



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

Mémoire de fin d'étude en vue de l'obtention du diplôme de Master

Option : biologie moléculaire

Thème

Etude rétrospective de la mucoviscidose et étude in silico de gène CFTR.

Présenté par :

NEKHAI .BENREKIA FEDWA SALIMA et CHEIKH CHAIMA

Devant le jury :

Président	DR MOKRANE KHEIR	MCB	USDB 1
Promoteur	Mme ABDUL HUSSAIN.A.S	MCA	USDB 1
Examineur	Mme ZEGGAR H	MCB	USDB 1

Promotion 2021-2022

Remerciements

*Le plus dur n'est pas de rédiger le rapport mais de remercier toutes les personnes qui m'ont soutenu pour réaliser ce projet. 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..*

.Aux membres du jury,

A Monsieur **MOKRANE KHEIR notre Président de jury,*

Nous vous remercions d'avoir accepté de présider le jury de ce mémoire, d'évaluer et d'enrichir ce travail. Nous vous exprimons notre profond respect...

A Madame **ZEGGAR H notre Jury de mémoire, On vous remercie d'avoir fait l'honneur d'accepter d'évaluer notre travail de mémoire. On a également été honoré de votre participation à notre jury de soutenance.*

*A notre directrice de mémoire, madame **ABDUL HUSSAIN.A.S***

Un grand honneur pour nous est, le fait que vous ayez accepté de nous encadrer et de diriger notre travail. On vous remercie pour tous vos efforts, et les progrès que vous nous avez permis de réaliser. Merci pour vos conseils avisés, pour votre disponibilité et pour la confiance que vous nous avez témoignée. Ainsi, Merci de nous avoir guidées et orientées tout au long de la réalisation de ce mémoire

Dédicace

Je dédie ce travail ;

A ceux qui m'ont soutenu, m'ont encouragé durant toute ma période d'étude, et pour leurs

*Sacrifices consentis. A ceux qui ont toujours voulu que je sois le meilleur : A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur, ma vie et mon bonheur (**maman** que j'adore).*

*A mon exemple éternel, celui qui s'est toujours sacrifié pour me voir réussir (**mon père**)*

*A l'homme de ma vie, mon soutien moral et source de joie et de bonheur (**mon mari Sofiane**)*

*A mon petit ange qui m'a ma donnée la force tout au long de cette année (**ma fille Anfel**)*

*A ma sœur : **Karima***

*A mes frère : **Faysal ; Amine***

A tous mes chères familles et amies.

Fedwa

Liste des figures:

Figure1: Principaux organes touchés par la mucoviscidose	16
Figure 2 : Parents, frères et sœurs d'un patient diagnostiquée Maladie de Wilson	19
Figure 3: Répartition de la prévalence de la mucoviscidose	20
Figure 4 : Évolution du nombre de patients depuis 1992.....	21
Figure 5: chromosome 07 (l'origine de la mucoviscidose).....	23
Figure 6 : Du gène à la protéine	24
Figure 7 : Structure de la protéine CFTR.....	25
Figure 8 : Biogenèse de la protéine CFTR	26
Figure 9 : Le mécanisme d'ouverture du canal CFTR.....	28
Figure 10 : Régulation des courants ioniques transmembranaires par la protéine CFTR...	29
Figure 11 : Classes de mutations de CFTR impliquées dans la mucoviscidose : pathogénie, fréquence et principales mutations	34
Figure 12 : Les différentes classes de mutations	37
Figure 13: différentes classes des mutations CFTR	37
Figure 14: Utilisation de BLAST sur la plate-forme NCBI.....	46
Figure 15: Utilisation de GenBank	47
Figure 16: Utilisation d'ORF Finder	48
Figure 17: Utilisation d'Ensembl	49
Figure 18: Utilisation de swiss-model	50
Figure 19: Séquence d'ARNm du gène CFTR sous format FASTA.....	52
Figure 20: Structure 3D de la protéine CFTR.....	53
Figure 21: Séquence protéique du gène CFTR.....	54
Figure22: Traduction de la séquence non mutée de l'exon 10.....	55
Figure 23: Traduction de la séquence mutée de l'exon 10.....	55
Figure 24 : représentation des résultats d'ensembl.....	56
Figure25 : Représentation graphique de lamucoviscidose.....	57

Liste des tableaux :

Tableau I: Répartition des patients atteints de mucoviscidose en Europe	21
Tableau II : Caractéristiques de différentes classes des mutations de gène CFTR....	35
Tableau III: Nomenclature et classification des mutations du gène CFTR	38
Tableau IV: Séquences des exons étudiés	45
Tableau V : Résultats de BLAST de gène CFTR.....	51

Liste des abréviations

ABC : ATP-Binding Cassette

AMPc : Adénosine Cyclique Mono Phosphate

ATP : Adenosine Tri-Phosphate

BLAST : Basic Local Alignment Search Tool

CF : Cystic fibrosis

CFTR : Cystic fibrosis Transmembrane conductance Regulator

OMS : Organisation mondiale de la Santé

TMD : Trans Membranaires hydrophobes Domain

NBD : Nucleotide Binding Domain

R : Regulator

RE : réticulum endoplasmique

Kb : kilobases

ERQC : Endoplasmique Reticulum Quality Control

KDa : kilodalton

NBD : Nucleotide Binding Domain

ENaC : Amiloride-sensitive sodium channel

Phe : Phénylalanine

OMIM : Online Mendelian Inheritance in Man

COSMIC : Catalogue Of Somatic Mutations In Cancer

Sommaire

<u>INTRODUCTION</u>	11
PARTIE THEORIQUE	12
<u>Chapitre I La mucoviscidose</u>	13
1.1: Historique	13
1.2 Histoire de la prise en charge de la maladie	14
1.3. Physiopathologie et manifestations cliniques.....	16
1.4expression de la mucoviscidose	17
1.5. Origines mucoviscidose... ..	18
<u>Chapitre II EPIDEMIOLOGIES DE LA MUCOVISCIDOSE</u>	
2.1 Épidémiologie Dans le monde	20
2.2 Épidémiologie En France.....	21
2.3 Épidémiologie En Algérie	22
<u>Chapitre III ASPECT MOLECULAIRE DU GENE CFTR</u>	23
3.1 Le gène CFTR.....	23
3.2 La protéine CFTR	25
3.3 Les mutations de la protéine CFTR.....	33
<u>Chapitre IV OUTILS BIO-INFORMATIQUE</u>	39
4.1 Définition de bio-informatique.....	39
4.2 Outils bio-informatique d'étude des gènes CFTR.....	40

PARTIE PRATIQUE	43
<u>Chapitre V MATERIELS ET METHODE</u>	44
5.1 Matériels.....	44
5.2 Méthode.....	45
<u>Chapitre VI RESULTATS ET DESSCUSION</u>	51
6.1 Les résultats de Blast n de l'ARNm CFTR	50
6.2 Les résultats de GenBank	52
6.3 Les résultats de swiss-prot d'ARNm du gène CFTR ID de CFTR sur swiss-model	53
6.4 Les résultats d'Ensembl.....	56
6.5 Les résultats d'ONIM	57
<u>Conclusion</u>	59
<u>Référence bibliographique</u>	60

Résumé :

La mucoviscidose est une maladie génétique souvent fatale affectant les organes exocrine (tractus respiratoire, digestif, génital et cutané) et particulièrement fréquente dans la population d'Europe du Nord et d'Amérique du Nord. Pour pionnière qu'elle soit dans des secteurs aussi variés que la génétique, la physiologie, ou la pharmacologie, la recherche sur la mucoviscidose se heurte néanmoins, aujourd'hui encore, à d'importantes zones d'ombre. En particulier, les facteurs responsables du déclenchement et de la sévérité de l'atteinte respiratoire dans la mucoviscidose restent mal identifiés. L'objectif principal de cette étude est l'analyse bio-informatique des mutations des molécules (gène CFTR : responsable de la maladie mucoviscidose: 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 bio-informatiques.

Mots-clés : Mucoviscidose, gène CFTR, mutations, séquençage, étude in silico.

Abstract :

Cystic fibrosis (CF) is a genetic disease, particularly frequent among northern Europeans and northern Americans which affects, often lethally, the function of exocrine organs (including the respiratory, digestive, genital and cutaneous mucosae). Despite its pioneering role in areas as diverse as genetics, physiology or pharmacology, CF research still faces essential unsolved questions. In particular, the factors involved in the appearance and severity of CF airway disease remain mostly unraveled so far. The main objective of this study is the bioinformatics analysis of molecule mutations (CFTR gene: responsible for the disease cystic fibrosis: 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 protein and disease phenotype using a few bioinformatics tools.

Keywords: Cystic fibrosis, CFTR gene, mutations, sequencing, in silico study

المخلص :

التليف الكيسي هو مرض وراثي قاتل في كثير من الأحيان يؤثر على أعضاء الغدد الصماء (الجهاز التنفسي والجهاز الهضمي والأعضاء التناسلية والجلد) ويتكرر بشكل خاص في سكان شمال أوروبا وأمريكا الشمالية. على الرغم من كونها رائدة في قطاعات متنوعة مثل علم الوراثة أو علم وظائف الأعضاء أو علم العقاقير ، فإن الأبحاث حول التليف الكيسي لا تزال حتى اليوم تواجه مناطق رمادية كبيرة. على وجه الخصوص ، لا تزال العوامل المسؤولة عن ظهور وشدة تورط الجهاز التنفسي في التليف الكيسي غير محددة بشكل جيد. الهدف الرئيسي من هذه الدراسة هو تحليل المعلومات الحيوية للطفرات الجزيئية (جين CFTR: المسؤول عن مرض التليف الكيسي: 1521-1523 delCTT (F508del) ، c.680T> G (L227R) ، c.1624G> T ، (G542X) ؛ C.3909C> G (N1303K) و c.1652C> A (G551D) وتأثيرها على النمط الظاهري للبروتين والأمراض باستخدام عدد قليل من أدوات المعلوماتية الحيوية.

الكلمات المفتاحية:

التليف الكيسي ، جين CFTR ، الطفرات ، التسلسل ، دراسة السيليكو

1 Introduction

La mucoviscidose est la maladie génétique autosomique récessive létale la plus fréquente dans la population d'origine caucasienne. Mucoviscidose, également appelé fibrose kystique, signifie « mucus visqueux ». C'est une maladie pouvant toucher l'ensemble des épithéliums glandulaires de l'organisme. (*Gibas et Jambeck, 2002*).

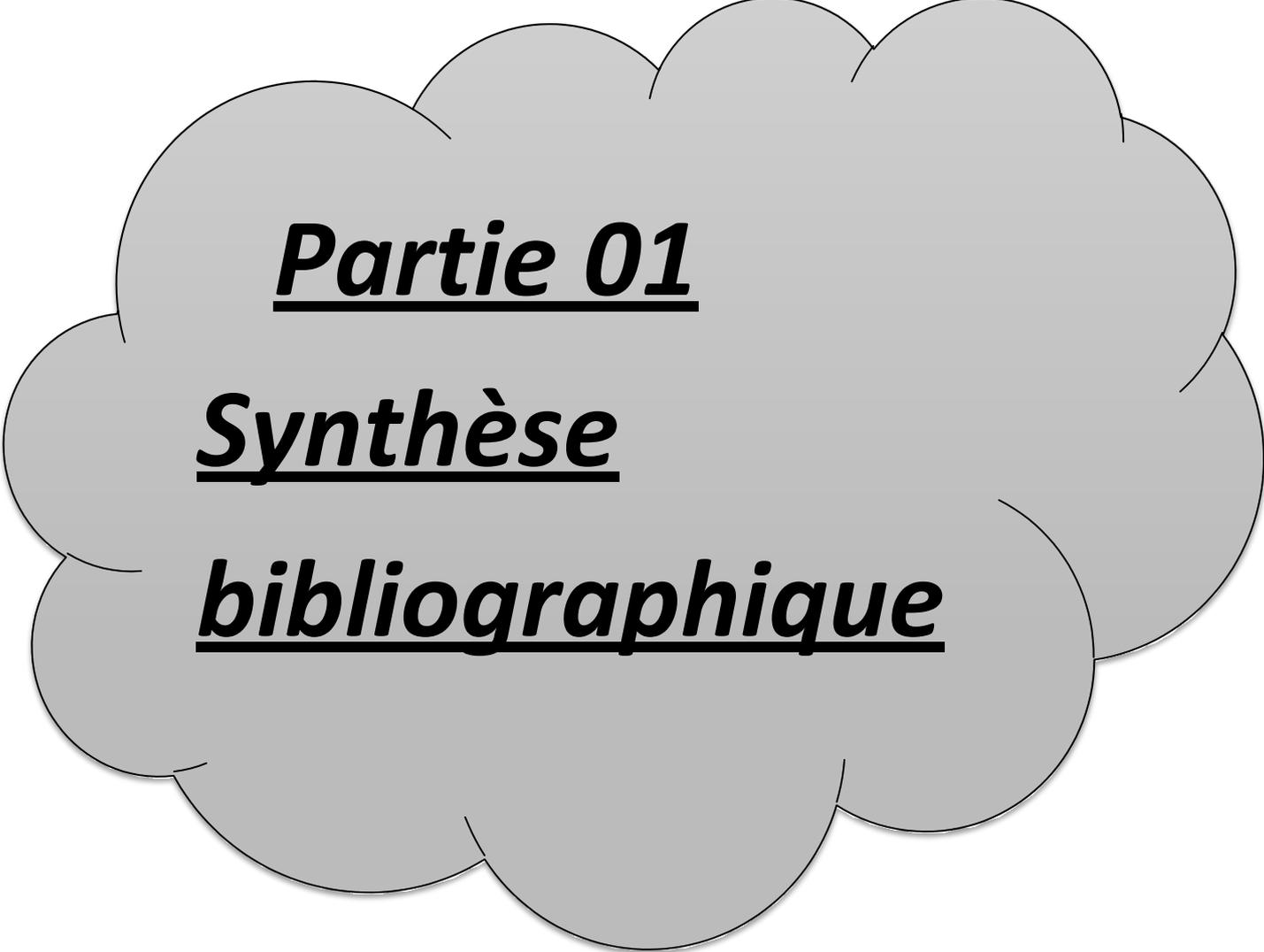
La protéine défectueuse à l'origine de la mucoviscidose est restée longtemps inconnue. En 1989, le gène sera localisé sur le bras long du chromosome 7 et dénommé CFTR pour Cystic Fibrosis Transmembrane conductance Regulator 1-3. Le gène CFTR code pour un canal chlorure de la membrane cytoplasmique situé au pôle apical des cellules épithéliales et dont la fonction est de réguler le transport de chlore à travers la membrane selon un gradient électrochimique.

Chez l'homme, l'expression pulmonaire de CFTR varie en fonction du type cellulaire. Elle est fortement exprimée dans les cellules ciliées de l'épithélium bronchique et dans les pneumocytes de type II au niveau alvéolaire. Le déficit en protéine CFTR joue également un rôle important dans la réponse immunoinflammatoire locale et dans la défense antimicrobienne. (*Gibas et Jambeck, 2002*).

La biologie moléculaire base sur l'étude de fonctionnement des gènes causant la maladie mucoviscidose étant due à des mutations ponctuelles de l'ADN.

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 connaissances pour mieux comprendre.

