



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

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

Du diplôme de Master

Option : Génétique

Thème

**Modélisation in silico de l'implication du gène HBS
dans la drépanocytose avec CRISPR-cas9**

Présenté par

Mr Khenfir Souhil

Mr Houari Abbas

Soutenu le : 27/09/2020

Devant le jury composé de :

M ^{me} Benazouz F.	MAA	USDB1	Présidente
M ^{me} Mokrane A.	MCB	USDB1	Examinatrice
M ^{me} Eddaikra A.	MCB	USDB 1	Promotrice
M ^{me} GuessaibiaN.	MCA	USDB 1	Co-promotrice

Promotion 2019-2020

Remerciements

On remercie Dieu, le tout puissant de nous avoir donné la santé, la volonté, le courage d'achever ce mémoire de fin d'étude. Ces quelques lignes ne suffiront pas à exprimer toute la gratitude et tous les moments que nous avons partagés ensemble, mais parfois un simple merci vaut bien plus que tous les discours du monde ; Ce mémoire nous a permis de rencontrer également des personnes merveilleuses et d'une rare générosité

*Nous exprimons notre profond remerciement et gratitude à notre promotrice **Mme Eddaikra A** ; Maitre de conférences au département de BPC de l'université de Blida 1, pour bien vouloir de diriger ce travail, et pour ses conseils et ses remarques constructives. Nous la remercions particulièrement pour sa disponibilité, ses encouragements. Que dieu la protège.*

*Nous exprimons notre profonde gratitude à **Mme Guessaibia N.** en qualité de co- promotrice et qui nous a accompagnés dans le départ de ce travail.*

*Nos remerciements les plus sincères s'adressent à, **M^{me} Benazouz F.** d'avoir accepté de présider notre jury de soutenance.*

*Nous tenons à remercier **M^{me} Mokrane A.** pour l'honneur qu'elle nous a fait en acceptant d'examiner notre travail.*

Dédicace

Je dédie ce modeste travail

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude ; je t'aime maman.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

*Aux personnes dont j'ai bien aimé la présence dans ce jour, à tous mes frères **Oussama, Ibrahim, yasser**, à ma petite-sœur **Ines**, mon cousin **youcef** je dédie ce travail dont le grand plaisir leurs revient en premier lieu pour leurs conseils, aides, et encouragements.*

*Aux personnes qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés, mes aimables amis : **Yahia, Oussama, Abdelhak, Amine, Tahar, Amar, Lotfi, Hamza.***

ABBES

Dédicace

Je dédie ce mémoire

A ma chère mère,

A mon cher père,

*Ma raison d'être, qui n'ont jamais cessé de m'épauler et
m'encourager depuis toujours, a leur sacrifices pour que
j'atteigne mes objectifs et que je progresse en continu*

A mon frère et mes sœurs

Puisse Dieu vous donner succès et réussite

A la mémoire de mon cher grand père

A toute ma famille

A tous mes amis

A tous ceux qui m'aiment

Et a vous chers lecteurs

Souhil

Listes des figures

	Titre	Page
Figure1	Diversité des séquences DR. Séquences du DR de six CRISPR appartenant à différentes espèces bactériennes. Le palindrome est surligné et la séquence 3' terminale AAC/G est entourée.	5
Figure2	Représentation schématique d'un locus du système CRISPR- Cas.	6
Figure3	Différentes types du système de CRISPR	7
Figure4	Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types I-A à I-F	8
Figure5	Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types II-A à II-C	8
Figure6	Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types III-A et III-B	9
Figure7	Motif adjacent au proto-espaceur ("Protospacer Adjacent Motif " - PAM).	10
Figure8	Schéma de la séquence polypeptidique et de l'organisation du domaine pour le type II-A Protéine Cas9 de <i>S. pyogenes</i> (SpyCas9).	11
Figure9	Mécanisme et les étapes du système CRISPR-Cas.	13
Figure10	Structure du complexe [Cas1 - Cas2] de <i>Escherichia coli</i> lié à une séquence d'ADN proto-espaceur constitué de 4 enzymes Cas1 et 2 enzymes Cas2.	13
Figure11	Système de réparation chez les eucaryotes (non homologous end joining, NHEJ) et (homology directed repair, HDR)	14
Figure12	Représentation schématique des nucléases guidées par l'ARNg	15
Figure13	Circulation des globules rouges HBA normaux et HBS anormaux dans les vaisseaux sanguins	18
Figure14	Structure tridimensionnelle de la molécule d'hémoglobine adult	19
Figure15	Structure du chromosome 11 et la localisation du gène HBB	20

Figure16	Drépanocytose au niveau du génotype	21
Figure17	Drépanocytose au phénotype moléculaire	22
Figure18	Schéma physiopathologique de la drépanocytose	22
Figure19	Méthode in silico de...	29
Figure20	Utilisation de genbank	31
Figure21	Tableau de bord de la requête sur NCBI	31
Figure22	Utilisation du blast sur NCBI	32
Figure23	Utilisation d'orf finder	33
Figure24	Utilisation de SWISS-MODEL	34
Figure25	Création de projet	35
Figure26	Introduire les coordonnées personnelles	36
Figure27	Séquence d'ARNm du gène HBB sous format FASTA.	38
Figure28	Affichage de la liste de tout les variants du gène HBB.	39
Figure29	Alignement de la séquence saine et la séquence muté	39
Figure30	Traduction de la séquence non mutée	40
Figure31	Traduction de la séquence mutée	41
Figure32	Représentation topographique HBB sur la base de données swiss-model.	42
Figure33	Séquence protéique du gène HBB	43
Figure34	Structure 3D de la protéine du gène HBB	43
Figure35	Structure 3D de l'hémoglobine	44
Figure36	Alignement de la séquence protéique du gène HBB et HBS	45
Figure37	Informations supplémentaires du variant (HBS)= E7V=E6V	46
Figure38	Intéraction du gène HBB avec d'autres gènes	47
Figure39	Prédiction du rôle des interactions de la protéine HBB	48

Liste des tableaux

N°	Titre	Page
Tableau I	Différentes espèces et leurs (Protospacer Adjacent Motif - PAM).	10
Tableau II	Echiquiers de Punnett : Dans le cas de deux parents hétérozygotes (AS)	23
Tableau III	Echiquiers de Punnett : Dans le cas d'un parent hétérozygote (AS) et l'autre parent normal (AA)	23
Tableau IV	Echiquiers de Punnett : Dans le cas d'un parent homozygote malade (SS) et l'autre parent est normal (AA)	24
Tableau V	Echiquiers de Punnett : Dans le cas d'un parent homozygote (SS) et l'autre parent est hétérozygote (AS)	24
Tableau VI	Echiquiers de Punnett : Dans le cas de deux parents homozygotes malades (SS):	25
Tableau VII	Séquence de l'exon étudié	27
Tableau VIII	outils bioinformatiques utilisés	28
Tableau IX	les résultats du design des ARN guides.	49
Tableau X	résultat du design du plasmide.	52
Tableau XI	design des primers (sens et anti sens)	54
Tableau XII	les séquences des 2 amorces	56
Tableau XIII	design de l'ADN recombinant	57

Liste des abréviations

SCD: Sickle cell disease

HBB: Sous unité beta de l'hémoglobine

HBG: Sous unite gamma de l'hémoglobine

HBA: Sous unite alpha de l'hémoglobine

Hb S: L'hémoglobine drépanocytaire

Hb A: L'hémoglobine d'un adulte normal

CSH : Cellule Souche Hématopoïétique

GCSH : greffe de cellule souche hématopoïétique

CRISPR: Clustered regularly interspaced palindromic repeats

Cas9 : CRISPR associated protein 9

SpCas9 : Cas9 de Streptococcus pyogenes

DR : Direct repeat

ARN : Acide ribonucléique

PAM : proto-spacer adjacent motif (le motif PAM)

Cr RNA : CRISPR ARN

Tracer RNA : Trans-activating CRISPR RNA

RNase : Ribonucléase

Cmr: Repeat-Associated Mysterious Proteins

Csm: CRISPR-Cas interference complex

ssODN: Single stranded oligodeoxyribonucleotide

E6V: Acide glutamique numéro 6 devient valine

E7V: Acide glutamique n°7 Devient valine

NHEJ: Non Homologous End joining

HDR: Homology-Directed Repair

Sg RNA: ARN guide ou single-guide RNA

dCas9 : dead Cas9 ou Cas9 catalytiquement inactive

KRAB : krüppel associated box

LCR : Région de contrôle du locus

DSB : Double strands breaks

iPSC : cellule souche pluripotente induite

CTD: C-terminal domain

BLAST n: Nucleotide basic local alignment search tool

BLAST p: Protein basic local alignment search tool

FASTA: Format de fichier

NCBI : National Center for Biotechnology Information

DDBJ: DNA Data Bank of Japan

EBI: The European Bioinformatics Institute

GLOSSAIRE

Terme	Définition
ADN	L'acide désoxyribonucléique ou ADN est une macromolécule biologique présente dans toutes les cellules ainsi que chez de nombreux virus. L'ADN contient toute l'information génétique, appelée génome, permettant le développement, le fonctionnement et la reproduction des êtres vivants.
ARN m	Acide nucléique essentiel dans le transport du message génétique et la synthèse des protéines.
ARN guide	ARN s'associant à une enzyme ou à un complexe protéique, qui, lorsqu'il s'apparie à une séquence d'ARN ou d'ADN complémentaire, permet à l'enzyme ou au complexe protéique de se positionner sur cet ARN ou ADN complémentaire.
Acides aminés	Les acides aminés sont des molécules qui entrent dans la composition des protéines grâce à leur assemblage par des liaisons que l'on appelle peptidiques.
Allogénique	Terme qui fait référence à l'ADN, aux cellules, aux tissus ou aux organes prélevés chez un donneur pour être administrés à un receveur qui est fortement, mais non entièrement, compatible avec le donneur sur le plan génétique
Bacériophage	Les bactériophages, ou phages (grec, phageton signifie nourriture/consommation), plus rarement virus bactériens, sont des virus n'infectant que des bactéries. Ils sont omniprésents dans l'ensemble de la biosphère.
Direct repeat (Les répétitions palindromiques successives)	Les répétitions directes sont un type de séquence génétique qui consiste en deux répétitions ou plus d'une séquence spécifique. En d'autres termes, les répétitions directes sont des séquences nucléotidiques présentes en plusieurs copies dans le génome.
Endonucléase	Une endonucléase est une nucléase, qui coupe un acide nucléique en fragments plus courts. Les endonucléases sont capables de couper au milieu de la chaîne, par opposition aux exonucléases qui n'attaquent que les nucléotides situés aux extrémités des fragments.
Espaceur	Fragment d'ADN de phage ou de plasmide invasifs, inséré entre deux courtes répétitions palindromiques successives du chromosome de nombreuses archées et bactéries.
Gène	Le gène est sélectionné d'un chromosome constitué d'ADN et conditionnant la transmission et l'expression d'un caractère héréditaire déterminé. Les gènes sont les unités responsables de l'hérédité, qui contrôlent les caractères ou aptitudes propres

	à un organisme.
Intégrases	Ce sont des protéines enzymatiques produits par les rétroéléments à ARN tels que les rétrovirus ou les rétrotransposons, et qui catalyse l'étape dite d'intégration du cycle réplicatif de ces agents infectieux. L'ADN étranger est ainsi "intégré" dans le génôme de l'organisme hôte.
Knock-out	En biologie moléculaire, le knock-out, en français « invalidation génique », est l'inactivation totale d'un gène. Ce terme fut initialement utilisé pour décrire la création de souris transgéniques chez qui un gène est invalidé par recombinaison homologue.
Knockt-in	Le knock-in est une technique de biologie moléculaire permettant d'étudier le rôle d'un gène. Le knock-in consiste à invalider un gène et de le remplacer par un autre gène, ou par un gène modifié. L'organisme synthétise le nouveau gène et l'assimile.
Locus	En génétique, un locus (pluriel «locus» ou «loci») est une position fixe (d'un gène ou d'un marqueur génétique) sur un chromosome.
Maladie monogénique	Les maladies monogéniques résultent d'une mutation qui affecte un seul gène. Parmi ces maladies, on peut citer la mucoviscidose également appelé fibrose kystique du pancréas.
Maladie héréditaire	Altération de l'état de santé transmissible aux descendants par les gamètes (cellules reproductrices) et résultant de la mutation (modification pathologique) d'un ou de plusieurs gènes.
Opéron	Un opéron est une unité d'ADN fonctionnelle regroupant des gènes qui opèrent sous le signal d'un même promoteur, une section d'ADN qui déclenche leur transcription. Les gènes sont ainsi transcrits en ARN messager ensemble et concourent à la réalisation d'une même fonction physiologique.
Promoteur	Séquence d'ADN favorisant l'initiation de la transcription et, le plus souvent, située en amont de la partie transcrite des gènes.
Plasmide	un plasmide est une molécule d'ADN distincte de l'ADN chromosomique, capable de réplication autonome et non essentielle à la survie de la cellule. Les plasmides sont bicaténaires (constitués de deux brins complémentaires) et généralement circulaires.
Rétrovirus	est un virus dont le génome est constitué d'ARN et dont le cycle viral comprend une étape de rétrotranscription au cours de laquelle un ADN est produit à partir de cet ARN. ... Cet ADN complémentaire du génome viral s'intègre dans le

	génomme de la cellule-hôte
Système immunitaire adaptatif	permet d'amplifier la réponse immunitaire et confère à la fois une réponse spécifique à l'antigène, et donc particulièrement adaptée à l'agent infectieux, et une réponse mémoire permettant une élimination plus efficace du même agent infectieux si l'organisme y est de nouveau confronté
trasposon	est une séquence d'ADN capable de se déplacer de manière autonome dans un génome, par un mécanisme appelé transposition. Cette transposition est rendue possible sous l'effet d'une enzyme, la transposase. Cette transposase coupe la chaîne d'ADN, qui est ensuite réparée.

Résumé

La drépanocytose (SCD) est une maladie causée par une mutation du gène de la β -globine (HBB), plus précisément à cause d'un polymorphisme nucléotidique où le gène muté deviendra (HbS).

Cette pathologie touche des millions de personnes dans le monde et se transmet selon un mode autosomique récessif, le seul traitement curatif disponible est la greffe de la moelle osseuse, qui est une intervention coûteuse et risqué car elle peut engendrer des complications et effets indésirables.

Une stratégie curative pour les β -hémoglobinopathies permet grâce à la correction génique par le système CRISPR Cas9, donc que peut-on savoir sur cette technologie et quels sont ces composants ?

Le but de ce travail (Modélisation in silico de l'implication du gène HBS dans la drépanocytose avec CRISPR-cas9) est de faire une étude in silico par outils bioinformatiques sur la mutation qui cause cette maladie, suite a cela et après avoir obtenu récolte de la séquence d'intérêt par le BLAST n (alignement nucléotidique), l'utilisation d'ORF FINDER pour connaitre les différents cadre de lecture du gène HBB, l'utilisation de swis-model nous permet de faire une modélisation de la structure tridimensionnelle de la protéine du gène HBB, et après l'obtention la séquence d'intérêt (séquence qui conduit à la drépanocytose) nous ferons une conception par la plateforme Benchling composée de plusieurs outils qui permettent d'appliquer des conceptions sur les composants du système CRISPR (L'ARN guide et l'endonucléase cas9, le vecteur plasmidique et l'ADN recombinant pour la réparation homologue). A cet effet, nous estimons que les résultats obtenus de cette recherche comporterons tous les éléments essentiels pour débiter la manipulation expérimentale in-vitro d'une thérapie génique sur des patients atteints de la drépanocytose.

Mots Clés : CRISPR-Cas9, Drépanocytose, HBB, Béta-globine, CRISPR, in silico, thérapie génique

Abstract

Sickle cell disease (SCD) is a condition caused by a mutation in the β -globin (HBB) gene, specifically due to a nucleotide polymorphism where the mutated gene becomes (HbS). This pathology affects millions of people around the world and is transmitted in an autosomal recessive mode; the only curative treatment available is a bone marrow transplant, which is an expensive and risky procedure because it can cause complications and adverse effects. A curative strategy for β -hemoglobinopathies would be through gene correction by the CRISPR Cas9 system, but what is this technology; and what are these components?

The goal of this study (In silico modeling of the involvement of the HBS gene in sickle cell disease with CRISPR-cas9) is to make an in silico study using bioinformatic tools of the mutation that causes this disease, harvesting the sequence of interest by the BLAST n (nucleotide alignment), the use of ORF FINDER to know the different reading frames of the HBB gene, the use of swiss-model which allows us to model the three-dimensional structure of the protein of the HBB gene , then after having had the sequence of interest (sequence which leads to sickle cell anemia) we will make a design by the Benchling platform which is composed of several tools which allow to make designs of the components of the CRISPR system (RNA guides and endonuclease cas9) vector (plasmid) and recombinant DNA for homologous repair. Towards the end, the results obtained from this research will include all the essential elements to begin the in-vitro experimental manipulation of a gene therapy on patients with sickle cell disease and really know if it is possible to cure them using our method of génotherapy.

Keywords: CRISPR-Cas9, Sickle cell disease, HBB, Beta-globin, CRISPR, in silico, gene therapy

المخلص

مرض الخلايا المنجلية (SCD) هو حالة ناتجة عن طفرة في جين β -globin (HBB) ، خاصة بسبب تعدد أشكال النوكليوتيدات حيث يصبح الجين طافر (HbS). يؤثر هذا المرض على ملايين الأشخاص حول العالم وينتقل بطريقة وراثية متنحية ؛ العلاج الوحيد المتاح هو زرع نخاع العظم ، وهو إجراء مكلف ومحفوف بالمخاطر لأنه يمكن أن يسبب مضاعفات وآثار سلبية. الاستراتيجية العلاجية لاعتلال الهيموجلوبين بيتا تكون من خلال تصحيح الجينات بنظام كريسبر كاس 9 ، ولكن ما هي هذه التقنية؟ وما هي هذه المكونات؟

الهدف من هذه الدراسة (في النمذجة السيليكونية لتورط الجين HBS في مرض الخلايا المنجلية باستخدام CRISPR-cas9) هو إجراء دراسة في السيليكو باستخدام أدوات المعلومات الحيوية للطفرة التي تسبب هذا المرض ، وحصاد تسلسل الاهتمام من قبل BLAST n (محاذاة النوكليوتيدات) ، واستخدام ORF FINDER لمعرفة إطارات القراءة المختلفة لجين HBB ، واستخدام النموذج السويسري الذي يسمح لنا بنمذجة البنية ثلاثية الأبعاد لبروتين جين HBB ، ثم بعد ذلك كان لدينا تسلسل الاهتمام (التسلسل الذي يؤدي إلى فقر الدم المنجلي) سنقوم بتصميم من خلال منصة Benchling التي تتكون من عدة أدوات تسمح بعمل تصميمات لمكونات نظام CRISPR (أدلة RNA و endonuclease cas9) ناقل (البلازميد) والحمض النووي المؤلف من أجل الإصلاح المتماثل. في النهاية ، ستشمل النتائج التي تم الحصول عليها من هذا البحث جميع العناصر الأساسية لبدء المعالجة التجريبية في المختبر للعلاج الجيني للمرضى المصابين بمرض الخلايا المنجلية ومعرفة ما إذا كان من الممكن علاجهم باستخدام طريقة العلاج الجيني لدينا.

كلمات مفتاحية: الكريسبر كاس9, مرض كريات الدم الحمراء المنجلية, الجين لبيتاغلوبين, البيتا غلوبين, تقنية الكريسبر, دراسة بالبيوانفورماتيك , التصحيح الجيني.

