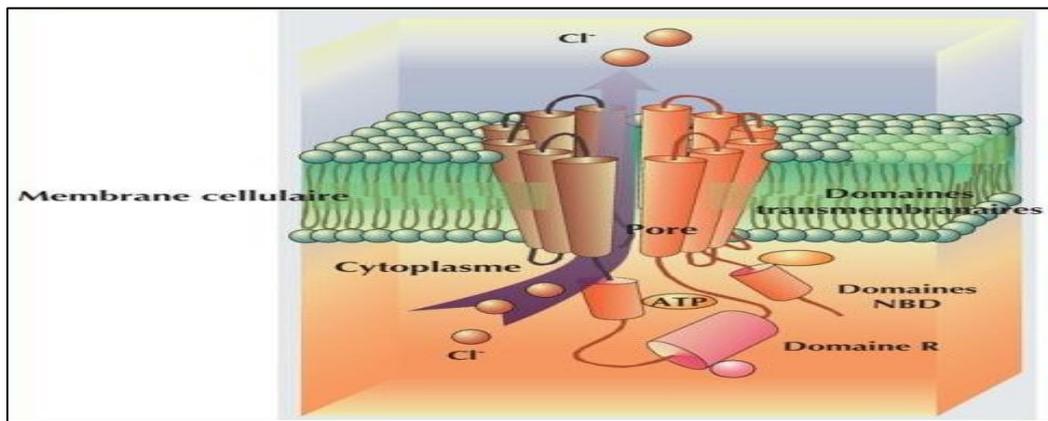
La protéine CFTR appartient à la famille des transporteurs ABC (ABCC7) (transporteurs à *ATP-binding cassette*), elle est constituée de 1480 acide aminés et a un poids moléculaire de 170 kDa. (**Hatton et al ,2013**)

Cette protéine comme les autres protéines de la famille ABC elle présente :

- Deux domaines transmembranaires hydrophobes (TMD), TM1 (exons 1–8).TM2 (exons14-18) (**Gaillard et al, 2009**), chacun composé de six hélices, qui délimitent un pore transmembranaire.
- Deux domaines intracytoplasmiques hydrophyles (NBD1 et NBD2) reliés aux TMD par quatre boucles intracellulaires (de CL1 à CL4) et capables de fixer l'ATP.

À cet égard, le 508e acide aminé joue un rôle crucial car il est localisé sur NBD1 à l'interface avec la première et la quatrième boucle cytoplasmique (CL4, CL1). De même, NBD2 interagit avec CL2 et CL3.

- Un domaine *Regulator* (R) intracellulaire qui est spécifique de CFTR au sein de la famille ABC. Ce dernier contient des résidus sérine susceptibles d'être phosphorylés par la protéine kinase A et la protéine kinase C, elles-mêmes régulées par l'adénosine cyclique monophosphate (AMPc) (**Figure 12**). (**Hatton et al ,2013**)



**Figure 12:** Structure de la protéine CFTR. (**Hatton et al ,2013**)

### III .1.3. Fonction de la protéine CFTR

La protéine CFTR a un rôle principal dans le transport actif de différents substrats, comme les acides aminés, les peptides, les protéines, et les ions au travers de la membrane plasmique. (**Bardin et al ,2018**)

#### III .1.3.1. Canal du chlore

L'augmentation intracellulaire de l'AMPc permet d'activer la kinase PKA (protéine kinase AMPc dépendante) capable alors de phosphoryler le domaine R.

La phosphorylation du domaine R par les kinases PKA et PKC induit un changement conformationnel augmentant l'accessibilité des domaines de liaison des nucléotides NBD1 et NBD2 à l'ATP.

La liaison de deux molécules d'ATP au niveau des deux NBDs permet un rapprochement de ces deux domaines, enfermant les molécules d'ATP dans une poche. Un signal est alors transmis à la région transmembranaire, permettant l'ouverture du canal qui est maintenue jusqu'à l'hydrolyse de l'ATP au niveau du NBD2.

L'hydrolyse de cette 2ème molécule d'ATP affaiblit la liaison NBD1-NBD2, ce qui déclenche la fermeture du canal. (Alix de Becdelievre, 2011)

### III .1.3.2. Autre fonctions

- ✓ Permet le passage d'autres anions comme I<sup>-</sup>, Br<sup>-</sup>, F<sup>-</sup>, et des molécules organiques comme le glutathion.
- ✓ inhibe le canal ENaC de ce fait limite la réabsorption de Na<sup>+</sup> dans les voies respiratoires, à l'inverse de l'épithélium sudoral où cette inhibition tonique n'a pas été démontrée. . (Hatton et al ,2013)

Le dysfonctionnement de la CFTR conduit à une gamme large et variable de présenter des manifestations et des complications au niveau des organes (**Figure 13**). (Egan, 2016)

- ❖ **Les poumons** : la mucoviscidose provoque un épaississement des sécrétions muqueuses qui recouvrent les bronches et une obstruction des petites bronches. Elle favorise également la survenue de surinfections répétées.
- ❖ **Le pancréas** : les sucs pancréatiques ne sont plus suffisamment sécrétés, ce qui conduit à une mauvaise absorption des graisses et des troubles nutritionnels. En ce sens, la mucoviscidose est également appelée fibrose kystique du pancréas.
- ❖ **Les glandes sudoripares** : augmentation de la concentration en ion chlorure dans la sueur.
- ❖ **Les glandes sexuelles chez l'homme** : obstruction des canaux déférents (empêchant le déplacement des spermatozoïdes produits par les testicules). (Elborn ,2016)

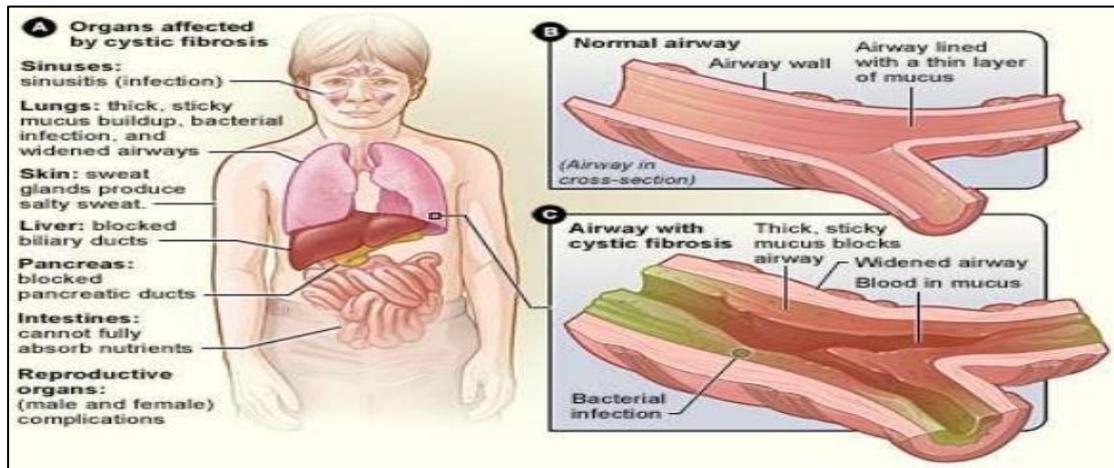


Figure 13: Organes touchés par la mucoviscidose. (Anonyme4)

### III .2 .Description et classification des mutations

Plus de 2000 mutations réparties tout au long du gène *CFTR*. La majorité est des mutations dites ponctuelles. Parmi ces mutations : mutation faux sens (40 %), de décalage du cadre de lecture (16 %), des défauts d'épissage (12 %), des mutations non-sens (8 %),(Covrol et, al 2015), des grandes insertions/délétions (3 %), des mutations dans le promoteur (0,8 %), enfin des variations de séquences dont on ne peut prédire le caractère pathogène (14 %).La plus fréquente est la délétion du 508<sup>e</sup> acide aminé p.Phe508del,F508del. Elle affecte 70 % des allèles dans la population caucasienne avec un gradient décroissant Nord-Sud. Vingt autres mutations seulement touchent plus de 0,1 % des allèles. (Hatton et al ,2013)

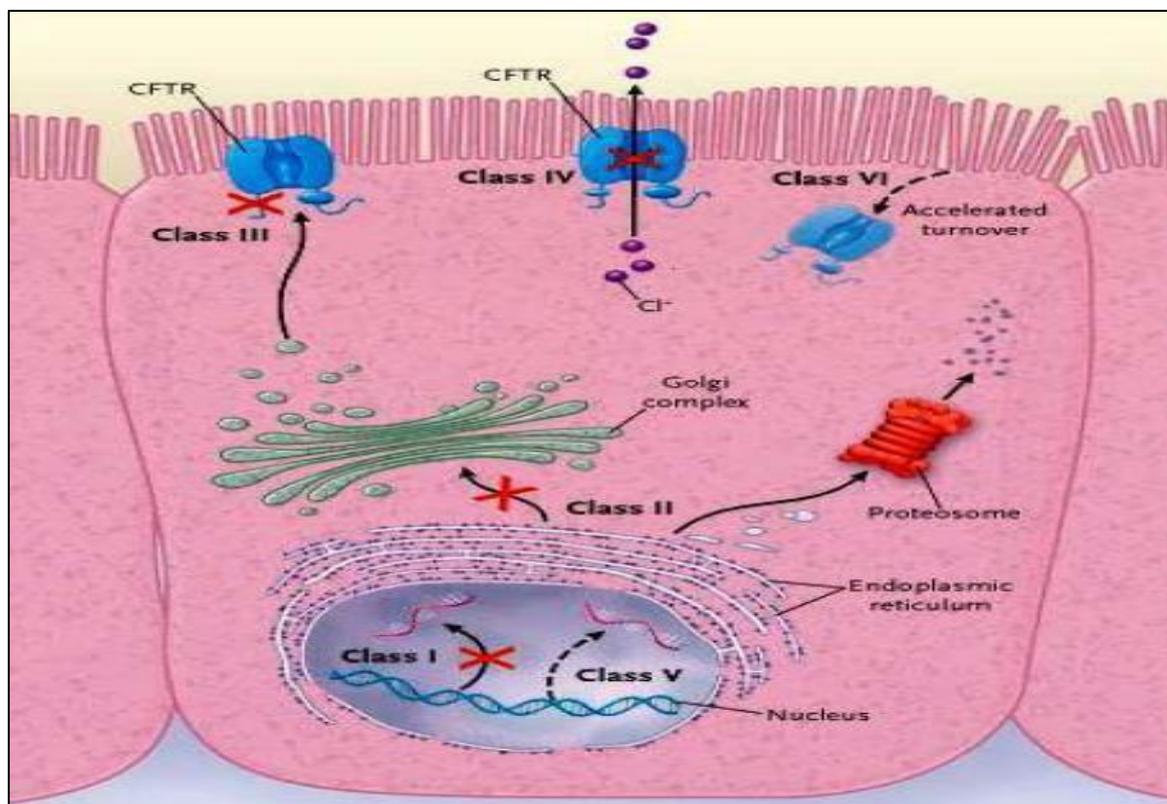
#### III .2 .1.Classification des mutations

Les mutations sont réparties en 6 classes une première classification a été faite par Welsh et Smith qui ont classés ces mutations en quatre groupes. Vankeerberghen a modifié ce classement pour établir 6 classes différentes. (Voir Tableau II et Figure 14).(Sediki.2015).

**Tableau II:** Caractéristiques de différentes classes des mutations de gène CFTR.

Les classes	Le défaut	Caractéristiques
I	Défaut de production de la protéine (absence de fabrication totale ou partielle)	<ul style="list-style-type: none"> <li>➤ mutations non-sens caractérisées par la présence d'un codon stop prématuré (PTC) – UAA, UAG ou UGA – et des décalages du cadre de lecture ou des mutations au site invariant d'épissage introduisant un PTC.</li> <li>➤ Dans la grande majorité des cas, le système du <i>Nonsense-Mediated mRNA Decay</i> (NMD) dégrade l'ARN messenger porteur du PTC. Dans le cas où ce mécanisme est outrepassé, la synthèse ribosomale est stoppée au PTC et la protéine « tronquée » est rapidement éliminée. <b>(Ouchtati A, Dali H, 2018)</b></li> </ul>
II	Défaut de maturation	<ul style="list-style-type: none"> <li>➤ La mutation la plus prévalent, F508del, est caractérisée par une dégradation de près de 100 % de la protéine nouvellement synthétisée.</li> <li>➤ Du fait d'un repliement anormal, cette protéine est dégradée par la voie du protéasome. Dans ce cas, la protéine n'est pas retrouvée à la membrane. <b>(Ouchtati A, Dali H, 2018)</b></li> </ul>
III	Mutations perturbant la régulation du canal Cl <sup>-</sup>	<ul style="list-style-type: none"> <li>➤ Cette classe est fréquemment localisée dans le domaine NBD1 et NBD2.</li> <li>➤ affecte l'ouverture du pore.</li> <li>➤ G551D en est l'exemple le plus fréquent.</li> <li>➤ Elle affecte un patient sur 25. <b>(Ouchtati A, Dali H,</b></li> </ul>
IV	Mutations altérant la conductance du canal Cl <sup>-</sup>	<ul style="list-style-type: none"> <li>➤ Les mutations de classe IV sont localisées pour la plupart dans les domaines membranaires qui forment le pore.</li> <li>➤ ces mutations sont également très rares.</li> </ul>

		<ul style="list-style-type: none"> <li>➤ <b>Exemple :</b> l'arginine 117 changé histidine (R117H), apparaît dans seulement 2% des cas des FC dans le monde entier. (Egan.2016)</li> </ul>
V	<b>Mutations altérant la stabilité de l'ARNm</b>	<ul style="list-style-type: none"> <li>➤ Elles peuvent induire une exclusion partielle mais parfois également permettre la production d'un ARN messenger normal. (Ouchtati A, Dali H, 2018)</li> </ul>
VI	<b>Mutations affectant la stabilité de la protéine</b>	<ul style="list-style-type: none"> <li>➤ Elles sont issues le plus souvent d'une délétion de la partie C-terminale de la protéine terminale</li> <li>➤ Les mutants de cette classe permettent la synthèse de protéines fonctionnelles mais instable au niveau membranaire et diminuent les propriétés de régulation des autres canaux.</li> <li>➤ C'est le cas de la mutation Q1412X. (Ouchtati A, Dali H, 2018)</li> </ul>



**Figure 14:** différentes classes des mutations CFTR. (Ouchtati A, Dali H, 2018)

Deux nomenclatures sont utilisées concernant le gène CFTR : la nomenclature traditionnelle, anciennement utilisée, les exons sont numérotés de 1 à 24 et la nomenclature **HGVS** (*Human Genome Variation Society*) qui suit les recommandations internationales (**Tableau III**). (Ouchtati A, Dali H, 2018)

**Tableau III:** Nomenclature et classification des mutations (Elborn, 2016)

classe	Nomenclature traditionnelle	Nomenclature HGVS	
		Le nom de l'ADN	Le nom de la protéine
I	G542X W1282X G551D 621+1G→T	c.1624G>T c.3846G>A c.1657C>T c.489+1G>T	Gly542X Trp1282X Arg553X -
II	F508del N1303K I507del R560T	c.1521_1523delCTT c.3909C>G c.1519_1521delATC c.1679G >C	Phe508del Asn1303Lys Ile507del Arg560Thr
III	G551D - - S549N	c.1652G>A - - c.1646G>A	Gly551Asp Gly178Arg Gly551Ser Ser549Asn
IV	R117H R347P R117C R334W	c.350G > A c.1040G> C c.349C>T c.1000C> T	Arg117His Arg347Pro Arg117Cys Arg334Trp
V	3849+10kbC→T 2789+5G→A 3120+1G→A	3717+10kbC c.2657+5G>A c.2988+1G >A	- - -
VI	4326delTC - 4279insA	- - -	- Gln1412X -

## III.2-Mutations spécifiques de la mucoviscidose

### III.2.Mutations du gène CFTR et différentes formes de mucoviscidose

L'existence de phénotypes différents de la maladie est liée à la nature de mutations présentes sur le gène *CFTR*, (**Bono-neri2018**) Les mutations des classes I, II et III sont associées à aucune fonction CFTR résiduelle; Par conséquent, les patients présentant ces mutations ont un phénotype sévère, tandis que la classe IV et V mutations ont une fonction résiduelle de la protéine CFTR, ce qui entraîne un phénotype pulmonaire moins sévère avec une insuffisance pancréatique (**Tableau IV**). (**Bono-neri2018**)

**Tableau IV:** Phénotype associé à chaque classe de mutation

	Les classes	Phénotype
Les mutations présentent un phénotype sévère	I	<ul style="list-style-type: none"> <li>➤ Cette mutation est responsable d'une forte incidence de meconiumileus.</li> <li>➤ la mutation <b>G542X</b> présentait une insuffisance pancréatique.</li> <li>➤ Les patients porteurs de la mutation <b>R1162X</b> a présenté des résultats de test élevés de chlorure de sueur (moyenne de 103 mEq / L), maladie pulmonaire légère à modérée, insuffisance pancréatique.</li> </ul>
	II	<p><b>Exemple : N1303K</b></p> <ul style="list-style-type: none"> <li>➤ Les personnes qui sont homozygotes pour cette mutation habituellement présente les résultats des tests de haute teneur en chlorure de sueur (moyenne de 98 mEq / L)</li> <li>➤ signes précoces de symptômes respiratoires.</li> <li>➤ une fonction pulmonaire réduite.</li> <li>➤ insuffisance pancréatique.</li> </ul>
	III	<p><b>Exemple : G551D</b>, lié à l'obstruction du passage de chlorure à travers le canal de la protéine CFTR</p> <ul style="list-style-type: none"> <li>➤ associée à une maladie pulmonaire.</li> <li>➤ insuffisance pancréatique</li> <li>➤ infection par PA</li> </ul>
Les mutations présentent un phénotype modéré	IV	<p><b>Exemple 1: P205S</b> caractérisé par des changements dans la conduction de chlorure à travers le canal de la protéine CFTR.</p> <ul style="list-style-type: none"> <li>➤ Une insuffisance pancréatique.</li> <li>➤ absence de symptômes gastro-intestinaux chez la plupart des patients.</li> </ul> <p><b>Exemple 2 :R117H</b>, est une mutation faux-sens, elle est considérée comme la deuxième mutation fréquente du gène CFTR derrière F508del.</p> <ul style="list-style-type: none"> <li>➤ caractérisé par une infertilité masculine.</li> <li>➤ Le risque d'atteinte respiratoire sévère liée à ce génotype est faible (1 l'âge adulte).</li> </ul>
	V	<p><b>Exemple : V 3272-26A&gt; G</b></p> <p>Résultat de la quantité insuffisante de la protéine CFTR normale présent dans la surface cellulaire.</p> <ul style="list-style-type: none"> <li>➤ associé au phénotype modéré de la maladie.</li> </ul>
	VI	<p><b>Exemple : R347H</b></p> <ul style="list-style-type: none"> <li>➤ une insuffisance pancréatique.</li> <li>➤ infection par PA.</li> </ul>

## II .4. Diagnostic et dépistage de la mucoviscidose

Le diagnostic de la mucoviscidose repose sur plusieurs tests paracliniques (voir tableau V)

**Tableau V:** Diagnostic paraclinique de la mucoviscidose (**Ouchtati A, Dali H, 2018**)

Diagnostic para clinique	Description
<b>Test de la sueur</b>	<ul style="list-style-type: none"> <li>➤ Il peut être réalisé dans l'âge de 15 jours pour un nourrisson à partir de 3 kg et reste positif toute la vie.</li> <li>➤ Ce test est l'examen de référence, il est indispensable pour confirmer le diagnostic de la mucoviscidose.</li> <li>➤ consiste à mesurer la concentration en -chlore d'un échantillon de sueur d'au moins 100 mg.</li> </ul>
<b>Différence de potentiel trans-épithéliale nasale</b>	<ul style="list-style-type: none"> <li>➤ Ce test consiste à mesurer la DDP trans-épithéliale par un voltmètre relié à deux électrodes en contact l'une avec la surface de l'épithélium nasal.</li> <li>➤ Elle permet en particulier d'identifier des formes atypiques de mucoviscidose et d'évaluer la fonction de la protéine CFTR en analysant le transport ionique.</li> <li>➤ Le seuil retenu pour le dosage de la TIR est 60 µg/l,</li> </ul>
<b>Dosage de la trypsine immuno-réactive (dépistage néonatal)</b>	<ul style="list-style-type: none"> <li>➤ le test le plus fréquemment rencontré, car il concerne tous les nouveaux nés depuis sa systématisation pour la première fois en France depuis 2002.</li> <li>➤ Le taux de la TIR est augmenté de façon non spécifique dans la mucoviscidose. Cette augmentation résulte d'une obstruction de canaux pancréatiques par des amas de protéines <i>in utero</i> qui génère un relargage d'enzymes dans le sang.</li> <li>➤ Le seuil retenu pour le dosage de la TIR est 60 µg/l, Dans les cas où cette valeur dépasse le seuil, cela conduit à réaliser un génotypage afin de rechercher les mutations les plus fréquentes du gène <i>CFTR</i>.</li> </ul>
<b>Analyse moléculaire du gène CFTR</b>	<p>La stratégie d'analyse moléculaire de gène <i>CFTR</i> est réalisée en 3 étapes :</p> <ul style="list-style-type: none"> <li>➤ La première étape consiste à réaliser une recherche des 30 mutations les plus fréquentes du gène. Par l'utilisation de plusieurs Kit commerciaux</li> <li>➤ La deuxième étape est réalisée en cas de génotype non établi, l'étude du gène est poursuivie par un balayage complet des séquences codantes du gène : analyse de toutes les mutations décrites dans les 27 exons.</li> <li>➤ Si le génotype reste toujours incomplet, une troisième étude sur les grands réarrangements est réalisée, se fait par PCR multiplexe avec des amorces fluorescentes ou par MLPA (Multiplex Ligation-dependant Probe Amplification)</li> </ul>

## IV: Étude *in Silico*: Contribution de la bioinformatique

### IV.1. Historique

La bioinformatique se définit actuellement comme un domaine de recherche qui analyse et interprète des données biologiques, au moyen de méthodes informatiques, afin de créer de nouvelles connaissances en biologie.

