



الجمهورية الجزائرية الديمقراطية الشعبية

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université Frères Mentouri Constantine1  
Faculté des Sciences de la Nature et de la Vie

جامعة الاخوة منتوري قسنطينة  
كلية علوم الطبيعة والحياة

Département : Biochimie et Biologie Moléculaire et Cellulaire  
Mémoire présenté en vue de l'obtention du Diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Alimentaires

Spécialité : Biochimie de la Nutrition

Intitulé :

---

Analyse bioinformatique des variants de l'hémoglobine humaine

---

Présenté et soutenu par : *ZIADI Yasmine*

*SEKEHAL Amira*

Jury :

Président du jury : Pr NECIB Y.

Rapporteur : Pr HAMIDECHI M. A.

Examineur : Dr MOKRANI M.

*Année universitaire  
2020-2021*

# *Remerciements*

## Liste des figures

Figure 1 : Structure basique de l'hème tétrapyrrolique .....	4
Figure 2 : Structure primaire de la sous-unité Alpha hémoglobine, sur la base de données UniProt, code P69905-1 .....	4
Figure 3: Schéma des quatre sous-unités formant la cavité centrale. 2,3-BPG est le 2,3 bisphosphoglycérate.....	5
Figure 4 : structure secondaire de la sous-unité Alpha hémoglobine. Source : base de données UniProt, code P69905-1.....	5
Figure 5 : Structure des différentes chaînes de l'hémoglobine humaine.....	6
Figure 6: Positions et longueurs des différents motifs structuraux de la sous-unité $\alpha$ hémoglobine. ....	6
Figure 7: Structure secondaire des hélices de la bêta globine. AA1 et AAn désignent le premier et le dernier acide aminé de la molécule.....	7
Figure 8: structure tertiaire de la sous-unité $\alpha$ humaine. Base de données SwissModel, code D1MGQ2. ....	7
Figure 9: Structure 3D de l'hémoglobine, Protein Data Bank ID 1hho. Chaînes $\alpha$ en bleu, chaînes $\beta$ en beige et hème lié à l'oxygène en rouge.....	8
Figure 10: Engagement des résidus aminoacides dans la fixation du 2,3 DPG sur l'HbA désoxygénée (ZAHER Y., 2011).....	11
Figure 11 : Tétramère d'hémoglobine et ses différents états T et R à l'interfac des dimères. (Dey S., Chakrabarti P, Janin J. 2011).....	12
Figure 12: Structures schématiques des gènes regroupés en clusters sur les deux chromosomes 11 (11p 15,5) et 16 (16p 13,3).....	13
Figure 13: Disposition des clusters des différents gènes sur les chromosomes 11 et 16. ....	14
Figure 14: Structure des gènes de globine.....	15
Figure 15 : Alignement multiple des séquences alpha-globuline par le logiciel Clustal oméga. ....	26
Figure 16 : Arbre phylogénétique des séquences alpha-globulines. Réalisé sur <a href="http://phylogeny.lirmm.fr/">http://phylogeny.lirmm.fr/</a> (Dereeper A. et al ,2008).....	29
Figure 17 : Alignement multiple des séquences bêta-globuline par le logiciel Clustal oméga. ....	30
Figure 18 : Arbre phylogénétique des séquences bêta-globulines. Réalisé sur <a href="http://phylogeny.lirmm.fr/">http://phylogeny.lirmm.fr/</a> (Dereeper A. et al , 2008).....	31
Figure 19 : Représentation 3D de l'Alpha-hémoglobine native (Code 1A00) sur PyMol. ....	32
Figure 20: Position du résidu N69 sur la molécule 1A00. ....	33
Figure 21 : Position du résidu C69 sur la molécule 1A00 après avoir substitué le résidu N69. ....	33
Figure 22 : Position naturelle du résidu D53 sur la sous-unité bêta de l'Hb. ....	33
Figure 23 : Position naturelle du résidu W53 sur la sous-unité bêta de l'Hb après avoir substitué le résidu naturel D53.....	34

## Liste des tableaux

Tableau 1 : Caractéristiques des HbA1 et HbA2 collectées à partir de la base de données UniprotKB.....	20
Tableau 2 : Caractéristiques des HbB collectées à partir de la base de données UniprotKB ...	23
Tableau 3 : Matrice de Id% des séquences alpha-globulines .....	27
Tableau 4 : Matrice de Id% des séquences bêta-globulines .....	30

## Sommaire

Introduction .....	1
Partie 1 : Revue bibliographique.....	2
1.Molécule d'hémoglobine (Hb) .....	2
1.1.Structure .....	2
1.2. Génétique de l'hémoglobine .....	12
2. Hémoglobinopathies .....	15
2.1.Hémoglobine S .....	16
2.2.Hémoglobine C .....	16
2.3.Hémoglobine E .....	17
3.Thalassémies .....	17
3.1. Bêta-Thalassémies .....	17
3.2. Alpha-Thalassémies .....	18
Partie 2 : matériel et méthodes.....	19
1.Matériel .....	19
1.1. Récupération des séquences d'hémoglobine a partir des banques de données .....	19
1.2. Prediction des substitutions entre résidus amino-acides : .....	19
2.Résultats :.....	20
2.1. Séquences de l'HbA : .....	20
2.2. Séquences de l'HbB : .....	23
2.3. Alignement multiple des séquences d'hémoglobine .....	25
2.4. Blast.....	31
Partie 3 : résultat et discussion .....	35
Conclusion .....	38
Références bibliographiques .....	39

# Introduction

La protéine de l'hémoglobine est l'une des plus étudiées en biochimie pathologique, surtout à cause des hémoglobinopathies associées aux variations génétiques modifiant ainsi son activité. Ces variations génétiques concernent des mutations sur une des chaînes de globine ou sont dues à un défaut dans la synthèse d'une des chaînes de globine. Ces changements biomoléculaires ne restent pas sans conséquences, bien au contraire, ils sont accompagnés le plus souvent de modifications structurelles et donc fonctionnelles qui inquiètent sérieusement la santé publique, car très répandus dans le monde (Couque N., 2013 ; Dey S. *et al.* 2011).

Les molécules d'hémoglobine dérivant de ces changements sont appelées les variants d'hémoglobine. Ce sont donc des individus mutés sur un résidu acide aminé par rapport à la molécule native. Certaines de ces mutations peuvent par exemple concerner la substitution du glutamate par une Valine (mutation sur le codon 6 du gène).

Le défi à relever de nos jours, vu le développement crucial des techniques biomoléculaires et biotechnologiques est d'arriver à lire et à analyser toutes les séquences protéiques de ces individus en utilisant les outils actuels de bioinformatique afin de prédire, dans les différentes populations humaines, et dans des délais très courts, les différents variants génétiques de l'hémoglobine et d'isoler ceux à risque qui peuvent avoir une mutation à effet inhibiteur de l'activité d'hémoglobine, donc sur la santé de ces individus (Wajcman H., 2005).

Dans notre travail, nous nous proposons de faire une lecture bioinformatique des protéines des hémoglobines humaines afin de prédire les mutations possibles parmi les résidus de ces séquences et de mesurer l'impact de cette substitution sur la santé des individus. Ainsi l'objectif principal de cette étude est de détecter les mutations les plus éventuelles (donc les plus possibles) compte tenu de la structure initiale de la protéine humaine, puis dans un deuxième temps de mesurer sa probabilité à affecter l'activité protéique de ces séquences.

Révue

Bibliographique



## 1. Molécule d'hémoglobine (Hb)

Le terme Hémoglobine, a pour la première fois été employé par le physiologiste Allemand Hoppe-Seyler en 1862 pour désigner la signification de pigment respiratoire. Elle est présente dans tout le monde du vivant, de la bactérie à l'homme. La myoglobine est à distinguer de l'hémoglobine lorsqu'elle est dans une cellule musculaire et on parlera de neuroglobine si elle est dans une cellule nerveuse. Toutes ces molécules sont caractérisées par un enroulement tridimensionnel caractéristique (*globin fold*) de six à huit hélices autour de l'hème (Wajcman , 2005). L'hémoglobine est une grosse molécule de protéine présente dans les globules rouges. En fait, l'hémoglobine est la substance qui donne au sang sa teinte rouge. Le biologiste moléculaire Max Perutz a découvert l'hémoglobine en 1959. Perutz a utilisé la cristallographie aux rayons X pour déterminer la structure spéciale de l'hémoglobine. Il découvrira également à terme la structure cristalline de sa forme désoxygénée. C'est la molécule porteuse d'oxygène aux milliards de cellules du corps, nécessaire à la vie des humains et des autres mammifères. Elle transporte à la fois l'oxygène et le dioxyde de carbone.

L'hémoglobine constitue jusqu'à 95% de la teneur en protéines des érythrocytes. Elle fait l'objet de recherches approfondies non seulement en raison de sa fonction clé dans la respiration aérobie mais aussi parce qu'elle constitue un modèle important de régulation allostérique et de liaison coopérative à l'oxygène (Dybas, 2020). On estime à 1,8 milliard d'années son origine (soit 0,5 milliard d'années après les algues bleues) bien avant la vie aérobie que nous connaissons aujourd'hui, appartenant à la superfamille des globines. (Wajcman, 2005).

### 1.1. Structure

Les hémoglobines se distinguent par la nature de leurs sous-unités constitutives. Par exemple chez l'homme, il a été constaté deux passages critiques dans l'évolution de sa structure : (i) de la vie embryonnaire à la vie fœtale et (ii) de la vie fœtale à la vie adulte. Au cours de la vie embryonnaire, deux chaînes Alpha coexistent :  $\zeta$  (prononcé zêta) et  $\alpha$ . Parallèlement, deux chaînes de type  $\beta$  existent :  $\varepsilon$  et  $\gamma$ . Les chaînes  $\gamma$  ( $^A\gamma$ : Ala136 et  $^G\gamma$  Gly136) seront remplacées par les chaînes  $\beta$  de l'adulte.

Classiquement, on distingue différentes chaînes de globines :

- Les chaînes  $\alpha$  avec 141 résidus

- Les chaînes  $\beta$  avec 146 résidus
- Les chaînes  $\gamma$  avec 146 résidus
- Les chaînes  $\delta$  avec 146 résidus.

De ce fait, il existe différentes combinaisons d'hémoglobines chez l'homme. On distinguera ainsi :

- Hémoglobine normale de l'adulte HbA  $\alpha_2\beta_2$  constituée de deux chaînes  $\alpha$  et deux chaînes  $\beta$ .
- Hémoglobine mineure de l'adulte HbA2  $\alpha_2\delta_2$
- Hémoglobine fœtale HbF  $\alpha_2\gamma_2$

L'hémoglobine, contenue dans les globules rouges, est un hétérotétramère constitué de deux types de sous-unités, formées chacune par le repliement de huit hélices autour d'une molécule d'hème. On y distingue quatre protomères similaires, de structure globulaire d'environ 64,458 KDa. (Wajcman . 2005).

Elle est constituée de quatre chaînes polypeptidiques identiques deux à deux et unies par liaisons non covalentes. On les distingue, chez l'homme, en deux types de chaînes : alpha et non alpha (bêta, gamma et delta). Celles-ci ont une structure spatiale très similaire et analogue à celle de la myoglobine.

L'hème constitue la partie non protéique de l'hémoglobine ; il est formé d'une structure aromatique et d'un atome de fer. C'est une ferro-protoporphyrine résultant de l'association d'une partie organique (porphyrine : structure aromatique, constituée de quatre noyaux pyrrol) et d'un atome de fer ferreux à l'état réduit ( $Fe^{++}$ ) pouvant réversiblement fixer un atome d'oxygène : l'atome de fer est lié par six valences (hexacoordiné). Quatre d'entre elles fixent le fer sur les quatre atomes d'azote de la porphyrine. Une valence du fer est liée à un des azotes d'une histidine de l'hélice F (His proximale), et la dernière à une histidine de l'hélice E (His distale)<sup>1</sup>.