Sommaire

Introduction.....	1
Chapitre 1 : Partie bibliographique.	
1. Généralités sur le CRISPR-Cas9.....	4
1.1 . Historique.....	4
1.1.1. Origine du CRISPR ...	4
1.1.2. Gènes cas.....	4
1.1.3 Origine des spacers.....	4
1.2. Définition.....	5
1.3. Différents types de systèmes CRISPR-Cas.....	6
1.3.1. Systèmes CRISPR-Cas de types I.....	7
1.3.2. Systèmes CRISPR-Cas de types II.....	8
1.3.3. Systèmes CRISPR-Cas de types III	9
1.4. Motifs PAM.....	9
1.5. Endonucléases Cas9	11
1.6. Mécanisme du système CRISPR-cas	11
1.7. Mécanisme de réparation.....	14

1.8. Edition des gènes.....	15
1.9. Conception d'un ARN guide unique.....	15
1.10. Types de cas9.....	16
1.11. Limitations du CRISPR-Cas9.....	16
1.12. Avenir de CRISPR	16
1.13. Applications de la technologie CRISPR	16
2. Drépanocytose	17
2.1 Définition de la drépanocytose	17
2.2 Épidémiologie de la maladie	18
2.3.1. Description du Gène HBB	19
2.3.2. La protéine du Gène HBB (bêta-globine)	20
2.4. Mutations spécifiques de la drépanocytose	21
2.5. Mode de transmission de la drépanocytose	23
Chapitre 2 : Matériel et méthodes.....	26
1. Matériel.....	27
1.1. Matériel biologique.....	27
1.2. Matériel non biologique.....	28
2. 2. Méthode.....	29
2.1 Obtention des données moléculaires.....	30
2.2 Aligement par Blast n.....	32

2.3. ORF Finder.....	33
2.4. Swiss-model.....	34
2.5. Benchling.....	35
Chapitre 3 : Résultats et discussion	37
1. Résultat.....	37
1.1. Résultat de genbank.....	38
1.2. Les résultats de Blast n.....	39
1.3. Résultat d'orf finder.....	40
1.4. Résultats de swiss-prot d'ARNm du gène HBB.....	42
1.4.1. La séquence protéique du gène HBB.....	43
1.4.2. Structure tridimensionnelle de la protéine HBB.....	43
1.4.3. Structure tridimensionnelle de l'hémoglobine.....	44
1.4.4. Alignement de la séquence protéique du HBB avec du variant HBS.....	45
1.5. Résultats de STRING	47
1.6. Résultats de Benchling.....	49
1.6.1. Design des ARN guides	49
1.6.2. Conception du plasmide	52
1.6.3. La conception des primers (les amorces) sur benchling	54
1.6.4. Design de l'ADN recombinant	57
2. Discussion	62
Conclusion et perspectives.....	64
Références bibliographiques.....	65

Introduction

La drépanocytose (SCD) est un terme générique utilisé pour divers troubles génétiques héréditaires qui entraînent une malformation de la structure des protéines de l'hémoglobine, conduisant à des globules rouges anormaux en forme de faucille et à un large éventail d'autres pathologies cliniques. La drépanocytose affecte 70 000 à 100 000 personnes aux États-Unis (**Hassell, 2010**) et environ 300 000 nouveau-nés dans le monde chaque année (**Piel, Steinberg and Rees, 2017**) ce qui en fait l'une des maladies monogéniques les plus courantes. Aux États-Unis, la durée de vie médiane des patients SCD est de 42 ans pour les femmes et de 38 ans pour les hommes (**Lanzkron, Patrick Carroll and Haywood, 2013**). Depuis l'identification de la mutation monogénique causale en 1956 par V.M. Ingram (**Ingram, 1956**) la génétique et la régulation des gènes impliqués ont fait l'objet de recherches intensives. La SCD est causée par l'héritage de deux copies anormales du gène de la β -globine (HBB), dont au moins une est la variante de l'hémoglobine S (HbS). La présence de deux copies du gène HbS (HbSS) provoque l'anémie falciforme, le cas le plus grave par rapport à l'hétérozygotie composée (**Frenette et al., 2007**). La variante HbS est le résultat d'une substitution nucléotidique unique de A à T dans le codon pour le sixième acide aminé dans la protéine β -globine, une sous-unité de la protéine hémoglobine tétramérique transportant l'oxygène ($\alpha_2\beta_2$) dans les globules rouges (**Frenette et al., 2007**). Cette mutation ponctuelle convertit un acide glutamique hydrophile en une valine hydrophobe en position six dans la β -globine, conduisant à un repliement anormal de l'hémoglobine. L'hémoglobine HbS qui en résulte a tendance à polymériser et à s'agréger, transformant les globules rouges en cellules rigides en forme de faucille. Ces cellules inflexibles ont tendance à se coller les unes aux autres et aux parois des vaisseaux sanguins, provoquant une vaso-occlusion, ce qui ralentit le flux sanguin et diminue l'apport d'oxygène aux tissus. Une telle mauvaise circulation sanguine peut endommager les organes et entraîner beaucoup de complications. Malheureusement il n'y a pas encore de remède universel; cependant, des traitements de soutien pour aider à réduire les complications de la maladie sont disponibles. Ces traitements comprennent les transfusions sanguines, les thérapies préventives telles que la prophylaxie à la pénicilline et la vaccination antipneumococcique, et la thérapie à l'hydroxyurée, qui diminue la polymérisation de l'HbS en augmentant les taux d'hémoglobine fœtale (HbF) (**Aliyu and Tumblin, 2005**). Compte tenu du fait que toutes les cellules sanguines sont dérivées de cellules souches hématopoïétiques (CSH), le seul traitement curatif disponible pour la drépanocytose a été la transplantation de cellules souches hématopoïétiques (HSCT) allogénique de donneurs sains pour reconstituer le

corps du patient avec des cellules sanguines saines (**Shenoy, 2011**). Cependant, même si la GCSH, mais cette méthode n'est pas disponible pour tous les patients en raison de la rare disponibilité de donneurs appariés et des effets secondaires associés, y compris à long terme toxicités telles que l'infertilité et les endocrinopathies. Pour guérir des maladies monogéniques comme les SCD pour lesquelles la mutation causale est déjà identifiée, une autre approche prometteuse serait de corriger directement la mutation au niveau du locus endogène sans avoir besoin de l'expression d'un gène exogène. Une telle thérapie basée sur l'édition génique précise pourrait offrir une alternative plus sûre aux thérapies basées sur l'expression transgénique actuelles grâce à la technologie communément appelées (CRISPR / Cas9); groupement régulier de courtes répétitions palindromiques espacées (CRISPR) - protéine associée à Cas9. Le but de ce mémoire est l'étude de la technologie CRISPR / Cas9 pour analyser la mutation HbS responsable de la drépanocytose et proposer une stratégie thérapeutiques potentielle.

1. Généralités sur le CRISPR-Cas9

1.1. Historique

1.1.1. Origine du CRISPR

En 1987, Ishino et coll. observent, après la séquence du gène *iap* dans le génome d'*Escherichia coli*, une structure constituée de la répétition de cinq séquences conservées de 29 nucléotides, séparées par des espaceurs ou spacers uniques de 32 nucléotides. Une seconde étude par cette même équipe montre que cette séquence n'est pas codante et qu'elle est retrouvée dans d'autres espèces proches. Aucune fonction ne lui est attribuée. Des structures similaires sont observées au cours des dix années suivantes par différentes équipes, chez des bactéries et des arches. L'acronyme CRISPR pour Cluster of Regularly Interspaced Short Palindromic Repeats est proposé par Jansen et coll. en 2002. Les répétitions, appelées DR « direct repeat », semblent posséder un palindrome qui pourrait leur conférer un rôle biologique (**fig.1**). L'étude publiée en 2002 par Tang et coll. démontre que la structure répétée est transcrite abondamment chez *Archaeoglobus fulgidus*. Au cours de l'exploration de petits ARN, souvent impliqués dans la régulation de l'expression de gènes, les auteurs observent des ARN d'une taille correspondant à un DR plus un spacer, ainsi que des multiples de ces ARN et hybridant à différentes positions sur trois locus CRISPR. Ils en concluent qu'un long transcrit précurseur est mûri par coupures progressives en petits ARN et proposent l'hypothèse selon laquelle ces molécules pourraient jouer un rôle dans le fonctionnement du CRISPR. Cette observation sera faite chez d'autres espèces confirmant l'importance de l'étape de transcription. (**Christine, 2018**)

1.1.2. Gènes cas

En 2002, Jansen et coll. montrent l'existence d'un groupe de gènes particuliers toujours situés près d'un CRISPR, et les appellent gènes cas pour CRISPR-associated. L'étude montre que dans un génome possédant plusieurs CRISPR avec le même DR, un seul est associé au groupe de gènes cas. L'équipe décrit également le leader, une séquence flanquant le CRISPR sur un côté et possédant un motif TATA, et qui sera plus tard identifié comme le promoteur pour la transcription du CRISPR. (**Christine, 2018**)

1.1.3. Origine des spacers

En 2005 sont publiées trois études montrant de manière indépendante que les spacers correspondent majoritairement à des séquences virales ou plasmidiques, Deux des études in silico explorent les séquences génomiques procaryotes disponibles et réussissent à assigner une origine probable à 2 % des spacers par homologie de séquence Bolotin et coll. décrivent de plus la présence d'une courte séquence Pu-Py-A-A proche de la séquence homologue du spacer, appelée protospacer, et qui s'aligne avec la fin du DR (...ACAAC) chez *Streptococcus thermophilus*. Cette séquence sera à nouveau décrite en 2008, puis appelée Protospacer Adjacent Motif ou PAM en 2009, Les nombreuses études effectuées depuis la découverte d'un CRISPR en 1987 ont permis d'établir que le système CRISPR-Cas est un mécanisme d'immunité adaptative chez les procaryotes L'utilisation de la bioinformatique va puiser dans les données de génomique de divers espèces Pour analyser de nombreux fonctions et traitements biologiques. (Christine, 2018)

<i>Thermofilum pendens</i>	GTATCAACAACGAATGAGTTGA ^{AAG} 25pb
<i>Verminephrobacter eiseniae</i>	TTTCTGAGCTGCCATCCGGCAGTG ^{AAC} 28pb
<i>Corynebacterium glutamicum</i>	GGAAATACCTCCGC GTGTGCGGAG AAGAG 29pb
<i>Desulfotomaculum reducens</i>	GTCAAAACACAAAAT ^{AATCCCTTTGGGAATTGA} ^{AAC} 37pb
<i>Azoarcus sp.</i>	GCATCGCCCC TCGGTGACGGGGGGCGTGGATTGA ^{AAC} 37pb
<i>Flavobacterium psychrophilum</i>	^{GTT} GGTAATTATAAGCT ^{AAAATACAATTTT} GAAAGCAATTCAC ^{AAC} 46pb

Figure 1 : Diversité des séquences DR. Séquences du DR de six CRISPR appartenant à différentes espèces bactériennes. Le palindrome est surligné et la séquence 3' terminale AAC/G est entourée. (Christine, 2018)

1.2. Définition

Le système CRISPR-Cas est constitué du locus CRISPR, de l'anglais clustered regularly interspaced short palindromic repeats, et de gènes cas (CRISPR-associated). Ce locus (**fig.2**) est composé de courtes séquences nucléotidiques répétées (appelées répétitions) intercalées de courtes séquences nucléotidiques variables et souvent uniques (appelées espaceurs, de l'anglais spacers) (Croteau Félix , Geneviève . Rousseau, 2018) , La taille des courtes répétitions varie selon les espèces microbiennes, mais elles sont habituellement de 20

à 50 paires de base et leurs séquences peuvent partiellement former des structures palindromiques, ces répétitions flanquent des espaceurs de tailles comparables, mais dont les séquences nucléotidiques sont souvent propres à chaque souche microbienne (**Croteau Félix , Geneviève . Rousseau, 2018**). CRISPR-Cas est un système immunitaire adaptatif utilisé par de nombreux microbes pour se défendre contre l'invasion d'acides nucléiques tels que les génomes viraux et autres éléments génétiques mobiles. Le système microbien utilise son locus CRISPR pour stocker de l'information génétique afin de produire des ARN guides. Ces derniers, de concert avec des endonucléases (Cas), empêchent des invasions futures (**Croteau Félix , Geneviève . Rousseau, 2018**).

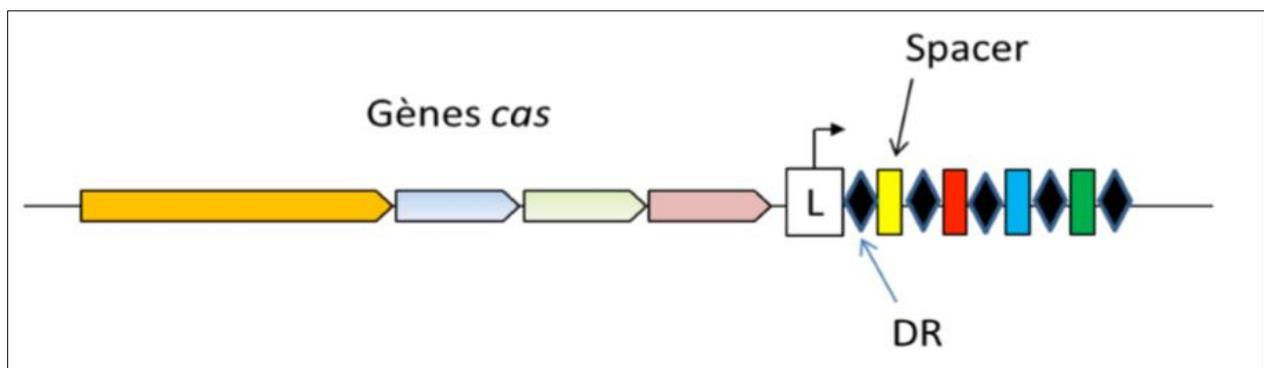


Figure 2 : Représentation schématique d'un locus du système CRISPR- Cas.

Le groupe de gènes cas est suivi du leader représenté par une boîte marquée L et d'une succession de séquences DR (losanges noirs) séparés par les spacers (rectangles colorés).(**Christine, 2018**).

1.3. Différents types de systèmes CRISPR-Cas

Il existe 3 types principaux (**fig.3**) du système CRISPR-Cas de l'immunité adaptative qui diffèrent par :

- les répertoires de gènes associés aux CRISPR (répertoires de cas gènes)
- l'organisation des opérons cas
- la structure des répétitions au sein des matrices CRISPR (**Jiang and Marraffini, 2015**)

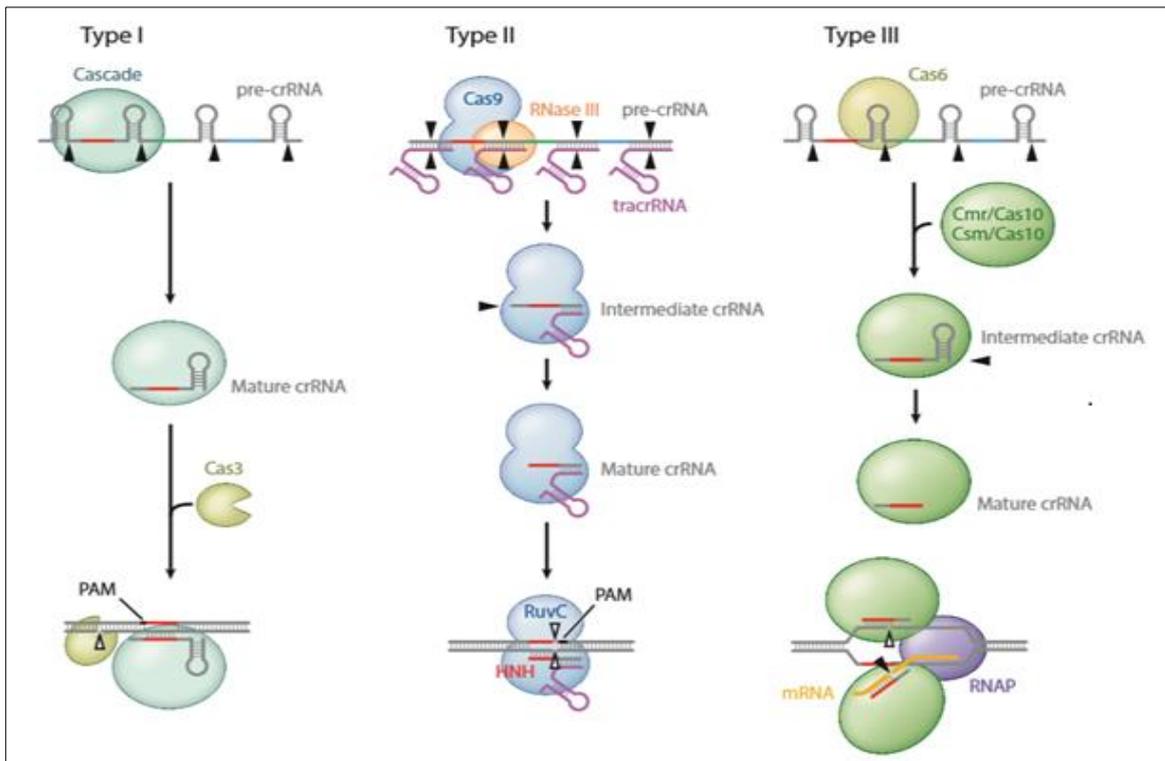


Figure 3 : Différentes types du système de CRISPR (Jiang and Marraffini, 2015)

1.3.1. Systèmes CRISPR-Cas de types I

Utilisent des complexes multi-protéiques pour le ciblage guidé par le crRNA.

-endoribonucléases de la famille Cas6 : clivage du pre-crRNA et biogenèse des crRNA des systèmes de types I-A, I-B, I-D, I-E et I-F (**fig.4**)

-endoribonucléase Cas5d : clivage du pre-crRNA et biogenèse des crRNA des systèmes de type I-C

-Chez *Escherichia coli*, les crRNAs sont incorporés dans un complexe multimérique appelée Cascade, complexe associé à CRISPR pour la défense antivirale), système CRISPR-Cas de type I-E nécessaire pour la protection contre les bactériophages.

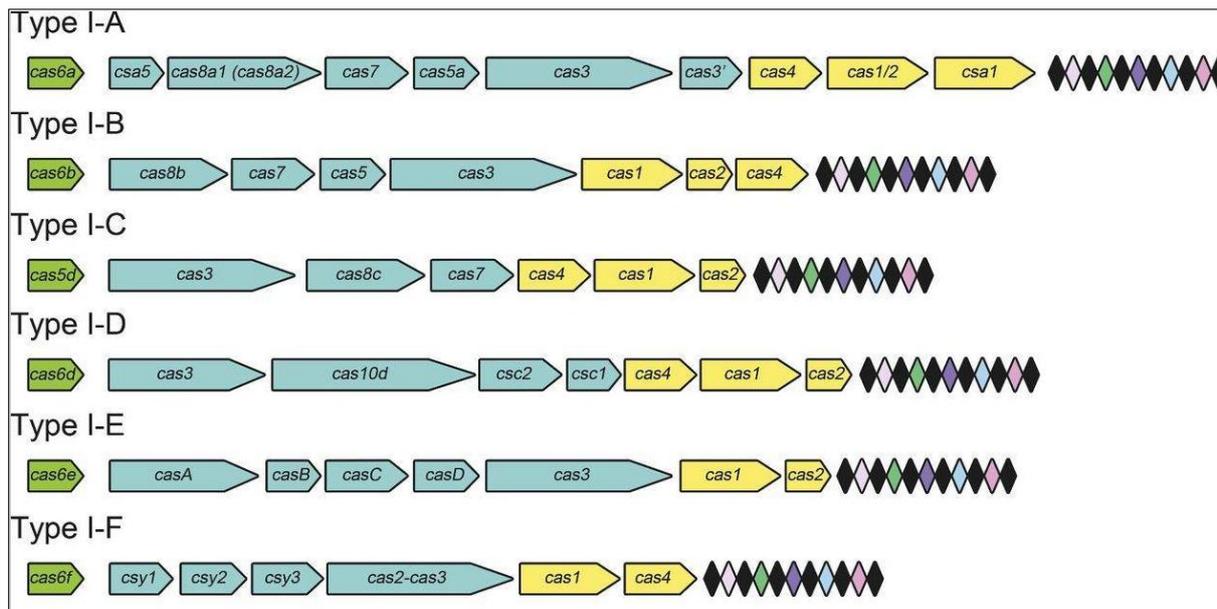


Figure 4 : Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types I-A à I-F (Charpentier *et al.*, 2015)

1.3.2. Systèmes CRISPR-Cas de types II

Ce sont les plus simples (**fig.5**) : un petit ARN agissant en trans appelé tracrRNA (trans-activating small RNA) s'apparie avec chaque répétition du pre-crRNA pour former un ARN double brin [tracrRNA:crRNA] clivé par la RNase III en présence de l'endonucléase Cas9. Ils génèrent des cassures double brin dans l'ADN invasif.

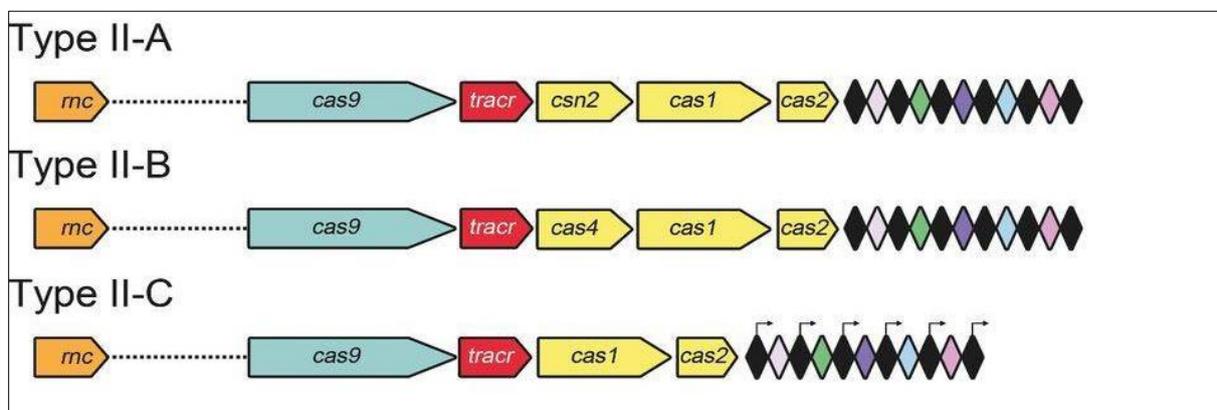


Figure 5 : Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types II-A à II-C (Charpentier *et al.*, 2015).

1.3.3. Systèmes de types III

Ce sont les plus complexes (**fig.6**). Ces locis contiennent des gènes codant des protéines Cas10 et des modules Csm (pour le type III-A) ou Cmr (pour le type III-B) qui forment ensemble les complexes de ciblage [Cas10-Csm] ou [Cas10-Cmr]. Comme dans les systèmes de types I, le clivage du pre-crRNA est réalisée par Cas6. Cependant, l'endonucléase Cas6 ne fait pas partie du complexe Cas10. Les systèmes de types III ont deux caractéristiques : la transcription de la cible est nécessaire pour l'immunité et les deux cibles d'ADN sont clivées. (**Jiang and Marraffini, 2015**)

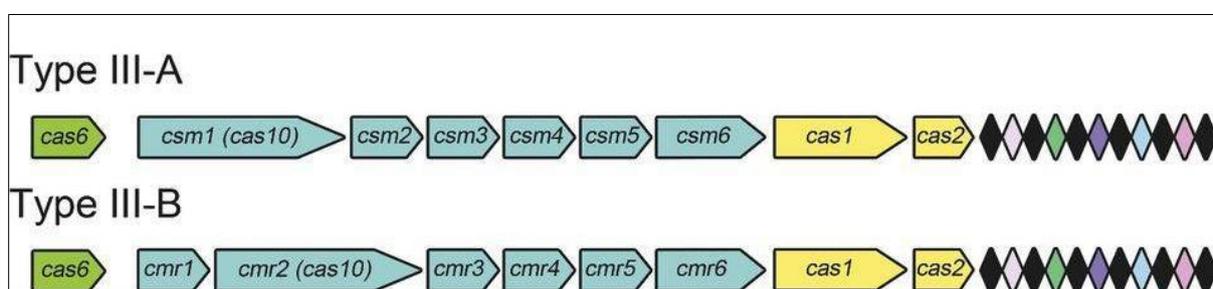


Figure 6 : Schéma représente la composition des gènes cas des systèmes CRISPR-Cas de types III-A et III-B (**Charpentier et al., 2015**)

1.4. Motifs PAM

Les séquences répétées (repeats) du locus CRISPR sont séparées par de courtes séquences d'ADN non répétitif - les séquences espaceurs (spacers) - qui proviennent de l'ADN invasif d'un plasmide ou d'un virus lors de la phase d'adaptation. Lors de l'interférence, la séquence nucléotidique espaceur (portée par le crRNA) doit être identique à une séquence du génome de ce plasmide ou de ce virus appelée proto-espaceur (protospacer - environ 20 nucléotides) afin que le complexe CRISPR-Cas bloque la réplication de l'ADN invasif. Dans les systèmes CRISPR-Cas de types I, de types II, il existe une séquence possédant un motif conservé qui suit immédiatement la séquence cible proto-espaceur. Cette séquence de l'ADN invasif de 2 à 5 nucléotides est appelée motif adjacent au proto-espaceur (Protospacer Adjacent Motif - PAM).(**Anders et al., 2014**)

-Le motif PAM (**fig.07**) est :

-nécessaire à l'acquisition des séquences espaceurs.