Le but de ce travail est : l'étude des mutations du gène CFTR et l'analyse bioinformatique dans la population algérienne.



Partie 01

Synthèse

bibliographique

Chapitre I: La mucoviscidose

1.1. Historique

En 1936 La mucoviscidose a été décrite scientifiquement comme maladie (déjà connue depuis longtemps). Au Moyen Âge on connaissait le sort funeste du nouveau né dont la mère remarquait le « baiser salé », c'est-à-dire le goût salé laissé par un baiser sur le front de l'enfant. Welsh cite : « Malheur à l'enfant chez qui un baiser sur le front a un goût salé. Il est ensorcelé et doit bientôt mourir. » (*Lindee, 2000*).

Selon Busch une des premières descriptions médicales des lésions pancréatiques rencontrées dans la mucoviscidose pourrait être un rapport d'autopsie fait par le professeur Pieter Pauw, à Leyde aux Pays-Bas en 1595, dans lequel il décrit une patiente chétive de 11 ans supposée ensorcelée et présentant un pancréas élargi, dur et blanc (*Lindee, 2002*).

Au début du XXe siècle apparaissent les premières observations associant maladie pulmonaire, diarrhée et anomalie pancréatique avec plusieurs cas dans la même famille. Ces descriptions se focalisent le plus souvent sur les problèmes digestifs, stéatorrhée et troubles pancréatiques, et leurs auteurs en font une forme de maladie cœliaque même si les problèmes bronchopulmonaires sont également notés. (*Gibas et Jambeck, 2002*).

C'est en 1936, dans une thèse écrite en allemand et présidée par le pédiatre suisse Guido Fanconi, que la maladie est décrite pour la première fois chez des enfants supposés atteints de maladie cœliaque, sous le nom de « fibrose kystique du pancréas et bronchectasie ». (*Gibas et Jambeck, 2002*).

La mucoviscidose ne fut considérée comme une entité pathologique distincte qu'en 1938 par la pédiatre américaine Dorothy Hansine Andersen, médecin au Babies' Hospital de New York, qui publia un article intitulé « fibrose kystique du pancréas et ses relations avec la maladie cœliaque ». C'est en pratiquant des autopsies sur des nourrissons qu'elle décrivit les caractéristiques cliniques et histologiques de la maladie, notamment l'obstruction intestinale néonatale, les complications respiratoires et digestives et les lésions histologiques spécifiques du pancréas. Elle relia cette maladie à un déficit en vitamine A et persista à soutenir cette théorie pendant de nombreuses années bien qu'elle ne fût jamais confirmée. Farber était persuadé que la maladie était due à une diffusion généralisée de mucus visqueux.

C'est au début des années 1980 que le lien physiopathologique fut fait entre d'une part l'anomalie de la sécrétion de mucus, entraînant des obstructions glandulaires avec anomalies

histologiques et d'autre part l'anomalie de la sueur, entraînant des sécrétions salées sans anomalie histologique. En 1983 Quinton, lui-même atteint de la mucoviscidose, montra que l'imperméabilité au chlore qu'il avait trouvée dans les glandes sudoripares était la cause de l'augmentation des électrolytes dans la sueur. Ceci fut considéré comme une étape majeure dans la compréhension de la maladie. Restait à localiser et à identifier le gène dont la mutation provoque la mucoviscidose, une tâche rendue difficile car seules la sémiologie clinique, le mode de transmission autosomique récessif et l'anomalie de transport du chlore étaient alors connus (*Regard et al.2018*).

En l'absence de toute connaissance sur la protéine défectueuse, et donc de marqueur chimique identifié, la récente technique du clonage positionnel - ou génétique inverse, qui permet d'identifier un gène sans connaître la protéine qu'il code, fut utilisée. Cette méthode consiste grâce à des analyses statistiques de liaison génétique à déterminer la région du chromosome où le gène a une forte probabilité de se trouver, afin de la séquencer et d'étudier les gènes exprimés. Le gène de la mucoviscidose est le premier gène à avoir été cloné uniquement par analyse de liaisons. Retrouvent un lien entre l'enzyme paraoxinase et le gène CF en étudiant des familles comportant plus d'un enfant atteint.

La même année, Tsui et al., suite à des études sur des souris hybrides, localisent le gène sur le bras long du chromosome 7 grâce à un marqueur RLFP lié à la fois à la paraoxinase et à la mucoviscidose. (*Jamie L. Sturgill.2018*)

En 1989, le gène impliqué dans la mucoviscidose est isolé par les équipes de Lap-Chi Tsui, Collins et Riordan. L'anomalie génétique à l'origine de la maladie est enfin découverte, il s'agit d'une mutation d'un gène localisé en 7q31 et contenant 27 exons, nommé cystic fibrosis codant une protéine transmembranaire appelée cystic fibrosis transmembrane conductance regulator composée de 1480 acides aminés. La découverte de l'anomalie génétique permit par la suite d'ajouter le génotypage au protocole diagnostique, et d'envisager la thérapie génique. (*Regard et al.2018*).

1.2 Histoire de la prise en charge de la mucoviscidose

Dans les années 1940, la maladie était considérée comme étant avant tout un problème nutritionnel avec une déficience en vitamine A. La prise en charge se résumait essentiellement à un régime riche en protéines, des injections intramusculaires de vitamine A à hautes doses, des extraits pancréatiques et des inhalations de pénicilline. À cette époque la durée de

survie reste courte, 64% des 28 patients d'une série de la Mayo Clinique n'atteignant pas l'âge de 7 ans. La première drogue antibactérienne, un sulfamidé commercialisé sous le nom de Prontosil, fut disponible en 1934 et la pénicilline sous forme injectable en 1944. D'autres antibiotiques suivirent et eurent un rôle essentiel dans le traitement des patients.

À cette date, le principal germe pathogène était *Staphylococcus aureus* et de nombreuses souches étaient encore sensibles à la pénicilline. En 1946 Di Sant Agnese et D. Andersen écrivent que l'amélioration du pronostic lors de cette décennie est due à « un régime approprié débuté rapidement et poursuivi de façon continue, l'utilisation de sulphadiazine pendant la période de toux chronique et de nébulisation de pénicilline. » Dans les années 1950, le sort des enfants atteints est toujours considéré comme sans espoir mais quelques centres spécialisés dans la prise en charge de la mucoviscidose voient le jour aux États-Unis et au Royaume-Uni. Le drainage postural est alors un des traitements traditionnels au stade de bronchectasie et en 1950 Young propose de le débiter dès le diagnostic posé. (*Lindee, 2002*).

Dans les années 1960, la création d'organisations nationales consacrée à la maladie permet une approche collaborative entre la communauté médicale et les parents et patients. Des résultats encourageants proviennent néanmoins de Londres, en 1969 David Lawson est le premier à suggérer qu'une antibioprophylaxie anti-staphylococcique continue pourrait améliorer la survie et en 1972 il écrit qu'« un diagnostic plus rapide par un dépistage néonatal amélioré est essentiel car actuellement le diagnostic est posé au stade alors que des lésions pulmonaires sont déjà présentes ». Elle écrit qu'« un traitement vigoureux et un contrôle de l'infection dès la petite enfance, peut prévenir la plupart de ces patients de devenir insuffisants respiratoires à l'enfance ». En 1983 le diagnostic anténatal devient disponible pour les familles ayant déjà un enfant affecté. (*Lindee, 2002*).

Une avancée majeure pour les patients au stade terminal de la maladie est la première transplantation cœur-poumon réalisée en 1984 à Harefield au Royaume-Uni et à Chapel Hill aux États-Unis. Il devient évident que les patients avec un apport normal en graisse ont un meilleur apport énergétique, un meilleur état nutritionnel et une meilleure croissance. Au début des années 1990, une vague d'enthousiasme suit l'identification du gène CFTR, certains patients pensant qu'un traitement curatif a été découvert. On met en évidence les risques de transmission d'agents infectieux tels que *Burkholderia cepacia* et *Pseudomonas aeruginosa*

entre patients, ce qui entraîne une disparition des réunions et groupes de patients, modifiant profondément leur vie sociale. (*Lindee, 2002*).

1.3 Physiopathologie et manifestations cliniques

Les manifestations cliniques de la mucoviscidose sont très hétérogènes tant entre les différentes familles qu'au sein d'une même famille.

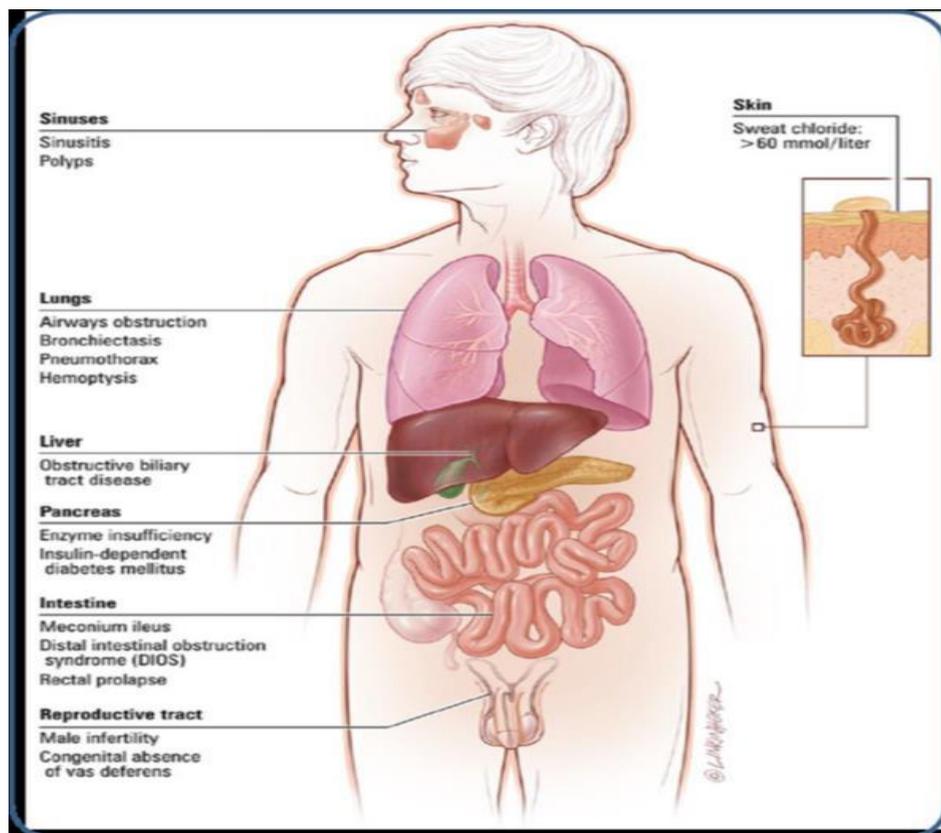


Figure.01. Principaux organes touchés par la mucoviscidose, d'après *Paranjape et Mogayzel(2018)*

L'excès de mucus et l'absence de clairance mucociliaire (mécanisme par lequel les particules inhalées supérieures à une certaine taille sont propulsées vers les voies digestives) favorisent l'infection bronchique par des bactéries opportunistes (*Staphylococcus aureus, Pseudomonas aeruginosa et Hemophilus influenzae*).

Cette infection engendre une réponse inflammatoire exacerbée au niveau des voies aériennes. Ce processus inflammatoire devient chronique et provoque la destruction des tissus

pulmonaires, entraînant ainsi une perte progressive de la fonction respiratoire jusqu'au décès du malade.

Les manifestations cliniques sont variables en fonction des patients et la maladie peut se déclarer à des âges différents. Dès la période pré-natale, il est parfois possible de voir, lors de la 2e échographie, la stagnation des matières fécales chez le fœtus, entraînant une dilatation des intestins (méconium).

1.4 expression de la mucoviscidose

❖ Chez le nouveau-né

- L'expulsion des premières selles peut être retardée,
- Diarrhée graisseuse.
- Toux
- Bronchiolite persistante.
- Persistance d'une jaunisse
- voire dans les cas les plus graves, cause d'une occlusion digestive
- la reprise de poids se fera plus tardivement : 2 à 3 semaines après la naissance au lieu de 8 à 10 jours.

❖ Chez l'enfant

Au niveau respiratoire, il présente des quintes de toux à répétitions, des bronchites fréquentes pouvant se compliquer d'une surinfection bactérienne. Les germes les plus fréquemment rencontrés sont l'*haemophilus influenzae* (jusqu'à l'âge de 5 ans), le staphylocoque aureus, et le *pseudomonas aeruginosa*. Ce dernier est le plus redoutable du fait des résistances existantes à de nombreux antibiotiques. (*Lubamba et al., 2012*).

Au niveau digestif, il existe une insuffisance pancréatique entraînant des troubles de la digestion : les graisses sont mal absorbées et par conséquent, l'enfant présente une diarrhée avec des selles abondantes, malodorantes, avec un aspect « huileux ». Ceci entraîne une dénutrition, avec des déficits nutritionnels notamment en vitamines A, D, E, K et a pour conséquence une courbe de poids inférieure à la normale avec des répercussions sur le développement de l'enfant. A l'inverse, il peut y avoir une constipation due à l'épaississement des sécrétions intestinales aggravée par la déshydratation. (*Lubamba et al., 2012*).

La peau de l'enfant, du fait des modifications du transport du chlore à travers les membranes, peut avoir un goût salé.

❖ **Chez l'adolescent et l'adulte**

Au niveau respiratoire, des infections pulmonaires à répétitions, des difficultés pour respirer se manifestent dans un premier temps lors des efforts, puis de façon constante. Les sécrétions sont opaques et purulentes. Les bactéries retrouvées sont les mêmes que chez l'enfant, présence aussi d'un champignon nommé *Aspergillus Fumigatus*. De même, les sinusites sont fréquentes chez les patients atteints de mucoviscidose pouvant se compliquer dans un second temps d'une polypose nasale. (*Pietrement et al., 2008*).

Au niveau digestif, l'insuffisance pancréatique se manifeste par une diarrhée chronique associée à des douleurs abdominales. L'atteinte du pancréas peut avoir pour autre conséquence, l'apparition d'un diabète nécessitant des injections d'insuline. De plus, toujours via le même mécanisme d'obstruction des canaux, une altération des voies biliaires est possible ayant pour conséquence, dans les cas les plus graves, l'altération du foie appelée cirrhose biliaire primitive. (*Pietrement et al., 2008*).