En effet, la bio-informatique peut être utilisée pour créer et partager un large éventail de bases de données et de logiciels, ce qui peut constituer un outil puissant scientifique, ce qui permet la comparaison, l'analyse et la visualisation des données biologiques provenant de l'étude de plusieurs organismes. (Stasinakis et al ; 2016)

Pour ce nouveau domaine scientifique, tout a vraiment commencé dans les laboratoires publics au milieu des années 1980, lorsque le laboratoire européen de biologie moléculaire (EMBL : European Molecular Biology Laboratory) et le département américain de la santé (NIH : National Institute of Health) ont créé les banques de données EMBL et GENBANK pour répertorier les séquences d'ADN découvertes par les biologistes, quatre ans plus tard, une nouvelle banque généralisée est apparue PIR (Protéine Information Resource) en USA et SwissProt (constituée à l'université de Genève à partir de 1986) (Boudjelal et Fettel, 2017)

En 1985, création d'un Programme d'alignement local de séquences "FASTA" (Pearson & Lipman).

En 1990, Programme d'alignement local de séquences "BLAST" (Legrand, 2016)

### IV.2. Description de quelques outils bioinformatiques utilisés dans les recherches biomédicales

Les variations de séquences au niveau des gènes sont le plus souvent responsables de troubles fonctionnels des protéines concernées menant à l'apparition de nombreuses maladies. Ces variations peuvent se produire aussi bien au niveau des régions géniques codantes (exons), qu'au niveau des introns et de ce fait avoir un effet sur le processus d'épissage alternatif, ce qui se traduit par diverses conséquences sur la structure et/ou la fonction de la protéine. L'étude *in silico* permet une meilleure compréhension de l'effet et de l'impact de différentes mutations sur le dysfonctionnement de la protéine. De nombreux logiciels et bases de données

ont été développés afin de prédire les différents effets et impacts de ces mutations délétères, par exemple

- ❖ **GenBank** : base de données généraliste américaine, qui contient des séquences nucléotidiques disponibles au public pour 370 000 espèces formellement décrits, Elle est construite et distribué par le Centre national pour l'information biotechnologique (NCBI), cette banque de donnée américain travaille en collaboration avec la DNA.)

Data Bank of Japan (DDBJ), European Bioinformatics Institute (EBI). ( **Dennis A. et al ,2017**)

- ❖ **BLAST** : (*Basic Local Alignment Search Tool*) Il Permet de trouver les régions similaires entre deux ou plusieurs séquences de nucléotides ou d'acides aminés et de réaliser un alignement de ces régions homologues. Ce programme permet de retrouver rapidement dans des bases de données, les séquences ayant des zones de similitude avec une séquence donnée.

**OMIM** : (*Online Mendelian Inheritance in Man*) : est une base de données de référence pour les maladies génétiques, née dans les années 1960 grâce au travail de Victor McKusick qui est souvent surnommé "the father of medical genetics" (**Rossi et al ,2011**), est le principal dépôt d'informations complètes, dont le commissaire sur les gènes et les phénotypes génétiques et les relations entre eux, un aperçu des caractéristiques cliniques de phénotypes. (**Amberger et al ,2018**)

- ❖ **Swiss model** : est un serveur Web de bioinformatique structurale dédié à la modélisation d'homologie de structures protéiques 3D. La modélisation par homologie est actuellement la méthode la plus précise pour générer des modèles de structure protéique tridimensionnels fiables. Elle est couramment utilisée dans de nombreuses applications pratiques. Les méthodes de modélisation d'homologie (ou comparative) utilisent des structures de protéines expérimentales ("modèles") pour construire des modèles de protéines liées à l'évolution ("cibles").

#### □ **COSMIC (the Catalogue Of Somatic Mutations In Cancer)**

Une base de données en ligne. Elle met en disposition les données des articles de la littérature scientifique. Elle est la ressource la plus vaste et la plus complète au monde pour explorer l'impact des mutations somatiques dans le cancer humain. Cette base de données est librement disponible sans restriction via son site web.

❖ **Ensembl** : Ensembl est un système bio-informatique d'annotation automatique de génomes. C'est un projet conjoint de l'European Bioinformatics Institute (EBI) et du Wellcome Trust Sanger Institute dont l'idée centrale est d'organiser de vastes champs d'information biologique autour de séquences génomiques.

Pour chaque génome analysé, Ensembl tente d'identifier par un processus automatique l'ensemble des gènes qu'il contient. Il s'appuie pour cela sur des données de séquences existantes (ARN, protéines), qu'il « raccroche » sur le génome, pour en déduire la structure des gènes. **(Herrero et al ,2016)**

# MATÉRIELS

# &



# MÉTHODES

Notre étude a été réalisée au sein de l'université de SAAD DAHLEB BLIDA-1 dans le cadre de la préparation du projet de fin d'étude intitulé «**étude in silico des mutations du gène CFTR provoquant la variabilité des formes de mucoviscidose**», menée sur une période de Cinq mois, allant du mois de février 2019 au mois juin 2019.

## I. Matériels

Cette étude a été réalisée grâce à :

- ❖ PC
- ❖ Connexion internet

Différents outils bioinformatiques

➤ Bases de données et logiciels bioinformatiques

- **GenBank** : La première étape de cette étude consiste à télécharger les séquences de la base de données nucléotidique GenBank.
- **Basic Local Alignment Search Tool (BLAST)** : pour l'alignement des séquences.
- **ORF Finder** : pour avoir la séquence protéique.
- **Online Mendelian Inheritance in Man(OMIM)** : pour la détection des mutations et pour voir la relation entre le génotype et le phénotype.
- **Swiss model** : est utilisé pour la visualisation des structures tridimensionnelles des protéines.
- **Ensembl** : Base de données spécialisée complémentaire à NCBI

❖ Séquence du gène CFTR

La séquence du gène CFTR ; d'une très grande taille, est identifiée dans la base de données sous l'ID **NC\_000007.14**.

**Tableau VI:** Séquences des exons étudiés

Ce tableau résume les séquences étudiées à travers les différentes bases de données

Séquence	Position	Numéro d'accèsion sur NCBI	Variant	ID sur NCBI
Gène CFTR	117479963..1 17668665	NC_000007.14	--	--
ARNm CFTR	1...6132	NM_000492.3	--	--
Exon10	1525...1716		c.1521_1523delCTT	rs113993960
Exon6	712...875		c.680T>G	--
Exon12	1717...1811		c.1624G>T	rs113993959
Exon24	4006...4095		c.3909C>G	rs 80034486
Exon12	1717...1811		c.1652G>A	rs75527207

## II. Méthodes

### II.1. Utilisation de GenBank

Consultée pour récupérer et télécharger la séquence de gène CFTR et les séquences des exons. En passant sur un moteur de recherche. (Figure 15)

1. Insérer le numéro d'accèsion.
2. Lancer la recherche.

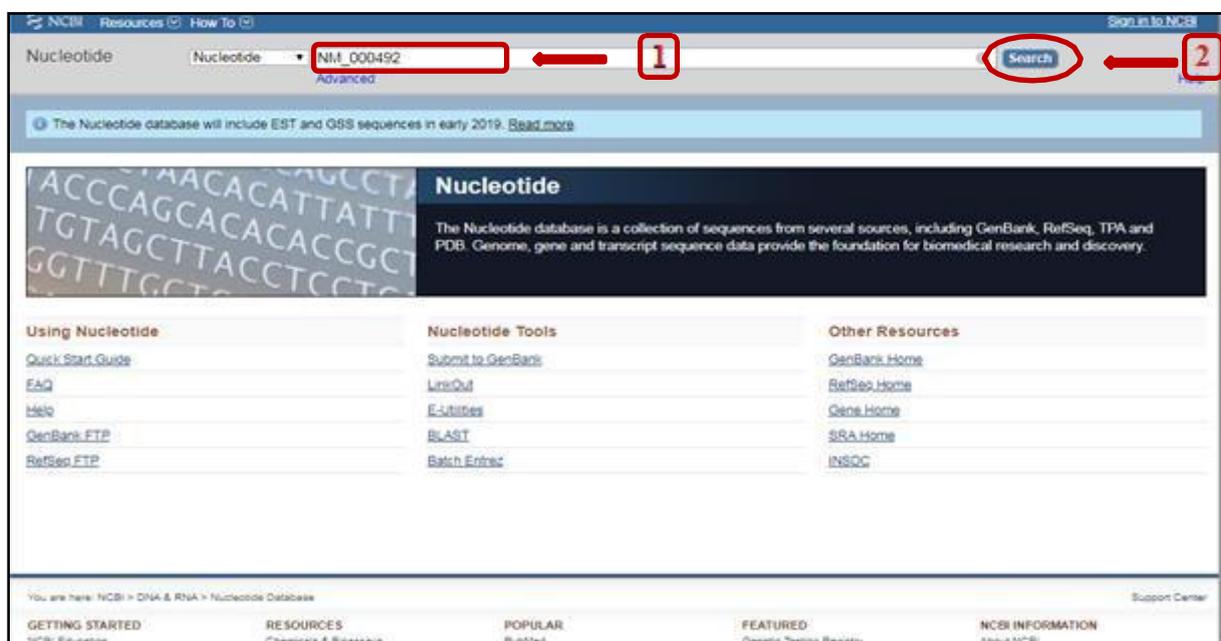


Figure 15: Utilisation de GenBank (personnelle, 2019)

#### II.1.A. Téléchargement des séquences

Une fois que la recherche de la séquence est lancée sur la plat-forme NCBI, un tableau de bord concernant la requête demandée est affichée sur la page de navigation, d'où on a la possibilité de télécharger les séquences sous format FASTA. (Figure 16)

The screenshot displays the NCBI GenBank interface for the Homo sapiens CF transmembrane conductance regulator (CFTR) mRNA sequence (NM\_000492.3). The interface includes a search bar at the top, a navigation menu, and a detailed view of the sequence. The sequence is 6132 bp long and is linear. The interface also provides options to analyze the sequence, such as Run BLAST, Pick Primers, Highlight Sequence Features, Find in this Sequence, and Show in Genome Data Viewer. Additionally, there are articles about the CFTR gene listed on the right side of the page.

**Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA**  
 NCBI Reference Sequence: NM\_000492.3  
[FASTA](#) [Graphics](#)

**LOCUS** NM\_000492 6132 bp mRNA linear PRI 21-MAY-2019  
**DEFINITION** Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA.  
**ACCESSION** NM\_000492  
**VERSION** NM\_000492.3  
**KEYWORDS** RefSeq; RefSeq Select.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
 Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 6132)  
**AUTHORS** Ruiz-Cabezas JC, Barros F, Sobrino B, Garcia G, Burgos R, Farhat C,  
 Castro A, Munoz L, Zambrano AK, Martinez M, Montalvan M and  
 Paz-Y-Mino C.  
**TITLE** Mutational analysis of CFTR in the Ecuadorian population using  
 next-generation sequencing  
**JOURNAL** Gene 696, 28-32 (2019)  
**PUBMED** [30763667](#)  
**REMARK** GeneRIF: Study from one hundred and forty-one suspected Ecuadorian  
 cystic fibrosis patients identified twenty-eight disease-causing  
 CFTR variants. The following six novel potentially  
 disease-associated variants were detected: 3 deletions

**Figure 16:** Tableau de bord de la requête sur NCBI (personnelle, 2019).

## II.2. Utilisation de Blast

Le Blaste est utilisé pour l'alignement entre les différentes séquences.

Nous avons utilisé pour notre étude le BLASTn et le BLASTp.

1. Choisir le type de blast selon le besoin.
2. Introduire le numéro d'accèsion ou bien la séquence sous forme FASTA.
3. Choisir la base de comparaison.
4. Lancer l'algorithme d'alignement. (Figure 17)

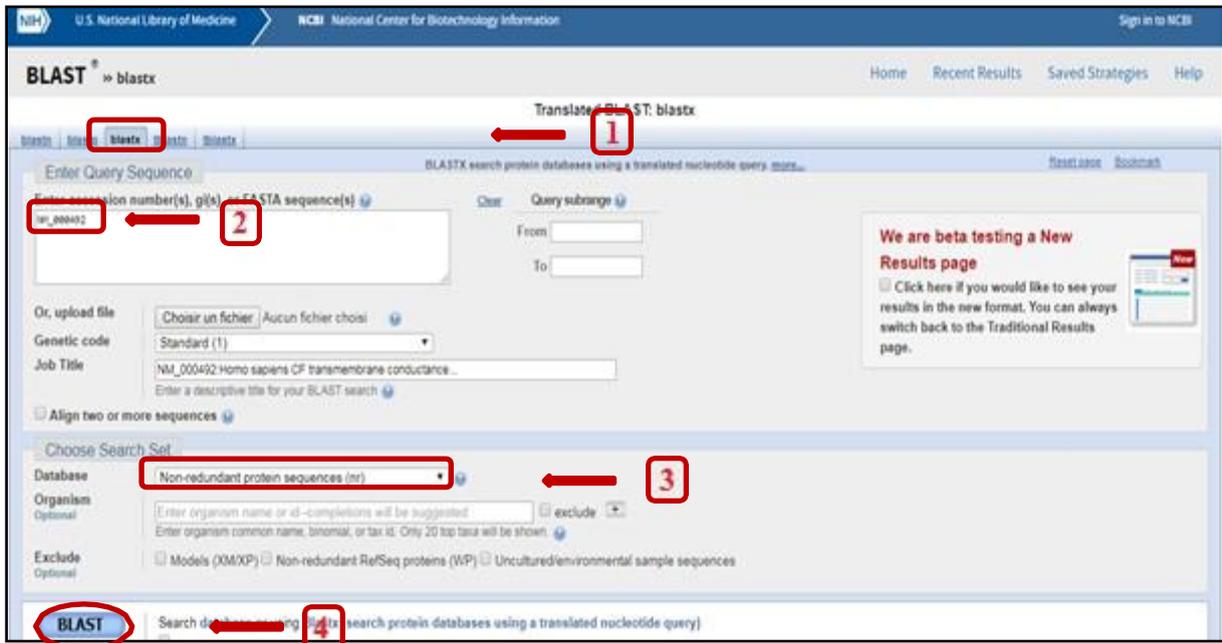


Figure 17: Utilisation de BLAST sur la plate-forme NCBI. (Personnelle, 2019)

### II. 3. Utilisation d’ORF Finder

Algorithme utilisé pour avoir les cadres de lectures ‘Open Frame Read’ de notre séquence. On passe sur la page d’accueil d’ORF Finder

1. Insérer la séquence.
2. Les ORFs peuvent commencer avec tous les codons.
3. Chercher les ORFs dans les cadre de lecture 1,2 et 3 du sens direct/indirect.
4. Ne renvoie que des ORF d’au moins 60 codons de long.
5. Utiliser le code génétique standard.
6. Lancer la recherche.

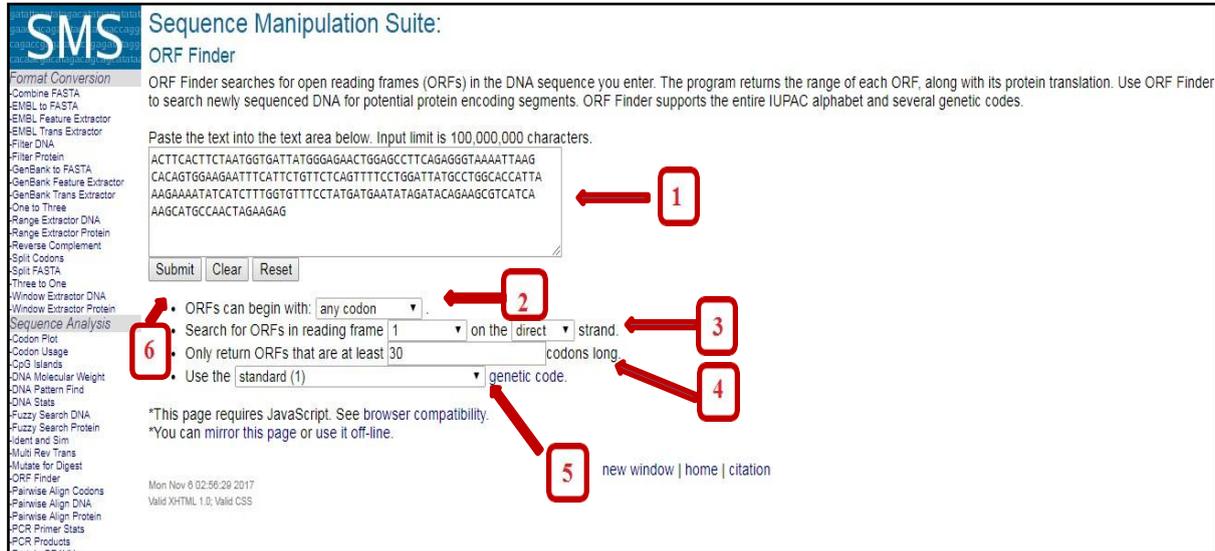


Figure 18: Utilisation d'ORF Finder (Personnelle, 2019)

## II.4. Utilisation d'Ensembl

On utilise cette base de données pour chercher une éventuelle relation entre la structure et la fonction de notre gène, pour cela, On passe sur la page d'accueil de la base de données Ensembl.

1. Choisir l'espèce Humaine.
2. Insérer le nom de gène.
3. Lancer la recherche. (Figure 19)

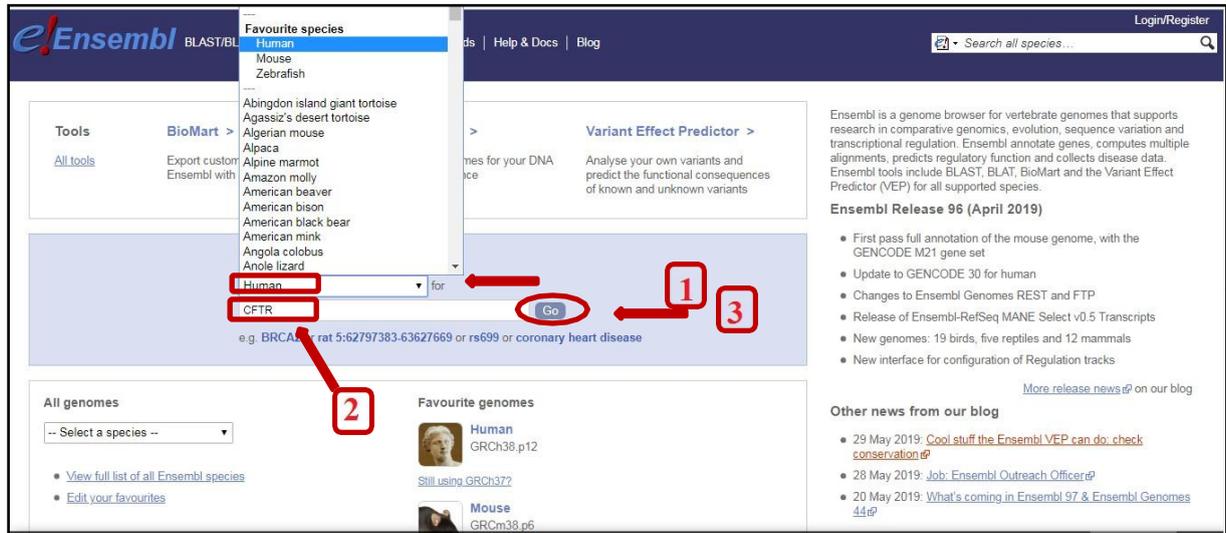


Figure 19: Utilisation d'Ensembl. (Personnelle, 2019)

### II.5. Utilisation de swiss-model

C'est une base de données spécialisée dans la quelle on retrouve différentes informations concernant notre requête (Structure 3D, séquence protéique, alignement multiple... etc)

On passe sur la page d'accueil de swiss-prot pour commencer la modélisation. (Figure 20)

1. On insère la séquence protéique ou bien le numéro d'accèsion.
2. Lancer la recherche de modèle.

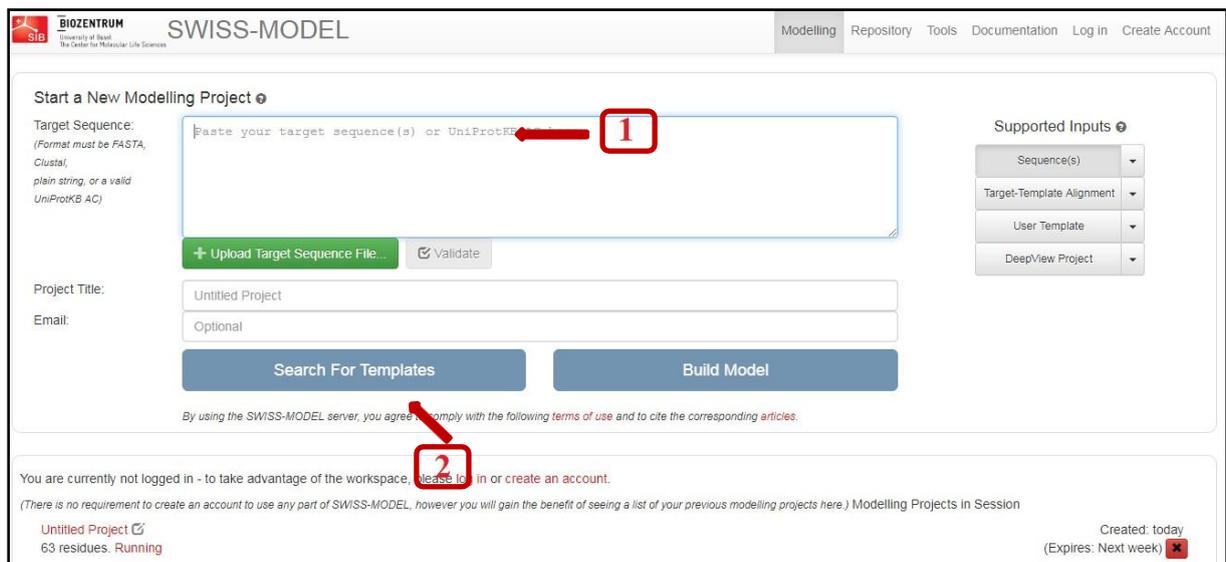


Figure 20: Utilisation de swiss-model. (Personnelle, 2019)

# RESULTATS

&



# DISCUSSION

## Résultats et discussion

### I. Les résultats de GenBank

Les résultats obtenus sous forme FASTA débutent par une ligne commençant par :

- Le signe « > » il est recommandé que toutes les lignes de texte soient plus courtes que 80 caractères.
- L'identifiant du gène.
- Quelques commentaires.
- Puis sur la deuxième ligne début la chaîne de séquence. (**Figure 21**)

The screenshot shows the NCBI GenBank FASTA viewer interface. At the top, there is a notification: "The Nucleotide database will include EST and GSS sequences in early 2019. Read more." Below this, the FASTA format is indicated. The main title is "Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA". Below the title, the NCBI Reference Sequence is given as "NCBI Reference Sequence: NM\_000492.3". There are links for "GenBank" and "Graphics". The sequence itself starts with ">NM\_000492.3 Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA" followed by the nucleotide sequence. On the right side, there are several interactive options: "Send to:", "Change region shown", "Customize view", "Analyze this sequence" (with sub-options like "Run BLAST", "Pick Primers", "Highlight Sequence Features", "Find in this Sequence", "Show in Genome Data Viewer"), "Articles about the CFTR gene" (with sub-options like "Restoration of F508-del Function by Transcomplementati", "Mutational analysis of CFTR in the Ecuadorian population using next-generation se", "Genetic and phenotypic traits of children and adolescents with cystic fit"), and "Pathways for the CFTR gene" (with sub-options like "Ub-specific processing proteases", "Deubiquitination", "Cargo recognition for clathrin-mediated endocytosis").