---

<sup>1</sup> <http://www.chups.jussieu.fr/polys/biochimie/SFbioch/POLY.Chp.3.2.html>

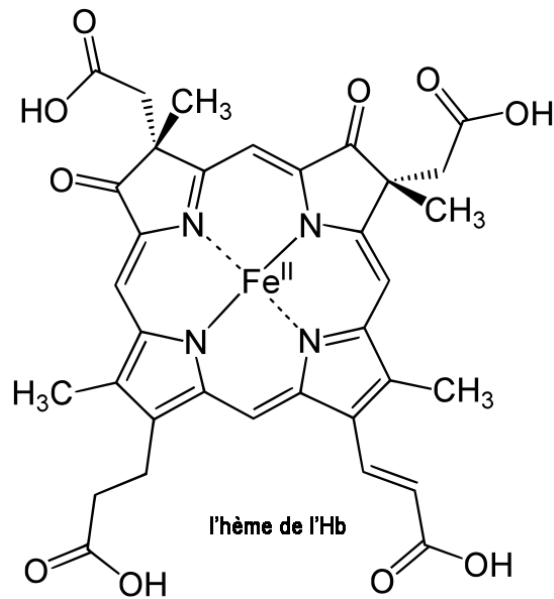


Figure 1 : Structure basique de l'hème tétrapyroliique<sup>2</sup>

La chaîne de la structure primaire est différente d'un type à l'autre. La chaîne alpha est longue de 142 résidus d'acides aminés avec une Arginine (R) en position C-terminale (figure 4), tandis que les chaînes bêta, delta et gamma possèdent 146 acides aminés et se terminent par une Histidine (H).

10	20	30	40	50
MVLSPADKTN	VKAAWGKVGA	HAGEYGAEAL	ERMFLSFPTT	KTYFPDFDLS
60	70	80	90	100
HGSAQVKGHG	KKVADALTNA	VAHVDDMPNA	LSALSDLHAH	KLRVDPVNFK
110	120	130	140	
LLSHCLLVTL	AAHLPAEFTP	AVHASLDKFL	ASVSTVLTSK	YR

Figure 2 : Structure primaire de la sous-unité Alpha hémoglobine, sur la base de données UniProt, code P69905-1

<sup>2</sup> <http://pst.chez-alice.fr/protstfo.htm>

Schématiquement les quatre sous-unités de l'hémoglobine formant leur cavité centrale peuvent être représentées ainsi :

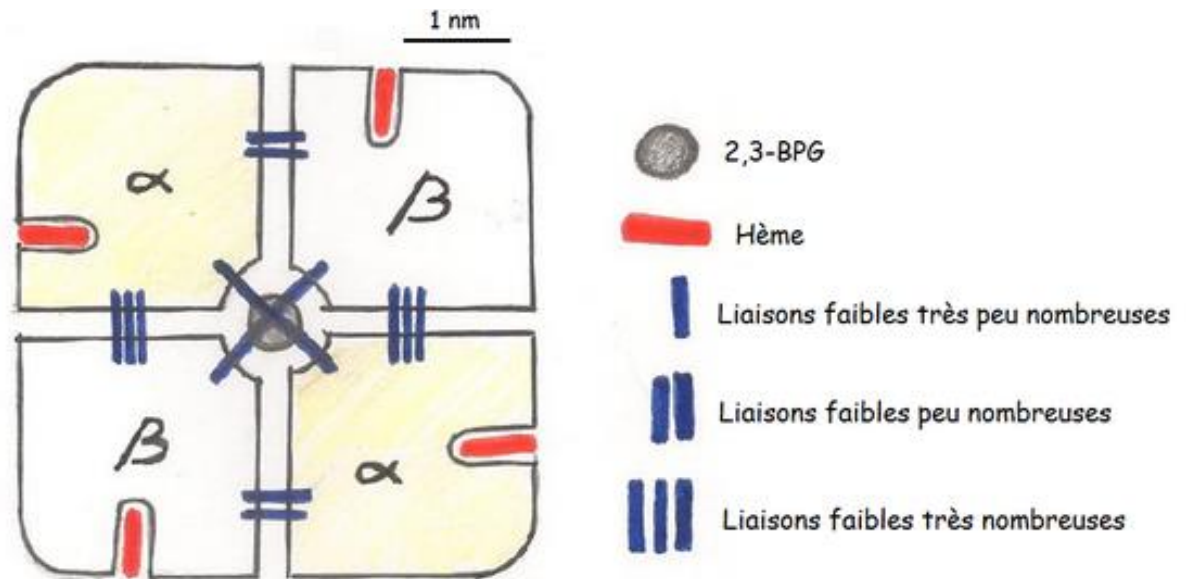


Figure 3: Schéma des quatre sous-unités formant la cavité centrale. 2,3-BPG est le 2,3 bisphosphoglycérate<sup>3</sup>.

Dans la structure secondaire, les chaînes de globine adoptent une configuration en hélices discontinues par repliement de la structure primaire. Ces chaînes enroulées sont constituées de sept ou huit segments hélicoïdaux (désignés par des lettres de A à H) séparés par des segments non hélicoïdaux ; ce sont les coudes (tours). Les chaînes sont formées chacune de 7 hélices  $\alpha$  pour les chaînes A et de 8 hélices  $\alpha$  pour les chaînes B.

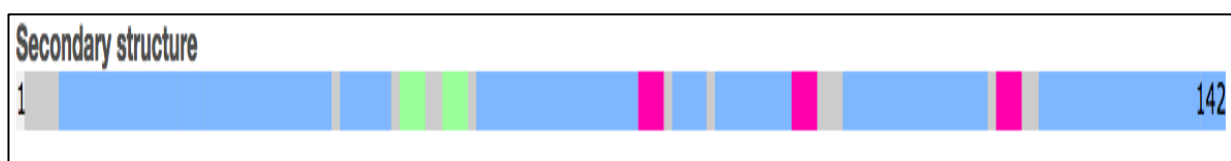


Figure 4 : structure secondaire de la sous-unité Alpha hémoglobine. Source : base de données UniProt, code P69905-1.

Légende :  Hélice alpha     Feuillet bêta     Tour     Indéterminée

<sup>3</sup> <http://tpe-julie-lise.e-monsite.com/pages/ii-composition/a-h.html>

La disposition des différentes hélices des deux chaînes  $\alpha$  et  $\beta$  humaines sont représentées sur la figure suivante :

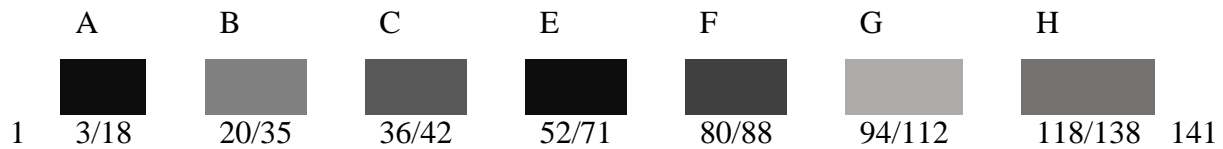


Figure 5 : Structure des différentes chaînes de l'hémoglobine humaine<sup>4</sup>.

Les motifs de structure secondaires (alpha, bêta et Tours) sont détaillés dans le tableau suivant :

Motif	Positions	Longueur	Motif	Positions	Longueur
Hélice	5-18	14	Hélice	77-80	4
Hélice	19-21	3	Hélice	82-90	9
Hélice	22-36	15	Tour	91-93	3
Hélice	38-43	6	Hélice	97-113	17
Feuillet bêta	45-47	3	Tour	115-117	3
Feuillet bêta	50-52	3	Hélice	120-138	19
Hélice	54-72	19	Hélice	139-141	3
Tour	73-75	3			

Figure 6: Positions et longueurs des différents motifs structuraux de la sous-unité  $\alpha$  hémoglobine.

Les huit hélices (Figure 6), réalisent par repliement, une structure de forme globulaire avec une poche hydrophobe près de la surface où se localise la molécule d'hème. La surface externe de la sous-unité est recouverte par les chaînes latérales hydrophiles ; ce qui permettra de meilleures interactions avec leur milieu. Les régions internes sont constituées par des résidus aminoacides hydrophobes (non polaires) tels que Isoleucine(T) , Tryptophane(W), etc. La cavité où se trouve la molécule d'hème présente une forme de V dont l'ouverture est occupée en partie par l'hélice C.

<sup>4</sup> <http://tpe-julie-lise.e-monsite.com/pages/ii-composition/a-h.html>

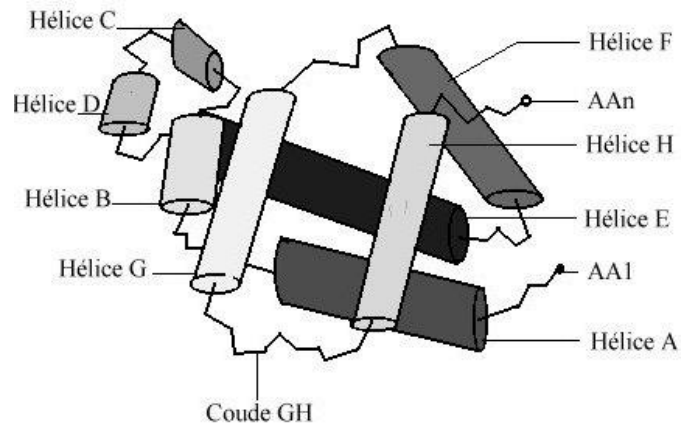


Figure 7: Structure secondaire des hélices de la bêta globine. AA1 et AAn désignent le premier et le dernier acide aminé de la molécule.

La structure tertiaire (3D) représente la conformation tridimensionnelle de la molécule de globine. Quatre liaisons interviennent au niveau des sommets des noyaux pyrroliques pour stabiliser la structure de l'hème qui échange un grand nombre de liaisons (soixantaine) à faible énergie avec les radicaux latéraux des acides aminés qui l'entourent. Une cinquième liaison située au niveau de l'His F8 dite « Histidine proximale » assure le lien entre l'hème et la globine. La sixième valence libre permet de fixer le ligand oxygène, qui est lui-même en rapport avec l'His E7 appelée encore « histidine distale ».

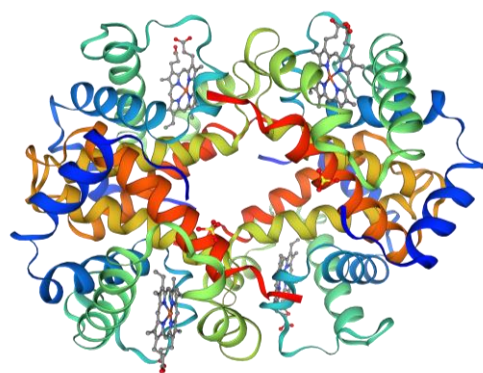


Figure 8: structure tertiaire de la sous-unité  $\alpha$  humaine. Base de données SwissModel, code D1MGQ2.

La stabilité de l'hémoglobine, ainsi que sa fonction sont dictées par des positions spécifiques voire critiques d'un ensemble de résidus aminoacides. Ainsi, le groupement azote de l'Histidine  $\beta 43$  va fixer le radical  $(\text{PO}_4)^{3-}$  au 2,3- diphosphoglycérate (2,3-DPG), alors que la Lysine  $\beta 82$  va fixer le radical  $\text{COO}^-$  sur le 2,3-DPG et l'Histidine F8 va constituer le site proximal de liaison avec le fer de l'hème tandis que l'Histidine E7 constituera le site distal à proximité de la liaison avec le fer (Zaher Y. 2011.)

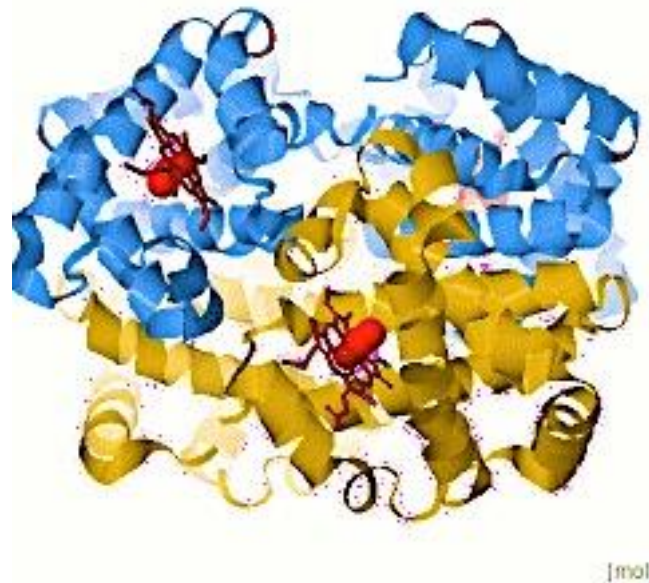


Figure 9: Structure 3D de l'hémoglobine, Protein Data Bank ID 1hho. Chaînes  $\alpha$  en bleu, chaînes  $\beta$  en beige et hème lié à l'oxygène en rouge.