Enfin, la mucoviscidose peut être responsable d'une stérilité chez l'homme, par l'obstruction des canaux permettant la circulation des spermatozoïdes. Chez la femme, présence d'une infertilité due à l'épaississement de la glaire cervicale gênant le passage de ces derniers. (*Elborn, 2016*)

1.5 Origines mucoviscidose

La mucoviscidose est une maladie génétique récessive létale (les deux parents doivent porter le gène défectueux pour transmettre la maladie). Le gène responsable de cette pathologie code pour une protéine appelée Cystic Fibrosis Transmembrane conductance Regulator (CFTR), découverte en 1989. Plus de 2 000 mutations ont été identifiées à ce jour, parmi lesquelles la plus fréquente (rencontrée chez environ 80 % des malades en France) est la mutation F508del. L'altération du canal CFTR modifie les échanges ioniques et rend le mucus plus visqueux et plus abondant. Bien que la qualité du mucus de nombreux organes soit affectée, les atteintes respiratoires sont la cause majeure de mortalité et de morbidité. (*Ouchtati A, Dali H, 2018*)

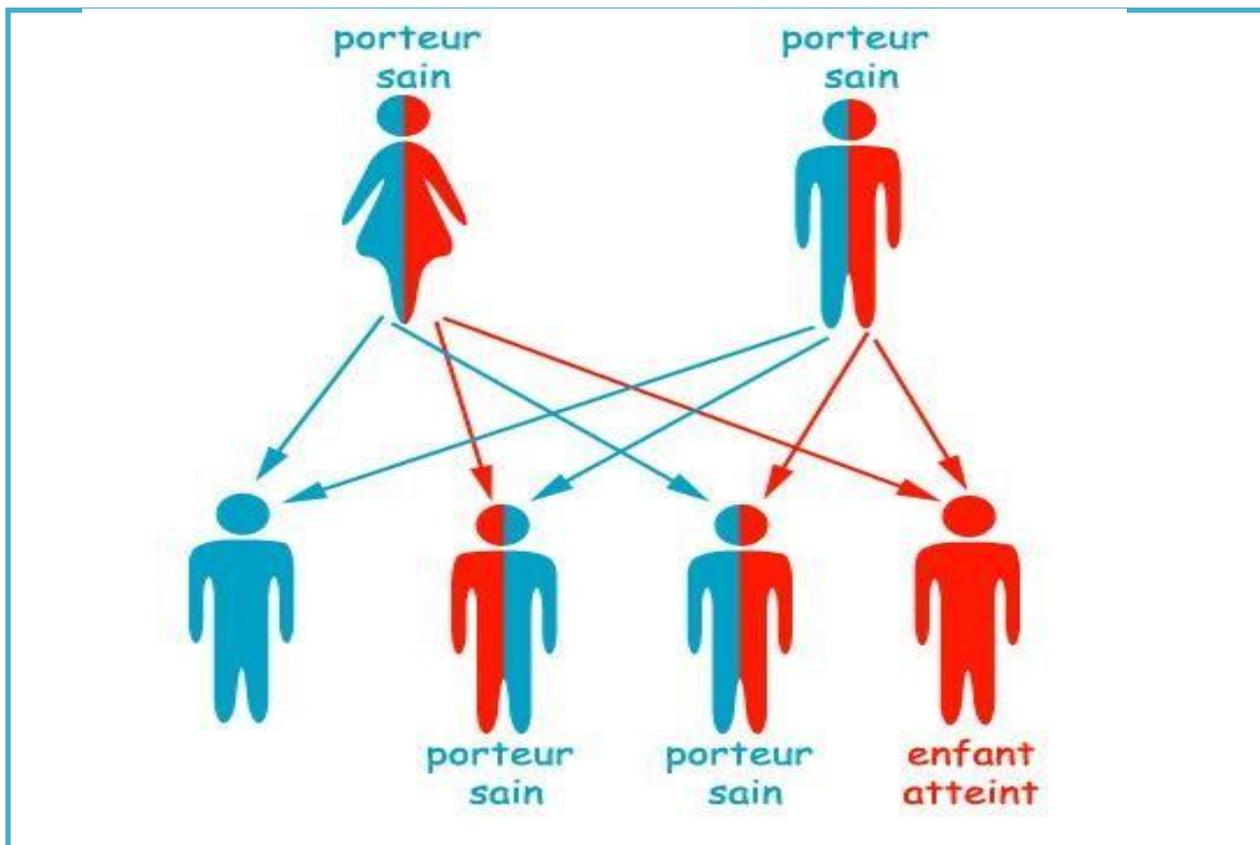


Figure02 : Parents, frères et sœurs d'un patient diagnostiquée Maladie de Wilson (cas le plus fréquent) (*Vincent Brossas2002*)

Dans la très grande majorité des cas, les parents d'une personne atteinte de maladie de Wilson sont porteurs d'un seul allèle anormal. Ils sont dits hétérozygotes ou porteurs sains car ils ne développeront pas la maladie.

Chapitre II: épidémiologie de la maladie

2.1 Épidémiologie Répartition dans le monde

Plus de 80 000 malades sont recensés à travers le monde. La mucoviscidose touche de façon préférentielle les populations caucasiennes. La prévalence y est estimée à environ 1/3000 naissance, et la proportion de sujets hétérozygotes porteurs sains à 1/30. (*Mirtajani et al,2017*).

La maladie affecte environ un Américain-blanc sur 3000, un Latino-américain sur 11 000 et un Afro-américain sur 15 000 - 20 000. La mucoviscidose est plus rare en Afrique et en Asie, avec une prévalence de 1 sur 100 000 en Iran et de 1 sur 350 000 au Japon .Il est important de noter que les données des pays moins développés pourraient être sous-estimées du fait de l'absence de registre national et d'un recensement insuffisant. L'OMS demande déjà à ces pays d'instaurer un registre national afin d'élaborer une approche stratégique à l'échelle mondiale.

La prévalence de la mucoviscidose est variable à l'échelle de la planète mais également entre régions au sein d'un même pays. Par exemple en France, la proportion de patients atteints de mucoviscidose est plus élevée à l'Ouest (particulièrement en Bretagne), au Nord et au Sud-Est du pays, ainsi que sur l'île de la Réunion (*Mirtajani et al ,2017*).

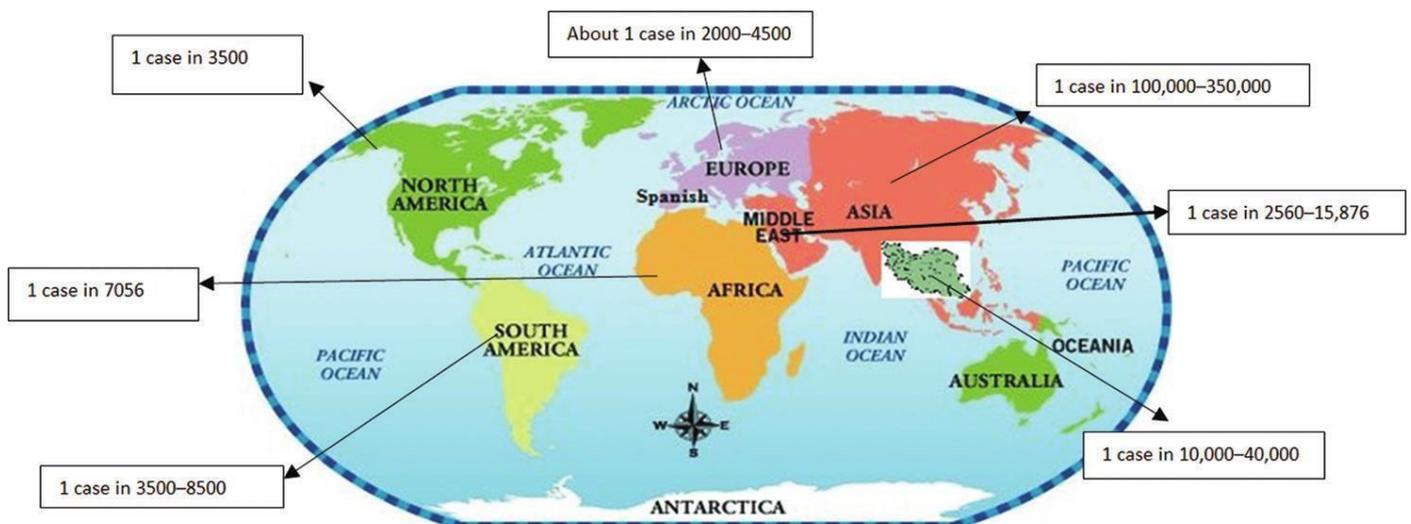


Figure 03 : Répartition de la prévalence de la mucoviscidose dans le monde. *Source :* *Geographical distribution of cystic fibrosis; The past 70 years of data analy*

Une incidence plus élevée est notée dans les pays européens à un taux de 1 sur 2000 à 3000 naissances (*Tableau I*).

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

pays	Population	Prévalence	Incidence
Italie	58057	0/872	1 :4238
France	60424	0/750	1 :4700
Portugal	10524	0/271	1 :6000
Finlande	5215	0/123	1 :25000
Espagne	40281	0/546	1 :3750
Allemagne	82425	0/829	1 :3300

2.2Épidémiologie et répartition en France

On compte 200 nouveau-nés malades chaque année en France pour un total de 6757 patients recensés en 2016 (couverture de 90%)

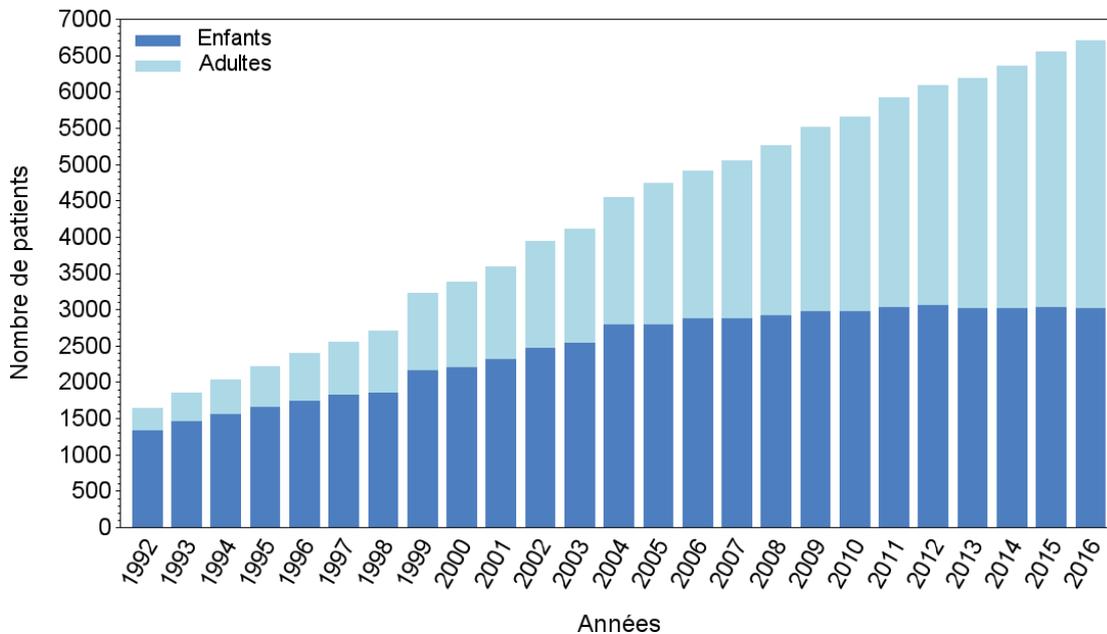


Figure 04 : Évolution du nombre de patients depuis 1992. *Registre français de la mucoviscidose – Bilan des données 2016*

2.3.épidémiologie et répartition en Algérie

La prévalence est aussi élevée en Algérie et dans d'autres pays du Maghreb. Pendant plusieurs années le diagnostic de cette affection été freiné (il a été dit que cette maladie n'existait pas en Algérie). Aujourd'hui, 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, le 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 (*Bendridi et al ,2014*).

Chapitre III : ASPECT MOLECULAIRE DU GENE CFTR

3.1 Le gène CFTR

Une analyse de liaison a permis de localiser le gène responsable de la mucoviscidose sur le bras long du chromosome 7 au locus q31 (Tsui et al., 1985) (figure 5).

C'est en 1989 que trois équipes ont identifié de manière concomitante le gène CFTR (Cystic Fibrosis Transmembrane Conductance Régulator) par clonage positionnel (Kerem et al., 1989 ; Riordan et al., 1989 ; Rommens et al., 1989).

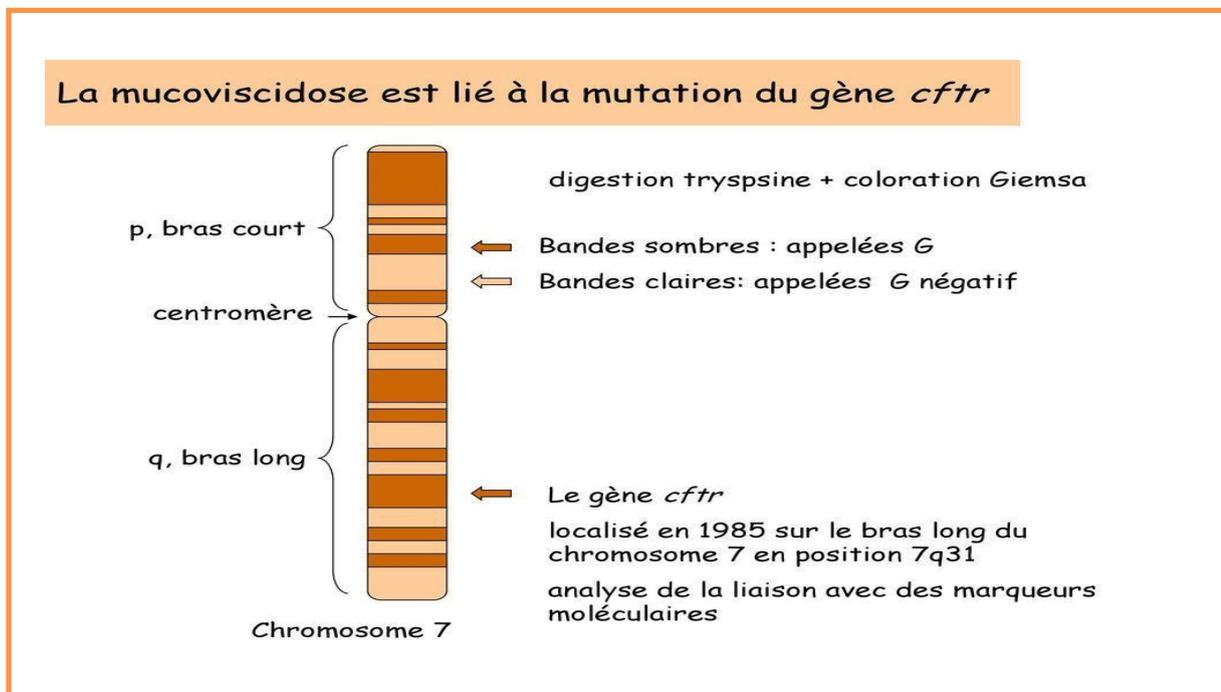


Figure05 : chromosome 07 (l'origine de la mucoviscidose) (Lamoril, 2008)

Ce gène se compose d'environ 250 000 paires de base et comprend 27 exons. Ces derniers sont transcrits en ARN messager de 6,5 kb. (figure06)(Tsui et al., 1985).