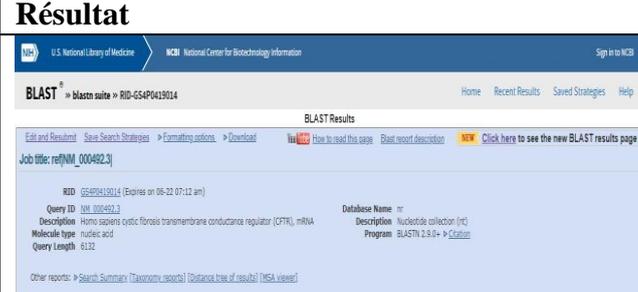
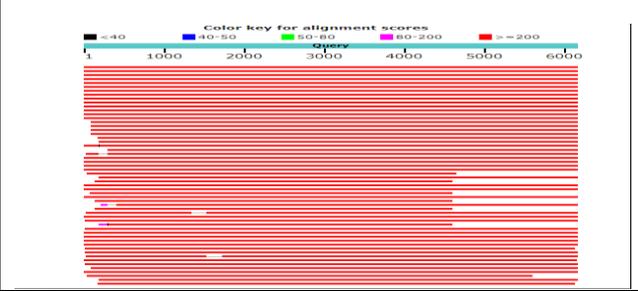
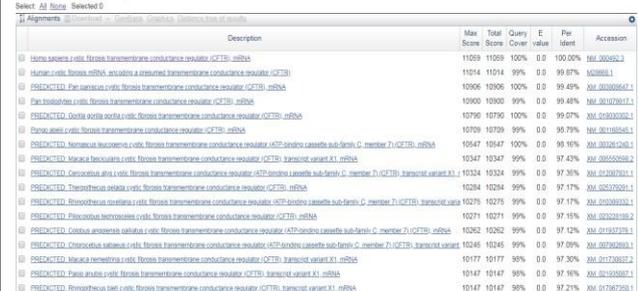
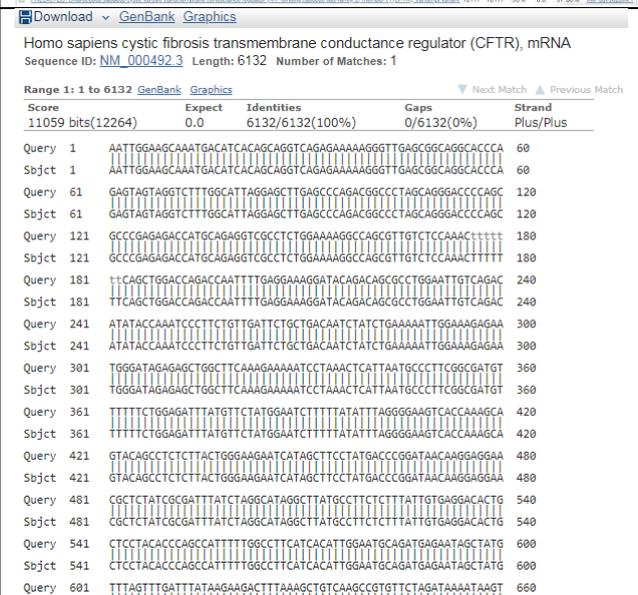
**Figure 21:** Séquence d'ARNm du gène CFTR sous format FASTA. (Personnelle, 2019)

La figure 21 nous donne la description de la séquence ARN du gène répertorié dans NCBI, ainsi que les différents identifiants. (Voir tableau VII)

## II. Les résultats de Blast n de l'ARNm CFTR

Les résultats de Blast n (alignement global nucléotidique) sont montrés dans le tableau VII

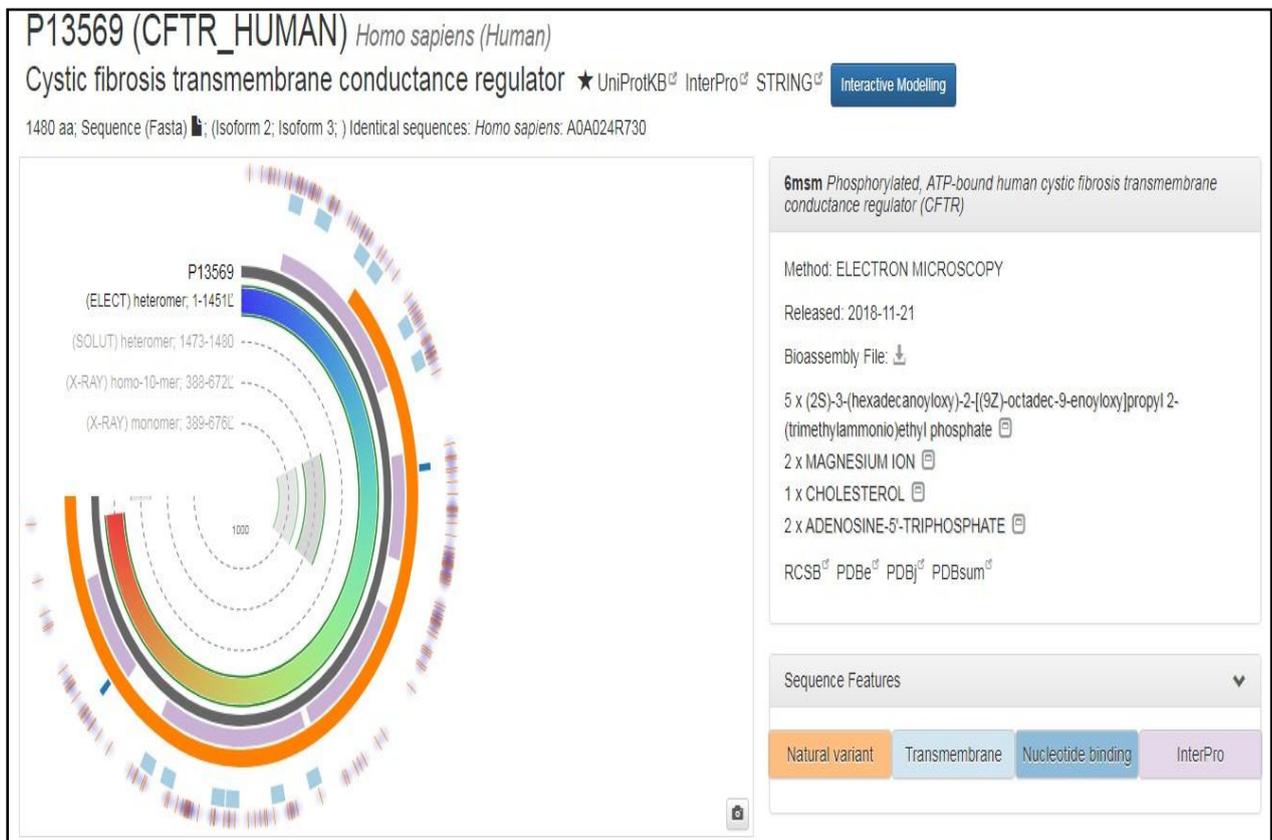
**Tableau VII:** Résultats de BLAST de gène CFTR

Résultat	Rôles	descriptifs
	<p>Récapitulatif de la requête.</p>	<ul style="list-style-type: none"> <li>Séquence soumise</li> <li>Identifiant</li> <li>Longueur</li> <li>Type</li> <li>Banque de donnée interrogée</li> <li>programme utilisé</li> </ul>
	<p>Représentation graphique des résultats.</p>	<ul style="list-style-type: none"> <li>les traits de couleur représentent à un alignement entre la séquence soumise et une séquence de la banque.</li> <li>la couleur correspond au score.</li> <li>La longueur correspond à la taille de l'alignement.</li> </ul>
	<p>Représentation global des résultats.</p>	<ul style="list-style-type: none"> <li>Chaque ligne correspond à un trait colorée la représentation graphique.</li> <li>Une liste des séquences avec leur numéro d'accèsion, le score d'alignement et l E-valu.</li> </ul>
	<p>Description des résultats, des alignements.</p>	<ul style="list-style-type: none"> <li>Query : la séquence soumise.</li> <li>Subject : la séquence trouvée dans la base de données.</li> <li>Les grandeurs propre à chaque alignement (score, E-valu, identité, positif, de gap).</li> </ul>

### III. Les résultats de swiss-prot d'ARNm du gène

**CFTR ID de CFTR sur swiss-model : P13569**

#### III.1. Représentation topographique des différents isoformes du gène CFTR



**Figure 22:** Représentation topographique des différents isoformes du gène CFTR sur la base de données swiss-model. (Personnelle, 2019)

Le motif correspond à une courte séquence (conservée) d'acides aminés (ou de nucléotides) que l'on peut associer aux rôles (fonctionnel ou structural) de la macromolécule, pendant que son motif structural est quelque peu différent : en effet, il tient compte du repliement dans l'espace de la chaîne polypeptidique et constitue une structure constituée de plusieurs régions de la protéine qui adoptent une structure secondaire, reliées entre elles. (Voir Figure 22)

### III.2. La séquence protéique du gène CFTR

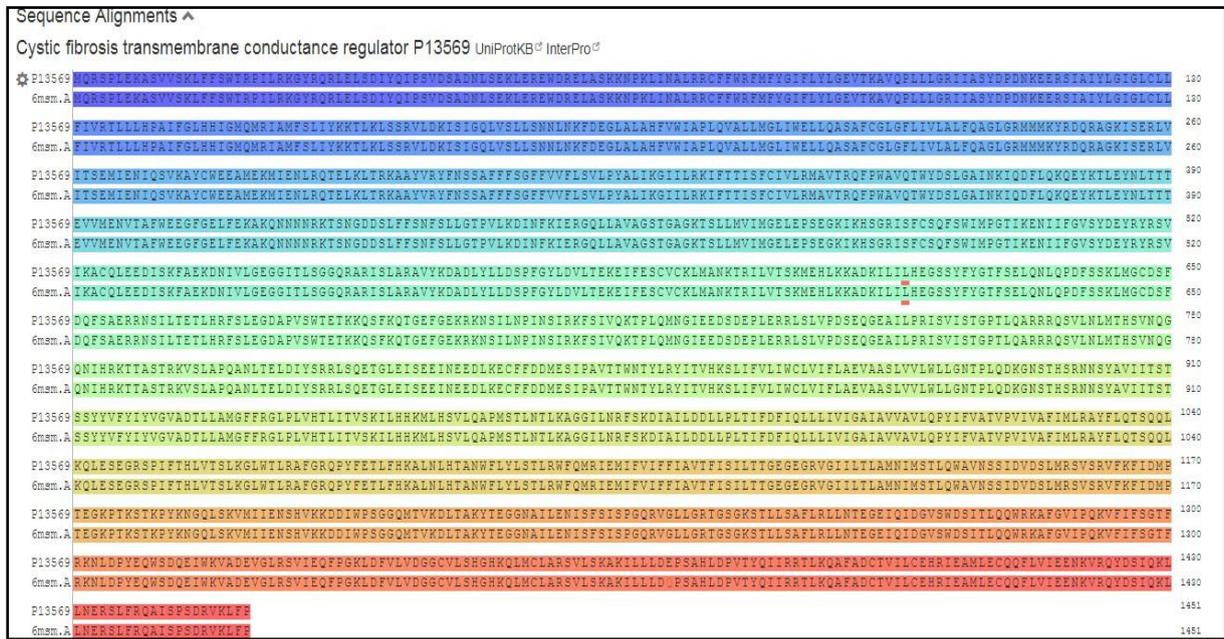


Figure 23: Séquence protéique du gène CFTR (personnelle, 2019).

### III.3. Structure tridimensionnelle de la protéine CFTR

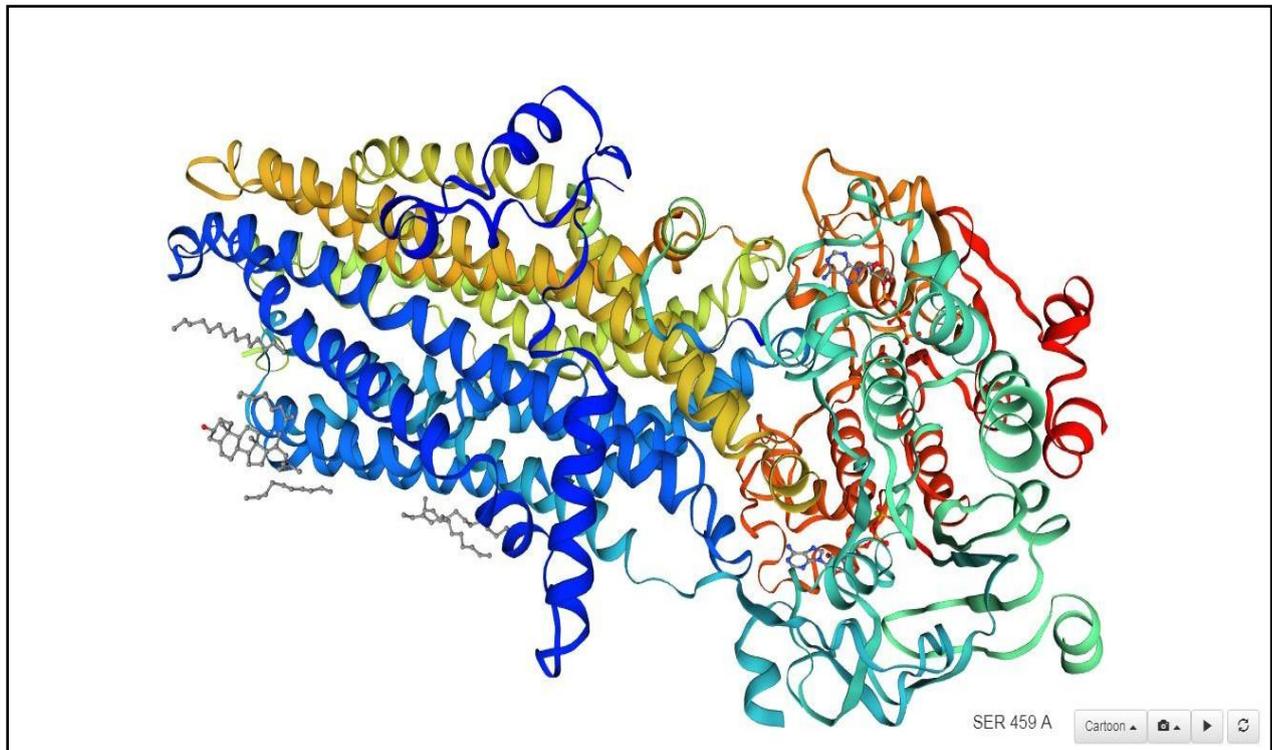
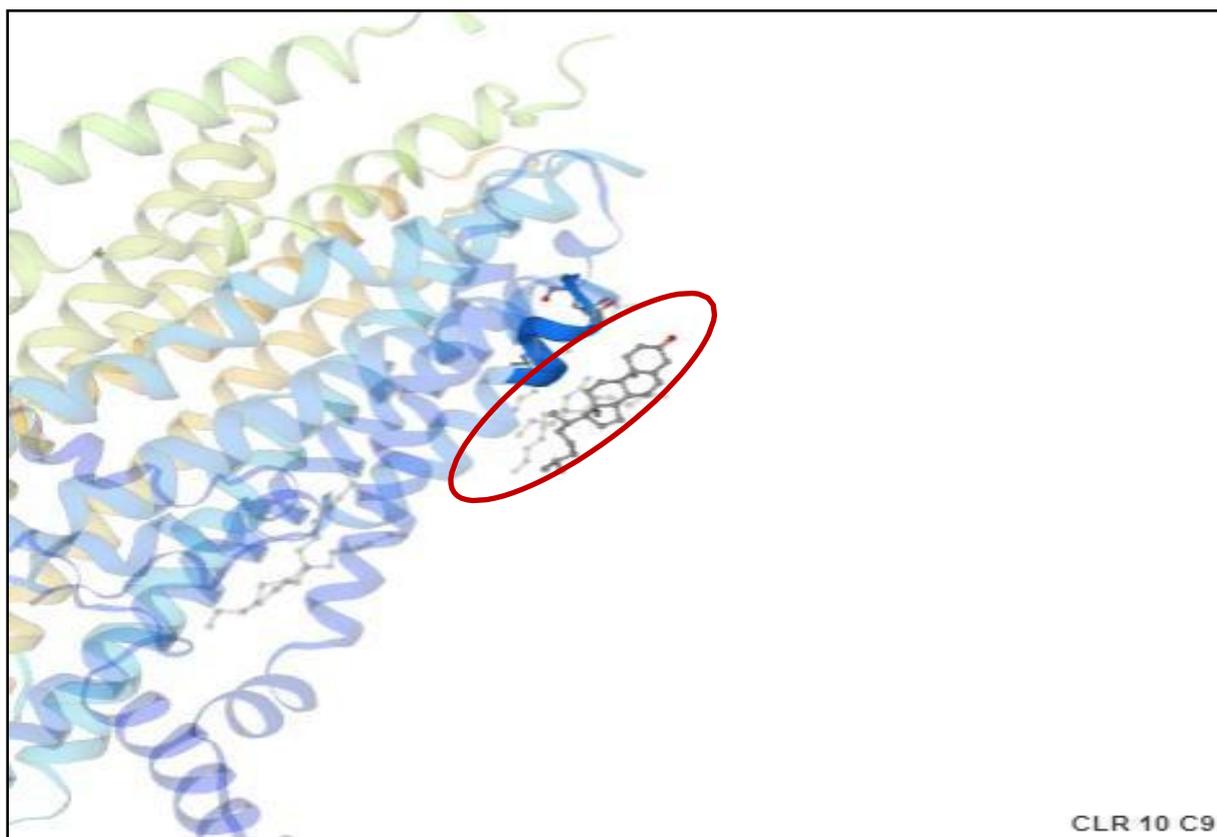


Figure 24: Structure 3D de la protéine CFTR. (Personnelle, 2019)

- ❖ Chaque couleur présent dans la séquence protéique (**Figure 23**) représente la même couleur dans la structure 3D (**Figure 24**).
- ❖ Pour savoir la position d'un n'importe quel acide amine sur la structure Tridimensionnel (**Figure 24**) on clique sur l'acide aminé, il va apparaitra sur la structure. (**Figure 23**).