Quant à la structure quaternaire (plusieurs sous-unités), elle est tétramérique (Figure 9). En effet certaines protéines sont constituées de plusieurs chaînes polypeptidiques, également appelées sous-unités. Lorsque celles-ci se rejoignent, elles confèrent à la protéine sa structure quaternaire. En général, les mêmes types d'interactions qui contribuent à la structure tertiaire (principalement des interactions faibles, telles que les liaisons hydrogène et les forces de dispersion de Londres) maintiennent également les sous-unités ensemble pour donner une structure quaternaire. Chacune des sous-unités est jointe à sa partenaire autour d'un axe de symétrie double, de sorte qu'une rotation de  $180^\circ$  amène une sous-unité en congruence (adéquation) avec sa partenaire. Une paire de chaînes est ensuite inversée et placée au-dessus de l'autre paire de sorte que les quatre chaînes se trouvent aux coins d'un tétraèdre.

Les chaînes polypeptidiques des sous-unités sont maintenues en place par les mêmes forces qui déterminent les liaisons hydrogène de la structure tertiaire, les ponts salins et parfois les ponts disulfures, sauf que les forces se situent entre les chaînes polypeptidiques des sous-unités plutôt qu'au sein de celles-ci. Les chaînes latérales aliphatiques et aromatiques hydrophobes des sous-unités peuvent s'agréger pour exclure l'eau. La structure quaternaire de l'hémoglobine se compose de quatre sous-unités peptidiques. Deux des sous-unités sont identiques et sont appelées sous-unités alpha. Les deux sous-unités restantes, appelées sous-unités  $\beta$ , sont identiques entre elles mais différentes des sous-unités  $\alpha$ .

Le double axe de symétrie qui relie les paires de chaînes  $\alpha$  et  $\beta$  traverse une cavité remplie d'eau au centre de la molécule. Cette cavité s'élargit lors de la transition de la structure R à la structure T pour former un site récepteur pour l'effecteur allostérique DPG entre les deux chaînes  $\beta$ . Le groupe hémique est coincé dans une poche de la globine avec ses chaînes latérales d'hydrocarbures à l'intérieur et ses chaînes latérales de propionate polaire à l'extérieur. Il y a neuf positions dans la séquence d'acides aminés qui contiennent le même acide aminé dans toutes ou presque toutes les espèces étudiées jusqu'à présent. Ces positions conservées sont particulièrement importantes pour la fonction de la molécule d'hémoglobine (Shaanan, B. 1983.)

La structure de l'hémoglobine adulte (HbA) humaine est telle que les chaînes sont assemblées selon une symétrie tétraédrique, de manière à ce que les sous-unités  $\alpha_1$  et  $\beta_1$  forment un dimère et  $\alpha_2$  et  $\beta_2$  l'autre dimère. Des contacts très rigides et apolaires prennent place dans le tétramère pour stabiliser la molécule. Il s'agit de liaisons électrostatiques fortes mais peu nombreuses, qui se mettent en place entre les sous-unités d'un même dimère ( $\alpha_1$ - $\beta_1$  et  $\alpha_2$ - $\beta_2$ ). D'autres liaisons faibles et peu rigides, impliquant essentiellement des résidus des hélices C et G et du segment FG, s'établissent entre les sous-unités de deux dimères voisins ( $\alpha_1$ - $\beta_2$  et  $\alpha_2$ - $\beta_1$ ) pour permettre le processus de transition allostérique. Un champ électrostatique permet d'établir un troisième contact au niveau de la « cavité centrale », moyennant une molécule de 2,3-DPG. Cette dernière liaison assure le contact entre les deux chaînes bêta tout en stabilisant la configuration désoxygénée de la molécule d'hémoglobine. Un groupe prothétique hémique contenant un ion fer (II) capable de transporter une molécule d'oxygène est associé à chacune des quatre sous-unités protéiques de l'hémoglobine. La globine est la protéine dont les groupes hémiques ont été éliminés. Le complexe d'hémoglobine avec l'oxygène est l'oxyhémoglobine ; celle sans oxygène est la désoxyhémoglobine. L'oxyhémoglobine, la principale hémoglobine du sang artériel, est rouge



vif ; la désoxyhémoglobine du sang veineux est violacée. Une longue exposition à l'oxygène convertira le fer hémique de l'hémoglobine en fer ferreux (III), ou état ferrique ( $Fe^{3+}$ ), pour donner de la méthémoglobine. La couleur brune du sang séché résulte de la conversion de l'hémoglobine en méthémoglobine. L'oxygène ne forme pas de complexes avec la méthémoglobine. Dans la méthémoglobinémie, une maladie sanguine héréditaire, le fer hémique de l'hémoglobine, dans les chaînes  $\alpha$  ou  $\beta$ , existe à l'état de fer ferreux (III). Les personnes ayant des ions fer (III) dans les groupes hème des chaînes  $\alpha$  et  $\beta$  seraient incapables de vivre car leur sang ne pourrait pas transporter l'oxygène.

Le tétramère de l'hémoglobine humaine comporte essentiellement trois zones de contact des sous-unités :

Zone 1 : Contacts entre sous-unités d'un même dimère  $\alpha 1\beta 1$  ou  $\alpha 2\beta 2$ . Concerne les hélices C, G et H ; avec un total de 34 acides aminés qui interagissent entre eux, mais la transition entre les deux états T et R de la forme quaternaire n'implique pas de grandes différences dans la structure spatiale des deux états. Lors de l'oxygénation d'un tétramère réduit, la première molécule d'oxygène se fixe sur une sous-unité  $\alpha$ , cette information serait ensuite transmise à la sous-unité  $\beta$  du même dimère pour donner une espèce intermédiaire où les sous unités  $\alpha 1\beta 1$  (ou  $\alpha 2\beta 2$ ) sont oxygénées avant que ne s'effectue la transition  $T \rightarrow R$  du tétramère (Chang J-G et al, 1998)

Zone 2 : Contacts entre chaînes polypeptidiques non homologues  $\alpha 1\beta 2$  ou  $\alpha 2\beta 1$

Il existe également une structure supra-quaternaire. Elle représente le mode de répartition des molécules d'hémoglobine à l'intérieur des érythrocytes. Les hémoglobines sont distantes entre elles de 8 Å et se répartissent essentiellement en périphérie pour donner la forme des hématies. Dans cette structure, l'hémoglobine peut être considérée comme une micelle avec des groupements hydrophiles situés à l'extérieur et des groupements hydrophobes localisés à l'intérieur de la molécule plus particulièrement au niveau de la poche de l'hème (Wajcman, 2005 ; Couque, 2013).

Zone 3 : Contact entre chaînes homologues. Les deux chaînes  $\beta$  sont séparées par une interface appelée « cavité centrale » qui est tapissée de résidus positivement chargés. Dans la forme T, une molécule de 2,3 diphosphoglycérate (2,3 DPG) s'y insère pour former un clamp électrostatique entre les deux chaînes  $\beta$ . Il a été montré par diffractions de rayons X que sept liaisons ioniques s'établissent entre les groupements négativement chargés du 2,3 DPG et les groupements positivement chargés de la protéine. Ces liaisons impliquent les groupements

NH2 terminaux des deux chaînes  $\beta$  (NA1), les deux His  $\beta$ 2 (NA2), les deux His  $\beta$ 143 (H21) et une des Lys  $\beta$ 82 (EF6) (Figure 10). Dans les chaînes  $\gamma$ , une Sér, au lieu d'une His, occupe la position 143 (H21), ce qui expliquerait une plus faible affinité du 2,3 DPG pour l'HbF. Dans la désoxyHb, des interactions électrostatiques, faisant intervenir l'ion  $\text{Cl}^-$ , s'établissent également entre l'extrémité C-terminale d'une chaîne  $\alpha$  et l'extrémité N-terminale de l'autre chaîne (Zaher Y., 2011).

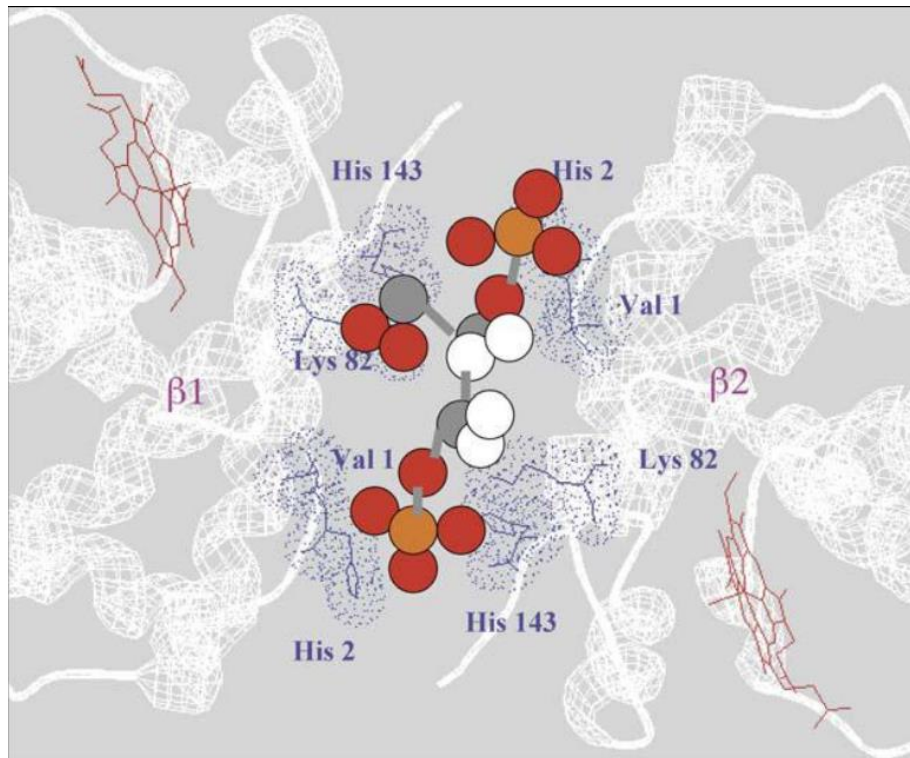


Figure 10: Engagement des résidus aminoacides dans la fixation du 2,3 DPG sur l'HbA désoxygénée (ZAHER Y., 2011).

L'approche enzymatique allostérique a démontré que la molécule d'hémoglobine est en équilibre entre deux états différents. L'état T dans lequel la molécule est tendue et présente une très faible affinité pour son ligand oxygène. L'état R, dans lequel la molécule est fortement relâchée présente la meilleure affinité pour l'oxygène ; les quatre sous-unités sont en parfaite coopération. La disposition des chaînes polypeptidiques est telle que des liaisons existent entre chaînes latérales de résidus appartenant aux sous-unités non homologues ( $\alpha$ 1- $\beta$ 1 et  $\alpha$ 2- $\beta$ 2). À l'inverse, il n'existe qu'un très petit nombre de contacts entre sous-unités identiques (Figure 11)

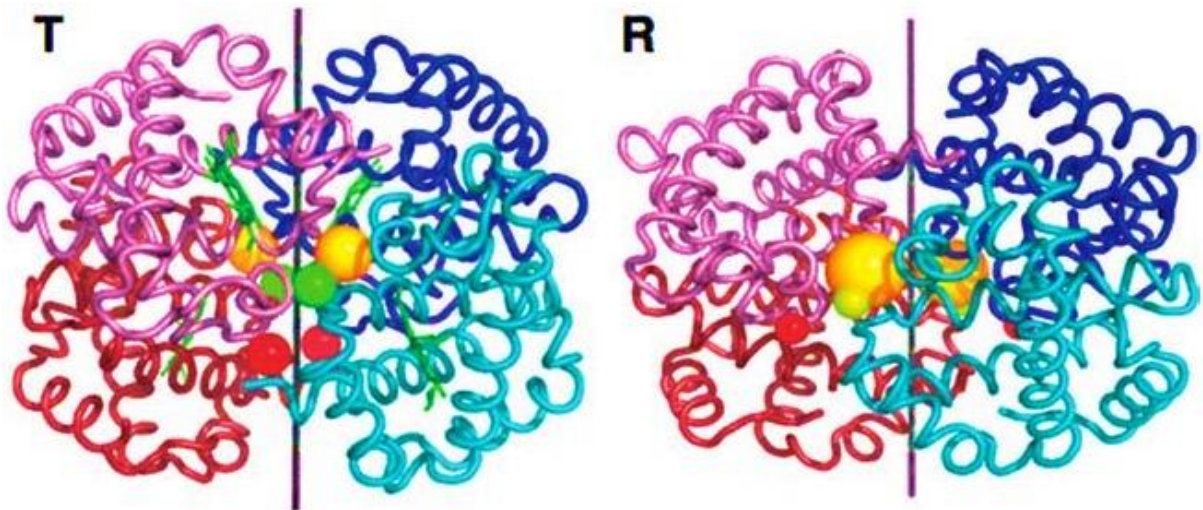


Figure 11 : Tétramère d'hémoglobine et ses différents états T et R à l'interface des dimères. (Dey S., Chakrabarti P, Janin J. 2011)

## 1.2. Génétique de l'hémoglobine

Chez l'homme plusieurs hémoglobines se succèdent au cours de la vie, et à tout moment, il en existe plusieurs simultanément. Ces hémoglobines se distinguent par la nature des sous-unités qui les constituent. (Grosveld, 1993).