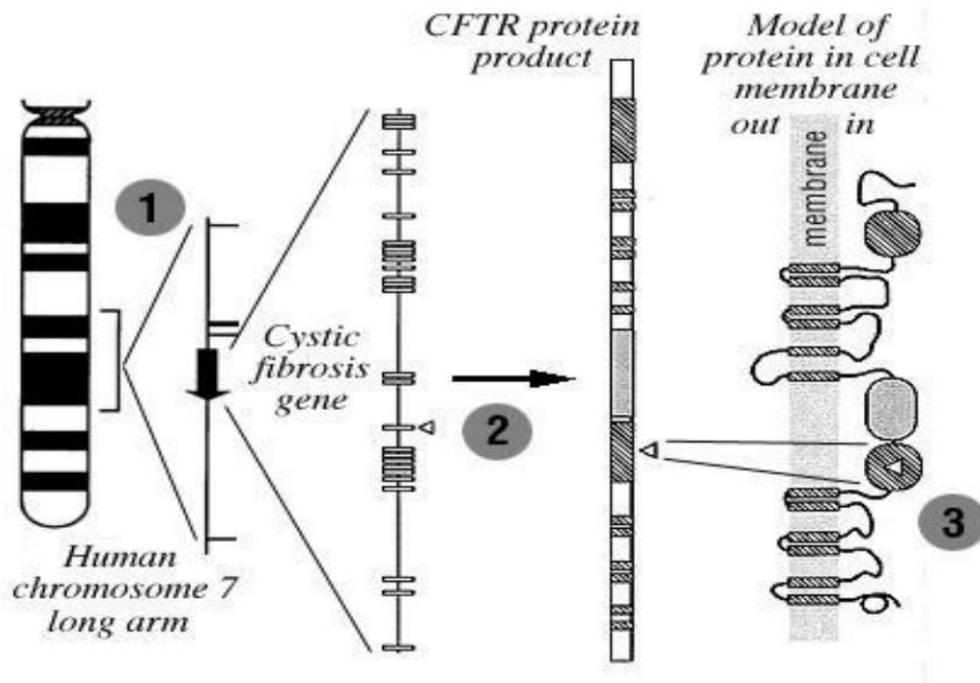


Figure 06. Du gène à la protéine (d'après Tsui dans *Science.ca*)

3.2 La protéine CFTR

Le gène CFTR code une protéine CFTR membranaire responsable du flux d'ions chlorures.

a) La structure et l'expression de la protéine CFTR

La protéine CFTR est une protéine transmembranaire de 1480 acides aminés (*Riordan et al., 1989*). Ses caractéristiques structurales ont permis de la classer dans la superfamille des transporteurs membranaires ABC (ATP-Binding Cassette) d'où son nom ABCC7 (*Hyde et al., 1990*).

Cette superfamille utilise l'ATP ou Adénosine Triphosphate comme source d'énergie pour transporter les substrats.

La protéine est constituée de 7 domaines.

- Il existe 2 domaines transmembranaires hydrophobes, chacun avec 6 étendues d'hélices alpha. Ceux-ci sont chacun connectés à un domaine de liaison aux nucléotides (NBD) dans le cytoplasme.
- Deux domaines intra cytoplasmiques hydrophyles (NBD1 et NBD2) reliés aux TMD par quatre boucles intracellulaires (de CL1 à CL4) et capables de fixer l'ATP.
- 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).

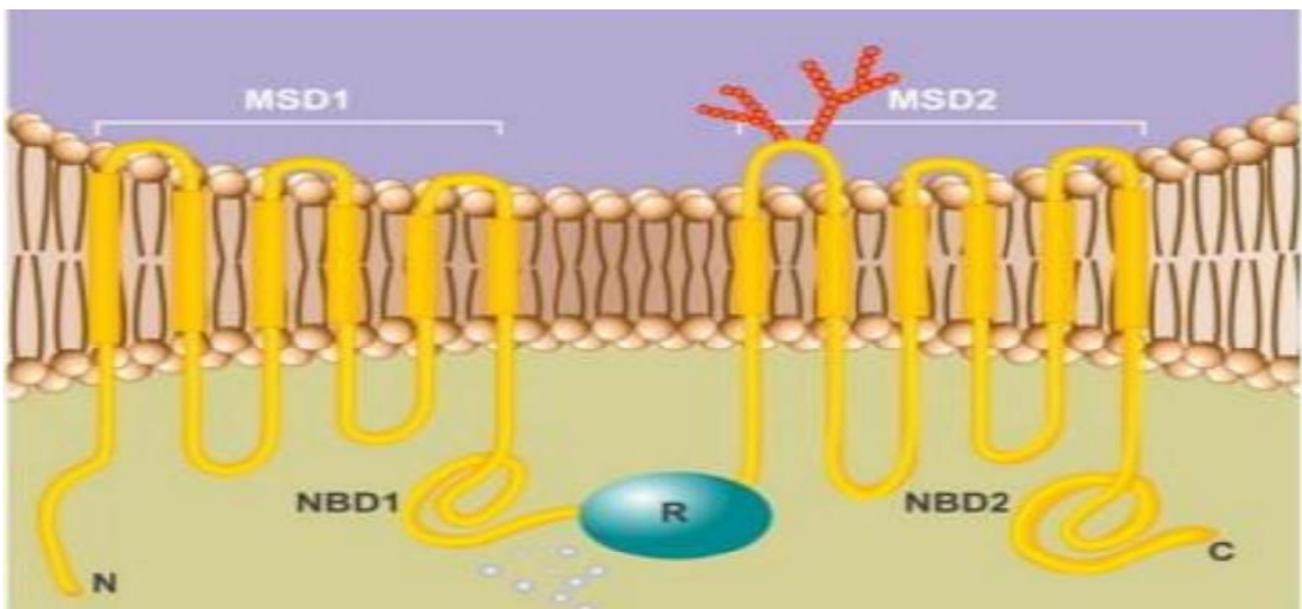


Figure07 : Structure de la protéine CFTR (*Lubamba et al. 2012*)

b. Biogenèse de la protéine CFTR

La maturation complète de la protéine CFTR s'effectue en environ 2 heures et pendant cette période, seulement 1/3 des précurseurs sont convertis en une protéine mature (*Gelman and Kopito, 2003*).

Comme la plupart des protéines, la protéine CFTR est assemblée au niveau du réticulum endoplasmique (RE) (**figure 8**).

C'est à ce moment-là qu'elle va subir un processus de maturation complexe comprenant différentes étapes de glycosylation. La protéine CFTR est synthétisée sous forme d'un précurseur non-glycosylé ou immature sur lequel a lieu l'addition co-translationnelle d'un oligosaccharide. Cette forme immature (ou forme A, 130kDa) est pliée et intégrée à la membrane du RE. Les cycles de repliement de la protéine CFTR sont contrôlés par le système ERQC (Endoplasmique Reticulum Quality Control) soit indirectement via des interactions avec des molécules chaperonnes, Hsp70/Hsc70, Hdj-2, Hsp90 et calnexine, soit directement via les N-glycannes indépendamment des chaperonnes, lesquelles jouent aussi la stabilité de la protéine CFTR (*Glozman et al., 2009*).

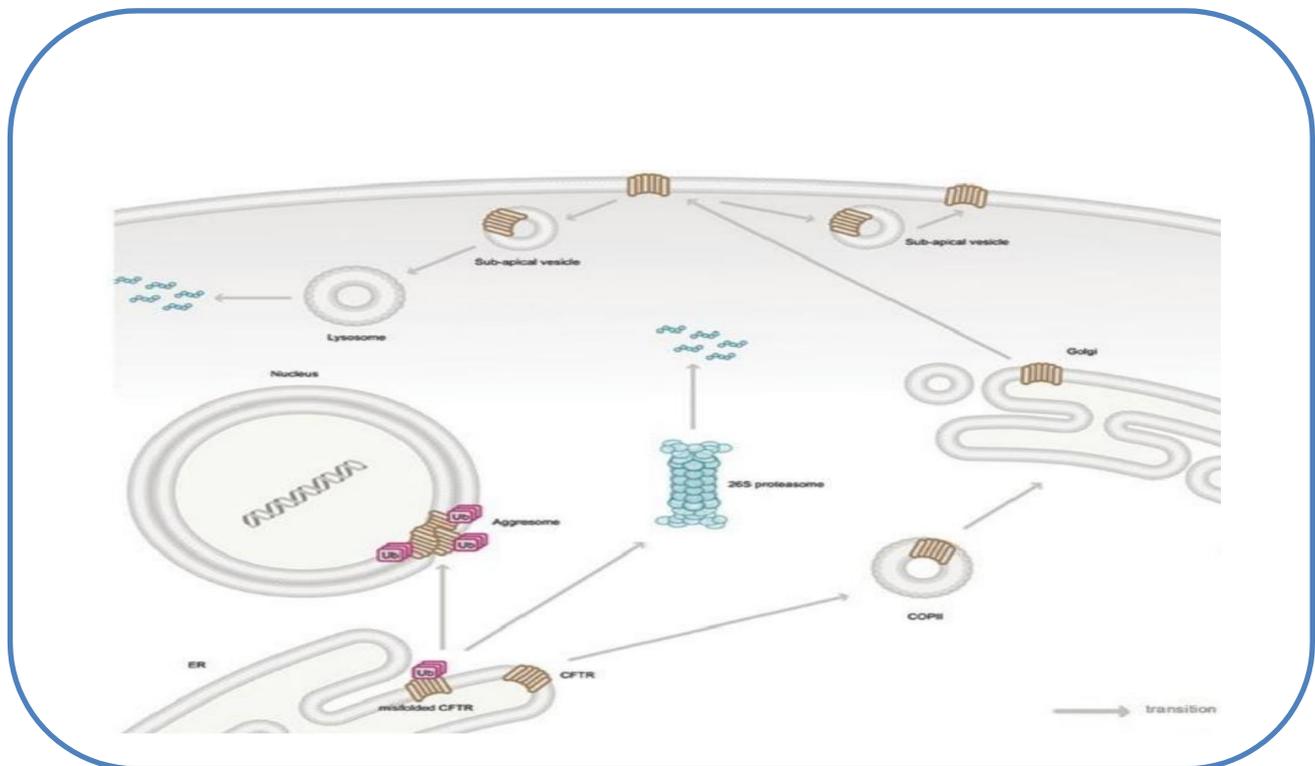


Figure 08 : Biogenèse de la protéine CFTR (*d'après Turnbull et al., 2007*)

Elle est ensuite glycosylée par addition de deux groupes de glucide sur la 4^{ème} boucle extracellulaire. Cela modifie son poids moléculaire à 150 kDa. Cette nouvelle forme immature est appelée forme B. Elle quitte le RE par des vésicules « coat complex II » (COPII) (*Figure 8 ; Wang et al., 2004 ; Turnbull et al., 2007*) en direction de l'appareil de Golgi. La dernière étape de la maturation a lieu au niveau du Golgi et se traduit par une modification des groupes de glycosylation. Cette nouvelle forme mature s'appelle la forme C et possède un poids moléculaire de 170 kDa. Cette forme mature est ensuite adressée à la membrane plasmique pour y exercer ses différentes fonctions. (*Turnbull et al., 2007*).

b) Le recyclage ou la dégradation de la protéine CFTR

Une grande proportion de la quantité de protéine CFTR (environ 70%) encore immature sous la forme A, est ubiquitinylée puis la protéine CFTR est dégradée par le système protéasome associé au reticulum endoplasmique. Cependant, une étude a montré que l'efficacité du processus de maturation de CFTR varie d'un type cellulaire à l'autre (*Varga et al., 2004*).

La demi-vie à la surface cellulaire de la protéine CFTR est très courte. En effet, CFTR est rapidement internalisée par endocytose et est recyclée à la surface et dégradée par les lysosomes et le protéasome. (*Lukacs et al., 1997*).

c) Le canal CFTR, une protéine multifonctionnelle

(1) Le canal chlorure

La principale fonction du canal CFTR est le transport des ions chlorures à travers la membrane plasmique de nombreuses cellules épithéliales telles que celles du poumon, du foie, du pancréas, du tractus digestif et de la peau. Ce transport se fait selon un gradient électrochimique. (*Devidas and Guggino, 1997*).

Par analogie avec les autres protéines canalaire, on considère que les domaines transmembranaires (TMD1 et TMD2) interviennent dans la formation du pore. Alors que le domaine TMD2 est peu impliqué, le domaine TMD1 joue un rôle important dans la

conductance et dans la sélectivité du canal chlorure Les segments membranaires M1, M5, M6, et M12 sont impliqués dans la zone formant le pore (*Schwiebert et al., 1998a*).

La partie du pore la plus resserrée a un diamètre d'environ 5.3 Å (*Hanrahan et al., 1998*).

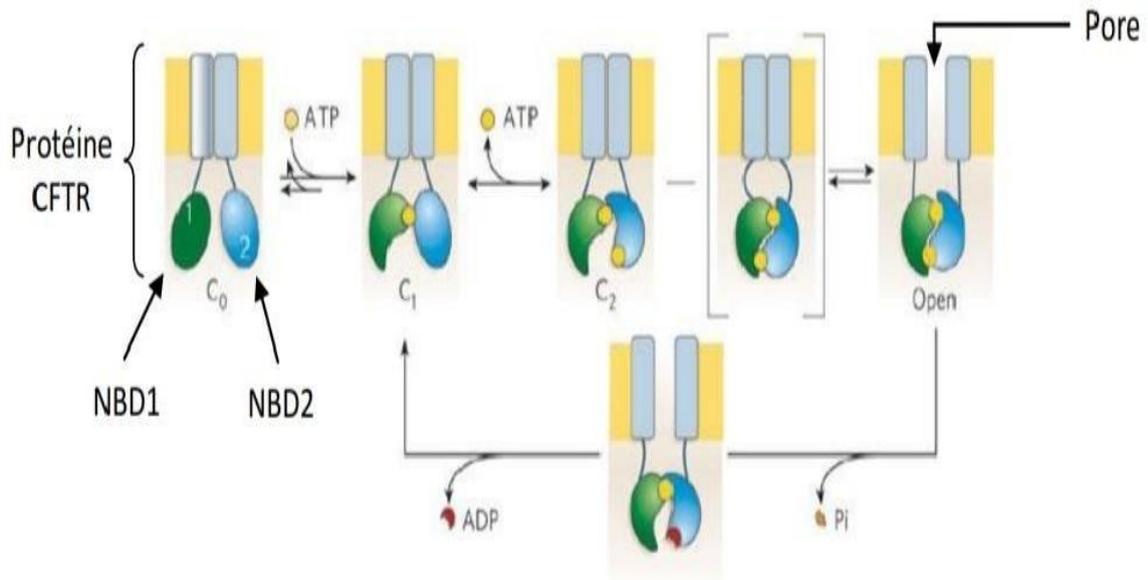


Figure 9 : Le mécanisme d'ouverture du canal CFTR (*d'après Gadsby et al., 2006*)

C'est un canal chlorure de faible conductance régulé par la voie AMPc. L'augmentation intracellulaire d'AMPc déclenche la phosphorylation de multiples résidus sérines dans le domaine R par les protéines kinases A et C. La liaison d'une molécule d'ATP sur chacun des NBDs permet un rapprochement de ces deux domaines (**figure 9**).

Cette interaction entraîne l'envoi d'un signal à la région transmembranaire permettant l'ouverture du canal. L'hydrolyse de l'ATP au niveau du domaine NBD2 met fin à l'interaction entre les domaines NBD1 et NBD2 entraînant la fermeture du canal (*Gadsby et al., 2006 ; Zhou et al., 2006*).

L'état déphosphorylé du domaine R maintient le canal fermé. La probabilité d'ouverture dépend du nombre de sérines phosphorylées (*Gadsby and Nairn, 1999*).

Cette phosphorylation est essentielle : le canal déphosphorylé est incapable d'hydrolyser l'ATP. Très rapidement, après la phosphorylation par les PKA, la déphosphorylation par des protéines phosphatases inactive le canal. Le domaine NBD1 est le site d'hydrolyse de l'ATP couplé à l'ouverture du canal et le domaine NBD2 est le site d'hydrolyse de l'ATP permettant la fermeture du canal (Aleksandrov *et al.*, 2002a, b).