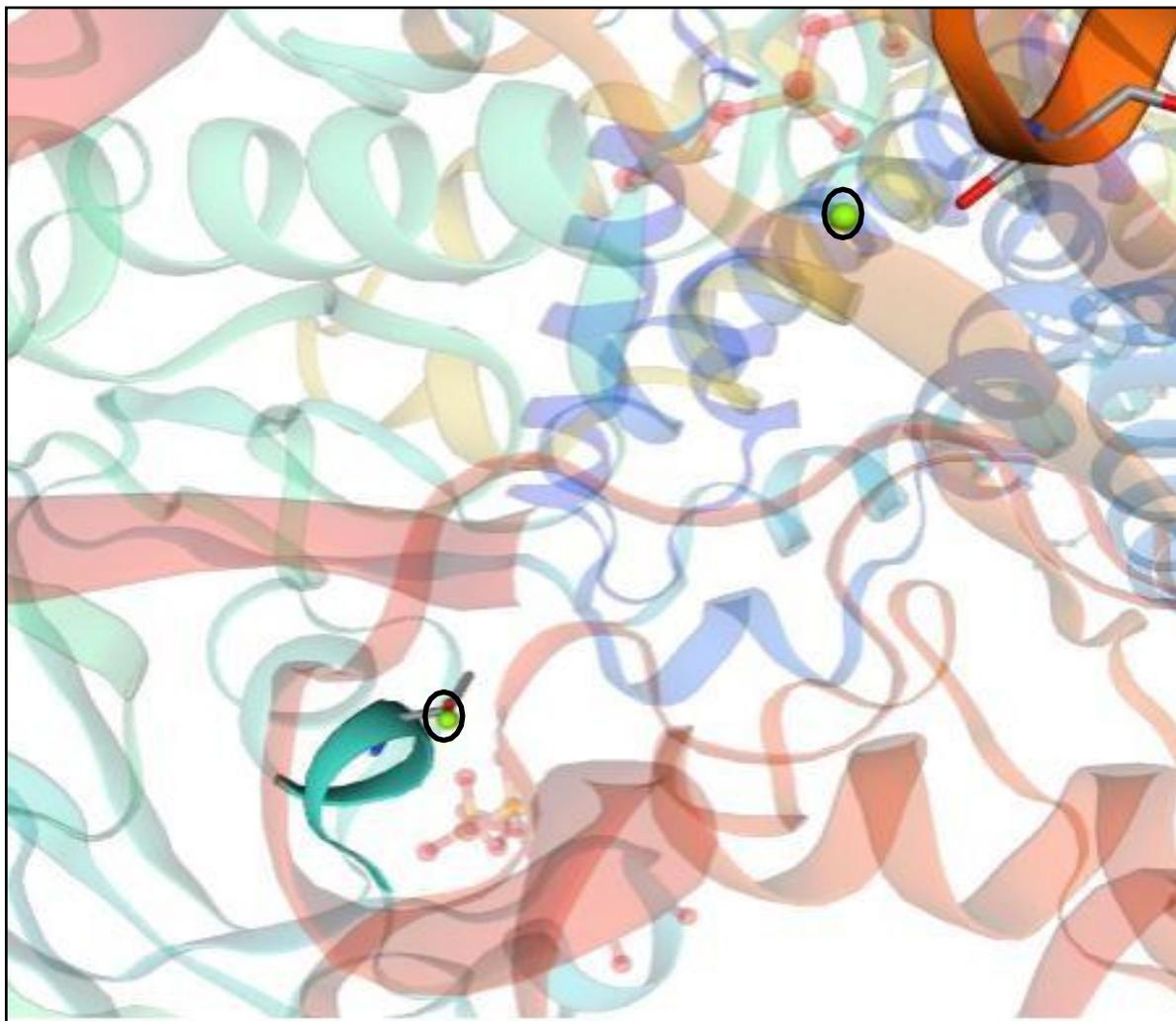
On peut voir aussi les ions de magnésium, le cholestérol et adénosine 5 triphosphate représentés dans la figure disponible sur cette base de données.

**La figure 25** nous montre la structure de la molécule de cholestérol présente au niveau de la protéine CFTR, Par l'utilisation le logiciel swiss model.



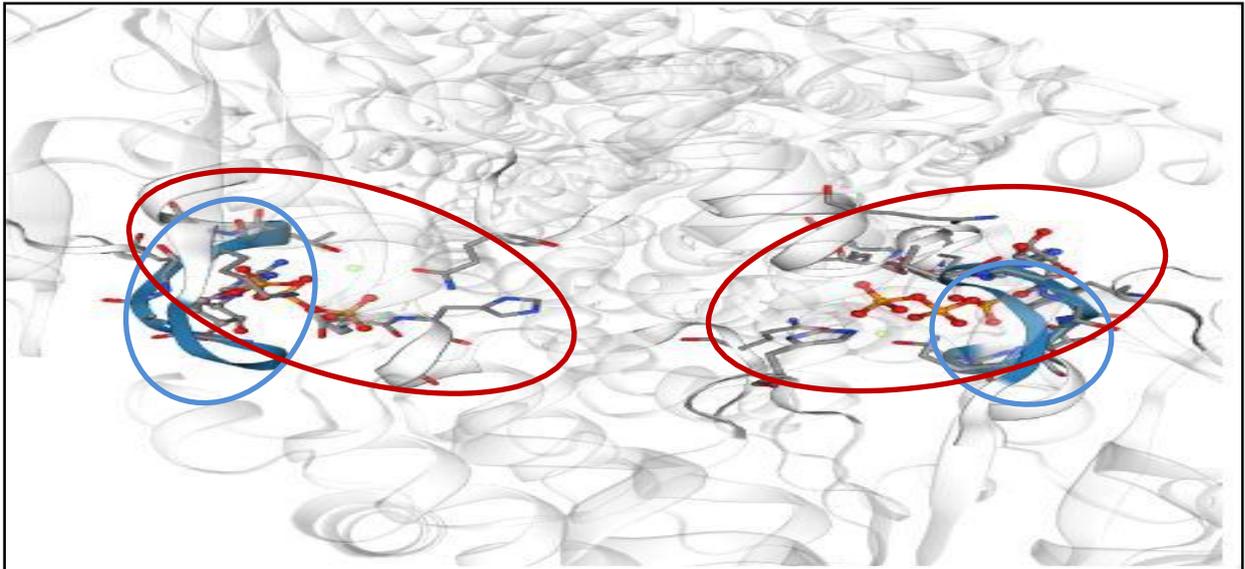
**Figure 25:** Structure moléculaire 3D de cholestérol sur le logiciel swiss-model. (Personnelle, 2019)

La figure 26 montre les deux ions de magnésium en vert présent au niveau de la protéine CFTR. Par l'utilisation de logiciel swiss model.



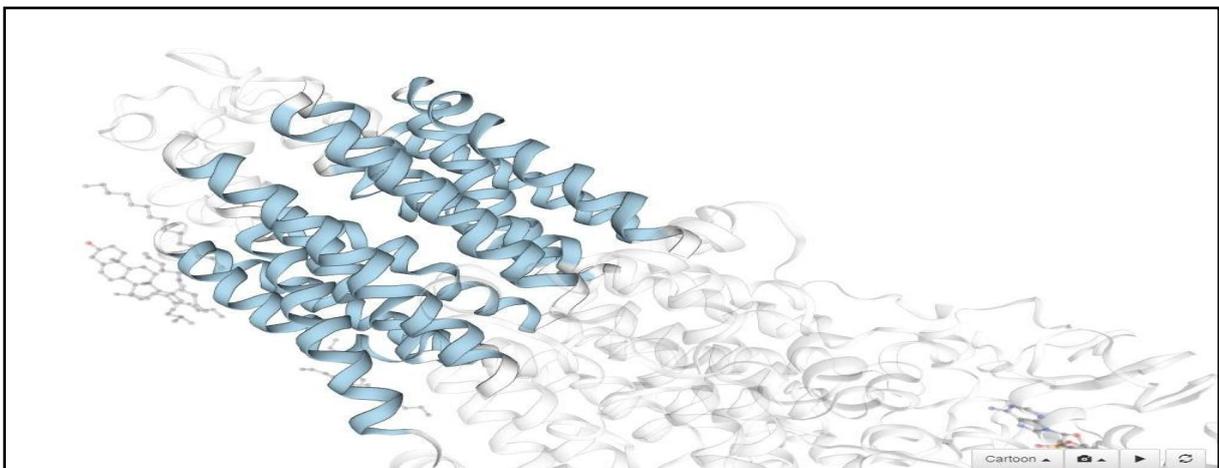
**Figure 26:** Représentation 3D des ions de magnésium sur swiss-model. (Personnelle, 2019)

La fixation de deux molécules d'ATP sur les deux domaines NBD1,NBD2 est nécessaire pour l'activation de la protéine CFTR pour l'ouverture du canal CFTR. Le logiciel swiss model montre les deux domaines NBDs en feuillets B en bleu .les deux molécules d'ATP par des couleurs différentes. **(Figure 27)**



**Figure 27:** Représentation 3D de la liaison des deux molécules d'ATP sur les deux domaines NBD1-NBD2 sur swiss-model. **(Personnelle, 2019).**

La protéine CFTR contient deux domaines transmembranaires hydrophobes TM1 et TM2, chaque domaine composé de six hélices, qui délimitent un pore transmembranaire. **(Gaillard et al ,2009).** La **Figure 28** montre les deux domaines transmembranaires avec les six hélices.



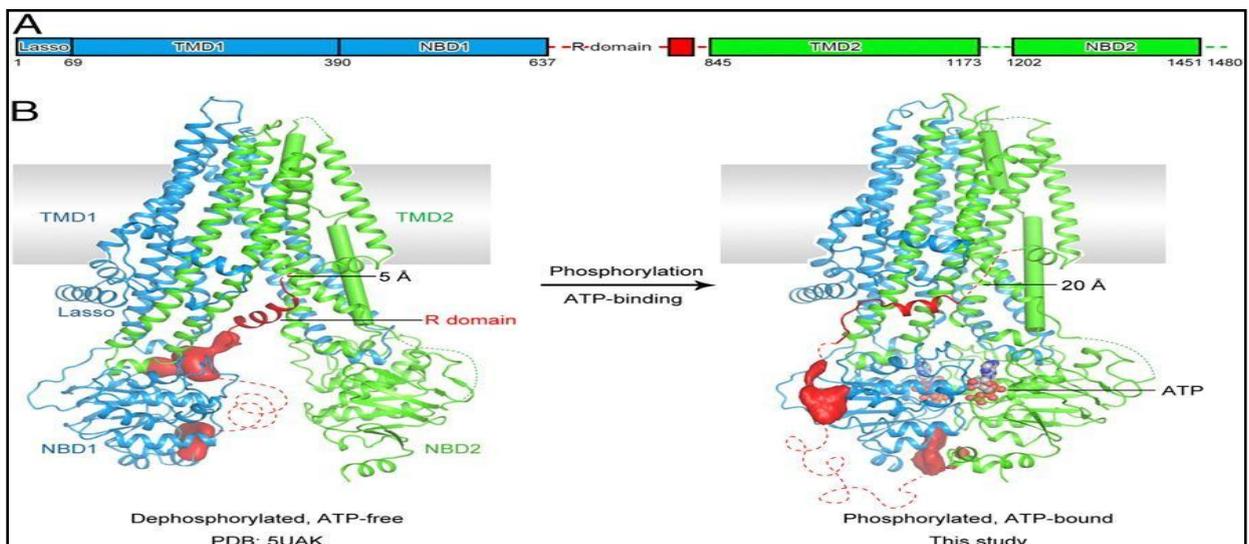
**Figure 28:** Représentation 3D des deux domaines transmembranaires de la protéine CFTR. **(Personnelle 2019)**

Plus de 2000 mutations ont été identifiées tout au long du gène CFTR, la **figure 29** nous montre tous les variants du gène CFTR en 3D qui sont présents dans la base de données swiss model



**Figure 29:** Variantes du gène CFTR dans swiss model (personnelle, 2019).

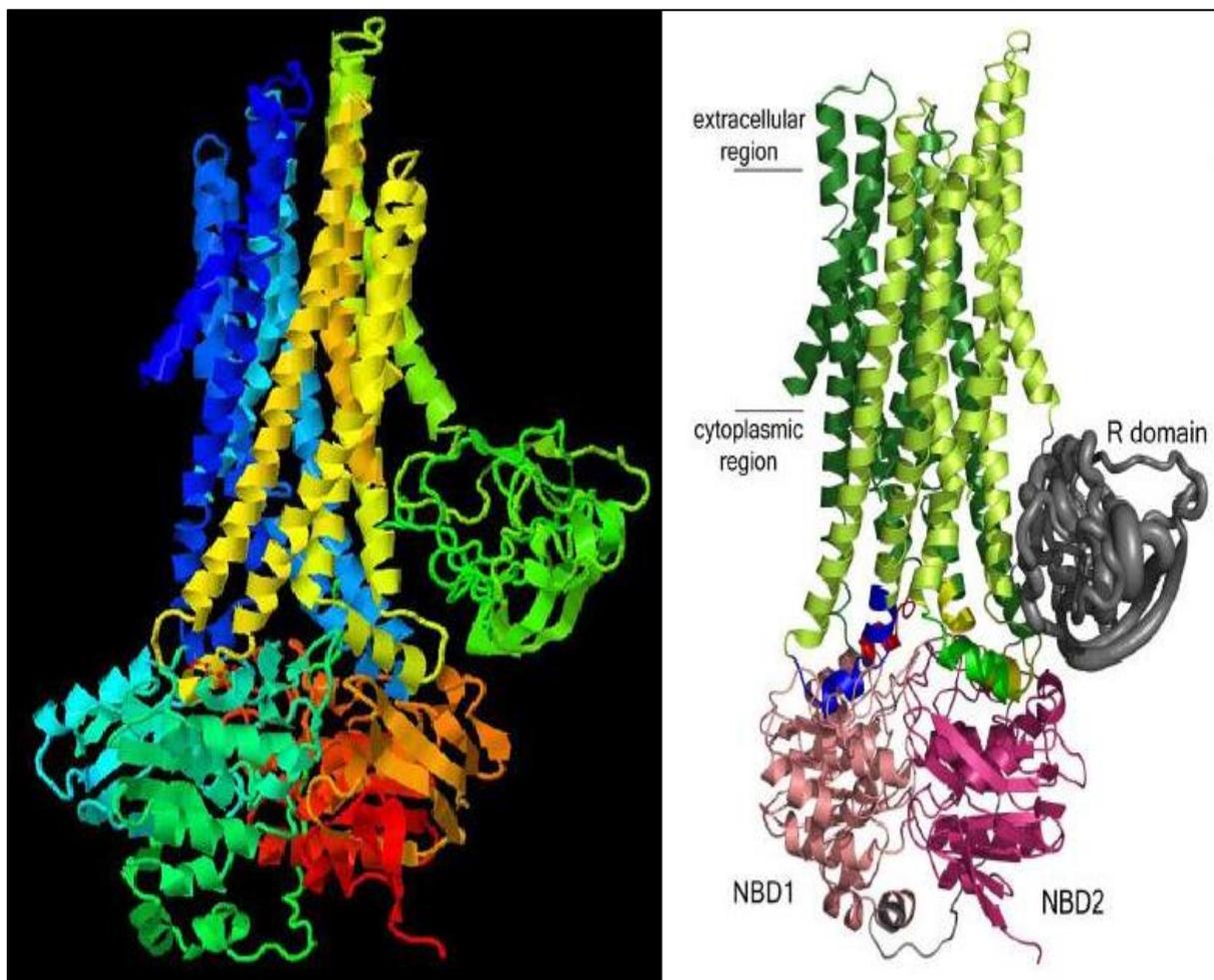
CFTR est un polypeptide unique contenant un motif lasso N-terminale, deux domaines transmembranaires (TMD), et deux domaines de fixation des nucléotides (NBD)(**Figure 30**).(Zhang et al, 2018)



**Figure 30:** Domaines de la protéine CFTR et les deux molécules d'ATP. (Zhang et al, 2018)

La **figure 31** obtenue avec MolUSc à partir de la structure 3D décrite par **Serohijos en 2011** et qui montre la structure tridimensionnelle de la protéine CFTR sauvage.

Cette structure montre une portion constituée de motifs en hélice et correspondant aux domaines transmembranaires (MSD1 et MSD2 - Membrane Spaning Domain). Ils sont surmontés par deux domaines intracellulaires capables de se lier à l'ATP, appelés NBD1 et NBD2 (Nucleotide Binding Domain). Enfin, une région régulatrice (R) leur est juxtaposée. (**Serohijos et al,2011**)

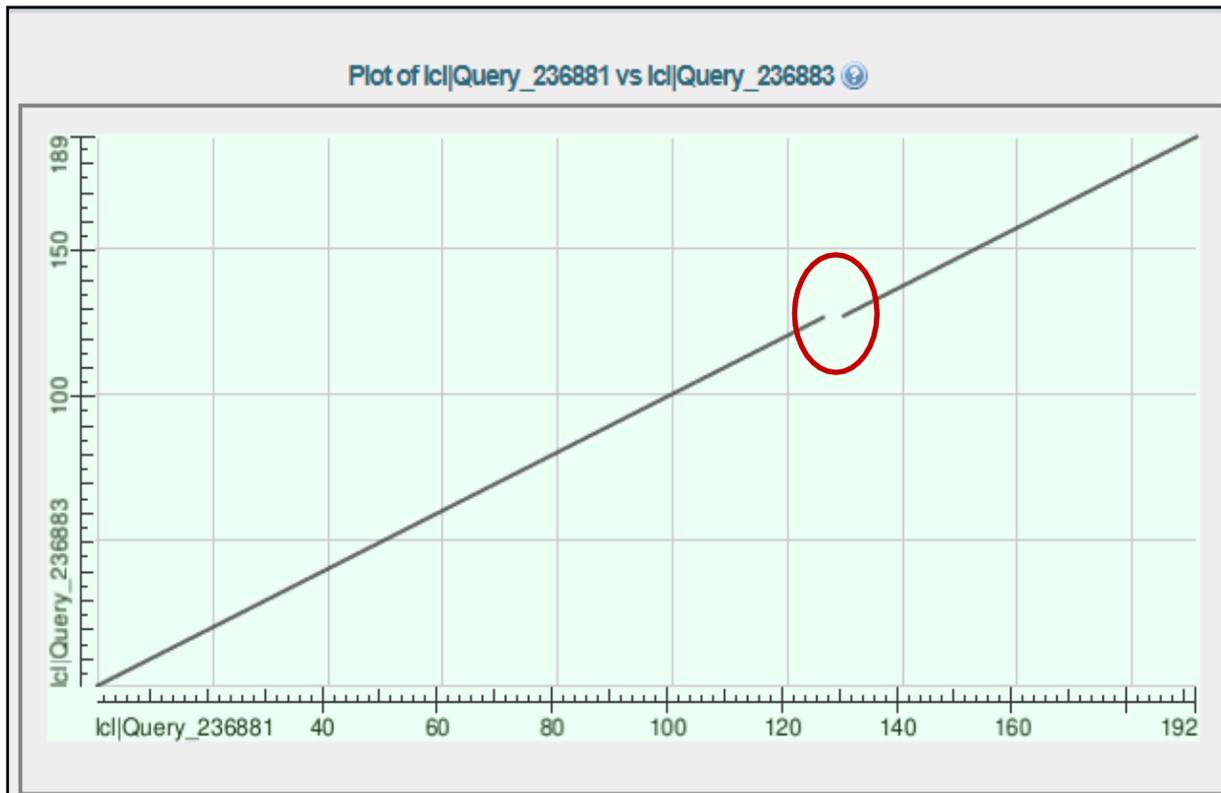


**Figure 31:** Modèle théorique de la structure de la protéine CFTR sauvage. (**Serohijos et al.2008**)

## II. Analyse de la mutation c.1521\_1523delCTT (F508del) présente au niveau de l'exon 10 de gène CFTR

ID sur NCBI : rs113993960

### II.1.Séquence sauvage/mutée



**Figure 32:** Dot matrix view de la mutation c.1521\_1523delCTT (F508del) (Personnelle, 2019)

- La séquence de requête est représentée sur l'axe des X et les nombres représentent les bases / résidus de la requête.
- Le sujet est représenté sur l'axe des ordonnées et les chiffres représentent à nouveau les bases / résidus du sujet.
- Les alignements sont affichés dans le graphique sous forme de lignes.
- Le nombre de lignes affichées dans le graphique est identique au nombre d'alignements trouvés par BLAST.

La superposition entre la séquence sauvage et la séquence mutée de l'exon 10 nous donne deux alignements :

1<sup>er</sup> alignement :(1-127)

2<sup>ème</sup> alignement : (131-192)

Les résultats des alignements nous a permis aussi l'identification de mutation de type délétion au niveau des bases 127-129. Comme le montre **la Figure 32**.

## II.2. Résultats de l'alignement nucléotidique

Les résultats de l'alignement nucléotidique montrent la disparition (délétion, gaps) de 3 nucléotides **CTT** à la position 1521\_1523.