-nécessaire à l'endonucléase Cas9 pour fixer l'ADN cible.

-nécessaire au phénomène d'interférence.

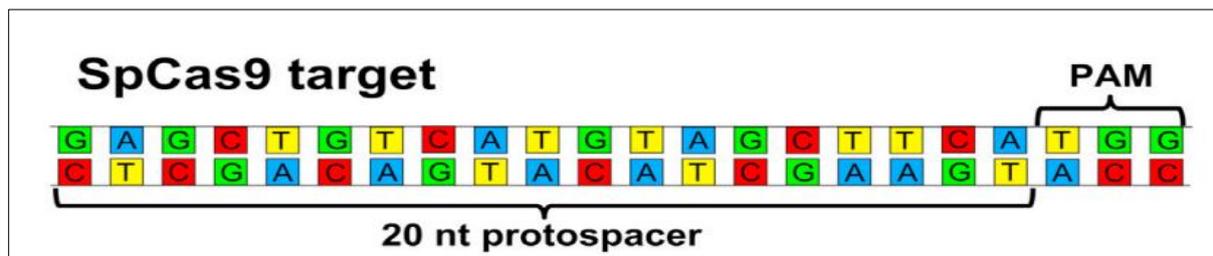


Figure 7 : Motif adjacent au proto-espaceur (Protospacer Adjacent Motif - PAM). (Siksnys and Gasiunas, 2016)

Tableau I: Différentes espèces et leurs (Protospacer Adjacent Motif - PAM). (Anders *et al.*, 2014)

Espèce et type d'endonucléase Cas9	Motif PAM – N : désigne n'importe quel nucléotide
Streptococcus pyogenes (SpCas9) - Type IIA	SpCas9 reconnaît le motif PAM 5'-NGG-3' située à 3 paires de base en 3' du site de clivage sur le brin d'ADN non complémentaire.
SpCas9 - D1135E	NGG
SpCas9 - VRER	NGCG
SpCas9 - EQR	NGAG
SpCas9 - VQR	NGAN ou NGNG
Streptococcus mutans - Type IIA	NGG
Streptococcus thermophilus B - Type IIA	NNAAAAW
Staphylococcus aureus - SaCas9	NNGRRT ou NNGRR(N)
Treponema denticola - Type IIA	NAAAAC
Lactobacillus buchneri - Type IIA	NAAAAN
Francisella novicida - Type IIB	NG
Neisseria meningitidis - Type IIC	NNNNGATT

1.5. Endonucleases Cas9

L'endonucléase Cas9 de *Streptococcus pyogenes* (SpyCas9) est spécifique des systèmes de type II-A, Elle contient deux domaine nucléase très conservés (RuvC) et (HNH). (**Fig.8**)

Le domaine RuvC est constitué de trois segments discontinus (RuvC-I à RuvC-III) : acides aminés 1 - 59, 718 - 769 et 909 - 1098) avec un lobe α -hélicoïdal inséré entre le premier et le deuxième segment, le domaine HNH (acides aminés 775 - 908) est inséré entre le deuxième et le troisième segment, elle contient aussi et domaine (Arg) région riche en arginine, domaine Topo-homologie, domaine C-terminal CTD (acides aminés 1099 - 1368) qui est impliqué dans la fixation du motif PAM (**Jinek et al., 2014**)

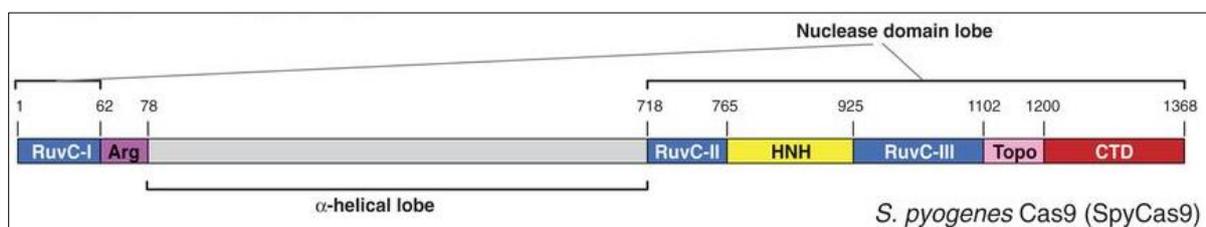


Figure08 : Schéma de la séquence polypeptidique et de l'organisation du domaine pour le type II-A Protéine Cas9 de *S. pyogenes* (SpyCas9). (**Jinek et al., 2014**)

1.6. Mécanisme du système CRISPR/Cas

Comme mentionné précédemment, les bactéries et les archaebactéries utilisent le système CRISPR/Cas pour se protéger contre l'invasion d'ADN étranger, tel que l'ADN viral. Trois types de systèmes CRISPR/Cas (I-III) ont été identifiés, le système de type II étant le mieux caractérisé. Dans le système de type II, de courtes séquences non répétitives appelées (spacers) sont intégrées dans chaque locus CRISPR. Ces (spacers) sont des vestiges d'une précédente invasion virale et servent de mémoire moléculaire qui peut être mobilisée pour améliorer l'issue de futures attaques (**fig.9**). Les espaceurs sont transcrits en ARN CRISPR (crRNA) et s'hybrident avec une autre molécule d'ARN, le crRNA transactivateur (tracrRNAs), pour former un complexe pré-crRNA-tracrRNA. Le traitement enzymatique produit une molécule mature de crRNA-tracrRNA ayant une homologie avec les séquences

virales qui guident la protéine de nucléase Cas pour couper les deux brins de l'ADN viral et désactiver le virus (**Wang et al., 2015**)

- Le système CRISPR-Cas passe par trois phases :

a) Phase d'acquisition : aussi appelée d'adaptation, un système CRISPR-Cas conserve de courtes séquences d'ADN invasif (les séquences espaceurs - spacers) comme éléments de mémoire des génomes des virus et des plasmides rencontrés. (**Wang et al., 2015**)

Cette phase commence par : La reconnaissance de l'ADN invasif par le complexe [Cas1 - Cas2] (**fig.10**). Elle est suivie par le clivage de la séquence proto-espaceur (30 à 40 paires de base selon le système CRISPR-Cas). Le complexe [Cas1 - Cas2] catalyse cette réaction via un mécanisme d'intégration nucléophile directe, semblable à celui des intégrases et des enzymes transposases des rétrovirus. La séquence espaceur est intégrée entre la séquence de tête (leader sequence) et la première séquence répétée de la matrice CRISPR. L'intégration de la séquence espaceur est couplée à la duplication de la séquence répétée adjacente. L'activité endonucléase de Cas2 n'intervient pas dans l'acquisition : la fonction principale de l'enzyme Cas2 est de former un échafaudage non catalytique au sein du complexe [Cas1-Cas2].

b) Phase de transcription - maturation : la matrice CRISPR (ensemble des séquences répétées - repeats - et des séquences espaceurs) est transcrite en un long transcrit primaire appelé pre-crRNA. Celui-ci est mûri en un ensemble de courts fragments d'ARN appelés crRNAs (CRISPR RNAs) : chacun d'entre eux contient une séquence unique complémentaire d'un fragment d'ADN invasif. (**Wang et al., 2015**)

c) Phase d'interférence : Les crRNAs sont utilisés comme ARN guide des protéines Cas pour cibler et hydrolyser l'ADN apparenté de virus ou de plasmides invasifs lors de nouvelles infections. (**Wang et al., 2015**)

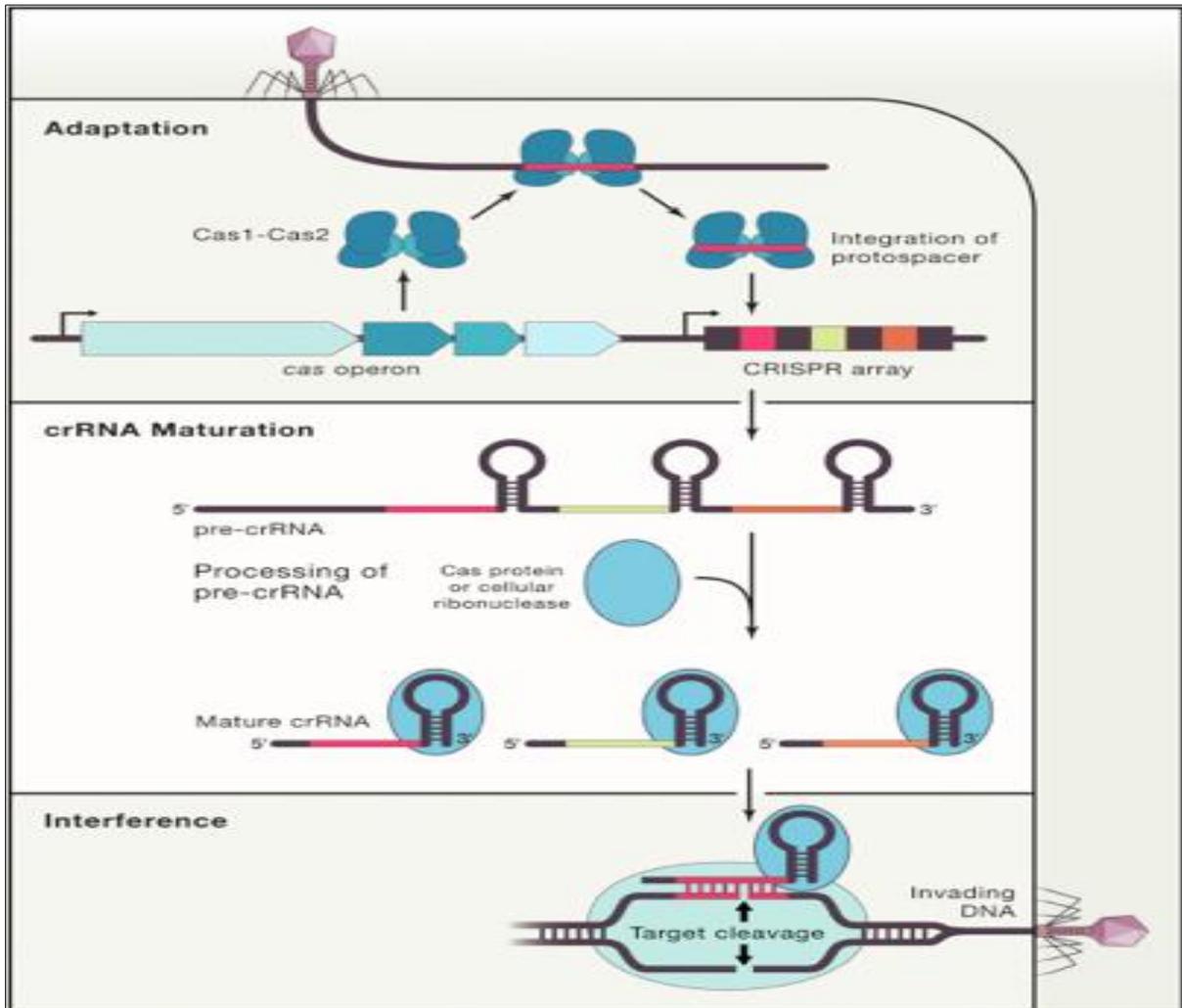


Figure 9 : Mécanisme et étapes du système CRISPR-Cas. (Hille *et al.*, 2018)

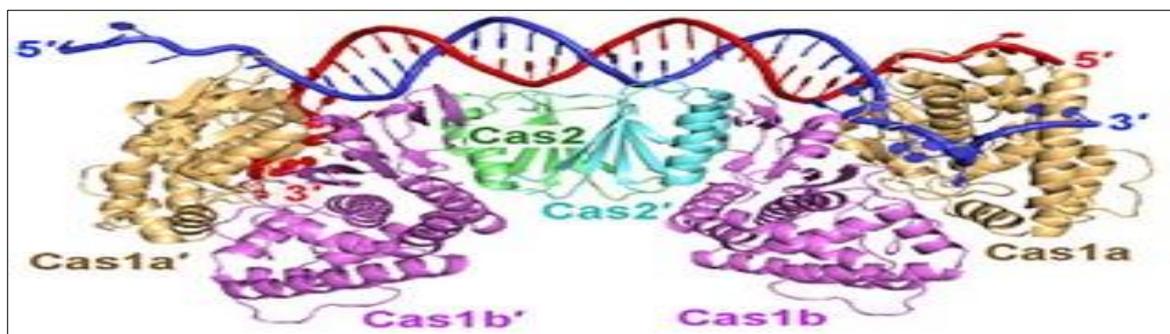


Figure 10 : Structure du complexe [Cas1 - Cas2] de *Escherichia coli* lié à une séquence d'ADN proto-espacer constitué de 4 enzymes Cas1 et 2 enzymes Cas2. (Wang *et al.*, 2015)

1.7. Mécanisme de réparation

Les cassures du double brin d'ADN produites par la Cas9 peuvent être réparées spontanément par jonction d'extrémités non-homologues (en anglais, non homologous end joining, NHEJ) ou en présence d'une séquence d'ADN donneur, par recombinaison homologue (en anglais, homology directed repair, HDR) (**Fig.11**). Cet ADN donneur doit contenir, de chaque côté de la coupure de l'ADN ciblé, des séquences homologues allant de 50 à 2000 nucléotides. Entre ces deux séquences homologues, l'ADN donneur peut comporter une longue séquence de nucléotides à introduire dans le gène ciblé. Cette partie de l'ADN donneur permet ainsi d'induire une mutation d'un ou plusieurs nucléotides dans le gène qui a été ciblé, ou d'introduire des segments de gène qui en sont absents (**Tremblay, 2015**)

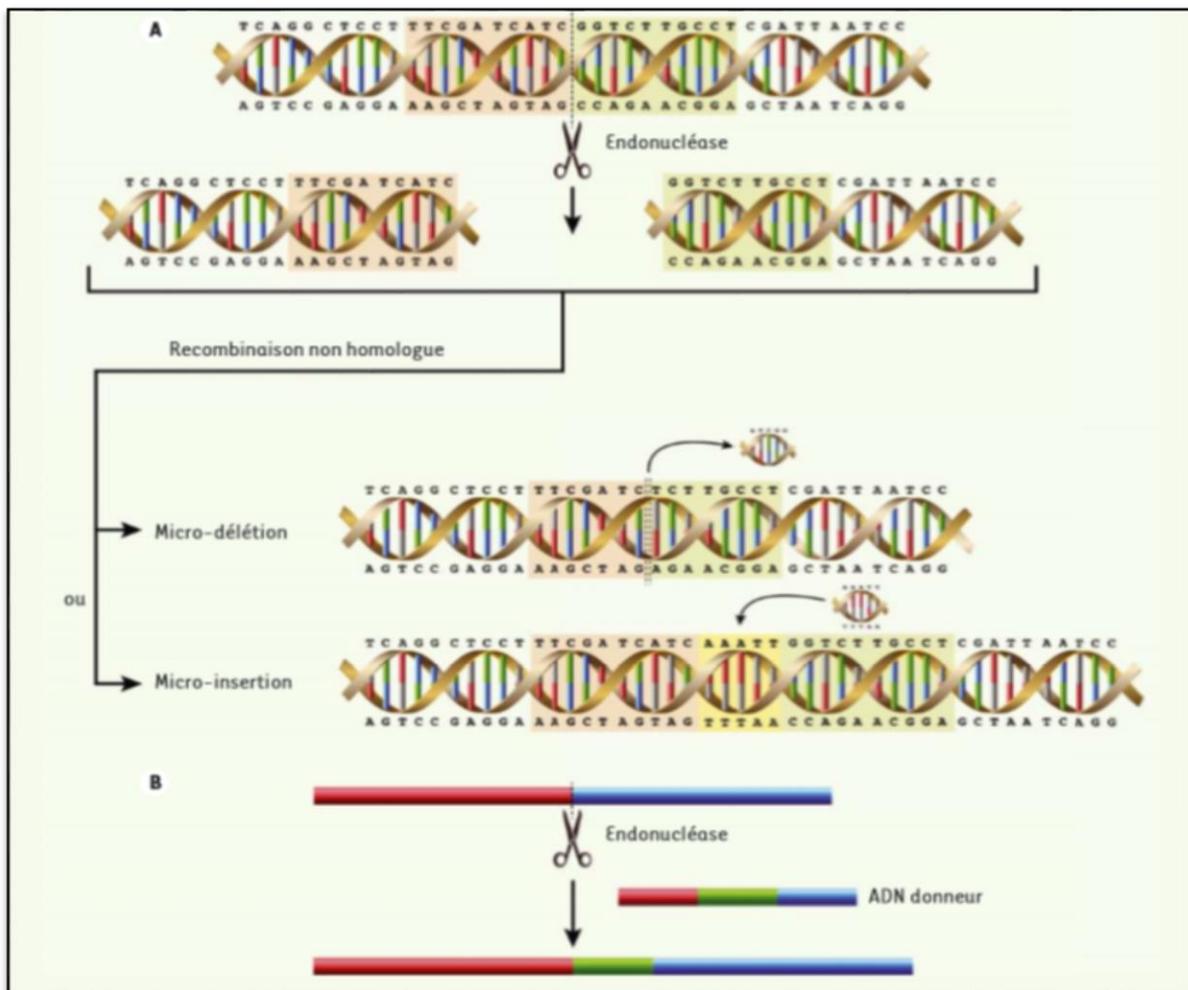


Figure11 : Système de réparation chez les eucaryotes (non homologous end joining, NHEJ) et (homology directed repair, HDR) (**Tremblay, 2015**)

1.8. Edition des gènes

En raison de sa simplicité et des possibilités d'adaptation, la technique CRISPR est rapidement devenue la plus populaire pour l'ingénierie des génomes. Elle nécessite un ARN spécifique pour recruter et diriger l'activité nucléase de Cas9 vers la séquence d'ADN cible du gène d'intérêt pour induire des cassures double brin dans l'ADN. Cela permet de modifier spécifiquement le gène ciblé dans des cellules de plantes, d'animaux et d'humains. Des variantes de la technique permettent également de réduire ou d'augmenter l'expression d'un gène choisi. Cette technologie peut donc être utilisée non seulement pour permettre de comprendre le rôle d'un gène, mais aussi de développer des thérapies pour des maladies héréditaires et acquises. Cet ARN guide est : -Soit constitué du tracrRNA plus un crRNA, tous les deux synthétisés chimiquement. -Soit un ARN guide unique (single guide RNA-sgRNA) qui correspond au tracrRNA et au crRNA en une seule construction (**Croteau Félix , Geneviève . Rousseau, 2018**)

1.9. Conception de l'ARN guide

Une séquence d'ADN génomique de 20 nucléotides est choisie dans le gène que l'on souhaite cibler. Par exemple, pour l'extinction d'un gène, on choisit une séquence cible située entièrement dans un exon constitutif du gène codant. Cette séquence cible doit se situer immédiatement avant le motif PAM (NGG) (**fig.12**) : la séquence cible peut donc être choisie sur l'un ou l'autre des 2 brins d'ADN à condition qu'elle soit suivie du motif PAM. (**Kim and Kim, 2014**)

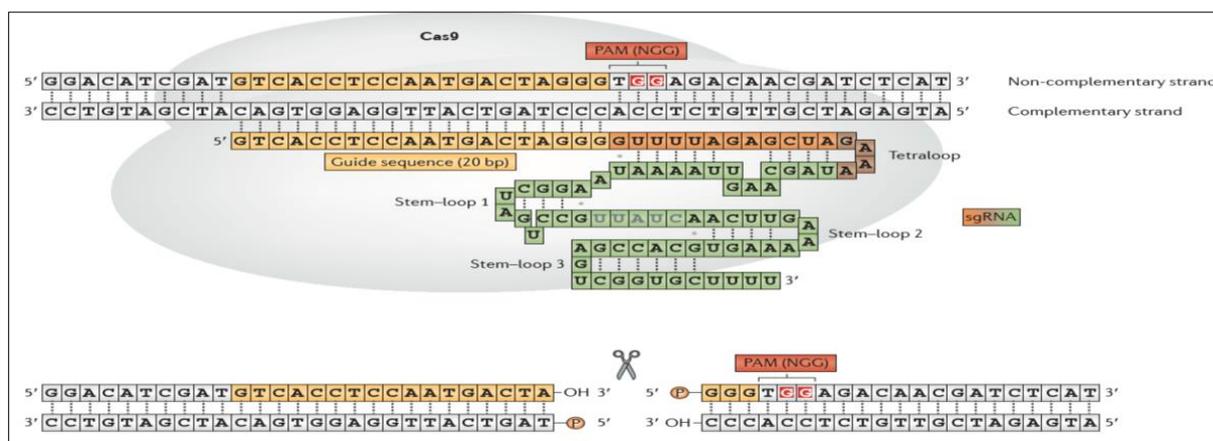


Figure12 : Représentation schématique des nucléases guidées par l'ARNg (**Kim and Kim, 2014**)

1.10. Types de cas9

La cas9 sauvage a été modifié en conséquence, de sorte que jusqu'à présent nous avons trois types de cas9 basés sur sa fonction. (**Sander and Joung, 2014**)

1) **La cas 9 sauvage** : induire des ruptures double brins, utilisé pour le knockout.