(2) Régulateur de l'activité d'autres canaux ioniques

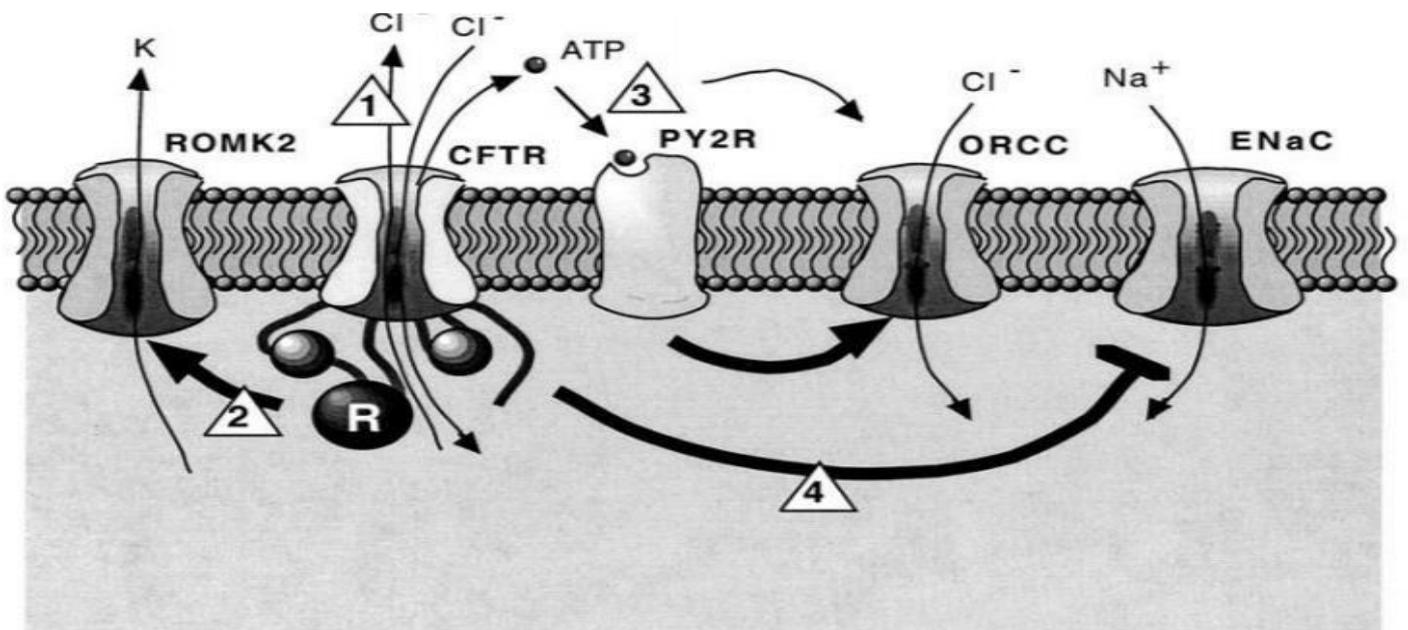


Figure 10 : Régulation des courants ioniques transmembranaires par la protéine CFTR

(Lyczak *et al.*, 2002)

La protéine CFTR régule également le transport d'autres ions via plusieurs mécanismes, d'où son nom de « régulateur transmembranaire de conductance ». (Figure 10).

La liste des protéines avec lesquelles le canal CFTR interagit dans son rôle de régulateur de la conductance continue de croître et comprend des canaux, des transporteurs, et des protéines liées à l'échafaudage du cytosquelette apical des cellules épithéliales. Ces protéines sont prêtes à participer à des fonctions sécrétrices qui sont connues pour être orchestrée par la protéine CFTR. Les enquêtes sur ces interactions ont mis en lumière l'augmentation de l'absorption de Na⁺ par le canal ENaC (Epithelial Na⁺ Channel) (figure 10, triangle n°4). Cette régulation se fait par la liaison de la queue C-terminale de la protéine ENaC intracellulaire domaine cytoplasmique de CFTR (NBD1 et le domaine R) (Stutts *et al.*, 1997 ;

Mall et al., 1998 ; Schreiber et al., 1999a, b ; Lyczak et al., 2002 ; Mall et al., 2004 ; Boucher, 2007).

Il a été démontré aussi que le canal CFTR régule la stabilité des canaux ENaC à la membrane (*Lu et al., 2007*).

Il a été montré que la protéine CFTR régule de manière positive, via les protéines kinases, la conductance des canaux chlorures *ORCC (Outwardly Rectifying Chloride Channel)* (*Egan et al., 1992*) et *CIC-3 (Gentsch et al., 2003)*, et diminue l'activité des canaux chlorures *CaCC (Ca²⁺- activated Chloride Channel)* via son domaine R (*Wei et al., 2001*).

Le canal chlorure rectifiant sortant ou l'*ORCC* est un canal chlorure régulé aussi par l'AMPc. La protéine CFTR est indispensable pour l'activation de ce canal (**figure 10**, triangle n°3). La libération d'ATP intracellulaire par la protéine CFTR activée permettrait, en effet, l'activation des canaux *ORCC (Schwiebert et al., 1999)*.

Le canal CFTR joue un rôle important dans la sécrétion d'ions bicarbonates (HCO_3) (*Hogan et al., 1997 ; Coakley et al., 2003 ; Krouse et al., 2004*).

en permettant leur transport (*Devor et al., 1999 ; Ishiguro et al., 2009*) mais aussi en stimulant probablement le canal échangeur $\text{Cl}^-/\text{HCO}_3^-$ (*Choi et al., 2001*) ou en inhibant le co-transporteur *NBC-1* ou $\text{Na}^+/\text{HCO}_3^-$ (*Park et al., 2002*).

Le domaine R ou domaine régulateur du canal CFTR permet, aussi, une régulation du canal potassique *ROMK2* (**figure 10**, triangle 2). De même, la protéine CFTR confère une sensibilité sulfonylurée sur le canal K^+ de rectification, vers l'intérieur du rein, *ROMK2*, par interaction avec le *NBD1* et R domaines, ainsi que le premier domaine transmembranaire, de CFTR. La protéine CFTR semble jouer un rôle dans le transport de l'eau à travers la membrane. En effet, une interaction entre la protéine CFTR et les aquaporines, des canaux perméables à l'eau, a été mise en évidence au niveau de la cornée (*Hamann, 2002*), du pancréas (*Burghardt et al., 2003*).

Récemment, les travaux de Pietrement C ont montré que l'interaction entre l'extrémité C-terminal de la protéine CFTR, le motif PDZ de la protéine *NHERF1* et l'aquaporine 9, facilite l'activation de l'aquaporine par l'AMPc (*Pietrement et al., 2008*).

Il existe d'autres canaux chlorures à la surface de la membrane apicale comme le canal

chlorure dépendant du calcium (CaCC). Ce dernier pourrait être influencé par la protéine CFTR via le récepteur PY2R (**figure 10**, triangle 3) qui nécessite une molécule d'ATP. Le CaCC pourrait aussi jouer un rôle dans la régulation du pH du liquide de surface de l'épithélium et un rôle de compensation dans la sécrétion d'ions chlorures chez la souris invalidée ou Knock out pour CFTR (*Nilius and Droogmans, 2003*).

La protéine CFTR est non seulement capable de réguler directement ou indirectement d'autres canaux ioniques, mais peut aussi avoir des effets modulateurs agonistes et antagonistes sur d'autres protéines. L'une de ces protéines est le canal K⁺ sensible à l'ATP, qui fait partie d'un sousgroupe de la superfamille des canaux K⁺ rectifiant entrants (KIR). Les canaux KIR sont localisés sur la membrane basolatérale des cellules épithéliales respiratoires, où ils interviennent dans le recyclage du potassium (*Loussouarn et al.1996*).

(3) Les autres fonctions de la protéine CFTR

D'autres fonctions pour CFTR ont été décrites comme

- le transport d'ATP.
- la modulation des phénomènes d'exocytose et d'endocytose.
- la régulation du pH des organelles intracellulaires.

Le canal CFTR est non seulement perméable aux ions chlorures, mais également au glutathion réduit (GSH) (*Linsdell and Hanrahan, 1998 ; Kogan et al., 2003*).

Différentes études ont émis l'hypothèse que la protéine CFTR aurait un rôle dans la défense anti-infectieuse via son domaine extracellulaire qui servirait de récepteur pour les bactéries *Pseudomonas aeruginosa* et *Salmonella enterica* (*Pier et al., 1998 ; Lyczak and Pier, 2002*). Ces travaux n'ont pas été confirmés. Une transactivation du facteur de transcription nucléaire NF- κ B résultant du dysfonctionnement de CFTR (*Tabary et al., 2001*) et l'observation de taux élevés d'IL-8 dans les voies respiratoires (*Vij et al., 2006 ; Perez et al., 2007*) ont suggéré un rôle de la protéine CFTR dans la réponse inflammatoire.

Les travaux de Boncoeur E. et al. ont montré que la protéine CFTR contrôle la dégradation et l'activité transcriptionnelle de NF- κ B en réponse à un stress oxydatif (*Boncoeur et al., 2008*).

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

- ❖ **Les poumons** : la mucoviscidose provoque un épaissement 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*)

Une étude plus récente vient de montrer que la surexpression de la protéine CFTR induit une répression de l'activité transcriptionnelle basale des promoteurs IL-8 et NF- κ B et de celle qui est induite par la cytokine IL-1 β (*Vij et al., 2009*).

Le rôle joué par la protéine CFTR dans la réaction inflammatoire, qui semble agir de façon très précoce, reste encore méconnu. La protéine CFTR a également un rôle dans le développement du tissu pulmonaire, non seulement dans la différenciation des cellules bronchiques, mais aussi dans leur prolifération (*Larson and Cohen, 2006*).

La protéine CFTR influencerait la contraction musculaire via un mécanisme indépendant des Rho-kinases qui implique l'ATP et le calcium (*Larson and Cohen, 2006*).

Dans une étude plus récente, les mêmes auteurs ont été montrés que la protéine CFTR joue un rôle dans l'organogenèse pulmonaire en modulant l'expression du peptide relié à l'hormone parathyroïdienne (PTH-rP) et la voie de signalisation Wnt/ β -caténine (*Cohen et al., 2008*).

D'autres travaux ont montré que l'homéostasie du calcium est affectée dans les cellules des muscles lisses des voies respiratoires chez les patients CF, suggérant un rôle de la protéine CFTR dans la régulation du calcium intracellulaire (*Michoud et al., 2009*).

Au niveau cardiaque, la protéine CFTR semble moduler la durée du potentiel d'action (*Tominaga et al., 1994*).

De plus, la stimulation des récepteurs β_3 diminue le potentiel d'action dans les myocytes de patients normaux et n'a pas d'effets chez les patients CF (*Leblais et al., 1999*).

Ces résultats suggèrent un rôle de la protéine CFTR dans la contraction musculaire cardiaque. La protéine CFTR intervient également dans le recyclage des membranes en favorisant les phénomènes d'exocytose et en inhibant ceux d'endocytose (*Ameen et al, 2007*).

L'absence de protéine CFTR normale pourrait ainsi être associée à une diminution de l'expression membranaire d'autres canaux. La protéine CFTR régule aussi le pH intracellulaire et acidifie les compartiments intracellulaires. Son dysfonctionnement a pour conséquence de modifier les sécrétions cellulaires et d'augmenter la viscosité du mucus. On observe, en particulier, un défaut de sialylation des mucines, ce qui ferait apparaître des récepteurs spécifiques pour *P.aeruginosa*. (*Ameen et al, 2007*).

3.3 Les mutations de la protéine CFTR

Des auteurs ont tenté de déterminer l'âge et le lieu d'origine des principales mutations responsables de la mucoviscidose par analyse génétique de différentes populations.

Près de 1 900 mutations du gène CFTR sont décrites actuellement. Les plus fréquentes sont les mutations faux-sens (40%), suivies des mutations par décalage du cadre de lecture (frameshift-16%), des mutations sur le site d'épissage (12%), des mutations non-sens (8%). En fonction de leurs répercussions, on parlera de mutations sévères ou modérées. (*Lubamba et al., 2012*).

Ces mutations peuvent être regroupées en 6 classes selon leurs conséquences fonctionnelles, entraînant une grande variabilité des manifestations cliniques et influençant l'évolution de la maladie :

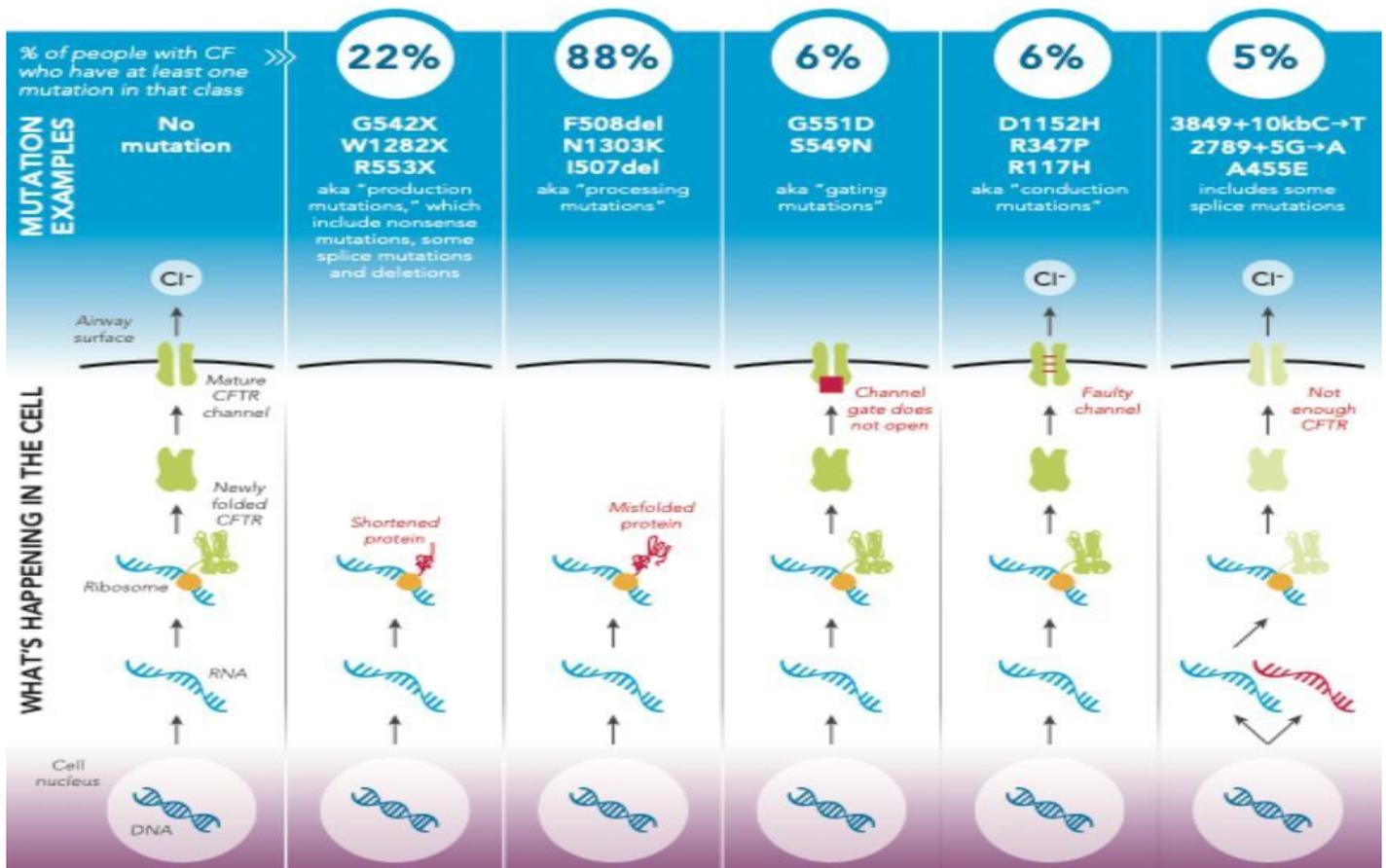


Figure 11. Classes de mutations de CFTR impliquées dans la mucoviscidose : pathogénie, fréquence et principales mutations. *Source : Cystic Fibrosis Foundation (2017).*

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 12-13*). (Sediki.2015).