➤ Hémoglobines embryonnaires : Durant la vie embryonnaire, deux chaînes de la famille  $\alpha$  coexistent :  $\zeta$  qui apparaît la première, puis  $\alpha$ . De même, il existe deux chaînes de type  $\beta$  :  $\epsilon$  spécifique à cette période initiale de la vie et les chaînes  $\gamma$  (ou fœtales). Ces diverses sous-unités permettent de réaliser les trois hémoglobines de l'embryon, l'Hb Gower 1 ( $\zeta 2\epsilon 2$ ), l'Hb Gower 2 ( $\zeta 2\epsilon 2$ ) et l'Hb Portland ( $\zeta 2\gamma 2$ ) (Wajcman, 2005).

L'hémoglobine embryonnaire est souvent symbolisée par l'Hb $\epsilon$ , qui ne doit pas être confondue avec l'hémoglobine E (hbE), qui est une mutation pathologique de l'HbA sur les sous-unités  $\beta$  (Zhenning, 2001)

➤ Hémoglobines Fœtales : L'hémoglobine fœtale (Hb F) de structure ( $\alpha 2 \gamma 2$ ) est détectable à partir de la 5<sup>ème</sup> semaine de vie intra-utérine. Parallèlement à cette modification de la nature des sous-unités de globine, il y a un changement du lieu où s'effectue l'érythropoïèse : sac vitellin dans la vie embryonnaire, puis foie et rate dans la vie fœtale et enfin moelle osseuse chez l'adulte.

L'Hb F est donc le constituant principal de la période foetale. Sa synthèse débute dès les stades précoces de la gestation et s'élève, entre les 8eme et 10eme semaines, à un taux de 90 %. Peu avant la naissance, entre les 32eme et 36eme semaines de gestation, les chaînes  $\gamma$  sont progressivement remplacées par les chaînes  $\beta$  de l'adulte (Wajcman, 2005).

Sur le plan phylogénétique et bioinformatique, la théorie de l'évolution postule que la majorité des mutations qui jalonnent l'histoire d'un gène sont neutres, c'est-à-dire qu'elles ne furent pas motivées par un avantage sélectif. L'occurrence de ces mutations et leur incorporation dans l'histoire d'un gène dans une espèce repose sur un phénomène appelé « dérive génétique » : la destinée d'un gène pour lequel existe une multitude d'allèles possibles est foncièrement aléatoire (Noirel J. 2010). Cette évolution conservée des différentes molécules Hb est le fruit d'une organisation moléculaire des gènes responsables de la synthèse de toutes ces structures. En effet, les gènes des chaînes de globine de l'hémoglobine sont regroupés en deux groupes (*clusters*) positionnés sur deux chromosomes différents (Figure 12) :

- les gènes du groupe  $\alpha$  codant pour les chaînes de type A sont sur le chromosome 16,
- ceux du groupe  $\beta$  codant pour les chaînes de type B sont sur le chromosome 11.

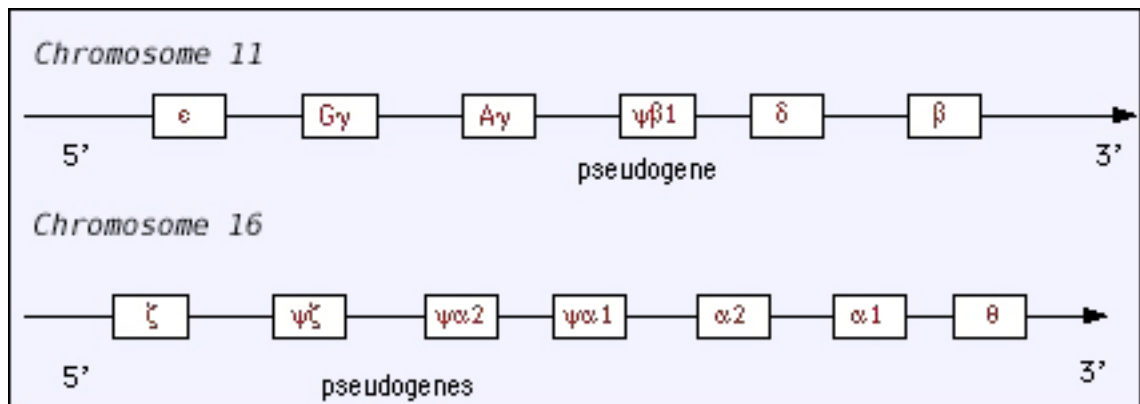


Figure 12: Structures schématiques des gènes regroupés en clusters sur les deux chromosomes 11 (11p 15,5) et 16 (16p 13,3).

Le groupe  $\alpha$  mesure 30 Kpb et le groupe  $\beta$  fait Kpb approximativement (Figure 13). Ce dernier compte, de plus, une région LCR (Locus Control Region).

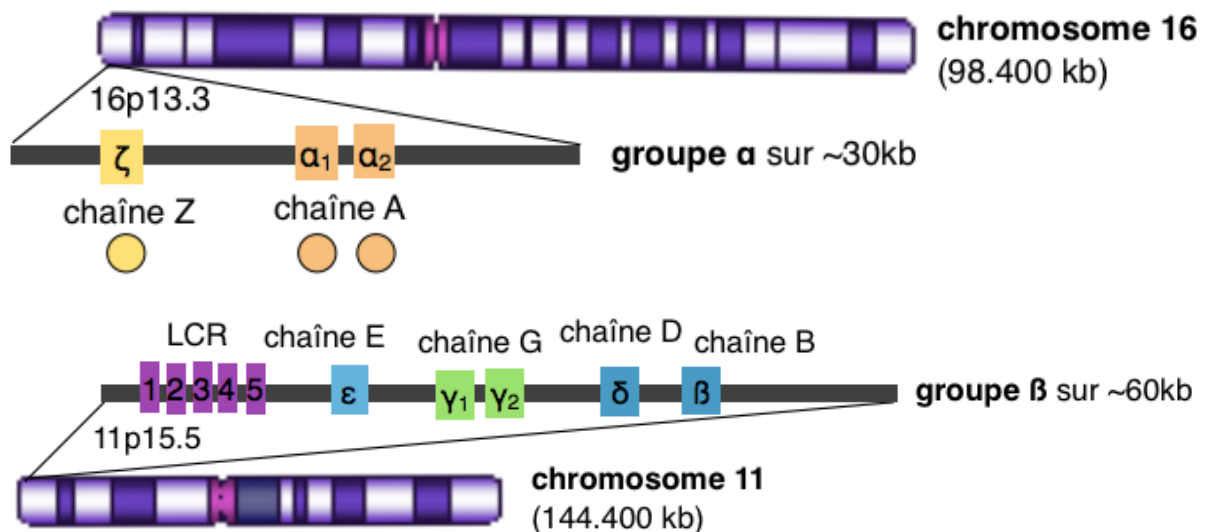


Figure 13: Disposition des clusters des différents gènes sur les chromosomes 11 et 16.

Les gènes de globine ont une structure similaire caractéristique des deux clusters avec trois exons et deux introns (avec des régions répétées en 3' et en 5' = *flanking regions*). Une différence concerne le deuxième intron (Figure 14) qui est beaucoup plus long que le premier dans la famille  $\beta$  (intron 1 = 120-140pb, intron 2 = 900pb) alors qu'ils sont du même ordre de grandeur, moins de 150 pb, dans la famille  $\alpha$ . Du fait de la présence de ce long exon les gènes de type  $\beta$  sont donc deux fois plus longs que les gènes du groupe  $\alpha$  (<http://pst.chez-alice.fr/protstfo.htm>).

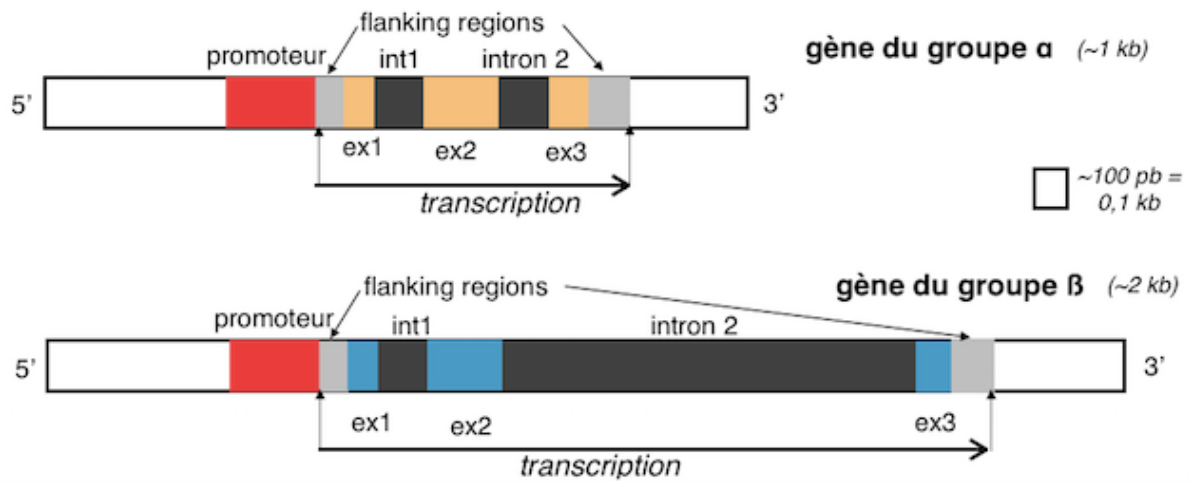


Figure 14: Structure des gènes de globine.

Les différentes séquences d'hémoglobines sont la conséquence de certaines anomalies liées à l'ADN. On distingue les hémoglobinopathies (polymorphisme des séquences d'ADN) et les thalassémies (anomalies de l'association des chaînes de globine) qui représentent un polymorphisme d'activité synthétique.

2. Hémoglobinopathies : Les hémoglobinopathies sont des maladies héréditaires génétiquement déterminées qui constituent un problème de santé publique dans de vastes parties du monde. Les hémoglobinopathies sont classiquement séparées en deux catégories :

- Soit un défaut qualitatif avec production en quantité normale d'une Hb « anormale »,
- Soit un défaut quantitatif de production de l'Hb normale, ce qui correspond à une thalassémie.

La différence est physiopathologique plus que génétique car les mêmes anomalies moléculaires sont le plus souvent rencontrées dans les deux pathologies (Couque.2016)

Les affections génétiques peuvent être résumées de la sorte :

- Production d'une hémoglobine de *structure* anormale : Mutants  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\gamma$  : Ce sont des variantes de l'hémoglobine.
- Défaut de production d'une chaîne de globine :  $\alpha$ ,  $\beta$  ou  $\delta$  : Thalassémies

2.1 Hémoglobine S : C'est la plus fréquente. Elle est majoritairement présente en Afrique (Bénin, Mali, Afrique centrale), aux Antilles et aux Etats-Unis, mais aussi chez les populations indiennes, arabes et méditerranéennes (Schmidt, 2012).

La drépanocytose est une maladie génétique de l'hémoglobine se transmettant sur un mode autosomique récessif. Elle résulte d'une mutation ponctuelle du 6<sup>ème</sup> codon du gène  $\beta$ -globine (chromosome 11). Cette mutation est responsable de la synthèse d'une hémoglobine anormale.

La polymérisation des molécules d'hémoglobine S mutées dans les globules rouges (GR) drépanocytaires est à l'origine d'une fragilisation et d'une perte de plasticité de ces hématies responsable de l'hémolyse et des phénomènes vaso-occlusifs (Mattionia, 2016)

2.2 Hémoglobine C : L'hémoglobine C (HbC) est l'une des mutations d'hémoglobine anormales les plus répandues dans le monde avec l'hémoglobine S et l'hémoglobine E. L'HbC est le deuxième plus répandu des hémoglobinopathies aux États-Unis et le troisième au monde. L'HbC est une variante structurelle de l'hémoglobine A normale (HbA) causée par une substitution des acides aminés de la lysine par l'acide glutamique en position six de la chaîne bêta de l'hémoglobine. (Schmidt, 2019).