Sequence ID: Query_236883 Length: 189 Number of Matches: 1				
Range 1: 1 to 189 <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>				
Score	Expect	Identities	Gaps	Strand
335 bits(181)	4e-97	189/192(98%)	3/192(1%)	Plus/Plus
Query 1	ACTTCACCTCTAATGGTGATTATGGGAGA	ACTTCACCTCTAATGGTGATTATGGGAGA	ACTTCACCTCTAATGGTGATTATGGGAGA	60
Sbjct 1	ACTTCACCTCTAATGGTGATTATGGGAGA	ACTTCACCTCTAATGGTGATTATGGGAGA	ACTTCACCTCTAATGGTGATTATGGGAGA	60
Query 61	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	120
Sbjct 61	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	AGTGGAAGAATTTTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA	120
Query 121	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	180
Sbjct 121	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	AATATCCTCTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC	177
Query 181	CAACTAGAAGAG	CAACTAGAAGAG	CAACTAGAAGAG	192
Sbjct 178	CAACTAGAAGAG	CAACTAGAAGAG	CAACTAGAAGAG	189

**Figure 33:** Délétion des trois nucléotides **CTT** à la position 1521-1523 sur l'exon 10. (Personnelle, 2019)

## II.3. Les résultats dans ORF Finder de la mutation

### c.1521\_1523deCTT (F508del)

#### I.3.a. Séquence nucléotidique et protéique de l'exon 10 sauvage

Présence des trois nucléotides **CTT** à la position 1521-1523 dans la séquence nucléotidique qui se traduit en **Phénylalanine** au niveau de la séquence protéique. (Voir figure 34)

```

ORF Finder results
Results for 192 residue sequence "Untitled" starting "ACTTCACTTC"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 192.
ACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAATTAAGCAC
AGTGGAGAAGAAATTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA
AATATCATCTTTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC
CAACTAGAAGAG

>Translation of ORF number 1 in reading frame 1 on the direct strand.
TSLLMVIMGELEPSEGGKIKHSGRISFCSQFSWIMPGTIKENIIFEVSYDEYRYSVIKAC
QLEE

```

**Figure 34:** Traduction de la séquence non mutée de l'exon 10. (Personnelle, 2019)

### I.3.b. Mutation F508del au niveau de l'exon 10

La délétion de trois nucléotides **CTT** à la position 1521-1523 dans l'exon 10, est traduite par une délétion de l'acide aminé **Phényle-alanine** à la position 508 de la séquence protéique. (Figure 35)

```

ORF Finder results
Results for 189 residue sequence "Untitled" starting "ACTTCACTTC"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 189.
ACTTCACTTCTAATGGTGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAATTAAGCAC
AGTGGAGAAGAAATTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA
AATATCATTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAA
CTAGAAGAG

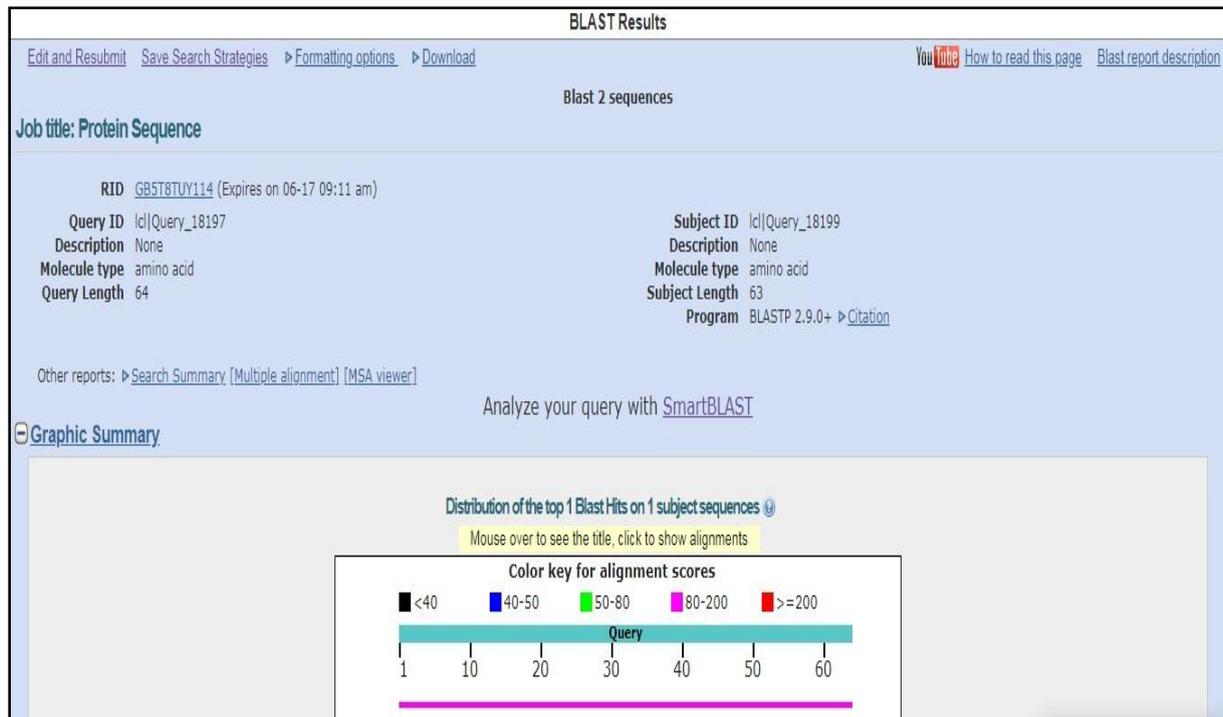
>Translation of ORF number 1 in reading frame 1 on the direct strand.
TSLLMVIMGELEPSEGGKIKHSGRISFCSQFSWIMPGTIKENIILVSYDEYRYSVIKACQ
LEE

```

**Figure 35:** Traduction de la séquence mutée de l'exon 10. (Personnelle, 2019)

## II.4. Résultat de l'alignement protéique entre la séquence mutée et non mutée

- La séquence protéique soumise à une taille de 64 acides aminés.
- la séquence de la base de données à une taille de 63 acides aminés.
- Le programme utilisé est le BLAST P.
- Le score d'alignement de cette séquence entre 80-200.



L'homologie de cette séquence avec la séquence model est de 98%, d'une similarité de 63/64. Cette séquence a un score égale à 126 .l'e-valu égale 9e-46. La mutation F508del caractérisée par la délétion d'une **Phénylalanine** à la position 508. (Voir figure 38)

unnamed protein product

Sequence ID: Query\_18199 Length: 63 Number of Matches: 1

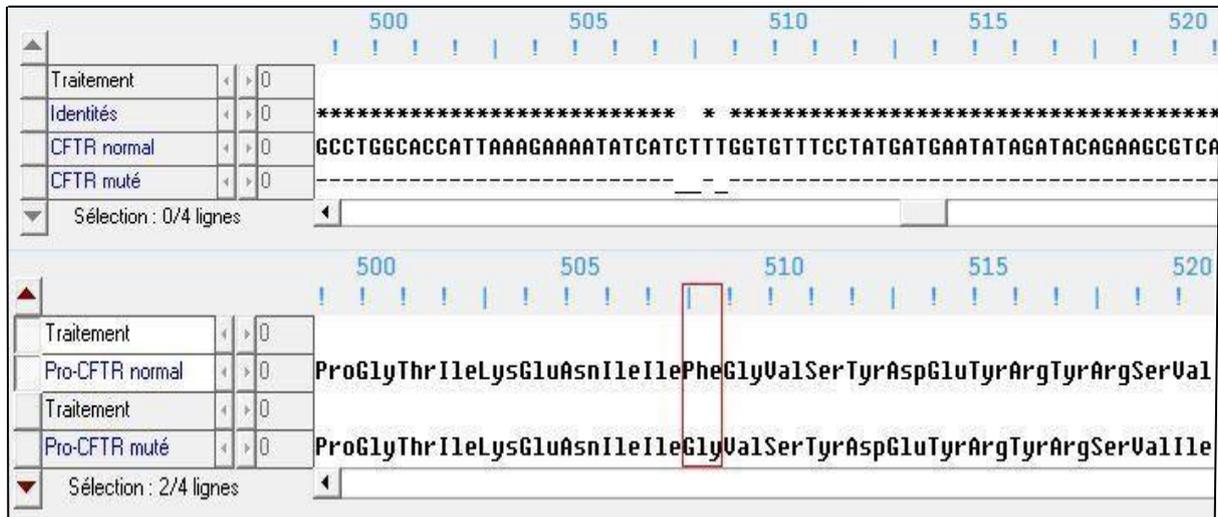
Range 1: 1 to 63 Graphics

Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
126 bits(316)	9e-46	Compositional matrix adjust.	63/64(98%)	63/64(98%)	1/64(1%)
Query 1	TSLLMVIMGELEPSEGKIKHSGRISFCSQFSWIMPGTIKENIIFGVSYDEYRYSVIKAC				60
Sbjct 1	TSLLMVIMGELEPSEGKIKHSGRISFCSQFSWIMPGTIKENIIFGVSYDEYRYSVIKAC				59
Query 61	QLEE 64				
Sbjct 60	QLEE 63				

Figure 36: Alignement de la séquence protéique de l'exon 10. (Personnelle, 2019)

La PHE 508 a disparu, il y a eu délétion de **CTT** (dernière base du 507 et 2 premières du 508) (voir figure 37) .Dans Anagène, un alignement avec discontinuités avec la molécule entière, permettra de retrouver leurs positions dans la séquence. (**Anonyme 5**)

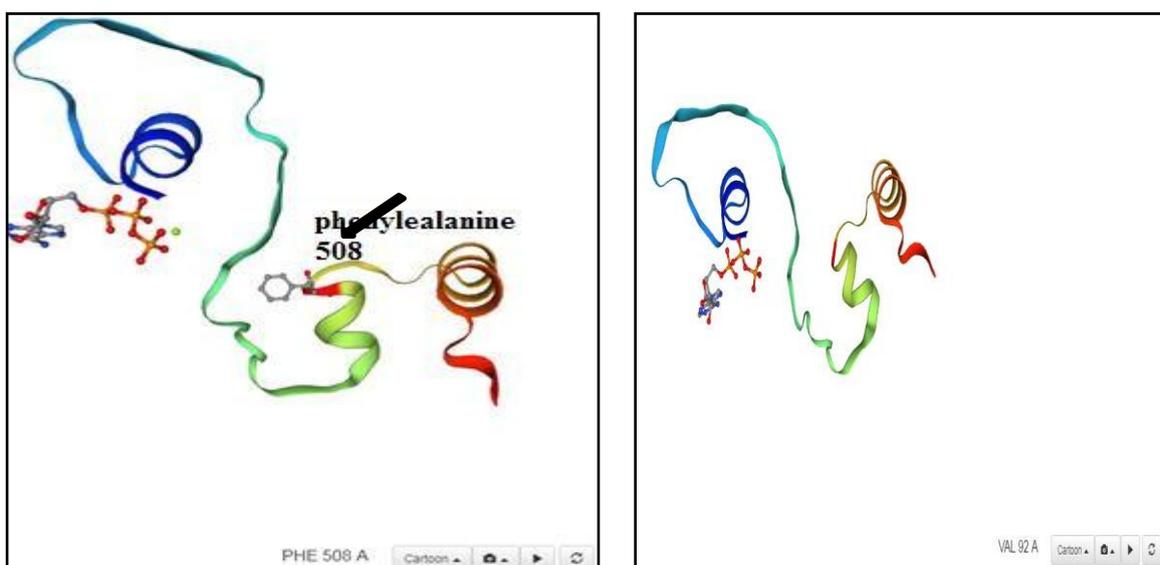


**Figure 37:** Résultats d'alignements nucléotidique et protéique de l'exon 10 avec Anagène. (**Anonyme5**)

## II.5. Structure 3D de la protéine sauvage et mutée de l'exon 10(F508del)

Présence d'acide aminé phénylalanine à la position 508 entre une glycine 509 et isoleucine dans la protéine normale (non mutée).

Absence de la **Phénylalanine** entre la **Glycine** et l'**Isoleucine** dans la protéine mutée.



**Figure 38:** structure 3D de la protéine de l'exon 10 sauvage et mutée (F508 del). (**Personnelle, 2019**)

### Protéine normale

Le peptide utilisé pour établir la structure tridimensionnelle comporte 26 acides aminés, qui correspondent au fragment Méthionine 498 - Alanine 523 de la CFTR. La structure est une hélice alpha, entre les acides aminés 4 et 25. La position de la phénylalanine 508 est indiquée par la couleur jaune. (Anonyme 5)

### Protéine modifiée

Le peptide utilisé pour établir la structure tridimensionnelle comporte 25 acides aminés, qui correspondent au fragment Méthionine 498 - Alanine 522 de la CFTR  $\Delta$ F508. La structure est une hélice alpha, entre les acides aminés 4 et 24. (Anonyme 5)

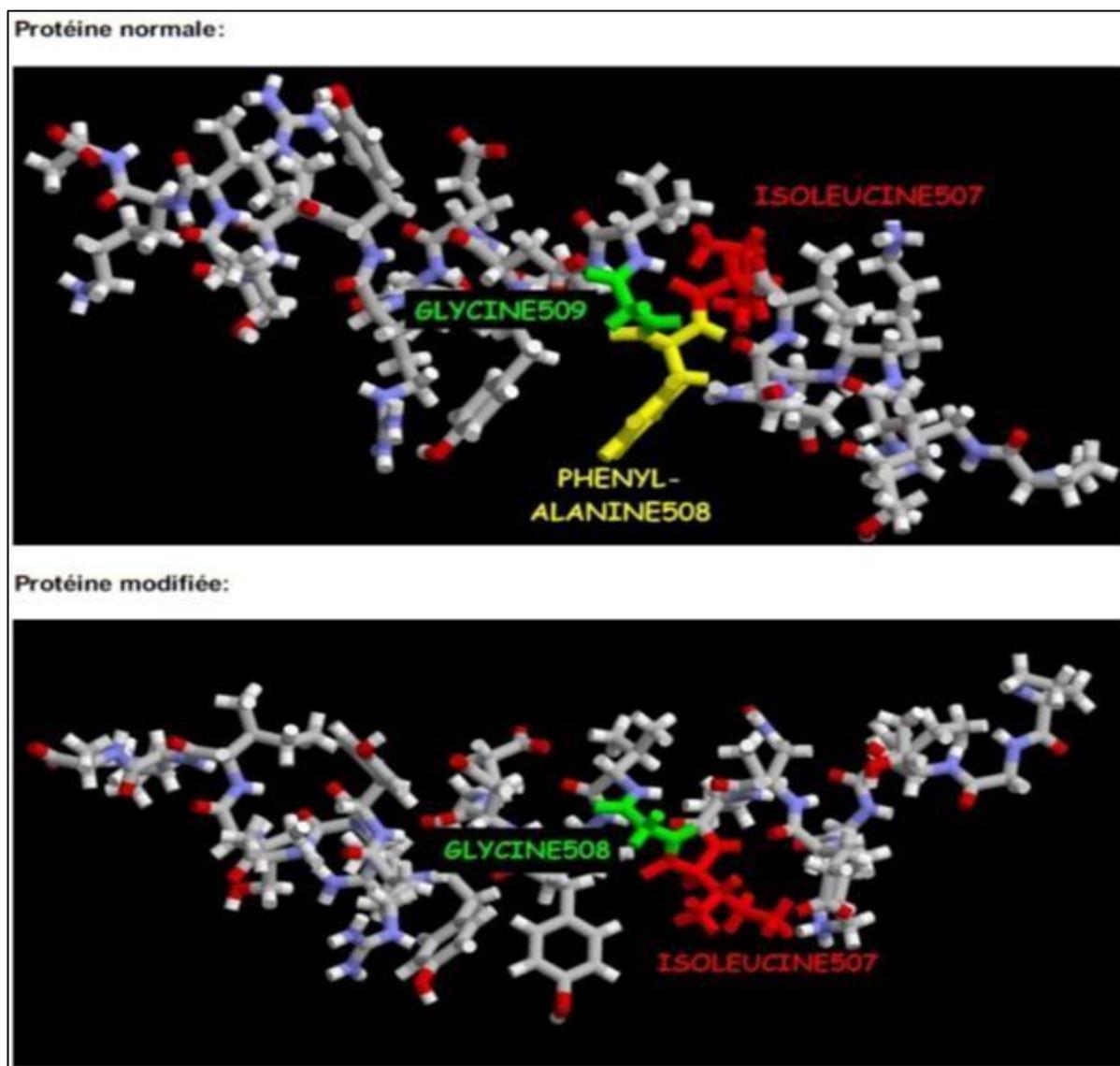
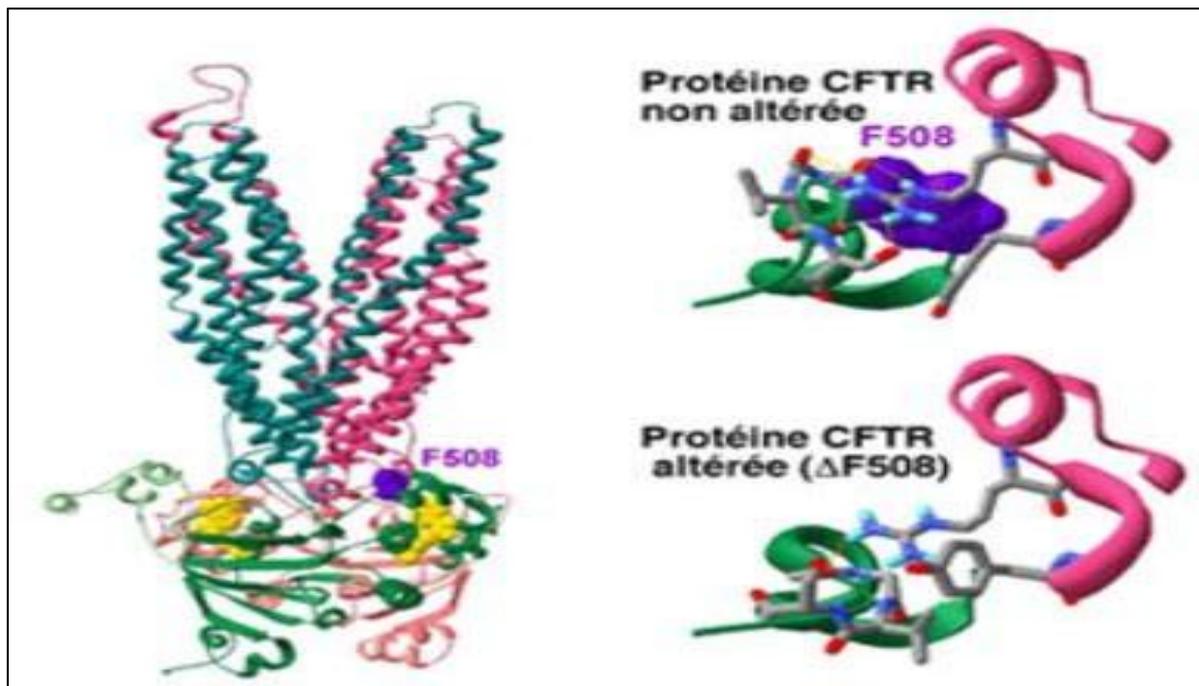


Figure 39: Délétion de la Phénylalanine à la position 508. (Anonyme6)

En haut, à droite, représentation agrandie de l'environnement de la **Phénylalanine F508** (en violet), acide aminé dont l'absence ( $\Delta F508$ ) (en bas à droite) est responsable de plus de 70% des cas de mucoviscidose. (**J. P. Mornon, et al, 2008**)



**Figure 40:** Représentation tridimensionnelle de la protéine CFTR sauvage et mutée (F508del). (**J. P. Mornon, et al, 2008**)

## II.6. Le phénotype de cette mutation

Cette mutation caractérisée par une dégradation totale de la protéine nouvellement synthétisée c'est-à-dire la perte total de la fonction de la protéine CFTR, avec un phénotype sévère de la maladie (insuffisance pancréatique, une fonction pulmonaire réduite, signes précoces des symptômes respiratoire, concentration élevée de teste de la sueur. (**Rosa et al ,2018**)

## II. Analyse de la mutation c.680T>G(L227R) présente au niveau de l'exon 6 du gène CFTR

### II.1. Résultat de l'alignement nucléotidique

La mutation c.680T>G au niveau de la base 680(2ème base du codant 34) est caractérisée par le changement d'une **Thymine** en **Guanine**. (Voir figure 41)

Sequence ID: Query_118819 Length: 164 Number of Matches: 1					
Range 1: 1 to 164 <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>					
Score	Expect	Identities	Gaps	Strand	
292 bits(323)	4e-84	163/164(99%)	0/164(0%)	Plus/Plus	
Query 1		GGACTTGCATTGGCACATTTTCGTGTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGG			60
Sbjct 1		GGACTTGCATTGGCACATTTTCGTGTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGG			60
Query 61		CTAATCTGGGAGTTGTTACAGGCGTCTGCCTTCTGTGGACTTGGTTTCCTGATAGTCCTT			120
Sbjct 61		CTAATCTGGGAGTTGTTACAGGCGTCTGCCTTCTGTGGACTTGGTTTCCTGATAGTCCTT			120
Query 121		GCCCTTTTTTCAGGCTGGGCTAGGGAGAATGATGATGAAGTACAG	164		
Sbjct 121		GCCCTTTTTTCAGGCTGGGCTAGGGAGAATGATGATGAAGTACAG	164		

**Figure 41:** Substitution d'une **Thymine** à la position 680 de l'exon 6. (Personnelle, 2019)

### II.2. Les résultats dans ORF Finder de la mutation c.680T>G(L227R)

#### II.2.a. Séquence sauvage

Après l'insertion de la séquence de l'exon 6 dans l'ORF Finder, les résultats montrent la présence de la **Thymine** dans la 2ème base du codant 34 dans la séquence d'ADN. Après la traduction de cette séquence on remarque la présence d'un acide aminé **Leucine** à la position 227 dans la séquence protéique non mutée non mutée. (Voir figure 42)

ORF Finder results	
<b>Results for 164 residue sequence "Untitled" starting "GGACTTGCAT"</b>	
>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 162.	
GGACTTGCATTGGCACATTTTCGTGTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGG	
CTAATCTGGGAGTTGTTACAGGCGTCTGCCTTCTGTGGACTTGGTTTCCTGATAGTCCTT	
GCCCTTTTTTCAGGCTGGGCTAGGGAGAATGATGATGAAGTAC	
>Translation of ORF number 1 in reading frame 1 on the direct strand.	
GLALAHFVWIAPIQLVALIMGLIWELLQASAFCGIGFLIVLALFQAGLGRMMMKY	

**Figure 42:** Traduction de la séquence non mutée de l'exon 6. (Personnelle, 2019)

### II.2.b. Mutation c.680T>G(L227R)

Les résultats dans ORF Finder du variant de l'exon 6 montrent un changement au niveau de la séquence nucléotidique d'une **Thymine** par une **Guanine** à la position 34, ce qui se traduit par un changement d'une **Leucine** par une **Arginine** au niveau de la séquence protéique. (Voir figure 43)

```

ORF Finder results
Results for 164 residue sequence "Untitled" starting "GGACTTGCAT"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 162.
GGACTTGCATTGGCACATTTTCGTGGATCGCTCCTTTGCAAGTGGCACTCCTCATGGGG
CTAATCTGGGAGTTGTTACAGGCGTCTGCCTTCTGTGGACGTGGTTTTCTGATAGTCCTT
GCCCTTTTTCAGGCTGGGCTAGGGAGAATGATGATGAAGTAC

>Translation of ORF number 1 in reading frame 1 on the direct strand.
GLALAHFVWNIAPLQVALLMGLIWELLQASAFCGRGLIVLALFQAGLGRMMMKY

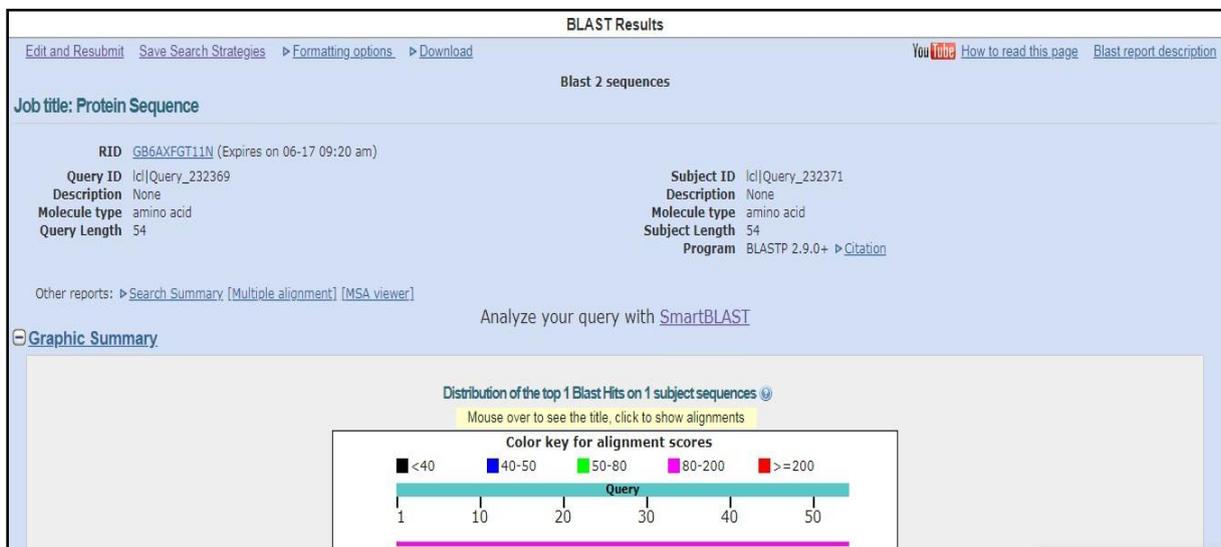
```

**Figure 43:** Traduction de la séquence mutée de l'exon 6. (Personnelle, 2019)

## II.3. Résultat de l'alignement protéique

### II.4. La séquence protéique soumise et la séquence de la base de données :

- Sont des molécules de type acide aminé.
- Leur taille est 54 acides aminés.
- Le programme utilisé est le blast P.
- Le score d'alignement de cette séquence entre 80-200.