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

<i>Classe du Mutation</i>	<i>caractérisations</i>
<i>classe 1</i>	<p>mutations altérant la production de la protéine. Ces mutations résultent en une perte de fonction, liée à un défaut de synthèse. Il n'y a pas de transcription du gène en acide ribonucléique messager (ARNm) stable. Sur le plan fonctionnel, ces mutants conduisent à une perte de la conductance au Cl⁻ du canal CFTR dans les épithéliums atteints.</p> <p>(Ouchtati A, Dali H, 2018)</p>
<i>classe 2</i>	<p>mutations perturbant le processus de maturation cellulaire de la protéine. La protéine mutante n'est donc pas adressée à la membrane, en général elle reste localisée dans le cytoplasme. Ainsi, la protéine est soit absente, soit présente en quantité réduite dans la membrane apicale. Les mutations de cette classe représentent la majorité des anomalies observée, notamment la mutation p.Phe508del ou F508del. Elle correspond à la délétion de 3 nucléotides au niveau du 10ème exon du gène, aboutissant à l'élimination d'une phénylalanine en position 508 de la protéine.</p> <p>(Ouchtati A, Dali H, 2018)</p>
<i>classe 3</i>	<p>mutations perturbant la régulation du canal chlore, les protéines mutées sont présentes à la membrane apicale, mais ne peuvent pas être activés ou ont une fonction de canal Cl⁻ anormale. (Ouchtati A, Dali H, 2018)</p>
<i>classe 4</i>	<p>mutations altérant la conduction et les</p>

	<p>mécanismes d'ouverture et de fermeture du canal Cl⁻. Les caractéristiques de ces canaux sont différentes de celles du canal CFTR endogène avec une diminution du flux d'ions et une sélectivité modifiée. Les mutations R117H appartiennent à cette classe (<i>Ouchtati A, Dali H, 2018</i>)</p>
<i>classe 5</i>	<p>mutations altérant la stabilité de l'ARNm CFTR. Ces mutations influencent la quantité d'ARNm, ainsi que la quantité de protéines fonctionnelles. (<i>Ouchtati A, Dali H, 2018</i>)</p>
<i>Classe6</i>	<p>mutations altérant la stabilité de la protéine mature et donc perturbant la régulation des autres canaux par CFTR 8. Ce sont principalement des mutations générant des protéines tronquées dans leur partie C-terminale comme Q1412X et 4326delTC. Les autres mutations retrouvées dans la mucoviscidose sont légèrement plus récentes. La détermination, par analyse génétique de l'origine géographique des mutations responsables de la mucoviscidose, ne semble pas envisageable compte tenu du fait que ces mutations existaient probablement chez les ancêtres des Européens, avant même la genèse des populations actuelles; et que les modifications successives de ces populations ont probablement effacé les traces de l'origine géographique de ces mutations génétiques. (<i>Ouchtati A, Dali H, 2018</i>)</p>

La nomenclature traditionnelle concernant le gène CFTR, anciennement utilisée, les exons sont numérotés de 1 à 24 (*Ouchtati A, Dali H, 2018*)

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

Classe de la mutation	nomenclature
Classe1	G542X W1282X G551D 621+1G→T
Classe2	F508del N1303K I507del R560T
Classe3	G551D - - S549N
Classe4	R117H R347P R117C R334W
Classe5	3849+10kbC→T 2789+5G→A 3120+1G→A
Classe6	4326delTC - 4279insA

Chapitre IV : OUTILS BIO-INFORMATIQUE :

4.1. La bio-informatique

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)**

Définition

- La bioinformation est l'information liée aux molécules biologiques : leur séquence, leur nombre, leur structure, leur fonction, leurs liens de "parenté", leurs interactions et leur intégration dans la cellule...
- Cette bioinformation est issue de diverses disciplines : la biochimie, la génétique, la génomique structurale, la génomique fonctionnelle, la transcriptomique, la protéomique, la biologie structurale (structure spatiale des molécules biologiques, modélisation moléculaire ...)
- Une définition de la bioinformatique : analyse de la bioinformation par des moyens informatiques.

Démarche

1. Compilation et organisation des données biologiques dans des bases de données (spécialistes ou généralistes).

2. Traitements systématiques des données : l'un des objectifs est de repérer et de caractériser une fonction et/ou une structure biologique importante. Les résultats de ces traitements constituent de nouvelles données biologiques obtenues "in silico".

3. Elaboration de stratégies :
 - ✓ apporter des connaissances biologiques supplémentaires en combinant les données biologiques initiales et les données biologiques obtenues "in silico".
 - ✓ ces connaissances permettent, à leur tour, de développer de nouveaux concepts en biologie.
 - ✓ concepts qui, pour être validés, peuvent nécessiter le développement de nouvelles théories et outils en mathématiques et en informatique.

4.2. Quelques Outils d'étude bio-informatique 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 :

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.

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**)



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).**



Partie 02
pratique

[matériel et méthodes]

Chapitre V

I. Matériels

Pour l'étude in silico les mutations de gène CFTR on est besoin du matériel suivant :

- PC
- Connexion internet

Et pour les outils bio-informatiques :

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 voire 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

Une 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 IV*)

Tableau IV: représentation des 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..17668665	NC_000007.14	-----	-----
ARNm CFTR	1...6132	NM_000492.3	-----	-----
Exon10	1525...1716	NM_000492.3	c.1521_1523delCTT	rs113993960
Exon6	712...875	NM_000492.3	c.680T>G	-----
Exon12	1717...1811	NM_000492.3	c.1624G>T	rs113993959
Exon24	4006...4095	NM_000492.3	c.3909C>G	rs 80034486
Exon12	1717...1811	NM_000492.3	c.1652G>A	rs75527207

II. Méthodes

1/utilisation de BLAST (Basic Local Alignment Search Tools)

Est un outil accessible en ligne, utilisé en bio-informatique pour rechercher des régions similaires entre deux ou plusieurs séquences nucléotidiques ou protéiques (figure14).

- Se rendre à l'adresse suivante : <https://blast.ncbi.nlm.nih.gov/Blast.cgi> .
- Choisir le type de blast selon le besoin (BLASTn et le BLASTp.)
- cliquer sur « Nucléotide Blast »
- On copie/colle la séquence recherchée dans le champ en haut de la page
- Et on clique enfin sur le bouton « BLAST ».
- Lancer l'algorithme d'alignement.
- Il faut patienter quelques secondes avant d'avoir le résultat.

Home Recent Results Save

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

BLAST+ 2.10.0 is released - Improved Composition-based statistics.

We have updated the BLAST process to improve the stability of BLAST results against changes in the number of results requested.

Mon, 23 Dec 2019 16:00:00 EST [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide → nucleotide

blastx
translated nucleotide → protein

tblastn
protein → translated nucleotide

Protein BLAST
protein → protein

Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

BLAST parameters search nucleotide databases using a nucleotide query. [View help](#) [Submit](#)

Enter Query Sequence

Enter accession number(s), g(x), or FASTA sequence(s)

OR upload file

Job Title

Align two or more sequences

Choose Search Set

Database Standard databases (nr, etc.) RefSeq databases Genomic + transcript databases

Organism

Exclude BLAST (SMPH) Uncultured environmental sample sequences

Limit to Sequences from type material

Enter query

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search database Nucleotide collection (nr/nt) using Blastn (Optimize for somewhat similar sequences)

Show results in a new window

Figure 14: Utilisation de BLAST sur la plate-forme NCBI. (Personnelle, 2022)

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 (**figure 16**).

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.

The screenshot shows the SMS ORF Finder web interface. On the left is a navigation menu with categories like 'Format Conversion' and 'Sequence Analysis'. The main content area has a title 'Sequence Manipulation Suite: ORF Finder' and a description of the tool. Below the description is a text input field containing a DNA sequence, with a red '1' pointing to it. Underneath the input field are 'Submit', 'Clear', and 'Reset' buttons. Below the buttons are configuration options: 'ORFs can begin with: any codon' (with a dropdown arrow and a red '2'), 'Search for ORFs in reading frame 1 on the direct strand.' (with dropdown arrows and a red '3'), 'Only return ORFs that are at least 30 codons long.' (with a text input field and a red '4'), and 'Use the standard (1) genetic code.' (with a dropdown arrow and a red '4'). At the bottom left, there is a red '6' pointing to the 'Submit' button. At the bottom right, there is a red '5' pointing to the footer links 'new window | home | citation'. The footer also includes the date 'Mon Nov 8 02:58:29 2017' and 'Valid XHTML 1.0, Valid CSS'.

Figure 16: Utilisation d'ORF Finder (Personnelle, 2022)

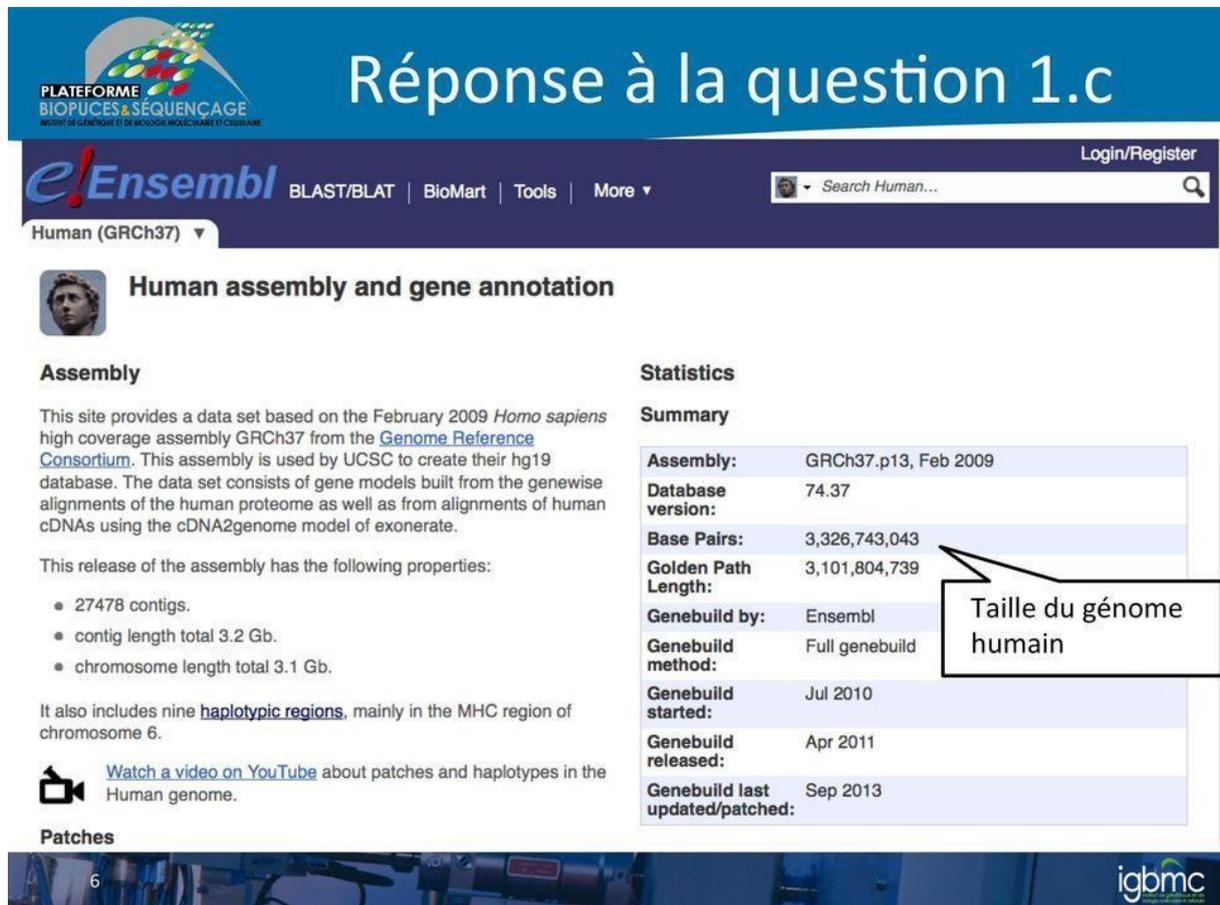
4. Utilisation d'Ensembl

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 17)



The screenshot shows the Ensembl website interface. At the top, there is a blue header with the text "Réponse à la question 1.c" and the logo for "PLATEFORME BIOPUCES & SÉQUENCAGE". Below the header, the Ensembl logo is visible along with navigation links for "BLAST/BLAT", "BioMart", "Tools", and "More". A search bar is present with the text "Search Human...". The main content area is titled "Human assembly and gene annotation" and includes a section for "Assembly" with a description of the GRCh37 assembly and a list of properties: 27478 contigs, total contig length of 3.2 Gb, and total chromosome length of 3.1 Gb. To the right, a "Statistics" section provides a summary of the assembly, including the assembly name (GRCh37.p13, Feb 2009), database version (74.37), base pairs (3,326,743,043), golden path length (3,101,804,739), and genebuild details (started in Jul 2010, released in Apr 2011, last updated in Sep 2013). A callout box points to the "Base Pairs" value with the text "Taille du génome humain".

Summary	
Assembly:	GRCh37.p13, Feb 2009
Database version:	74.37
Base Pairs:	3,326,743,043
Golden Path Length:	3,101,804,739
Genebuild by:	Ensembl
Genebuild method:	Full genebuild
Genebuild started:	Jul 2010
Genebuild released:	Apr 2011
Genebuild last updated/patched:	Sep 2013

Figure 17: Utilisation d'Ensembl. (Personnelle, 2022)

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 18)**.

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

SWISS-MODEL Repository

[Workspace] [Repository] [Modelling] [Tools]

SWISS-MODEL Repository Model Details

Model Overview [+/-]
Click on the bars to get more details about individual Models or experimental structures

1 737

Sequence [+/-]
P53354 Alpha-amylase I (1,4-alpha-D-glucan glucanohydrolase)
Aedes aegypti (Yellowfever mosquito).