L'hémoglobinose C homozygote (CC) est caractérisée par un syndrome anémique modéré, une splénomégalie, une lithiase biliaire. Le taux d'hémoglobine se situe entre 10 et 13 g/dL, il y a 100-200 G/L de réticulocytes, 50 à 90 % d'hématies en cible. A l'électrophorèse de l'Hb : HbC > 90 %, HbA2absente, HbF 3 à 8 %. Elle ne justifie aucun traitement.

Le double hétérozygotisme SC doit être considérés comme la drépanocytose homozygote SS, bien que la symptomatologie soit un peu moins sévère que chez les sujets SS. Parmi les complications, on note une perte progressive de la vision à l'âge adulte due à une rétinopathie. L'anémie est modérée, les réticulocytes = 100-200 G/L, il y a 30-50 % d'hématies en cible et de rares drépanocytes. A l'électrophorèse de l'hémoglobine : HbS 40-50 %, HbC 40-50 %, HbA absente ; HbF normale ou  $\pm$  augmentée (Aubry 2019).

2.3 Hémoglobine E : C'est la seconde hémoglobinopathie dans le monde après l'hémoglobinopathie S. Elle s'observe en Asie du sud-est, en particulier en Thaïlande, au Cambodge, au Laos où 35 % de la population est atteinte. Au total, 30 à 50 millions de personnes sont porteuses. Elle est due au remplacement d'un acide glutamique par une lysine. Comme la drépanocytose, l'hémoglobinopathie E confère une certaine résistance au paludisme.

L'hémoglobinoase E hétérozygote (AE) se caractérise par une microcytose sans anémie ou une pseudo polyglobulie microcytaire. Le taux d'hémoglobine est normal. Il y a plus de 10 % d'hématies en cible. A l'électrophorèse de l'hémoglobine : HbE 30-40 %, HbA 60-70 %.

L'hémoglobinoase E homozygote (EE) entraîne une microcytose avec anémie absente ou modérée. Le taux d'hémoglobine est de 12 à 15 g/dL, il y a plus de 30 % de globules rouges en cible. Le double hétérozygotisme HbE/ $\beta$  thal a un aspect de thalassémie majeure (Aubry, 2019).

### 3. Thalassémies

Les thalassémies constituent un groupe hétérogène d'anomalies moléculaires qui a en commun un défaut de synthèse d'une des chaînes de globine. Sur le plan physiopathologique, ces anomalies sont à l'origine d'un déséquilibre de synthèse avec accumulation de chaînes non appariées responsable de dysérythroïèse. Alors les anomalies les plus fréquentes :  $\alpha$  et  $\beta$  thalassémies (Couque, 2016).

3.1. Bêta-Thalassémies : La Bêta thalassémie se produit en raison de la synthèse insuffisante des réseaux de bêta-hémoglobine et d'un excès d'alpha réseaux. Il y a deux gènes sur le chromosome 11 qui sont exigés pour produire la bêta région du réseau d'hémoglobine, qui est hérité d'un parent. Le nombre de mutations géniques correspond à la gravité de la condition comme suit :

- Une mutation génique : signes modérés ou symptômes, désignés sous le nom du trait de mineur ou de thalassémie alpha de bêta thalassémie.
- Deux mutations géniques : modérées aux symptômes sévères, désignés sous le nom du commandant de bêta thalassémie ou de l'anémie de Cooley (Smith, 2015).

Les bêta-thalassémies, appelées aussi « maladies des globules rouges » se caractérisent par l'absence ou le défaut congénital de la synthèse de la chaîne  $\beta$  de l'hémoglobine. Donc dans ce cas, les chaînes bêta de l'hémoglobine sont produites en quantité insuffisante ou nulle,



ce qui provoque une production insuffisante d'hémoglobine globale. Seule la synthèse de l'HbA est entravée. Près de 200 allèles ont été décrits, concernant soit le gène de la chaîne  $\beta$ , soit, beaucoup plus rarement, des gènes régulateurs. Elles touchent chaque année 200 000 enfants à la naissance (Higgs et al., 2012). Elles se présentent sous trois formes cliniques différentes :

- La forme homozygote ou thalassémie majeure ou maladie de Cooley ;
- La forme hétérozygote ou thalassémie mineure ;
- La forme intermédiaire (Aubry, 2019).

Le type de thalassémie le plus courant et le plus grave est la bêta-thalassémie. Elle peut être mineure, intermédiaire ou majeure en fonction du nombre de gènes touchés. La forme la plus sévère est la forme majeure, également appelée anémie de Cooley, se traduit par une anémie marquée et commence le plus souvent à se manifester entre l'âge de 6 et 24 mois. Dans la forme intermédiaire, l'anémie est moins sévère et se manifeste plus tardivement. La forme mineure, ou maladie de Rietti-Greppi-Micheli, n'entraîne, elle, en général, quasiment aucun symptôme. Elle est le plus souvent découverte par hasard. (Jeff, 2018).

### 3.2. Alpha-Thalassémies

Dans les régions tropicales d'Afrique ou d'Asie, les  $\alpha$ -thalassémies sont extrêmement fréquentes car les porteurs de trait  $\alpha$ -thalassémique, partiellement résistant à l'infection, ont subi une pression de sélection par le paludisme (Couque, 2016).

Les  $\alpha$ -thalassémies correspondent à un défaut de synthèse des chaînes  $\alpha$  de globine de l'hémoglobine. Les  $\alpha$ -thalassémies sont dues dans la majorité des cas à des délétions, on peut distinguer deux types de délétions : les délétions de type 1 qui entraînent la perte des deux gènes HBA1 et HBA2 sur le même chromosome, les délétions de type 2 qui entraînent une perte de la partie intergénique entre HBA1 et HBA2 (Weatherall, 2001).

## Matériel et méthodes

## 1. Matériel :

### 1.1. Récupération des séquences d'hémoglobine a partir des banques de données

Des séquences de l'hémoglobine humaine ont été téléchargées à partir du portail informatique ExPASy<sup>5</sup> (Expert Protein Analysis System) et de PMut<sup>6</sup> qui est un outil bioinformatique en ligne pour l'annotation des mutations pathologiques sur les protéines.

Le portail ExPASy est un ensemble de ressources bioinformatiques SIB (Swiss Institute of Bioinformatics) qui donne accès à des bases de données scientifiques et à des outils logiciels (c'est-à-dire des ressources) dans différents domaines des sciences de la vie, notamment la protéomique, la génomique, la phylogénie, la biologie des systèmes, la génétique des populations, la transcriptomique, etc.

Les bases de données hébergées dans ce portail sont nombreuses (Enzyle, genomics, proteomics, Swiss2Dpage, etc.). Nous avons, pour la réalisation de notre travail, consulté le domaine de données Proteomics, hébergeant la base de données universelle UniprotKB<sup>7</sup> afin d'extraire les séquences qui intéressent notre travail.

La banque de données PDB<sup>8</sup> (Protein Data Bank) du RCSB (Research Collaboratory for Structural Bioinformatics) a été consultée pour télécharger les formes natives (naturelles) des molécules d'hémoglobines. C'est une collection mondiale de données sur la structure tridimensionnelle (ou structure 3D) de macromolécules protéiques (et même nucléiques) déterminées principalement par cristallographie et par spectroscopie RMN.

### 1.2. Prediction des substitutions entre résidus amino-acides :

PMut, quant à lui, il permet la prédiction rapide et précise (taux de réussite d'environ 80% chez l'homme) du caractère pathologique des mutations ponctuelles d'acides aminés en se basant sur l'utilisation de réseaux neuronaux (Intelligence Artificielle). Le programme permet également le balayage rapide des points chauds de mutation, qui sont obtenus par trois procédures : (1) balayage d'alanine, (2) mutation massive et (3) mutations génétiquement accessibles. Une interface graphique pour les structures Protein Data Bank (PDB) et une base de données contenant des profils de points chauds pour toutes les structures PDB non redondantes sont également accessibles à partir du serveur PMut.

---

<sup>5</sup> <https://www.expasy.org/>

<sup>6</sup> <http://mmb.irbbarcelona.org/PMut/>

<sup>7</sup> <https://www.uniprot.org/>

<sup>8</sup> [www.rcsb.org/](http://www.rcsb.org/)

## 2. Résultats :

### 2.1. Séquences de l'HbA :

Des séquences HbA native ont été importées à partir de la base de données UniprotKB. Le tableau 1 suivant résume ces données.

Tableau 1 : Caractéristiques des HbA1 et HbA2 collectées à partir de la base de données UniprotKB

Code	Taille	Séquence
P69905	142	>P69905_HbA1 MVLSPADKTNVKA AWGKVG AHAGEY- GAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVA- DALTNAVAHVDDMPNAL- SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT- PAVHASLDKFLASVSTVLTISKYR
D1MGQ2	142	>D1MGQ2_HbA2 MVLSPADKTNVKA AWGKVG AHAGEY- GAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVA- DALTNAVAHVDDMPNAL- SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT- PAVHASLDKFLASVSTVLTISKYR

Nous notons que les séquences natives de l'HbA comportent plusieurs variants naturels qui sont également reportées sur la base de données UniprotKB. Nous pouvons, par exemple citer le cas de :

- VAR\_002719 : est un variant de la séquence codée P69905 et où le résidu V (Val) est substitué par E (Glu) en position numéro 2.
- VAR\_002733 : est un autre variant de la séquence codée P69905 et où le résidu K (Lys) est substitué par M (Met) en position numéro 17.

Sur le site PMut, nous avons, en tant qu'utilisateurs, sélectionné le site de mutation (Position 51 sur la séquence protéique) et nous avons analysé une seule mutation (Faux-sens) en effectuant un balayage complet des mutations à cette position. Le programme va simuler une mutation ponctuelle massive tout au long de la séquence (analyse Mutation Hot-Spot),

aidant à détecter les régions où les mutations devraient avoir un impact pathologique important.

PMut nous a permis de prédire une mutation Faux-sens en position 51 qui concerne la substitution du résidu originaire H (His) par le résidu D (Asp). Ce choix de substitution est dicté par la nature physico-chimique très différente de ces deux résidus amino-acides. En effet, l'histidine est acide aminé qualifié de basique alors que l'acide aspartique est fortement acide. L'histidine présente un  $pK_a = 1,8$  ;  $pK_b = 9,2$  ;  $pK_r = 6,0$  ; une occurrence moyenne dans les protéines de 2,1% et une hydrophobicité du radical égale à  $-3,2$  ; alors que l'acide aspartique présente un  $pK_a = 2,1$  ;  $pK_b = 9,8$  ;  $pK_r = 3,9$  ; une occurrence moyenne dans les protéine de 5,5% et une hydrophobicité du radical égale à  $-3,5$  qui fait de lui un résidu plus hydrophile que l'histidine (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

Cette mutation provoquée *in silico*, en position 51 substituant l'histidine par l'acide aspartique est qualifiée de neutre par le scan de PMut avec un score de 0,04 et donc NON PATHOLOGIQUE pour l'homme.

Une deuxième mutation a été proposée cette fois-ci pour un autre acide aminé non basique. En effet une mutation en position 77 a concerné la substitution du résidu originaire M (Met) par le résidu A (Ala). Ce choix de substitution est dicté par la nature physico-chimique très similaire de ces deux résidus amino-acides. La méthionine présente un  $pK_a = 2,3$  ;  $pK_b = 9,2$  ; une occurrence moyenne dans les protéines de 1,7% et une hydrophobicité du radical égale à  $+1,9$  ; alors que l'alanine présente un  $pK_a = 2,3$  ;  $pK_b = 9,7$  ; une occurrence moyenne dans les protéines de 9,0% et une hydrophobicité du radical égale à  $+1,8$  qui fait de lui un résidu légèrement moins hydrophile que la méthionine (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

La mutation provoquée *in silico*, en position 77 substituant la méthionine par l'alanine est qualifiée elle aussi de neutre par le scan de PMut avec un score de 0,38 et donc NON PATHOLOGIQUE pour l'homme.

La troisième substitution a été provoquée en position N°5 et a concerné la proline qui a été mutée par la phénylalanine. Ce choix de substitution est dicté par la nature physico-

chimique très distincte de ces deux résidus amino-acides. La proline présente un  $pK_a = 2,0$  ;  $pK_b = 10,6$  ; une occurrence moyenne dans les protéines de 4,6% et une hydrophobicité du radical égale à -1,6 ; alors que la phénylalanine présente un  $pK_a = 1,8$  ;  $pK_b = 9,1$  ; une occurrence moyenne dans les protéines de 3,5% et une hydrophobicité du radical égale à +2,8 qui fait de lui un résidu nettement moins hydrophile que la proline (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

Cette mutation par substitution, en position 5 remplaçant la proline par la phénylalanine est qualifiée de PATHOLOGIQUE par le scan de PMut avec un score de 0,51 et donc PATHOLOGIQUE pour l'homme.