L'homologie de cette séquence avec la séquence model est de 98%, d'une similarité de 53/54. Cette séquence a un score égal à 101. L'e-valoré est de  $2e-36$ . La mutation 227 caractérisée par le changement de la **Leucine** en **Arginine** à la position 227 de l'exon 6. (Voir figure 45)

unnamed protein product						
Sequence ID: Query_232371 Length: 54 Number of Matches: 1						
Range 1: 1 to 54 <a href="#">Graphics</a>			▼ Next Match ▲ Previous Match			
Score	Expect	Method	Identities	Positives	Gaps	
101 bits(252)	2e-36	Compositional matrix adjust.	53/54(98%)	53/54(98%)	0/54(0%)	
Query 1	GLALAHFVWIAPLQVALLMGLIWELLQASAFQGLGFLIVLALFQAGLGRMMMKY	54				
	GLALAHFVWIAPLQVALLMGLIWELLQASAFQGLFLIVLALFQAGLGRMMMKY					
Sbjct 1	GLALAHFVWIAPLQVALLMGLIWELLQASAFQGRGFLIVLALFQAGLGRMMMKY	54				

Figure 44: Alignement de la séquence protéique de l'exon 6. (Personnelle, 2019)

## II.5. Structure tridimensionnelle de la protéine (L227R)

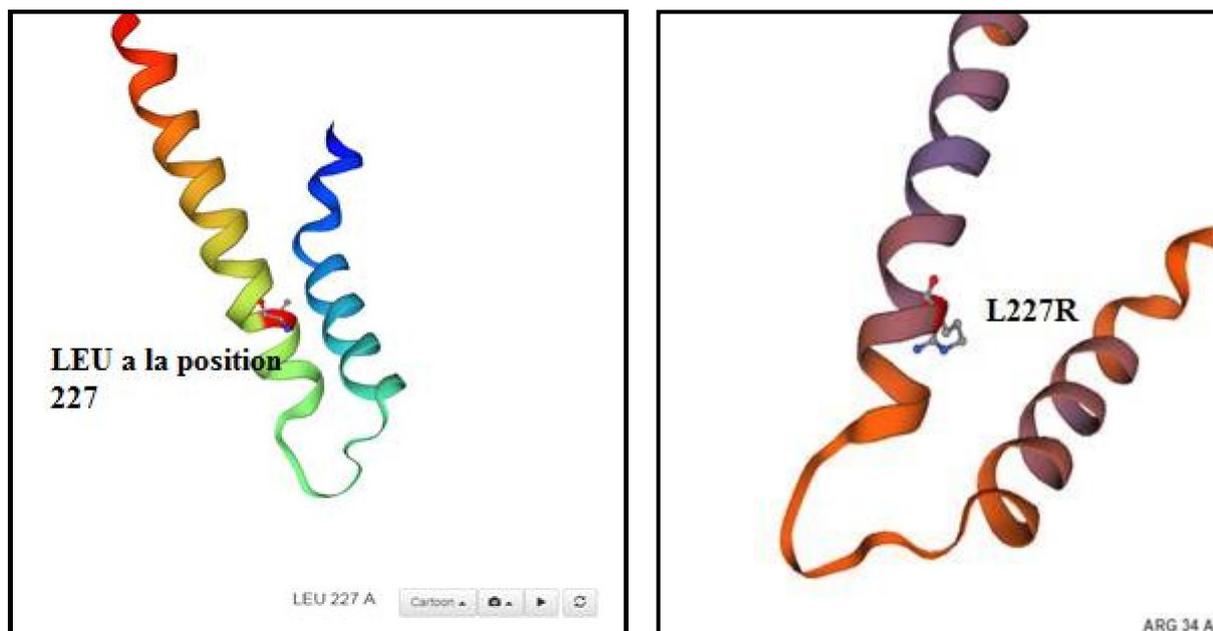
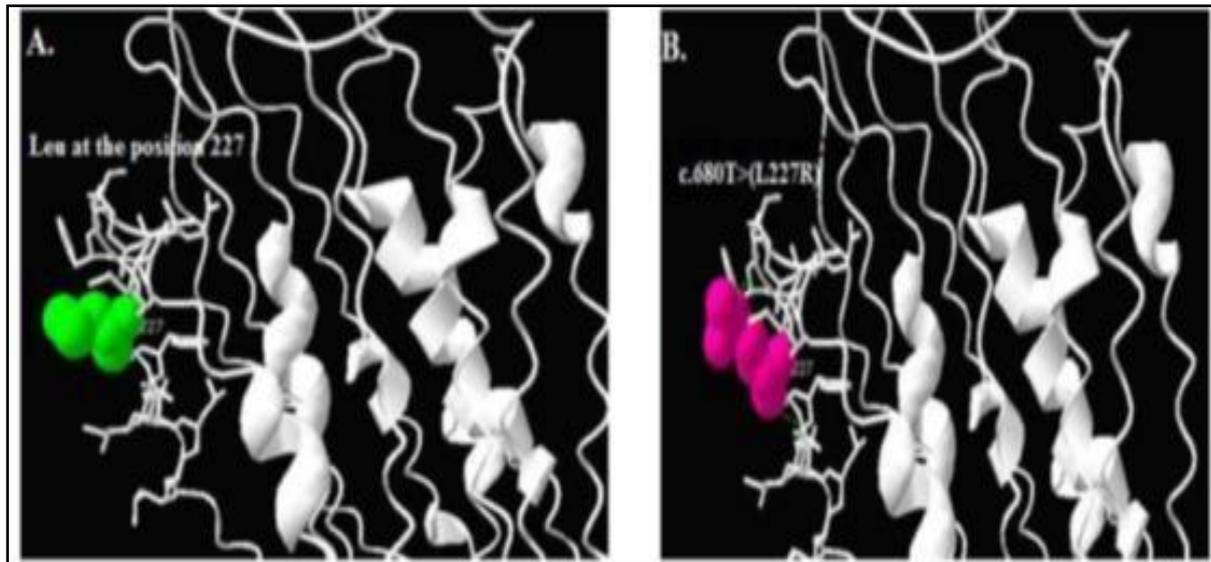


Figure 45: Structure 3D d'exon 6 et la substitution de la **Leucine** en **Arginine** à la position 227. (Personnelle, 2019)



**Figure 46:** Substitution d'une **Leucine** à la position 227. (Seddiki, 2015)

Le caractère pathogène de la mutation c.680T>G (L227R) a été évalué par différents logiciels. SIFT prédit le changement de la **Leucine** en position 207 en **Arginine** comme intolérant, ce qui pourrait par conséquent avoir un impact sur la protéine en induisant une perte totale voir partielle de sa fonction. (seddiki.2015)

Les logiciels PolyPhen-2 et Align-GVGD prédisent respectivement que ce changement est probablement endommageant et interférant avec la fonction de la protéine CFTR. L'analyse par le logiciel I-Mutant2.0 montre que la substitution de la **Leucine** en **Arginine** à la position 227 de la protéine CFTR affecte sensiblement la stabilité de cette dernière. Enfin, Mutation Taster confirme le caractère pathogène de cette mutation. (Seddiki.2015)

### III. Analyse de la mutation c.1624G>T (G542X) présente au niveau de l'exon 12

ID sur NCBI : rs113993959

#### III.1. Résultats de l'alignement nucléotidique

La mutation c.1624G>T au niveau de la base 1624 dans l'exon 12 (1er base du codant 14) est caractériser par le changement d'une **Guanine** par une **Thymine**.

Sequence ID: Query\_131909 Length: 95 Number of Matches: 2

Range 1: 1 to 95 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
168 bits(185)	4e-47	94/95(99%)	0/95(0%)	Plus/Plus
Query 1	GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG	60		
Sbjct 1	GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG	60		
Query 61	AGTGGAGGTCAACGAGCAAGAATTTCTTTAGCAAG	95	*	
Sbjct 61	AGTGGAGGTCAACGAGCAAGAATTTCTTTAGCAAG	95		

**Figure 47 :** Substitution d'une **Guanine** à la position 1624 de l'exon 12. (Personnelle, 2019)

## III.2. Résultats des ORF Finder

### III.2.a. Sauvage

**ID sur NCBI :** [AIS25620.1](#)

Les résultats dans ORF Finder montrent la présence de la **Guanine** dans la 1ère base du codant 14 de l'exon 12 dans la séquence d'ADN. Après la traduction de cette séquence on remarque la présence d'une **Glycine** à la position 542 dans cette séquence non mutée.

```

ORF Finder results
Results for 95 residue sequence "Untitled" starting "GACATCTCCA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 93.
GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG
AGTGGAGGTCAACGAGCAAGAATTTCTTTAGCA
                                     *

>Translation of ORF number 1 in reading frame 1 on the direct strand.
DISKFAEKDNIVLGEggITLsggQRARISLA
  
```

**Figure 48:** Traduction de la séquence non mutée de l'exon 12. (Personnelle, 2019)

### III.2.b. Mutation c.1624G>T (G542X)

Les résultats dans ORF Finder du variant de l'exon 12 montrent un changement au niveau de la séquence nucléotidique de la **Guanine** en **Thymine** à la position 1624, ce qui se traduit par un changement d'une **Glycine** en **codant stop** au niveau de la séquence protéique.

```

ORF Finder results
Results for 95 residue sequence "Untitled" starting "GACATCTCCA"

No ORFs were found in reading frame 1.

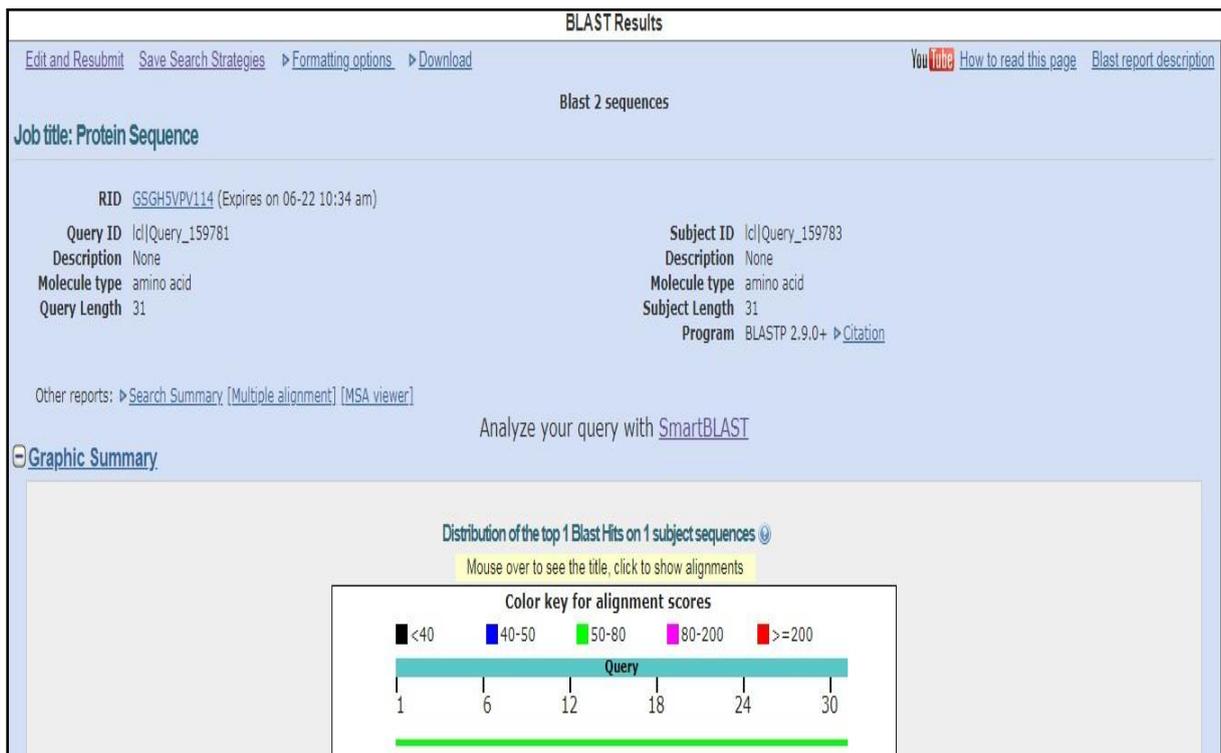
```

**Figure 49:** Traduction de la séquence mutée de l'exon 12. (Personnelle, 2019)

### III.3. Résultats d'alignement protéiques

La séquence protéique soumise et la séquence de la base de données :

- Sont des molécules de type acide aminé.
- Leur taille est 31 acides aminés.
- Le programme utilisé est le blast P.
- Le score d'alignement de cette séquence entre 50-80.

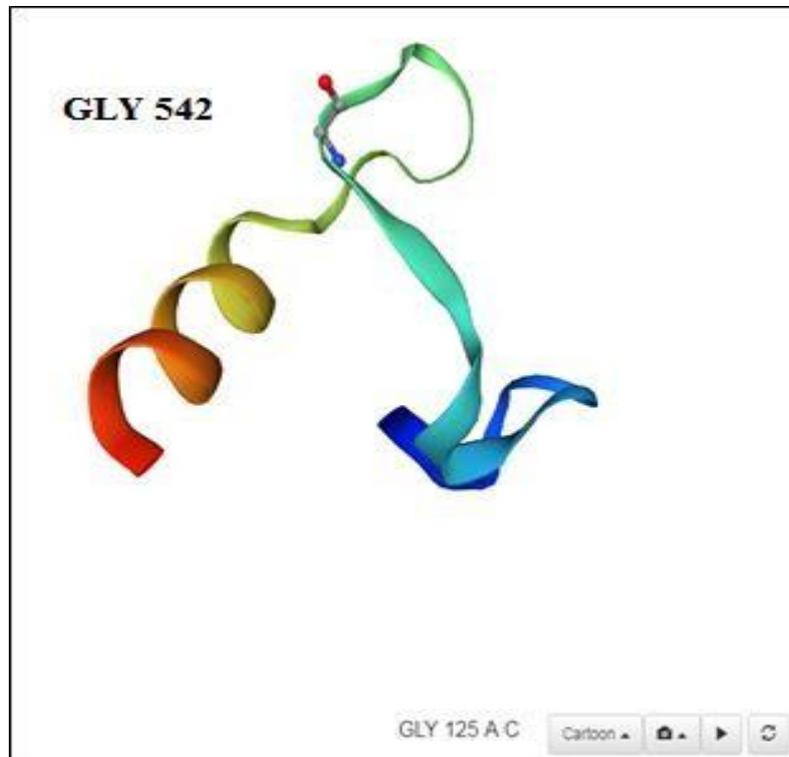


L'homologie de cette séquence avec la séquence de la base de données est de 97%, d'une similarité de 30/31. Cette séquence a un score égal à 58.8. L'e-valu égale  $4^{-20}$ . La mutation G542X caractérisée par le changement de d'une **Glycine** en **codant stop** à la position 542. (Voir figure 50)

Score	Expect	Method	Identities	Positives	Gaps
58.5 bits(140)	4e-20	Compositional matrix adjust.	30/31(97%)	30/31(96%)	0/31(0%)
Query 1	DISKFAEKDNI	VLGE	EGGITLSGGQRARISLA	31	
	DISKFAEKDNI	VL	EGGITLSGGQRARISLA		
Sbjct 1	DISKFAEKDNI	VL*	EGGITLSGGQRARISLA	31	

**Figure 50:** Alignement local de la séquence protéique de l'exon 12(G542X). (Personnelle, 2019)

### III.4. La structure 3D de la protéine G542X



**Figure 51:** Structure 3D de l'exon 12 et le changement de **Glycine** par un **codant stop**.

(Personnelle, 2019)

### III.5. Le phénotype associé à la mutation G542X

Cette mutation caractérisée par une perte totale ou partielle de la fonction protéique, avec un phénotype sévère de la maladie (forte incidence de meconium ileus, insuffisance pancréatique). (Rosa et al ,2018).

## IV. Analyse de la mutation c.3909C>G (N1303K) présente au niveau de l'exon 24

ID sur NCBI : rs 80034486

### IV.1. Résultats de l'alignement nucléotidique

La mutation c.3909C>G au niveau de la base 3909 (3ème base du codant 12) est caractérisée par le changement d'une **Cytosine** en une **Guanine**.