Domain [+/-]
Link to: [InterPro]

Alpha-amylase_C
Alpha-amylase

Model 3D Structure [+/-]
Based on template: 1clv [SMTL] [PDB] [SCOP] [CATH]
Sequence identity: 49%
Residue range: 261 to 736
[display] [download]

Alignment [+/-]

TARGET	261	QPFFFP	GHSGIVHLFE	WKFSDIAREC	EN
1clvA	2	xk--danfas	grnsivhlfe	wkwndiadc	er
TARGET		sssss	hhhhhhh	h	
1clvA		sssss	hhhhhhh	h	
TARGET	307	LVSPSRAWWE	RYQPISFEIK	SRSGNERQFS	DM
1clvA	49	lvadgrpwe	ryqpvsiin	trsgdesaft	dm

Template Selection [+/-]
Template Search [+/-] [start new template identification in Workspace]
Quality [+/-] [start new quality assessment in Workspace]

Model Quality Assessment
Anolea

Annotations on the right:
 - Overview of available models for target protein
 - UniProt annotation (via REST)
 - PFAM domain annotation (via InterPro / DAS)
 - In page visualization (using ASTEX plugin)
 - Tools integration with SWISS-MODEL Workspace
 - Model Quality Estimation

Figure 18: Utilisation de swiss-model. (Personnelle ,2022)

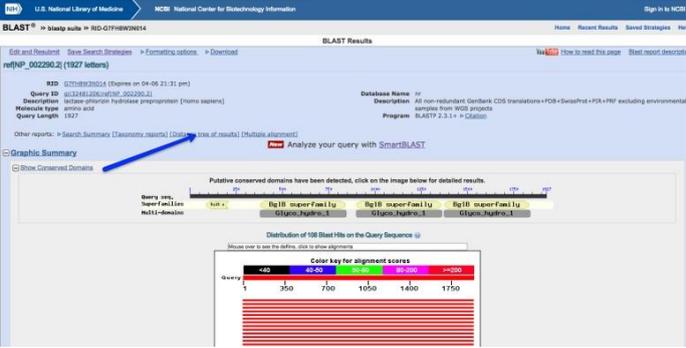
Chapitre VI

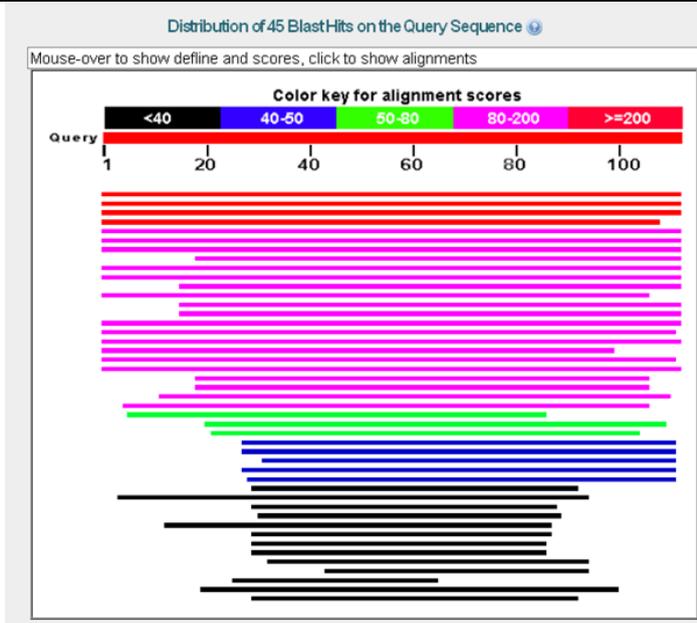
Résultats et Discussions

1. Les résultats de Blast n de l'ARNm CFTR :

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

Tableau V : Résultats de BLAST de gène CFTR

RESULTATS	DESCRIPTIFS	ROLES
	<ul style="list-style-type: none">▪ Séquence soumise▪ Identifiant▪ Longueur▪ Type▪ Banque de donnée interrogée▪ programme utilisé	Récapitulatif de la requête.



- les traits de couleur représentent à un alignement entre la séquence soumise et une séquence de la banque.
- la couleur correspond au score.
- La longueur correspond à la taille de l'alignement

Représentation graphique des résultats

Alignments
 Select All [Get selected sequences](#) [Distance tree of results](#) [Multiple alignment](#)

```

->  ref|NP\_001823.1 U G M colipase isoform 1 preproprotein [Homo sapiens]
ref|XP\_527612.1 U G M PREDICTED: colipase isoform 3 [Pan troglodytes]
ref|XP\_002311312.1 U G M PREDICTED: colipase [Pan troglodytes]
length=112

GENE ID: 1208 CLPS | colipase, pancreatic [Homo sapiens] (Over 10 PubMed links)
Score = 228 bits (582), Expect = 2e-75, Method: Compositional matrix adjust.
Identities = 112/112 (100%), Positives = 112/112 (100%), Gaps = 0/112 (0%)
Query 1  HEKILLLLVALSVAYAAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Sbjct 1  HEKILLLLVALSVAYAAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Query 61  SENSECSVRTLYGIYKPCERGLTCRCDKTIUGSITNTNFGICHDAGRSKQ 112
Sbjct 61  SENSECSVRTLYGIYKPCERGLTCRCDKTIUGSITNTNFGICHDAGRSKQ 112

->  ref|XP\_00278909.1 G M PREDICTED: colipase-like [Nomascus leucogenys]
length=112

GENE ID: 100593647 LOC100593647 | colipase-like [Nomascus leucogenys]
Score = 225 bits (574), Expect = 3e-74, Method: Compositional matrix adjust.
Identities = 109/112 (97%), Positives = 112/112 (100%), Gaps = 0/112 (0%)
Query 1  HEKILLLLVALSVAYAAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Sbjct 1  HEKILLLLVALSVAYAAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Query 61  SENSECSVRTLYGIYKPCERGLTCRCDKTIUGSITNTNFGICHDAGRSKQ 112
Sbjct 61  SENTRCSVRTLYGIYKPCERGLTCRCDKTIUGSITNTNFGICHDAGRSKQ 112

->  ref|XP\_002816946.1 G M PREDICTED: colipase-like [Pongo abelii]
length=112

GENE ID: 100448898 LOC100448898 | colipase-like [Pongo abelii]
Score = 223 bits (568), Expect = 2e-73, Method: Compositional matrix adjust.
Identities = 110/112 (98%), Positives = 111/112 (99%), Gaps = 0/112 (0%)
Query 1  HEKILLLLVALSVAYAAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Sbjct 1  HEKILLLLVALSVA AAPGPGCIINLENGELCHNSAQCSNCCQHSSALGLARCTSHA 60
Query 61  SENSECSVRTLYGIYKPCERGLTCRCDKTIUGSITNTNFGICHDAGRSKQ 112
Sbjct 61  SENSECSVRTLYGIYKPCERGLTCRCDKTIUGAITNTNFGICHDAGRSKQ 112

```

- Requête : la séquence soumise.
- Sujet : la séquence trouvée dans la base de données.
- Les grandeurs propre à chaque alignement (score, E-valu, identité, positif, de gap).

Description des résultats, des alignements

2. 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 19**).

The Nucleotide database will include EST and GSS sequences in early 2019. [Read more](#)

FASTA

Send to:

Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA

NCBI Reference Sequence: NM_000492.3
[GenBank](#) [Graphics](#)

>NM_000492.3 Homo sapiens CF transmembrane conductance regulator (CFTR), mRNA
 AATTGGAAGCAATGACATCACAGCAGGTGAGAGAAAAGGGTTGAGCGGACAGCCAGAGTAGTAGG
 TCTTTGGCATTAGGAGCTTGAGCCAGACGGCCCTAGCAGGGACCCAGCGCCGAGAGACCATGCAGAG
 GTCGCCTCTGGAAAAGGSCAGCGTGTCTCCAACTTTTTTCAGCTGGACAGACCAATTTGAGGAAA
 GGATACAGACAGCGCCGGAATGTCAGACATATACCAATCCCTCTGTTGATTCTGCTGACAATCTAT
 CTGAAAAATGGAAAAGAAATGGGATAGAGAGCTGGCTTCAAGAAAAATCCTAACTCATTAAATGCCT
 TCGCGATGTTTTCTGGAGATTTATGTTCTATGGAACTTTTTATATTTAGGGAAAGTCAACAAAGCA
 GTACAGCTCTCTTACTGGGAAGAAATCATAGCTTCTATGACCCGGATAACAAGGAGGAAACGCTCTATCG
 CGATTTATCATGAGCATAGGCTTATGCTTCTCTTTATGTTGAGGACAGCTCTCTACACCCAGCCATTT
 TGCCCTTCATCACATTGGAATGAGATGAGAGATGCTATGTTAGTTGATTTATAAGAAAGCTTTAAAG
 CTGCAGCGGTGTTCTAGATAAATAAGTATGGACAACCTGTTAGTCTCTTTCCAAACCACTGAACA
 AATTTGATGAAGGACTTGCATGGCACATTTGTTGGATCGCTCCTTTGCAAGTGGCACTCTCATGGG
 GCTAATCTGGAGTTGTTACAGGCGTCTGCTCTCTGGACTTGGTTTCTGATAGTCTTGCCTTTTT
 CAGGCTGGCTAGGGAGAAATGATGAAAGTACAGAGATCAGAGAGCTGGGAAGATCAGTGAAGACTTG
 TGATTACCTCAGAAATGATTGAAATATCCAATCTGTTAAGGCATCTGCTGGGAAGAAAGCAATGGAAAA
 AATGATTGAAAACCTTAAGCAACACAGAACTGAACTGACTCGGAAGGACGCTATGAGAGACTTCAAT
 AGCTCAGCCTCTCTCTCAGGGTTCTTGGGTTGTTTATCTGTTGCTTCTATGCACTAATCAAAG
 GAATCATCTCCGAAAAATATCACCACTCTCATTCTGCATTGTTCTGCGCATGGCGGTCACTCGGCA
 ATTTCCCTGGGCTGTACAACAATGGTATGACTCTCTGGAGCAATAAACAATAACAGGATTTCTACAA
 AAGCAAGAAATAAAGCAATTGGAATAAATAAAGCACTACAGAGTATGATGGAGAAATGAACAGCT
 TCTGGGAGGAGGAAATTTGGGAAATTTGAGAAAGCAAAACAATAAACAATAAAGAAACTCTAA
 TGGTGATGACAGCTCTCTCAGTAATTTCTCACTCTTGGTACTCCTGTCTGAAAGATATAATTTCT
 AAGATAGAAAGGACAGTTGTTGGGCTTCTGGAATCCACTGGAGCAGGCAAGACTCACTCTAATGG
 TGATTATGGGAGAACTGGAGCTTCAAGGGTAAAAATAAGCACAGTGGAAAGATTTCACTCTGTTCTCA
 GTTTCTCGGATATGCTGGCACCATTAAAGAAAAATCATCTTTGGTGTTCCTATGATGAATATAGA
 TACAGAAAGCGTCATCAAGCATGCCAAGTGAAGAGGACATCTCAAGTTTGCAGAGAAAGCAATATAG
 TTCTTGGAGAAAGTGGAAATCACACTGAGTGGAGGCAACGAGCAAGAAATTTTTCAGCAAGAGCAGTATA
 CAAAGATGCTGTTGATTTATAGACTCTCCTTTGGATACCTAGATGTTTAAACAGAAAAAGAAATA
 TTTGAAAGCTGTGTGTAACCTGATGGCTAACCAAACTAGGATTTTGGTCACTCTAAAATGGAACTT
 TAAAGAAAGCTGACAAAAATTAATTTGCATGAAAGTAGCAGCTATTTTATGGGACATTTGAGAACT
 CCAAAATCTACAGCAGACTTAGCTCAAACTCATGGATGTTGTTCTTTCGACCAATTTAGTCAGAAA

Analyze this sequence
 Run BLAST
 Pick Primers
 Highlight Sequence Features
 Find in this Sequence
 Show in Genome Data Viewer

Articles about the CFTR gene
 Restoration of F508-del Function by Transcomplementati [Cell Physiol Biochem. 2019]
 Mutational analysis of CFTR in the Ecuadorian population using next-generation se [Gene. 2019]
 Genetic and phenotypic traits of children and adolescents with cystic fit [J Bras Pneumol. 2018]

Pathways for the CFTR gene
 Ub-specific processing proteases
 Deubiquitination
 Cargo recognition for clathrin-mediated endocytosis

Figure 19: Séquence d'ARNm du gène CFTR sous format FASTA. (Djijli,I, 2019)

La figure 19 nous donne la description de la séquence ARN du gène répertorié dans NCBI, ainsi que les différents identifiants. (Voir tableau V)

3. Les résultats de swiss-prot d'ARNm du gène CFTR ID de CFTR sur swiss-model : P13569

a. Structure tridimensionnelle de la protéine CFTR

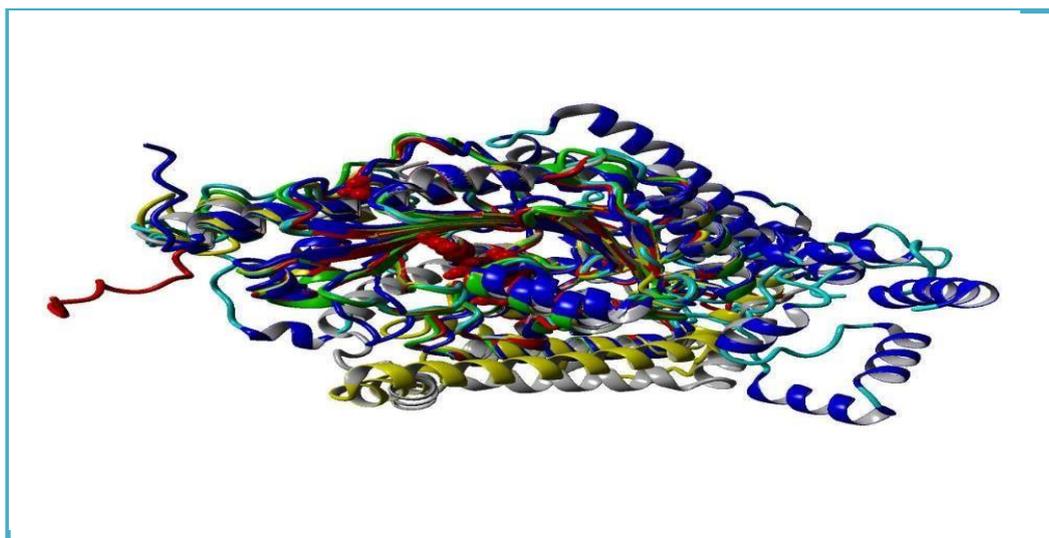


Figure 20: Structure 3D de la protéine CFTR. (Personnelle, 2022)

b. La séquence protéique du gène CFTR

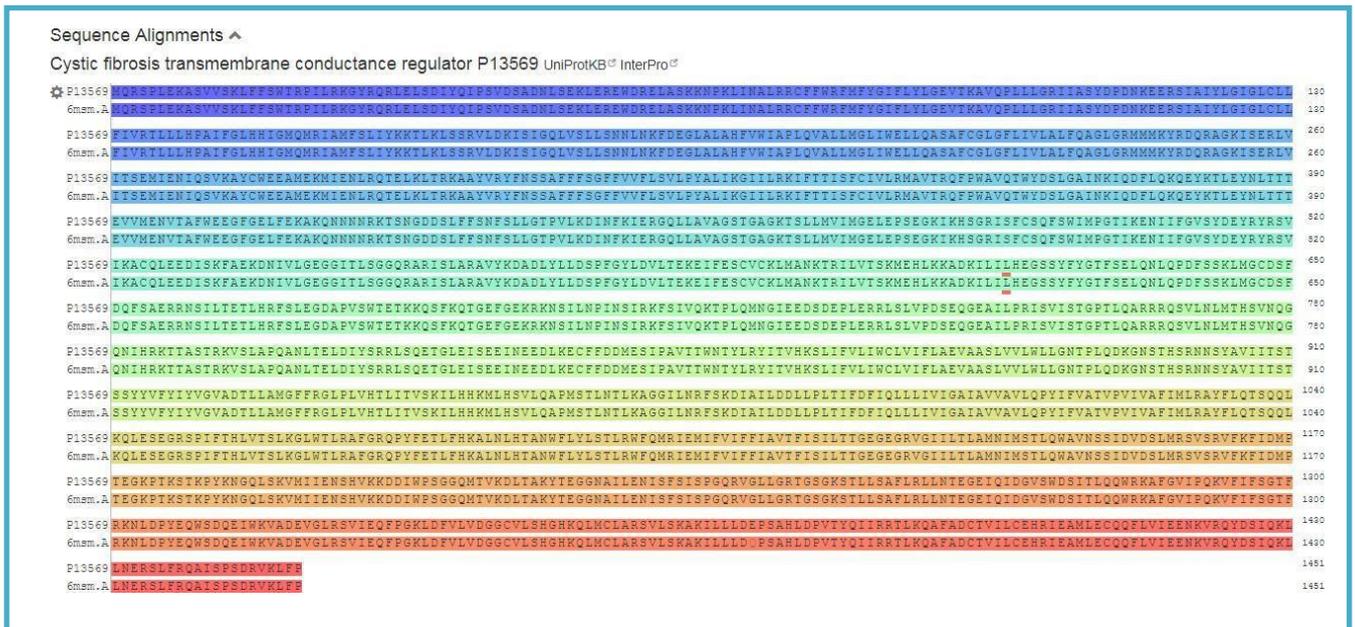


Figure 21: Séquence protéique du gène CFTR (Dijli,I, 2019).