La substitution suivante a concerné encore une fois la proline (position N°5) mutée par la tyrosine. La proline présente un  $pK_a = 2,0$  ;  $pK_b = 10,6$  ; une occurrence moyenne dans les protéines de 4,6% et une hydrophobicité du radical égale à -1,6 ; alors que la tyrosine présente un  $pK_a = 2,2$  ;  $pK_b = 9,1$  ;  $pK_r = 10,1$  ; une occurrence moyenne dans les protéines de 3,5% et une hydrophobicité du radical égale à -1,3 qui fait de lui un résidu un peu moins hydrophile que la proline (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

La mutation, *in silico*, en position 5 substituant la proline par la tyrosine est qualifiée de neutre par le scan de PMut avec un score de 0,50 et donc NON PATHOLOGIQUE pour l'homme.

La dernière mutation provoquée *in silico* vise à substituer l'asparagine par le cystéine en position N° 69. L'asparagine présente un  $pK_a = 2,0$  ;  $pK_b = 8,8$  ; une occurrence moyenne dans les protéines de 4,4% et une hydrophobicité du radical égale à -3,5 ; alors que la cystéine présente un  $pK_a = 1,8$  ;  $pK_b = 10,8$  ;  $pK_r = 8,3$  ; une occurrence moyenne dans les protéines de 2,8% et une hydrophobicité du radical égale à +2,5 qui fait de lui un résidu nettement moins hydrophile que l'asparagine (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

Le scan de PMut a révélé une mutation PATHOLOGIQUE pour l'homme avec un score de 0,51.

## 2.2. Séquences de l'HbB :

Les séquences natives HbB ont été enregistrées à partir de la base de données UniprotKB. Les principales caractéristiques de ces données sont reportées sur le tableau suivant.

Tableau 2 : Caractéristiques des HbB collectées à partir de la base de données UniprotKB

Code	Taille	Séquence
P68871	147	> P68871_subunit beta MVHLTPEEKSAVTALWGKVVDEVGGEAL- GRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGA FSDGLAHLNLDKGTFFATLSEHCDKLLHVDPEN- FRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANA- LAHKYH
D9YZU5	147	>D9YZU5 subunit Beta MVHLTPEEKSAVTALWGKVVDEVGGEAL- GRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGA FSDGLAHLNLDKGTFFATLSEHCDKLLHVDPEN- FRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANA- LAHKYH
Q9UK54	128	> Q9UK54_subunit beta NVDEVGGEAL- GRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGA FSDGLAHLNLDKGTFFATLSEHCDKLLHVDPEN- FRLLGNVLVCVLDHFGKEFTPPVQAAYQKVVAGVANA- LAHKYH

Notons, à ce stade, que les séquences natives de HbB P68871 comportent plusieurs variants naturels qui sont également reportées sur la base de données UniprotKB. Nous pouvons, par exemple citer le cas de :

- VAR\_002907: est un variant de la séquence codée P68871 et où le résidu E (Glu) est substitué par K (Lys) en position numéro 27.
- VAR\_003056: est un autre variant de la séquence codée P68871 et où le résidu V (Val) est substitué par G (Gly) en position numéro 127.

Sur le site PMut, nous avons, sélectionné le site de mutation (Position 53 sur la séquence protéique) et nous avons analysé une seule mutation (Faux-sens) en effectuant un balayage complet (comme pour le cas de la sous-unité alpha) des mutations à cette position.

La première substitution a été provoquée en position N°53 et a concerné le remplacement de l'Acide Aspartique par la Méthionine. Ce choix de substitution est du à la nature physico-chimique très distincte de ces deux résidus amino-acides. L'Acide Aspartique présente un  $pK_a = 2,1$  ;  $pK_b = 9,8$  ;  $pK_r = 3,9$  ; une occurrence moyenne dans les protéines de 5,5% et une hydrophobicité du radical égale à -3,5 ; alors que la Méthionine présente un  $pK_a = 2,3$  ;  $pK_b = 9,2$  ; une occurrence moyenne dans les protéines de 1,7% et une hydrophobicité du radical égale à +1,9 qui fait de lui un résidu nettement moins hydrophile et plus hydrophobe que l'acide aspartique (<http://cyan1.grenet.fr/podcastmedia/PACES-2014-2015/UE1-BCH4-2a-aa-proprietes-pc.pdf>).

Après avoir scanné la séquence avec l'outil PMut, celui-ci a révélé une mutation PATHOLOGIQUE pour l'homme avec un score de 0,54.

Deux autres substitutions ont été réalisées également *in silico* sur la même position, donc touchant le même résidu Acide Aspartique N°53. Ces deux substitutions ont été réalisées en remplaçant l'Acide Aspartique par le Tryptophane d'une part, et par la Tyrosine d'autre part. Les caractéristiques physicochimiques de ces deux résidus ont permis ce choix substitutionnel. En effet, le Tryptophane a un  $pK_a = 2,4$  ;  $pK_b = 9,4$  ; une occurrence moyenne dans les protéines de



<sup>9</sup>1,1% et une hydrophobicité égale à -0,9. La Tyrosine, possède un pKa = 2,2 ; pKb = 9,1 ; pKr = 10,1 ; une occurrence moyenne dans les protéines de 3,5% et une hydrophobicité égale à -1,3.

Le scan PMut a révélé des mutations PATHOLOGIQUES pour ces deux substitutions.

Concernant la séquence d'hémoglobine bêta codée Q9UK54 sur UniprotKB, une mutation *in silico*, a été réalisée en position 34 où l'Acide Aspartique a été substitué par une Isoleucine. Nous avons choisi cette substitution à cause de la nature physico-chimique différente de ces deux résidus amino-acides. L'Acide Aspartique présente un pKa = 2,1 ; pKb = 9,8 ; pKr = 3,9 ; une occurrence moyenne dans les protéines de 5,5% et une hydrophobicité du radical égale à -3,5 ; alors que l'Isoleucine a un pKa = 2,4 ; pKb = 9,7 ; une occurrence moyenne dans les protéines de 4,6% et une hydrophobicité du radical égale à +4,5. L'Isoleucine étant nettement plus hydrophobe que l'Acide Aspartique.

Après avoir scanné la séquence grâce à l'outil PMut, il s'avère que cette substitution de Asp par Ile est PATHOLOGIQUE avec un score de 0,51.

### 2.3. Alignement multiple des séquences d'hémoglobine :

L'alignement multiple des alpha-globulines a été réalisé dans un but de comparer les séquences de la sous unité alpha humaine avec celles de trois autres vertébrés (RAT : *Rattus norvegicus*, Taureau : *Bos taurus* et la souris : *Mus musculus*) et de dresser une phylogénie comparative.

La structure primaire des séquences alpha-globulines (écrites sous format Fasta) est la suivante :

```
>Human_P69905
MVLSPADKTNVKAAWGKVG AHAGEYGAELERMFLSFPTTK-
TYFP HFDLSHGSAQVKGHGKKVADALTNAV AHVDDMPNAL-
```

```

SALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFT-
PAVHASLDKFLASVSTVLTSKYR
>MOUSE_P01942
MVLSGEDKSNIAAWGKIGGHGAEYGAELERMFAFPPTTKTYFPHFDVSH-
GSAQVKGHGKKVADALASAAGHLDDLPGAL-
SALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFT-
PAVHASLDKFLASVSTVLTSKYR
>BOVIN_P01966
MVLSAADKGNVKAAWGKVGGAHAEYGAELERMFLSFPTTK-
TYFPHFDLSHGSAQVKGHGAKVAAALTKAVEHLDDLPGAL-
SELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFT-
PAVHASLDKFLANVSTVLTSKYR
>RAT_P01946
MVLSADDKTNKNCWKGKIGGHGGGEYGEEALQRMFAAFPTTKTYFSHID-
VSPGSAQVKAHGKKVADALAKAADH-
VEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTPAM-
HASLDKFLASVSTVLTSKYR

```

Le programme informatique (logiciel) Clustal oméga a permis cet alignement. Le résultat est donné dans la figure suivante :

```

RAT      MVLSADDKTNKNCWKGKIGGHGGGEYGEEALQRMFAAFPTTKTYFSHIDVSPGSAQVKAHG 60
MOUSE    MVLSGEDKSNIAAWGKIGGHGAEYGAELERMFAFPPTTKTYFPHFDVSHGSAQVKGHG 60
Human    MVLSPADKTNVKAAWGKVGGAHAGEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 60
BOVIN    MVLSAADKGNVKAAWGKVGGAHAEYGAELERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 60
          ****  ** *:.* .***:*.**.* ** ***:*** :***** *:.*: * ***.**
          .**

RAT      KKVADALAKAADHVEDLPGALSTLSDLHAHKLRVDPVNFKFLSHCLLVTLACHHPGDFTP 120
MOUSE    KKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTP 120
Human    KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 120
BOVIN    AKVAAALTKAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTP 120
          *** **:*.* .*:.*:.* ** *****:***.*****.* *.:***

RAT      AMHASLDKFLASVSTVLTSKYR      142
MOUSE    AVHASLDKFLASVSTVLTSKYR      142
Human    AVHASLDKFLASVSTVLTSKYR      142
BOVIN    AVHASLDKFLANVSTVLTSKYR      142
          *:*****.*****

```

Figure 15 : Alignement multiple des séquences alpha-globuline par le logiciel Clustal oméga.

– Le symbole (\*) = une identité parfaite des acides aminés des individus.

- Le symbole (:) = une similarité entre les résidus similaires physico-chimiquement
- Le symbole (.) = une substitution non conservatrice ; résidus non similaires physico-chimiquement.

Un motif protéique commun (positions : 92 – 141) a été obtenu avec l’outil de la base de données Prosite du portail ExPASy qualifiant une zone fortement similaire à toutes les séquences analysées :

R-V-D-P-V-N-F-K-x(0,1)-L-x(0,1)-S-H-[CS]-L-L-V-T-L-A-[ACS]-H-x-P-[AGS]-[DE]-F-T-P-A-[MV]-H-A-S-L-D-K-F-L-A-[NS]-V-S-T-V-L-T-S-K-Y-R

L’alignement multiple des séquences a été suivi par le calcul d’une matrice du pourcentage d’identité (Id%) réalisé par l’outil Clustal 2.1. Les résultats obtenus sont reportés dans la matrice suivante :

Tableau 3 : Matrice de Id% des séquences alpha-globulines

	Rat	Souris	Homme	Taureau
Rat				
Souris	84,51			
Homme	78,17	85,92		
Taureau	77,46	86,62	88,03	

L’identité entre résidus amino-acides la plus importante a été observée entre la séquence de l’homme et celle du taureau avec un score de 88,03% ; alors que la plus basse est observée entre la séquence du rat et celle du taureau avec un score égal à 77,46% ; mais cette valeur reste extrêmement élevée et révèle le fort lien entre les structures primaires des séquences alignées et donc leur appartenance à une même famille protéique d’alpha-globulines.

L’alignement multiple a révélé que la position N°5 ne comporte pas le même acide aminé chez tous les organismes testés. En effet, chez l’homme, nous retrouvons la Proline (dont les mutations par substitutions avec la Phénylalanine et avec la Cystéine ont causé des pathologies avec des scores PMut de 0,51 chacune) ; cependant, chez le rat et le taureau, nous

---

constatons une Alanine alors que la séquence de la souris contient une Glycine à cette même position.

Les mutations observées dans la séquence de la souris (P01942) concernent les substitutions dans les sites :

- A27Y : le résidu Alanine en position 27 est substitué par la Tyrosine
- M33A : le résidu Méthionine en position 33 est substitué par une Alanine
- K61A : le résidu Lysine en position 61 est substitué par une Alanine

Pour le cas du taureau, la séquence P01966 a révélé les mutations suivantes :

- G58C
- S82A
- N132S

En fin, la séquence P01946 du rat a montré les substitutions suivantes :

- G52A
- K91A
- F99A

Nous constatons que le site N°5 de la Proline chez l'homme n'est pas du tout concerné par d'éventuelles mutations chez les trois vertébrés (souris, rat et taureau) analysés au cours de cette étude. Cependant, remarquons également que des mutations par substitutions peuvent être observées chez ces trois vertébrés dans des sites différents de celui de l'homme et concernant des résidus autres que la Proline.

Pour le cas de la deuxième mutation observée chez l'homme en position 69 qui a concerné la substitution de l'Asparagine par une Alanine avec un score de 0,51 ; aucune mutation à ce site n'a été observée également chez les trois sujets vertébrés. En effet en position N°69 le rat et le taureau possèdent dans leurs séquences une Lysine alors que la souris contient dans cette même position une Sérine.