Sequence ID: Query_44527 Length: 90 Number of Matches: 1				
Range 1: 1 to 90 <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>				
Score	Expect	Identities	Gaps	Strand
159 bits(175)	2e-44	89/90(99%)	0/90(0%)	Plus/Plus
Query 1	AAAGTATTTATTTTTCTGGAACATTTAGAAAAA	CTGGATCCCTATGAACAGTGGAGT	60	
Sbjct 1	AAAGTATTTATTTTTCTGGAACATTTAGAAAAA	AGTGGATCCCTATGAACAGTGGAGT	60	
Query 61	GATCAAGAAATATGGAAAGTGCAGATGAG	90		
Sbjct 61	GATCAAGAAATATGGAAAGTGCAGATGAG	90		

**Figure 52:** Substitution d'une Cytosine à la position 3909 sur l'exon 24.

(Personnelle, 2019)

### IV.2. Résultats dans ORF Finder de la mutation c.3909C>G (N1303K)

#### IV.2.a. Sauvage

Les résultats dans ORF Finder montrent la présence d'une **Cytosine** à la position 3903(3ème base du codant 12) de l'exon 24 dans la séquence d'ADN. Après la traduction de cette séquence on remarque la présence d'une **Asparagine** à la position 3909 dans la séquence non mutée.

```

ORF Finder results
Results for 90 residue sequence "Untitled" starting "AAAGTATTTA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 90.
AAAGTATTTATTTTTCTGGAACATTTAGAAAAA*CTTGGATCCCTATGAACAGTGGAGT
GATCAAGAAATATGGAAAGTTGCAGATGAG

>Translation of ORF number 1 in reading frame 1 on the direct strand.
KVFIFSGTFRKNLDPYEQWSDQEINKVADE

```

**Figure 53:** Traduction de la séquence non mutée de l'exon 24. (Personnelle, 2019)

#### IV.2.b. Mutation c.3909C>G (N1303K)

Les résultats dans ORF Finder du variant de l'exon 24 montrent un changement au niveau de la séquence nucléotidique d'une **Cytosine** en **Guanine** à la position 3909, ce qui se traduit par un changement d'une **Asparagine** en **Lysine** à la position 1303 au niveau de la séquence protéique.

```

ORF Finder results
Results for 90 residue sequence "Untitled" starting "AAAGTATTTA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 90.
AAAGTATTTATTTTTCTGGAACATTTAGAAAAAAGTTGGATCCCTATGAACAGTGGAGT
GATCAAGAAATATGGAAAGTTGCAGATGAG *

>Translation of ORF number 1 in reading frame 1 on the direct strand.
KVFIFSGTFRKKLDPYEQWSDQEINKVADE

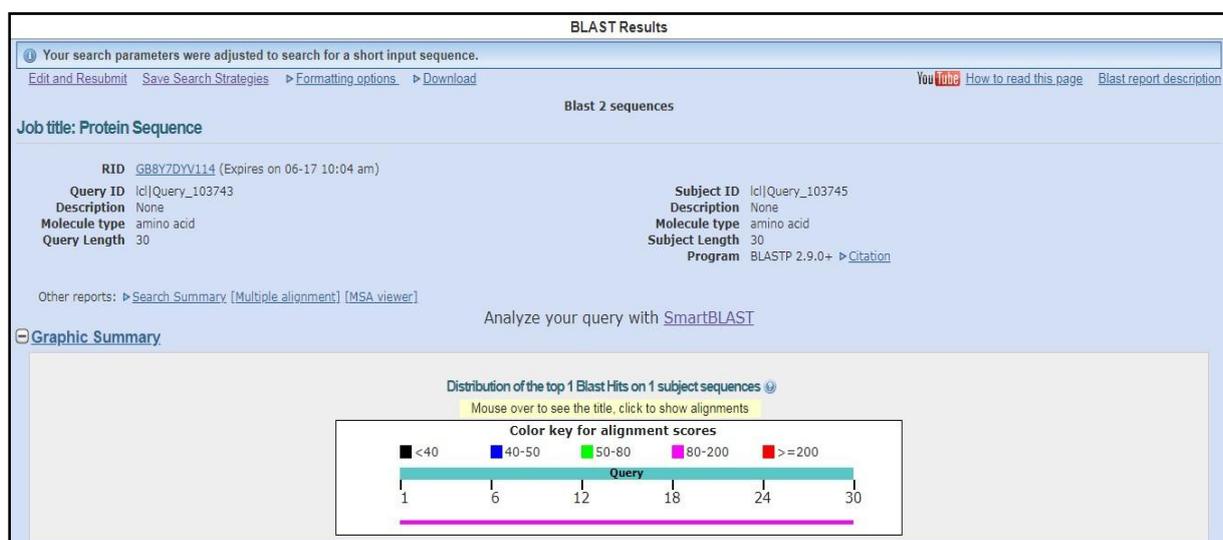
```

**Figure 54:** Traduction de la séquence mutée de l'exon 24. (Personnelle, 2019)

### IV.3. Résultats de l'alignement protéique

La séquence protéique soumise est la séquence de la base de données :

- Sont des molécules de type acide aminé.
- Leur taille est 30 acides aminés.
- Le programme utilisé est le blast P.
- Le score d'alignement entre 80-200.

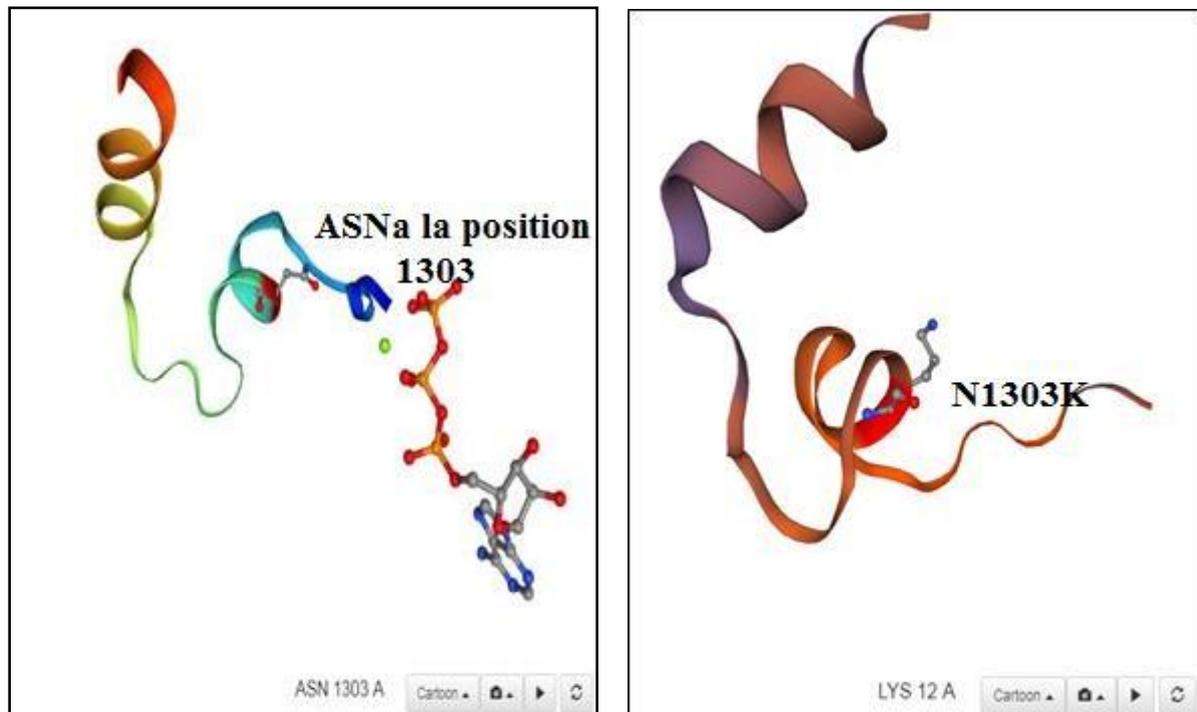


L'homologie de cette séquence avec la séquence model est de 970(%, d'une similarité de 29/30. Cette séquence a un score égal à 101. l'e-valu égale 6e-35. La mutation **N1303K** caractérisée par le changement de d'une **Asparagine** en **Lysine** à la position 1303. (Voir **figure 56**)

unnamed protein product				
Sequence ID: Query_186633 Length: 30 Number of Matches: 1				
Range 1: 1 to 30 <a href="#">Graphics</a> <span style="float: right;">▼ Next Match ▲ Previous Match</span>				
Score	Expect	Identities	Positives	Gaps
101 bits(231)	6e-35	29/30(97%)	29/30(96%)	0/30(0%)
Query 1	KVFIFSGTFRK	NIDPYEQWSDQEIWKVADE	30	
Sbjct 1	KVFIFSGTFRK	NIDPYEQWSDQEIWKVADE	30	

**Figure 55:** Alignement protéique de l'exon 24 (N1303K). (Personnelle, 2019)

#### IV.4 Structure 3D de la protéine N1303K



**Figure 56:** Structure 3D de l'exon 24 et le changement d'une **Asparagine** en **Lysine**. (personnle,2019)

Selon des études précédentes, l'analyse de la mutation c.3909C>G (N1303K) par différents outils bioinformatique montre que la substitution de L'**Asparagine** en **Lysine** en position 1303 est intolérante (Intolerated). Ce résultat signifie que cette dernière a un impact entraînant une perte partielle ou totale de la fonction protéique. Avec un phénotype sévère de la maladie. (Seddiki ,2015)

## V. Analyse de la mutation c.1652G>A (G551D) présente au niveau de l'exon 12

ID sur NCBI : rs75527207

### V.1. Résultats de l'alignement nucléotidiques

La mutation c.1652G>A au niveau de la base 1652(2ème base du codant 23) de l'exon 12 est caractérisée par le changement d'une **Guanine** en **Adénine**.

Sequence ID: Query_53219 Length: 95 Number of Matches: 4				
Range 1: 1 to 95 <a href="#">Graphics</a>		▼ Next Match ▲ Previous Match		
Score	Expect	Identities	Gaps	Strand
168 bits(185)	4e-47	94/95(99%)	0/95(0%)	Plus/Plus
Query 1	GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG	60		
Sbjct 1	GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG	60		
Query 61	AGTGGAGGTCAACGAGCAAGAATTTCTTTAGCAAG	95		
Sbjct 61	AGTGGAGATCAACGAGCAAGAATTTCTTTAGCAAG	95		
	*			

**Figure 57:** Substitution d'une **Guanine** à la position 1625 de l'exon 12. (Personnelle, 2019)

### V.2. Les résultats dans ORF Finder de la mutation c.1652G>A (G551D)

#### V.2.a. Séquence sauvage

Les résultats dans ORF Finder montrent la présence d'une **Guanine** dans la 2ème base du codant 23 de l'exon 12 dans la séquence d'ADN. Après la traduction de cette séquence on remarque la présence d'une **Glycine** en position 551 dans la séquence protéique non mutée.

```

ORF Finder results
Results for 95 residue sequence "Untitled" starting "GACATCTCCA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 93.
GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG
AGTGGAGGTCAACGAGCAAGAATTTCTTTAGCA      *

>Translation of ORF number 1 in reading frame 1 on the direct strand.
DISKFAEKDNIVLGE G G I T L S G G Q R A R I S L A
                    *

```

**Figure 58:** Traduction de la séquence non mutée de l'exon 12. (Personnelle, 2019)

### V.2.b. Mutation c.1652G>A (G551D)

Les résultats de l'ORF Finder du variant de l'exon 12 montrent un changement au niveau de la séquence nucléotidique d'une **Guanine** en **Adénine** à la position 1652, ce qui se traduit par un changement d'une **Glycine** en **Acide Aspartique** à la position 551 au niveau de la séquence protéique.

```

ORF Finder results
Results for 95 residue sequence "Untitled" starting "GACATCTCCA"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 93.
GACATCTCCAAGTTTGCAGAGAAAGACAATATAGTTCTTGGAGAAGGTGGAATCACACTG
AGTGGAGATCAACGAGCAAGAATTTCTTTAGCA      *

>Translation of ORF number 1 in reading frame 1 on the direct strand.
DISKFAEKDNIVLGE G G I T L S G D Q R A R I S L A
                    *

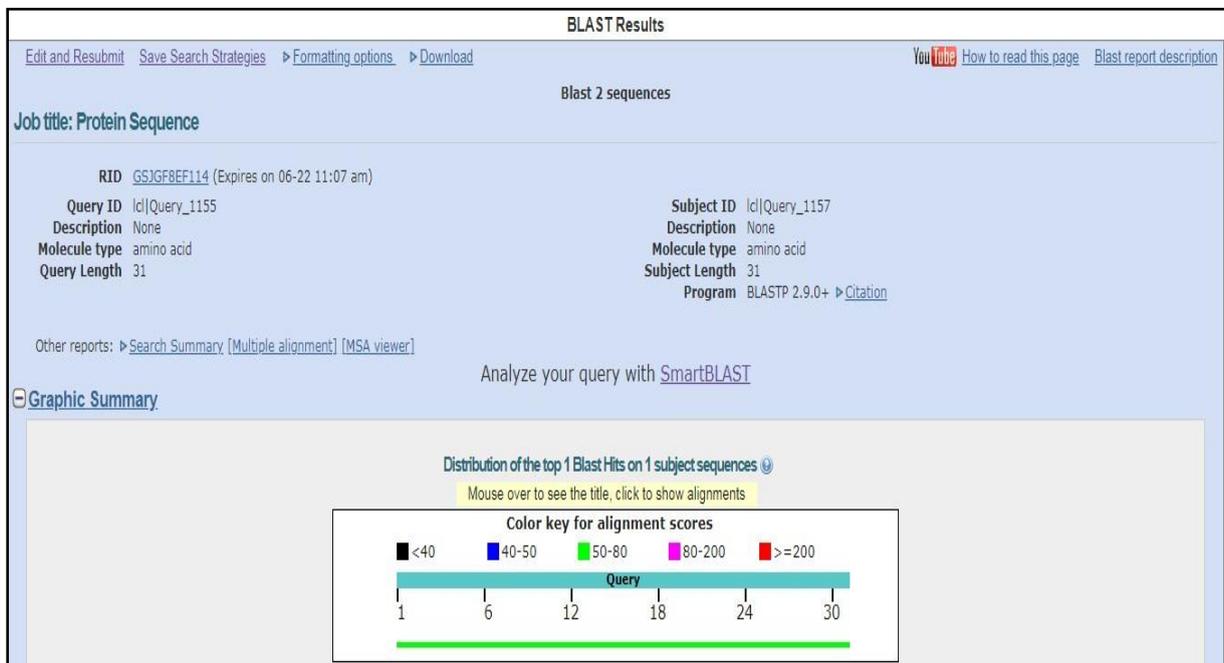
```

**Figure 59:** Traduction de la séquence mutée de l'exon 12. (Personnelle, 2019)

### V.3. Résultats de l'alignement protéique

La séquence protéique soumise et la séquence de la base de données :

- Sont des molécules de type acide aminé.
- Leur taille est 31 acides aminés.
- Le programme utilisé est le blast P.
- Le score d'alignement de cette séquence entre 50-80.



L'homologie de cette séquence avec la séquence modèle est de 97%, d'une similarité de 30/31. Cette séquence a un score égal à 59,7, l'e-value égale  $1 \times 10^{-20}$ . La mutation **G551D** caractérisée par le changement de d'une **Glycine** en **Acide Aspartique** à la position 551. (Voir figure 61)

unnamed protein product  
Sequence ID: Query\_1157 Length: 31 Number of Matches: 1

Range 1: 1 to 31 [Graphics](#) ▼ Next Match ▲ Previous Match

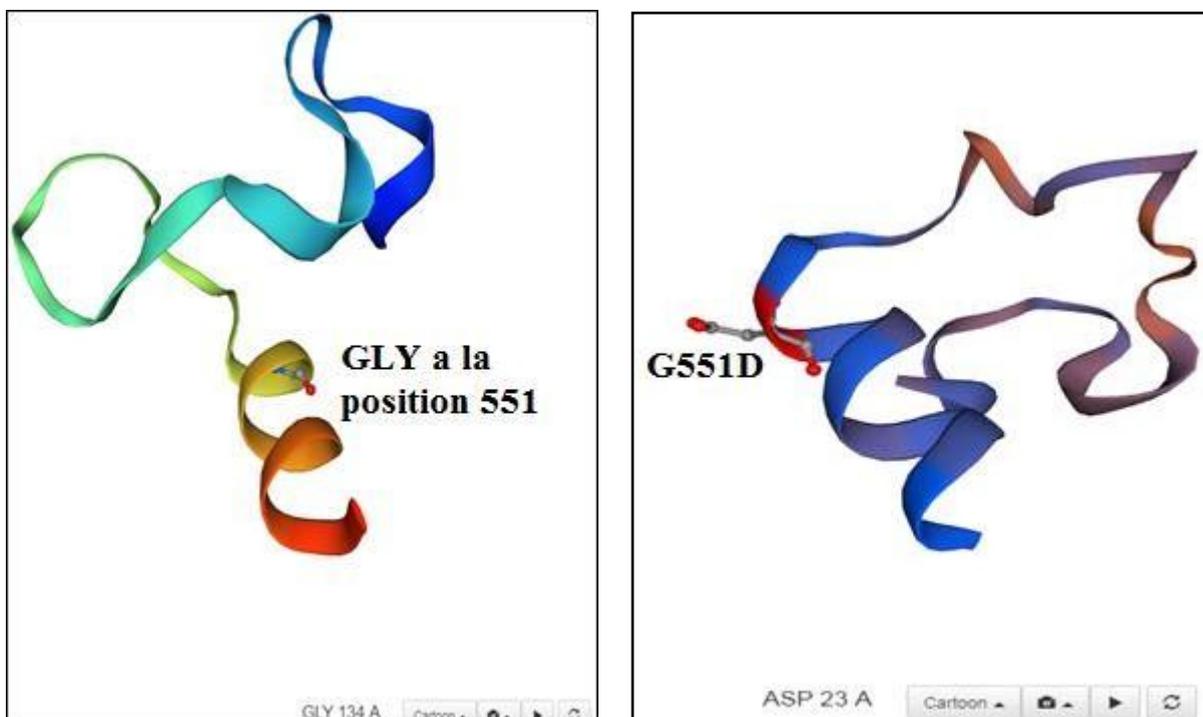
Score	Expect	Method	Identities	Positives	Gaps
59.7 bits(143)	1e-20	Compositional matrix adjust.	30/31(97%)	30/31(96%)	0/31(0%)

```

Query 1 DISKFAEKDNIVLGEGGITLSGGCRARISLA 31
      DISKFAEKDNIVLGEGGITLSGCRARISLA
Sbjct 1 DISKFAEKDNIVLGEGGITLSGGCRARISLA 31
  
```

**Figure 60 :** Alignement local de la séquence protéique G551D. (Personnelle, 2019)

#### V.4. La structure 3D de la protéine G551D



**Figure 61:** Structure 3D de l'exon 12 et le changement d'acide aminé **Glycine** en **Acide Aspartique**. (Personnelle, 2019)

### V.5. Le phénotype associé à la mutation G551D

La substitution de la **Glycine** en **Acide Aspartique** montre une perturbation de la régulation du canal CFTR (**Ouchtati A, Dali H ,2018**), qui caractérisée par un phénotype sévère de la maladie (maladie pulmonaire sévère, insuffisance pancréatique, infection par *Pseudomonas Aeruginosa*. (**Rosa et al, 2018**))

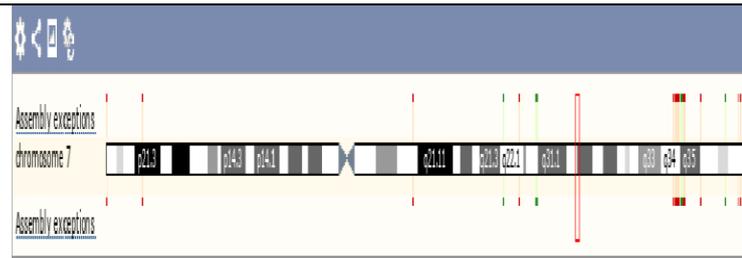
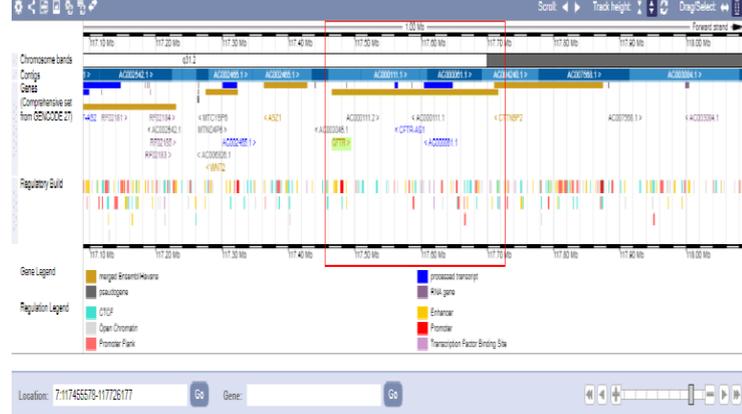
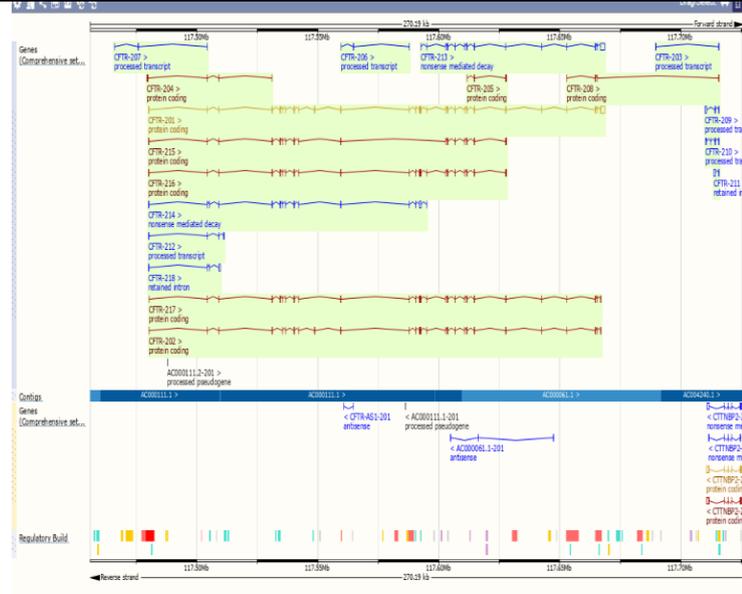
Le tableau VIII résume les résultats d'une étude sur les mutations du gène CFTR dans la population algérienne, seulement Cinq mutations ont été identifiées dans la population algérienne (**Seddiki ,2015**)

mutations	E/II	Remplacement d'un nucléotide	Substitution d'acides aminés	La fréquence (%)
<b>c.1521_1523delCTT (F508del)</b>	Exon11	del CTT	del phe507/508	21.42
<b>c.680&gt;TG(L227R)</b>	Exon6	T→G812	Leu→Arg 227	9.52
<b>c.1624G&gt;T (G542X)</b>	Exon12	G→T1756	Gly→ Stop 542	9.52
<b>c.3909C&gt;G (N1303K)</b>	Exon24	C→G4041	Asn→Lys1303	4.76
<b>c.1652G&gt;A (G551D)</b>	Exon12	G→A1652	Gly→Asp542	2.38