- ❖ Chaque couleur présent dans la séquence protéique (**Figure 21**) représente la même couleur dans la structure 3D (**Figure 20**).
- ❖ Pour savoir la position d'un n'importe quel acide amine sur la structure Tridimensionnel (**Figure 21**) on clique sur l'acide aminé, il va apparaitra sur la structure. (**Figure 20**).

c. Les résultats dans ORF Finder de la mutation c.1521_1523deCTT (F508del)

c.1. 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 (**figure22**).

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  
AGTGAAGAATTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA  
AATATCATCTTTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGC  
CAACTAGAAGAG
```

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

Figure 22: Traduction de la séquence non mutée de l'exon 10. (Personnelle, 2022)

c. 2. 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 (**figure23**).

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  
AGTGAAGAATTCATTCTGTTCTCAGTTTTCTGGATTATGCCTGGCACCATTAAAGAA  
AATATCATTGGTGTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAA  
CTAGAAGAG
```

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

figure23 : Traduction de la séquence mutée de l'exon 10. (Personnelle, 2022)

4. Les résultats d'Ensembl

Les résultats d'ensembl présente dans la **figure24**.

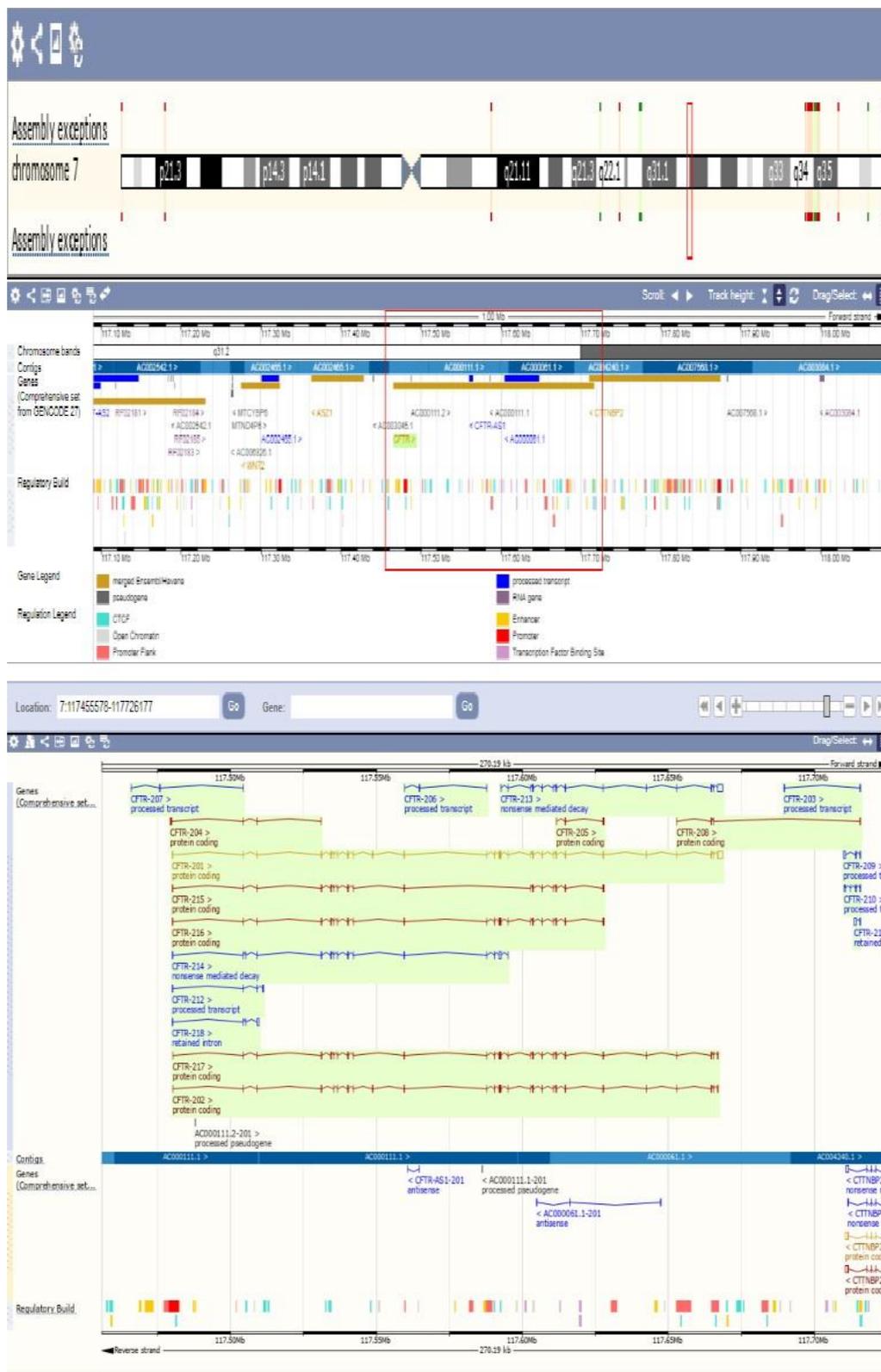


Figure24 : représentation des résultats d'ensembl (Djijli.i2019).

Ensembl donne une description :

- d'image chromosomique
- des régions de chromosome.
- du numéro de chromosome.
- De la région d'intérêt (rectangle en rouge).
- Les bandes de chromosome.
- La séquence génomique.
- La région d'intérêt.
- La position de gène
- La transcription de la région d'intérêt :
 - les traits correspondent aux exons.
 - les lignes correspondent aux introns.

5. Les résultats d'ONIM

ONIM donne une 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.

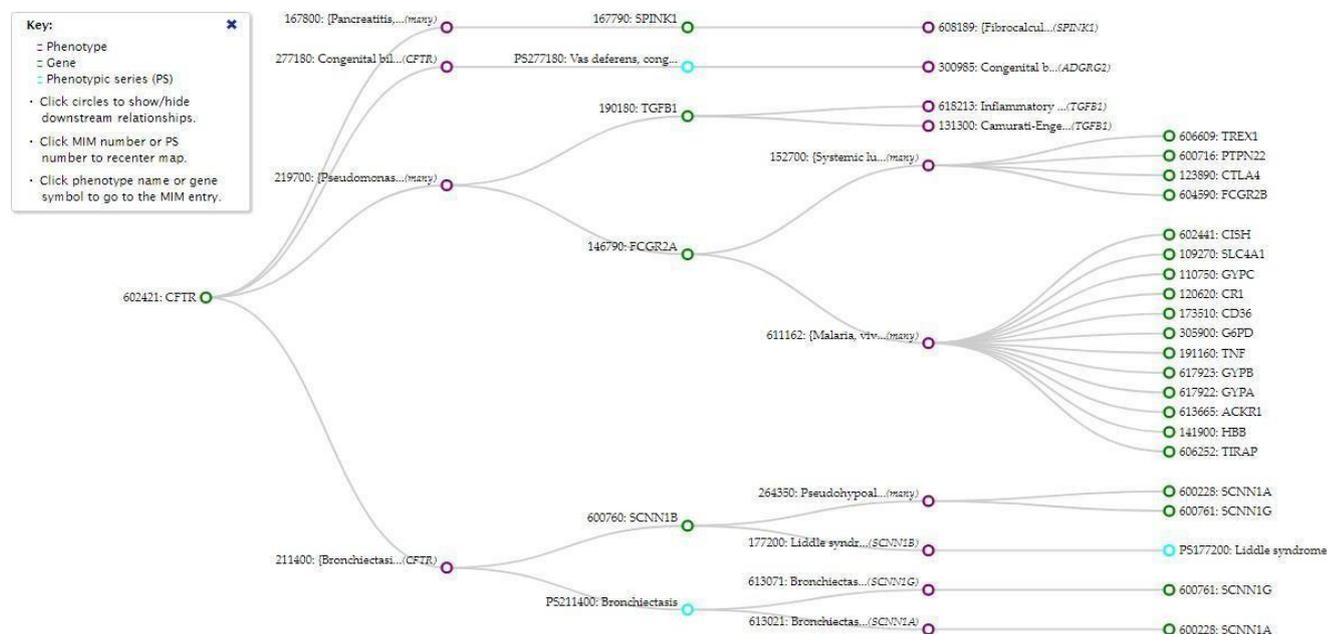


Figure25 : Représentation graphique linéaire de la relation phénotype génotype de la mucoviscidose. (Personnelle, 2019)

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

Conclusion

Le gène CFTR a été cloné en septembre 1989. Très rapidement la mutation la plus fréquente du gène a été identifiée – la délétion F508del – et aujourd’hui plus de 1800 mutations ont été rapportées dans ce gène. Le développement des techniques de biologie moléculaire nous permet désormais en quelques jours d’identifier l’une de ces 1800 mutations et de confirmer ou d’infirmier dans les meilleurs délais un diagnostic de mucoviscidose chez une personne suspectée d’être porteuse de la maladie. Nous pouvons également écarter, chez les conjoints de sujets hétérozygotes, la probabilité d’être porteurs et modifier de façon considérable cette probabilité (si on a pu écarter 95 % des mutations, le risque d’être porteur passe de 1/30 à 1/600).

La connaissance du gène, de la protéine et la meilleure compréhension de la physiopathologie de la maladie, permettent d’entrevoir le développement de thérapies spécifiques des dysfonctionnements de la protéine CFTR par la prise en charge du diagnostic moléculaire de la mucoviscidose.

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.

Référence bibliographique

1. 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.
2. Djilji.I, Étude in silico des mutations du gène CFTR provoquant la variabilité des formes de mucoviscidose, mémoire de master, spécialité génétique, Université saad dahleb blida, Faculté des Sciences de la Nature et de la Vie, Département de Biologie et physiologie animale
3. Ikpa PT, Bijvelds MJ, de Jonge HR. Cystic fibrosis toward personalized therapies. *Int J Biochem Cell Biol.* 2014;52:192–200. doi: 10.1016/j.biocel.2014.02.008.
4. 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
5. S. Dimassi, M. Tilla, D. Sanlaville, Anomalies chromosomiques ; *Journal de pédiatrie et de puériculture* ; vol 1269 page 1-22 :2017.
6. Oujidi Bachija, thèse de doctorat, impact de la pathologie cardiovasculaire dans la prise en charge des broncho-pneumopathies chronique obstructive au service de pneumo-phtisiologie B du CHUYd'oran université des sciences et de la technologie d'Oran, faculté de médecine, département de médecine, soutenue le24-04-2018.
7. M. Susan Lindee, Genetic disease since 1945, Macmillan Magazines Ltd, vol1, page 236-241.2000.
8. Anderson MP, Gregory RJ, Thompson S, Souza DW, Paul S, Mulligan RC et al. :Demonstration that CFTR is a chloride channel by alteration of its anion selectivity.*Science* 1991;253:202-4.
9. Recommandation : Registre Français de la Mucoviscidose. Rapport 2007.
10. Cystic Fibrosis Genetic Analysis Consortium. Population variation of common.cystic fibrosis mutations. *Hum Mutat* 1994;4:167-77.
11. 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
12. Cutting GR. : Modifier genes in Mendelian disorders: the example of cystic.fibrosis. *Ann NY Acad Sci* 2010;1214:57-69.

13. Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S, Wilde RG, Karp G, Takasugi J, Chen G, Jones S, Ren H, Moon YC, Corson D, Turpoff AA, Campbell JA, Conn MM, Khan A, Almstead NG, Hedrick J, Mollin A, Risher N, Weetall M, Yeh S, Branstrom AA, Colacino JM, Babiak J, Ju WD, Hirawat S, Northcutt VJ, Miller LL, Spatrick P, He F, Kawana M, Feng H, Jacobson A, Peltz SW, Sweeney HL (2007) PTC124 targets genetic disorders caused by nonsense mutations. *Nature* 447: 87-91
14. Tabcharani JA, Rommens JM, Hou YX, Chang XB, Tsui LC, Riordan JR, Hanrahan JW (1993) Multi-ion pore behaviour in the CFTR chloride channel. *Nature* 366: 79-82.
15. Schwarzenberg SJ, Thomas W, Olsen TW, Grover T, Walk D, Milla C, Moran A (2007) Microvascular complications in cystic fibrosis-related diabetes. *Diabetes care* 30: 1056-1061.
16. Stutts MJ, Canessa CM, Olsen JC, Hamrick M, Cohn JA, Rossier BC, Boucher RC (1995) CFTR as a cAMP-dependent regulator of sodium channels. *Science* 269: 847-850.
17. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A (2008) Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Human mutation* 29: 1342-1354
18. Wills PJ, Hall RL, Chan W, Cole PJ (1997) Sodium chloride increases the ciliary transportability of cystic fibrosis and bronchiectasis sputum on the mucus-depleted bovine trachea. *The Journal of clinical investigation* 99: 9-13

Webographies :

https://tel.archives-ouvertes.fr/tel-01344122/file/2015_BONINI_archivage.pdf

<https://pubmed.ncbi.nlm.nih.gov/30726326/>

<http://campus.cerimes.fr/genetique-medicale/enseignement/genetique31/site/html/cours.pdf>

<https://www.sciencedirect.com/science/article/abs/pii/S1762419304001638>

https://utl-landerneau.com/wp-content/uploads/2019/01/2019_01_22_La-Mucoviscidose.pdf

www.genet.sickkids.on.ca/cftr/app

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