2) **Dead cas9 (dcas9)** : pas de coupure, utilisé pour le knock-down (répression) et le knock-in (activation), basé sur l'utilisation de plasmide pour cibler deux domaines (KRAB : domaines de répression transcriptionnelle et VP64 : domaine activateur de transcription)

3) **Nickase** : elle induit des cassures à un seul brin. (Modifier la fonction d'un gène)

1.11. Limitations du CRISPR-cas9

Il est difficile de délivrer le matériel CRISPR / Cas aux cellules matures en grand nombre, ce qui reste un problème pour de nombreuses applications cliniques. Les vecteurs viraux sont la méthode d'administration la plus courante, et n'est pas 100% efficace, donc même les cellules qui absorbent CRISPR / Cas peuvent ne pas avoir d'activité d'édition du génome. Il n'est pas précis à 100%, et les modifications «hors cible» off-targets, bien que rares, peuvent avoir de graves conséquences, en particulier dans les applications cliniques. (**Tremblay, 2015**)

1.12. Avenir de CRISPR

Les applications futures de la technologie CRISPR sont essentiellement aussi infinies que les formes de vie elles-mêmes. Alors que les initiatives actuelles sont principalement axées sur la thérapeutique ou la technologie alimentaire, il existe également des applications moins importantes mais très réelles du système CRISPR-Cas9. (**Donohoue, Barrangou and May, 2018**)

1.13. Applications de la technologie CRISPR

Les industries touchées par CRISPR comprennent la médecine, l'alimentation, l'agriculture et l'espace biotechnologique industriel. Parce que le système d'édition de gènes CRISPR-Cas9 est si facile à fabriquer et à utiliser, les chercheurs de diverses disciplines scientifiques

peuvent y accéder pour modifier génétiquement l'organisme de leur choix. Produits pharmaceutiques et biotechnologie: L'avenir de la médecine sera écrit avec CRISPR; en permettant une identification précise et rapide des cibles génétiques potentielles pour des tests précliniques efficaces. Parce qu'il peut être utilisé pour «éliminer» différents gènes, CRISPR offre aux chercheurs un moyen plus rapide et plus abordable d'étudier des centaines de milliers de gènes pour voir lesquels sont affectés par une maladie particulière. Alimentation et agriculture: "Si vous avez mangé du yaourt ou du fromage, il est probable que vous ayez mangé des cellules CRISPR." - Rodolphe Barrangou, ancien scientifique de Danisco et rédacteur en chef du CRISPR Journal. L'édition de gènes peut rendre l'agriculture plus efficace. Il peut réduire les pénuries alimentaires mondiales pour les cultures de base comme les pommes de terre et les tomates. Et il peut créer des cultures résilientes, imperméables aux sécheresses et autres impacts environnementaux. Biotechnologie industrielle: En réorganisant les microbes à l'aide de CRISPR, les chercheurs peuvent créer de nouveaux matériaux. En altérant les microbes pour accroître la diversité, créer de nouveaux matériaux biosourcés et fabriquer des biocarburants plus efficaces. Des produits chimiques actifs dans les parfums à ceux impliqués dans le nettoyage industriel, CRISPR pourrait avoir un grand impact ici en créant de nouveaux matériaux biologiques plus efficaces. (Donohoue, Barrangou and May, 2018)

2. Drépanocytose

2.1. Définition de la maladie

La drépanocytose (SCD) (**fig.13**), une des maladies génétiques les plus courantes, est causée par des mutations du gène HBB codant la sous-unité bêta de l'hémoglobine. Après la naissance, les tétramères d'hémoglobine (**fig.14**) contiennent deux sous-unités alpha et deux bêta globines codées par le gène HBB qui débute son expression chez le nouveau-né. Avant cela, la bêta globine codée par l'un des deux gènes HBG qui sont exprimées au stade fœtal et normalement réduites jusqu'à l'inactivité après la naissance. Alors qu'une mutation ponctuelle dans le codon 6 (GAG > GTG), entraînant la substitution de l'acide glutamique à l'acide aminé valine) dans le gène HBB crée un trait SCD (Cai *et al.*, 2018).

2.2. Epidémiologie

La prévalence du trait en Europe est estimée à environ 1/150 : en Afrique centrale et de l'ouest (15-25%), dans les DOM d'Amérique (10-12%) et certaines régions méditerranéennes (1-15%), une forte prévalence est observée dans les zones étant ou ayant été impaludées, cette anomalie offrant une protection contre le neuro-paludisme. (Santin, 2013)

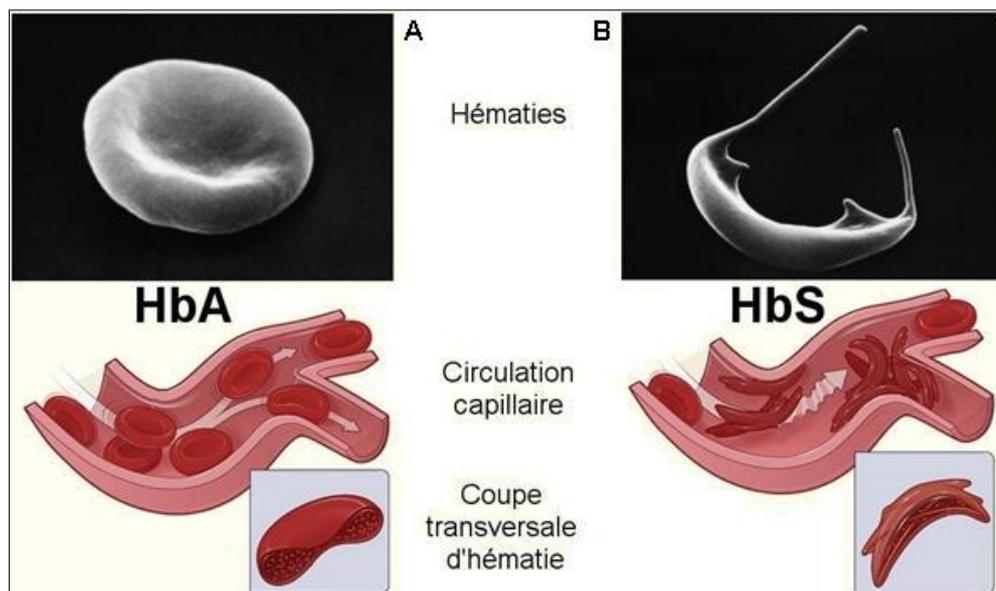


Figure13 : Circulation des globules rouges HbA normaux et HbS anormaux dans les vaisseaux sanguins. (Itim et noui, 2017)

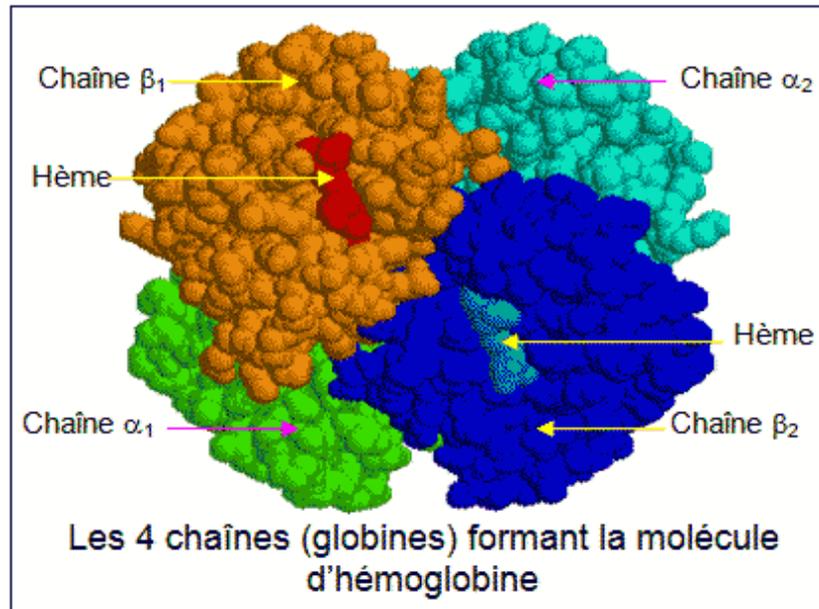


Figure14: Structure tridimensionnelle de la molécule d'hémoglobine adult
(Itim et noui, 2017)

2.3. Description du gène

(Le gène (HBB) sous-unité bêta de l'hémoglobine) se trouve au niveau du chromosome 11, (11p15.4) (**fig.15**), il contient 3 exons et est régulé par un promoteur 5' adjacent, qui contient des boîtes TATA, CAAT et CACCC dupliquées, et un élément régulateur en amont connu sous le nom de région de contrôle du locus (LCR). HBB est contenu dans le cluster de gènes HBB, qui comprend également les gènes codant pour la chaîne delta-globine, les chaînes gamma A et G, et HBBP1 (un pseudo-gène HBB) et epsilon. (**Wong, 2006**)

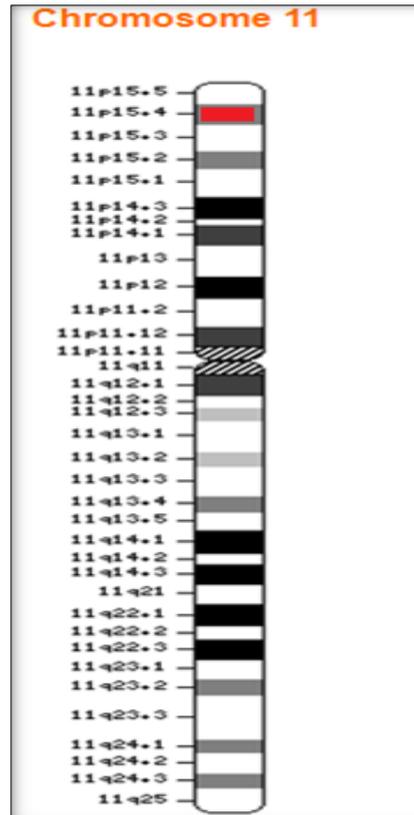


Figure15 : Structure du chromosome 11 et la localisation du gène HBB (**11p15.4**) (Itim et noui, 2017)

2.4. Protéine bêta-globine

Le gène HBB fournit des instructions pour fabriquer une protéine appelée bêta-globine. La bêta-globine est un composant (sous-unité) d'une protéine plus grande appelée hémoglobine (**fig.14**), qui est située à l'intérieur des globules rouges. Chez l'adulte, l'hémoglobine se compose normalement de quatre sous-unités protéiques: deux sous-unités de bêta-globine et deux sous-unités d'une autre protéine appelée alpha-globine, qui est produite à partir d'un autre gène appelé **HBA**. Chacune de ces sous-unités protéiques est attachée (liée) à une molécule contenant du fer appelée hème; chaque hème contient une molécule de fer en son centre qui peut se lier à une molécule d'oxygène. L'hémoglobine dans les globules rouges se lie aux molécules d'oxygène dans les poumons. Ces cellules voyagent ensuite dans la circulation sanguine et fournissent de l'oxygène aux tissus dans tout le corps (**Wong, 2006**)

2.5. Mutations spécifiques de la drépanocytose

La drépanocytose (Hb S/S) causée par la variante homozygote du gène HBB est la cause la plus courante de cette maladie. Elle est caractérisé par la présence de deux allèles d'hémoglobine S (HbS; p.Glu6Val dans HBB) conduisant la substitution de l'acide aminé d'acide glutamique en valine (**fig.16**) et (**fig.17**). Cette substitution entraîne la formation de polymères moléculaires insolubles qui s'agrègent aux globules rouges les déformant (figure 18), les rendant cassants et augmentant l'adhérence à l'endothélium. Cela peut entraîner une occlusion veineuse, une diminution de la perfusion tissulaire et une ischémie; qui contribuent à un dysfonctionnement multi-organique L'hémoglobine polymérisée est également nocive pour la membrane des globules rouges, entraînant une déshydratation cellulaire, des dommages oxydatifs et une adhérence accrue aux cellules endothéliales. Les autres facteurs contribuant à la physiopathologie de la drépanocytose comprennent: la leucocytose (entraînant une production accrue de cytokines nuisibles et une altération du flux sanguin), des anomalies de la coagulation et une régulation vasculaire anormale. Le résultat net de ces anomalies est un raccourcissement de la durée de vie des globules rouges ou une hémolyse et une occlusion vasculaire intermittente et un état d'inflammation chronique. La gravité des manifestations de la maladie varie, même chez les individus présentant les mêmes variantes pathogènes de l'HBB. (Wong, 2006)

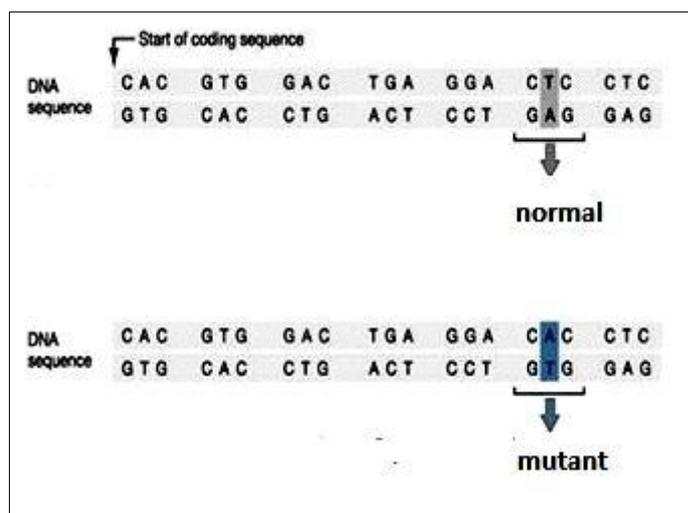


Figure 16 : Drépanocytose au niveau du génotype (Itim et noui, 2017)

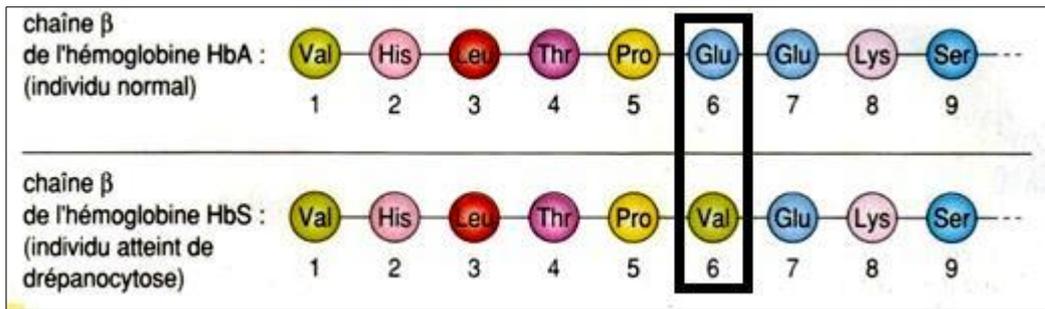


Figure 17 : Drépanocytose au phénotype moléculaire (Itim et noui, 2017)

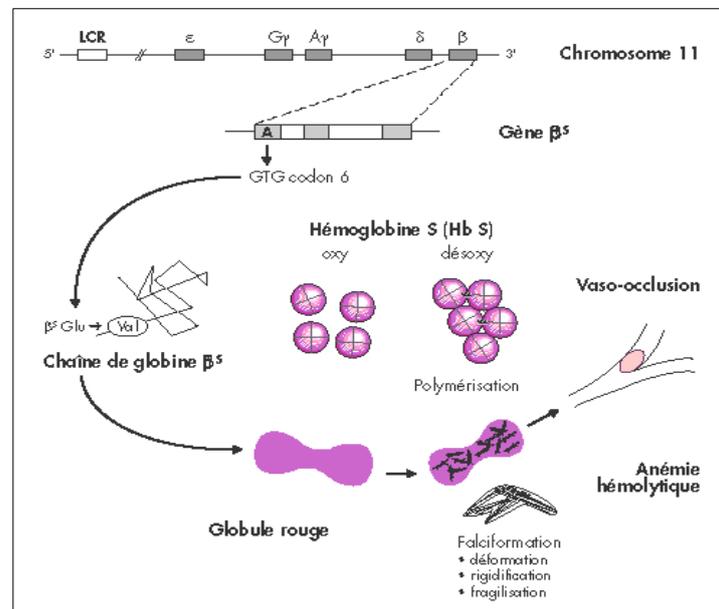


Figure 18 : Schéma physiopathologique de la drépanocytose (Itim et noui,2017)

On explique la polymérisation de la désoxyhémoglobine S par le fait que la valine n°6 est un résidu hydrophobe qui remplace un acide aminé hydrophile, l'acide glutamique. Les globines étant entourées par un film d'eau, la présence d'un site hydrophobe crée un point de "collage" entre 2 molécules d'hémoglobines voisines. Celui-ci s'établit entre la leucine 88 et la phénylalanine 85 d'une chaîne alpha d'une molécule d'hémoglobine et la valine 6 de la chaîne b de l'hémoglobine voisine, d'où création d'une structure cristalline en fibres. (Bender, 2020)

2.6. Mode de transmission

La drépanocytose (SCD) est héritée de manière autosomique récessive. (Wong, 2006)

2.6.1. Echiquiers de Punnett (tableaux de croisement)

Tableau II: Le cas de deux parents hétérozygotes (AS)

Gamètes ♀ \ Gamètes ♂		A	S	
		A	(AA) [A]	(AS) [A]
		S	(AS) [A]	(SS) [S]

A : Allele sain, S : Allele muté

- 25 % des enfants seront homozygotes malades (SS)
- 50 % des enfants seront hétérozygotes porteurs (AS)
- 25 % des enfants seront normaux (AA)

Tableau III : Le cas d'un parent hétérozygote (AS) et l'autre parent normal (AA)

Gamètes ♀ \ Gamètes ♂		A	A	
		A	(AA) [A]	(AA) [A]
		S	(AS) [A]	(AS) [A]

- 50 % des enfants seront porteurs hétérozygotes (AS)
- 50 % des enfants seront normaux (AA)

Tableau IV: Le cas d'un parent homozygote malade (SS) et l'autre parent est normal (AA)

Gamètes ♀ \ Gamètes ♂		A	A
		S	S
S	(AS)	(AS)	
	[A]	[A]	
S	(AS)	(AS)	
	[A]	[A]	

A : Allele sain, S : Allele muté

-100 % des enfants seront hétérozygotes porteurs (AS)

Tableau V: Le cas d'un parent homozygote (SS) et l'autre parent est hétérozygote (AS)

Gamètes ♀ \ Gamètes ♂		A	S
		S	S
S	(AS)	(SS)	
	[A]	[S]	
S	(AS)	(SS)	
	[A]	[S]	

A : Allele sain, S : Allele muté

-50 % des enfants seront homozygotes malades (SS)

-50 % des enfants seront hétérozygotes porteurs (AS)

Tableau VI : Dans le cas de deux parents homozygotes malades (SS):

Gamètes $\begin{matrix} \text{♀} \\ \text{♂} \end{matrix}$		Gamètes	
		S	S
S	S	(SS)	(SS)
	[S]	[S]	[S]
S	S	(SS)	(SS)
	[S]	[S]	[S]

A : Allele sain, S : Allele muté

-100 % des enfants seront homozygotes malades (SS)

Matériel et Méthodes

Notre étude est une étude *in silico* réalisée pendant la période du COVID 19 sur une période de cinq (5) mois allant du mois d'Avril 2020 au mois d'Aout 2020.

Notre principal objectif est une modélisation *in silico* par les outils d'analyses bioinformatique de la technologie de CRISPR des séquences modifiées d'un gène responsable d'une maladie génétique.

Initialement la thématique que nous avons proposée était « *l'étude in silico par technologie CRISPR-cas9 du gène TP53 provoquant l'hépatoblastome* ».

Cependant, nous n'avons pas trouvé un modèle de technologie CRISPR du gène TP53 responsable de l'hépatoblastome qui répond à notre méthodologie *in silico*. Pour cela, notre travail a été réorienté vers le gène HBB responsable de la drépanocytose et dont la thématique s'intitule « *Modélisation in silico de l'implication du gène HBS dans la drépanocytose avec CRISPR-cas9* »

1. Matériel

1.1. Matériel biologique

Les échantillons biologiques de notre travail sont des séquences d'ADN et d'ARNm du gène HBB ayant le polymorphisme « rs 334 ». Les caractéristiques des données sont représentées par le tableau ci-après :

Tableau VII : Séquences et l'exon à étudier :

Séquence	Position	Numéro d'accension NCBI	Variant	SNP
Gène HBB	5225303 -5227 232	NC_000011.10		
ARNmHBB	5 225464 -5 227071	NM_000518.5		
Exon 1	1..142	NM_000518.5	c.20A>T (p.Glu6Val)	rs334

1.2. Matériel non biologique

Notre étude a nécessité l'utilisation du matériel informatique (ordinateur portable) et d'une connexion internet à haut débit.

Afin de réaliser notre étude *in silico* et répondre à notre problématique, nous avons utilisés des outils bio-informatiques pour le traitement de nos échantillons moléculaires et qui sont des bases de données à accès libre (Open Source) nécessitant des interfaces web, des logiciels, des softwares et des algorithmes. Ces outils sont résumés dans le tableau ? ci après.

Tableau VIII : outils bioinformatiques utilisés

Bases de données	But d'utilisation	url
NCBI	Recherche de séquences et études cliniques	https://www.ncbi.nlm.nih.gov/
ORF Finder	Analyse, reconnaissance du cadre de lecture	https://www.bioinformatics.org/sms2/orf_find.html
Swiss model	Analyse, modélisation tridimensionnel des protéines	https://swissmodel.expasy.org/
BLAST (Basic Local Alignment Search Tool)	pour l'alignement des séquences.	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Benchling	Pour faire la conception des composants du CRISPR-Cas9	https://www.benchling.com/crispr/

2. Méthode

Le protocole de notre étude est schématisé par l'organigramme par la (fig.19) ci-dessous.