L'arbre phylogénétique dérivant de cet alignement multiple est représenté par la figure suivante :

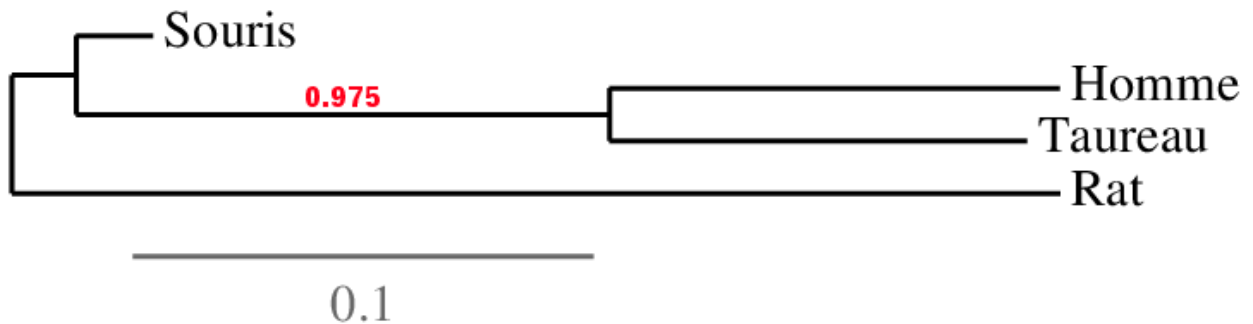


Figure 16 : Arbre phylogénétique des séquences alpha-globulines. Réalisé sur <http://phylogeny.lirmm.fr/> (Dereeper A. et al ,2008)

Le dendrogramme montre une forte similarité globale entre les séquences humaine et bovine (Id% = 88,03). La souris est plus similaire à la paire (clade ou cluster) formée par l'homme et le taureau mais le rat aurait une séquence plus éloignée ; il peut faire office de racine d'arbre ou de groupe externe (Out group).

Les séquences bêta-hémoglobine humaine et certains vertébrés (Rat, Souris et Taureau) ont fait l'objet d'alignement multiple avec l'outil Clustal Oméga. Les structures primaires de ces séquences (au format Fasta) sont les suivantes :

>P68871\_Homme

MVHLTPEEKSAVTALWGKVVNDEVGGEAL-  
GRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLNDLKG  
TFATLSELHCDKLHVDPENFRLGNVLCVLAHFFGKEFTPPVQAAYQKVVAGVA-  
NALAHKYH

>P02091\_RAT

MVHLTDAEKAAVNGLWGKVNPDVVGGEALGRLLVVYPWTQRYFDSFGDLSSA-  
SAIMGNPKVKAHGKKVINAFAFNDGLKHLNDLKGTFASLSELHCDKLHVDPEN-  
FRLGNMIVIVLGHHLG KEFTPCAQAQAFQKVVAGVASALAHKYH

>P02088\_Souris

MVHLTDAEKAAVSCLWGKVNSEVVGGEALGRLLVVYPWTQRYFDSFGDLSSA-  
SAIMGNAKVKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPEN-  
FRLGNMIVIVLGHHLG KDFTPAQAQAFQKVVAGVATALAHKYH

>P02070\_Taureau

MLTAEKAAVTAFWGKVKVDEVGGEAL-

GRLLVVYPWTQRFFESFGDLSTADAVMNNPKVKAHGKVKVLDSEFSNGMKHLDDLK

TFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFGKE

FTPVL-

QADFQKVVAGVANALAHRYH

Le résultat de l'alignement multiple de ces séquences a donné :

```

P02091_RAT      MVHLTDAEKAAVNGLWGKVPDDVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNPK      60
P02088_Souris  MVHLTDAEKAAVSCLWGKVNSEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNAK      60
P68871_Homme   MVHLTPEEKSAVTALWGKVNDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK      60
P02070_Taureau --MLTAEKAAVTAFWGKVKVDEVGGEALGRLLVVYPWTQRFFESFGDLSTADAVMNNPK      58
                **  **:* . :****: *:*****:*****: .*:.* *

P02091_RAT      VKAHGKVINAFNDGLKHLNPKGTFAHLSELHCDKLHVDPENFRLLGNMIVIVLGHHLG      120
P02088_Souris  VKAHGKVVITAFNDGLNHLDSLKGTFAHLSELHCDKLHVDPENFRLLGNMIVIVLGHHLG      120
P68871_Homme   VKAHGKVLGAFSDGLAHLNPKGTFAHLSELHCDKLHVDPENFRLLGNLVCVLAHHFG      120
P02070_Taureau VKAHGKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVVLARNFG      118
                *****: :*.:*: **.****** *****:*****:* **.:*:

P02091_RAT      KEFTPCAQAQAFQKVVAGVASALAHKYH 147
P02088_Souris  KDFTPAAQAQAFQKVVAGVATALAHKYH 147
P68871_Homme   KEFTPPVQAAYQKVVAGVANALAHKYH 147
P02070_Taureau KEFTPVLQADFQKVVAGVANALAHRYH 145
                *:* ** :*****.****:*

```

Figure 17 : Alignement multiple des séquences bêta-globuline par le logiciel Clustal oméga.

- Le symbole (\*) = une identité parfaite des acides aminés des individus.
- Le symbole (:) = une similarité entre les résidus similaires physico-chimiquement
- Le symbole (.) = une substitution non conservatrice ; résidus non similaires physico-chimiquement.

Le calcul d'une matrice du pourcentage d'identité (Id%) réalisé par l'outil Clustal 2.1 a été effectué. Les résultats obtenus sont reportés dans la matrice suivante :

Tableau 4 : Matrice de Id% des séquences bêta-globulines

	Rat	Souris	Homme	Taureau
Rat				
Souris	91,84			
Homme	81,63	80,27		
Taureau	75,17	73,79	84,14	

L'identité entre résidus amino-acides la plus importante a été observée entre la séquence de la souris et celle du rat avec un score de 91,84% ; alors que la plus basse est observée entre la séquence du taureau et celle de la souris avec un score égal à 73,79% ;

Ces valeurs restent extrêmement élevées et révèlent le fort lien entre les structures primaires des séquences alignées et donc leur appartenance à une même famille protéique de bêta-globulines.

L'alignement multiple a révélé que la position N°53 ne comporte pas le même acide aminé chez tous les organismes testés. En effet, chez l'homme, nous retrouvons l'Acide Aspartique (dont les mutations par substitutions avec la Méthionine, avec le Tryptophane et avec la Tyrosine ont causé des pathologies avec des scores PMut de 0,54 chacune) ; cependant, chez le rat et la souris, nous constatons une Sérine à cette même position. La séquence bêta-globuline du taureau contient un Acide Aspartique comme dans la séquence humaine.

L'arbre phylogénétique dérivant de cet alignement multiple est représenté par la figure suivante :

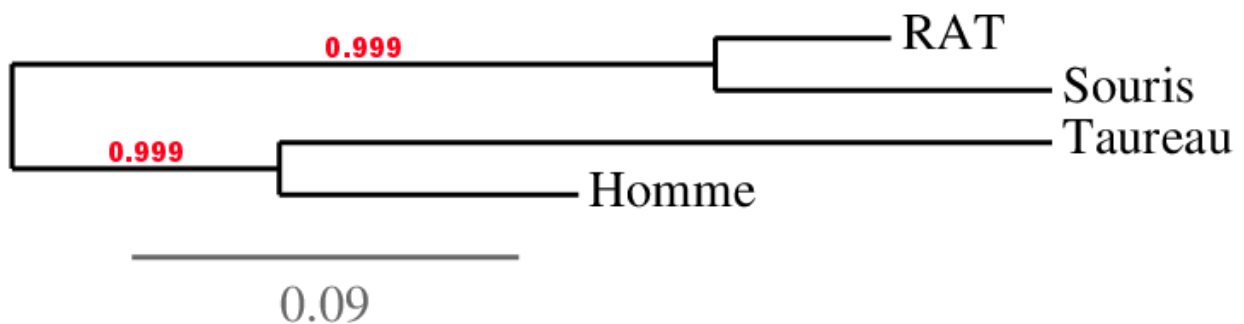


Figure 18 : Arbre phylogénétique des séquences bêta-globulines. Réalisé sur <http://phylogeny.lirmm.fr/> (Dereeper A. et al, 2008)

L'arbre phylogénétique montre une forte similarité globale entre la séquence du rat et celle de la souris (Id% = 91,84). L'homme est plus similaire au taureau.

#### 2.4.Blast

La séquence d'alpha-globuline codée P69905 présentait deux mutations possibles qui sont pathologiques pour l'homme selon le scan effectué par PMut. La première mutation proposée par PMut était P5F (substitution de la Proline en position 5 par une Phénylalanine).

Suite à cette proposition *in silico*, nous avons effectué un PSI-Blast sur le site de NCBI afin de vérifier si cette protéine (mutée en position N°5) existe dans la base de données. Le résultat du PSI-Blast a révélé une séquence qui possède une Phénylalanine en 5<sup>ème</sup> position codée 1BAB\_A contenant 143 acides aminés.

La deuxième mutation se situe en position N°69, dans laquelle un résidu Asparagine est substitué par une Cystéine. Le résultat du PSI-Blast a révélé une absence de cette séquence mutée *in silico* en position 69 pour laquelle le résidu Asparagine était substitué par une Cystéine.

L'analyse moléculaire de ces substitutions a été effectuée également avec le logiciel PyMol pour visualiser en 3D les mutations des résidus concernés par cette étude. En effet, une molécule native d'Alpha-hémoglobine a été téléchargée à partir de la banque de données internationale PDB sous le code 1A00. Sa structure 3D sur PyMol est la suivante après avoir caché les molécules d'eau et coloré les quatre sous-unités différemment :

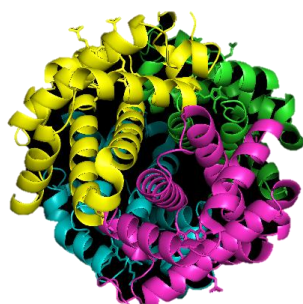


Figure 19 : Représentation 3D de l'Alpha-hémoglobine native (Code 1A00) sur PyMol.

La disposition naturelle du résidu Asparagine est représentée dans la figure suivante après avoir effectué un zoom sur ce dernier :



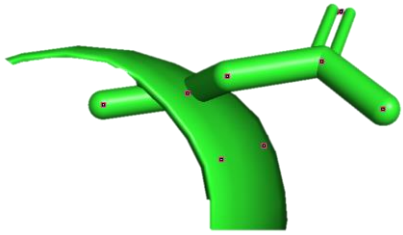


Figure 20: Position du résidu N69 sur la molécule 1A00.

La substitution du résidu N par une Cystéine (C) est représentée comme suit :



Figure 21 : Position du résidu C69 sur la molécule 1A00 après avoir substitué le résidu N69.

Dans ce cas de la bêta-globuline, nous avons obtenu trois mutations possibles du résidu Acide Aspartique (D) en position N°53 par la Méthionine (M) ou par le Tryptophane (W) ou enfin par la Tyrosine (Y).

La position native du résidu D53 est reportée sur la figure suivante :

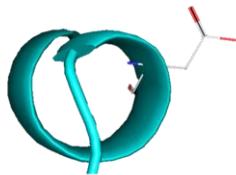


Figure 22 : Position naturelle du résidu D53 sur la sous-unité bêta de l'Hb.

Dans le cas de la substitution D53W, sur le logiciel PyMol nous avons obtenu la figure suivante :

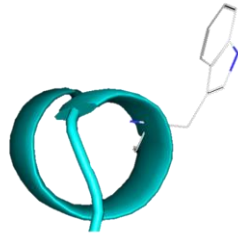


Figure 23 : Position naturelle du résidu W53 sur la sous-unité bêta de l'Hb après avoir substitué le résidu naturel D53.

## Discussion

### Partie 3 : résultat et discussion

Parmi les hémoglobinopathies, nous pouvons distinguer deux pathologies différentes :

- Des anomalies de structure, une hémoglobine anormale est présente, entraînant ou non des signes fonctionnels. Responsables des drépanocytoses dues à une seule mutation.
- Des anomalies de synthèse provoquant des thalassémies dont l'anémie qui se situe au premier plan. Des mutations multiples sont observées dans les différentes anémies.