### Les résultats d'Ensembl

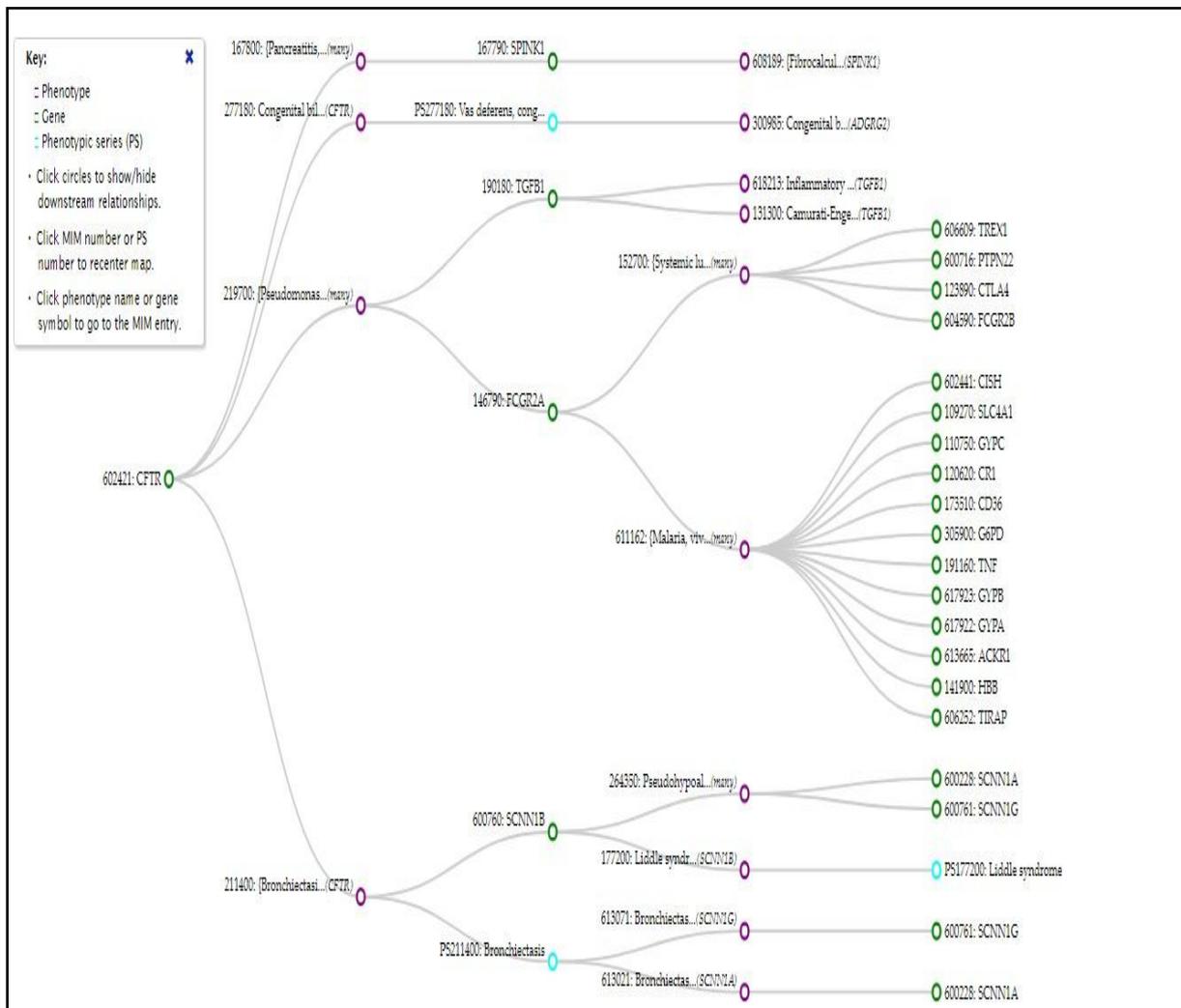
Les résultats de l'Ensembl se présentant comme dans le tableau IX :

**Tableau IX:** Résultats dans Ensembl

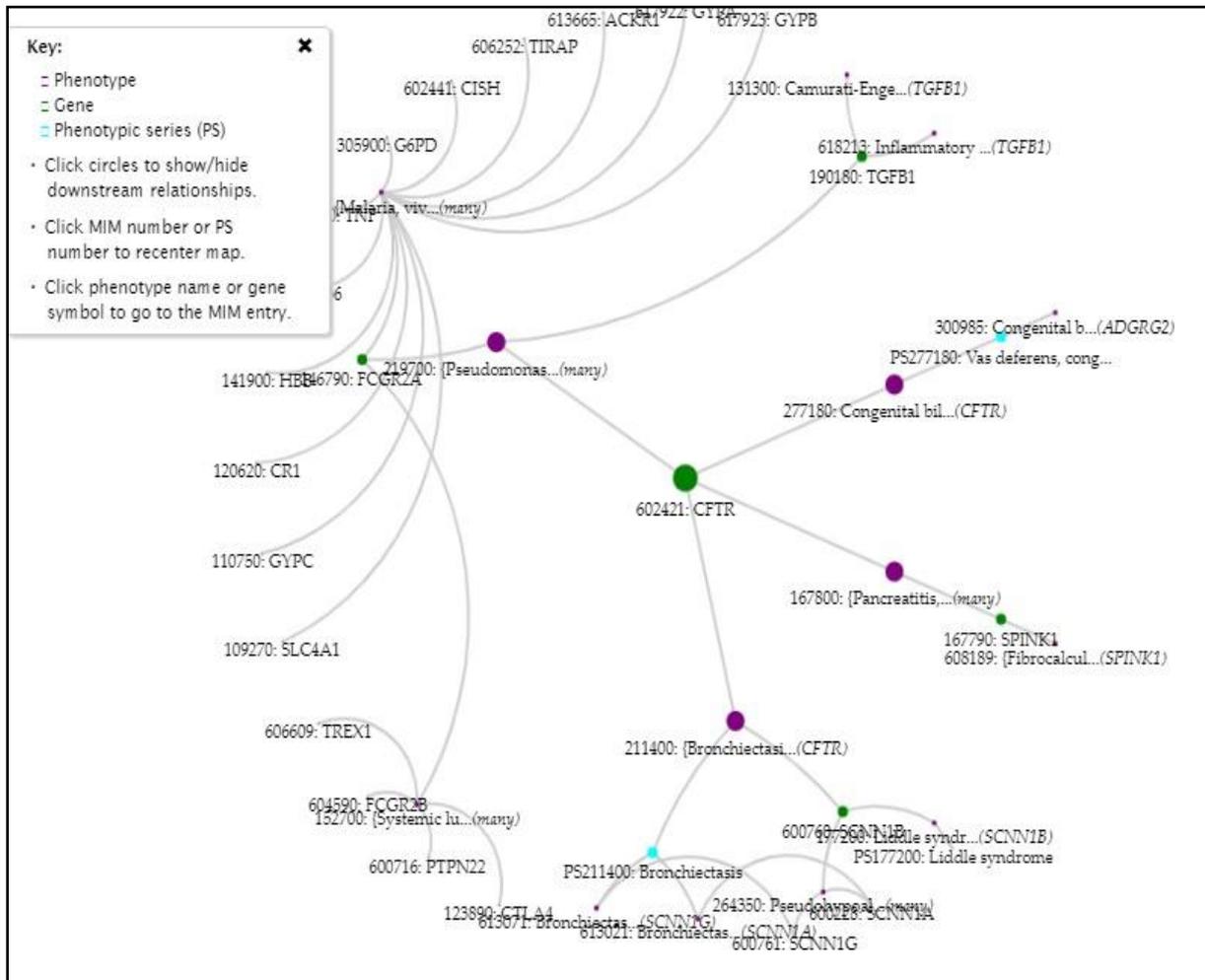
Résultat	Rôle	Descriptifs
	Image chromosomique	<ul style="list-style-type: none"> <li>Les régions de chromosome.</li> <li>Le numéro de chromosome.</li> <li>La région d'intérêt (rectangle en rouge).</li> </ul>
	Image vue d'ensemble	<ul style="list-style-type: none"> <li>Les bandes de chromosome.</li> <li>La séquence génomique.</li> <li>La région d'intérêt.</li> <li>La position de gène.</li> </ul>
	Représentation de région	<ul style="list-style-type: none"> <li>La transcription de la région d'intérêt :                             <ul style="list-style-type: none"> <li>-les traits correspondent aux exons.</li> <li>-les lignes correspondent aux introns.</li> </ul> </li> <li>La séquence de la région d'intérêt.</li> </ul>

### Les résultats dans OMIM

Représentation graphique des relations phénotype / gène associées à cette entrée. Les séries phénotypiques (lorsqu'elles sont disponibles) sont affichées avec les gènes pertinents et les phénotypes suivants à une profondeur de 4 nœuds.



**Figure 62:** Représentation graphique linéaire de la relation phénotype génotype de la mucoviscidose. (Personnelle, 2019)



**Figure 63:** Représentation graphique circulaire de la relation phénotype génotype de la mucoviscidose. (personnelle,2019)

On remarque de la représentation graphique les différentes relations phénotypo-génotypiques entre la mucoviscidose et l'expression d'autres maladies associées pour les quelles sont montrés la nature de la mutation en cause ainsi que leurs identifiants dans la base de données.

## Conclusion

Le gène CFTR est situé sur le bras long du chromosome 7 dans la position 117479963..117668665, qui contient 27 exons .

La déficience du canal CFTR, entraîne des perturbations hydro-électrolytiques qui sont à l'origine de l'hyper viscosité des sécrétions muqueuses. La gravité de la maladie dépend de la ou des mutations du gène CFTR.

À ce jour ,ont été identifié près de 2000 mutations de gène CFTR, ces mutations sont regroupées en six classes en fonction de leurs effets sur la fonction de la protéine, soit en réduisant la quantité de la protéine ou en réduisant la fonction de la protéine .

La fréquence de la maladie, varie selon l'origine géographique des patients. Loin d'être une pathologie européenne, la mucoviscidose est présente en Algérie. Seulement cinq mutations ont été identifiées et qui sont les plus fréquentes à savoir :c.1521-1523delCTT(F508del), c.680T>G(L227R),c.1624G>T(G542X);C.3909C>G(N1303K) et c.1652C>A(G551D).

Cette étude est basée sur la compréhension des mécanismes moléculaires à l'origine du défaut du canal CFTR. Les effets délétères de cinq mutations spécifiques à la population Algérienne c.1521\_1523delCTT (F508del), c.680T>G (L227R), c.1624G>T (G542X),c.3909A >C (N1303K) ,c.1652G>A (G551D),ont été étudiés en utilisant d'une combinaison de logiciels d'étude *in silico* et modélisation 3D. La mutation c.680T>G (L227R) a été prédite comme altérant les relations hydrophobes au cœur même de la protéine ou avec les lipides de la membrane.

D'autre part, la c.3909A>C (N1303K) a été prédite comme délétère au niveau structure et fonction du canal CFTR, les changements physicochimiques entre les acides aminés natifs et mutés perturbent fortement la stabilité de la protéine.

## **Référence bibliographique**

Alix de BECDELIEVRE, Contribution à l'amélioration des connaissances sur la relation Génotype-phénotype dans la mucoviscidose et caractérisation phénotypique de l'inflammation pulmonaire, Thèse de doctorat de Génétique Humaine, Université Paris-Est - Val de Marne, soutenue 2011.

A Ouchtati, H Dali, le diagnostic de la mucoviscidose à Constantine, mémoire de master, spécialité génétique, Université des Frères Mentouri Constantine, Faculté des Sciences de la Nature et de la Vie, Département de Biologie Animale, soutenue le 28-06-2018.

Aurélié Hatton, Anne Nguyen-Khoa, Agnès Mogenet, Frédérique Chedevert, Aleksander Edelman, Isabelle Sermet-Gaudelus, Mucoviscidose : vers une stratégie thérapeutique personnalisée, *mt pédiatrie*, vol 16, page 333-340, 2013.

A.W.R. Serohijos, T. Hegedus, A.A. Aleksandrov, L. He, L. Cui, N.V. Dokholyan, J.R. Canal, Canal transporteur des ions Chlorure : la protéine CFTR, 23/09/2011

Boudjellal Sara, Fettel hadjer, étude bioinformatique de trois formes alléliques du gène APO intervenant dans l'expression de la maladie d'Alzheimer forme sporadique, mémoire de master, Université de Blida 1 ; département biologie et physiologie cellulaire, spécialité bioinformatique, soutenue 2017.

Carlo Castellani, Baroukh M. Assael, Cystic fibrosis: à clinical view, *Cell. Mol. Life Sci*, page 1-12, 2016.

C. Lacoste, A. Fabre, C. Pécheux, N. Lévy, M. Krahn, P. Malzac, N. Bonello-Palot, C. Badens, P. Bourgeois ; Le séquençage d'ADN à haut débit en pratique clinique ; *ARCPED* ; vol 4395 ; page 1-11 ; 2017.

C. Hamard, X. Mignard, N. Pecuchet, N. Mathiot, H. Blons, P. Laurent-Puig, K. Lero, A. Lupo, J. Chapron, F. Giraud, J. Arrondeau, F. Goldwasser, M. Alifano, D. Damotte, M. Wislez ; IHC, FISH, CISH, NGS dans les cancers bronchiques non à petites cellules : quelles évolutions dans l'ère des biomarqueurs ? ; *Revue de Pneumologie clinique* ; vol 708 ; page 1-12. 2018.

ChuSY, WengCY. Introduction to Genetic/Rare Disease and the Application of Genetic Counseling. *pub med*, vol 64(5), page 11-17, 2017.

Cynthia Gibas et per Jambek, introduction à la bioinformatique, Edition O'Reilly, 2002.

Dennis A. Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell and Eric W. Sayers ; GenBank ; *Nucleic Acids Research*, Vol. 45, p37-42, 2017

D. Gaillarda, C. Clavel, K. Bessaci-Kabouya M. Abély, Les formes atténuées de la mucoviscidose : génétique - suivi prolongé nécessaire, *Archives de pédiatrie* ; vol 16, page 387-390, 2009.

H Lafoeste, L regard, C Martin, G, Chassagnon, P, R, Burgel, Complications aiguës respiratoires et non respiratoires chez l'adulte mucoviscidosique, *pneumologie clinique*, vol 74, page 267-278, 2018.

Harriet Corvol, Kristin E. Thompson, Olivier Tabary, Philippe Le Rouzic, and Guillot, Translating the genetics of cystic fibrosis to personalized medicine. *Translational Research*, page 1-10, 2015

*Méga Guide Stages IFSI*, 423. Maladies génétiques, Pages 1338-1340. 2015

J. P. Moron, P. Lehn, I. Callebaut, Atomic model of human cystic fibrosis transmembrane conductance regulator : Membrane-spanning domains and coupling interfaces, *Cellular and Molecular Life Sciences*, 65 : page 2594-2612 (2008).

Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M. J. Searle, Ridwan Amode, Simon Brent, William Spooner, Eugene Kulesha, Andrew Yates and Paul Flicek ; *Ensembl comparative genomics resources* ; Database update ; vol 096 ; p1-17, 2016.

Jamie L. Sturgill. Sphingolipids and their enigmatic rôle in asthma, advances in biological regulation, vol 70, page 74-81, 2018.

Joanna S. Amberger and Ada Hamosh ; Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes ; *Curr Protoc Bioinformatics* ; vol 58 ; p1- 20. 2018

J Stuart Elborn, Cystic fibrosis. The Lancet, Vol 388, Issue 10059, Pages 2519-2531 ; 19–25 November 2016,

J. Lamoril, N. Ameziane, J.-C. Deybach, P. Bouizegarène, M. Bogard, Notions de génétique moléculaire pour comprendre l'hérédité, *Immuno-analyse et biologie spécialisée* vol 23(6), page 331—352, 2008.

Katiana Murieli da Rosa, Eliandra da Silveira de Lima, Camila Correia Machado, Thiane Rispoli, Victória d'Azevedo Silveira, Renata Ongaratto, Talitha Comaru, Leonardo Araújo Pinto, Genetic and phenotypic traits of children and adolescents with cystic, fibrosis in Southern Brazil, *J Bras Pneumol*. vol 44(6):page 498-504, 2018.

Khayatha, S. Qia, F. de Blay, Bronchopneumopathie chronique obstructive (BPCO) et environnement intérieur, *Revue des Maladies Respiratoires*, vol 1103, page 1-9, 2016.

L. Regard, H. Lafoeste, C. Martina, G. Chassagnon, P.-R. Burgela, Vieillir avec la mucoviscidose : comorbidités classiques et émergentes chez l'adulte mucoviscidosique. *Pneumologie clinique*, vol 707, page 1-13, 2018.

Marie E. Egan, MD, Genetics of Cystic Fibrosis Clinical Implications, *Clin Chest Med*, vol 37(1) page 9-16, 2015.

Marie-Françoise Bendridi, Rachida Boukari, Hicham Mansour. Débat pour tout méditerranéen Quels sont les partenariats existants ? Quels sont les manques ?, *médecine/sciences; (hors série n° 1)* vol 30 page 40-46, 2014.

Martina Gentsch, PhD; and Marcus A. Mall, MD, Ion Channel Modulators in Cystic Fibrosis, *chest*, vol 154(2):page 383-393, 2018.

Seyed Bashir Mirtajani, Poopak Farnia, Maryam Hassanzad, Jaleedin Ghanavi, Parissa Farnia, Ali Akbar Velayati. Geographical distribution of cystic fibrosis; The past 70 years of data analysis. *Biomed Biotechnol Res J*, vol 1(2), page 105-112, 2017.

Francine Bono Neri, MA, RN, PNP, Cara Romano, MS, RN, CRRN, & Anne Isedeh, MSN, RN, FNP-BC, Cystic Fibrosis: Advancing Along the Continuum, *J Pediatr Health Care*, vol 33(3): page 242-254, 2018.

M. Susan Lindee, *Genetic diseases since 1945*, Macmillan Magazines Ltd, vol 1, page 236-241. 2000.

M. SUSAN LINDEE, Genetic Disease in the 1960s: A Structural Revolution, *American Journal of Medical Genetics (Semin. Med. Genet.)*, vol 115, page 75–82, 2002.

Oujidi Bachija, thèse de doctorat, impact de la pathologie cardiovasculaire dans la prise en charge des broncho-pneumopathies chroniques obstructives au service de pneumo-phtisiologie B du CHU d'Oran université des sciences et de la technologie d'Oran, faculté de médecine, département de médecine, soutenue le 24-04-2018.

Panagiotis K ; Stasinakis ; Despoina Nicolaou ; Modeling of DNA and Protein Organization Levels with Cn3D Software ; *Biochemistry and Molecular Biology Education* ; vol 100 ; page 1-4 ; 2016.

Pauline Bardin, Florence Sonnevile, Olivier Tabary, Mucoviscidose : dans la ligne des miR, *Med Sci (Paris)* ; vol 34 ; page 554-62. 2018.

René Caquet, *Caryotype ; 250 examens de laboratoire (11e édition)*, Pages 80-82, 2010.

S. Dimassi, M. Tilla, D. Sanlaville, Anomalies chromosomiques ; *Journal de pédiatrie et de puériculture* ; vol 1269 page 1-22 : 2017.

Sediki FZ, étude génétique de la mucoviscidose dans un échantillon de la population algériennes, thèse de doctorat, spécialité génétique moléculaire et cellulaire, université des sciences et de la technologie d'Oran, faculté des sciences de la nature et de la vie, département de génétique moléculaire appliquée, soutenue le 25-05-2015.

Shutao Dai and Yan Long ; Genotyping Analysis Using an RFLP Assay ; *Methods in Molecular Biology*, vol. 1245 ; page 91-99 ; 2014.

Simona Rossi , Aristotelis Tsirigios , Antonio Amoroso , Nicoletta Mascellani , Isidore Rigoutsos , George A. Calin , Stefano Volinia, OMiR: Identification of associations between OMIM diseases and microRNAs ; *Genomics* vol 97, page 71–76. 2011.

Stefano Fagioli, Erica Daina, Lorenzo D'Antiga, Michele Colledan, Giuseppe Remuzzi, Monogenic diseases that can be cured by liver transplantation, *Journal of Hepatology* vol 59, page 595–612, 2013.

Sylvain Legrand, Blast Basic Local Alignment Search Tool, cour, université de Lille 1 science et technologie, page 1-52, Septembre 2016.

Zhe Zhang, Fangyu Liua, and Jue Chen, Molecular structure of the ATP-bound, phosphorylated human CFTR, *PNAS*, vol. 115, page 12757–12762, | December 11, 2018.

V. Malan, S. Roman ; Diagnostic des anomalies chromosomiques par CGH array en pathologie constitutionnelle : la fin du caryotype en première intention ; *Archives de pédiatrie* vol 19 ; page 437–442 ; 2012

Zeina Akiki, Biological Markers For Chronic Obstructive Pulmonary Disease And Asthma, thèse de doctorat, Spécialité: Santé Publique–Epidémiologie, université de paris Saclay, soutenue, le 11 avril 2016

Zhe Zhanga, Fangyu Liua, and Jue Chena, Molecular structure of the ATP-bound, phosphorylated human CFTR, *PNAS*, vol. 115, page 12757–12762 ; December 11, 2018 .

### **Webographies :**

**Anonyme 1 :** <http://biofaculte.blogspot.com/2015/12/5-exercices-corriges-en-genetique-humain.html>

**Anonyme 2 :** [http://www.jpbinagine.com/Sharjah/3\\_C4\\_2016/32genevol/doc32/Chap1/Chap114.html](http://www.jpbinagine.com/Sharjah/3_C4_2016/32genevol/doc32/Chap1/Chap114.html)

**Anonyme 3 :** [https://www.news-medical.net/life-sciences/Restriction-Fragment-Length-Polymorphism-\(RFLP\)-Technique.aspx](https://www.news-medical.net/life-sciences/Restriction-Fragment-Length-Polymorphism-(RFLP)-Technique.aspx)

**Anonyme 4 :** <http://www.actuscimed.com/2017/12/thelancetdiabetesandendocrinology.html>

**Anonyme 5 :** <https://www2.ac-lyon.fr/enseigne/biologie/spip.php?article123>

**Anonyme 6 :** <http://svtmarcq.over-blog.com/article-variation-genetique-et-sante-1ere-s-103390766.html>