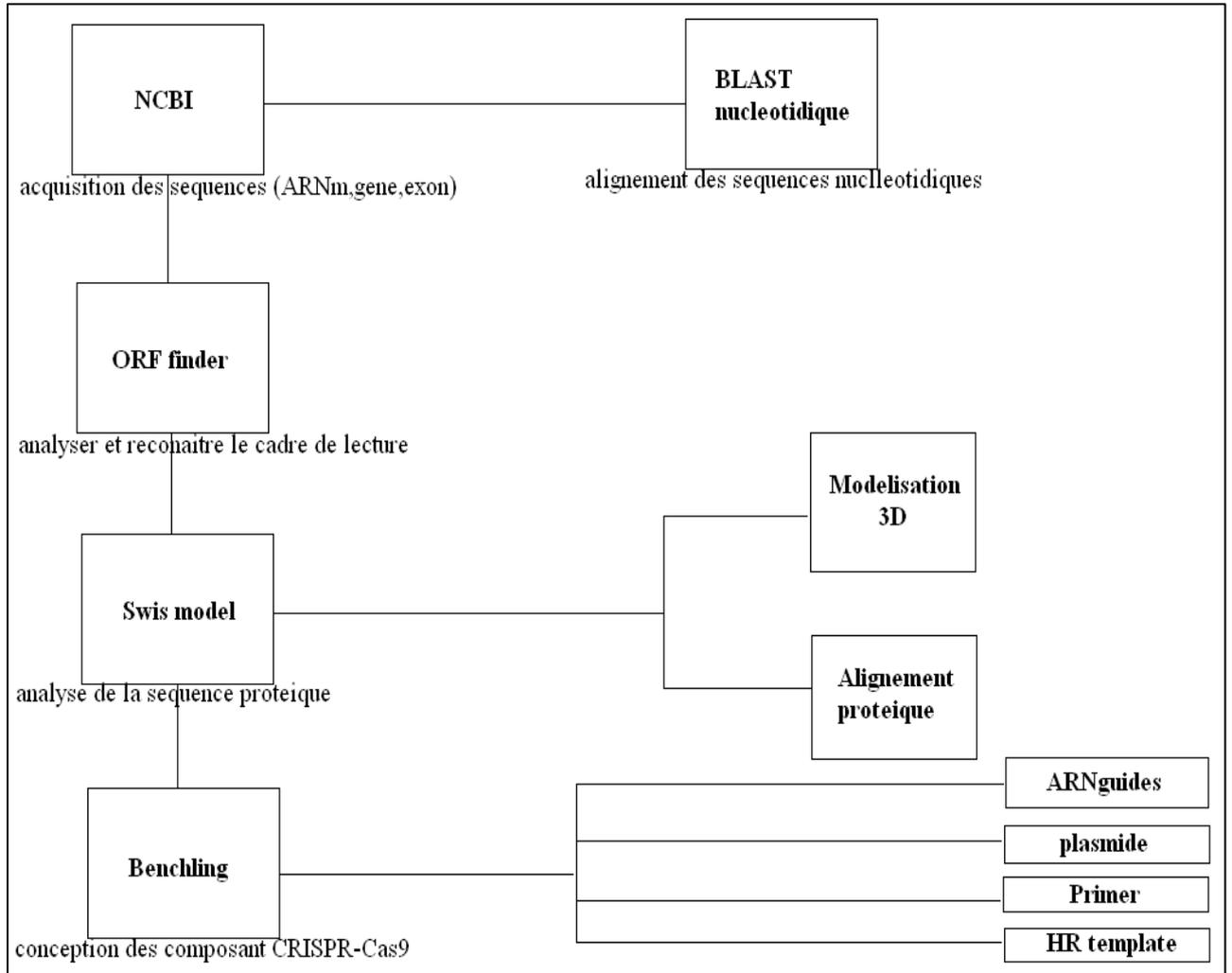


Figure 19: Modélisation in silico de l'implication du gène HBS dans la drépanocytose avec CRISPR-cas9.

2.1. Obtention des données moléculaires

2.1.1 .Base de données NCBI

Principe

Base de données du National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/gene/>) est une division du National Institutes of Health (NIH, institutions nationales pour la recherche médicale et biomédicale des États-Unis). NCBI Gene intègre des informations diverses sur les gènes d'un grand nombre d'espèces. Cette dernière conduit des recherches dans la biologie informatique, développe des logiciels pour analyser des données de génome et fournir des informations biomédicales

Protocol d'analyse

-La première étape consiste à interroger la base de données NCBI et écrire notre requête dans l'onglet recherche du moteur de recherche NCBI pour récupérer et télécharger la séquence de gène HBB et la séquence de l'exon n°1

-La deuxième étape consiste à aller vers le lien nucléotide puis insérer le numéro d'accession.

-Puis nous activons la recherche de la requête

La figure ci-dessous (**fig.20**) illustre notre analyse.

-Une fois que la recherche de la séquence est lancée sur la plate-forme NCBI, après avoir cliqué sur send to; un tableau concernant la requête demandée est affichée sur la page de navigation, d'où on a la possibilité de télécharger les séquences sous format FASTA. (**fig.21**)



Figure20 : Utilisation de genebank (<https://www.ncbi.nlm.nih.gov/>).

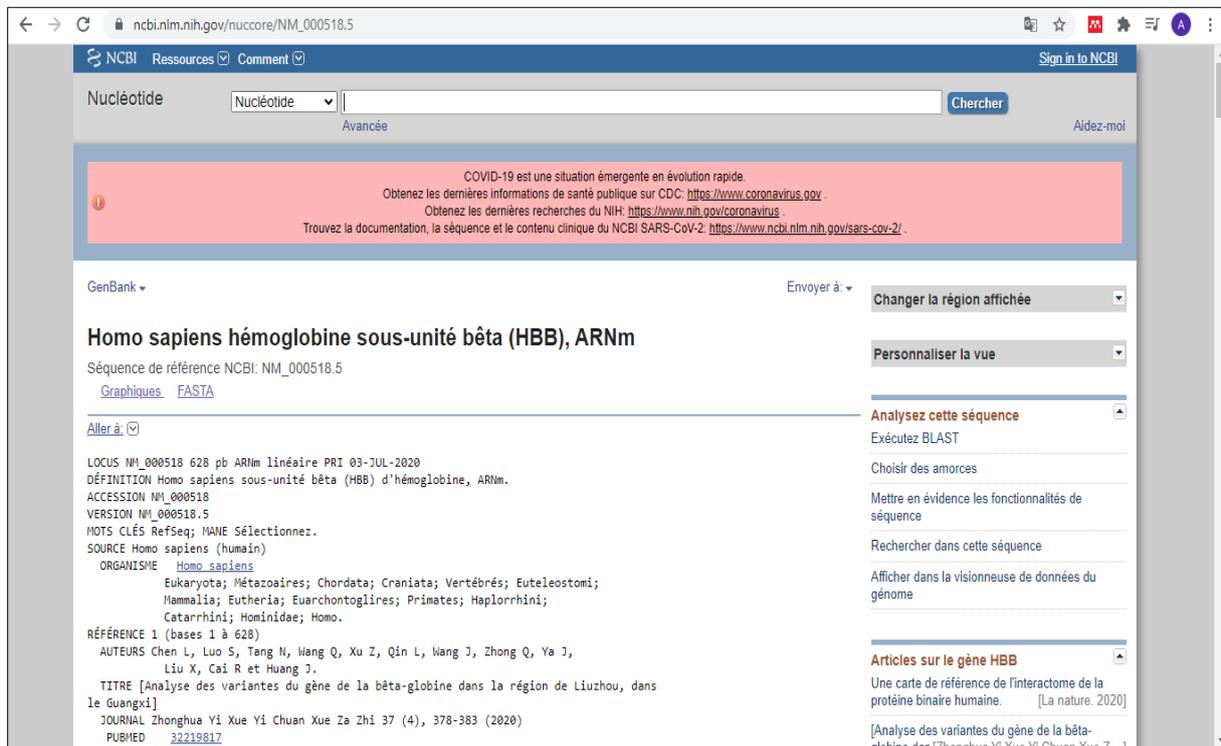


Figure21 : Tableau de bord de la requête sur NCBI

(https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5).

2.2. Alignement par Blast

Principe

Le Blast est utilisé pour l'alignement entre les différentes séquences. Nous avons utilisé pour notre étude le BLASTn et le BLASTp.

Protocol d'analyse

- Nous choisissons le type de blast selon le besoin.
- Nous choisissons le numéro d'accèsion ou bien la séquence sous forme FASTA.
- Nous choisissons la base de comparaison.
- Nous lançons l'algorithme d'alignement. **(Figure22)**

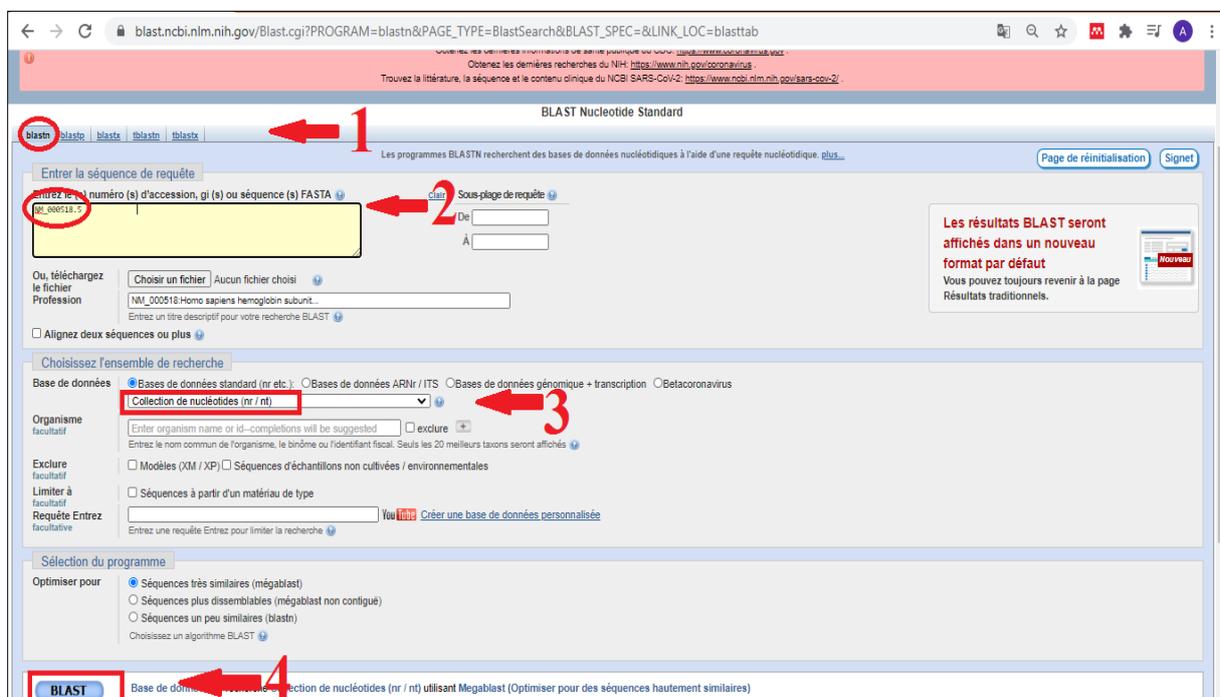


Figure22 : Utilisation du blast sur NCBI

(https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

2.3. ORF Finder

Principe

Algorithme utilisé pour avoir les cadres de lectures ‘Open Frame Read’ de notre séquence.

Protocol d’analyse

On passe sur la page d’accueil d’ORF Finder

- 1-nous insérons la séquence.
- 2-Les ORFs peuvent commencer avec tous les codons.
- 3-nous cherchons les ORFs dans les cadre de lecture 1,2 et 3 du sensdirect/indirect.
- 4-ne renvoie que des ORF d’au moins 60 codons de long.
- 5-nous utilisons le code génétique standard.
- 6-nous lançon la recherche.

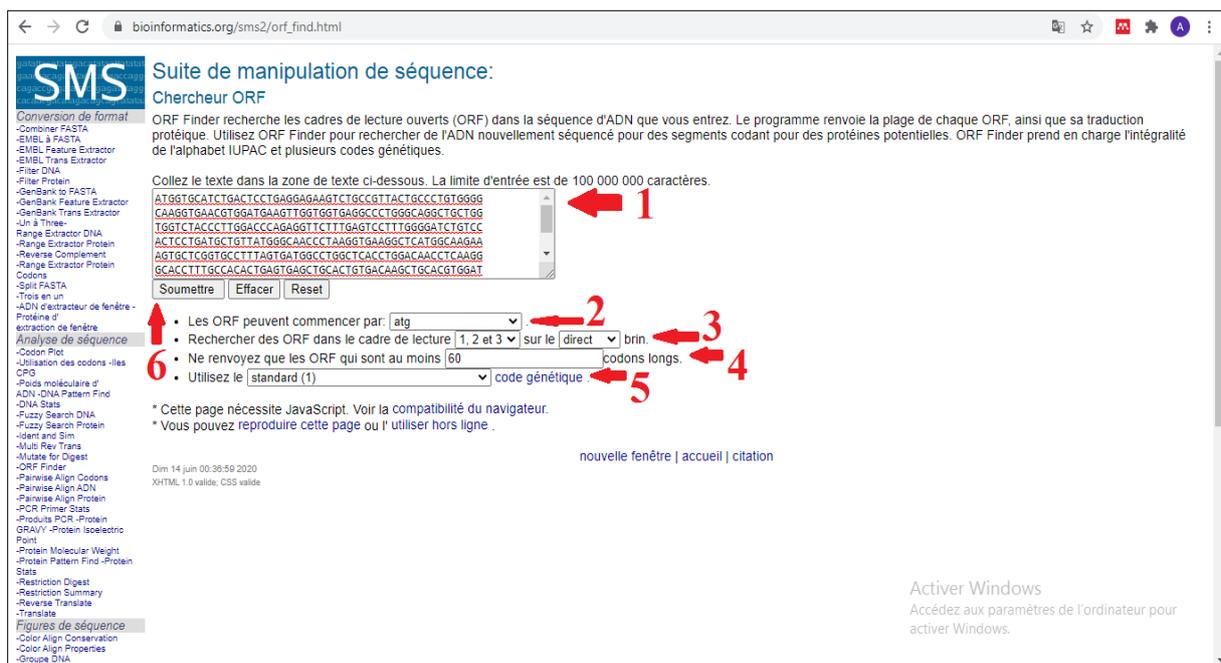


Figure23 : Utilisation d’orf finder

2.4. SWISS-MODEL

Principe

C'est une base de données qui donne des informations concernant notre requête (Structure 3D, séquence protéique, alignement multiple...etc)

Protocol d'analyse

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

(fig.24)

-On insère la séquence protéique ou bien le numéro d'accession.

-Ensuite on lance la recherche du modèle.

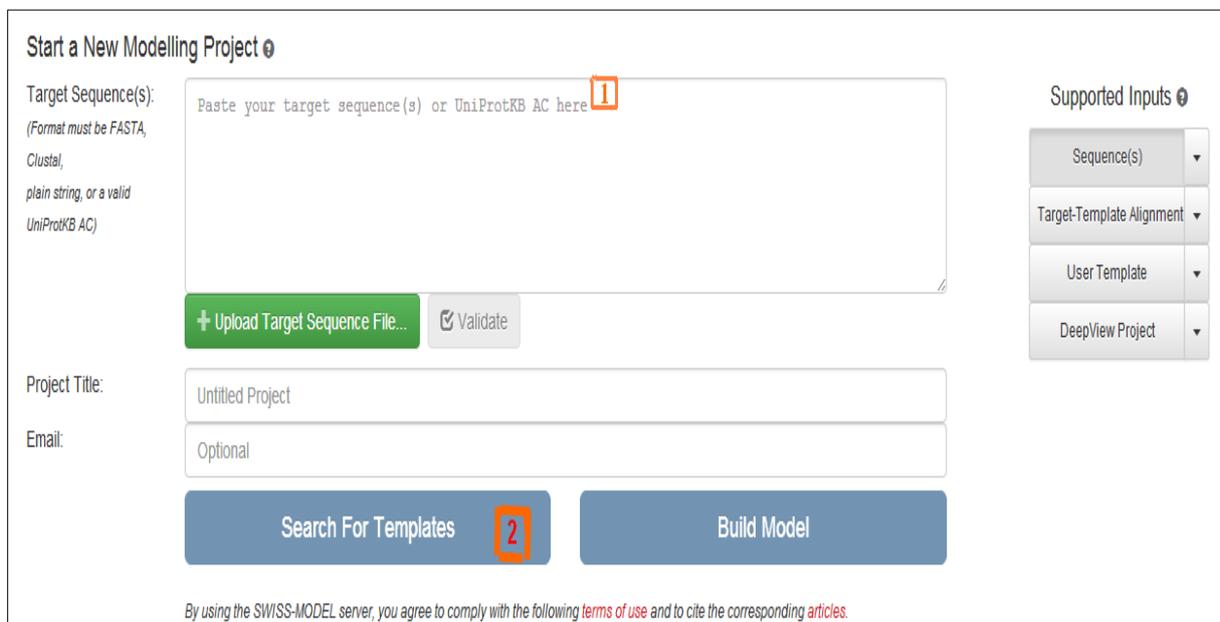


Figure24 : Utilisation de SWISS-MODEL (<https://swissmodel.expasy.org/interactive>)

2.5. Benchling

Principe

C'est une plateforme informatique qui propose des services et outils bioinformatique qui servent à faire un design (conception) 'des ARN guides à partir du gène étudié, aussi à construire le plasmide, les primers (les amorces) et le design de la séquence servant comme matrice pour la réparation homologue.

Protocol d'analyse

-Tout d'abord il faudra créer votre nouveau projet et le nommer. **(Figure 25) (figure26)**

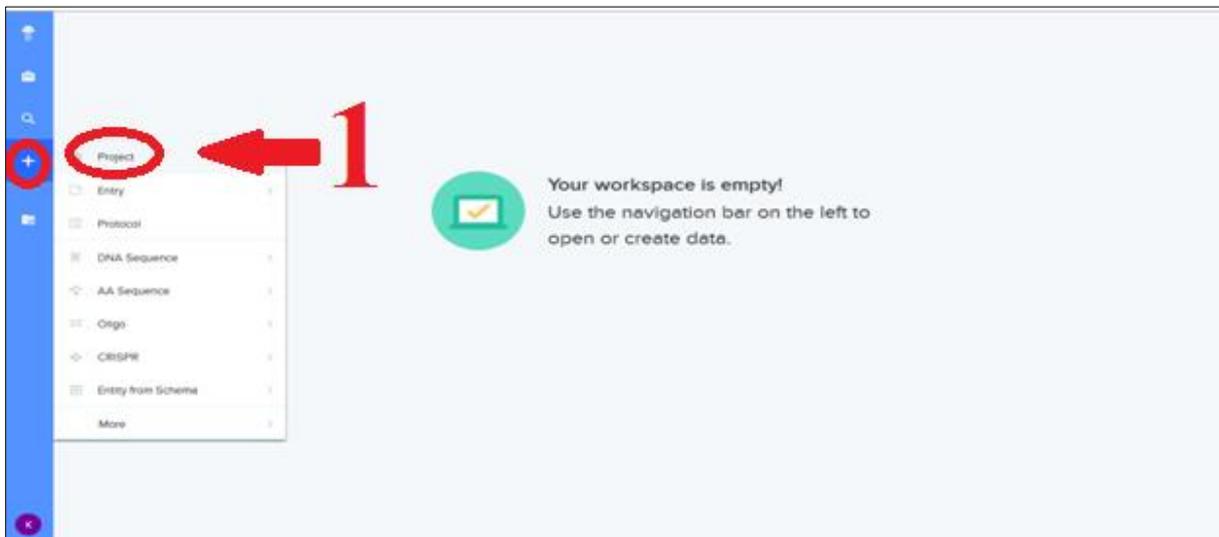


Figure25: Création de projet (<https://www.benchling.com/crispr/>)

Create project

Name* Owner

Description

Manage collaborators

Collaborators

User / Organization	Permissions	Auditor
Khenfir Souhil	ADMIN	<input checked="" type="checkbox"/>

Figure26: Introduire les coordonnées personnelles (<https://www.benchling.com/crispr/>)

- Après avoir introduit vos coordonnées cliquer sur ‘create project’.
- 1-ensuite il faut importer la séquence intérêt de la banque de donné.
- 2-mettre l’identifiant de la séquence d’intérêt (HBS)
- 3- clique sur importer (import)

Résultats

Et

Discussions

1. Résultats

1.1. Résultat 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

La (fig.27) ci-après illustre la description de la séquence ARNm du gène HBB répertorié dans NCBI.

The screenshot shows the NCBI GenBank FASTA page for the Homo sapiens hemoglobin subunit beta (HBB) mRNA sequence. The URL is https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5?report=fasta. The page features a search bar at the top, a navigation menu, and a prominent red banner with COVID-19 information. The main content area displays the FASTA sequence for the HBB gene, with the title "Homo sapiens hémoglobine sous-unité bêta (HBB), ARNm" and the reference sequence identifier "Séquence de référence NCBI: NM_000518.5". The sequence is shown in a monospaced font, with line lengths kept under 80 characters. On the right side, there are several interactive options, including "Analysez cette séquence", "Exécutez BLAST", and "Articles sur le gène HBB".

Figure27 : Séquence d'ARNm du gène HBB sous format FASTA (https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5)

Cette figure nous donne la description de la séquence ARNm du gène HBB répertorié dans NCBI.