La lecture des éventuelles hémoglobinopathies repose sur des renseignements fins et précis dont la structure moléculaire du tétramère et l'organisation génique y compris les régions régulatrices. En fait, il existe plusieurs variants protéiques (silencieux ou pathologiques) des chaînes polypeptidiques formant la molécule de l'hémoglobine (soit les chaînes alpha, soit les chaînes bêta) et les différentes localisations possibles d'une mutation se manifestent par des signes hématologiques spécifiques (Lapie D. et Elion J., 2005)

Les mutations par substitutions proposées *in silico* dans ce travail ont concerné les deux chaînes alpha et bêta de l'hémoglobine. Certaines substitutions ont été supposées par le scan de PMut être pathologiques pour l'homme. En effet la substitution N69C a apporté des changements conformationnels dans la structure 3D de la molécule, changements qui pourraient se montrer sévères par rapport à la fonction de la molécule. En effet, toute molécule protéique n'est fonctionnelle que pour une conformation confortable énergétiquement donnée. Si une conformation est introduite à la molécule protéique par changement des propriétés du milieu telles que le changement du pH, de la pression osmotique ou de la température peuvent entraîner une modification plus ou moins importante vis à vis de la fonction naturelle de la protéine cible. De plus, il est noté sur la matrice de substitutions Blosum62 que cette mutation est affectée d'un score égal à -3, une valeur fortement négative révélant une nette différence des propriétés physico-chimiques des deux résidus et donc une contribution totalement différente par rapport à l'activité protéique.

Sur le site NCBI, en consultant la table des substitutions<sup>10</sup> de l'Asparagine (N) sur la base de la matrice Blosum62, nous notons également la différence dans le potentiel H-Bonds de ces deux résidus. En effet, il est égal à 5 pour le résidu N et est nul (égal à zéro) pour le

---

<sup>10</sup> [https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa\\_explorer.cgi?mode=table&key=blosum&self=N](https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi?mode=table&key=blosum&self=N)

résidu C. Cette nette différence est synonyme d'interactions différentes avec le milieu de la protéine.

La substitution apportée sur la sous-unité bêta en position N°53 du résidu D par l'acide aminé W a montré sur PyMol une différence nette des deux structures. Ces changements conformationnels auraient des effets négatifs vis à vis de la structure et donc de la fonction de la protéine mutée. Sur la matrice de substitutions Blosum62, cette mutation montre un score égal à -4, une valeur fortement négative exprimant une forte différence des propriétés physico-chimiques des deux résidus et donc un rôle totalement différent par rapport à l'activité protéique.

Sur le site NCBI, en consultant la table des substitutions<sup>11</sup> de l'Aspartate (D) sur la base de la matrice Blosum62, nous notons également la différence dans le potentiel H-Bonds de ces deux résidus. En effet, il est égal à 4 pour le résidu D et égal à 1 pour le résidu W. Cette différence est synonyme d'interactions différentes avec le milieu de la protéine pour les deux résidus.

Des travaux de prédiction des mutations d'acides aminés *in silico* similaires à notre étude ont été effectués sur un variant d'une hémoglobine Charlieu par Joly P. *et al.* (2010) où une substitution a été observée sur le codon 106 du gène alpha1 (CTC > CCG), sans aucune expression phénotypique mais associée à une forte microcytose. (Joly P et all, 2010).

D'autres auteurs tels que Marden MC *et al.* (2007) ont substitué ce même résidu amino-acide chez 282 espèces animales *in silico* et noté qu'il s'agit d'une position importante pour la structure de la molécule (Marden M.C *et al.* 2007).

---

<sup>11</sup> [https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa\\_explorer.cgi?mode=table&key=blosum&self=N](https://www.ncbi.nlm.nih.gov/Class/Structure/aa/aa_explorer.cgi?mode=table&key=blosum&self=N)

## Conclusion

## Conclusion

L'étude des hémoglobinopathies ont toujours été portées sur des cas observés cliniquement avec des phénotypes irréguliers significatifs de pathologie. Les analyses bioinformatiques permettent de prédire ces accidents par étude des éventuelles substitutions par mutations ponctuelles en se basant sur la structure primaire de la protéine cible.

Dans notre étude, nous sommes arrivés à proposer, *in silico*, des substitutions de certains résidus par d'autres telles que la mutation ponctuelle de l'asparagine en position 69 par une cystéine. Cette substitution a provoqué une modification importante quant à la conformation tridimensionnelle de la protéine d'hémoglobine.

Cette étude a été portée sur un effectif de séquences protéiques réduit, mais peut être confirmée sur des échantillons plus importants afin de corroborer les résultats de tous nos prédictions.

Ainsi, en termes de perspectives, nous préconisons une modélisation bioinformatique par apprentissage automatique afin de proposer un modèle prédictif plus crédible permettant de détecter les substitutions les plus attendues surtout sur le plan thermodynamique.

## Références bibliographiques



## Références bibliographiques

Aubry p, Gaüzère B. 2019. Hémoglobinoses. Centre René La busquière, Institut de Médecine Tropicale, Université de Bordeaux, 33076 Bordeaux (France).

Aubry P, Gaüzère B.2020.Thalassémies.

Chang J-G et all. 1998. Hb Koln [b98(FG5) Val → Met : the first case found in a Chinese family. Hemoglobin. Vol. 22, (5 & 6), pp. 535-536).

Couque N. 2013. Diagnostic d'une hémoglobinopathie. Feuillet de Biologie. 311 : 5-18.

Couque N, Trawinski E, and Elion J.2016.Génétique Des Maladies de L'hémoglobine. Revue Francophone Des Laboratoires, 2016.481 ,49–60.

Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008 Jul 1;36 (Web Server issue):W465-9. Epub 2008 Apr 19.

Dey S., Chakrabarti P, Janin J. 2011. A survey of hemoglobin quaternary structures. Proteins: Structure, Function and Bioinformatics. 79 (10) : 2861-2870.)

Dreyfus B et all. 1992.Hématologie Medecine-Sciences, Flammarion.

Dybas J et all.2020. «Probing the Structure-Function Relationship of Hemoglobin in Living Human Red Blood Cells», Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 239 ,118530.

Grosveld F, Dillon N, Higgs D. 1993. The regulation of human globin gene expression, in the Hemoglobinopathies. Bailliere's editor.

Higgs DR, Engel JD.2012. Stamatoyannopoulos G. « Thalassemia » Lancet ; 379 :373-83.

Jeff.2018. Thalassémie - Symptômes et traitement. Le Journal des Femmes Santé.

Joly P., Szymanowicz A., Neyron M. J., Zine A., Francia A. 2010. Hb Charlieu [ $\alpha$ 1 Cd106 (G13) Leu > Pro] : A new alpha-thalassemic hemoglobin variant found during an unexplained microcytosis. Immuno-analyse et biologie spécialisée. 25 : 104-109

Labie D. et Elion J., 2005. Molecular and pathophysiological bases of haemoglobin diseases. EMC-Hématologie 2 : 220–239.

- Marden M. C., Dewilde S., Kiger L. Hamdane D., Uzan J., Burmester T. et al. 2007. Exploiting a list of protein sequences. *Gene*. 398: 35-41.
- Mattioni S et all. 2016. La Drépanocytose En France *Revue Francophone Des Laboratoires*, 2016.481, 61–66.
- Noirel J. 2010. Évolution in silico des protéines monomériques et dimériques. Thèse Doctorat. 272 pg.
- Schmidt Kayla L.2019.Tim R Randolph, «Development of a Microscopic Method to Identify Hemoglobin C Conditions for Use in Developing Countries, American Society for Clinical Laboratory Science, ascls.119.001412.
- Schmidt M. 2012. Complémentarité des techniques d'électrophorèse capillaire et de CLHP dans le diagnostic des hémoglobinopathies. Mémoire pour le Diplôme d'études spécialisées de Biologie Médicale. Université de Lille 2.
- Shaanan, B. 1983. Structure of human oxyhaemoglobin at 2.1 resolution. *J. Mol. Biol.* 171: 31).
- Wajcman H.2005. Hémoglobines : structure et fonction. EMC-Hématologie. 2(3) : 145-157.
- Wajcman H.2013.Hémoglobines : structure et fonction .Hôpital Henri-Mondor ,51avenue du Maréchal-De-Lattre-de-Tassigny, 94010 Créteil cedex, France.
- Weatherall DJ.2001. Phenotype-genotype relationships in monogenic disease: lessons from the thalassemiases», *Nat Rev Genet.*;2(4):245-55.
- Zaher Y. 2011. Hémoglobines instables : de la physiopathologie à la thérapeutique. Thèse de Doctorat en Pharmacie. Université Mohammed V. 157pg.
- Zhenning He et J. Eric Russell.2001. Expression, purification, and characterization of human hemoglobins Gower-1 ( $\zeta 2\epsilon 2$ ), Gower-2 ( $\alpha 2\epsilon 2$ ), and Portland-2 ( $\zeta 2\beta 2$ ) assembled in complex transgenic–knockout mice, *Blood*, vol. 97: 1099-1105.

## Résumé

## Résumé

L'objectif principal de cette étude bioinformatique des molécules d'hémoglobine humaine est de prédire les éventuelles mutations *in silico* par substitutions de certains résidus amino-acides de la séquence native (non mutée) et de mesurer l'effet de cette mutation ponctuelle sur l'activité de la protéine d'hémoglobine par comparaison aux mutations de certains vertébrés choisis dans cette étude (Taureau, rat et souris). Des résultats significatifs ont été observés après avoir scanné les séquences substituées avec l'outil bioinformatique PMut. Ainsi, une première substitution du résidu N par une cystéine et une deuxième substitution du résidu D par un tryptophane ont causé des modifications conformationnelles importantes dans la structure 3D observée sur le logiciel PyMol.

Mots clés : Hémoglobine – variant – Alignement – Structure 3D

### Abstract

The main aim of this bioinformatics study of human hemoglobin molecules is to predict possible in silico mutations by replacement of certain amino acid residues in the native (non-mutated) sequence and to measure the effect of this point mutation on the activity of the hemoglobin protein compared to mutations in certain vertebrates selected in this study (Taurus, rat and mouse). Significant results were observed after scanning the substituted sequences with the PMut bioinformatics tool. Thus, a first substitution of the N residue by a cysteine and a second substitution of the D residue by a tryptophan caused significant conformational modifications in the 3D structure observed on the PyMol software.

Keywords: Hemoglobin - variant - Alignment - 3D structure

### ملخص

الهدف الرئيسي من هذه الدراسة المعلوماتية الحيوية لجزيئات الهيموجلوبين البشرية هو التنبؤ بالطفرات المحتملة في السيليكو عن طريق استبدال بعض بقايا الأحماض الأمينية من التسلسل الأصلي (غير المتحور) وقياس تأثير هذه الطفرة النقطية على نشاط بروتين الهيموجلوبين مقارنة بطفرات بعض الفقاريات المختارة في هذه الدراسة (الثور والجرذ والفأر). وقد لوحظت نتائج مهمة بعد مسح التسلسلات المستبدلة باستخدام أداة المعلوماتية الحيوية PMut. وبالتالي، فإن الاستبدال الأول لبقايا N بواسطة السيستين والاستبدال الثاني لبقايا D بواسطة التربتوفان تسبب في تعديلات مطابقة مهمة في الهيكل ثلاثي الأبعاد الذي لوحظ في برنامج PyMol .

الكلمات المفتاحية: الهيموجلوبين - البديل - المحاذاة - الهيكل ثلاثي الأبعاد

Date de soutenance :	Rédigé par : ziadi yasmine Sekehal amira
<p style="text-align: center;">Thème :</p> <p style="text-align: center;"><b>Analyse bioinformatique des variants de l'hémoglobine humaine</b></p>	
<p style="text-align: center;">Mémoire de fin de cycle pour l'obtention du diplôme de Master en BIOCHIMIE DE LA NUTRITION</p>	
<p>L'objectif principal de cette étude bioinformatique des molécules d'hémoglobine humaine est de prédire les éventuelles mutations <i>in silico</i> par substitutions de certains résidus amino-acides de la séquence native (non mutée) et de mesurer l'effet de cette mutation ponctuelle sur l'activité de la protéine d'hémoglobine par comparaison aux mutations de certains vertébrés choisis dans cette étude (Taureau, rat et souris). Des résultats significatifs ont été observés après avoir scanné les séquences substituées avec l'outil bioinformatique PMut. Ainsi, une première substitution du résidu N par une cystéine et une deuxième substitution du résidu D par un tryptophane ont causé des modifications conformationnelles importantes dans la structure 3D observée sur le logiciel PyMol.</p>	
<p>Mots clés : Hémoglobine – variant – Alignement – Structure 3D</p>	
<p>Jury d'évaluation :</p> <p>Président du jury : Pr NECIB Y.</p> <p>Rapporteur : Pr HAMIDECHI M. A.</p> <p>Examineur : Dr MOKRANI M.</p>	
<p style="text-align: center;">Année universitaire : 2020-2021</p>	