1.2. Résultats de Blast n : Après avoir fait le blast nous obtiendrons une liste de prédiction des variants par score. Parmi ces séquence là nous en choisisons celle qui possède la mutation d'intérêt (GAG en GTG)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Homo sapiens hemoglobin subunit beta (HBB) mRNA	1160	1160	100%	0.0	100.00%	NM_000518.5
PREDICTED: Pan troglodytes hemoglobin subunit beta (HBB) mRNA	1155	1155	100%	0.0	99.84%	XM_508242.4
PREDICTED: Pan paniscus hemoglobin subunit beta (LOC100976465) mRNA	1149	1149	100%	0.0	99.68%	XM_003819029.3
Homo sapiens hemoglobin_beta mRNA (cDNA clone MGC.14540 IMAGE.4292125) complete cds	1142	1142	99%	0.0	99.52%	BC007075.1
Human messenger RNA for beta-globin	1140	1140	99%	0.0	99.52%	V00497.1
PREDICTED: Gorilla gorilla gorilla hemoglobin subunit beta (LOC101126932) mRNA	1133	1133	100%	0.0	99.20%	XM_019036164.1
Homo sapiens hemoglobin beta mRNA complete cds	1133	1133	99%	0.0	99.36%	AY509193.1
PREDICTED: Pongo abelii hemoglobin subunit beta (HBB) mRNA	1098	1098	99%	0.0	98.25%	XM_002822127.4
PREDICTED: Nomascus leucogenys hemoglobin subunit beta (HBB) mRNA	1088	1088	100%	0.0	97.93%	XM_004090649.3
PREDICTED: Hylobates moloch hemoglobin subunit beta (HBB) mRNA	1077	1077	100%	0.0	97.61%	XM_032166808.1
Homo sapiens hemoglobin beta chain variant (HBB) mRNA HBB-Dothan allele complete cds	1064	1064	93%	0.0	99.32%	EU694432.1
Homo sapiens hemoglobin beta subunit variant (HBB) mRNA complete cds	1064	1064	92%	0.0	99.83%	AF181989.1
Homo sapiens beta globin chain variant (HBB) mRNA complete cds	1057	1057	92%	0.0	99.48%	AF349114.1
PREDICTED: Trachypithecus francoisi hemoglobin subunit beta (LOC117081249) mRNA	1038	1038	100%	0.0	96.51%	XM_033207068.1
PREDICTED: Colobus angolensis palliatus hemoglobin_beta (HBB) mRNA	1038	1038	100%	0.0	96.50%	XM_011963601.1
PREDICTED: Rhinopithecus roxellana hemoglobin subunit beta (LOC104662339) mRNA	1033	1033	100%	0.0	96.35%	XM_010363344.2
PREDICTED: Rhinopithecus bieti hemoglobin subunit beta (LOC108521664) mRNA	1033	1033	100%	0.0	96.35%	XM_010363344.2

Figure28 : Affichage de la liste de tout les variants du gène HBB.

Sequence ID: [AY136510.1](#) Length: 534 Number of Matches: 1

Range 1: 1 to 534

Score	Expect	Identities	Gaps	Strand
970 bits(525)	0.0	531/534(99%)	0/534(0%)	Plus/Plus

```

Query 51  ATGGTGCATCTGACTCCTGAGGAGAAGTCTGCCGTACTGCCCTGTGGGGCAAGGTGAAC 110
Sbjct 1   ATGGTGCACCTGACTCCTGTGAGAAGTCTGCCGTACTGCCCTGTGGGGCAAGGTGAAC 60
Query 111 GTGGATGAAGTTGGTGGTGAAGCCCTGGCAGGCTGCTGGTGTCTACCTTGACCACAG 170
Sbjct 61  GTGGATGAAGTTGGTGGTGAAGCCCTGGCAGGCTGCTGGTGTCTACCTTGACCACAG 120
Query 171  AGGTTCTTTGAGTCCTTTGGGGATCTGCCACTCCTGATGCTTTATGGGCAACCTAAG 230
Sbjct 121  AGGTTCTTTGAGTCCTTTGGGGATCTGCCACTCCTGATGCTTTATGGGCAACCTAAG 180
Query 231  GTGAAGGCTCATGGCAAGAAAGTCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC 290
Sbjct 181  GTGAAGGCTCATGGCAAGAAAGTCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC 240
Query 291  AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCATGTGACAAGCTGCACGTGGAT 350
Sbjct 241  AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCATGTGACAAGCTGCACGTGGAT 300
    
```

Figure29 : Alignement de la séquence saine et la séquence muté

après avoir choisit la séquence de l'échantillon malade nous apercevrons une homologie de 99% entre HBB et HBS sachant que la différence responsable de la maladie se manifeste au niveau du codon numéro 6.

1.3. Résultat d'orf finder

Présence du codon GAG dans la séquence nucléotidique qui se traduit en **acide glutamique (E)** au niveau de la séquence protéique. (**fig.30**)

```

ORF Finder results
Results for 444 residue sequence "Untitled" starting "ATGGTGCATC"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to b
ATGGTGCATCTGACTCCGAGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAC
GTGGATGAAGTTGGTGGTGGCCCTGGGCAGGCTGCTGGTGGTCTACCCTTGGACCCAG
AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAG
GTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC
AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT
CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGC
AAAGAATTCACCCACCCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAAT
GCCCTGGCCACAAGTATCACTAA

>Translation of ORF number 1 in reading frame 1 on the direct strand.
MVHLTEPKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGRNVLGAFSDGLAHLDNAIKGTATLSELHCDKLHVDPENFRLLGNVLCVLAHFFG
KEFTPPVQAAYQKVVAGVANALAHKYH*

No ORFs were found in reading frame 2.

No ORFs were found in reading frame 3.

```

Figure30 : Traduction de la séquence HBB sans le variant

Cette figure montre bien le variant responsable de la drépanocytose au niveau du codon 6 :
(c.20A>T (p.Glu6Val))

```

ORF Finder results
Results for 444 residue sequence "Untitled" starting "ATGGTGCATC"

>ORF number 1 in reading frame 1 on the direct strand extends from base 1 to base 444.
ATGGTGCATCTGACTCCGTGTEAGAAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAAC
GTGGATGAAGTTGGTGGTTGGCCCTGGGCAGGCTGCTGGTGGTCTACCCCTGGACCCAG
AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAG
GTGAAGGCTCATGGCAAGAAAGTGTCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC
AACCTCAAGGGCACCTTTGCCACACTGAGTGAAGTGCACCTGTGACAAGCTGCACGTGGAT
CCTGAGAACTTCAGGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGC
AAGAATTCCACCCACCAAGTGCAGGCTGCCTATCAGAAAAGTGGTGGCTGGTGTGGCTAAT
GCCCTGGCCCAAGTATCACTAA

>Translation of ORF number 1 in reading frame 1 on the direct strand.
MVHLTVKSAVTALWGKVNVDVGGGALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH*

No ORFs were found in reading frame 2.
No ORFs were found in reading frame 3.

```

Figure31 : Traduction de la séquence mutée

La substitution du nucléotide A en T du codon 6 dans l'exon 1, est traduite par une substitution de l'acide aminé **acide glutamique** par l'acide aminé **valine** dans la séquence protéique.

1.4. Résultat de SWISS-PROT

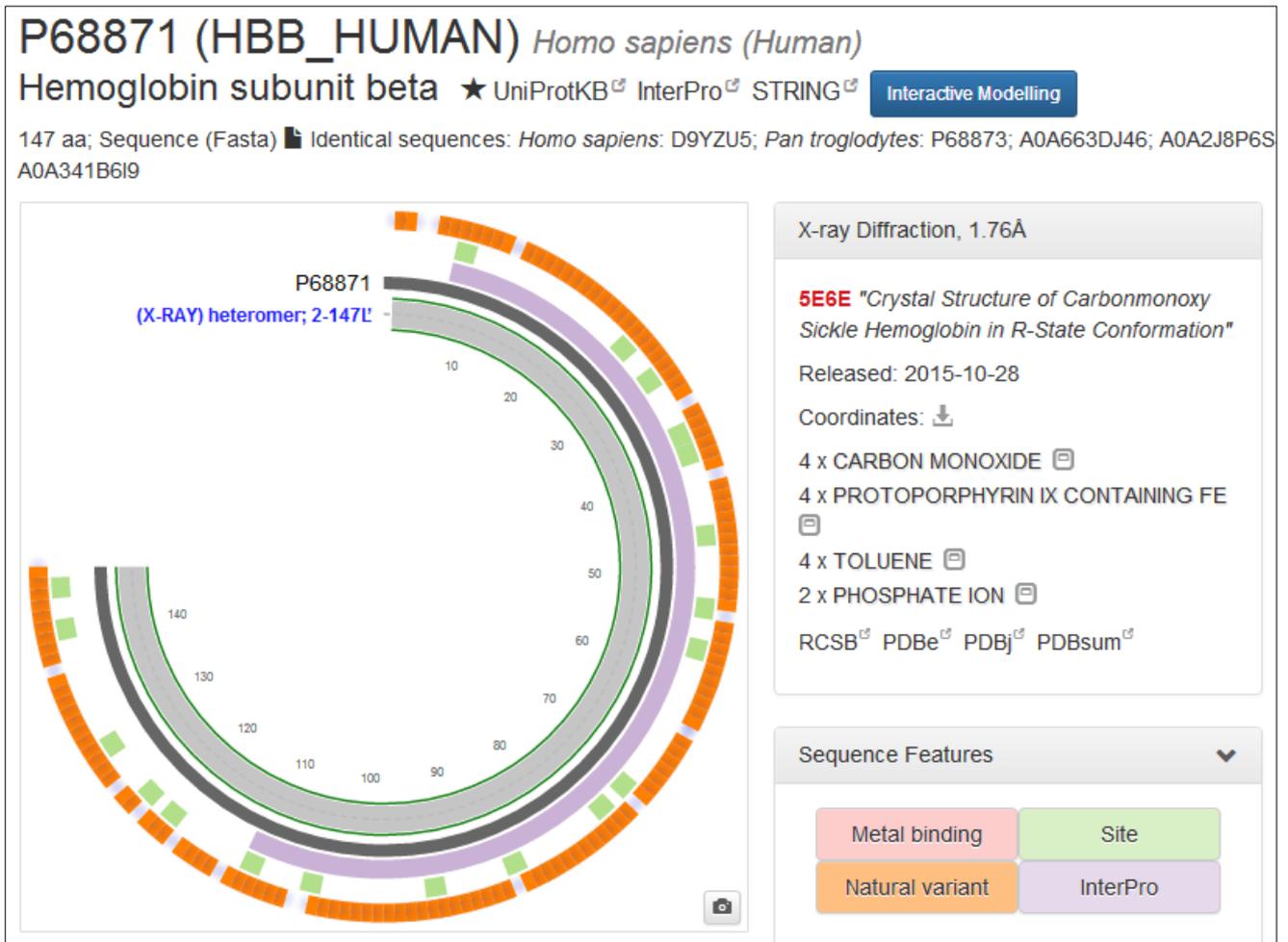


Figure32 : modélisation moléculaire représentant l'homologie de structure protéique entre la protéine normale HBB et la protéine mutée HBS

Pour la modélisation par l'outil SWISS PROT nous avons utilisé l'ARNm du gène HBB. Cette modélisation a générée l'identifiant : P68871. **(Fig.32)** ci-après montre la représentation topographique des différents isoformes du gène HBB. Nous pouvons voir ainsi, les caractéristiques de la séquence du gène HBB et les différents motifs de la molécule qui sont mis en évidence par un code couleur qui permet de distinguer chaque domaine et de comparer son absence/présence au niveau des autres protéines.

1.4.1. Séquence protéique du gène HBB

Sequence Alignments ^

Hemoglobin subunit beta P68871 UniProtKB InterPro

Toggle Identical (BD)

P68871	VHLTPEEKSAVTALWGKVNVEVGGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGA	71
5e6e. (BD)	VHLTPVEKSAVTALWGKVNVEVGGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGA	70
P68871	FSDGLAHLNLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLCVLAHFGKEFTPPVQAAAYQKVVAGVANA	141
5e6e. (BD)	FSDGLAHLNLDNLKGTFFATLSELHCDKLHVDPENFRLLGNVLCVLAHFGKEFTPPVQAAAYQKVVAGVANA	140
P68871	LAHKYH	147
5e6e. (BD)	LAHKYH	146

Figure33 : Séquence protéique du gène HBB

1.4.2. Structure tridimensionnelle de la protéine HBB

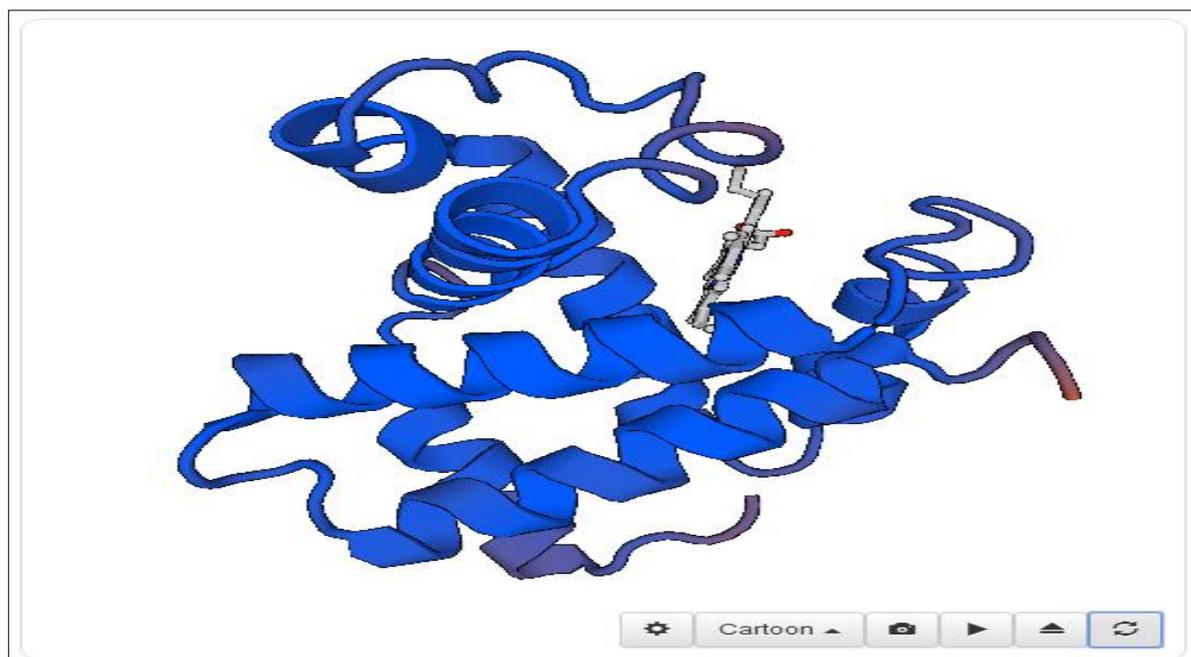


Figure34 : Structure 3D de la protéine du gène HBB

1.4.3. Structure tridimensionnelle de l'hémoglobine

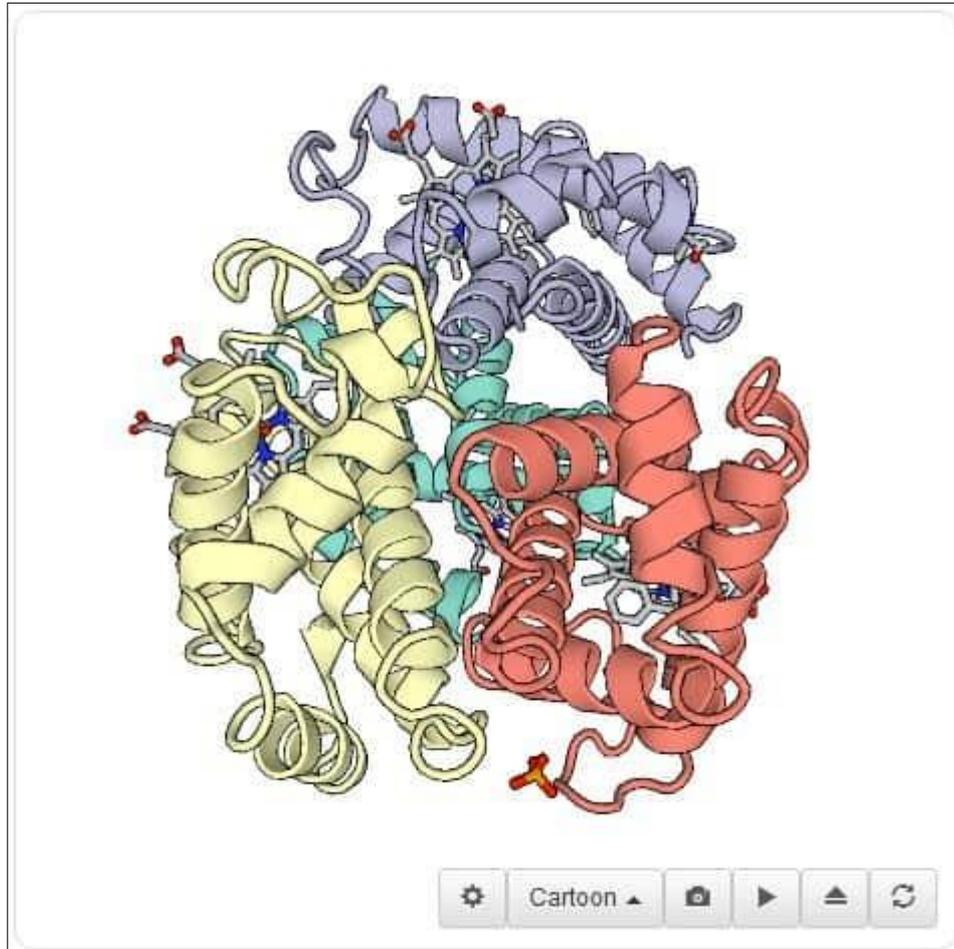


Figure35 : Structure 3D de l'hémoglobine

La couleur bleu représente la protéine du gène HBB dans la structure 3D de l'hémoglobine (**fig35**).

Pour savoir la position d'un n'importe quel acide aminé sur la structure Tridimensionnel (**Fig33**) on clique sur l'acide aminé, il va apparaître sur la structure.

1.4.4. Alignement de la séquence protéique (HBB) le variant (HBS)

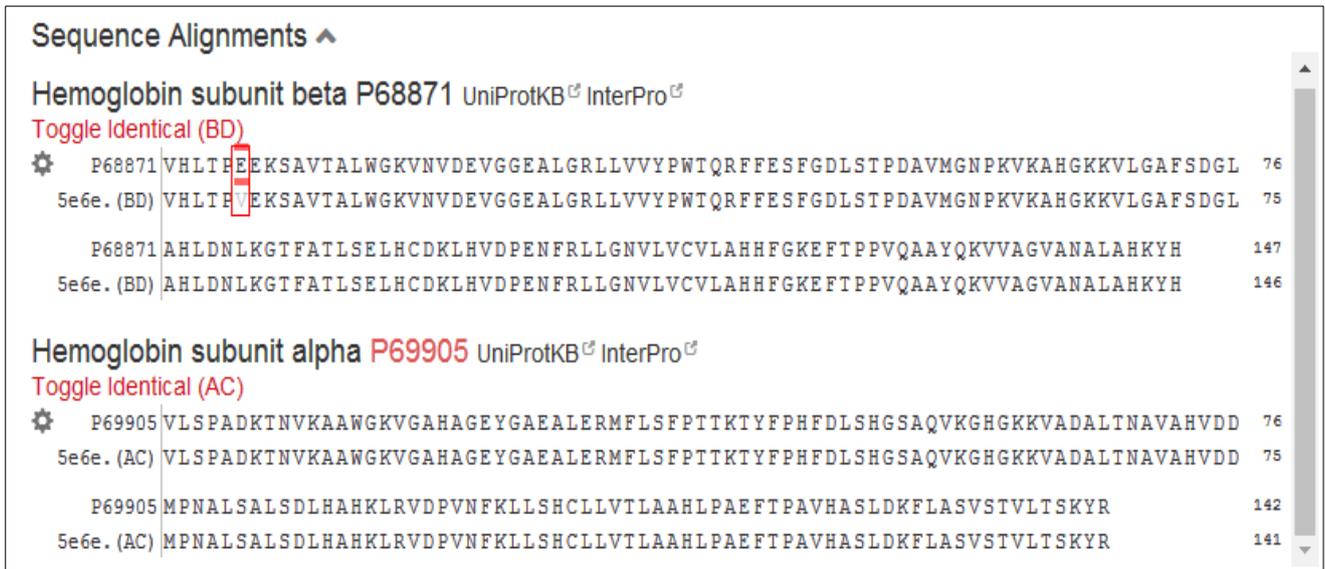


Figure36 : Alignement de la séquence protéique du gène HBB et HBS

L’alignement a détecté le changement de l’acide aminé E (acide glutamique) en V (Valine) ce qui confirme que la mutation A en T est effectivement responsable de la maladie (SCD) nous apercevons aussi qu’il y’a une homologie entre la séquence HBA de l’échantillon malade ce qui affirme que le gène HBA n’est pas responsable de la maladie.

Juste après avoir visualisé les structures protéiques en 3D, nous allons passer à plus d'informations dans la même plate-forme mais avec un autre outil : Swiss-prot

The screenshot shows the UniProtKB/Swiss-Prot variant page for P68871: Variant p.Glu7Val. The page is titled "UniProtKB/Swiss-Prot P68871: Variant p.Glu7Val" and is for the Hemoglobin subunit beta gene (HBB). The variant information is as follows:

- Variant position:** 7
- Type of variant:** Disease [Disclaimer]
- Residue change:** From Glutamate (E) to Valine (V) at position 7 (E7V, p.Glu7Val).
- Physico-chemical properties:** Change from medium size and acidic (E) to medium size and hydrophobic (V)
- BLOSUM score:** -2
- Involvement in disease:** Sickle cell anemia (SKCA) [MIM:603903]. Characterized by abnormally shaped red cells resulting in chronic anemia and periodic episodes of pain, serious infections and damage to vital organs. Normal red blood cells are round and flexible and flow easily through blood vessels, but in sickle cell anemia, the abnormal hemoglobin (called Hb S) causes red blood cells to become stiff. They are C-shaped and resembles a sickle. These stiffer red blood cells can lead to microvascular occlusion thus cutting off the blood supply to nearby tissues. [ECO:0000269|PubMed:1195378, ECO:0000269|PubMed:13464827, ECO:0000269|PubMed:16001361, ECO:0000269|PubMed:24100324, ECO:0000269|Ref.10]. Note=The disease is caused by mutations affecting the gene represented in this entry.
- Variant description:** In SKCA, Hb S, at heterozygosity confers resistance to malaria.
- Other resources:** Variant rs334 [dbSNP | Ensembl]

Figure37 : Informations supplémentaires du variant (HBS)= E7V=E6V

Le variant E6V se situe à la position 7 (plus le codon d'initiation) acide glutamique en valine qui conduit à une maladie qui s'appelle la drépanocytose qui est caractérisé par une structure anormale des globules rouge (la forme faucille).

La forme hétérozygote de ce variant confère une résistance contre la malaria.

1.5. Résultats de STRING

La dernière étape consiste à présenter une vue générale des différentes interactions de la protéine HBB grâce a STRING sur la même interface (expasy).

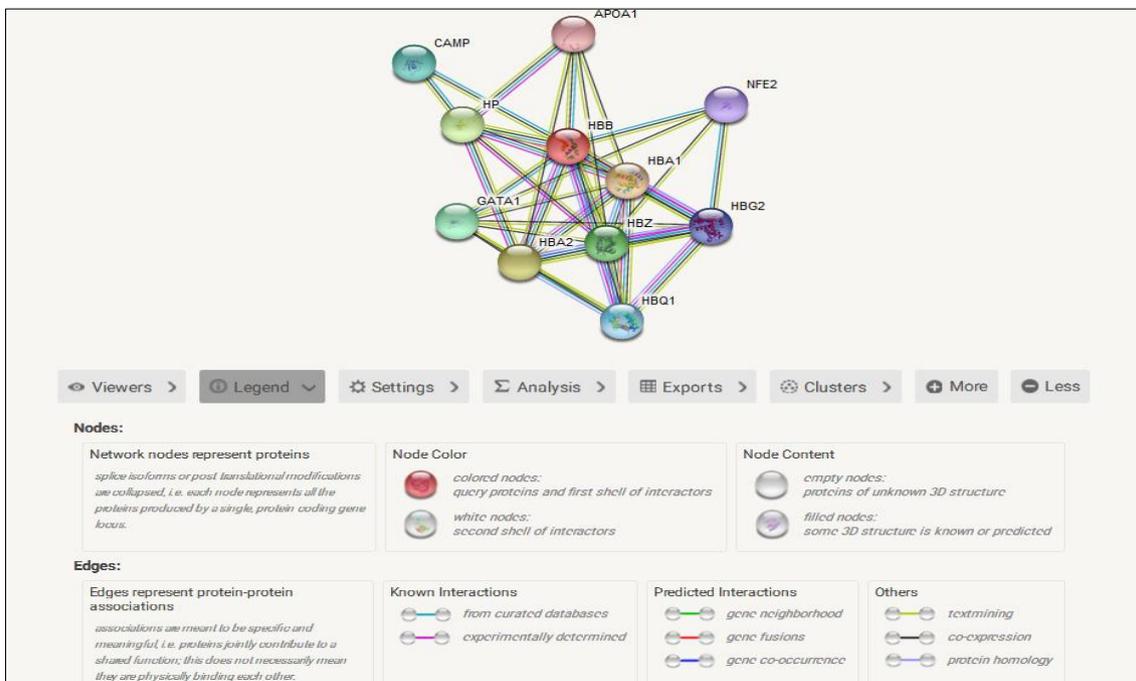


Figure38 : Interaction du gène HBB avec d'autres gènes

On peut visualiser les différentes Interactions gène-gènes ou bien protéines-protéines, chaque couleur représente un composant et/ou une interaction spécifique

-Il est possible aussi d'acquérir une prédiction du rôle de ces interactions

Your Input:		Neighborhood	Gene Fusion	Cocurrence	Coexpression	Experiments	Databases	Textmining	[Homology]
● HBB	Hemoglobin subunit beta; Involved in oxygen transport from the lung to the various peripheral tissues; Belongs to the globin family (147 aa)								
Predicted Functional Partners:									
● HBA1	Hemoglobin subunit alpha; Involved in oxygen transport from the lung to the various peripheral tissues; Belongs to the globin fa...				●	●	●	●	●
● HBA2	Hemoglobin subunit alpha 2; Involved in oxygen transport from the lung to the various peripheral tissues; Hemoglobin subunits ...				●	●	●	●	●
● HP	Haptoglobin; As a result of hemolysis, hemoglobin is found to accumulate in the kidney and is secreted in the urine. Haptoglobi...				●	●	●	●	●
● HBZ	Hemoglobin subunit zeta; The zeta chain is an alpha-type chain of mammalian embryonic hemoglobin (142 aa)				●	●	●	●	●
● GATA1	Erythroid transcription factor; Transcriptional activator or repressor which probably serves as a general switch factor for erythr...				●	●	●	●	●
● CAMP	Cathelicidin antimicrobial peptide; Binds to bacterial lipopolysaccharides (LPS), has antibacterial activity; Endogenous ligands (...)				●	●	●	●	●
● HBQ1	Hemoglobin subunit theta 1 (142 aa)				●	●	●	●	●
● HBG2	Hemoglobin subunit gamma-2; Gamma chains make up the fetal hemoglobin F, in combination with alpha chains; Belongs to th...				●	●	●	●	●
● NFE2	Transcription factor NF-E2 45 kDa subunit; Component of the NF-E2 complex essential for regulating erythroid and megakaryo...				●	●	●	●	●
● APOA1	Apolipoprotein A-I; Participates in the reverse transport of cholesterol from tissues to the liver for excretion by promoting chol...				●	●	●	●	●

Figure39 : Prédiction du rôle des interactions de la protéine HBB

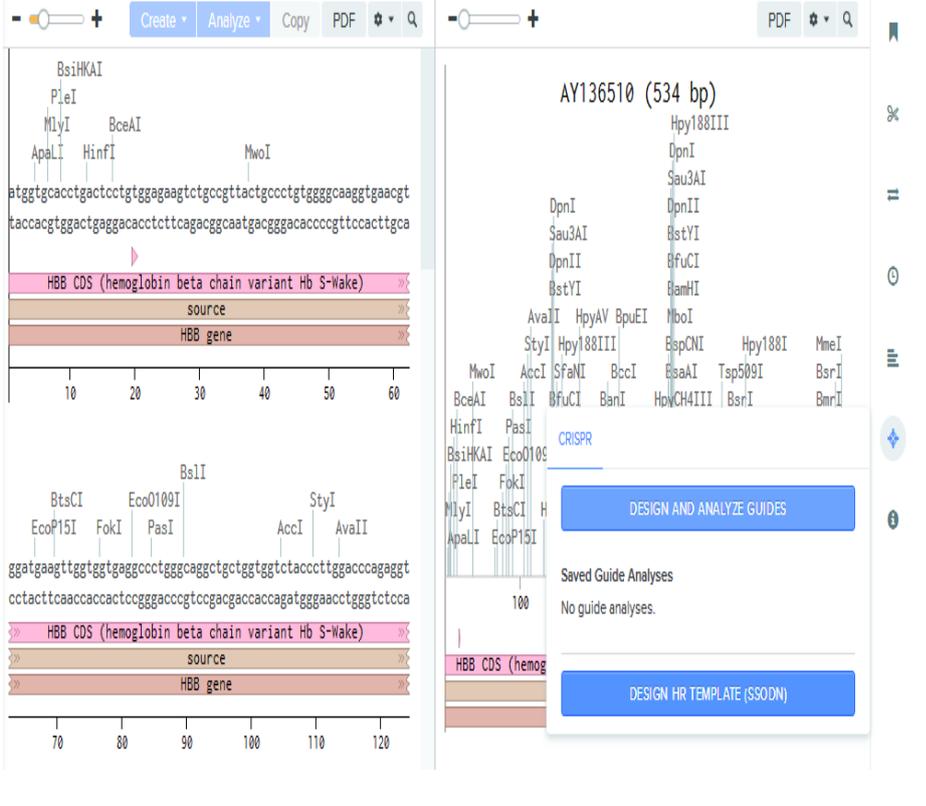
HBB est primordial pour le transport d'oxygène des poumons jusqu'aux tissus périphériques, sur la figure est représenté les prédictions des fonctions du reste de la famille des sous unités d'hémoglobine

1.6. Résultats de Benchling

1.6.1. Design des ARN guides

Après avoir eu la séquence il faut commencer le design des ARN guides, (DESIGN AND ANALYZE GUIDES), les deux ARN guide qu'on va construire doivent être complémentaires à la la région d'intérêt, L'outil du design va nous prédire tout les guides possible de la séquence HBS de taille 543pb (qui cause la drépanocytose).

Tableau IX : les résultats du design des ARN guides.

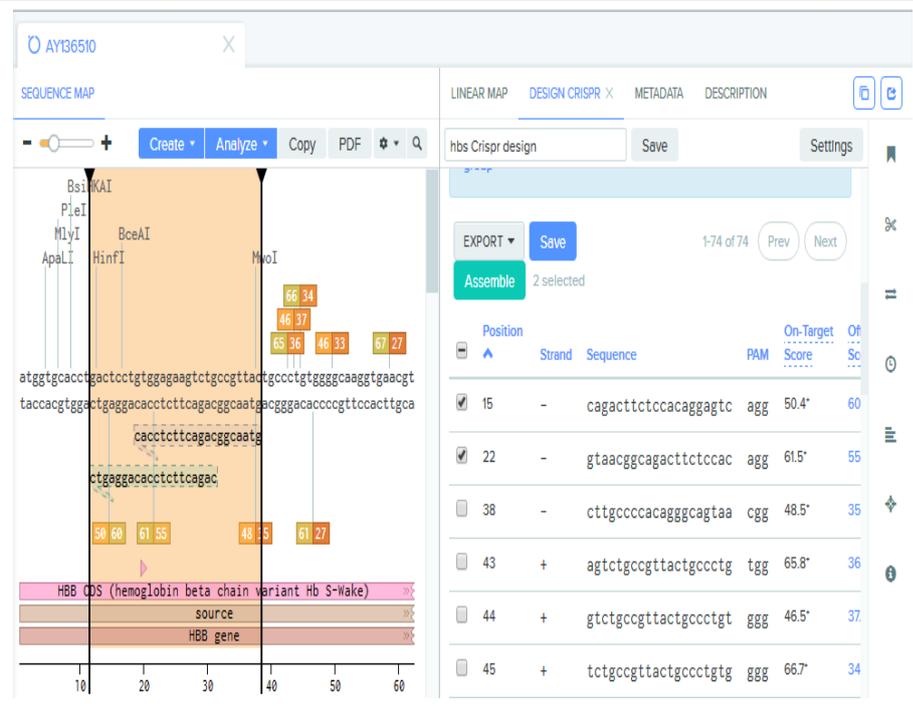
Résultat de du design des ARN guides	titre
	<p>Sélection du design de l'ARNg. (DESIGN AND ANALYZE GUIDES)</p>

<p>Design CRISPR Guides: Guide parameters</p> <p>Design Type</p> <ul style="list-style-type: none"> <input checked="" type="radio"/> Single guide Wild-type Cas9, single gRNA (higher efficiency) <input type="radio"/> Paired guides Double Cas9 nickase, two gRNAs (lower off-target effects) <input type="radio"/> Guides for "base editing" (Komor et al., 2016) C → T (or G → A) substitution, no dsDNA breaks <p>Guide Length: 20</p> <p>Genome: GRCh38 (hg38, Homo sapiens)</p> <p>PAM: NGG (SpCas9, 3' side)</p> <p>Advanced Settings</p> <p><input checked="" type="checkbox"/> Save these as my default CRISPR settings</p> <p>Finish</p>	<p>Insertion des coordonnées du design des guides</p>
<p>SEQUENCE MAP</p> <p>LINEAR MAP DESIGN CRISPR X METADATA DESCRIPTION</p> <p>Untitled Save Settings</p> <p>To get started, create a target region by selecting it on the sequence map and pressing </p> <p>Target: 1 444 </p> <p>Region</p> <p>Target 1: 444</p> <p>Region</p> <p>HBB CDS (hemoglobin beta chain variant Hb S-Wake)</p> <p>source</p> <p>HBB gene</p> <p>HBB CDS (hemoglobin beta chain variant Hb S-Wake)</p> <p>source</p> <p>HBB gene</p>	<p>Ciblage de la séquence intérêt et création des guides</p>



Position	Strand	Sequence	PAM	On-Target Score	Off-Target Score
15	-	cagacttctccacaggagtc	agg	50.4*	60.8
22	-	gtaacggcagacttctccac	agg	61.5*	55.6
38	-	ctgtcccccacaggcagtaa	cgg	48.5*	35.1
43	+	agtctgcggttactgcctgt	tgg	65.8*	36.3
44	+	gtctgcggttactgcctgt	ggg	46.5*	37.6
45	+	tctgcggttactgcctgtg	ggg	66.7*	34.3
47	-	cagttcaccttccccaca	ggg	61.1*	27.2

Affichage de la liste des ARNg potentiels



Choix des deux ARNg

Les ARN guides sont essentiels car leur rôle est d'orienter la Cas9 pour cibler la séquence d'intérêt dans le génome, sans eux il sera impossible de la trouver, ils doivent être

complémentaires et sont choisis selon un nombre maximal de off targets et on targets, pour que l'enzyme coupe exactement là où nous voulons, et pour que la Cas9 ne fasse pas de faute qui peuvent engendrer des défauts de cadre de lecture. Nous avons choisis deux régions des guides proches et qui englobe notre mutation. Sans oublier la séquence PAM qui est capital pour la reconnaissance du début de notre région d'intérêt, nous avons choisis NGG car elle propose une reconnaissance optimale avec la Cas9

1.6.2. Conception du plasmide

Après avoir sauvegardé on clique sur Assemble.

-ici nous choisirons d'importer un plasmide, le px330 qui est le plus travaillé chez les mammifères.

Le px330 : pX330 Plasmide de laboratoire de Zhang pour exprimer un ARN guide chimérique (ARNg) avec Cas9, pour les codons humains. Aussi connu sous le nom de pX330-U6-Chimeric_BB-CBh-hSpCas9.

Tableau X : résultat du design du plasmide.

Résultat		Titre																			
<table border="1"> <thead> <tr> <th>Position</th> <th>Strand</th> <th>Sequence</th> <th>PAM</th> <th>On-Target Score</th> <th>Off-Target Score</th> </tr> </thead> <tbody> <tr> <td>15</td> <td>-</td> <td>cagacttctccacaggagtc</td> <td>agg</td> <td>50.4*</td> <td>60</td> </tr> <tr> <td>22</td> <td>-</td> <td>gtaacggcagacttctccac</td> <td>agg</td> <td>61.5*</td> <td>55</td> </tr> </tbody> </table>		Position	Strand	Sequence	PAM	On-Target Score	Off-Target Score	15	-	cagacttctccacaggagtc	agg	50.4*	60	22	-	gtaacggcagacttctccac	agg	61.5*	55	Initiation de l'assemblage des guides	
Position	Strand	Sequence	PAM	On-Target Score	Off-Target Score																
15	-	cagacttctccacaggagtc	agg	50.4*	60																
22	-	gtaacggcagacttctccac	agg	61.5*	55																

	<p>Choix du plasmide et les régions d'insert et de terminaison</p>
	<p>Visualisation circulaire du plasmide et ces composants</p>

Le plasmide comporte tout nos composants essentiels: 2 enzymes Cas9, 2 ARN guides, et différents enzymes importantes à l'édition du génome, donc sa conception doit y être, parce qu'il est essentiel pour la transfection, bien entendu après avoir sauvegardé notre plasmide, nous pourrons ensuite le commander dans un kit complet des laboratoires spécialisés (zhang lab).

1.6.3. Conception des primers (les amorces)

Après la conception des plasmides nous passerons à la conception des primers (les amorces).

Une fois le plasmide acquis il doit être amplifié pour un maximum de rendement, et obtenir les résultats souhaités, cela grâce aux amorces, qui permettent le clonage des plasmides à partir d'un seul, les primers aussi font partie du kit commandé.

Tableau XI : design des primers (sens et anti sens)

Résultat	titre																														
<p>The screenshot shows the Benchling web interface for primer design. The main window displays a sequence map for AY136510 with various restriction sites. A context menu is open over the sequence, showing options like 'Create Primer...' with 'Forward' and 'Reverse' sub-options. The primer design table is visible on the right side of the interface.</p> <table border="1"> <thead> <tr> <th>Position</th> <th>Strand</th> <th>Sequence</th> <th>PAM</th> <th>On-Target Score</th> <th>Off-Target Score</th> </tr> </thead> <tbody> <tr> <td>15</td> <td>-</td> <td>cagacttctccacaggagtc</td> <td>agg</td> <td>50.4*</td> <td>66.7</td> </tr> <tr> <td>22</td> <td>-</td> <td>gtaacggcagacttctccac</td> <td>agg</td> <td>61.5*</td> <td>63.1</td> </tr> <tr> <td>38</td> <td>-</td> <td>cttgcccacagggcagctaa</td> <td>agg</td> <td>48.5*</td> <td>38.3</td> </tr> <tr> <td>43</td> <td>+</td> <td>agttgcggttactgccttg</td> <td>agg</td> <td></td> <td></td> </tr> </tbody> </table>	Position	Strand	Sequence	PAM	On-Target Score	Off-Target Score	15	-	cagacttctccacaggagtc	agg	50.4*	66.7	22	-	gtaacggcagacttctccac	agg	61.5*	63.1	38	-	cttgcccacagggcagctaa	agg	48.5*	38.3	43	+	agttgcggttactgccttg	agg			<p>Conception pour la création du brin sens (forward) et le brin anti sens (reverse).</p>
Position	Strand	Sequence	PAM	On-Target Score	Off-Target Score																										
15	-	cagacttctccacaggagtc	agg	50.4*	66.7																										
22	-	gtaacggcagacttctccac	agg	61.5*	63.1																										
38	-	cttgcccacagggcagctaa	agg	48.5*	38.3																										
43	+	agttgcggttactgccttg	agg																												

SEQUENCE MAP

LINEAR MAP DESIGN CRISPR X DESIGN PRIMER X DESCRIPTION METADATA

Single Primer

Strand Forward Reverse

Bases 5' `cgc-gactctcttggagagctcccttca` 3'

Saved primers exist with the same bases:

forward primer (houari / HBS VARIANT)

3' Location 38

Overhang 0

Cut Site

Verify at 37 °C

T_m 66.4°C

GC Content 58.06%

Length 31 bp

Activater Windows

Accédez aux paramètres de l'ordinateur pour activer Windows.

BASES 534 START 12 END 38 LENGTH 27 GC 55.56% MELTING TEMP 62.11°C

Représentation du primer (amorce) du brin sens

SEQUENCE MAP

LINEAR MAP DESIGN CRISPR X DESIGN PRIMER X DESCRIPTION METADATA

Single Primer

Bases 5' `aac-gaacccagcactttccacagagctagg` 3'

Saved primers exist with the same bases:

reverse primer (houari / HBS VARIANT)

3' Location 8

Overhang 0

Cut Site

Verify at 37 °C

T_m 67.6°C

GC Content 51.43%

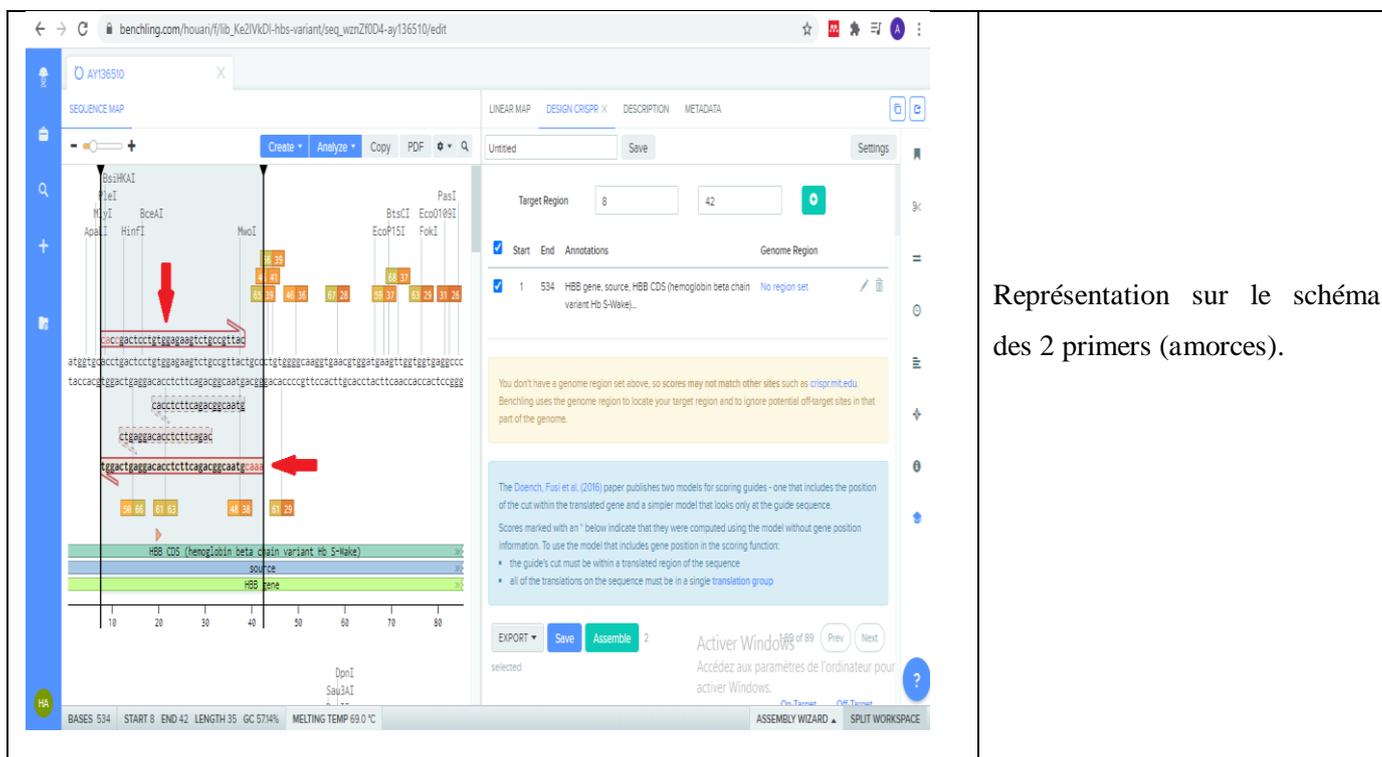
Length 35 bp

Activater Windows

Accédez aux paramètres de l'ordinateur pour activer Windows.

BASES 534 START 8 END 38 LENGTH 31 GC 54.84% MELTING TEMP 65.6 °C

Représentation du primer (amorce) du brin anti-sens.



Représentation sur le schéma des 2 primers (amorces).

Comme il a été décrit précédemment le rôle des amorces (sens et anti sens) est d'amplifier les plasmides pour plus de rendement, donc leur conception est primordiale pour l'expérimentation, Tout d'abord nous allons rajouter à l'extrémité 3' (AAAC) pour diminuer l'activité hors cible, puis à l'extrémité 5' aussi ajouterons (CACC) pour le clonage, cela améliore la transcription médiée par le promoteur U6 (<https://benchling.com/protocols/5DmqRd/crispr-mediated-gene-disruption-in-ch12f3-2-cells/sbs>)

Tableau XII : les sequence des 2 amorces

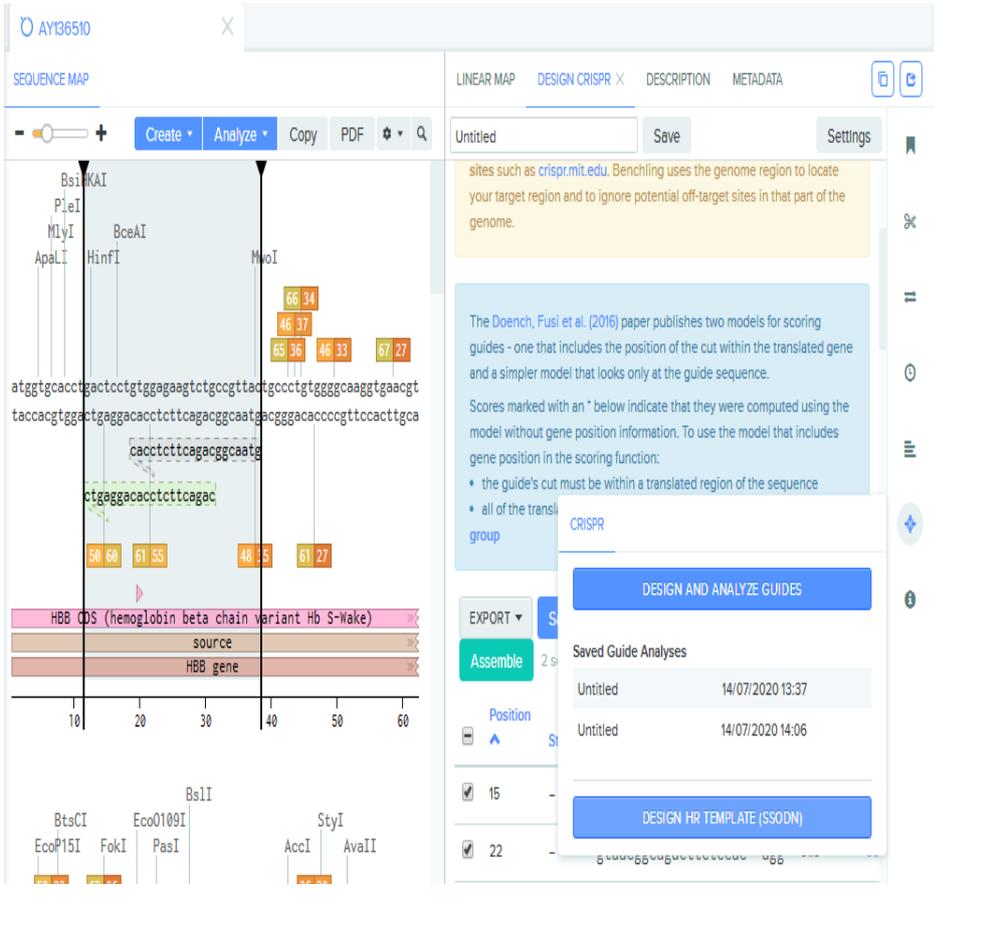
Ce tableau montre les séquences des amorces des deux brins :

Les amorces	La séquence
Brin sens	5'CACCGACTCCTGTGGAGAAGTCTGCCGTTAC'3
Brin anti-sens	5'CAAAGTAACGGCAGACTTCTCCACAGGAGTCAGGT'3

1.6.4. Design de l'ADN recombinant

Vers la fin la séquence coupée par la Cas9 possédant le codon muté (GTG) doit être remplacé par un recombinant homologue possédant le codon sain (GAG), pour cela benchling offre un outil de conception de ce dernier, grâce à la barre à droite nous choisissons le design des homologues recombinants. (DESIGN HR TEMPLATE) (SSODN=single-stranded donor oligonucleotides)

Tableau XIII: design de l'ADN recombinant

Résultat	titre
 <p>The screenshot displays the Benchling interface for a CRISPR design project. On the left, a 'SEQUENCE MAP' shows the HBB gene structure with various restriction enzyme sites (BsiKI, PstI, MlyI, BceAI, ApaI, HinfI, MhoI, BtsCI, EcoO109I, PasI, StyI, AccI, AvaII) and a sequence alignment. The central part shows a 'DESIGN CRISPR' tool with a 'CRISPR' dropdown menu open, highlighting 'DESIGN HR TEMPLATE (SSODN)'. The right sidebar contains a 'Saved Guide Analyses' table with two entries: 'Untitled' (14/07/2020 13:37) and 'Untitled' (14/07/2020 14:06). The bottom of the interface shows a 'Position' section with a list of 15 and 22 items.</p>	<p>sélection de la création des recombinants homologues</p>

	<p>Step 1: Introduce Desired Edits</p> <p>Select and type in the Sequence Map to introduce the changes (e.g. point mutations, insertions, deletions) you want to make to the genome.</p> <p>Silent mutations to remove the target site will be added in a later step.</p> <p>Knock-in Edits</p> <p>No edits introduced yet.</p> <p>Next</p>	<p>sélection du codon muté</p>
	<p>Step 1: Introduce Desired Edits</p> <p>Select and type in the Sequence Map to introduce the changes (e.g. point mutations, insertions, deletions) you want to make to the genome.</p> <p>Silent mutations to remove the target site will be added in a later step.</p> <p>Knock-in Edits</p> <p>No edits introduced yet.</p> <p>Next</p>	<p>délétion du codon muté (GTG)</p>

<p>SEQUENCE MAP</p> <p>atggtgcacctgactcctgagagaagtctgccgttactgccctgtggggcaaggtgaactgga taccacgtggactgagga ctgagga caccocgttccacttgcacct</p> <p>Insert Bases Insert Part</p> <p>50 60 61 55</p> <p>HBB CDS (hemoglobin beta chain variant Hb S-Wake) >>> source >>> HBB gene >>></p>	<p>insertion du codon corrigé (GAG)</p>
<p>SEQUENCE MAP</p> <p>atggtgcacctgactcctgagagaagtctgccgttactgccctgtggggcaaggtgaactgga taccacgtggactgagga actccttcagacggaatgacgggacaccocgttccacttga</p> <p>ctgagga actccttcagac</p> <p>50 60 61 55 48 35 61 27</p> <p>HBB CDS (hemoglobin beta chain variant Hb S-Wake) >>> source >>> HBB gene >>></p> <p>Step 1: Introduce Desired Edits</p> <p>Select and type in the Sequence Map to introduce the changes (e.g. point mutations, insertions, deletions) you want to make to the genome.</p> <p>Silent mutations to remove the target site will be added in a later step.</p> <p>Knock-in Edits</p> <p>Changed t to a at 20</p> <p>Next</p>	<p>sauvegarde de la substitution pour le design</p>

	<p>Step 2: Adjust HR Arms</p> <p>Adjust the region to use as the HR template by clicking and dragging the ends of the selection on the Sequence Map.</p> <p>A 200 bp region around your mutations has already been selected. At least 50 bp on each side flanking the mutations is recommended.</p> <p>Template Region</p> <p>1 - 200 Reset to Default</p> <p>Template Length: 200 bp Left Arm Length: 19 bp Right Arm Length: 180 bp</p> <p>Knock-in Edits</p> <p>Changed t to a at 20</p> <p>Back Next</p>	<p>insertion des deux régions flanquant</p>
	<p>Step 4: Summary</p> <p>Knock-in Edits</p> <p>Changed t to a at 20</p> <p>Template Range 1 to 200</p> <p>Copy the template or its reverse complement to your clipboard.</p> <p>To design a template for the same knock-in edits but with a different guide, click here.</p> <p>Back</p>	<p>obtention de la séquence HR</p>

Une fois que le CRISPR a réalisé les deux coupures double brins, la séquence sélectionnée possédant le codon muté GTG, sera remplacé, et donc une séquence homologue possédant le codon GAG sain, sera inséré dans la même région, et cela grâce aux (recombinant homologue) (HR template), qui est une séquence de 200pb, constitué d'une séquence saine, et de deux séquences flanquant complémentaires à ce qui est en amont et en aval de la séquence excisé, pour l'insertion exacte des recombinants.

Remarque : Après avoir fini et sauvegardé les conceptions sur benchling, il sera possible d'exporter le protocole, grâce à un e-mail envoyé de l'équipe benchling. Ces coordonnées seront nécessaire pour faire la commande du kit complet des composants nécessaires de l'édition génétique par CRISPR Cas9, pour la suite de l'expérimentation in-vitro, un des laboratoires les plus renommés faisant ces commandes est Zhang lab.

2. Discussion

Les cellules souches pluripotentes induites par l'homme (iPSC) et l'édition du génome fournissent un moyen précis de générer des cellules à correction génétique pour la modélisation des maladies et les thérapies cellulaires. Les iPSC humains générés par des patients atteints de drépanocytose (SCD) ont une mutation ponctuelle homozygote dans le gène HBB codant pour les protéines β -globines adultes et sont utilisés comme système modèle pour améliorer les stratégies de thérapie génique humaine. Le système CRISPR / Cas9 est beaucoup plus efficace pour stimuler le ciblage génique du locus HBB endogène près de la mutation ponctuelle SCD dans les iPSC humains que les nucléases à doigt de zinc et les TALEN. En utilisant un ARN guide spécifique et Cas9, nous avons facilement corrigé un allèle du gène SCD HBB dans les iPSC humains par recombinaison homologe avec une matrice d'ADN donneur contenant l'ADN HBB de type sauvage et une cassette de sélection qui a ensuite été retirée pour éviter une éventuelle interférence de HBB. **(Huang *et al.*, 2014)**

Dans notre étude grâce aux outils bioinformatique, nous avons pu réaliser une conception complète de tout les composants nécessaire pour débiter la manipulation expérimentale, bien entendu nous avons été limité par les conditions de l'expérimentation pratiquement inexistantes; ce qui nous a obligé de faire un travail purement in-silico. En revanche les résultats obtenus ont été conformes à nos attentes et cela grâce à la présence des séquences PAM sur deux régions en amont et en aval de notre mutation au niveau de l'exon 1 du gène, ce qui nous permettra de faire deux coupure double brins et éventuellement couper un morceau du gène et le remplacer par un oligonucléotide identique possédant une Adénine au niveau du sixième codon au lieu d'une thymine et donc réparer la mutation du gène HbS pour qu'il devient HBB.

1. Un groupe de chercheurs avaient réussi à traiter la drépanocytose, D'après eux :

1.1." Les cellules SCD HEL ont été conçues pour convertir l'HbS / HbS en HbB / HbS, ce qui correspond à la conversion de SCD en génotype SCT. Les études de validation mentionnées ont démontré la reproductibilité, l'efficacité et la faisabilité de cette approche CRISPR / Cas9 pour les études cliniques" **(Wen *et al.*, 2017)**

- Ce qui concorde avec notre vision sur l'expérience Il est effectivement possible de faire l'édition génétique par CRISPR / Cas9 sur le gène HBS pour le convertir en gène sain HBB.

1.2." bien que les deux sites de mutation soient distants de plus de 10 nucléotides. Ces résultats indiquent qu'il est possible d'éditer simultanément deux sites cibles séparés dans le même gène par l'approche CRISPR / Cas9. L'édition simultanée du génome de deux sites cibles peut être utile pour traiter des maladies portant deux ou plusieurs sites de mutation ponctuelle proches l'un de l'autre" (**Wen et al., 2017**)

-cela confirme notre approche d'utiliser deux ARN guides produisant deux coupures double brins.

1.3." l'individu avec SCT peut avoir des phénomènes pathologiques dans des circonstances extrêmes. Ainsi, la correction des HSPC des patients SCD du génotype Hbs / HbS au génotype HBB / HBB est toujours préférée dans la future étude " (**Wen et al., 2017**)

- Coïncidant avec notre but de la recherche ; ceci prouve qu'il est idéal d'éditer les deux mutations homozygotes et non pas un seul allèle pour la guérison totale des patients.

2. d'après Christopher Vakulskas et son équipe qui ont voulu traiter la drépanocytose ; mais en ayant le minimum de régions hors cible :

"Ils ont conçu un écran bactérien non biaisé, et l'ont utilisé pour isoler une Cas9 de haute fidélité qui dépasse la spécificité de ciblage et correspond à l'efficacité d'édition cible de Cas9 de type sauvage." (**Vakulskas, 2020**)

-Ce groupe a développé une nouvelle technique qui nécessite plus d'essais ; en mettant l'accent sur la Cas9 de *Streptococcus pyogenes* et Cas12a (Cpf1) de l'espèce *Acidaminococcus* pour la réduction des événements hors cible du clivage, alors que nous avons choisis le type classique et plus fiable qui est la Cas9 sauvage (Wild cas9) (**Vakulskas, 2020**)

CONCLUSION

Le gène HBB est situé sur le bras court du chromosome 11, la mutation 6A>T (Glu6>Val) entraîne la drépanocytose, plusieurs outils et logiciels ont été utilisés pour étudier la maladie au niveau génétique et moléculaire (présentation en 3D des protéines issues des gènes mutés et sains, et leur alignement pour cibler la position exacte du polymorphisme nucléotidique) pour bien sûr traiter cette pathologie, c'est ici qu'un protocole de conception prédictif du CRISPR-Cas9 et de ces composants (conception de l'ARN guide, du plasmide, du primer et du recombinaisonnologue) a été mis en place comme première étape cruciale pour ensuite passer à l'expérimentation *in vitro* sans rencontrer de difficultés majeures pendant la manipulation génétique.

Après avoir construit tous les éléments essentiels *in-silico*, nous obtiendrons le kit complet des composants d'un des laboratoires spécialisés en édition par CRISPR (Zhang lab est le plus renommé) pour débiter la thérapie génique d'un individu souffrant de drépanocytose, et confirmer que notre approche est fiable comme traitement.

En perspectives il est intéressant de:

- travailler concrètement avec des cellules de sujets atteints de drépanocytose.
- cloner différentes cellules souches de différentes régions pour connaître si cela a une influence sur les résultats.
- obtenir plus d'échantillons malades pour savoir l'impact de la variabilité génétique des individus sur la recherche.
- tester différents vecteurs et savoir si cela affecte l'édition.
- introduire des enzymes pour tenter de minimiser les régions hors cible.
- utiliser d'autres outils d'analyse et conception de CRISPR-Cas9, pour savoir si cela retentit sur les données

Reference bibliographique

Aliyu, Z. Y. and Tumblin, A. R. (2005) ‘Current therapy of sickle cell disease Fe ta St or ti Fo u Fe rra ta St or ti Fo u’, *Haematologica*, 90(4), pp. 7–10.

Anders, C. *et al.* (2014) ‘Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease’, *Nature*. Nature Publishing Group. doi: 10.1038/nature13579.

Bender, M. A. (2020) ‘Summary’, pp. 1–35.

Cai, L. *et al.* (2018) ‘A Universal Approach to Correct Various HBB Gene Mutations in Human Stem Cells for Gene Therapy of Beta-Thalassemia and Sickle Cell Disease’, *Stem Cells Translational Medicine*, 7(1), pp. 87–97. doi: 10.1002/sctm.17-0066.

Charpentier, E. *et al.* (2015) ‘Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity’, *FEMS Microbiology Reviews*, 39(3), pp. 428–441. doi: 10.1093/femsre/fuv023.

Itim Hanane, Noui Chiema, ‘La drépanocytose : causes , symptômes et Traitement’, mémoire de master, université des frères mentouri constantine; biologie animal; génétique moléculaire, soutenue 2017

Christine, P. (2018) ‘BIOLOGIE A UJOURD ’ HUI’, 211(4), pp. 247–254.

Croteau Félix R., Geneviève M. Rousseau, S. M. (2018) ‘Le système CRISPR-Cas Au-delà de l’édition génomique’, *Médecine/Sciences*, 34, pp. 813–819.

Donohoue, P. D., Barrangou, R. and May, A. P. (2018) 'Advances in Industrial Biotechnology Using CRISPR-Cas Systems', *Trends in Biotechnology*. Elsevier Ltd, 36(2), pp. 134–146. doi: 10.1016/j.tibtech.2017.07.007.

Frenette, P. S. *et al.* (2007) 'Sickle cell disease: old discoveries, new concepts, and future promise Find the latest version: Science in medicine Sickle cell disease: old discoveries, new concepts, and future promise', *Journal of Clinical Investigation*, 117(4), pp. 850–858. doi: 10.1172/JCI30920.850.

Hassell, K. L. (2010) 'Population Estimates of Sickle Cell Disease in the U.S.', *American Journal of Preventive Medicine*. Elsevier Inc., 38(4 SUPPL.), pp. S512–S521. doi: 10.1016/j.amepre.2009.12.022.

Hille, F. *et al.* (2018) 'The Biology of CRISPR-Cas: Backward and Forward', *Cell*, 172(6), pp. 1239–1259. doi: 10.1016/j.cell.2017.11.032.

Huang, X. *et al.* (2014) 'Production of Gene - Corrected Adult Beta Globin Protein in Human Erythrocytes Differentiated from Patient Ipscs after Genome Editing of the Sickle Point Mutation', pp. 1–15. doi: 10.1002/stem.1969.

Ingram, V. M. (1956) 'A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin', *Nature*, 178(4537), pp. 792–794. doi: 10.1038/178792a0.

Jiang, W. and Marraffini, L. A. (2015) 'CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems', *Annual Review of Microbiology*, 69(1), pp. 209–228. doi: 10.1146/annurev-micro-091014-104441.

Jinek, M. *et al.* (2014) ‘Structures of Cas9 endonucleases reveal RNA-mediated conformational activation’, *Science*, 343(6176). doi: 10.1126/science.1247997.

Kim, H. and Kim, J. (2014) ‘REVIEWS A guide to genome engineering with programmable nucleases’, *Nature Publishing Group*. Nature Publishing Group, 15(5), pp. 321–334. doi: 10.1038/nrg3686.

Lanzkron, S., Patrick Carroll, C. and Haywood, C. (2013) ‘Mortality rates and age at death from sickle cell disease: U.S., 1979-2005’, *Public Health Reports*, 128(2), pp. 110–116. doi: 10.1177/003335491312800206.

Piel, F. B., Steinberg, M. H. and Rees, D. C. (2017) ‘Sickle cell disease’, *New England Journal of Medicine*, 376(16), pp. 1561–1573. doi: 10.1056/NEJMra1510865.

Sander, J. D. and Joung, J. K. (2014) ‘CRISPR-Cas systems for editing, regulating and targeting genomes’, *Nature Biotechnology*. Nature Publishing Group, 32(4), pp. 347–350. doi: 10.1038/nbt.2842.

Santin, A. (2013) ‘Drépanocytose et complications aiguës’, 1.

Shenoy, S. (2011) ‘Hematopoietic stem cell transplantation for sickle cell disease: current practice and emerging trends.’, *Hematology / the Education Program of the American Society of Hematology*. American Society of Hematology. Education Program, 2011, pp. 273–279. doi: 10.1182/asheducation-2011.1.273.

Siksnys, V. and Gasiunas, G. (2016) ‘Rewiring Cas9 to Target New PAM Sequences’, *Molecular Cell*. Elsevier Inc., 61(6), pp. 793–794. doi: 10.1016/j.molcel.2016.03.002.

Tremblay, J. P. (2015) ‘CRISPR, un système qui permet de corriger ou de modifier l’expression de gènes responsables de maladies héréditaires’, *Medecine/Sciences*, 31(11), pp. 1014–1022. doi: 10.1051/medsci/20153111016.

Vakulskas, C. (2020) ‘Could a CRISPR Therapeutic Cure Sickle-Cell Disease?’, *Genetic Engineering and Biotechnology News*, 40(3), pp. 17–18. doi: 10.1089/gen.40.03.05.

Wang, J. *et al.* (2015) ‘Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems’, *Cell*. Elsevier Inc., 163(4), pp. 840–853. doi: 10.1016/j.cell.2015.10.008.

Wen, J. *et al.* (2017) ‘Cellular function reinstatement of offspring red blood cells cloned from the sickle cell disease patient blood post CRISPR genome editing’. *Journal of Hematology & Oncology*, pp. 1–11. doi: 10.1186/s13045-017-0489-9.

Wong, A. Y. H. (2006) ‘American Journal of pediatrics’, 151(9), pp. 839–845. doi: 10.1038.

Webographie :

<https://benchling.com/protocols/5DmqRd/crispr-mediated-gene-disruption-in-ch12f3-2-cells/sbs>

<https://www.ncbi.nlm.nih.gov/>

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

<https://swissmodel.expasy.org/interactive>

<https://www.benchling.com/crispr/>